

2015

Combining DI-ESI-MS and NMR datasets for metabolic profiling

Darrell D. Marshall

University of Nebraska-Lincoln

Shulei Lei

University of Nebraska-Lincoln

Bradley Worley

University of Nebraska-Lincoln, bradley.worley@huskers.unl.edu

Yuting Huang

University of Nebraska-Lincoln

Aracely Garcia- Garcia

University of Nebraska-Lincoln

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/chemistrypowers>

Marshall, Darrell D.; Lei, Shulei; Worley, Bradley; Huang, Yuting; Garcia, Aracely Garcia-; Franco, Rodrigo; Dodds, Eric D.; and Powers, Robert, "Combining DI-ESI-MS and NMR datasets for metabolic profiling" (2015). *Robert Powers Publications*. 56.
<http://digitalcommons.unl.edu/chemistrypowers/56>

This Article is brought to you for free and open access by the Published Research - Department of Chemistry at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Robert Powers Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Darrell D. Marshall, Shulei Lei, Bradley Worley, Yuting Huang, Aracely Garcia- Garcia, Rodrigo Franco, Eric D. Dodds, and Robert Powers



Published in final edited form as:

Metabolomics. 2015 April ; 11(2): 391–402. doi:10.1007/s11306-014-0704-4.

Copyright Springer Science+Business Media New York 2014

Combining DI-ESI-MS and NMR datasets for metabolic profiling

Darrell D. Marshall^{1,†}, Shulei Lei^{1,†}, Bradley Worley^{1,†}, Yuting Huang¹, Aracely Garcia-Garcia^{2,3}, Rodrigo Franco^{2,3}, Eric D. Dodds^{1,*}, and Robert Powers^{1,2,*}

¹Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304

²Redox Biology Center, University of Nebraska-Lincoln, Lincoln, NE 68583-0905

³School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, Lincoln, NE 68583-0905

Abstract

Metabolomics datasets are commonly acquired by either mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR), despite their fundamental complementarity. In fact, combining MS and NMR datasets greatly improves the coverage of the metabolome and enhances the accuracy of metabolite identification, providing a detailed and high-throughput analysis of metabolic changes due to disease, drug treatment, or a variety of other environmental stimuli. Ideally, a single metabolomics sample would be simultaneously used for both MS and NMR analyses, minimizing the potential for variability between the two datasets. This necessitates the optimization of sample preparation, data collection and data handling protocols to effectively integrate direct-infusion MS data with one-dimensional (1D) ¹H NMR spectra. To achieve this goal, we report for the first time the optimization of (i) metabolomics sample preparation for dual analysis by NMR and MS, (ii) high throughput, positive-ion direct infusion electrospray ionization mass spectrometry (DI-ESI-MS) for the analysis of complex metabolite mixtures, and (iii) data handling protocols to simultaneously analyze DI-ESI-MS and 1D ¹H NMR spectral data using multiblock bilinear factorizations, namely multiblock principal component analysis (MB-PCA) and multiblock partial least squares (MB-PLS). Finally, we demonstrate the combined use of backscaled loadings, accurate mass measurements and tandem MS experiments to identify metabolites significantly contributing to class separation in MB-PLS-DA scores. We show that integration of NMR and DI-ESI-MS datasets yields a substantial improvement in the analysis of neurotoxin involvement in dopaminergic cell death.

Keywords

DI-ESI-MS; NMR; Metabolomics; Multivariate Statistics; Multiblock PCA; Multiblock PLS

*To whom correspondence should be addressed: Robert Powers, University of Nebraska-Lincoln, Department of Chemistry, 722 Hamilton Hall, Lincoln, NE 68588-0304, rpowers3@unl.edu, Phone: (402) 472-3039, Fax: (402) 472-9402. Eric D. Dodds, University of Nebraska-Lincoln, Department of Chemistry, 711 Hamilton Hall, Lincoln, NE 68588-0304, edodds2@unl.edu, Phone: (402) 472-3592, Fax: (402) 472-9402.

[†]Equal contribution

Electronic supplementary material

Additional information as noted in text. A link to the electronic supplementary material is provided below.

INTRODUCTION

The analysis of metabolomics samples is routinely carried out using either mass spectrometry (MS) (Dettmer, Aronov, Hammock 2007) or nuclear magnetic resonance (NMR) spectroscopy (Nicholson, Lindon, Holmes 1999). However, NMR and MS have distinct, complementary sets of strengths and limitations. The advantages of NMR for metabolomics include relatively high-throughput, non-destructive data acquisition, minimal sample handling, simple methods for quantitation of metabolite alterations, and redundant spectral information to improve the accuracy of metabolite identification (Pan, Raftery 2007; t'Kindt, Scheltema, Jankevics, Brunker, Rijal, Dujardin, Breitling et al. 2010; Zhang, Halouska, Gaupp, Lei, Snell, Fenton, Barletta et al. 2013). However, one-dimensional (1D) ^1H NMR is limited by low sensitivity ($\sim 1 \mu\text{M}$), low information content (~ 0.02 ppm resolution over a ~ 10 ppm spectral width) and low dynamic range, all of which reduce the observable set of metabolites. These deficiencies of NMR are strengths of MS (Lenz, Wilson 2007; Pan, Raftery 2007). For instance, MS has a much higher sensitivity compared to NMR, readily measuring concentrations in the nanomolar (nM) range. MS also boasts higher resolution ($\sim 10^3$ to 10^4) and dynamic range ($\sim 10^3$ to 10^4). As a result, MS-based metabolomics studies can potentially detect a much greater subset of the metabolome than NMR.

More often than not, MS metabolomics relies on hyphenated analytical platforms, such as GC-MS (Kuehnbaum, Britz-McKibbin 2013) or LC-MS (Crockford, Holmes, Lindon, Plumb, Zirah, Bruce, Rainville et al. 2006), to reduce peak overlap and improve coverage of the metabolome. Peak overlap in the mass spectrum occurs because of the relatively narrow molecular-weight distribution of the metabolome (Kell 2004). Ion suppression is also a significant concern given the complexity and heterogeneity of metabolomics samples. The competition for charge between co-eluting analytes may lead to altered or missing metabolite signals (Metz, Page, Baker, Tang, Ding, Shen, Smith 2008). While the coupling of a chromatographic separation to MS potentially alleviates these issues, it also increases analysis time and requires additional sample preparation in comparison with NMR.

The extra sample processing required by chromatography may lead to variations in the observed metabolome not relevant to the biological system under study (Canelas, ten Pierick, Ras, Seifar, van Dam, van Gulik, Heijnen 2009). As an example, the chemical derivatization step required by GC-MS may individually bias metabolite concentrations. The derivatization yields may differ for each metabolite, and no derivatizing agent exists that will universally and efficiently label all metabolites in any given biological sample (Kanani, Chrysanthopoulos, Klapa 2008). Compound decomposition during derivatization or separation may also contribute to this bias (Xu, Zou, Ong 2009), and co-eluting matrix compounds may further suppress the ionization of true analytes in LC-MS (Taylor 2005). In fact, there is now a growing body of evidence suggesting that, with a judicious choice of instrumental conditions, direct infusion electrospray MS (DI-ESI-MS) may achieve equal or greater ion transmission efficiency in metabolic fingerprinting relative to LC-MS (Draper, Lloyd, Goodacre, Beckmann 2014). DI-ESI-MS requires less sample pre-treatment and allows for shorter instrument cycle times than LC-MS and GC-MS, and does not require post-acquisition alignment of retention times (Kopka 2006; Lange, Tautenhahn, Neumann,

Gropl 2008). Thus, DI-ESI-MS is an attractive choice of analytical platform to complement NMR for high-throughput metabolic fingerprinting and profiling.

Ion sources such as direct analysis in real time (DART) and desorption electrospray ionization (DESI) have also been utilized for MS analysis of metabolomics samples (Chen, Pan, Talaty, Raftery, Cooks 2006; Gu, Pan, Xi, Asiago, Musselman, Raftery 2011). DESI and DART are ambient ionization methods and provide an abundance of reproducible data (Chen, Pan, Talaty, Raftery, Cooks 2006; Gu, Pan, Xi, Asiago, Musselman, Raftery 2011). However, DESI and DART are not nearly as widely accessible as DI-ESI, which is nearly universally available in modern instrumentation facilities. DI-ESI-MS is also easily automated and can be performed with minimal sample preparation, which is indispensable to high-throughput studies (Draper, Lloyd, Goodacre, Beckmann 2014).

The combination of NMR and MS techniques for metabolic fingerprinting and profiling is a growing trend (Pan, Raftery 2007) and has been shown to improve metabolomics coverage (Barding, Beni, Fukao, Bailey-Serres, Larive 2013). A number of metabolomics studies combine 1D ^1H NMR experiments with LC-MS (Atherton, Bailey, Zhang, Taylor, Major, Shockcor, Clarke et al. 2006; Jung, Jung, Kim, Ryu, Hwang 2013) or GC-MS (Barding, Beni, Fukao, Bailey-Serres, Larive 2013). In these cases, samples and data for each instrumental platform are handled in effective isolation. Most importantly, such studies require the preparation of separate sets of metabolite samples that meet the specific needs of NMR and MS instrumentation (Beltran, Suarez, Rodriguez, Vinaixa, Samino, Arola, Correig et al. 2012). Nevertheless, such parallel approaches greatly enhance the structural characterization and quantitation of metabolites (Dai, Xiao, Liu, Hao, Tang 2010). An alternative approach is to use MS as the primary analytical tool, relying on NMR to validate the results or confirm the identification of key metabolites (Mullen, Wheaton, Jin, Chen, Sullivan, Cheng, Yang et al. 2012). To date, a limited number of metabolomics studies actually integrate NMR spectral data with information obtained from direct infusion ion sources.

The combined multivariate statistical analysis of data from multiple instrumental platforms is a nascent and underutilized practice in the metabolomics field. Most studies that integrate NMR and DI-ESI-MS data still perform separate analyses of their respective data matrices and combine the results in an attempt to enhance the total information content. For example, Chen *et al.* performed independent principal component analyses (PCA) on NMR and MS datasets and combined the scores from each analysis into a three dimensional (3D) scores plot (Chen, Pan, Talaty, Raftery, Cooks 2006). While the resulting combined scores yielded greater between-class separations than the original NMR or MS scores, such an analysis *completely ignores* the highly informative correlations that exist between the two datasets. Similarly, Gu *et al.* (Gu, Pan, Xi, Asiago, Musselman, Raftery 2011) replaced binary class designations in an orthogonal projections to latent structures (OPLS) analysis of MS data with scores from a PCA of the corresponding NMR spectra. While the resulting OPLS-R class separations were greater than the original OPLS-DA separations, such an analysis carries no statistical guarantee of success for any other dataset.

Multiblock bilinear factorizations such as Consensus PCA, Hierarchical PCA, Hierarchical PLS and Multiblock PLS provide a powerful framework for analyzing a set of multivariate observations from multiple analytical measurements containing potentially correlated variables (Smilde, Westerhuis, de Jong 2003; Westerhuis, Kourti, Macgregor 1998; Wold 1987). Such algorithms provide analogous information to classical PCA and PLS in situations where extra knowledge is available to subdivide the measured variables into multiple “blocks”. As a result, the correlation structures of each block *and* the between-block correlations may be simultaneously utilized. Due to the existence of trends common to each block, this use of between-block correlations during modeling will ideally bring the model loadings (latent variables) into better agreement with the true underlying biology (hidden variables). In short, multiblock algorithms provide an ideal means of integrating 1D ¹H NMR and DI-ESI-MS datasets for metabolic fingerprinting studies (Xu, Correa, Goodacre 2013).

