# CIMR: Chemical Analysis Group

## Metabolomics Standards Initiative (MSI)

**Sponsor:** Metabolomics Society http://www.metabolomicssociety.org/
**Reference:** http://msi-workgroups.sourceforge.net/bio-metadata/reporting/pbc/
**Version:** $Revision: 1.1 $
**Date:** 2006/06/19
**Authors:** Alexander Amberg, Aventis <Alexander.Amberg@aventis.com>
Dave Barrett, Univ. of Nottingham <david.barrett@nottingham.ac.uk>
Mike Beale, Met-RO <mike.beale@bbsrc.ac.uk>
Clare Daykin, Univ of Nottingham <clare.daykin@nottingham.ac.uk>
Teresa W-M. Fan, University of Louisville, KY <twfan@ucdavis.edu>
Oliver Fiehn, UC Davis <ofiehn@ucdavis.edu>
Jules Griffin, Cambridge <jlg40@mole.bio.cam.ac.uk>
Richard Higashi, UC Davis <rmhigashi@ucdavis.edu>
Joachim Kopka, Max Planck MPI <Kopka@mpimp-golm.mpg.de>
John Lindon, Imperial College <j.lindon@imperial.ac.uk>
Andrew Lane, Univ. of Louisville <anlane01@gwise.louisville.edu>
Andrew Nicholls, GSK <andrew.w.nicholls@gsk.com>
Michael Reily, Pfizer <Michael.Reily@pfizer.com>
Lloyd W. Sumner, The Noble Foundation <lwsumner@noble.org>
**Credits:** *Co-chair*: Lloyd W. Sumner.



## Table of Contents



# 1. This document

This document forms part of the the standards for reporting metabolomics experiments developed under the Metabolomics Society (http://www.metabolomicssociety.org/Metabolomics Standards Initiative (MSI). It should be read in the context of top level document for those standards ??? .

The current version of the document is work in progress. ???.

# 2. Scope and Goals

## 2.1. Scope of The Chemical Analysis Group

The scope of our efforts will be to identify, develop and disseminate best chemical analysis practices in all aspects of metabolomics. The proposed standards will be consistent with good analytical chemistry practices with extra provisions for metabolomic data (the main difference being large numbers of data-sets and the need to be able compare them electronically) and will be in alignment with those typically required by quality analytical journals.

The aim will not be to prescribe how to perform a metabolomics experiments, but to formulate a minimum set of reporting standards that describe the experiments (what the experiments are and how they were actually executed). Consequently, there will be no attempt to restrict or dictate specific practices, but to develop consistent and appropriate descriptors to support the dissemination and re-use of metabolomic data. Such reporting standards will specify the data identified as necessary for complete and comprehensive reporting in a range of identified contexts, such as submission to academic journals and public databases. Data exchange standards will be developed to provide a transparent technical vehicle which meets or exceeds the requirements of reporting standards.

## 2.2. The Goals of The Chemical Analysis Group are to:

1. work cooperatively on a consensus draft for a minimum core set of necessary data related to the chemical analyses associated with metabolomics experiments.
2. include key persons from the Group#s specialist area to participate in the discussion in an inclusive manner.
3. reach out and evaluate previous and relevant work in their specialist areas including similar work in transcriptomics and proteomics studies, and recent metabolomics standardization efforts.
4. pay careful attention to the distinction of best practice (which will change), reporting standards (which should have longer validity) and data exchange standards (which support reporting).
5. respond to documents from the other groups and produce an advanced draft ready for discussion in February 2006
6. respond to documents from the other groups and produce a final draft ready for discussion in June 2006
7. Invite editorial boards of Metabolomics, Phytochemistry, Analytical Chemistry, _____??? To review and advise on the practicality, acceptability, and support of standards.

   Specific editorial contacts:
   a. Roy Goodacre (Metabolomics)
   b. Dieter Strack (Phytochemistry)
   c. John Yates (Analytical Chemistry)
   d. ???

# 3. Related Work

## 3.1. Related Literature

[1] Bino, R. J. and Hall, R. D. and Fiehn, O. and Kopka, J. and Saito, K. and Draper, J. and Nikolau, B. J. and Mendes, P. and Roessner-Tunali, U. and Beale, M. H. and Trethewey, R. N. and Lange, B. M. and Wurtele, E. S. and Sumner, L. W.. *Potential of metabolomics as a functional genomics tool*. *Trends In Plant Science*. 9. 9. 418-425. 2004.

