Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results

Arthur L. Castle, Oliver Fiehn, Rima Kaddurah-Daouk and John C. Lindon

Informatics standards and controlled vocabularies are essential for allowing information technology to help exchange, manage, interpret and compare large data collections. In a rapidly evolving field, the challenge is to work out how best to describe, but not prescribe, the use of these technologies and methods. A Metabolomics Standards Workshop was held by the US National Institutes of Health (NIH) to bring together multiple ongoing standards efforts in metabolomics with the NIH research community. The goals were to discuss metabolomics workflows (methods, technologies and data treatments) and the needs, challenges and potential approaches to developing a Metabolomics Standards Initiative that will help facilitate this rapidly growing field which has been a focus of the NIH roadmap effort. This report highlights specific aspects of what was presented and discussed at the 1st and 2nd August 2005 Metabolomics Standards Workshop.

Keywords: *metabolomics; metabonomics; minimum information standards; metabolic profiling*

INTRODUCTION

The Metabolomics Standards Workshop was sponsored by the NIH and organized with the help of the Metabolomics Society to bring together multiple ongoing standards efforts with the National Institutes of Health (NIH) research community [1–4]. The following is a briefing on this workshop and the full proceedings and presentations can be viewed at the NIH and Metabolomics Society websites [5, 6].

The study of global metabolic changes in biological systems has been termed metabolomics [7] and metabonomics [8]. Metabolomics uses comprehensive analytical methods to study metabolism, metabolites and other small molecules within cells and tissues. These collective small molecule constituents, or metabolome, along with the proteome and genome, comprise the building blocks of living systems. Application of metabolomics technology has rapidly expanded from early metabolomics studies involving plant physiology and animal toxicity on to many other disciplines and forms an important component in studying systems biology of organisms along with genomics and proteomics [7–10]. Metabolites within metabolomes can also be shared among organisms forming an ecosystem that describe functional interactions between organisms and their environments [9, 11].

Global metabolomics approaches are possible because of multiple high throughput technologies

Corresponding author. Arthur L. Castle, PhD Program Director, Metabolomics and Informatics, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health (NIH). 6707 Democracy Blvd, Room 791, MSC 5460, Bethesda, Maryland 20892-5460, Tel: 301-594-7719; Fax: 301-480-0475. E-mail: castlea@mail.nih.gov

Arthur Castle is the NIH Program Director for the Roadmap Metabolomics Technology Development program and the NIDDK Program Director for Metabolomics and Informatics.

Dr Oliver Fiehn is an Associate Professor at the UC Davis Genome Center and chairs the efforts on 'Standards for Reporting Metabolomics Studies' for the Metabolomics Society.

Rima Kaddurah-Daouk is an Associate Research Professor at Duke University Medical Center and the current President of the Metabolomics Society.

John Lindon is a Professor in the Biological Chemistry section of the Biomedical Sciences Division of Imperial College London, UK. His research interests lie in the development and application of analytical and chemometric techniques to studies of endogenous and xenobiotic metabolism.

including nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry (MS) and molecular nano-sensors (MNS) and is dependent on informatics systems that can support this wealth of data collections [12–15]. Informatics standards and controlled vocabularies are essential to allow the application of information technology to help exchange, manage, interpret and compare large data collections [3, 15–17]. The challenge is to work out how best to describe, but not prescribe, the use of these technologies and methods. Accepted minimum standards need to be developed for a reporting structure of metabolomic data and the underlying study designs. These standards will be different from current best practice protocols that are rapidly changing and which also might report on far more details and parameters. What may be the best methods and technologies today may become antiquated tomorrow.

In general, informatics of metabolomics can be broken into information about the experiment; the analytical aspects of instrumentation and sample processing; and the chemometric identification, verification and description of the metabolome components [2, 3, 15]. Understanding the biological context of how samples were acquired is essential since the metabolome is very dynamic and dependent on many internal and external environmental influences [3, 18]. The analysis and presentation of metabolomics information needs to be described so that this information can be used by others to confirm, refute or validate experimental findings, and tools need to be developed to implement the use of these standards [3, 15]. Careful examination of the metabolomics technologies, methodologies and resources in use or being developed will help define the need for informatics standards and associated tools.