The successful integration of DI-ESI-MS data with 1D ¹H NMR data for metabolic fingerprinting and profiling necessitates improving sample preparation, data collection and data processing protocols. Our described optimization of sample preparation protocols enabled the utilization of a single sample for both NMR and MS analysis. To further diminish the impact of sample handling, samples were infused directly into the mass spectrometer without pre-source separation. Electrospray source conditions were then optimized in order to maximize the performance of DI-ESI-MS and minimize ion suppression and/or enhancement (matrix effects). Multiblock PCA (MB-PCA) and multiblock PLS (MB-PLS) were used to analyze the collected NMR and mass spectral data, allowing the identification of key metabolites that significantly contributed to class separation from the resulting scores and loadings. Finally, NMR, accurate mass and MS/MS data were collected to enhance the accuracy and efficiency of metabolite identification. Our resulting protocol for combining DI-ESI-MS with 1D ¹H NMR for metabolic fingerprinting and profiling is summarized in Figure 1.

MATERIALS AND METHODS

Samples and reagents

All standard reagents and isotopically labeled chemicals were obtained from Sigma Aldrich (St. Louis, MO), Fischer Scientific (Fair Lawn, NJ) and Cambridge Isotopes (Andover, MA). A standard metabolite mixture was prepared by mixing six compounds together: caffeine, L-histidine, β-alanine, L-glutamine, (S)-(+)-ibuprofen, and L-asparagine at concentrations of 10 mM in double distilled water (ddH₂O)/methanol/FA (49.75:49.75:0.5). The solution was diluted by a factor of 1000 for MS analysis. Metabolite extracts from *Escherichia coli* Mach1 were prepared as previously described in detail (Zhang, Halouska, Gaupp, Lei, Snell, Fenton, Barletta et al. 2013). To generate analytical replicates, each of the metabolite extracts were separated into three aliquots of 100 μL and then diluted ten-fold with ddH₂O/methanol/FA (49.75:49.75:0.5) containing 10 μM caffeine as an internal mass reference. A complete description of the preparation of standard samples is available in the supporting information.

Preparation of metabolomics samples from human dopaminergic neuroblastoma cells

Human dopaminergic neuroblastoma cells (SK-N-SH) with different neurotoxin treatments were used as a metabolomics test system for developing the methodology for integrating NMR and MS data. SK-N-SH cell lines were cultured in DMEM/F12 medium consisting of 100 units/mL penicillin-streptomycin and 10% fetal bovine serum. 1.5×10^6 SK-N-SH cells were plated on a 10 cm dish for overnight incubation with 5% CO₂ at 37 °C. The cells were then treated with 1-methyl-4-phenylpyridinium (MPP⁺), 50 μM 6-hydroxydopamine (6-OHDA), 0.5 mM paraquat or 4.0 μM rotenone for 24 h. The cells were washed twice with 5 mL PBS washes. 1.0 mL of cold methanol (−80 °C) was immediately added to simultaneously lyse and quench the cells, which were then incubated at −80 °C for 15 minutes to facilitate cell lysis and metabolite extraction. The cells were detached from each dish using a cell scraper and cellular detachment was confirmed using an inverted microscope. The methanol and detached cell debris were transferred to 2.0 mL microcentrifuge tubes, which were centrifuged for five minutes at 15,294 g at 4 °C to separate the metabolite extract from the cell debris. The cell debris was then washed with 500 μL of 80%/20% methanol/water and then with 500 μL of 100% ddH₂O. Supernatants from each of the three extractions were finally combined into 2.0 mL microcentrifuge tubes.

Cell extract samples were then split into two portions: 1.8 mL for NMR analysis and 200 μL for MS analysis (Figure 1A). The MS portions were diluted tenfold with a solution of H₂O/methanol/FA (49.75:49.75:0.5) containing 20 μM reserpine as an internal mass reference. The NMR portions were placed in a RotoSpeed vacuum to remove the organic solvent, followed by freezing and lyophilization. Lyophilized NMR-bound metabolite extracts were then resuspended in 600 μL of 50 mM deuterated potassium phosphate buffer at pH 7.2 (uncorrected) containing 50 μM of 3-(trimethylsilyl)propionic acid-2,2,3,3-d₄ (TMSP-d₄) and transferred to 5 mm NMR tubes.

NMR data acquisition and preprocessing

The NMR data was collected and processed according to our previously described protocol (Zhang, Halouska, Gaupp, Lei, Snell, Fenton, Barletta et al. 2013). A Bruker Avance DRX 500 MHz spectrometer equipped with a 5 mm triple-resonance cryogenic probe (¹H, ¹³C, ¹⁵N) with a z-axis gradient, BACS-120 sample changer, and an automatic tuning and matching accessory was utilized for automated NMR data collection.

Following acquisition, the 1D ¹H NMR spectra were processed in our MVAPACK software suite (<http://bionmr.unl.edu/mvapack.php>), which provides a uniform data handling environment that is highly tuned for NMR chemometrics (Worley, Powers 2014a). A 1.0 Hz exponential apodization function and a single round of zero-filling were applied prior to Fourier transformation. Spectra were then automatically phased and normalized using phase-scatter correction (PSC) (Worley, Powers 2014b). Finally, chemical shift regions containing spectral baseline noise or solvent signals were manually removed. Binning of the processed NMR spectra was performed using an adaptive intelligent binning algorithm that minimizes splitting signals between multiple bins (De Meyer, Sinnaeve, Van Gasse, Tsiporkova, Rietzschel, De Buyzere, Gillebert et al. 2008).

Direct infusion ESI-Q-TOF-MS acquisition and preprocessing

Standard metabolite mixture and E. coli metabolome extracts—The standard metabolite mixtures were used to initially optimize DI-ESI-MS source conditions. DI-ESI-MS data were collected on a Synapt G2 HDMS quadrupole time-of-flight MS instrument (Waters, Milford, MA) equipped with an ESI source. All MS experiments were carried out at a flow rate of 10 $\mu\text{L}/\text{min}$ for 1 min and mass spectra were acquired over a mass range of m/z 50 to 1200.

Human dopaminergic neuroblastoma cell extracts—Mass spectra of the SK-N-SH samples were acquired in positive ion mode over a mass range of m/z 50 to 1200. Spectra were acquired for 0.5 min using the following optimized source conditions: 2.5 kV for ESI capillary voltage, 60 V for sampling cone voltage, 4.0 V for extraction voltage, 80 °C for source temperature, 250 °C for desolvation temperature, 500 L/h for desolvation gas, and 15 $\mu\text{L}/\text{min}$ flow rate of injection.

The initial stages of mass spectral data processing were performed using MassLynx V4.1 (Waters Corp., Milford, MA). A background subtraction was performed on all spectra (Figure 1B): reference spectra of either paraquat, MPP^+ , rotenone, or 6-OHDA in $\text{ddH}_2\text{O}/\text{methanol}/\text{FA}$ (49.75:49.75:0.5) at 10 ppm were used as backgrounds. Background subtraction of each spectrum was performed in a class-dependent manner (*e.g.* the MPP^+ MS reference spectrum was used as background for MPP^+ treated cells). As a result, mass spectral signals from the drugs themselves are guaranteed to not influence subsequent analyses. The background-subtracted mass spectra were then loaded into MVAPACK for binning and normalization. All mass spectra were linearly re-interpolated onto a common axis that spanned from m/z 50 to 1200 in 0.003 m/z steps, resulting in 383,334 variables prior to processing. Based on the low probability of observing a metabolite in the mass range m/z 1100 to 1200 (Figure S1), the region was removed prior to binning. Mass spectra were uniformly binned using a bin width of 0.5 m/z , resulting in a data matrix of 2,095 variables. Finally, the MS data matrix observations were normalized using probabilistic quotient (PQ) normalization (Dieterle, Ross, Schlotterbeck, Senn 2006).

Multivariate statistical analysis

Using functions available in the latest version of MVAPACK, the NMR and mass spectral data were joined into a single multiblock data structure and modeled using MB-PCA and MB-PLS. More specifically, the CPCA-W algorithm (Westerhuis, Kourti, Macgregor 1998) was used to generate the MB-PCA model. MB-PLS with super-score deflation (Westerhuis, Coenegracht 1997) was used to generate the MB-PLS model. Both blocks were scaled to unit variance prior to modeling, and equal contribution of each block to the models (fairness) was ensured by further scaling each block by the square root of its variable count (Smilde, Westerhuis, de Jong 2003). For the purposes of comparison, PCA and PLS models of the independent NMR and MS data matrices were also constructed. All PLS models were trained on a binary discriminant response matrix (*i.e.*, PLS-DA), in which untreated cells were assigned to one class and all drug-treated cells were assigned to a second class.

Cross validation of multivariate models

All PCA and MB-PCA models were cross-validated using a leave-one-out (LOOCV) procedure in MVAPACK during model fitting (Lei, Zavala-Flores, Garcia-Garcia, Nandakumar, Huang, Madayiputhiya, Stanton et al. 2014). PLS-DA and MB-PLS-DA models were cross-validated using a Monte Carlo leave-*n*-out (MCCV) procedure (Bove, Prou, Perier, Przedborski 2005). The results of cross-validation were summarized by per-component Q^2 values, where the number of model components was chosen such that cumulative Q^2 was a strictly increasing function of component count. Response permutation tests of all supervised models were performed with 1,000 permutations each to assess the statistical significance of model R^2_Y and Q^2 values (Kamel, Hoppin 2004). CV-ANOVA significance tests (Eriksson, Trygg, Wold 2008) were also performed to supplement the results of the permutation tests. Results of all permutation tests, along with cross-validated scores plots of all supervised models, are provided in the supplementary information (Figures S6–S11).

Metabolite identification by DI-ESI-MS and MS/MS

Metabolite identifications were achieved by obtaining accurate m/z and further verified using MS/MS experiments. A modified static mode nano-electrospray ionization (nESI) source was used for metabolite identification. Samples were loaded into home-pulled borosilicate emitters fabricated from PYREX 100 mm capillary melting point tubes (Corning, Tewksbury, MA, USA). Emitters were pulled with a vertical micropipette puller (David Kopf Instruments, Tujunga, CA, USA). Each emitter was examined under a microscope (American Optical Company, Buffalo, NY, USA) in order to maintain reproducible tip geometries. The capillary was filled with a metabolite extract and placed in a home-made sprayer mounted to the Synapt G2 nESI source XYZ stage such that the capillary potential was applied by a platinum wire in direct contact with the sample solution.

