

A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis

Timothy Sangster,^{†,a} Hilary Major,^b Robert Plumb,^c Amy J. Wilson^{‡,a} and Ian D. Wilson^{§,a}

DOI: 10.1039/b604498k

Metabonomic/metabolomic studies can involve the analysis of large numbers of samples for the detection of biomarkers and confidence in the analytical data, generated by methods such as GC and HPLC-MS, requires active measures on the part of the analyst. However, quality control for complex multi-component samples such as biofluids, where many of the components of interest in the sample are unknown prior to analysis, poses significant problems. Here the repeat analysis of a pooled sample throughout the run, thereby enabling the analysis to be monitored and controlled using targeted inspection of the data and pattern recognition, is advocated as a pragmatic solution to this problem.

Introduction

The acquisition of robust and meaningful global metabolite profiles from complex biological samples, including biofluids, such as plasma and urine, or tissue extracts poses an interesting set of problems for the analyst. Obtaining such profiles forms the core of the rapidly emerging sciences known variously as metabonomics and metabolomics, where the hope is that such techniques will uncover important biomarkers of *e.g.* toxicity or disease.^{1,2} Variability in the samples can arise from a number of sources including physiological differences (*e.g.* strain, gender, age, diurnal and hormonal effects *etc*) (see ref. 3 for a discussion of this topic) and variability in the analytical method itself (both sample preparation and analysis). For certain techniques, such as ¹H NMR

spectroscopy, where sample preparation for biofluids is minimal, analytical reproducibility has been demonstrated to be very good.⁴ However, analytical methods for metabolomics that employ either HPLC or GC-MS generally require more sample preparation, which in the case of GC-based analysis is often extensive, followed by a chromatographic separation and then mass spectrometry. Chromatographic techniques are liable, to a greater or lesser extent (depending upon the technique and sample type), to degradation of performance over time as columns become contaminated, and the response of mass spectrometers can also decline with time for similar reasons. In conventional target compound analysis these factors are controlled by the incorporation of internal standards, most often a deuterated version of the analyte, into the analytical procedure to counter, if not entirely eliminate, such effects. In addition, specific quality control samples (QCs) are employed to monitor the performance of the method. In the case of validated methods, *e.g.* drug analysis, the QCs are blank matrix samples spiked with known concentrations of the analyte designed to cover the range of concentrations that can be determined with reasonable accuracy and precision. These samples are usually found at the beginning and end of the sample set and also scattered randomly through the analytical run. Examination of the QC

data at the end of the analysis against a set of predefined criteria enables the analytical scientist to decide whether or not to accept or reject the batch. In such targeted analysis, retention is also monitored but, given the specificity of the technique, some variation through the run is more easily tolerated. Such rigor is not merely good analytical practice but, where data is to be used to support drug registration, is covered by regulatory guidance.⁵

This approach is not viable for metabonomics analysis for the following reasons.

(1) The samples typically contain 100s to 1000s of components covering a wide range of concentrations and structural types, of variable and unknown MS response.

(2) The bulk of the analytes in the sample are unknown prior to the analysis and indeed, because of limitations in the state of our knowledge at the moment, potential biomarkers may remain unidentified at the end.

(3) By definition, stable isotope labelled internal standards cannot be used where the identity of the analytes are unknown, even if it were a realistic economic proposition to prepare them all and spike them in.

(4) The post analysis processing of the data, if it makes use of the 3-dimensional data set provided by mass, retention and intensity information, cannot easily

^aDept. of Drug Metabolism and Pharmacokinetics, AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire, UK SK10 4TG.

E-mail: ian.wilson@astrazeneca.com

^bWaters Corporation, Atlas Park, Manchester, UK M22 5PP

^cWaters Corporation, 34 Maple Street, Milford, MA, 01757, USA

† Present address: Huntingdon Life Science, East Millstone, New Jersey 08875-2360, USA.

