Anal. Chem. **2007,** *79,* 89−97

Monitoring Diet Effects via Biofluids and Their Implications for Metabolomics Studies

Haiwei Gu,[†] Huanwen Chen,[‡] Zhengzheng Pan,[‡] Ayanna U. Jackson,[‡] Nari Talaty,[‡] Bowei Xi,[§] Candice Kissinger,[⊥] Chester Duda,[⊥] Doug Mann,[⊥] Daniel Raftery,*[,‡] and R. Graham Cooks*[,‡]

Department of Physics, Department of Chemistry, and Department of Statistics, Purdue University, West Lafayette, Indiana 47907, and Bioanalytical Systems, Inc., 2701 Kent Avenue, West Lafayette, Indiana 47906

The effect of diet on metabolites found in rat urine samples has been investigated using nuclear magnetic resonance (NMR) and a new ambient ionization mass spectrometry experiment, extractive electrospray ionization mass spectrometry (EESI-MS). Urine samples from rats with three different dietary regimens were readily distinguished using multivariate statistical analysis on metabolites detected by NMR and MS. To observe the effect of diet on metabolic pathways, metabolites related to specific pathways were also investigated using multivariate statistical analysis. Discrimination is increased by making observations on restricted compound sets. Changes in diet at 24-h intervals led to predictable changes in the spectral data. Principal component analysis was used to separate the rats into groups according to their different dietary regimens using the full NMR, EESI-MS data or restricted sets of peaks in the mass spectra corresponding only to metabolites found in the urea cycle and metabolism of amino groups pathway. By contrast, multivariate analysis of variance from the score plots showed that metabolites of purine metabolism obscure the classification relative to the full metabolite set. These results suggest that it may be possible to reduce the number of statistical variables used by monitoring the biochemical variability of particular pathways. It should also be possible by this procedure to reduce the effect of diet in the biofluid samples for such purposes as disease detection.

The field of metabolomics[1] is becoming increasingly important in the understanding of biological processes given that metabolic changes are observed in a variety of diseases, such as diabetes,[2] breast cancer,[3] coronary heart disease,[4] and inborn diseases.[5,6] One rapidly growing area of metabolomics is its application to nutritional research, including the understanding of individual variability in nutritional requirements and the development of individually customized nutrition.[7,8] Mass spectrometry (MS) and nuclear magnetic resonance (NMR) have proven to be powerful analytical tools in metabolomics analysis, especially when combined with multivariate statistical analyses.[9,10] A few studies to date have been performed that combine both analytical techniques with statistical analysis.[6,11] Together, these methods can be used to distinguish the effects of diet, age, gender, diurinal variation, and genetic strain on the metabolic profile of biofluid samples.[12−15]

Due to their high sensitivity, high-throughput characteristics and applicability to complex mixtures, mass spectrometric methods have demonstrated promise in metabolomics.[16−18] For example, we recently demonstrated that desorption electrospray ionization (DESI) MS can be used effectively for metabolomics research with little or no sample preparation.[6,19] DESI is an ambient ionization technique that combines features of electrospray ionization (ESI) and desorption ionization to permit analysis directly from a surface with virtually no sample preparation.[20] This

* To whom correspondence should be addressed. (Raftery) Tel: (765) 494-6070. Fax: (765) 494-6070. E-mail: raftery@purdue.edu. (Cooks) Tel: (765) 494-5262. Fax: (765) 494-9421. Email: cooks@purdue.edu.

[†] Department of Physics, Purdue University.

[‡] Department of Chemistry, Purdue University.

[§] Department of Statistics, Purdue University.

[⊥] Bioanalytical Systems, Inc.

(1) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nat. Rev. Drug Discovery* **2002**, *1*, 153−161.

(2) Wang, C.; Kong, H. W.; Guan, Y. F.; Yang, J.; Gu, J. R.; Yang, S. L.; Xu, G. W. *Anal. Chem.* **2005**, *77*, 4108−4116.

(3) Whitehead, T. L.; Kieber-Emmons, T. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 165−174.

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

(5) Constantinou, M. A.; Papakonstantinou, E.; Benaki, D.; Spraul, M.; Shulpis, K.; Koupparis, M. A.; Mikros, E. *Anal. Chim. Acta* **2004**, *511*, 303−312.

(6) Pan, Z.; Gu, H.; Talaty, N.; Chen, H. W.; Hainline, B. E.; Cooks, R. G.; Raftery, D. *Anal. Bioanal. Chem.* **2006**, in press.

(7) German, J. B.; Roberts, M. A.; Watkins, S. M. *J. Nutr.* **2003**, *133*, 4260−4266.

(8) Gibney, M. J.; Walsh, M.; Brennan, L.; Roche, H. M.; German, B.; van Ommen, B. *Am. J. Clin. Nutr.* **2005**, *82*, 497−503.

(9) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *39*, 1−40.

(10) Wagner, S.; Scholz, K.; Donegan, M.; Burton, L.; Wingate, J.; Volkel, W. *Anal. Chem.* **2006**, *78*, 1296−1305.

(11) 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.

(12) Plumb, R.; Granger, J.; Stumpf, C.; Wilson, I. D.; Evans, J. A.; Lenz, E. M. *Analyst* **2003**, *128*, 819−823.

(13) Bell, J. D.; Sadler, P. J.; Morris, V. C.; Levander, O. A. *Magn. Reson. Med.* **1991**, *17*, 414−422.

(14) Moser, V. C.; Phillips, P. M.; McDaniel, K. L.; Marshall, R. S.; Hunter, D. L.; Padilla, S. *Toxicol. Sci.* **2005**, *86*, 375−386.

(15) Phipps, A. N.; Stewart, J.; Wright, B.; Wilson, I. D. *Xenobiotica* **1998**, *28*, 527−537.

(16) Flad, T.; Tolson, J. *Anal. Bioanal. Chem.* **2005**, *381*, 24−27.

(17) Vaidyanathan, S.; Jones, D.; Broadhurst, D. I.; Ellis, J.; Jenkins, T.; Dunn, W. B.; Hayes, A.; Burton, N.; Oliver, S. G.; Kell, D. B.; Goodacre, R. *Metabolomics* **2005**, *1*, 243−250.

(18) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrom. Rev.* **2005**, *24*, 613−646.

(19) Chen, H. W.; Pan, Z.; Talaty, N.; Cooks, R. G.; Raftery, D. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1577−1584.