The optimum nESI source parameters were slightly different from those of the normal DI-ESI source. The nESI source parameters were as follows: capillary spray voltage 1.20 kV, sampling cone voltage 40 V, extraction voltage 5 V, and source temperature 80 °C. Spectra were collected for 2 min in positive ion mode over a mass range of m/z 50 to 1200. MS/MS collision energies (CE) were optimized for each metabolite to yield maximum fragmentation. Mass calibration of the instrument was performed by external calibration with sodium acetate, which was infused under the same conditions as the samples. The mass signals of sodium acetate cluster ions (having the general formula $[(C_2H_3O_2Na)_n + Na]^+$) occur every m/z 82.0031. Such cluster ions were used to externally calibrate the instrument from m/z 104.9928 ($n = 1$ cluster) to m/z 1171.0328 ($n = 14$ cluster). All metabolite spectra were smoothed, centroided, and internally mass corrected relative to the $[M+H]^+$ ion for reserpine (m/z 609.2812) using MassLynx V4.1. Accurate m/z values were searched against the following online metabolite MS databases: Human Metabolome Database (HMDB, <http://www.hmdb.ca/>, 41,514 metabolites) (Wishart, Jewison, Guo, Wilson, Knox, Liu, Djoumbou et al. 2013; Wishart, Knox, Guo, Eisner, Young, Gautam, Hau et al. 2009; Wishart, Tzur, Knox, Eisner, Guo, Young, Cheng et al. 2007) and the general Metabolite and Tandem MS Database (METLIN, <http://metlin.scripps.edu>, 242,766 metabolites) (Smith, O'Maille, Want, Qin, Trauger, Brandon, Custodio et al. 2005) with a threshold

window of 20 ppm. The wide window was used to guarantee a thorough search, but a more stringent mass tolerance (~ 1 ppm) was used when making the final assignment.

RESULTS AND DISCUSSION

Optimization of Metabolomics Sample Preparation and MS Source Conditions

Acquisition of DI-ESI-MS data of the highest reliability, reproducibility and information content necessitated the identification of instrumental parameters that yielded maximal ion transmission efficiency. Using a standard metabolite mixture, we first optimized several critical ion source conditions, namely: SCV, ECV, desolvation temperature, desolvation gas flow rate, and cone gas flow rate. Each parameter was sequentially optimized by systematically varying its setting within a predefined range and searching for a maximum ion intensity based on the sum of all detectable spectral signal intensities. After an optimal setting was identified, the parameter value was held fixed while the next parameter was then varied. The process was repeated until an optimized setting was achieved for all five source parameters.

Initial optimization using the standard metabolite mixture indicated that changes to the SCV setting had the largest impact on ion transmission. Ion intensities increased from 5.18×10^3 to 2.69×10^5 for β -alanine, 1.83×10^4 to 1.39×10^6 for glutamine, 1.41×10^5 to 5.46×10^6 for L-histidine, 1.28×10^4 to 6.84×10^6 for caffeine, and 7.21×10^3 to 7.26×10^5 for L-asparagine as the SCV was reduced from 100 V to 20 V. Ibuprofen was not observed in any of the mass spectra. Changes to all of the other source parameters were found to have a minimal impact on ion intensity: no other parameter increased the ion intensities by more than a factor of five. For example, the ion intensity only varied from 7.21×10^5 to 3.03×10^5 for caffeine when the desolvation gas flow was changed from 500 L/h to 1000 L/h. The optimal source parameters for the standard metabolite mixture were determined to be: SCV of 40 V, ECV of 6.0 V, desolvation temperature of 150 °C, desolvation gas flow of 500 L/h, and a cone gas flow of 0 L/h.

Further optimization of the SCV and ECV settings was pursued by applying DI-ESI-MS to a biological matrix of metabolites extracted from Mach1 *E. coli*. Three compounds from the cellular extract were randomly selected based on their equal distribution within the typical mass range (m/z 50 to 1200) of known metabolites. These three compounds had molecular ion peaks corresponding to m/z 118.09, m/z 437.21, and m/z 704.53. The impact of ECV on ion intensity was tested over a range of 2.0 to 10 V, which identified an optimal ECV of 4.0 V. Importantly, the ion intensity of these selected molecular ion peaks were not significantly affected by varying ECV. ECV values optimized against the standard metabolite mixture and the *E. coli* cellular extract were found to be equivalent to within experimental error. The impact of SCV on ion intensity was also tested using *E. coli* cellular extracts. SCV was varied over a range of 20 to 100 V (Figure S2A), which identified an optimal SCV range of 40 to 60 V by examining a subset of ion peak intensities corresponding to m/z 118.09, m/z 437.21, and m/z 704.53 (Figure S2B). An SCV of 40 V, consistent with the value obtained with the standard metabolite mixture, was found to maximize the total information content because intense signals were observed for all metabolites.

Although ion intensity is also positively correlated with metabolite concentrations, simply injecting a highly concentrated metabolite sample may increase the likelihood of ion suppression (Annesley 2003; Skazov, Nekrasov, Kuklin, Simenel 2006) due mostly to increased salt concentrations. The signals from low-mass and polar compounds are more likely to be suppressed by other metabolites as the total sample concentration increases. It was therefore necessary to determine an optimal sample size for DI-ESI-MS analysis of metabolomics samples in order to maximize both ion intensity and information content. *E. coli* cellular extracts that were optimized to maximize signal intensity in NMR-based metabolomics (Zhang, Halouska, Gaupp, Lei, Snell, Fenton, Barletta et al. 2013) were used to determine the optimal sample size for DI-ESI-MS. A mass spectrum (Figure S3A) was collected for a series of sample dilutions (1:10, 1:25, 1:50, and 1:100) in ddH₂O/methanol/FA (49.75:49.75:0.5). For each dilution factor, the total number of spectral peaks above the noise threshold was calculated (Figure S4) and the relative intensities of molecular ion peaks at *m/z* 118.09, 437.21, and 705.53 were monitored. A bar graph summarizing these results is presented in Figure S3B. More spectral signals and higher signal intensities for the three monitored molecular ions were observed for the 10x sample relative to all the other dilution factors. Thus, a ten-fold dilution of a metabolomics sample previously optimized for 1D ¹H NMR experiments was deemed suitable for DI-ESI-MS in our combined MS and NMR metabolic fingerprinting protocol. Put simply, a single sample may be prepared and split for NMR and MS analysis, where the preparation of the MS-bound sample only requires a ten-fold dilution into a compatible solvent, such as ddH₂O/methanol/FA (49.75:49.75:0.5).

NMR and MS data pretreatment

Data preprocessing and pretreatment are critical components of any multivariate statistical analysis and have been extensively reviewed (Worley, Powers 2013). Our protocols for NMR data preprocessing have been previously reported (Halouska, Powers 2006; Zhang, Halouska, Gaupp, Lei, Snell, Fenton, Barletta et al. 2013; Zhang, Halouska, Schiaffo, Sadykov, Somerville, Powers 2011) and were utilized as a basis for NMR data handling.

A minimalistic set of pretreatment steps was pursued for both NMR and MS data matrices. To decrease the time required for computing PCA models, NMR spectra were AI-binned (De Meyer, Sinnaeve, Van Gasse, Tsiporkova, Rietzschel, De Buyzere, Gillebert et al. 2008) and mass spectra were uniformly binned in preparation for PCA. The data were then normalized to correct for random errors in dilution factors or experimental parameters. Binned NMR and mass spectra were normalized using Standard Normal Variate (SNV) and Probabilistic Quotient (PQ) normalization methods, respectively (Dieterle, Ross, Schlotterbeck, Senn 2006). Full-resolution NMR spectral data was used for PLS, which permitted the creation of backscaled pseudo-spectral loadings that greatly enhance model interpretability (Cloarec, Dumas, Craig, Barton, Trygg, Hudson, Blancher et al. 2005; Cloarec, Dumas, Trygg, Craig, Barton, Lindon, Nicholson et al. 2005). However, because the full-resolution mass spectral data matrix contained over 300,000 variables, binned mass spectra were used in PLS modeling to reduce computation time. While the NMR spectra were binned to mask chemical shift variations that reduce the effectiveness of PCA, binning of the mass spectra was performed solely to decrease the time required for model

computation. For the mass spectral data, a uniform bin width of 0.5 m/z was used based on the mass distribution of all metabolites cataloged in the HMDB (Figure S1). As noise is known to decrease the interpretability of scores (Halouska, Powers 2006), all spectral regions found by manual inspection to contain only baseline noise were removed prior to modeling. The final NMR data matrices for PCA and PLS contained 159 and 16,138 variables, respectively. Likewise, the MS data matrix contained 2,095 (PCA and PLS) variables.

Despite the marked reduction in MS variable count incurred from uniform binning, the resulting binned data matrix retained enough information to differentiate signals arising from distinct metabolites. Based on available data of accurate m/z of metabolites in the HMDB, 85% of metabolites have an m/z difference with their “neighbor” metabolites (the metabolites with the closest m/z) greater than 0.5 m/z (Figure S1). Also, the number of metabolites identified from a direct-infusion MS analysis of cell extracts generally ranges from 200 to 400 metabolites (Draper, Lloyd, Goodacre, Beckmann 2014), which would be divided over 2,095 bins. This implies that on average the 0.5 m/z bin size would be expected to capture no more than one mass isotopic distribution or one metabolite per bin.

Classical PCA and PLS modeling

PCA of the binned NMR data matrix ($N = 29$, $K = 159$) resulted in 10 principal components having cumulative R^2_X (degree of fit) and Q^2 (predictive ability) metrics of 0.95 and 0.46, respectively. Overall, no patterns were readily discernable in the NMR PCA scores (Figure 2A) due to high within-class variation in the data. However, scores for paraquat treatment were found to significantly separate from all other classes ($p < 0.002$) along PC1 (Worley, Halouska, Powers 2013). Scores from PCA of the binned MS data matrix ($N = 29$, $K = 2,095$) were found to exhibit markedly less within-class variation compared to the NMR data (Figure 2B). Three significant components were identified from the binned MS data, yielding fairly low cumulative R^2_X and Q^2 metrics of 0.34 and 0.16. While paraquat treatment still separated from other drug treatments in MS PCA scores space, the greatest separations were observed between treated and untreated cells ($p < 1.5 \times 10^{-9}$). These differing patterns of separation in NMR and MS PCA scores suggested that multiblock analyses could provide further information, ideally separating both control and paraquat scores from all other classes.

PLS-DA of the full-resolution NMR ($N = 29$, $K = 16,138$) and MS ($N = 29$, $K = 2,095$) data matrices both resulted in two-component models. With the exception of the algorithmically forced separation between control and treatment classes, similar clustering patterns were observed when compared to the PCA scores. Leave- n -out cross-validation metrics from the NMR ($R^2_Y = 0.92$, $Q^2 = 0.64$) and MS ($R^2_Y = 0.99$, $Q^2 = 0.93$) PLS-DA models indicated reasonable levels of fit and predictive ability. Further validation by CV-ANOVA (Eriksson, Trygg, Wold 2008) indicated reliable models with p values of 1.5×10^{-6} and 2.9×10^{-15} for NMR and MS data, respectively. Response permutation tests for both PLS-DA models returned p values equal to zero, supporting the CV-ANOVA significance test results.

Multiblock PCA and PLS modeling

Identification of consensus directions in the NMR and MS data matrices that maximally captured data matrix variations (MB-PCA) or data-response correlations (MB-PLS) resulted in more informative models than those calculated against either NMR or MS *in vacuo*. Using MB-PCA, five significant components were identified ($Q^2 = 0.23$) that cumulatively explained comparable amounts of variation in the NMR ($R^2_X = 0.85$) and MS ($R^2_X = 0.50$) blocks relative to the individual PCA models. As expected, MB-PCA combined the information from both blocks to dramatically increase class separations in super-scores space (Figure 2C). More specifically, both control and paraquat classes were separated from other drug treatments, predominantly along PC1. Furthermore, MPP⁺ treatment exhibited significant separation from 6-OHDA and Rotenone treatments, which was not expected from examination of the individual NMR or MS PCA scores.

MB-PLS of the data yielded similar improvements in model information content. Two significant components were identified ($R^2_Y = 0.98$, $Q^2 = 0.89$) that clearly separated untreated and paraquat treatment classes from all other classes in scores space. CV-ANOVA testing resulted in a p value of 1.7×10^{-12} and response permutation testing yielded a p value equal to zero, indicating a reliable MB-PLS-DA model.

