Between-Person Comparison of Metabolite Fitting for NMR-Based Quantitative Metabolomics

ARTICLE *in* ANALYTICAL CHEMISTRY · NOVEMBER 2011

Impact Factor: 5.64 · DOI: 10.1021/ac202123k · Source: PubMed

5 AUTHORS, INCLUDING:

Gregory Tredwell
Australian National University
13 PUBLICATIONS **122** CITATIONS
SEE PROFILE

Volker Behrends
University of Roehampton
20 PUBLICATIONS **293** CITATIONS
SEE PROFILE

Manuel Liebeke
Max Planck Institute for Marine Microbiology
46 PUBLICATIONS **949** CITATIONS
SEE PROFILE

Jake Bundy
Imperial College London
82 PUBLICATIONS **2,849** CITATIONS
SEE PROFILE

analytical chemistry

ARTICLE

pubs.acs.org/ac

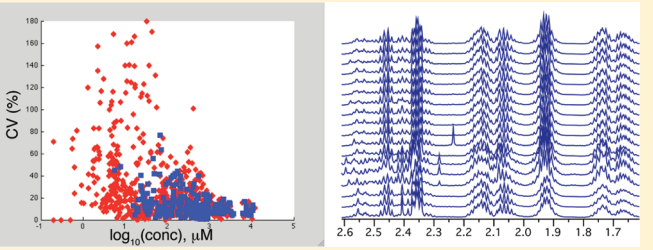
Between-Person Comparison of Metabolite Fitting for NMR-Based Quantitative Metabolomics

Gregory D. Tredwell, Volker Behrends, Florian M. Geier, Manuel Liebeke, and Jacob G. Bundy*

Department of Surgery and Cancer, Imperial College London, Sir Alexander Fleming Building, London SW7 2AZ, U.K.

Supporting Information

ABSTRACT: Nuclear magnetic resonance (NMR) spectroscopy is widely used as an analytical platform for metabolomics. Many studies make use of 1D spectra, which have the advantages of relative simplicity and rapid acquisition times. The spectral data can then be analyzed either with a chemometric workflow or by an initial deconvolution or fitting step to generate a list of identified metabolites and associated sample concentrations. Various software tools exist to simplify the fitting process, but at least for 1D spectra, this still requires a degree of skilled operator input. It is of critical importance that we know how much person-to-person variability affects the results, in order to be able to judge between different studies. Here we tested a commercially available software package (Chenomx' NMR Suite) for fitting metabolites to a set of NMR spectra of yeast extracts and compared the output of five different people for both metabolite identification and quantitation. An initial comparison showed good agreement for a restricted set of common metabolites with characteristic well-resolved resonances but wide divergence in the overall identities and number of compounds fitted; refitting according to an agreed set of metabolites and spectral processing approach increased the total number of metabolites fitted but did not dramatically increase the quality of the metabolites that could be fitted without prior knowledge about peak identity. Hence, robust peak assignments are required in advance of manual deconvolution, when the widest range of metabolites is desired. However, very low concentration metabolites still had high coefficients of variation even with shared information on peak assignment. Overall, the effect of the person was less than the experimental group (in this case, sampling method) for almost all of the metabolites.



Nuclear magnetic resonance (NMR) spectroscopy is widely used for metabolic profiling (metabolomics)[1,2] and has been the subject of major investments such as the Human Metabolome Database.[3] This continued popularity is due to a number of inherent properties of NMR; in particular, proton NMR is a near-universal detector and reports on metabolites from all chemical classes simultaneously, with high precision.[4,5] Furthermore, because there is no physical separation of metabolites and because, disregarding relaxation effects, the intensity of a resonance is proportional to the actual numbers of nuclei giving rise to it, NMR is particularly powerful for making between-metabolite comparisons and can detect even subtle relative concentration changes.[6]