PLATFORMS, ACQUISITION, PROCESSING AND ANALYSIS

Metabolomics information is derived from heterogeneous platforms that vary in their technical and scientific foundations with each technology and biological matrix presenting its own set of challenges in providing concise descriptions that allow for proper interpretations of the metabolic information acquired [3, 16, 19, 20]. All platforms infer some aspects of metabolite abundance, spatial and temporal changes in concentrations, transformation of mass between molecular species or the transfer of information between pathways, tissue, organs or organisms [21–23]. Definitive descriptions of all these aspects are not always possible and these technologies need to be described so that a confidence in the aspects of the measures acquired can be determined. Ultimately, a quantitative or qualitative metabolite measurement is made and its description using a standard such as the recently developed International Chemical Identifier (InChI) coding will be essential for interoperability of informatics systems [24].

MS and NMR spectroscopy are the most common technology platforms used in metabolomics research [25–34]. They are often coupled with other chromatography techniques or modified to provide multiple measurement dimensions based on physical or chemical properties of metabolites [31–36]. Other detection platforms such as infra red spectroscopy or electrochemical detectors can also be used to provide additional measurement dimensions of metabolite characteristics [37]. Each detection or separation technique has different abilities in providing high sample throughput, resolution in identifying metabolites and sensitivity limitations. In addition to technologies that can measure many metabolites at once, new probes and sensors that are specific to individual metabolites are being produced with high throughput biotechnology techniques. These probes are useful in defining the temporal and spatial aspects of metabolites within biological systems [12, 13]. Some of these techniques can also be used in conjunction with isotope tracing studies to measure the flux of metabolite conversions between molecular species [29, 30]. All techniques need to be described in enough detail to reproduce results, and common informatics standards for describing these platform configurations are needed [34, 38, 39].

Metabolites are identified through the use of data libraries that contain physical and chemical properties of many metabolites as detected by each technology platform [3, 26, 34, 38–45]. Pure metabolites standards and well-characterized biological matrix of metabolites can also be used to confirm metabolite characteristics and calibrate equipment to improve library identification [38, 40–45]. Often, libraries for a single platform cannot uniquely identify a metabolite due to the lack of unambiguous physical and chemical properties being detected or resolution of the technology to decipher small differences. Identical mass, isomeric and chiral properties can be especially difficult to decipher [27, 28].

Therefore, the full description of metabolomes will likely require multiplexing multiple technologies which present many different dimensions of chemical information for every molecule in a complex sample and their temporal spatial relationships. Cooperation with instrument suppliers will be essential in describing these technologies. Computers will need to decipher and collapse all these dimensions in order to infer the chemical make up of a metabolome. Leveraging of extensive libraries that must communicate with common ontologies will likely be required for rapid metabolomics analysis [14].

Comparing metabolomics information derived from different platforms is a formidable challenge since definitive identification of most metabolites within complex biological fluids is difficult. Some initial studies have achieved a correlation between NMR data sets and between MS- and NMR–based data sets for improved biomarker identification using analysis of covariance across a cohort of sample sets [35, 36].

Currently, the extent of libraries, standards and analysis tools are insufficient to easily decipher the complete chemical make up of metabolomes. The current technologies; however, can detect signals from many, if not most, metabolites and confirmatory analysis can be used to identify the metabolite sources of these signals. The statistical analysis of these signals without source identification may provide useful biomarkers for classification and complex multivariate signals or signatures have demonstrated utility in toxicological applications; however, molecular identification is important for mechanistic analysis [14, 46, 47]. Describing these unknowns for future identification, mechanistic follow-up or as classification tools is a large challenge [2, 3, 14, 48]. Publishing all raw signal data is not yet practical, but properties of the informative signals need to be described such that exactly the same technology is not needed to reproduce similar signals if this information is to be easily verified and expanded upon by other investigators. Some access to raw signal data, however, could be valuable in improving analysis techniques. Therefore, it was recommended that raw data be kept within local laboratory information management systems (LIMS) systems; however, development of systems with some public access to this data and protocol information via internet portals should be encouraged [43, 44, 49, 50].

EXPERIMENTAL DESIGNS AND DESCRIPTIONS

Metabolomics profiles are dependent on both the biological state of the organism from which they were obtained and any subsequent processing techniques. As with genomics and proteomics, there is a need to describe experiments such that computers can aid optimally in data mining [2–4, 16, 17, 22, 38]. This requires the development of both ontologies, minimal reported information about the experiment and tools to implement them. While each field has specific experimental techniques and terminology, any harmonization across fields will be beneficial to advance system biology approaches. This will be especially important with genomics and proteomics in order to study pathways and networks.