Metabolite identification by combining NMR, accurate mass and MS/MS

Identification of key metabolites from a metabolomics sample is undoubtedly a nontrivial undertaking. The difficulties of metabolite assignment are further compounded by spectral overlap present in both 1D ¹H NMR and DI-ESI-MS data. However, the combination of these two complementary forms of spectral information may aid in overcoming the ambiguities encountered during assignment. The ability of our approach to aid in metabolite identification was demonstrated using NMR and MS data obtained from the treatment of human dopaminergic neuroblastoma cells with known neurotoxic agents.

A first-pass identification of biologically important metabolites was performed by examination of the backscaled NMR block loadings from MB-PLS-DA (Figure 3A). ¹H NMR chemical shifts of loading 'signals' that contributed significantly to class separation were used to query the Human Metabolome Database (HMDB) (Wishart, Jewison, Guo, Wilson, Knox, Liu, Djoumbou et al. 2013) for matching metabolites. To confirm the NMR results, accurate mass and MS/MS experiments were performed, guided by information obtained from the backscaled MS block loadings (Figure 3B). Effectively, the MB-PLS-DA MS block loadings identified specific masses to pursue for more focused and detailed analyses. For accurate mass measurements, reserpine was used as an internal m/z reference, because it is located in a region containing a minimal number of known metabolites and it can be ionized in both positive and negative modes. Accurate masses were used to conduct elemental composition analyses with MassLynx 4.1 based on a restricted list of elements (C, H, O, N, P, S, Na and K), resulting in a set of possible molecular formulas and associated compounds. Metabolite assignments consistent with both accurate mass and NMR chemical shifts were retained for further study by collision-induced dissociation MS/MS. It is noteworthy that many compounds were identified as sodium adducts, which is expected for

DI-ESI-MS (Lin, Yu, Yan, Hang, Zheng, Xing, Huang 2010), particularly when sample purification steps have been kept to a minimum.

As an illustration, examination of backscaled MS block loadings identified a significantly increased mass spectral signal at m/z 203.0534. This accurate mass is consistent with the molecular formula $C_6H_{12}O_6Na$ (theoretical exact $m/z = 203.0532$) to within 1 ppm. The corresponding potassium adduct of $C_6H_{12}O_6$ was also observed at m/z 219.0266, which is within 2.5 ppm of theoretical exact m/z 219.0271. The molecular ion peak m/z 203.0534 was selected for further examination by MS/MS using collision-induced dissociation, which yielded product ions of m/z 123, 141, and 159, among others (Figure 4). These three peaks were assigned to $[C_6H_{12}O_6-CO_2+Na]^+$, $[C_6H_{12}O_6-CO_2-H_2O+Na]^+$ and $[C_6H_{12}O_6-CO_2-2H_2O+Na]^+$ respectively. Together with the elemental composition suggested by accurate mass measurement, the neutral losses of CO_2 and multiple H_2O indicate a likelihood that the precursor ion was a polyhydroxy carboxylic acid. Possible structures are inset in Figure 4. Using these procedures, our NMR analysis of human dopaminergic neuroblastoma cells treated with paraquat identified an increase in metabolites associated with the Pentose Phosphate Pathway (PPP) (Lei, Zavala-Flores, Garcia-Garcia, Nandakumar, Huang, Madayiputhiya, Stanton et al. 2014).

CONCLUSIONS

We report an optimized protocol for combining 1D 1H NMR and DI-ESI-MS datasets for the purposes of high-throughput metabolic fingerprinting and profiling. By splitting metabolite extracts optimized for NMR acquisition and diluting the MS-bound aliquots ten-fold in H_2O /methanol/FA (49.57:49.75:0.5), we obtained samples suitable for direct infusion electrospray ionization, thus avoiding the use of pre-source chromatographic separations. We also optimized several DI-ESI-MS ion source conditions in order to maximize the quality of the MS metabolomics data. The optimal source parameters were determined to be: SCV of 40 V, ECV of 4.0 V, desolvation temperature of 150 °C, desolvation gas flow of 500 L/h, and a cone gas flow of 0 L/h. The acquired mass spectra were preprocessed with background subtraction, followed by uniform binning with a 0.5 Da bin size and spectral noise region removal. Using multiblock bilinear factorization algorithms that capitalize on the availability of blocking information, we achieved greater levels of model interpretability with the NMR and MS data than available from single-block PCA and PLS methods. We also demonstrated the combined use of NMR and MS/MS data for the rapid and accurate identification of metabolites significantly perturbed in backscaled MB-PLS-DA loadings. In summary, we present a unique means of increasing metabolome coverage in a high throughput manner by leveraging the complementary information provided by MS and NMR, without the encumbrances and liabilities of pre-source chromatographic separations. Our protocol for using combined NMR and MS metabolomics data was successfully demonstrated on a study examining the impact of neurotoxins known to induce dopaminergic cell death, an important model relevant to Parkinson's disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Jiantao Guo for providing us with the *Escherichia coli* Mach1 cell line. This manuscript was supported in part by funds from the National Institute of Health (R01 AI087668, R21 AI087561, R01 CA163649, P20 RR-17675, P30 GM103335), the University of Nebraska, the Nebraska Tobacco Settlement Biomedical Research Development Fund, and the Nebraska Research Council. The research was performed in facilities renovated with support from the National Institutes of Health (RR015468-01).

References

- Annesley TM. Ion suppression in mass spectrometry. *Clin Chem.* 2003; 49:1041–1044.10.1373/49.7.1041 [PubMed: 12816898]
- Atherton HJ, et al. A combined 1H-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR- α null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome. *Physiol Genomics.* 2006; 27:178–186.10.1152/physiolgenomics.00060.2006 [PubMed: 16868074]
- Barding GA, Beni S, Fukao T, Bailey-Serres J, Larive CK. Comparison of GC-MS and NMR for Metabolite Profiling of Rice Subjected to Submergence Stress. *J Proteome Res.* 2013; 12:898–909.10.1021/pr300953k [PubMed: 23205590]
- Beltran A, et al. Assessment of Compatibility between Extraction Methods for NMR- and LC/MS-Based Metabolomics. *Anal Chem.* 2012; 84:5838–5844.10.1021/ac3005567 [PubMed: 22697410]
- Bove J, Prou D, Perier C, Przedborski S. Toxin-induced models of Parkinson's disease. *NeuroRx.* 2005; 2:484–94. [PubMed: 16389312]
- Canelas AB, et al. Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal Chem.* 2009; 81:7379–89.10.1021/ac900999t [PubMed: 19653633]
- Chen H, Pan Z, Talaty N, Raftery D, Cooks RG. Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation. *Rapid Communications in Mass Spectrometry.* 2006; 20:1577–1584.10.1002/rcm.2474 [PubMed: 16628593]
- Cloarec O, et al. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Analytical Chemistry.* 2005; 77:1282–1289.10.1021/Ac048630x [PubMed: 15732908]
- Cloarec O, et al. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in H-1 NMR spectroscopic metabolomic studies. *Analytical Chemistry.* 2005; 77:517–526.10.1021/Ac048803i [PubMed: 15649048]
- Crockford DJ, et al. Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabolomic toxicology studies. *Anal Chem.* 2006; 78:363–71.10.1021/ac051444m [PubMed: 16408915]
- Dai H, Xiao C, Liu H, Hao F, Tang H. Combined NMR and LC-DAD-MS Analysis Reveals Comprehensive Metabonomic Variations for Three Phenotypic Cultivars of *Salvia miltiorrhiza* Bunge. *J Proteome Res.* 2010; 9:1565–1578.10.1021/pr901045c [PubMed: 20067324]
- De Meyer T, et al. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Analytical Chemistry.* 2008; 80:3783–3790.10.1021/Ac7025964 [PubMed: 18419139]
- Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev.* 2007; 26:51–78.10.1002/mas.20108 [PubMed: 16921475]
- Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabolomics. *Analytical Chemistry.* 2006; 78:4281–4290.10.1021/Ac051632c [PubMed: 16808434]
- Draper J, Lloyd AJ, Goodacre R, Beckmann M. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics.* 2014; 9:4–29.10.1007/s11306-012-0449-x
- Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS (R) models. *Journal of Chemometrics.* 2008; 22:594–600.10.1002/Cem.1187

- Gu H, Pan Z, Xi B, Asiago V, Musselman B, Raftery D. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: Application to the detection of breast cancer. *Anal Chim Acta*. 2011; 686:57–63.10.1016/j.aca.2010.11.040 [PubMed: 21237308]
- Halouska S, Powers R. Negative impact of noise on the principal component analysis of NMR data. *Journal of Magnetic Resonance*. 2006; 178:88–95.10.1016/j.jmr.2005.08.016 [PubMed: 16198132]
- Jung J-Y, Jung Y, Kim J-S, Ryu DH, Hwang G-S. Assessment of peeling of Astragalus roots using 1H NMR- and UPLC-MS-based metabolite profiling. *J Agric Food Chem*. 2013; 61:10398–10407.10.1021/jf4026103 [PubMed: 24073592]
- Kamel F, Hoppin JA. Association of pesticide exposure with neurologic dysfunction and disease. *Environ Health Perspect*. 2004; 112:950–958.10.1289/ehp.7135 [PubMed: 15198914]
- Kanani H, Chrysanthopoulos PK, Klapa MI. Standardizing GC-MS metabolomics. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2008; 871:191–201.10.1016/j.jchromb.2008.04.049
- Kell DB. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol*. 2004; 7:296–307.10.1016/j.mib.2004.04.012 [PubMed: 15196499]
- Kopka J. Current challenges and developments in GC-MS based metabolite profiling technology. *J Biotechnol*. 2006; 124:312–322.10.1016/j.jbiotec.2005.12.012 [PubMed: 16434119]
- Kuehnbaum NL, Britz-McKibbin P. New Advances in Separation Science for Metabolomics: Resolving Chemical Diversity in a Post-Genomic Era. *Chem Rev*. 2013; 113:2437–2468.10.1021/cr300484s [PubMed: 23506082]
- Lange E, Tautenhahn R, Neumann S, Gropl C. Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. *BMC Bioinformatics*. 2008; 9:375. [PubMed: 18793413]
- Lei S, et al. Alterations in energy/redox metabolism induced by mitochondrial and environmental toxins: A specific role for Glucose-6-Phosphate-Dehydrogenase and the Pentose Phosphate Pathway in Paraquat Toxicity. *ACS Chem Biol*. 2014 under revision.
- Lenz EM, Wilson ID. Analytical Strategies in Metabolomics. *J Proteome Res*. 2007; 6:443–458.10.1021/pr0605217 [PubMed: 17269702]
- Lin L, et al. Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabolomics? A serum metabolomic study of kidney cancer. *The Analyst*. 2010; 135:2970.10.1039/c0an00265h [PubMed: 20856980]
- Metz TO, et al. High-resolution separations and improved ion production and transmission in metabolomics. *TrAC, Trends Anal Chem*. 2008; 27:205–214.10.1016/j.trac.2007.11.003
- Mullen AR, et al. Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature*. 2012; 481:385–8.10.1038/nature10642 [PubMed: 22101431]
- Nicholson JK, Lindon JC, Holmes E. “Metabonomics”: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. 1999; 29:1181–1189. [PubMed: 10598751]
- Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem*. 2007; 387:525–527.10.1007/s00216-006-0687-8 [PubMed: 16955259]
- Skazov RS, Nekrasov YS, Kuklin SA, Simenel AA. Influence of experimental conditions on electrospray ionization mass spectrometry of ferrocenylalkylazoles. *Eur J Mass Spectrom*. 2006; 12:137–142.10.1255/ejms.795
- Smilde AK, Westerhuis JA, de Jong S. A framework for sequential multiblock component methods. *J Chemom*. 2003; 17:323–337.10.1002/cem.811
- Smith CA, et al. METLIN: a metabolite mass spectral database. *Therapeutic drug monitoring*. 2005; 27:747–51. [PubMed: 16404815]
- t’Kindt R, et al. Metabolomics to unveil and understand phenotypic diversity between pathogen populations. *PLoS Negl Trop Dis*. 2010; 4:e904.10.1371/journal.pntd.0000904 [PubMed: 21152055]