There are two classes of methods used for analyzing NMR metabolomic data. Chemometric approaches use pattern-recognition methods to analyze whole spectra, and individual metabolites need only be identified a posteriori following identification of discriminatory peaks.[7] Conversely, deconvolution or fitting methods aim to match resonances to metabolite standards, described as "quantitative metabolomics" or "targeted profiling".[8,9] Chemometric methods have several advantages, in particular, spectra can be analyzed without prior knowledge of what metabolites are present, even for wholly novel compounds. However there are also many advantages to the quantitative metabolomics approach. First, the data analysis problem is reduced from hundreds or thousands of bins/data points (with unhelpful statistical properties such as high correlation between variables) to, typically, tens of variables that represent actual discrete entities. Second, knowledge of chemical identities permits analyses based on prior knowledge, for instance, chemical similarity or pathway relatedness.[10−12] Third, quantitative data are much more valuable if metabolomic data are to be used as part of an input for a systems biology study. Finally, if studies are to be compared between different laboratories, an essential underpinning of science, then quantitative data on named metabolites are ideal.

There are a number of different software options for metabolite fitting, reviewed by Wishart.[8] In our own lab, we have made use of a commercial package, NMR Suite (Chenomx, Edmonton, Canada). This software is not completely automated (although its use in fully automated studies has recently been explored[13]): rather, it provides computer-assisted manual fitting. This allows judging of individual peak shifts and shapes by the analyst but means there is an element of subjectivity in the procedure. When a single person fits a set of spectra, the reproducibility is very good: independently fitted technical repeats are more similar to each other even than to other samples within the same control

Received: August 12, 2011
Accepted: October 11, 2011
Published: October 11, 2011

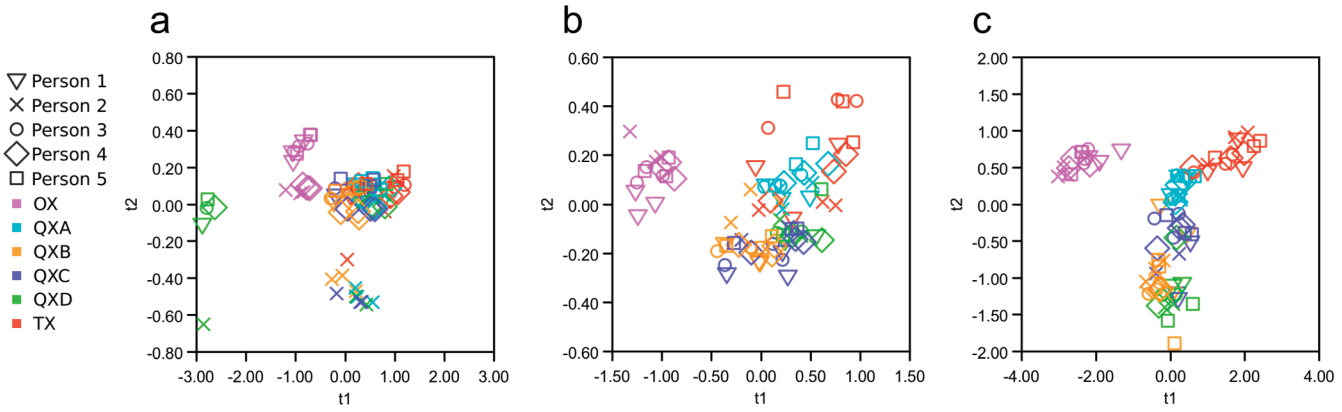


Figure 1. Principal component analysis scores plots of (a) 16 metabolites fitted independently from round one; (b) same as for part a but with one metabolite (threonine) and one outlying sample (QXB1) excluded; (c) 37 metabolites fitted based on exemplar spectra in round two, with one outlying sample (QXB1) excluded. Data were log transformed and mean centered.

group.[10] However, in order to be able to compare between studies, one needs to know the between-person reproducibility. A previous study has assessed the precision and accuracy of targeted profiling using the Chenomx software for a small number of metabolites.[14] Three separate analysts assigned concentrations for nine metabolites spiked into human urine samples, and the study concluded that all three analysts produced similar results with coefficients of variation for individual metabolite concentrations ranging from 1 to 41%.[14]

We expand upon this work by comparing the Chenomx NMR Suite software between five different analysts and report here how the reproducibility varied for all detectable metabolites that we could assign, for a different sample type. In order to obtain information on the widest possible range of metabolites, it was necessary to share information on peak assignment; however, a core group of easily assigned and high-concentration metabolites compared well between different analysts even with no prior sharing of information.