‡ Present address: School of Clinical Medical Sciences, University of Newcastle Upon Tyne, Framlington Place, Newcastle Upon Tyne, UK NE2 4HH.

§ Ian Wilson is a joint recipient of the 2005 SAC Gold Medal.

tolerate significant changes in chromatographic retention or mass spectrometer response.

A partial solution to this problem is to accept that it is not possible to control the analysis of all of the compounds and instead to opt for the control of a limited number of them for which deuterated internal standards are available. This is done in the hope that if these analytes "behave" themselves then the system is under control and thus the analysis of all of the remaining analytes is also under control. The selected standards can be spiked in to samples prior to analysis as pseudo internal standards, or run simply as QCs alongside the test samples. Whilst such an approach might have been sustainable a few years ago, increasing experience of quantitative HPLC-MS, where ion suppression/ion enhancement and source contamination can cause highly variable responses during analysis, has done much to dim such optimism amongst the authors. In addition to MS-related effects, changes in column selectivity are also likely as columns age, which might well lead to differential changes in retention for *e.g.* bases *vs.* acids *etc.*, especially if in HPLC silanolophilic interactions are involved in the retention mechanisms of any of the analytes.

Against this backdrop we have attempted to formulate a pragmatic strategy for controlling the multi-component, multi-parametric, analytical process encountered in metabolomics. At the heart of this approach are the samples themselves that, between them, contain all of the analytes that will be encountered during the analysis.

We therefore advocate taking aliquots from every sample which are then combined in a representative pool sample. The pool sample is then split to form a multi-sample QC set which is analysed at the beginning, end and randomly through the analytical run. The same pool sample can also be used for performing a system suitability test prior to beginning the main analytical run if required. For batches of 100 or so samples the QCs would represent a minimum of 10% of the total analysed (more if the batch was small).

Post analysis, the pool sample QC data can be examined visually for gross changes to give a rapid assessment of

how well the run has gone. Similarly, a small number of selected components can be rapidly screened for peak shape, intensity, mass accuracy and retention time against predetermined acceptance criteria. Assuming that these criteria are met then the whole data set can be taken forward for initial multivariate statistical analysis, using an unsupervised method such as principal components analysis (PCA) with the QC data expected to cluster closely together, and show no time related trends (supervised methods should not be used as these will "force" the QCs to cluster together potentially masking variability). If statistical analysis reveals more subtle, time-related, changes the analyst can use the results to determine if there was a gradual change during the analysis or whether a sudden deterioration had occurred at some point midway through the analysis.

Here we give two examples of the use of this strategy for the control of GC-MS analysis of rat plasma samples and the reversed-phase gradient HPLC-ToF-MS of human urine.

For GC-MS analysis, plasma samples from 4 different strains of rat (100 μ L) were protein precipitated using 3 volumes of acetonitrile, followed by centrifugation, and then 100 μ L of each supernatant was evaporated to dryness prior to derivatisation. In addition, prior to analysis, 50 μ L of each original sample were pooled to generate the QC and aliquots of 100 μ L of this pooled sample were taken through the same process. All samples were then subjected to a double derivatisation procedure involving methoxylamine hydrochloride and then MSTFA at 37 °C.⁶ Capillary GC-MS, using a 20 m \times 180 μ m \times 0.18 μ m DB5 column, with a temperature gradient from 85 to 320 °C, was then used to generate profiles in both EI and CI modes (separate batches of sample were used for the EI and CI runs). The resulting GC-MS data were processed using the Waters MarkerLynx Application Manager. The results of the PCA of the GC-EIMS and GC-CIMS data are shown in Fig. 1 and 2 respectively. As these figures show, despite extensive and lengthy sample preparation and subsequent GC-MS, the QC samples generally cluster closely together in the PCA scores plot, providing a degree of confidence that the results obtained for the test