- Taylor PJ. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clin Biochem.* 2005; 38:328–334.10.1016/j.clinbiochem.2004.11.007 [PubMed: 15766734]
- Westerhuis JA, Coenegracht PMJ. Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares. *Journal of Chemometrics.* 1997; 11:379–392.10.1002/(Sici)1099-128x(199709/10)11:5<379::Aid-Cem482>3.3.Co;2-#
- Westerhuis JA, Kourti T, Macgregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J Chemom.* 1998; 12:301–321.10.1002/(sici)1099-128x(199809/10)12:5<301::aid-cem515>3.0.co;2-s
- Wishart DS, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res.* 2013; 41:D801–7.10.1093/nar/gks1065 [PubMed: 23161693]
- Wishart DS, et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 2009; 37:D603–10.10.1093/nar/gkn810 [PubMed: 18953024]
- Wishart DS, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 2007; 35:D521–6.35/suppl_1/D521. [PubMed: 17202168]
- Wold, S. PLS modeling with latent variables in two or more dimensions. 1987.
- Worley B, Halouska S, Powers R. Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical Biochemistry.* 2013; 433:102–104.10.1016/j.ab.2012.10.011 [PubMed: 23079505]
- Worley B, Powers R. Multivariate analysis in metabolomics. *Curr Metabolomics.* 2013; 1:92–107.10.2174/2213235x11301010092
- Worley B, Powers R. MVAPACK: A Complete Data Handling Package for NMR Metabolomics. *ACS Chemical Biology.* 2014a; 9:1138–1144.10.1021/cb4008937 [PubMed: 24576144]
- Worley B, Powers R. Simultaneous phase and scatter correction for NMR datasets. *Chemometrics and Intelligent Laboratory Systems.* 2014b; 131:1–6.10.1016/j.chemolab.2013.11.005 [PubMed: 24489421]
- Xu F, Zou L, Ong CN. Multiorigin of Chromatographic Peaks in Derivatized GC/MS Metabolomics: A Confounder That Influences Metabolic Pathway Interpretation. *J Proteome Res.* 2009; 8:5657–5665.10.1021/pr900738b [PubMed: 19835423]
- Xu Y, Correa E, Goodacre R. Integrating multiple analytical platforms and chemometrics for comprehensive metabolic profiling: application to meat spoilage detection. *Anal Bioanal Chem.* 2013; 405:5063–5074.10.1007/s00216-013-6884-3 [PubMed: 23512189]
- Zhang B, et al. Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes. *Journal of Integrated OMICS.* 2013; 2:120–137.
- Zhang B, Halouska S, Schiaffo CE, Sadykov MR, Somerville GA, Powers R. NMR Analysis of a Stress Response Metabolic Signaling Network. *J Proteome Res.* 2011; 10:3743–3754.10.1021/pr200360w [PubMed: 21692534]

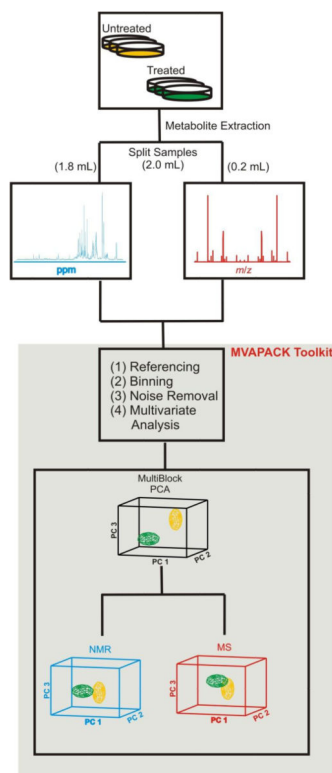


Figure 1.

A flow chart illustrating our protocol for combining NMR and MS datasets for metabolomics. A) 2.0 mL of a single metabolite extract was split into 1.8 mL and 0.2 mL for NMR and MS analysis, respectively. B) Spectral binning of the NMR data used adaptive intelligent binning. First, the background is subtracted from the MS spectrum followed by spectral binning. Spectral binning of the MS data used fixed binning with a set bin width of 0.5 m/z . C) Baseline noise removal and normalization separately applied to the NMR and MS data sets. D) The NMR and MS datasets were modeled by MB-PCA and MB-PLS. E) The resulting block scores and loadings were analyzed for significantly contributing metabolites.

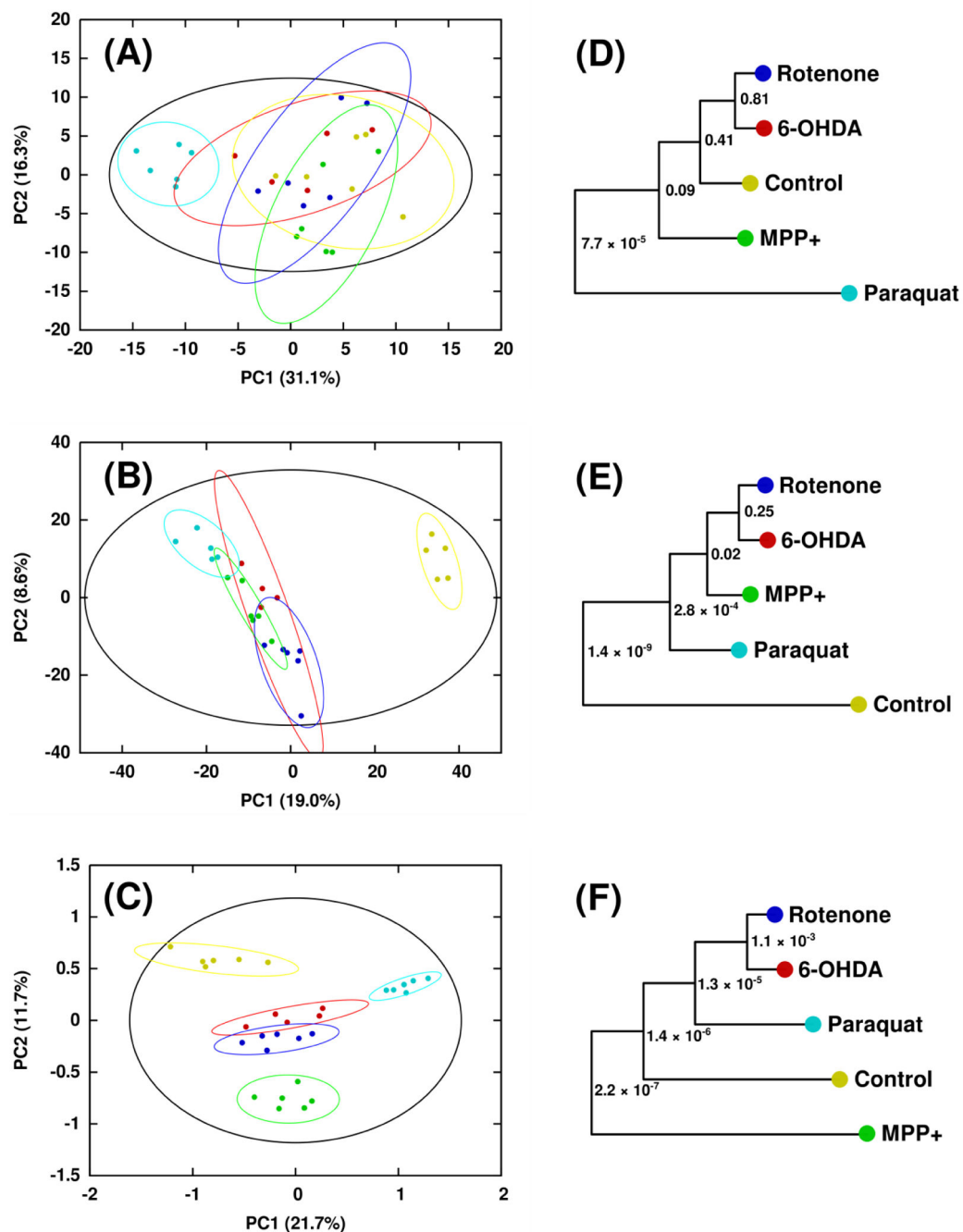


Figure 2. Scores generated from (A) PCA of ^1H NMR *in vacuo*, (B) PCA of DI-ESI-MS *in vacuo*, and (C) MB-PCA of ^1H NMR and DI-ESI-MS. Separations between classes are greatly increased upon combination of the two datasets via MB-PCA. Symbols designate the following classes: Control (●), Rotenone (●), 6-OHDA (●), MPP+ (●), and Paraquat (●). Corresponding dendrograms are shown in (D–F). The statistical significance of each node in the dendrogram is indicated by a p value (Worley, Halouska, Powers 2013).

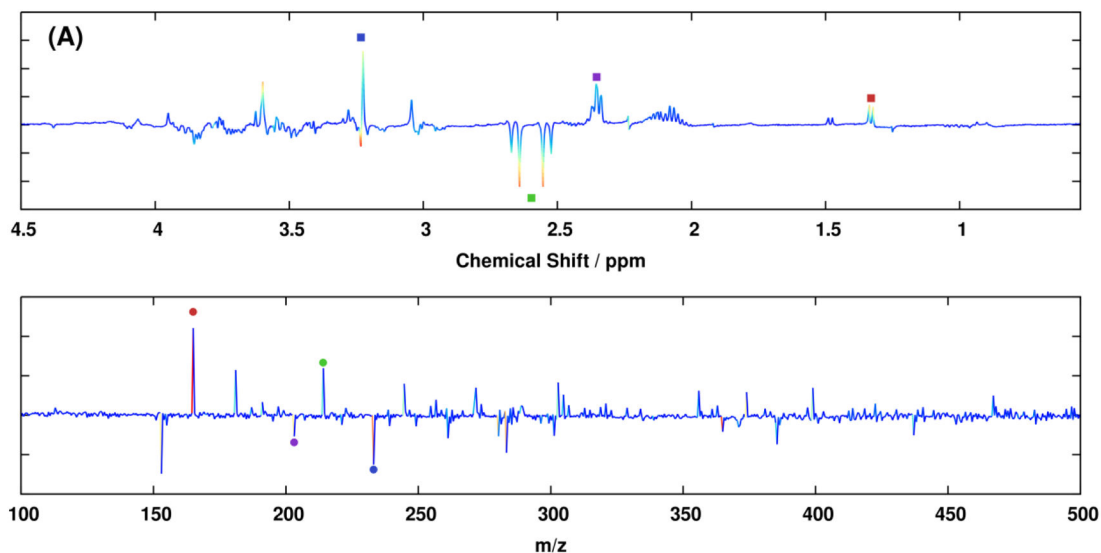


Figure 3. Backscaled MB-PLS-DA first component loadings generated from (A) the ^1H NMR block and (B) the DI-ESI-MS block that compare control with drug treatment. The peaks in the loadings are labeled with the same colored symbol (^1H NMR, square; MS, circle) and were assigned to the following metabolites: lactate (■), glutamate (■), hexose (■), citrate (■), heptose (●), hexose (●), phosphoaspartate (●), and an ambiguous metabolite (●).