METHODS

We took spectra from an earlier study on the bioprocessing yeast *Pichia pastoris*.[15] The data set consisted of six classes, representing cells sampled by one of six different methods: unquenched (centrifuged) cell extracts (OX); cell extracts made with four different methods based on rapid quenching with cold solvents (QXA, QXB, QXC and QXD); and a total quenched extract (TX) of cells + broth combined. Each class contained three replicates, giving a total of 18 spectra. All participants then fit these data independently in two rounds. In round one, to mimic completely independent analysis, all participants processed the spectra and then fitted metabolites using NMR Suite 6 (Chenomx, Edmonton, Canada) completely independently. Four of us chose to process and fit the spectra within NMR Suite, while one processed the spectra using iNMR 3 (Nucleomatica, Molfetta, Italy), with export as JCAMP files and subsequent import of the processed files into NMR Suite. Metabolite assignment was carried out individually; online data sources such as BioMagResBank[16] and HMDB[3,17] could be used in addition to the NMR Suite internal database, if chosen. The only guidance given was that additional standard files for two high-concentration metabolites (trehalose and arabitol) were shared with all participants.

In round two, we shared both processed spectra and "template profiles" (i.e., previous examples of spectra fitted using NMR Suite) containing assigned metabolites. The spectra were processed using iNMR 3 (with a zero-filling factor of 2 and exponential line broadening of 0.5 Hz) and exported as JCAMP files, which were then given to all participants. The processing was fully automated (using the iNMR software routines for metabolomic phase correction, baseline correction, and chemical shift referencing), and so we did not have this step repeated by individual participants. The spectra were then fitted a second time in NMR Suite after first importing one of two template profiles for each spectrum (one for all TX samples and one for all remaining samples), ensuring that all participants fitted exactly the same set of 37 metabolites to the spectra. (Metabolite assignment, including 2D NMR and spiking of authentic standards, is described in Tredwell et al.[15]) In addition, all participants viewed the fitting of one spectrum together.

Data analysis was carried out using Matlab (Mathworks, Massachusetts) and Aabel 3 (Gigawiz, Tulsa). The spectra can be downloaded (both as raw and processed files) from the Supporting Information as well as the Chenomx template files for six additional metabolites not in the standard database (trehalose, UDP-*N*-acetylglucosamine, methionine sulfoxide, arabitol, nicotinamide mononucleotide, and histidinol).

RESULTS AND DISCUSSION

In round one, participants processed and fitted the spectra independently. The only information given was that two of the metabolites present were arabitol and trehalose. In total, 77 metabolites were fitted in at least one sample (Figure S1 in the Supporting Information). However, while the majority of metabolite assignments were fitted to all samples, when compared across all 5 of us, just 16 metabolites were fitted in almost all cases (Figure S1 in the Supporting Information), and so we initially analyzed the data from these 16 compounds only (Table S1 in the Supporting Information).

In round two, all spectra were processed by automation and participants refitted all spectra. In addition to the 16 metabolites assigned from round one, a further 21 metabolites with confirmed assignments were included (Table S2 in the Supporting Information). The full metabolite list was saved as a Chenomx