Biological reference information is crucial to understanding the metabolome. Nutritional and biorhythm effects can greatly complicate and, if not analysed properly, could confound metabolomics experiments [9, 51, 52]. It is necessary to describe in enough detail the environment and state of the organism so that experiments can be compared or repeated to produce similar results. This is the core biological context of the measurement event and the type of measurement made.

The procedures used in processing samples can affect the reported metabolite concentrations [47, 53]. Even small changes in centrifugation of plasma have been shown to change recovery of some metabolites and many metabolites are rapidly degraded or created after samples have been extracted. Contamination from processing vessels is also a potential confounding factor. Some of these difficulties can be ameliorated by using standard reference material; however, enough descriptive information needs to be provided to allow for proper interpretation of results and knowledge about how sensitive metabolites are to procedural changes is also important.

A balance between necessary descriptions of sample collection, preparation protocol and instrumentation and the burden to investigators in describing all details needs to be developed. This balance will likely depend on the use of the information. The level of descriptive detail involved in standard operating procedures (SOPs), good laboratory practices (GLP) and good manufacturing practices (GMP) are generally used mostly for regulatory submission and not basic science where

such stringent record-keeping may consume too many resources to implement, limit productivity and reduce compliance to standards. Therefore, minimal information standards for sample collection, preparation and technology description might vary depending on the use of this information with submission to regulatory agencies having specific requirements [54]. Minimum information standards should not include unnecessary and burdensome details that are unnecessary to reproduce similar results; however, investigators are highly encouraged to record these details for their laboratory management. Developing a greater depth of reported information for 'minimum standards' may be implemented by individual academic principal investigators leading by example.

RESOURCES, INTEGRATION AND INTEROPERABILITY

A network of integrated databases and resources is needed to meet the challenge of metabolomics research. Properties of these databases can be summarized into four major categories: reference databases that have a large collection of single or highly limited data types; repositories that have many data types focused around a theme such as an organism, disease or technology; indexes and tools that access many other databases or resources and laboratory management information systems that help organize complex experiments [2, 14, 55]. While many databases have several of these properties, one usually dominates. The types of data may include information about a metabolite and its biological relevance; the instrumentation, technology and methodology used to make the measurement; and the experimental design and context in which the observation was made.

Public signal pattern databases for MS and NMR exist to connect signal observations with chemical entities [40, 41, 43, 56]. Chemical indexes such as PubChem, Chem Abstracts, Chemical Entities of Biological Interest (ChEBI) provide the ability to search for chemical entities and their basic properties and link to other sources [57–59]. These links are dependent on proper unambiguous interpretation of chemical nomenclature such as the INChI standard [24]. Once identified, these metabolites then need to be linked to biological context of written literature, stored experimental data, and biological knowledge databases of biological pathways and networks [57–59]. Pathway databases provide the system

biology relationships between metabolites, proteins and genes, and pattern databases support the link between observation from technology platforms and chemical indexes [60, 61]. Public pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) can provide information on tens of thousands of pathways from hundreds of organisms [60]. Projects such as Lipid Maps have several databases of these types connected around a lipid theme [43, 44]. Other projects have collections centered around organisms or tissues [42, 62].

The reference information within these databases ultimately comes from data reduction of experimental data. Any laboratory that produces metabolomics information needs to manage large collections of metabolite measurements and metadata supporting these measurements. Information from these LIMS systems needs to flow from experiment to reference and theme, and back again to interpret metabolomics information.

Flow of information from one database to another requires common language centred on biology and chemistry [17, 22, 44]. Minimum information about studies, technology used and metabolites measured are needed to compare experimental data while reference databases are needed to connect the metabolites to biological pathways and networks [2–4, 7, 16, 17, 22, 57–59]. Semantic compatibility is necessary for the flow of information from laboratory databases to reference databases and to reduce and reuse this information in further analysis. Syntax between all these databases must have commonality or be translated for information to flow. Therefore, informatics standards and developed tools to implement these standards are essential for efficient metabolomics research.