[2] Jenkins, H. and Hardy, N. and Beckmann, M. and Draper, J. and Smith, A. R. and Taylor, J. and Fiehn, O. and Goodacre, R. and Bino, R. J. and Hall, R. and Kopka, J. and Lane, G. A. and Lange, B. M. and Liu, J. R. and Mendes, P. and Nikolau, B. J. and Oliver, S. G. and Paton, N. W. and Rhee, S. and Roessner-Tunali, U. and Saito, K. and Smedsgaard, J. and Sumner, L. W. and Wang, T. and Walsh, S. and Wurtele, E. S. and Kell, D. B.. *A proposed framework for the description of plant metabolomics experiments and their results*. *Nature Biotechnology*. 22. 12. 1601-1606. 2004.

[3] Jenkins, Helen and Johnson, Helen and Kular, Baldeep and Wang, Trevor and Hardy, Nigel. *Toward Supportive Data Collection Tools for Plant Metabolomics*. *Plant Physiol. %R*

*10.1104/pp.104.058875*. 138. 1. 67-77. 2005.

[4] Lindon, J. C. and Nicholson, J. K. and Holmes, E. and Keun, H. C. and Craig, A. and Pearce, J. T. M. and Bruce, S. J. and Hardy, N. and Sansone, S. A. and Antti, H. and Jonsson, P. and Daykin, C. and Navarange, M. and Beger, R. D. and Verheij, E. R. and Amberg, A. and Baunsgaard, D. and Cantor, G. H. and Lehman-McKeeman, L. and Earll, M. and Wold, S. and Johansson, E. and Haselden, J. N. and Kramer, K. and Thomas, C. and Lindberg, J. and Schuppe-Koistinen, I. and Wilson, I. D. and Reily, M. D. and Robertson, D. G. and Senn, H. and Krotzky, A. and Kochhar, S. and Powell, J. and van der Ouderaa, F. and Plumb, R. and Schaefer, H. and Spraul, M.. *Summary recommendations for standardization and reporting of metabolic analyses*. *Nature Biotechnology*. 23. 7. 833-838. 2005.

[5] Orchard, S. and Hermjakob, H. and Apweiler, R.. *The proteomics standards initiative*. *Proteomics*. 3. 7. 1374-1376. 2003.

[6] Orchard, Sandra and Hermjakob, Henning and Taylor, Chris and Aebersold, Ruedi and Apweiler, Rolf. *Human Proteome Organisation Proteomics Standards Initiative Pre-Congress Initiative*. *PROTEOMICS*. 5. 18. 4651-4652. 2005.

[7] Quackenbush, John. *Data standards for 'omic' science*. 22. 5. 613. 2004.

## 3.2. Related Internet Sites

http://www.smrsgroup.org/
http://www.niddk.nih.gov/fund/other/metabolomics2005/
http://www.metabolomicssociety.org/nih.html
http://www.mged.org/Mission/index.html#DefinedMGEDStandards
http://psidev.sourceforge.net/
http://www.mpdg.org/

# 4. Proposed Minimum Information Set for Reporting Chemical Analysis

## 4.1. Proposed Minimum Metadata for Sample Preparation

### 4.1.1. Sample harvesting protocol

| | |
|---|---|
| harvesting method | |
| harvesting time | |
| harvesting duration | to include time, temp |
| $LN_2$ | |
| lyophilization | |
| fresh tissue processing | |
| sample storage prior to further preparation | (-80C for 2 weeks). All temperatures should be measured if possible; however temp setpoints are acceptable assuming quality monitoring was performed and no abnormalities recorded. |

### 4.1.2. Extraction

| | |
|---|---|
| solvent | |
| volume | |
| quantity tissue | 1ml MeOH per 6mg lyophilized tissue |

### 4.1.3. Extract concentration, and resuspension processes

Dried under nitrogen, resuspended in H2O or pyridine,

### 4.1.4. Sample enrichment (if relevant)

SPE (column, sorbent, manufacturer)
Desalting, MWCO etc.

### 4.1.5. Derivatization

| | |
|---|---|
| OMS/ TMS | (temperatures & duration) |

## 4.2. Proposed Minimum Metadata Relative to Chromatography

### 4.2.1. Chromatography Instrument

| | |
|---|---|
| Manufacturer | |
| model number | |
| software package | |
| software package version number or date | |

### 4.2.2. Auto-injector

| | |
|---|---|
| Injector model/type | |
| software version | |
| method name | |
| injection volume | |
| wash cycles (volumes) | |
| solvent | |

### 4.2.3. Separation column and pre/guard column

| | |
|---|---|
| manufacturer | |
| product # | |
| stationary media composition | (support and coating e.g. silica C8 etc) |
| stationary media physical parameters | (i.e. coating thickness for GC/MS & particle size and pore size for LC/MS) |
| internal diameter | |
| length | |
| column temperature | |
| flow rate | |