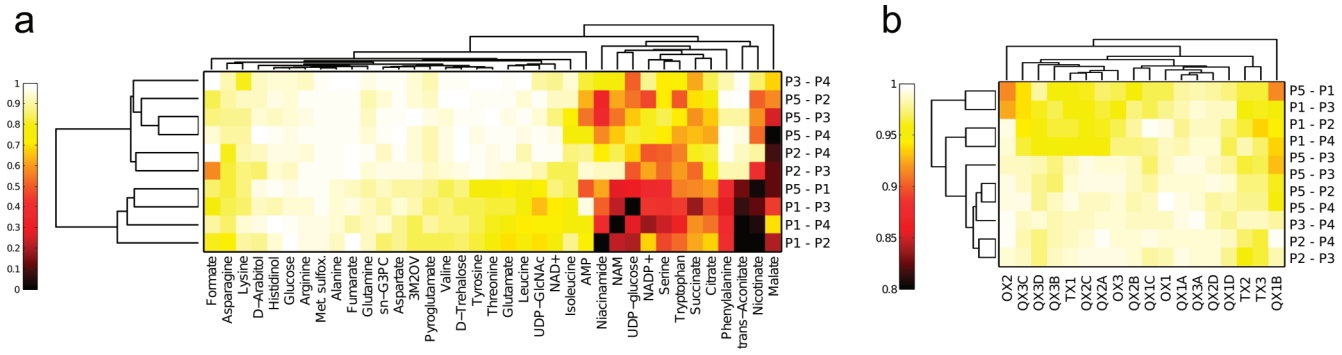


Figure 2. Correlations (Spearman's ρ) between people for both metabolites and samples ordered by hierarchical clustering in both dimensions: (a) association between different people for individual metabolites and (b) association between different people for individual samples.

"template" profile, ensuring all metabolites were included in round two by all participants.

We used principal components analysis (PCA) to give a quick comparison of the overall variability of different participants (Figure 1). For round one, the samples were clustered by the extraction method along PC 1, with one outlying sample (QXB1), which suffered losses during sample preparation and as a result had low concentrations for all metabolites. Person 2 was separated along PC2 (Figure 1a), with the loadings indicating differences in threonine. Inspection of the fitted spectra revealed that person 2 had incorrectly fitted lactate as well as threonine to a doublet resonance at 1.32 ppm, highlighting the potential for user bias and the need for 2D NMR spectra for robust assignments. When threonine and the outlying sample QXB1 were excluded (Figure 1b), there were no longer any clear differences between people (Figure 1b) and data points were largely clustered by sample type in both PC1 and PC2, with the TX and QX samples overlapping slightly.

The data for round two clustered largely by the sampling method and not by person, and again sample QXB1 was a clear outlier (data not shown). With this sample excluded, there was very clear clustering according to sample method rather than by person (Figure 1c). It should be borne in mind that the current study is of a very controlled set of samples: yeast cells grown under the same conditions and differing only in sampling methods. This therefore represents a very conservative comparison between people and experimental treatments, and it is likely that the results would look even better if we had used highly variable samples such as urine or cell growth media.

The multivariate analysis for both rounds indicates that the biological variation between samples was greater than person-to-person variation, since there was clustering of sample classes rather than of different analysts. We examined this in more detail using two-way analysis of variance (ANOVA) for all metabolites, with experimental treatment ("method") and individual variability in fitting ("person") as factors. For round one, "method" had a greater contribution to variance than "person" for all metabolites except threonine (Table S1 in the Supporting Information). For round two (Table S2 in the Supporting Information), 17 out of 37 metabolites had a nonsignificant "person" effect. This means that 20 out of 37 metabolites had significant differences in fitting for at least one of the participants. This may sound like a poor outcome; however, we can also use ANOVA to compare the relative contributions of "method" and "person". For round two, three metabolites, malate, nicotinate, and UDP-glucose, were more affected by person than sampling method, but this still means that for the large majority of 34 compounds, the relative effect of individual analyst variability was smaller than experimental treatment, even for highly controlled and very similar samples.

To look more closely at differences between people when fitting metabolite concentrations, we first performed pairwise rank correlations (Spearman's ρ) of individual metabolite concentrations over all 18 samples, and second, pairwise correlations of individual samples over all metabolites (i.e., 16 for round one and 37 for round two). Round one results are shown as Supporting Information (Figure S2), and round two results are shown in Figure 2. For round one, correlations were very poor for threonine and phenylalanine for one person, but otherwise pairwise correlations for metabolites were high across all participants. For round two, the majority of metabolite correlations were good, especially for the same 16 metabolites fitted from round one (Figure 2a). However, a number of metabolites did not correlate well between people (particularly person 1), with malate showing the worst correlations between people.