10.1021/ac060946c CCC: $37.00
Published on Web 11/21/2006

feature has given rise to numerous biological applications of the technique.[21−23]

A new spray ionization technique, extractive electrospray ionization (EESI), which has promising applications, has been developed. EESI is related to DESI, but was developed for the direct analysis of trace compounds in the solution phase, especially when complex mixtures are of interest.[24] The method is also related to an electrospray experiment reported by Fenn and Fursteneau in which compounds present on particulates in air were detected by electrosraying a solvent.[25] DESI utilizes one spray source, in contrast to EESI, which is a direct solution analysis method and utilizes two spray sources, eliminating the use of a surface on which the analyte is first collected. One spray source nebulizes the sample while the other provides charged solvent droplets. The two spray sources used are angled with respect to each other and to the mass spectrometer inlet to introduce the analyte of interest directly to the source.[26] During EESI, the sample and solvent are continuously introduced into the source, making rapid analysis possible. Varying the angles of the spray sources affects the sensitivity and long-term stability of the technique.[24] No sample preparation is required for this method, and stable signals have been observed for as long as 7 h for raw urine,[24] a feature that is beneficial for metabolomics analysis, since large numbers of samples can be analyzed successively, each in a short period of time.

^{1}H NMR spectroscopy has been widely used to study the metabolic variation in biofluid samples, and its capabilities for metabolomics are well established.[9,27] It is a rapid, quantitative method used to extract metabolic information from a sample with minimal or no preparation.[19,28,29] Thousands of peaks resulting from a large collection of metabolites can be simultaneously displayed in the NMR spectra. The challenge of analyzing spectra of complex metabolite mixtures that may contain hundreds of NMR-detectable compounds can be addressed by processing the data using multivariate statistical methods.

A variety of these statistical methods have proven useful for the analysis of complex biofluids. Principal component analysis (PCA) is often used because it provides an effective unsupervised method to differentiate (but by itself not to further describe) the chemical changes in biofluids. PCA is a dimension-reduction technique that transforms correlated variables into a smaller number of orthogonal variables called principal components (PCs).[30,31] Each spectrum is represented by a single point in a score plot with the PCs as the coordinate axes. Similar spectra will be clustered together, and different spectra will be separated along at least one PC axis. The loading plot contains the information about compounds that contribute to the separation. For further analysis of the compounds within the loading plot, the Pearson correlation may be applied to determine the degree to which variables are related to one another.[32] Values range from +1 to −1, indicating positive and negative correlations, respectively. The identification and validation of metabolite species may possibly lead to the rapid identification of biomarkers of disease in appropriate situations.[30,33]

Although much concern has been focused on the extraction and analysis of samples, studies on the effects of the diet of subjects during the sample collection period are also vitally important, since diet is known to play a major role in metabolomics research.[7,15,34] Previous research demonstrates that diet has remarkable effects on urinary composition, and significant changes in hippuric acid, succinate, citrate, and *N*-methylnicotinamide are often observed.[13,15,34,35] In the current study, we show that changes in the diet of rats can be easily observed using PCA and monitored most clearly by examining compounds formed by specific metabolic pathways. Some pathways, such as the urea cycle and metabolism of amino groups (UCMAG) are quick to follow changes in diet. As a result, this presents an opportunity to focus the metabolic analysis on specific pathways.

MATERIALS AND EXPERIMENTS

Animal Study and Sample Collection. To assess the influence of diet variations, urine samples were obtained from four male BALB/c rats for three consecutive days. The rats were acclimated for a period of 4 days before experiments were initiated. Each rat was housed in a metabolism cage with free access to water and rotated daily through the three diets: overnight fast, normal diet (Harlan Teklad 2018 Vegetarian Rodent Diet, 18% protein and 5% fat), and turkey cat food diet (Marsh Gourmet Sliced Turkey in Gravy, Marsh Supermarkets; stored in a refrigerator throughout the course of the study) in a different order for each rat. In total, 12 urine samples were collected and stored at −80 °C until NMR and MS analysis was performed. Rats were treated according to protocols approved by a local Institutional Animal Care and Use Committee.

Sample Preparation and Instrumentation for NMR Studies. A Bruker DRX 500-MHz spectrometer equipped with a room-temperature HCN probe was used to acquire one-dimensional ^{1}H spectra. Samples were prepared by mixing 300 μL of undiluted rat urine with 300 μL of 0.5 M potassium phosphate buffer solution (pH 7.4) containing 10 mM of 3-(trimethylsilyl)propionic-(2,2,3,3-d_4) acid sodium salt (TSP) in D_2O, which was used as the frequency standard (δ = 0.00). Water peaks were suppressed using a standard 1D-NOESY (nuclear Overhauser effect spectroscopy) pulse sequence coupled with water presaturation. For each

(20) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306*, 471−473.
(21) Cooks, R. G.; Ouyang, Z.; Takats, Z.; Wiseman, J. M. *Science* **2006**, *311*, 1566−1570.
(22) Talaty, N.; Takats, Z.; Cooks, R. G. *Analyst* **2005**, *130*, 1624−1633.
(23) Wiseman, J. M.; Puolitaival, S. M.; Takats, Z.; Cooks, R. G.; Caprioli, R. M. *Angew. Chem., Int. Ed.* **2005**, *44*, 7094−7097.
(24) Chen, H.; Venter, A.; Cooks, R. G. *Chem. Commun.* **2006**, 2042−2044.
(25) Fuerstenau, S.; Kiselev, P.; Fenn, J. *Proc. 47th ASMS Conf. Mass Spectrom. Allied Top.*; Dallas, Texas, 1999, ThOE 3:00.
(26) Takats, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Anal. Chem.* **2004**, *76*, 4050−4058.
(27) Sandusky, P.; Raftery, D. *Anal. Chem.* **2005**, *77*, 7717−7723.
(28) de Graaf, R. A.; Behar, K. L. *Anal. Chem.* **2003**, *75*, 2100−2104.
(29) Ramadan, Z.; Jacobs, D.; Grigorov, M.; Kochhar, S. *Talanta* **2006**, *68*, 1683−1691.
(30) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, 1998.
(31) Yang, K.; Trewn, J. *Multivariate Statistical Methods in Quality Management*; McGraw-Hill Companies Inc.: New York, 2004.
(32) Kranowski, W. J. *Principles of Multivariate Analysis: A User's Perspective*; revised edition; Oxford University Press: Oxford, UK, 2000.
(33) Jonsson, P.; Johansson, A. I.; Gullberg, J.; Trygg, J.; A, J.; Grung, B.; Marklund, S.; Sjostrom, M.; Antti, H.; Moritz, T. *Anal. Chem.* **2005**, *77*, 5635−5642.
(34) Vigneau-Callahan, K. E.; Shestopalov, A. I.; Milbury, P. E.; Matson, W. R.; Kristal, B. S. *J. Nutr.* **2001**, *131*, 924s−932s.
(35) Gavaghan, C. L.; Nicholson, J. K.; Connor, S. C.; Wilson, I. D.; Wright, B.; Holmes, E. *Anal. Biochem.* **2001**, *291*, 245−252.

spectrum, 32 transients were collected, resulting in 32k data points using a spectral width of 6000 Hz. An exponential weighting function corresponding to 0.3-Hz line broadening was applied to the free induction decay before applying Fourier transformation.