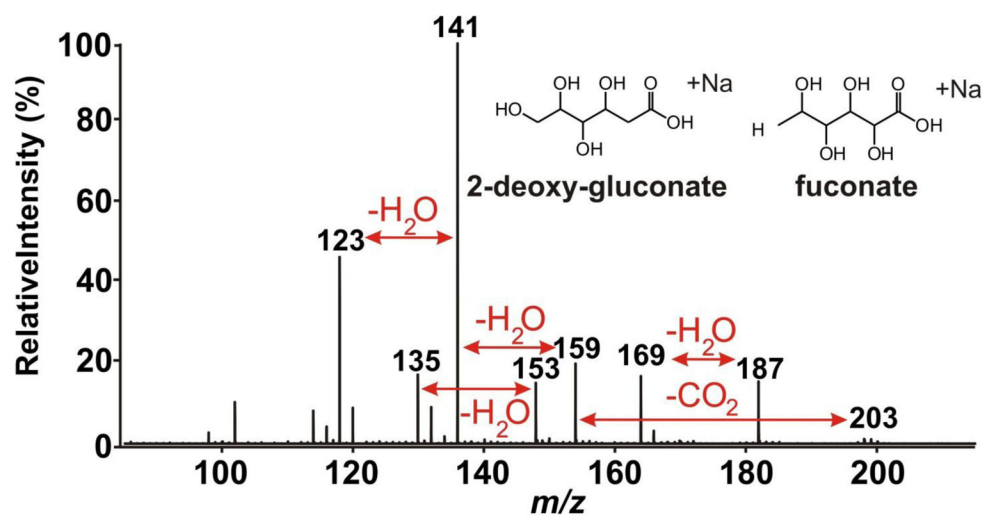


Figure 4. Direct injection *static*-nESI-MS/MS CID spectrum of m/z 203. Fragment ions at m/z 123, 141, and 159 are consistent with a polyhydroxy carboxylic acid, such as 2-deoxy-gluconate or fuconate (inset). A complete assignment of all peaks to the putative structure(s) was not possible, most likely due to selection of multiple multiple isobaric and / or isomeric species prior to dissociation.

Supplementary Information for:

Combining DI-ESI-MS and NMR Datasets for Metabolic Profiling

Darrell D. Marshall,¹ Shulei Lei,¹ Bradley Worley,¹ Yuting Huang,¹ Aracely Garcia-Garcia,^{2,3}

Rodrigo Franco,^{2,3} Eric D. Dodds,¹ and Robert Powers^{1,2*}

¹Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304

*²Redox Biology Center, and ³School of Veterinary Medicine and Biomedical Sciences, University
of Nebraska-Lincoln, Lincoln, NE 68583-0905*

TABLE OF CONTENTS

Supplementary Methods. Further description of procedures used to prepare metabolomics samples.

Figure S1. Mass distribution of known metabolites in the Human Metabolome Database.

Figure S2. Effects of sampling cone voltage on total ion intensity in DI-ESI-MS experiments.

Figure S3. Effects of sample dilution on total ion intensity in DI-ESI-MS experiments.

Figure S4. Number of detected peaks in DI-ESI-MS spectra versus dilution factor.

Figure S5. Representative spectra from DI-ESI-MS and 1D ^1H NMR experiments.

Figures S6 - S8. PLS-DA and MB-PLS-DA cross-validated scores for NMR, MS and combined datasets.

Figures S9 – S11. Response permutation test results for NMR, MS and combined PLS-DA and MB-PLS-DA models.

SUPPLEMENTARY METHODS

Preparation of standard bacterial metabolomics samples

Escherichia coli Mach1 starter culture was grown at 37 °C in LB media. The culture was shaken at 200 rpm for 12 hours and grown to an approximate O.D.₆₀₀ of 12.0. Then 1 mL of the starter culture was added to three 250 mL flasks containing 25 mL of LB media at 37 °C. The bacteria were allowed to grow to a stationary growth phase (6 h) with an O.D.₆₀₀ of 10.75. *E. coli* cells were then harvested by transferring 5 mL from each 250 mL flask to three corresponding 15 mL falcon tubes, centrifuged for 10 min at 2057 g at 25 °C. The supernatant media was discarded. After two washing steps with PBS, the supernatant was discarded and the cell pellets were washed with 10 mL of cold distilled water three times. The cell pellets were re-suspended with 1 mL of distilled water and transferred to 1.6 mL microcentrifuge tubes and centrifuged at 15,294 g for 5 minutes at 25 °C. The cell pellets were then lysed with a Sonic Dismembrator model 100 (Fisher Scientific) for 5 minutes, then centrifuged for 10 minutes at 15,294 g at 25 °C. 700 µL of the metabolite extract was transferred to a second 1.6 mL microcentrifuge tube. Another extraction was performed on the remaining cell debris by adding 1.0 mL of 50/50 ddH₂O/methanol to the first microcentrifuge tube, which was then centrifuged for 10 minutes at 15,294 g at 25 °C. 800 µL of the supernatant was extracted and combined with the initial 700 µL extract. The resulting 1.5 mL cell lysate was frozen, lyophilized (Freezone 4.5, Kansas City, MO, USA), and then resuspended in 1.0 mL of ddH₂O/methanol/FA (49.75:49.75:0.5). To generate analytical replicates, each of the three metabolomics samples was separated into 3 aliquots of 100 µL and then diluted ten-fold with ddH₂O/methanol/FA (49.75:49.75:0.5) containing 10 µM caffeine as an internal mass reference.

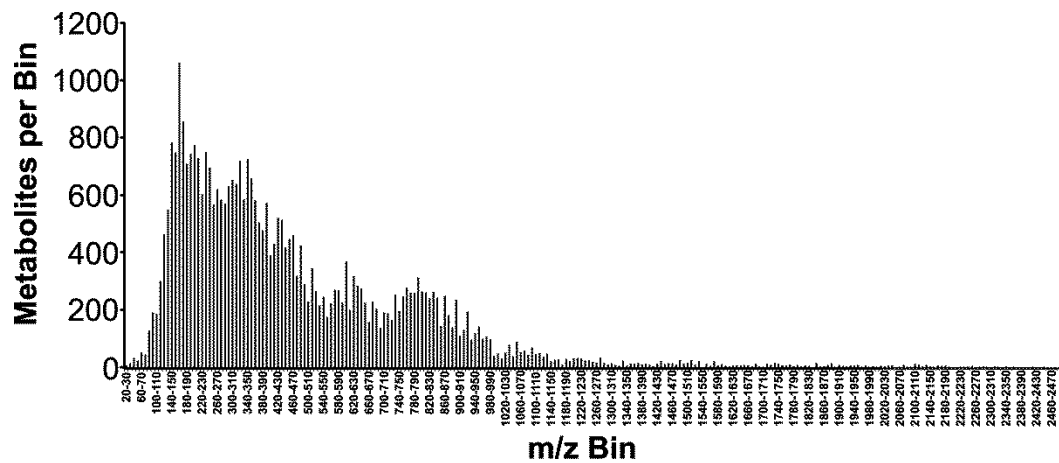


Figure S1. Mass distribution of known metabolites from the Human Metabolome Database (HMDB, <http://www.hmdb.ca/>).

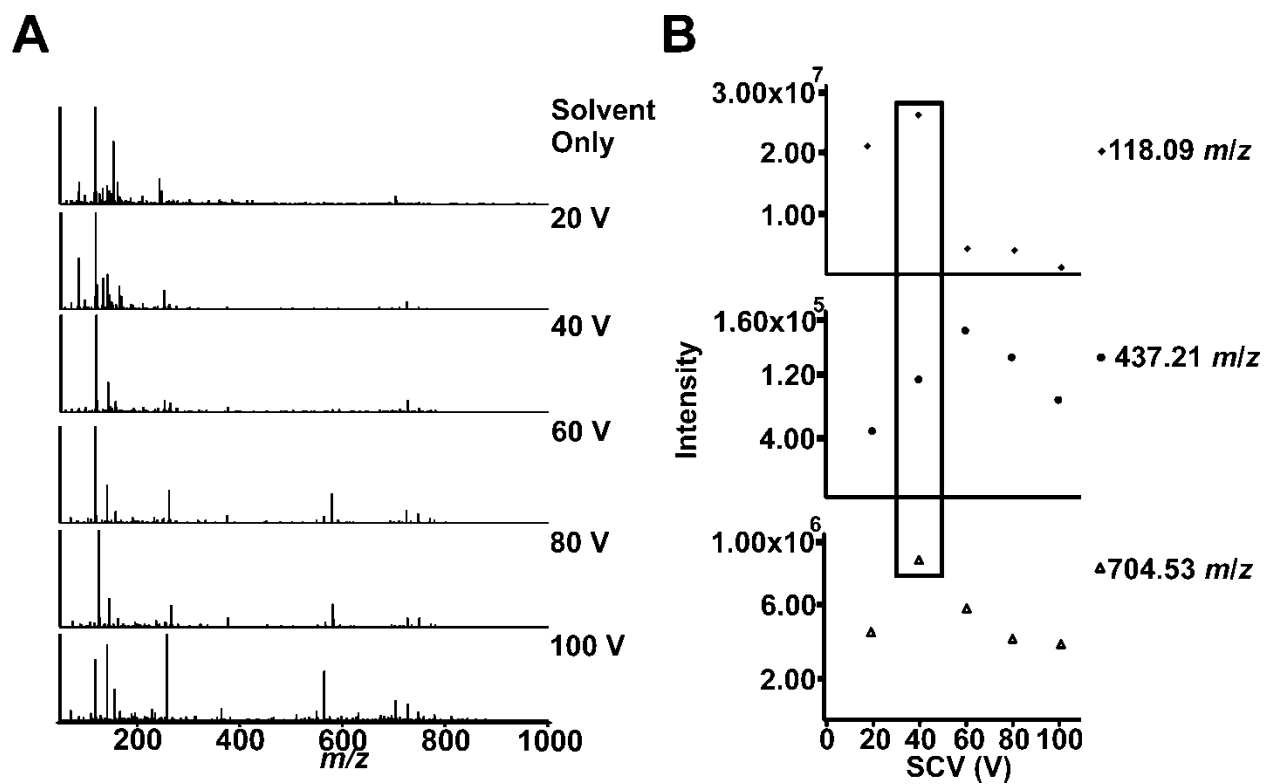


Figure S2. The effects of sampling cone voltage on DI-ESI-MS on total ion count for *E. coli* Mach1 metabolite extract. (A) Mass spectra at varying SCVs (B) Intensity change plotted against SCV for three spectrometric peaks at *m/z* 118.09, 437.21, and 704.53.