The correlations (ρ) of individual samples between people were, as expected, much higher than for the comparison of metabolites (Figure 2b). Generally the TX samples, which are slightly more complex than the other samples due to the presence of extracellular metabolites and media components, and the QXB1 sample, which contained low concentrations for all metabolites, resulted in slightly lower correlations.

Finally, we calculated mean concentrations and percent coefficients of variation (% CV) between people (n = 5) for each metabolite in individual samples, giving 288 data points (18 samples × 16 metabolites) for round one and 666 data points (18 samples × 37 metabolites) for round two (Figure 3). Comparing the same 16 metabolites for both rounds shows that there was no real improvement in round two, with the majority of data points below 20% coefficient of variation for both rounds (Figure 3a). When including all of the metabolites from round two, however, the data appear worse, as the distribution of % CV is higher (Figure 3b). We wanted to know if this was because round two was including additional lower-concentration metabolites, which could not be fitted as precisely and so we plotted mean metabolite concentrations (n = 18, i.e., average across all spectra) against mean % CV for both rounds one and two (Figure 3d,e). There is a clear relationship between concentration and between-person variability, and it was indeed the case that round two included a number of metabolites with very low concentrations and concomitantly high % CV. However, there were also seven additional metabolites in round two with mean coefficient of variation <20% and three additional high-concentration