After phasing and baseline correction using Bruker's XWIN-NMR software, NMR spectral regions were binned to 1000 buckets of equal width to remove the errors resulting from the small fluctuations of chemical shifts due to pH or ion concentration variations. Cloarec and co-workers have recently reported an alternative approach that utilizes the full-resolution data to improve the interpretability of statistical results, although it relies on the supervised statistical method, O-PLS-DA (orthogonal projection on latent structure discriminant analysis).[36] The spectral region from 4.5 to 6 ppm was removed to eliminate the variations in the water resonance suppression as well as the urea signal. Each spectrum was normalized by the integration of the whole spectrum. Noise effects were reduced for the datasets by an iterative (threshold-based) approach. All remaining regions were imported into Pirouette software (v. 3.11; InfoMetrix, Woodinville, WA), where mean-centered PCA was performed.

Instrumentation for Extractive Electrospray Ionization Mass Spectrometry Studies. EESI-MS experiments were carried out using a Thermo Finnigan LCQ (San Jose, CA) mass spectrometer coupled with a home-built EESI source.[24] The two sprayers were set in such a manner that both the angle between the sample nebulizer and MS inlet (α) and the angle between the two sprayers (β) were equal to 90°; this was found to minimize carryover of the urine samples. One hundred-fold-diluted urine samples were examined without any further sample pretreatment. Samples were infused at a rate of 1 μL/min by a syringe pump into the sample nebulizer and dispersed under ambient conditions. The spray solvent (methanol/water/acetic acid, 45:45:10) was infused by another syringe pump at an infusion rate of 5 μL/min. Charged solvent droplets were guided into the sample cloud so that analytes could be extracted into the solvent. The resulting droplets were directed into the atmospheric interface of the mass spectrometer, where evaporation of the solvent yielded analyte ions for MS analysis. All MS spectra were recorded for exactly 1.5 min and converted into txt format for further statistical processing.

To confirm the structures of those compounds that best differentiated the spectra, collision-induced dissociation (CID) was performed in the positive ion detection mode of EESI-MS. To obtain CID spectra, a window of 1.0 m/z units was used to isolate the parent ions, and 25−35% (manufacturer's units) collision energy was applied. To reduce the instability of EESI mass spectra and demonstrate the reproducibility of the technique, five replicate spectra were collected sequentially for each sample.

Similar to the procedure used for the analysis of NMR spectra, the mass spectral region between m/z 100 and 400 was reduced to 1000 buckets of equal width. The data were normalized by integration of each spectrum prior to statistical analysis using Pirouette software. For pathway analysis, mean-centered PCA was applied to 42 compounds known to be associated with the purine metabolism and 19 related to UCMAG with m/z values ranging from 100 to 400. The presence of these compounds in urine samples was confirmed by CID experiments, relevant literature, or the METLIN metabolite database.[37]

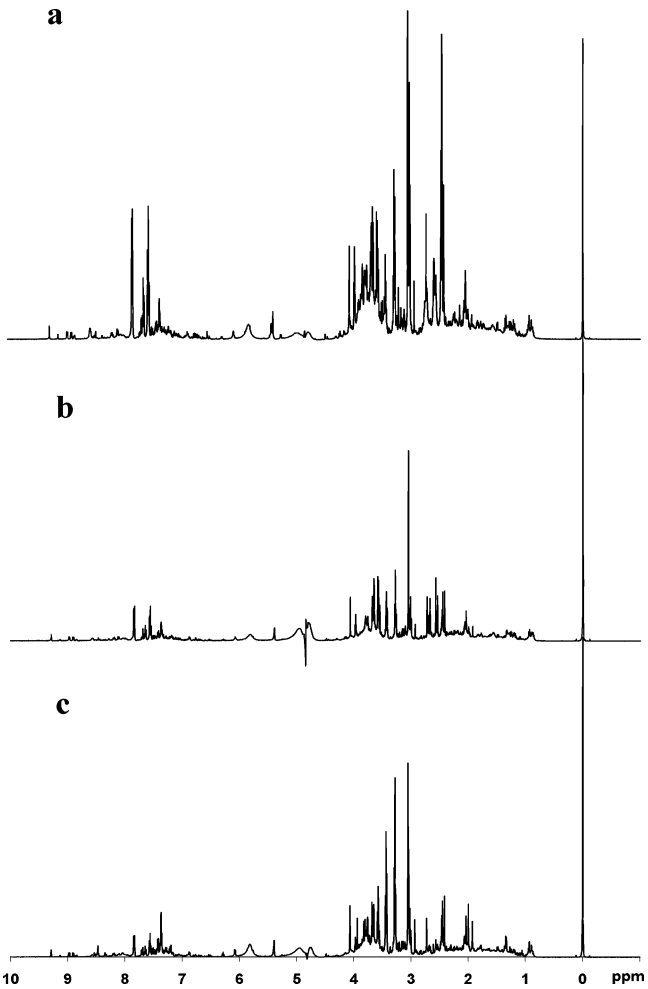

Figure 1. Typical ^{1}H NMR spectra of urine from rats with different diets: (a) normal diet (b) overnight fast (c) turkey diet.

Principal Component Analysis. The variability in the spectral profiles was studied by PCA and by multivariate analysis of variance (MANOVA). To give a simple qualitative measurement of the separation of the urine samples, we first applied a multivariate normal model to the scores from the PCA results using the p-value. Wilks' lambda (Λ),[31] which in this study is an indicator of the strength of the dietary effect, was also calculated for each full score plot and every two clusters in the score plot. The Wilks' Λ was used as the level of discrimination since the p-values used to test the null hypothesis in MANOVA was <0.01 for all score plots. Because Wilks' Λ values do not require a normal distribution assumption, which is difficult to verify for this sample size, it is likely to be a more appropriate measure of clustering than p-values. Wilks' Λ values <0.1 will indicate a stronger treatment effect and, thus, better clustering. In the current study, MANOVA analysis was performed using the R program (version R 2.2.0).

RESULTS AND DISCUSSION

The effect of diet on metabolic composition of rat urine was determined using PCA of ^{1}H NMR and EESI-MS spectra. Figures

(36) Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *Anal. Chem.* **2005**, *77*, 517−526.

(37) http://metlin.scripps.edu/.