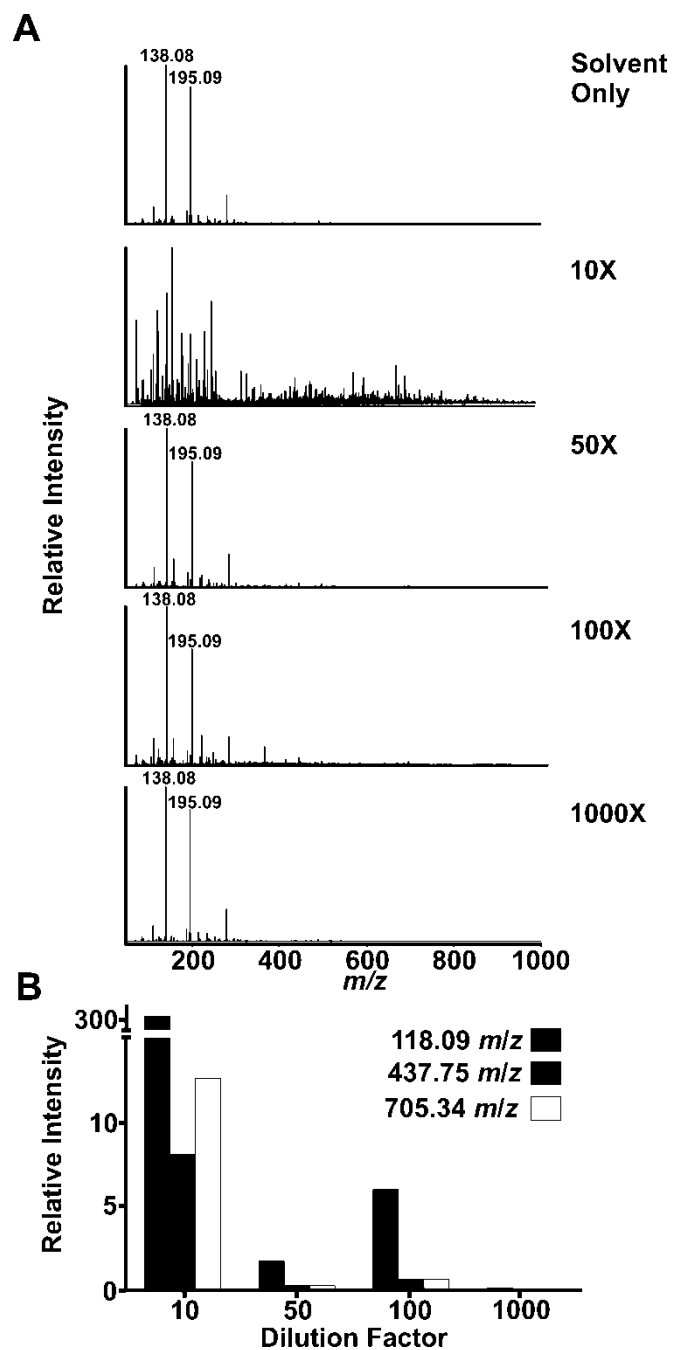


Figure S3. (A) DI-ESI-MS spectra of *E. coli* Mach1 metabolite extract spiked with $[\text{caffeine}+\text{H}]^+$ m/z 195.09 and its primary fragment m/z 138.07 (internal mass reference) at various dilutions; $\text{H}_2\text{O}/\text{methanol}/\text{FA}$ (49.75:49.75:0.5) and 10 μM caffeine (solvent blank), 10 fold dilution, 50 fold dilution, 100 fold dilution, and 1000 fold dilution. (B) Bar graph of relative intensities for 10, 50, 100, 1000 times dilution for three spectrometric peaks at m/z 118.09 (red), 437.21 (blue), and 704.53 (green) normalized to $[\text{caffeine}+\text{H}]^+$.

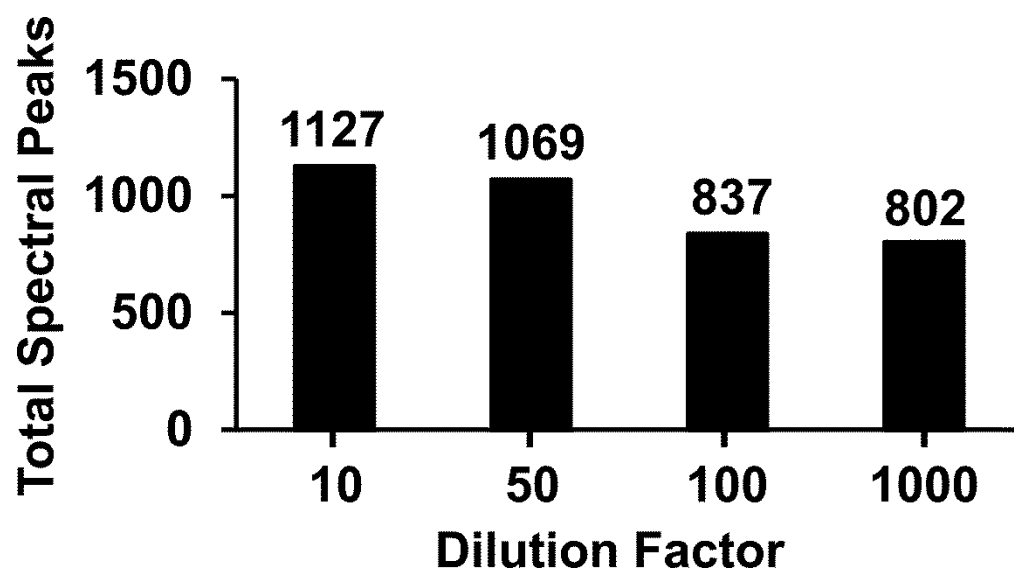


Figure S4. Total number of peaks in the MS spectrum above the noise threshold is plotted as a function of dilution factor.

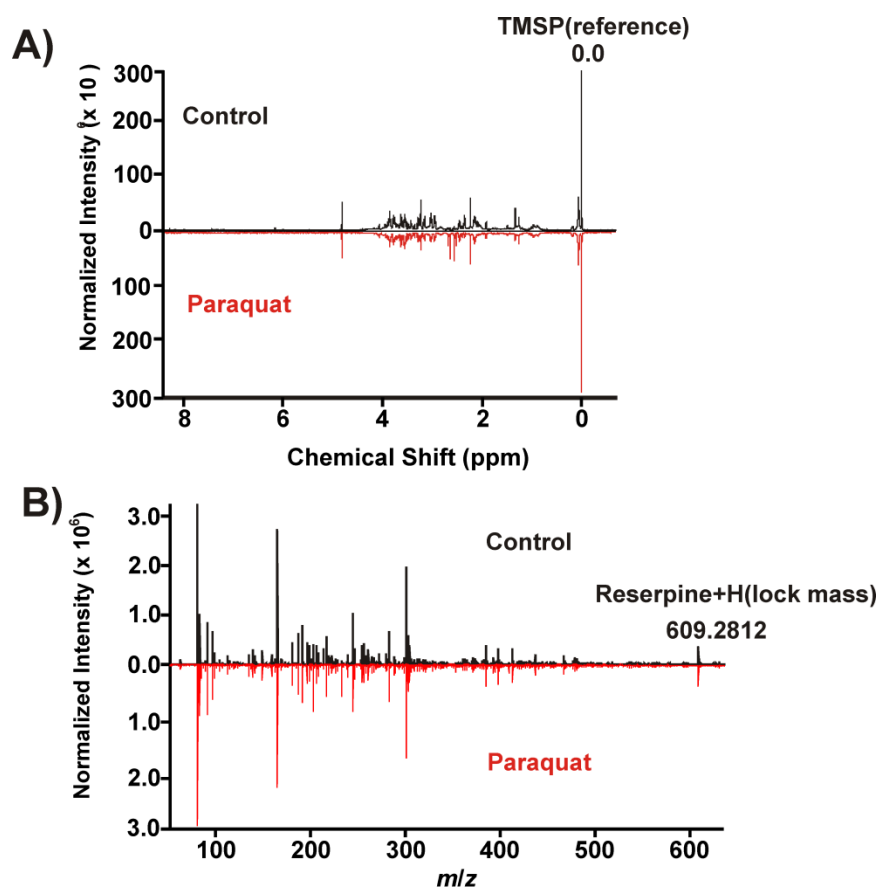


Figure S5. nano-DI-ESI-MS and 1D ^1H NMR spectra of control and paraquat treated samples.

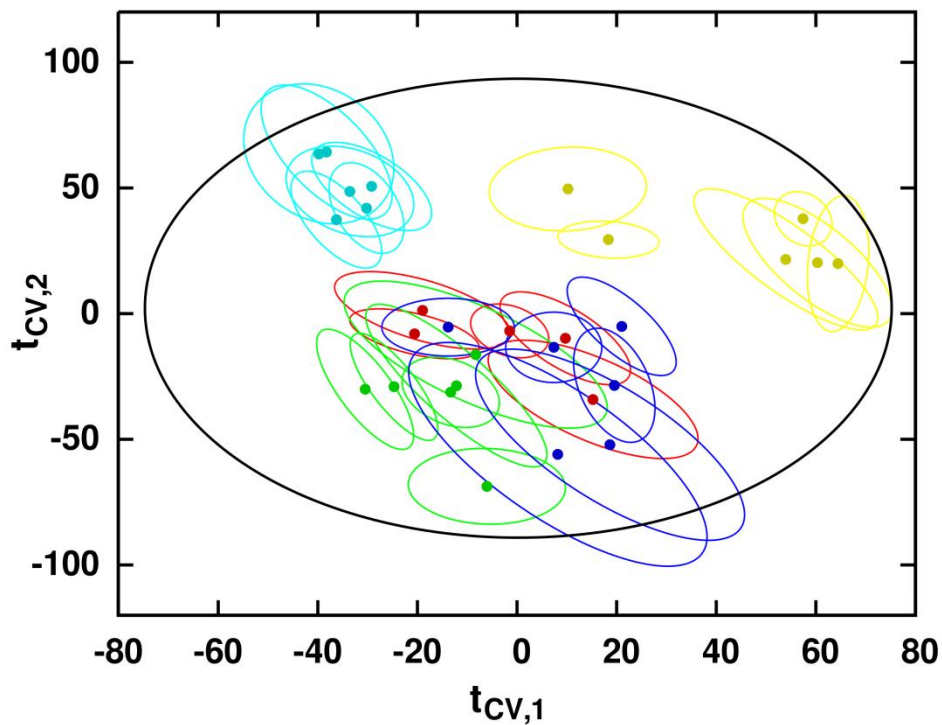


Figure S6. PLS-DA scores computed from single Monte Carlo leave- n -out cross-validation of the *in vacuo* ^1H NMR dataset. Symbol colors designate the following classes: Control (●), Rotenone (●), 6-OHDA (●), MPP+ (●), and Paraquat (●). Symbols designate the mean value of an observation from all iterations of a Monte Carlo cross-validation, and ellipses designate the 95% confidence ellipse for the respective observation.