### 4.2.4. Separation parameters

| | |
|---|---|
| method name | |
| injector temperature | |
| split or splitless mode | |
| spli ratio | (if relevant) |
| mobile phase compositions | |
| mobile phase flow rates | |
| thermal/solvent/solute gradient profiles | |

### 4.2.5. Quality Control to validate chromatography performance

| | |
|---|---|
| | Minimum should include description whether or not QC was performed and how it was measured |
| validation sample | |
| internal standards | |
| chromatographic resolution | |
| cycles per column/injector/septum/blank | |

### 4.2.6. Data acquisition

| | |
|---|---|
| SOP Protocol name | |
| date | |
| operator | |
| publication reference | (can be journal or website URL but should be publicly accessible and link should be stable) |

# 4.3. Proposed Minimum Metadata Relative to Mass Spectrometry

## 4.3.1. Instrument

manufacturer
model #
operational software name
operational software version

## 4.3.2. Ionization source

| | |
|---|---|
| Ionization mode | (EI, APCI, ESI#.) |
| polarity | |
| vacuum pressure | |
| skimmer/focusing lens voltages | (e.g. capillary voltage etc.) |
| gas flows | (e.g. nebulization gas, cone gas etc. |
| source temperature | |

## 4.3.3. Mass Analyzer

| | |
|---|---|
| Type | |
| m/z range | |
| calibration | |
| resolution | |
| mass accuracy | |
| logic program for data acquisition | |
| spectral acquisition rate | |
| vacuum pressure | |
| lock spray | (concentration, lock mass, flow rate, frequency tec) |

## 4.3.4. Quality Control

tune
sensitivity
mass accuracy
resolution

## 4.3.5. Data acquisition

SOP Protocol name
date
operator
data acquisition rate

# 4.4. Proposed Minimum Metadata Relative to Metabolite Identification

- Most metabolites are not novel identifications *(i.e. previously characterized & identified using high rigor)*
- Propose a minimum of two independent domains relative to an authentic compound for metabolite identifications (retention time & mass spectrum- unit res molecular ion, or high res. Mass or fragmentation pattern?), (retention time & chemical shifts) , (accurate mass & tandem MS), (retention time & UV spectrum)
- More than two adds additional confidence (Rt, m/z, UV, chemical shift- one chemical shift is not very good- needs more detail such as multiplicity, 2D connectivity, etc)

# 4.5. Proposed Minimum Metadata Relative to Nuclear Magnetic Resonance

## 4.5.1. Instrument

| | |
|---|---|
| manufacturer | |
| model # | |
| magnetic field strength | in Tesla {example 14.1 T Varian Inova ; 18.8 T Bruker Avance} |

## 4.5.2. Hardware

| | |
|---|---|
| VT control | |
| pulsed field gradients | (z or x,y,z) |
| max gradient strength | |
| no. shims | |
| no channels | |
| probe type | (e.g. 10mm $^{31}$P, 5mm HCN coldprobe, 3mm flow-probe, etc.) |
| solution or solid-state | |
| automation or manual operation | |
| autotune or manual tune | |
| LC-NMR: sample handler | |
| LC-NMR: injection volumes | |
| LC-NMR: wash cycles | |

## 4.5.3. Sample property

| | |
|---|---|
| temperature | |
| volume | |
| extract/powder/intact organisms | |
| tissue or cells | |
| type of NMR tube | (e.g. conventional, Shigemi,, mircocell etc. ) |
| pH | |
| solvent | ($D_2O$, $CD_3OD$, $CDCl_3$, etc) |

## 4.5.4. Acquisition & Data Processing Parameter

### For 1-D NMR:

| | |
|---|---|
| observed nucleus | |
| pulse sequence name | |
| pulse sequence implementation | (e.g. gradient selection, sensitivity enhancement) |
| solvent saturation *or* | |
| decoupling method | |
| excitation pulse width | |
| spectral width | |
| acquisition time | |
| interpulse delay | (or recycle time) |
| digitization parameter | |
| number of transients | |
| technique | For solvent suppression |
| excitation maximum | For solvent suppression |
| excitation bandwidth | For solvent suppression |

### Additional parameters for 2-D NMR:

| | |
|---|---|
| observed nucleus in F2 and F1 | |
| pulse sequence | |
| excitation pulse widths | for relevant nuclei |
| spectral width in F2 and F1 | |
| solvent saturation method | |
| observed nucleus for F2 and F1 | |
| number of transients in t2 | |
| number of increments in t1 | |
| acquisition times for t2 and t1 | |

| | |
|---|---|
| tion | |

### For homonuclear NMR

| | |
|---|---|
| heteronuclear decoupling | presence or absence (e.g. isotope-enriched samples) |
| decoupling mode | |
| decoupling bandwidth | |
| spin lock field strength | (in Hz) |
| decoupling bandwidth | |
| spin lock field duration | (in sec) |
| mixing time | (NOESY, ROESY etc.) |