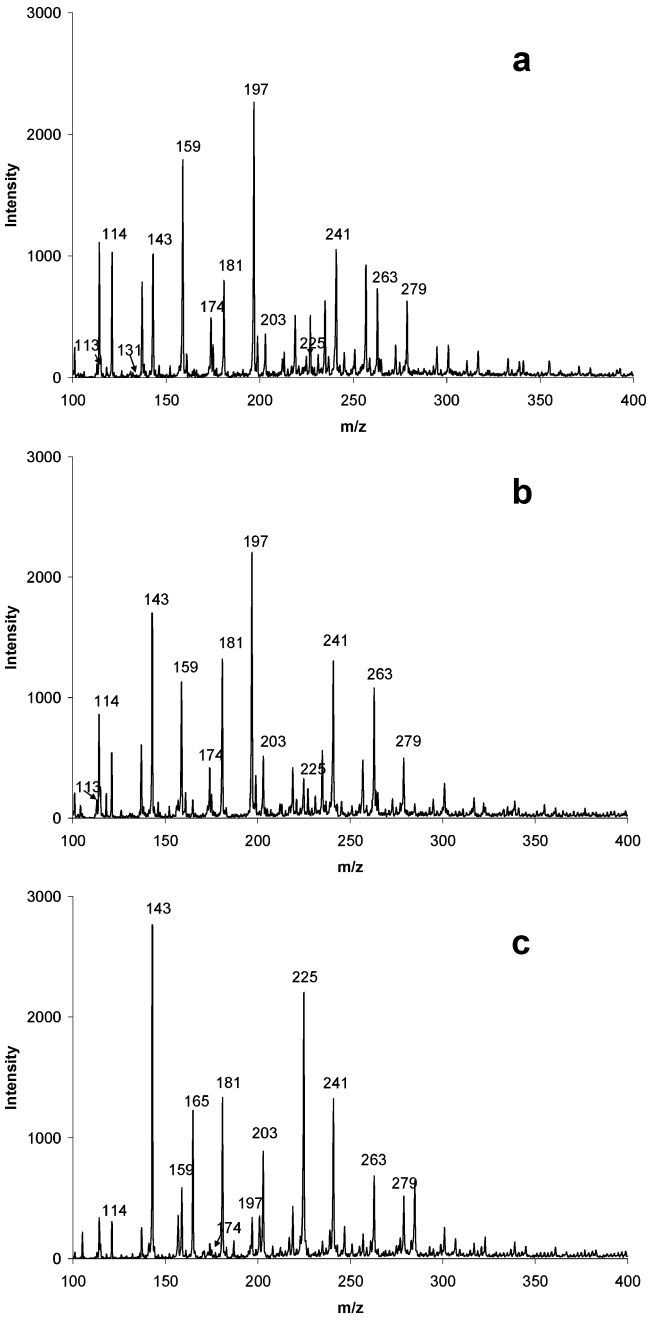

Figure 2. Typical EESI-MS data. Mass spectra collected using LCQ on 100-fold-diluted rat urine samples and methanol/water/acetic acid (45:45:10) spray solvent. (a) normal diet, (b) overnight fast, and (c) turkey diet.

1 and 2 depict typical ^{1}H NMR and EESI-MS spectra and illustrate the pronounced variation among the spectra from the three diets. For both techniques, the spectra share common features but are still unique to each diet. Application of PCA to each spectrum will identify which metabolites are most influential in causing the observed variations among the spectra.

As shown in Figure 1, ^{1}H NMR spectra show a large number of isolated and overlapped peaks caused by the hundreds of metabolites present in the samples. The three spectra in Figure 1 illustrate the chemical shifts of metabolites that are responsible for the distributions in the score plots of PCA results. In the ^{1}H NMR spectra, the aliphatic regions are dominated by peaks from trimethylamine oxide (TMAO), taurine, creatinine, glucose, succinate, dimethylamine and α-ketoglutarate, whereas hippurate and phenylalanine generate large resonances visible in the aromatic region. These assignments are based on previous work reported in the literature.[5,38] There is a larger variation in the aliphatic region than in the aromatic region; therefore, it is anticipated that the aromatic region has a smaller effect on the statistical classification.

Compared to the NMR spectra, the EESI mass spectra show more variations among the three types of samples. For example, changes in intensities of peaks that are provisionally assigned for creatinine (m/z 114), alloxan (m/z 143), gluconic acid (m/z 197), and 3-hydroxykynurenine (m/z 225) are significant in Figure 2. For instance, the intensity of the gluconic acid signal, m/z 197, changes by a factor of almost 8 (from 2195, 2254, 343, arbitrary units) for the normal, overnight fast, and turkey diets, respectively. Figure 3 illustrates this variance in peak intensity for gluconic acid and three other metabolites prominent in each spectrum for the different diets. In Figure 3, the urine of rats treated with the turkey diet have higher ion abundances for alloxan and 3-hydroxykynurenine, whereas peaks for gluconic acid are lower for the turkey diet as compared to the other two diets. Moreover, for glucose, the difference between rats with different diets is much smaller than for the other compounds. These results are also confirmed by PCA results presented later. The variation between rats fed the same diet is also indicated in Figure 3 by the size of the corresponding error bars. Overall, these variations among the individual rats are relatively small, with the largest variation being observed for alloxan in the turkey and normal diets and gluconic acid in the normal diet and overnight fast.

Assignments of peaks that showed pronounced variations in intensities as well as those specific to the purine metabolism and the UCMAG were confirmed through tandem mass spectrometry experiments. Figure 4 illustrates typical EESI tandem mass spectra recorded by CID spectra for the four compounds in Figure 3. The CID data were collected at collision energies ranging from 25 to 35% with a methanol/water/acetic acid (45:45:10) spray solvent in the positive ion mode. For example, the presence of protonated alloxan was confirmed with a standard alloxan solution, which showed fragment ions with m/z 143, 126, 114, and 84, corresponding to losses of $C_4H_3O_4N_2$ (protonated parent ion), OH, COH, and NHCOHNH, respectively.

PCA Results of ^{1}H NMR Spectra. To display the quantitative metabolite variations due to diet and obtain a more accurate analysis, PCA was performed using the full, processed ^{1}H NMR spectra. As shown in Figure 5a, PCA separated the 12 rat urine samples into three groups according to the dietary treatments in the score plot of PC1 versus PC2. The first two PCs explain more than 90% of the total variance. Figure 5b illustrates this variation in 1-D loading plots of PC1 and PC2 resulting from the NMR spectra. The variation within the score plot can be attributed to the alterations of metabolite resonance signals in the NMR spectra. From the two loading plots, the species that are most responsible for differentiation in the NMR spectra are creatinine (3.05 s), glucose (3.42 t, 3.54 dd), 2-oxoglutarate (2.45 t, 3.01 t), TMAO (3.26 s), and taurine (3.28 t, 3.43 t), which contribute strongly to

(38) Feng, J. H.; Li, X. J.; Pei, F. K.; Chen, X.; Li, S. L.; Nie, Y. X. *Anal. Biochem.* **2002**, *301*, 1−7.