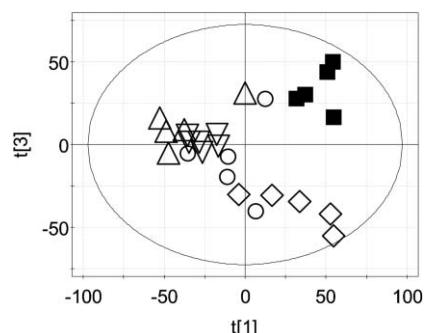


Fig. 1 All samples were plasma from 20 week old male rats. The figure shows the PCA scores plot of PC1 versus PC2 obtained from the GC-EIMS data. Key: (■) QCs; (○) Wistar-derived rats; (◇) Zucker (fa/fa) rats; (△) Zucker lean/fa cross; (▽) Zucker lean rats.

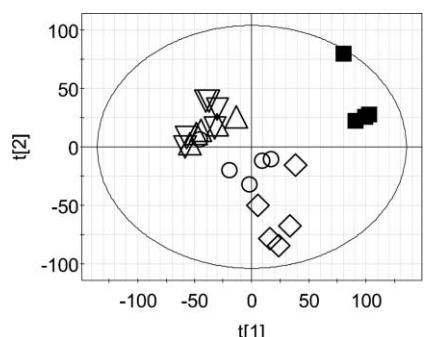


Fig. 2 All samples were plasma from 20 week old male rats. The figure shows the PCA scores plot of PC1 versus PC2 obtained from the GC-CIMS data. Key: (■) QCs; (○) Wistar-derived rats; (◇) Zucker (fa/fa) rats; (△) Zucker lean/fa cross; (▽) = Zucker lean rats.

samples are suitable for further data analysis with the aim of finding biomarkers. The one QC sample which does not cluster with the others in the GC-CIMS example (Fig. 2) was the second injection of the batch (the first having been automatically discarded). Such behaviour in the first few samples run in both GC and LC-MS is not unusual in our experience and, as discussed below, has led us to a change our analytical practice.

For UPLC-MS analysis, urine samples from a set of human samples (100 μ L) were diluted by the ratio 1 : 4 with 0.1% formic acid, followed by centrifugation, and then injected onto the LC-MS. In addition, prior to analysis, 50 μ L of each original sample were pooled to generate a QC sample and aliquots of 100 μ L of this pooled sample were taken through the same process. The samples were analysed on a Waters ACQUITY UPLC system

with a 10 cm × 2.1 mm, 1.7 µm BEH C18 ACQUITY column coupled to a QToF Micro mass spectrometer. The column was maintained at approximately 40 °C and elution was performed using a gradient of 0.1% formic acid and acetonitrile. The resulting LC-MS data were processed using the Waters MarkerLynx Application Manager, with statistical analysis in SIMPCA-P. The results for the PCA of the UPLC-MS data are shown in Fig. 3, and once again show that, although there is some variability in the QCs, these samples nevertheless cluster closely together and indicate that the analysis is under control.

In any QC approach the question of course then arises as to how "tight" the data should be to be considered acceptable. Currently our practice is to use the QC data as a means of rejecting batches as, if the QCs are widely scattered in the scores plot, it is fairly easy to decide that the analysis was not fit for purpose. However, we are not then advocating the blind acceptance of the remaining batches but, having filtered out obviously bad data, it then seems reasonable to invest more time analysing the results from runs that appear, on the basis of the PCA

result, to have been under good analytical control. For potential biomarkers detected in the samples it would then be reasonable to examine the reproducibility of the method for that component, taking into account the intensity of the ion of interest (so that intense ions giving good signal-to-noise (S/N) ratios should perhaps be required to exhibit a higher degree of precision than those of low intensity). In this respect the FDAs guidance on bioanalytical method validation⁵ would probably provide a good starting point. Thus, coefficients of variation of less than 15% would be required for ions with good S/N and 20% for those with an S/N of perhaps only 3 times greater than background.