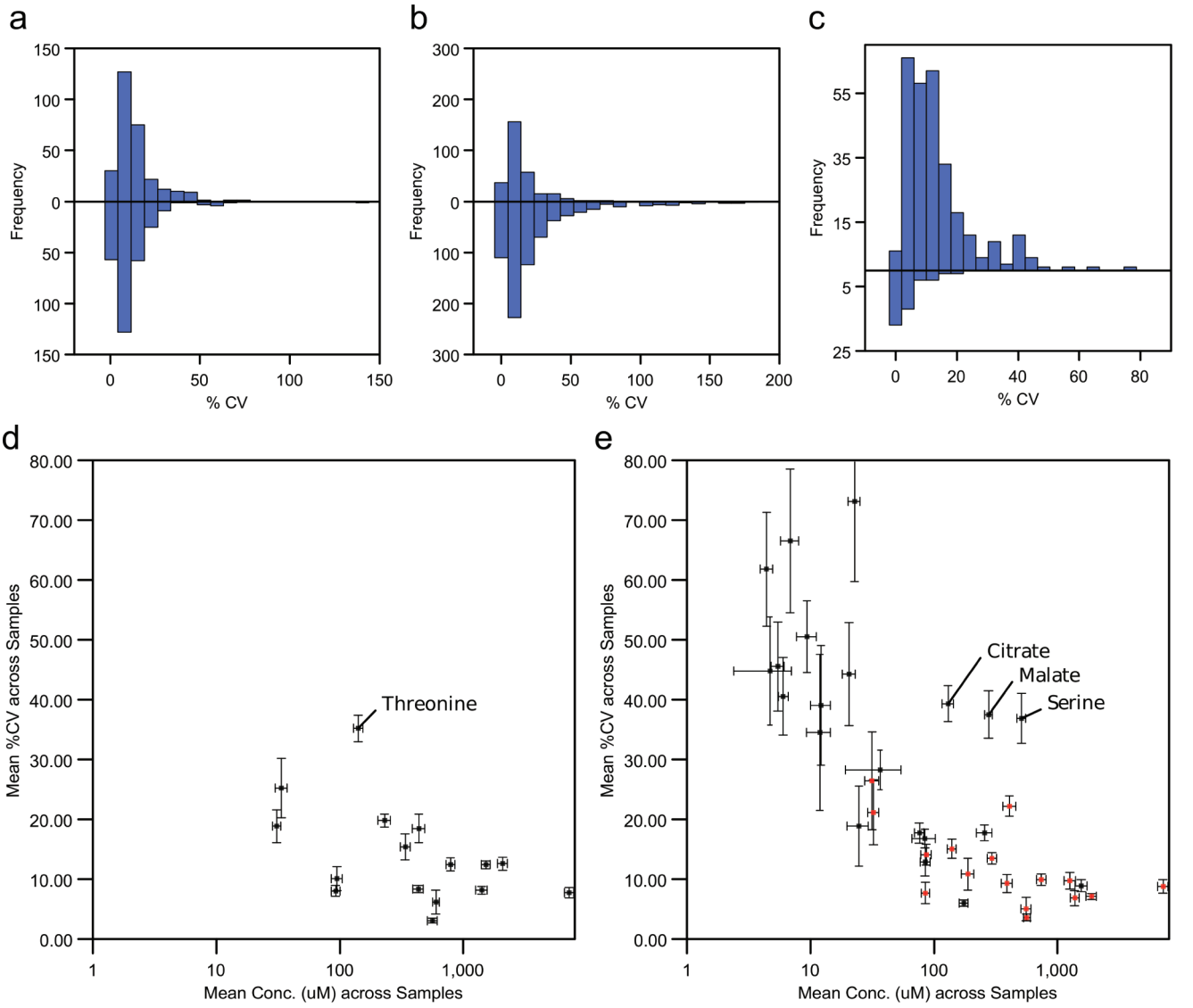


Figure 3. (a) Distribution of coefficients of variation between people, calculated for each metabolite for each spectrum separately (n = 5). Comparison of 16 metabolites fitted independently from round one (top) to the same 16 metabolites fitted based on exemplar spectra from round two (bottom); (b) comparison of 16 metabolites fitted independently from round one (top) to 37 metabolites fitted based on exemplar spectra from round two (bottom); (c) comparison of 16 metabolites fitted independently from round one (top) to the within-person variation of 37 metabolites fitted in one spectrum five times independently. (d) Relationship between mean metabolite concentration and mean coefficient of variation for 16 metabolites fitted independently from round one across samples; (e) same as for part d but for 37 metabolites fitted based on exemplar spectra from round two. Red data points refer to the same 16 metabolites as in part d.

metabolites that nonetheless had relatively poor % CV (citrate, malate, and serine). In these cases, the greater variability can be attributed to the fact that their resonances appear in highly complex regions with overlapping signals from other metabolites. These results are consistent with Slupsky et al.[14] who found the same concentration dependent variation between three analysts for the fitting of nine different metabolites spiked into human urine.

What implications do these results have for metabolomics studies in general? Some of the fitted metabolites have extremely poor reproducibility across different individuals (Figures 2 and 3). This seems at first far worse than similar studies of 1D spectra of replicate samples, which have reported a median CV of around 3%.[18] However, it must be remembered that these are the extreme values: the median % CV for round two (Figure 3b) was 14% which, while still high, is much more acceptable, particularly when it is remembered this represents a between-person metric. It should also be borne in mind that a majority of studies will probably have data fitted by one person only. In order to compare to this common situation, we had one participant fit one spectrum five times independently. Unsurprisingly, the results were far more precise, with a median % CV of just 2.4% and almost all metabolites (32 out of 37) below 10% (Figure 3c). Between-laboratory comparisons have been made for NMR metabolomic data,[19,20] and the between-person contribution to variation is not greater than the typical between-lab contribution (and the within-person contribution is less than both).

Overall, we found a good agreement between fitted metabolite concentrations among five analysts, certainly good enough to permit comparisons of studies between different people or even between different groups. The main sources of variation were

dx.doi.org/10.1021/ac202123k |*Anal. Chem.* 2011, 83, 8683–8687

incorrect assignments, overlapping signals, or low metabolite concentrations, although generally this variation was still less than the biological variation, even for a set of very similar cell extracts. Sharing prior information on spectral assignment enabled reproducible fitting of a larger number of metabolites, but the differences in spectral processing between people were not a large source of error. We propose a number of suggestions to improve the reliability of this type of data. First, metabolite assignments need to be rigorously confirmed before fitting (e.g., with 2D NMR experiments and spike-in experiments, as appropriate). Second, automated processing was not an important source of error and could be beneficial for studies with large numbers of samples. Finally, we recommend that, where possible, small discrete studies (e.g., tens of samples) be fitted by just one person to maximize precision; however, between-person comparisons are also useful in highlighting possible "problem" metabolites with a high variance contribution from the fitting process. Where possible, making NMR metabolomic data sets publicly available would permit evaluation and reanalysis by other researchers.