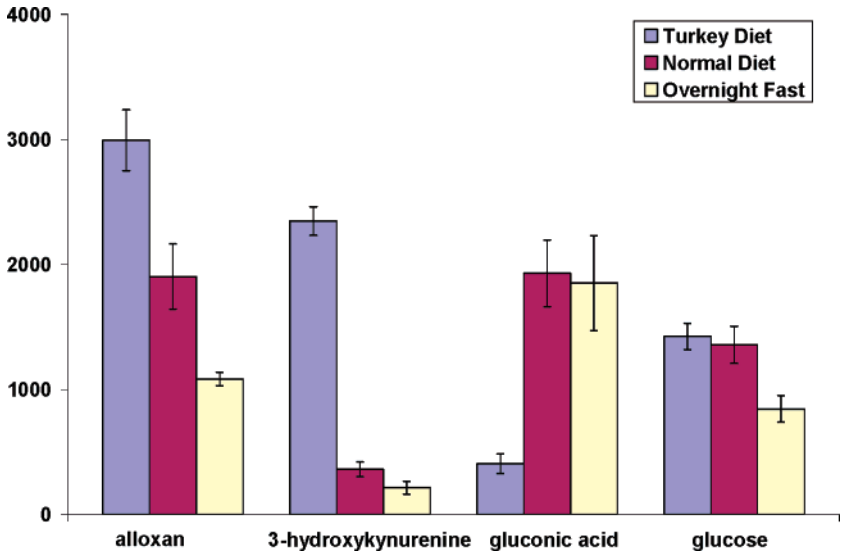

Figure 3. Plots of intensities of a small set of compounds in EESI-MS of different samples: alloxan (*m*/*z* 143), 3-hydroxykyurenine (225), gluconic acid (*m*/*z* 197), and glucose (*m*/*z* 181). See Figure 2 for spectra.

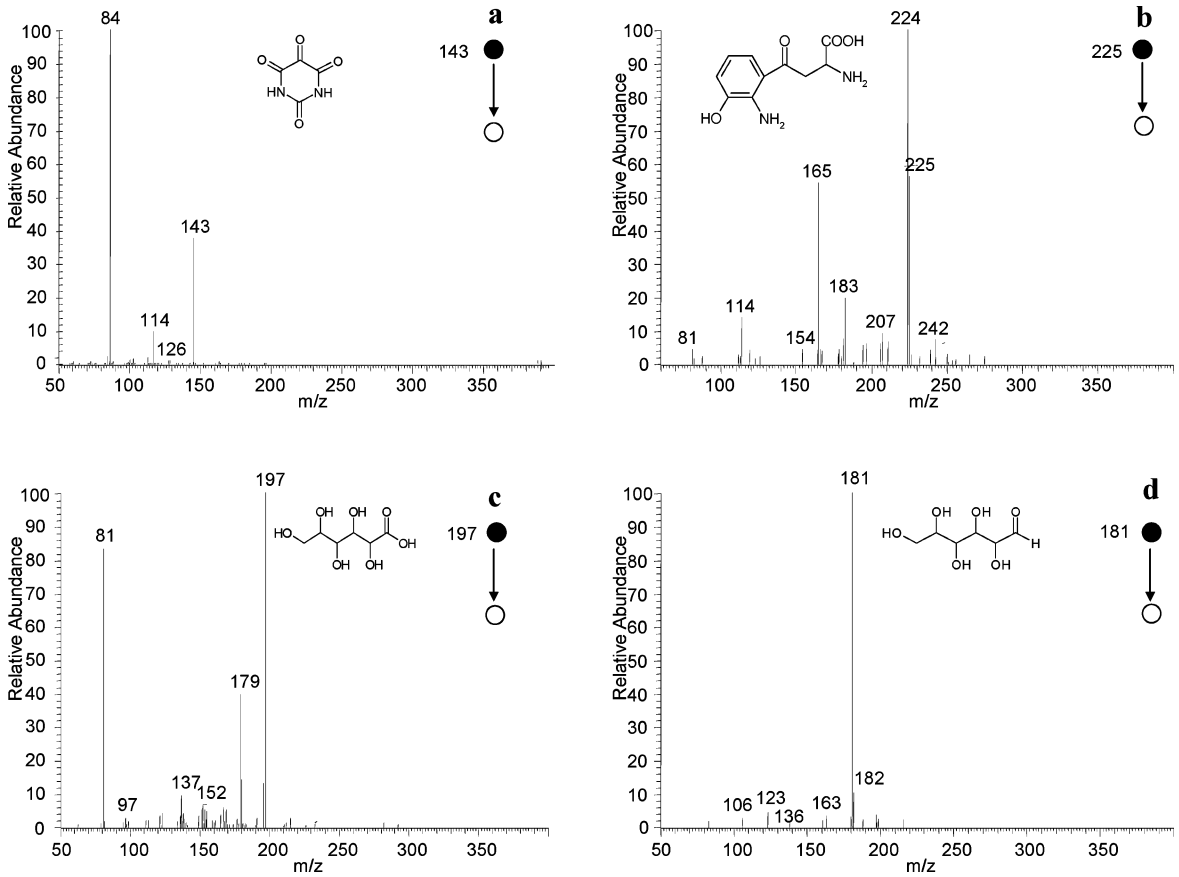

Figure 4. Typical EESI tandem mass spectrometry data. CID was used to record product ion spectra collected in the positive ion mode with a collision energy ranging from 25 to 35% using methanol/water/acetic acid (45:45:10) spray solvent in the LCQ mass spectrometer. Samples were diluted 100-fold to conserve sample. (a) Alloxan (CID *m*/*z* 143), (b) 3-hydroxykyurenine (CID *m*/*z* 225), (c) gluconic acid (CID *m*/*z* 197), and (d) glucose (CID *m*/*z* 181).

the aliphatic region. Additional, smaller changes are seen in the aromatic region.

Wilks' Λ values presented in Table 1 represent the quality of the separation or clustering for the score plot of Figure 5a. The Λ value for spectra within a cluster is 1, since the same diet treatment is being evaluated. Since Λ values are <0.1 for the remaining comparisons, it is reasonable to claim that the classification in the score plot is of good quality. Two terms are important for the calculation of Λ values: one is the variation among spectra in each cluster; another is the difference among clusters. The former is determined by many factors, such as health, interaction between rats, and the reproducibility of the instrument. However, this term is expected to be small because the rats chosen were of the same strain and were allowed to