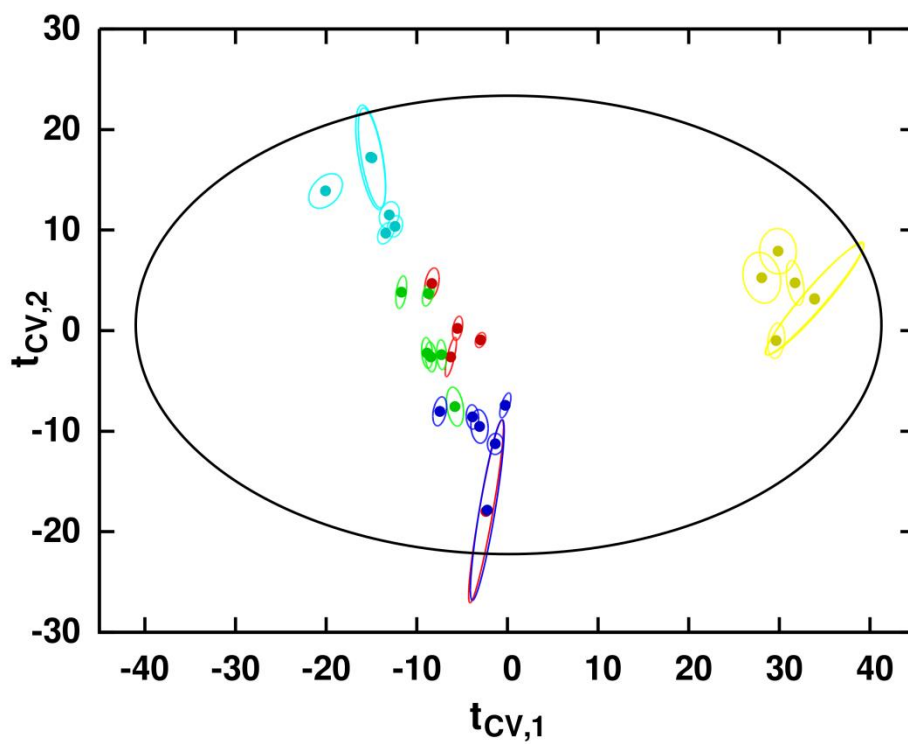


Figure S7. PLS-DA scores computed from a single Monte Carlo leave- n -out cross-validation of the *in vacuo* DI-ESI-MS dataset. See Figure S6 for color, symbol and ellipse definitions.

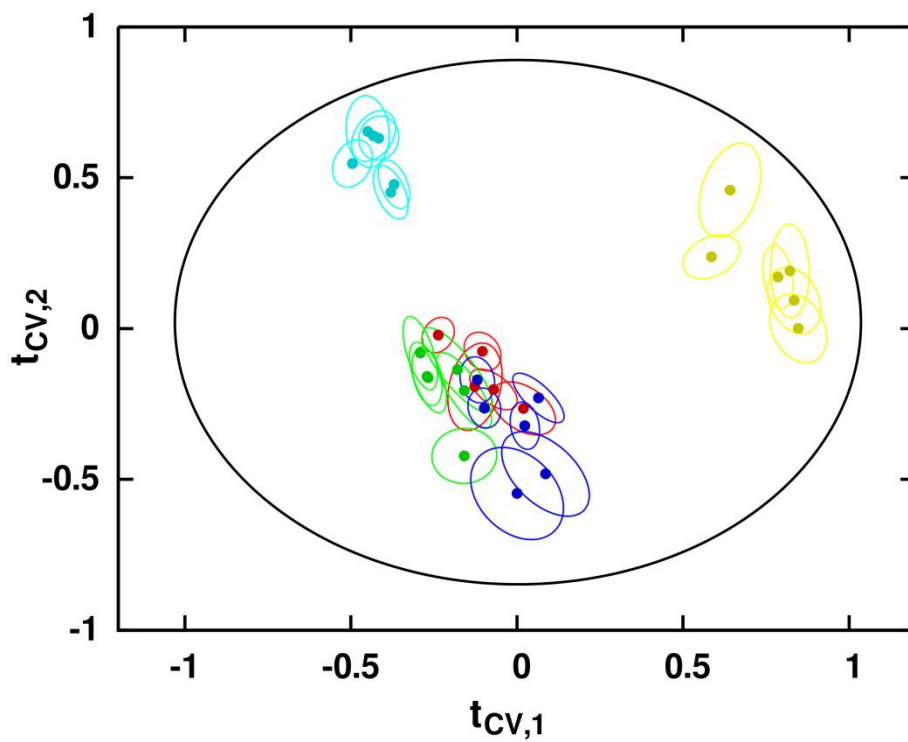


Figure S8. MB-PLS-DA super scores computed from a single Monte Carlo leave- n -out cross-validation of the combined MS and NMR datasets. See Figure S6 for color, symbol and ellipse definitions.

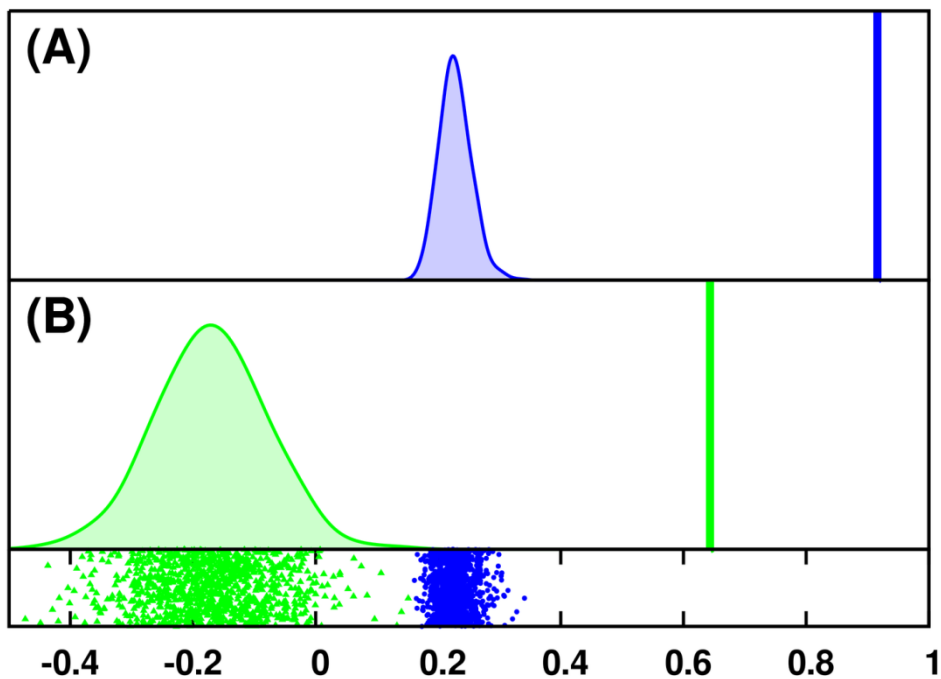


Figure S9. Response permutation test results for the PLS-DA model of the *in vacuo* ^1H NMR dataset. Model fit (R^2) statistics (A) are shown in blue, and model predictive ability (Q^2) statistics are shown in green. True values of R^2 and Q^2 are represented by vertical bars, and null distributions are computed through kernel density estimation of the values from permutation. Scatter plots of the permutation R^2 and Q^2 statistics are shown in the lower pane.

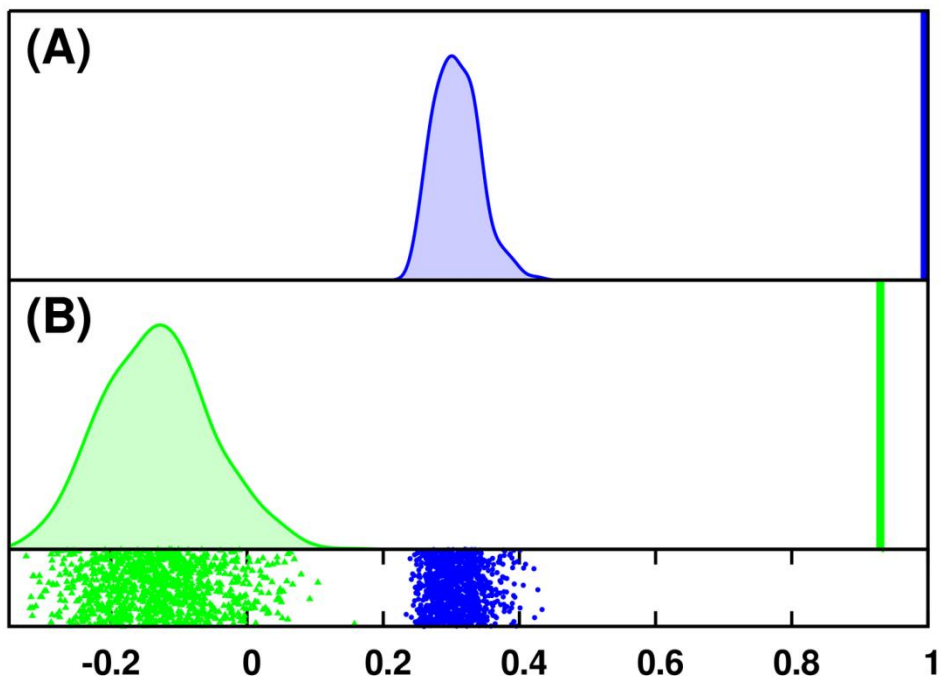


Figure S10. Response permutation test results for the PLS-DA model of the *in vacuo* DI-ESI-MS dataset. Model fit (R^2) statistics (A) are shown in blue, and model predictive ability (Q^2) statistics are shown in green. True values of R^2 and Q^2 are represented by vertical bars, and null distributions are computed through kernel density estimation of the values from permutation. Scatter plots of the permutation R^2 and Q^2 statistics are shown in the lower pane.

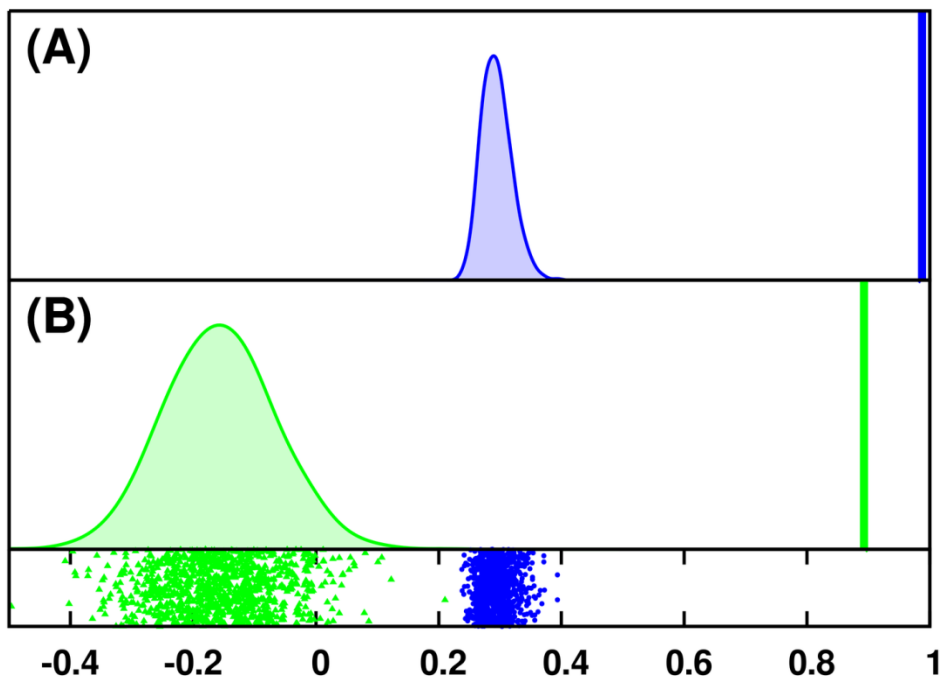


Figure S11. Response permutation test results for the MB-PLS-DA model of combined MS and NMR datasets. Model fit (R^2) statistics (A) are shown in blue, and model predictive ability (Q^2) statistics are shown in green. True values of R^2 and Q^2 are represented by vertical bars, and null distributions are computed through kernel density estimation of the values from permutation. Scatter plots of the permutation R^2 and Q^2 statistics are shown in the lower pane.