### For heteronuclear NMR

| | |
|---|---|
| direct or indirect detection | |
| proton decoupling mode | (Waltz, Garp, Wurst, Stud etc.) |
| proton decoupling effective band width | |
| evolution time | for constant time experiments |
| editing mode | (cf. INEPT-based experiments) |
| heteronuclear spin lock strength | |
| heteronuclear spin lock mixing time | (e.g. HCCH-TOCSY) |

## 4.5.5. Data Processing

| | |
|---|---|
| degree of zero filling | |
| degree of linear prediction | |
| apodization parameters and window function in all dimensions | (exponential, Gaussian, sine bell etc) |
| Baseline or baseline corrections | (dc offset, linear or non-linear corrections) |
| first point multipliers | |
| any shifting of the fids | |

## 4.5.6. Quality Control

| | |
|---|---|
| calibration (chemical shift & concentration) standard used | (e.g. DSS, TMS) |
| duplications | |
| standard error/deviation of quantification | |
| state the 50% and 1% line widths on the reference | (cf. DSS, TSP or TMS methyl peak) |
| for X nuclei: | external reference<br>external conditions<br>correction made for susceptibility effects |
| shift referencing method for indirect dimension in 2D experiments | (direct or indirect based on # ratios) |
| For quantation, state method used | (spiking with substance x at relative concentration y; intensity normalized to reference line- correction for saturation effects- T1 values measured?) |
| Any relaxation agents added | (type, amount)<br>For direct X-detection (esp. 13C or 31P), correction for NOEs as well as saturation?<br>For light water samples, what corrections are made for non-linear excitation profile and method? |
| gradient selection | |
| sensitivity enhancement | |
| solvent saturation or decoupling method | |
| excitation pulse width | |

| | | |
|---|---|---|
| spectral width | | |
| acquisition time | | |
| interpulse delay | | |
| digitization parameter | | |
| number of transients | | |
| apodization parameter | | |
| zero-filling parameter | | |
| for 2-D NMR | observed nucleus in F2 | |
| | observed nucleus in F1 | |
| | pulse sequence | |
| | excitation pulse widths | for relevant nuclei |
| | spectral width in F2 | |
| | spectral width in F1 | |
| | solvent saturation method | |
| | observed nucleus for F2 and F1 | |
| | number of transients in F2 | |
| | number of increments in F1 | |
| | acquisition time for F2 | |
| | acquisition time for F1 | |
| | apodization parameters for F2 | |
| | apodization parameters for F1 | |
| | linear prediction for F2 | |
| | zero-filling parameters for F2 | |
| | zero-filling parameters for F1 | |
| | decoupling mode | |
| | decoupling bandwidth | |
| | decoupling power | |
| | spin lock field strength | (in Hz) |
| | spin lock field duration | (in sec) (# water suppression method, macromolecule suppression method (e.g.spin-echo sequence)) |

### 4.5.7. Data acquisition

SOP Protocol name
date
operator

pH Markers??? (imidazole # this is used from Chenomx-software # calculates the difference between imidazole and creatinine)

Internal such as histidine in $^{1}H$ NMR; inorganic phosphate ($^{31}P$ NMR)- calibration with respect to external standard (see above).

## 4.6. Proposed Minimum Metadata Relative to Stable Isotopes & Flux Analysis ?

| | |
|---|---|
| compound | |
| element/isotope | |
| position(s) labeled, % | e.g. [$^{13}$C-1]-D-glucose 98%, [$^{15}N_2$]-L-glutamine (99%) |
| chemical purity | of the labeled compound(s) |
| concentration | of the compound used in the experiment |
| fraction of total present | (requires detailed breakdown of media composition for cell and tissue studies, including analysis of any added FCS or other growth supplements) |
| mode of labeling pulse | continuous addition; top up etc. |
| No. moles isotope | added during the experiment |

### 4.6.1. Data analysis

method for determining positional labeling

| | |
|---|---|
| method for determining fractional labeling | |
| standard error of the estimates | |
| estimated isotope recovery | in observable fractions (and fraction of total isotope supplied) |

## 4.7. Proposed Minimum Metadata Relative to Capillary Electrophoresis ?

## 4.8. Proposed Minimum Metadata Relative to Electrochemical Detection ?

## 4.9. Proposed Minimum Metadata Relative to Infrared Spectroscopy ?

## 4.10. Proposed Minimum Metadata for Data Export ?