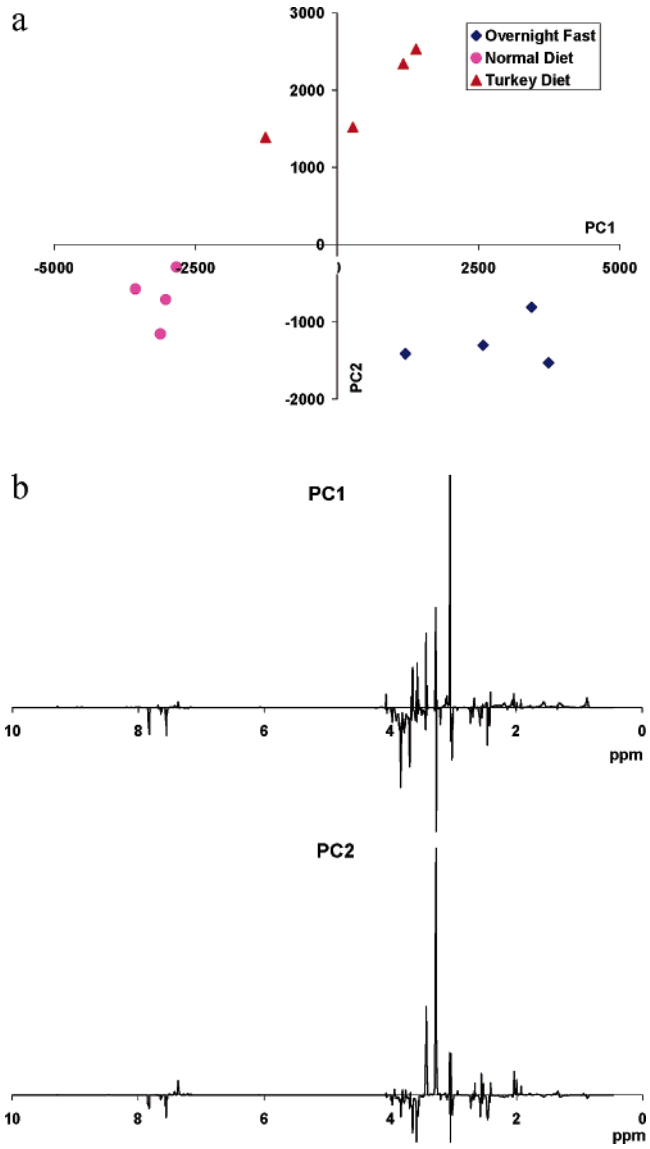

Figure 5. Results of mean-centered PCA results for NMR data of rat urine samples: (a) score plot with an overall $\Lambda = 0.005$ and (b) loading plots for PC1 and PC2.

Table 1. Wilks' Λ for Score Plot Based on NMR Spectra[a]

	turkey diet	normal diet	overnight fast	full plot
turkey diet	1	0.091	0.024	0.005
normal diet	0.091	1	0.047	
overnight fast	0.024	0.047	1	

[a] See Figure 5a for score plot.

interact throughout the study, thus minimizing metabolic differences due to gut microflora.[39] In addition, the process of acquiring and processing the data is kept consistent during the study. The latter term, variation between clusters, is expected to be the most influential to the observed classification in the score plot, which we assume is determined by the different dietary regimens. The small error bars seen in Figure 3 add further evidence that these effects are relatively small, as compared to the observed diet effects.

(39) Robertson, D. G. *Toxicol. Sci.* **2005**, *85*, 809−822.

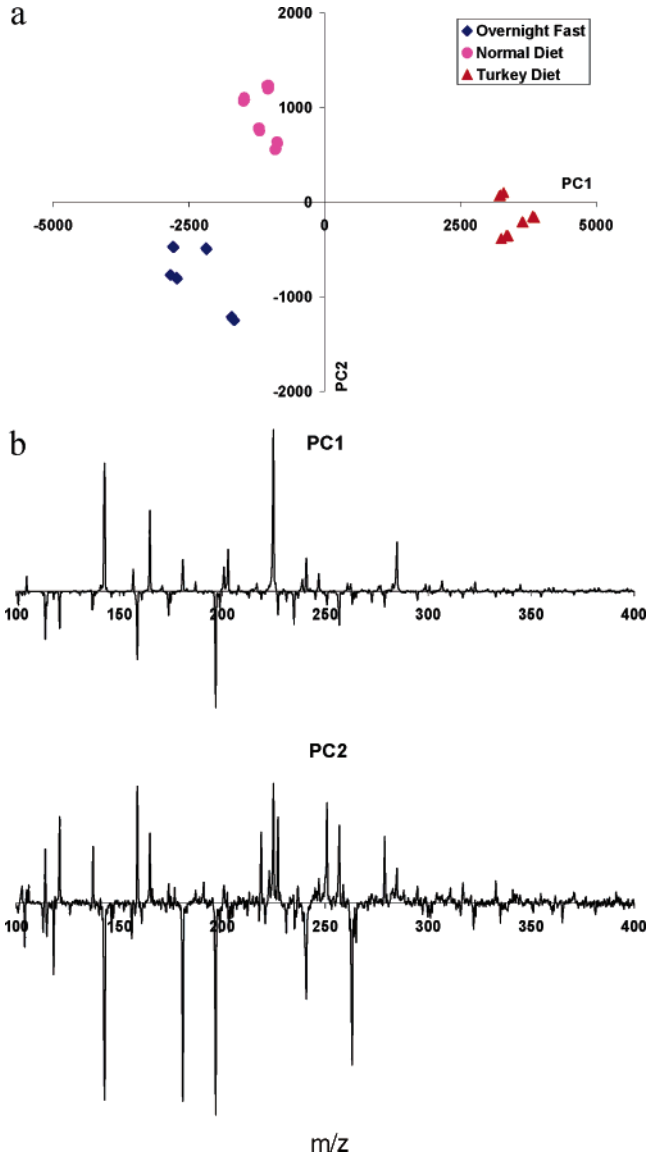

Figure 6. Plots of mean-centered PCA results for EESI-MS data of rat urine samples recorded using methanol/water/acetic acid as spray solvent with five measurements for each sample: (a) score plot illustrating reproducibility of the EESI technique and separation of diets with an overall $\Lambda = 0.001$ and (b) loading plots for PC1 and PC2.

PCA Results of Extractive Electrospray Ionization Mass Spectra. PCA was carried out using the EESI mass spectral data over the m/z range of 100−400. Five replicate measurements were performed for each sample. In Figure 6a, good reproducibility is indicated; each cluster contains 20 spectra. The reproducibility is evident as the five spectra for each sample are clustered tightly together to give the appearance of fewer data points. Improved classification is obtained when compared with the score plot of the NMR spectra (Figure 5a). Table 2 gives Λ values for the score plot of the EESI mass spectral data. It is found that Figure 6a has a somewhat tighter cluster when the same diet is evaluated and there is better separation between different diets than Figure 5a, which is evident by the smaller Λ values. The high-quality separation of diets in Figure 6a explains the large differences observed for EESI mass spectra of urine samples from rats fed different diets.