CURRENT EFFORTS IN METABOLOMICS STANDARDS

Developing informatics standards for metabolomics will require cooperation from many diverse communities. The Standard Metabolic Reporting Structure (SMRS) and an Architecture for Metabolomics (ArMET) consortia, and the recent MetaboMeeting have already built considerable consensus within the metabolomics community [3, 4, 15]. With the help of the Metabolomics Society, the European Bioinformatics Institute (EBI) and NIH funded researchers; this collaboration should continue to grow. Each discipline that uses metabolomics

technology will need to develop and harmonize their efforts with other disciplines. In addition, equipment suppliers and other industrial producers or users of metabolomics technologies should be involved in development and implementation of standards. It is expected that mistakes will be made as standards are developed. Therefore, it must be a broadly based community effort that is adaptable and extensible. These efforts in turn will need to be consolidated with proteomics, genomics and system biology standards. Further research into the integration of transcriptomics, proteomics and metabolomics data sets is required.

Metabolomics studies produce large amounts of data that could be reused by other investigators to help confirm, interpret or validate research findings. Leveraging of this information depends on having reporting standards with common ontology and structure on how metabolomics experiments should be reported. To this end, it was agreed that working groups be established under the guidance of the Metabolomics Society with the help of the EBI. These groups would involve the community in developing a 'Reporting Standards in Metabolomics' (RSM) document. Noting that a standard that is broadly developed by the community and generally accepted as necessary, and not overly burdensome is more likely to be followed.

The scope of this document will be to formulate minimum reporting standards that describe metabolomics experiments, but do not prescribe how to perform these experiments. An oversight committee and five subcommittees were established as outcomes of this workshop [63]. The subcommittee areas are (i) biological sample context, (ii) chemical analysis, (iii) data analysis, (iv) ontology and (v) data exchange. A draft of these standards will be discussed at the June 2006 Metabolomics Society meeting [64]. Investigators wanting to contribute to these efforts can find the appropriate contact information at the Metabolomics Society website [63].

Informatics standards along with large data collections and physicals standards are needed to form an infrastructure necessary to support the rapid identification of metabolites, and the use of metabolomics. As with any infrastructure, all parts are needed for the whole to work effectively and rapid concurrent development of these resources will speed the use of metabolomics in prognostic, diagnostics and quality assurance medical, nutritional and environmental applications.

Key Points

- The reporting of minimal information about experimental design, technology and analysis with common ontologies are necessary to facilitate computer-assisted use of metabolomics information.
- These informatics standards must describe and not prescribe the use of this technology and
- It should be harmonized with other informatics standards efforts in genomics and proteomics to build towards comprehensive system biology descriptions.
- A Metabolomics Standards Initiative has begun to address these challenges.

References

1. National Institutes of Health. NIH Roadmap for Medical Research Public Meetings and Workshops. http://nihroadmap.nih.gov/publicmeetings.asp (28 March 2006, date last accessed).

2. Jenkins H, Hardy N, Beckmann M, *et al*. A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* 2004;**22**: 1601–6.

3. Lindon JC, Nicholson JK, Holmes E, *et al*. Standard Metabolic Reporting Structures working group. Summary recommendations for standardization and reporting of metabolic analyses. *Nat Biotechnol* 2005;**23**:833–8.

4. Standard Metabolic Reporting Structures (SMRS) Group. MetaboMeeting 1.0 July 18th-19th, 2005 — Cambridge, UK. http://smrsgroup.sourceforge.net/metabomeeting. html (28 March 2006, date last accessed).

5. National Institutes of Health. NIH VideoCasting. http://www.videocast.nih.gov/ (28 March 2006, date last accessed).

6. Metabolomics Society. http://www.metabolomicssociety. org/ (28 March 2006, date last accessed).

7. Fiehn O. Metabolomics–the link between genotypes and phenotypes. *Plant Mol Biol* 2002;**48**:155–71.

8. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* 2002;**1**:153–61.

9. Gibney MJ, Walsh M, Brennan L, *et al*. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr* 2005;**82**:497–503.

10. Ishii N, Robert M, Nakayama Y, *et al*. Toward large-scale modeling of the microbial cell for computer simulation. *Biotechnol* 2004;**113**:281–94.

11. Nicholls AW, Mortishire-Smith RJ, Nicholson JK. NMR spectroscopic-based metabonomic studies of urinary metabolite variation in acclimatizing germ-free rats. *Chem Res Toxicol* 2003;**16**:1395–404.

12. Deuschle K, Fehr M, Hilpert M, *et al*. Genetically encoded sensors for metabolites. *Cytometry A* 2005;**64**:3–9.

13. Richieri GV, Ogata RT, Kleinfeld AM. A fluorescently labeled intestinal fatty acid binding protein. Interactions with fatty acids and its use in monitoring free fatty acids. *J Biol Chem* 1992;**267**:23495–501.