For both HPLC and GC, when the potential biomarkers have been identified by statistical analysis of the data, it is possible to re-examine the QC data specifically to look at the variability of the results obtained for those specific ions in the QC-data set. Once satisfied that the results are unlikely to be artefacts of the analysis, it may then be worth devoting the time to the identification of these interesting metabolites with the aim of developing specific and

validated methods for them to prove the hypothesis that they are indeed biomarkers for the biological state under investigation.

On the basis of our observations made using this approach we know that the first few analytical runs are the most variable (e.g. see the GC-CIMS data above). Whilst the reasons for this are not clear the consequences are obvious, and we would therefore strongly advocate that, prior to beginning an analytical run, several QC samples are run first to effectively "condition" the chromatographic system. The data from these initial runs should not form part of total QC data set used subsequently to "validate" the quality of the metabolite profiles generated, but could be used as supporting data to show system suitability.

The methodology outlined here has been specifically designed for use on relatively small batches of samples (from a few tens of samples up to a few hundred) that could be accommodated in a single analytical run on one instrument. Such a sample size would be typical of the sorts of numbers generated in toxicological studies in animal species, investigative studies in disease models or

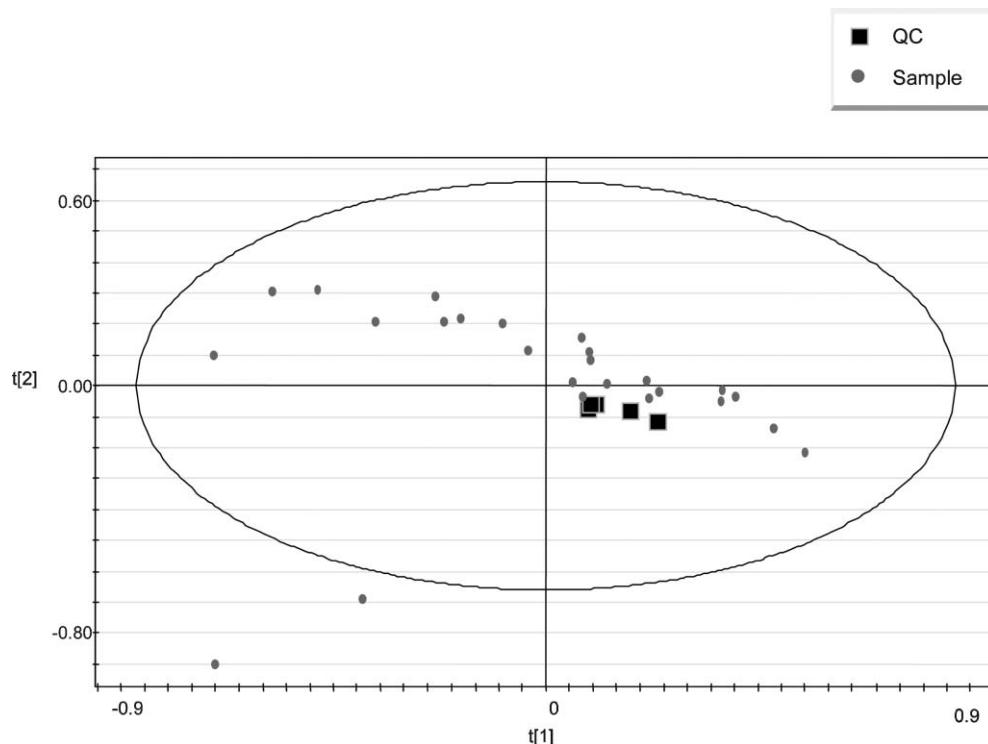


Fig. 3 The PCA scores plot obtained following the UPLC-MS analysis of human urine samples (grey circles). To generate QCs 50 µL of each original sample were pooled (black squares). The samples were analysed by reversed-phase UPLC on a 10 cm × 2.1 mm, 1.7 µm BEH C18 ACQUITY column coupled to a QToF Micro mass spectrometer.

small scale studies in humans. In such applications, the QC policy described here seems to be appropriate to demonstrate “within day” analytical control. Whether or not the use of such QCs would enable “split” batches, run on different days, to be combined is less clear. This is clearly an area that requires further investigation as there are circumstances when it would be highly desirable to be able to combine “between” day datasets (e.g., after instrument failure partway through a run, or where there are more samples than can be easily accommodated in a single analytical run).