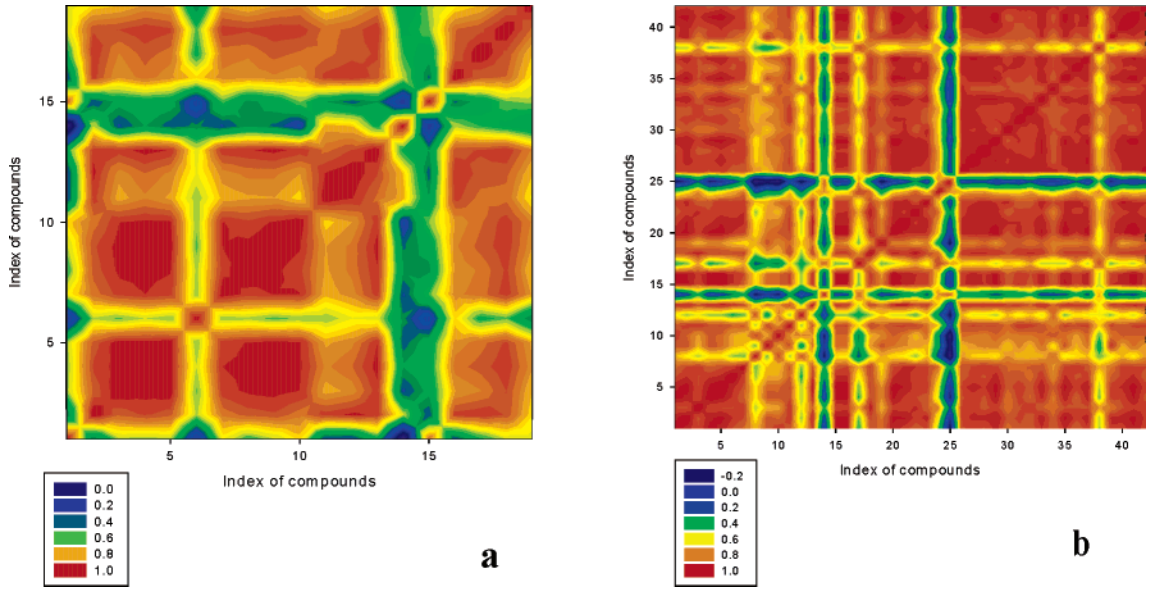

Figure 7. Pearson correlation among (a) 19 molecules related to the UCMAG (urea cycle and metabolism of amino groups), and (b) 42 molecules related to purine metabolism. The scale indicates the range of correlations observed.

Table 2. Wilks' Λ for Score Plot Based on EESI- Mass Spectra[a]

	turkey diet	normal diet	overnight fast	full plot
turkey diet	1	0.010	0.009	0.001
normal diet	0.010	1	0.035	
overnight fast	0.009	0.035	1	

[a] See Figure 6a for score plot.

The molecules that contribute most to the differences in spectral patterns were determined using the same methodology as that used for ^{1}H NMR, and these data are presented in Figure 6b and Supporting Information Figure S1. The principal compounds that show variations in MS include glucose (m/z 181), creatinine (m/z 114), alloxan (m/z 143), gluconic acid (m/z 197), cystine (m/z 240), 3-hydroxykynurenine (m/z 225), γ-L-glutamylcysteine (m/z 251), and carnosine (m/z 227). The concentrations of alloxan, 3-hydroxykynurenine, and 5-dihydro-1*H*-imidazole-5-carboxylate are higher in urine samples from rats on the turkey diet than from rats on the other two diets; conversely, the concentration of urinary gluconic acid is lower from rats on the turkey diet. However, for glucose, the loading value for PC1 is small compared to its PC2 value; thus, the effect of PC2 is not negligible, even though PC2 contains only 7% of the total variance in the spectra (PC1 explains 85%). The results are in agreement with those presented in Figure 3; spectra for the turkey diet show higher intensities for ions corresponding to alloxan and 3-hydroxykynurenine and lower intensities for gluconic acid, as indicated, whereas the differences among the three diet regimens for glucose are blurred. NMR and EESI-MS give similar clustering. However, with the exception of glucose and creatinine, they select for different information due to their differences in sensitivity, selectivity, and detection method. These differences are also complicated by spectral overlaps that are different for the two techniques. However, the results here indicate that the PCA of NMR data and EESI mass spectral data could be cross-validated in terms of classification.

PCA of Compounds in the Urea Cycle and Metabolism of Amino Groups and Those Related to Purine Metabolism. The effect of the three diets was further examined by monitoring compounds associated with specific metabolic pathways. Metabolic pathways are composed of a series of chemical reactions occurring in living systems to generate certain compounds. The concentrations of enzymes that catalyze these reactions can be changed at the gene level by changes induced by diet.[40] All the reactants for the pathway reactions come from food intake, either directly or indirectly. As a result, we might expect that metabolites in some pathways will more strongly express differences induced by diet intake than those associated with other pathways. We have focused on purine metabolism and the UCMAG for this analysis.

A question one might ask is whether the metabolites in an individual pathway are correlated to each other. The Pearson correlation can be used to address this question.[30,32] The Pearson correlation was calculated for each pair of metabolites identified by MS in each of the two metabolic pathways (19 compounds for UCMAG and 42 for purine metabolism) across the set of 12 urine samples. As is shown in Figure 7, the Pearson correlation matrices indicate that most of the compounds within each of these two metabolic pathways are highly and positively correlated, and this is especially so for metabolites that are directly linked by enzymes in the pathway. Correlation values above 0.9 are not uncommon. Interestingly, there are several places where there is a negative correlation, and these indicate the possibility of a change in enzymatic activity that couples two negatively correlated metabolites.

Figure 8a shows the PCA results for those compounds present in the UCMAG that are responsible for ions with m/z 100−400. In the score plot (Figure 8a), there are three clusters that follow the diet regimens, similar to the classification that results from the full spectrum analysis. The Wilks' Λ for the reduced score plot (Figure 8a) is summarized in Table 3. It is shown that the

(40) Eder, K.; Flader, D.; Hirche, F.; Brandsch, C. *J. Nutr.* **2002**, *132*, 3400−3404.