14. Mendes P. Emerging bioinformatics for the metabolome. *Brief Bioinform* 2002;**3**:134–45.

15. Jenkins H, Johnson H, Kular B, *et al*. Toward supportive data collection tools for plant metabolomics. *Plant Physiol* 2005;**138**:67–77.

16. Brazma A, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;**29**: 365–71.

17. Harris MA, Clark J, Ireland A, *et al*. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;**32**:D258–61.

18. Fiehn O, Wohlgemuth G, Scholz M. Automatic annotation of metabolomic mass spectra by integrating experimental metadata. *Proc Lect Notes Bioinformatics* 2005;**3615**:224–39.

19. Birkemeyer C, Luedemann A, Wagner C, *et al*. Metabolome analysis: the potential of *in vivo* labeling with stable isotopes for metabolite profiling. *Trends Biotechnol* 2005;**23**:28–33.

20. Villas-Boas SG, Hojer-Pedersen J, Akesson M, *et al*. Global metabolite analysis of yeast: evaluation of sample prepara-tion methods. *Yeast* 2005;**22**:1155–69.

21. Nielsen J, Oliver S. The next wave in metabolome analysis. *Trends Biotechnol* 2005;**23**:544–6.

22. Hucka M, Finney A, Sauro HM, *et al*. SBML Forum The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003;**19**:524–31.

23. Schad M, Mungur R, Fiehn O, Kehr J. Metabolic profiling of laser microdissected vascular bundles of *Arabidopsis thaliana*. *Plant Methods* 2005;**1**:2.

24. International Union of Pure and Applied Chemistry. The IUPAC International Chemical Identifier (InChITM). http://www.iupac.org/inchi/ (28 March 2006, date last accessed).

25. Cooks RG, Busch KL, Glish GL. Mass spectrometry: analytical capabilities and potentials.. *Science* 1983;**222**:273–91.

26. Fredenhagen A, Derrien C, Gassmann E. An MS/MS library on an ion-trap instrument for efficient dereplication of natural products. Different fragmentation patterns for $[M+H]+$ and $[M+Na]+$ ions. *J Nat Prod* 2005;**68**: 385–91.

27. Lee SH, Williams MV, Blair IA. Targeted chiral lipidomics analysis. *Prostaglandins Other Lipid Mediat* 2005;**77**:141–57.

28. Steiner WE, Clowers BH, Hill HH, Jr. Rapid separation of phenylthiohydantoin amino acids: ambient pressure ion-mobility mass spectrometry (IMMS). *Anal Bioanal Chem* 2003;**375**:99–102.

29. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable *Escherichia Coli* adaptation to loss of key metabolic enzymes. *J Biol Chem* 2005;**281**:8024–33.

30. Brunengraber H, Kelleher JK, Des Rosiers C. Applications of mass isotopomer analysis to nutrition research. *Annu Rev Nutr* 1997;**17**:559–96.

31. Griffin JL. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for character-ization of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 2003;**7**:648–54.

32. Lindon JC, Nicholson JK, Wilson ID. Directly coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *J Chromatogr B Biomed Sci Appl* 2000;**748**:233–58.

33. Tang H, Wang Y, Nicholson JK, Lindon JC. Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Anal Biochem* 2004;**325**:260–72.

34. Davies AN, Lambert J, Lancashire RJ, *et al*. Guidelines for the representation of pulse sequences for solution-state nuclear magnetic resonance spectrometry (IUPAC recommendations 2001). *Pure Appl Chem* 2001;**73**: 1749–64.

35. Cloarec O, Dumas ME, Craig A, *et al*. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal Chem* 2005;**77**:1282–89.

36. Crockford DJ, Holmes E, Lindon JC, *et al*. Statistical HeterospectroscopY (SHY), a new approach to the integrated analysis of NMR and UPLC-MS datasets: application in metabonomic toxicology studies. *Anal Chem* 2006;**78**:363–71.

37. Kristal BS, Vigneau-Callahan KE, Matson WR. Simultaneous analysis of the majority of low-molecular-weight, redox-active compounds from mitochondria. *Anal Biochem* 1998;**263**:18–25.

38. Orchard S, Hermjakob H, Taylor CF, *et al*. Further steps in standardisation. Report of the second annual Proteomics Standards Initiative Spring Workshop (Siena, Italy, 17–20th April 2005). *Proteomics* 2005;**5**: 3552–5.