There are, in addition, the QC requirements of larger epidemiological or clinical studies, to be considered. Such types of study are more problematic because the large numbers of samples collected (often over a period of some years) means that analysis in a single batch is not possible. In such cases, as there is a need to also ensure between batch as well as within batch data quality, it would almost certainly not be practicable to use a pool QC prepared from the samples themselves, and instead the use of a single bulk sample prepared at the start of the study and split into a large number of sub-aliquots and stored with the study samples may be preferable. However, such an approach assumes sample stability over the collection period of the study.

Clearly, if confidence is to be placed in the data generated from complex sample

analysis of the type encountered in metabolomics studies, some assurance of the quality of the data is required. This is especially the case if data are to be submitted in support of regulatory studies, but also forms an important part of any analytical study in this area. There are already initiatives working towards the standardisation of the reporting of metabolomics data,⁷ and quality control procedures need to form part of this debate. Probably the best way to use the approach described above is as an initial screen of the analytical results. Thus, if the QC data is highly variable the run fails and re-analysis is required. In contrast, if the QC data are close, this does not necessarily mean that the analysis is satisfactory, but allows provisional acceptance of the run, and justifies devoting more time to a more exhaustive interrogation of the data with more advanced statistical procedures.

Whilst this approach of repeat analysis of a pooled sample can be criticized in any number of ways it at least has the advantages of ease of implementation, speed (it can be performed by the analyst at the instrument) and relevance to the samples being analysed. We therefore offer it as one possible route towards a viable QC policy for monitoring global metabolite profiling that covers that part of the analytical process involving sample preparation and chromatographic/mass spectroscopic analysis, in the hope of stimulating a debate on what we

believe to be an important problem facing investigators in this area.

References

- 1 *Metabonomics in Toxicity Assessment*, ed. D. Robertson, J. C. Lindon, J. K. Nicholson and E. Holmes, CRC Press, Boca Raton, 2005.
- 2 J. T. Brindle, H. Antti, E. Holmes, G. Tranter, J. K. Nicholson, H. W. L. Bethell, S. Clarke, P. M. Schofield, E. McKilligin, D. E. Mosedale and D. J. Grainger, *Nat. Med.*, 2002, **8**, 1439–1445.
- 3 M. E. Bolland, E. G. Stanley, J. C. Lindon, J. K. Nicholson and E. Holmes, *NMR Biomed.*, 2005, **18**, 143–162.
- 4 H. C. Keun, T. M. D. Ebbels, H. Antti, M. Bolland, O. Beckonert, G. Schlotterbeck, H. Senn, U. Niederhauser, E. Holmes, J. C. Lindon and J. K. Nicholson, *Chem. Res. Toxicol.*, 2002, **15**, 1380–1386.
- 5 *FDA Guidance for Industry, Bioanalytical Method Validation*, Food and Drug Administration, Centre for Drug valuation and Research (CDER), May 2001.
- 6 O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey and L. Willmitzer, *Nat. Biotechnol.*, 2000, **18**, 1157–1161.
- 7 J. C. Lindon, J. K. Nicholson, E. Holmes, H. C. Keun, A. Craig, J. T. M. Pearce, S. J. Bruce, N. Hardy, S.-A. Sansone, A. Antti, P. Jonsson, C. Daykin, M. Navarange, R. D. Beger, E. R. Verheij, A. Amberg, D. Baunsgaard, G. H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J. N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I. D. Wilson, M. D. Reilly, D. G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer and M. Spraul, *Nat. Biotechnol.*, 2005, **23**, 833–838.