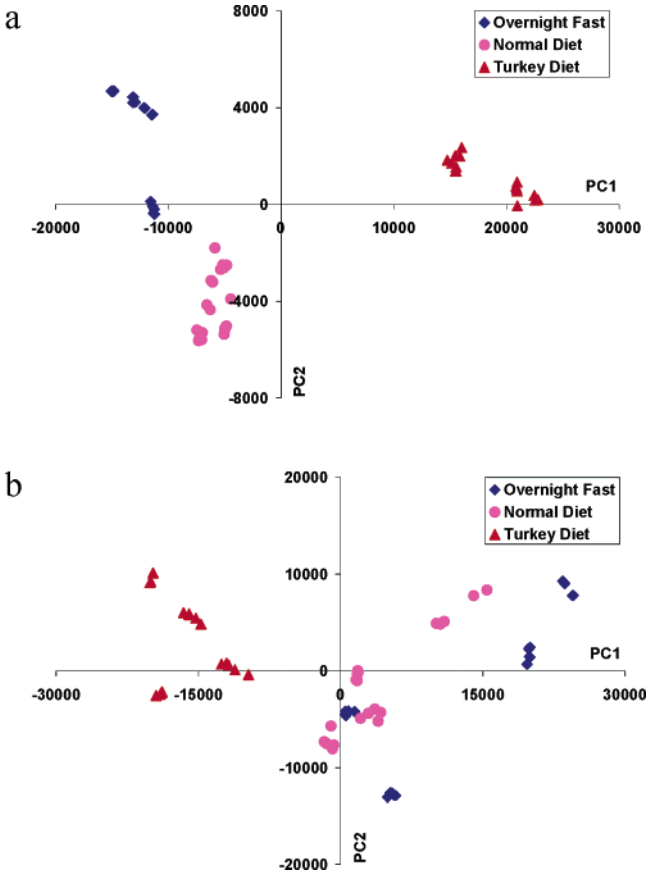

Figure 8. Score plots of mean-centered PCA results of EESI-MS data monitoring compounds in (a) the UCMAG and (b) purine metabolism. (a) Score plot of the UCMAG compounds illustrating tight clustering and good separation based on diet using compounds in the UCMAG illustrating the significant effect of diet. Overall $\Lambda = 0.003$. (b) Score plot of purine metabolism compounds illustrating poor clustering and separation based on diet using compounds in purine metabolism. Overall $\Lambda = 0.106$.

Table 3. Wilks' Λ for Score Plot Based on PCA of 19 Compounds from the Urea Pathway[a]

	turkey diet	normal diet	overnight fast	full plot
turkey diet	1	0.020	0.019	0.003
normal diet	0.020	1	0.093	
overnight fast	0.019	0.093	1	

[a] See Figure 8a for score plot.

clustering is of good quality, although Λ values are slightly higher than for the analysis using the full mass spectra. The loading plot (Figure S1a) illustrates that creatinine, guanidinoacetate, and 5-dihydro-1*H*-imidazole-5-carboxylate are the main compounds that contribute to the classification seen in the score plot. These results suggest that 19 metabolites in the UCMAG are enough to express most of the variations in metabolic profiles caused by different diets.

Figure 8b shows the PCA results for 42 compounds that are related to purine metabolism and that give ions with m/z 100−400. In the score plot (Figure 8b), only rats on the turkey diet are separated, whereas the data points representing the overnight fast and normal diet are mixed. Compared to Figures 6a and 8a, Figure 8b gives the worst separation, because Λ values in Table

Table 4. Wilks' Λ for Score Plot Based on PCA of 42 Compounds from the Purine Metabolism[a]

	turkey diet	normal diet	overnight fast	full plot
turkey diet	1	0.104	0.107	0.106
normal diet	0.104	1	0.478	
overnight fast	0.107	0.478	1	

[a] See Figure 8b for score plot.

4 are larger than 0.1. For example, the level of discrimination between overnight fast and normal diet is 0.48. One point worth noting here is that even the *p*-value for purine metabolism is <0.01, which indicates that the mean values for samples representing the different groups are well-separated. The compounds that strongly influence the separation between diets were identified using the loading plot (Figure SI1b). 5-Dihydro-1*H*-imidazole-5-carboxylate, xanthosine, and allantoin can separate the turkey diet from the other two diets somewhat, but the normal diet and overnight fast diets cannot be differentiated by PCA.

The present study suggests that metabolites of the UCMAG are more affected by diet as compared to metabolites of purine metabolism. Excess nitrogen is converted to urea and removed from the human body by dominant reactions in the UCMAG.[41,42] Animals cannot transform atmospheric nitrogen into forms that can be used by the body, and thus, diet is the main source for amino acids containing nitrogen, which is important in formation of tissues. Currently, dietary alteration is being applied as a clinical treatment for diseases caused by urea cycle defects,[43] as well as for a number of genetic metabolic diseases.[6] Purine metabolism involves the synthetic process of purine and pyrimidine nucleotides.[41,44] Indeed, the nutritional requirement for nucleotides is mostly relieved by nucleotide sources within the body; thus, it is expected and found that diet will have much less effect on the concentrations of compounds related to purine metabolism.

CONCLUSION

In this study, we have demonstrated (1) the effect of diet on metabolites in purine metabolism and the urea cycle that center on the metabolism of amino groups, (2) the ability of EESI-MS to serve as an analytical tool for metabolomics analysis, and (3) the PCA results of NMR data and EESI mass spectral data can be cross-validated. Rats of identical species fed with the same diet are clustered together whereas different treatments are separated when PCA is applied to the full NMR and MS spectra. MS in combination with NMR is a promising tool for discovering and quantifying metabolites in biofluid samples. In this study, the PCA of EESI-MS and full ^{1}H NMR spectra gave similar clusters according to the diet treatments. EESI yields reproducible results with good clustering, which demonstrates that EESI has promising applications for metabolomic analysis. Depending on the type or amount of the sample, either EESI or DESI can be picked for the particular metabolomics application. DESI has proved successful

(41) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; W. H. Freeman and Company: New York, 2001.
(42) Mori, M.; Gotoh, T.; Nagasaki, A.; Takiguchi, M.; Sonoki, T. *J. Inherited Metab. Dis.* **1998**, *21*, 59−71.
(43) Leonard, J. V. *J. Pediatr.* **2001**, *138*, S40−S44.
(44) Zöllner, N. *P. o. t. N. S.*, 41, 329−342.

in analyzing dried spots of blood, urine, and tissues to differentiate diseased and normal samples.[6,19]

Diet is an important factor that is necessary to consider in metabolomics research. From a more specific analysis, we find that performing PCA on different metabolic pathways indicates that diet affects the UCMAG more than it does purine metabolism. Chemical variations caused by diet are well-expressed by the UCMAG. Significant variations were observed in creatinine, guanidinoacetate, 5-dihydro-1*H*-imidazole-5-carboxylate, xanthosine, and allantoin. This approach may be useful for tailoring metabolomic analyses to specific metabolic pathways, and may also suggest a method to control the influence of diet in future metabolomics studies.

This paper also shows that the influence of diet can be observed by PCA in NMR and EESI mass spectra after 24 h. Although diets can be varied, in clinical applications, some variability of subjects, including age or heredity, cannot be eliminated. The distortion caused by these factors can be significant and make statistical analysis less certain. Focusing on a specific pathway may reduce the number of compounds needed for statistical analysis and perhaps limit the effect of diet and hopefully some other potential factors.

ACKNOWLEDGMENT

This work is supported by NIH/NIGMS Grant no. 5R01 GM58008-07 and the NIH Roadmap Initiative on Metabolomics Technology, NIH/NIDDK 3 R21 DK070290-01.

SUPPORTING INFORMATION AVAILABLE

Supporting information for this text includes loading plots of the mean-centered PCA results of EESI-MS data monitoring the specific pathways.

Received May 23, 2006. Accepted September 22, 2006.

AC060946C