■ ASSOCIATED CONTENT

Supporting Information. Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: j.bundy@imperial.ac.uk. Fax: +44 20 75943226.

■ ACKNOWLEDGMENT

We thank Toby Athersuch and Tim Ebbels (Imperial College London) for their helpful comments and discussion of the manuscript. This study was supported by the BBSRC (Grant References BB/E20372/1 and BB/G020434/1) and the NERC (Grant NE/H009973/1). F.M.G. was funded by the AXA Research Trust.

■ REFERENCES

(1) Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L. *Chem. Soc. Rev.* **2011**, *40*, 387–426.
(2) Zhang, S.; Nagana Gowda, G. A.; Ye, T.; Raftery, D. *Analyst* **2010**, *135*, 1490–1498.
(3) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res.* **2007**, *35*, D521–6.
(4) Fan, T. W. M. *Prog. NMR Spectrosc.* **1996**, *28*, 161–219.
(5) Nicholson, J. K.; Wilson, I. D. *Prog. NMR Spectrosc.* **1989**, *21*, 449–501.
(6) Keun, H. C.; Ebbels, T. M.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Chem. Res. Toxicol.* **2002**, *15*, 1380–1386.
(7) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–479.
(8) Wishart, D. S. *Trends Anal. Chem.* **2008**, *27*, 228–237.
(9) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. *Anal. Chem.* **2006**, *78*, 4430–4442.
(10) Bundy, J. G.; Sidhu, J. K.; Rana, F.; Spurgeon, D. J.; Svendsen, C.; Wren, J. F.; Sturzenbaum, S. R.; Morgan, A. J.; Kille, P. *BMC Biol.* **2008**, *6*, 25.
(11) Chagoyen, M.; Pazos, F. *Bioinformatics* **2011**, *27*, 730–731.
(12) Xia, J.; Wishart, D. S. *Bioinformatics* **2010**, *26*, 2342–2344.
(13) Mercier, P.; Lewis, M. J.; Chang, D.; Baker, D.; Wishart, D. S. *J. Biomol. NMR* **2011**, *49*, 307–323.
(14) Slupsky, C. M.; Rankin, K. N.; Wagner, J.; Fu, H.; Chang, D.; Weljie, A. M.; Saude, E. J.; Lix, B.; Adamko, D. J.; Shah, S.; Greiner, R.; Sykes, B. D.; Marrie, T. J. *Anal. Chem.* **2007**, *79*, 6995–7004.
(15) Tredwell, G. D.; Edwards-Jones, B.; Leak, D. J.; Bundy, J. G. *PLoS One* **2011**, *6*, e16286.
(16) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402–8.
(17) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res.* **2009**, *37*, D603–10.
(18) Parsons, H. M.; Ekman, D. R.; Collette, T. W.; Viant, M. R. *Analyst* **2009**, *134*, 478–485.
(19) Ward, J. L.; Baker, J. M.; Miller, S. J.; Deborde, C.; Maucourt, M.; Biais, B.; Rolin, D.; Moing, A.; Moco, S.; Vervoort, J.; Lommen, A.; Schafer, H.; Humpfer, E.; Beale, M. H. *Metabolomics* **2010**, *6*, 263–273.
(20) Viant, M. R.; Bearden, D. W.; Bundy, J. G.; Burton, I. W.; Collette, T. W.; Ekman, D. R.; Ezernieks, V.; Karakach, T. K.; Lin, C. Y.; Rochfort, S.; de Ropp, J. S.; Teng, Q.; Tjeerdema, R. S.; Walter, J. A.; Wu, H. *Environ. Sci. Technol.* **2009**, *43*, 219–225.