39. Markley JL, Bax A, Arata Y, *et al*. Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol* 1998;**280**:933–52.

40. National Institute of Standards and Technology. NIST Chemistry WebBook. http://webbook.nist.gov/chemistry/ (28 March 2006, date last accessed).

41. The Scripps Research Institute. METLIN Metabolite Database. http://metlin.scripps.edu/ (28 March 2006, date last accessed).

42. American Academy of Forensic Sciences, Toxicology Section. Mass Spectrometry Database Committee. http://www.ualberta.ca/~gjones/mslib.htm (28 March 2006, date last accessed).

43. LIPID MAPS Consortium. LIPID Metabolites And Pathways Strategy. http://www.lipidmaps.org/ (28 March 2006, date last accessed).

44. Fahy E, Subramaniam S, Brown HA, *et al*. A compre-hensive classification system for lipids. *J Lipid Res* 2005;**46**: 839–861.

45. Adamski M, Blackwell T, Menon R, *et al*. Data manage-ment and preliminary data analysis in the pilot phase of the HUPO Plasma Proteome Project. *Proteomics* 2005;**5**: 3246–61.

46. Lindon JC, Keun HC, Ebbels TM, *et al*. The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* 2005;**6**:691–9.

47. Robertson DG. Metabonomics in toxicology: a review. *Toxicol Sci* 2005;**85**:809–22.

48. Bino RJ, Hall RD, Fiehn O, *et al*. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 2004;**9**: 418–25.

49. Zolnai Z, Lee PT, Li J, *et al*. Project management system for structural and functional proteomics: Sesame. *J Struct Funct Genomics* 2003;**4**:11–23.

50. University of Wisconsin-Madison. Center for Eukaryotic Structural Genomics – Sesame Page. http://www. uwstructuralgenomics.org/sesame.htm (28 March 2006, date last accessed).

51. Bollard ME, Stanley EG, Lindon JC, *et al*. NMR–based metabonomic approaches for evaluating physiological influences on biofluid composition.NMR. *Biomed* 2005;**18**: 143–62.

52. Vigneau-Callahan KE, Shestopalov AI, Milbury PE, *et al*. Characterization of diet-dependent metabolic serotypes: analytical and biological variability issues in rats. *J Nutr* 2001; **131**:924S–32S.

53. Shurubor YI, Matson WR, Martin RJ, Kristal BS. Relative contribution of specific sources of systematic errors and analytical imprecision to metabolite analysis by HPLC–ECD. *Metabolomics* 2005;**1**:159–168.

54. US Food and Drug Administration. NCTR Center for Toxicoinformatics – ArrayTrack Microarray Database and Analysis. http://www.fda.gov/nctr/science/centers/ toxicoinformatics/ArrayTrack/ (28 March 2006, date last accessed).

55. Reiser L, Mueller LA, Rhee SY. Surviving in a sea of data: a survey of plant genome data resources and issues in building data management. *Plant Mol Biol* 2002;**48**:59–74.

56. National Institute of Advanced Industrial Science and Technology (AIST), Japan. Spectral Database for Organic Compounds SDBS. http://www.aist.go.jp/RIODB/ SDBS/ (28 March 2006, date last accessed).

57. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health. The PubChem Project. http://pubchem.ncbi.nlm.nih.gov/ (28 March 2006, date last accessed).

58. American Chemical Society. Chemical Abstract Service Home Page, http://www.cas.org/ (28 March 2006, date last accessed).

59. European Bioinformatics Institute. Chemical Entities of Biological Interest (ChEBI), http://www.ebi.ac.uk/chebi/ (28 March 2006, date last accessed).

60. Kanehisa Laboratories. KEGG: Kyoto Encyclopedia of Genes and Genomes. http://www.genome.ad.jp/kegg/ (28 March 2006, date last accessed).

61. Cold Spring Harbor Laboratory and European Bioinformatics Institute. Reactome – a curated knowledge-base of biological pathways. http://www.reactome.org/ (28 March 2006, date last accessed).

62. Max Plank Institute of Molecular Plant Physiology. The Comprehensive System Biology Project (CSB). http://csbdb.mpimp-golm.mpg.de/ (28 March 2006, date last accessed).

63. Metabolomics Society. Metabolomics Society – Standardization in Metabolomics Experiments. http:// www.metabolomicssociety.org/mstandards.html (28 March 2006, date last accessed).

64. Metabolomics Society. Metabolomics Society – Scientific Events. http://www.metabolomicssociety.org/boston2006. html (28 March 2006, date last accessed).