Metabolomics Standards: Chemical Analysis Group

1.0 PARTICIPANTS

Co-chairs:

a. Lloyd W. Sumner, The Noble Foundation
b. Teresa W-M. Fan, University of Louisville, KY

Contributors:

c. Alexander Amberg, Aventis
d. Dave Barrett, Univ. of Nottingham
e. Mike Beale, MetRo
f. Clare Daykin, Univ of Nottingham
g. Oliver Fiehn, UC Davis
h. Jules Griffin, Cambridge
i. Richard Higashi, UC Davis
j. Joachim Kopka, Max Planck MPI
k. John Lindon, Imperial College
l. Andrew Lane, Univ. of Louisville
m. Andrew Nicholls, GSK
n. Michael Reily, Pfizer

Email List: Alexander.Amberg@aventis.com; mike.beale@bbsrc.ac.uk; andrew.w.nicholls@gsk.com; j.lindon@imperial.ac.uk; anlane01@gwise.louisville.edu; lwsumner@noble.org; clare.daykin@nottingham.ac.uk; david.barrett@nottingham.ac.uk; ofiehn@ucdavis.edu; rmhigashi@ucdavis.edu; twfan@ucdavis.edu; Kopka@mpimp-golm.mpg.de; Jules Griffin [jlg40@mole.bio.cam.ac.uk]; Reily, Michael [Michael.Reily@pfizer.com]

2.0 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.
    a. Specific editorial contacts
        i. Roy Goodacre (Metabolomics)
        ii. Dieter Strack (Phytochemistry)
        iii. John Yates (Analytical Chemistry)
        iv. ???

3.0 RELATED WORK

3.1 Related Literature

- Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of Metabolomics as a Functional Genomics Tool. Trends in Plant Science 9: 418-425
- Jenkins H, Hardy N, Beckmann M, Draper J, Smith A, Taylor J, Fiehn O, Goodacre R, Bino R, Hall R, Kopka J, Lane G, Lange B, Liu J, Mendes P, Nikolau B, Oliver S, Paton N, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner L, Wang T, Walsh S, Wurtele E, Kell D (2004) A proposed framework for the description of plant metabolomics experiments and their results. Nat Biotechnol 22: 1601-1606

- Jenkins H, Johnson H, Kular B, Wang T, Hardy N (2005) Toward Supportive Data Collection Tools for Plant Metabolomics. Plant Physiology 138: 67-77
- Lindon J, Nicholson J, Holmes E, Keun H, Craig A, Pearce J, Bruce S, Hardy N, Sansone S, Antti H, Jonsson P, Daykin C, Navarange M, Beger R, Verheij E, Amberg A, Baunsgaard D, Cantor G, Lehman-McKeeman L, Earll M, Wold S, Johansson E, Haselden J, Kramer K, Thomas C, Lindberg J, Schuppe-Koistinen I, Wilson I, Reily M, Robertson D, Senn H, Krotzky A, Kochhar S, Powell J, van der Ouderaa F, Plumb R, Schaefer H, Spraul M, (2005) Summary recommendations for standardization and reporting of metabolic analyses. Nat Biotechnol 23: 833-838
- Orchard S, Hermjakob H, Apweiler R (2003) The proteomics standards initiative. Proteomics 3: 1374-1376
- Orchard S, Hermjakob H, Taylor C, Aebersold R, Apweiler R (2005) Human Proteome Organisation Proteomics Standards Initiative. Pre-Congress Initiative. Proteomics 5: 4651-4652
- Quackenbush J (2004) Data standards for 'omic' science. Nat Biotechnol 22: 613-614

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/

4.0 PROPOSED MINIMUM INFORMATION SET FOR REPORTING CHEMICAL ANALYSIS

4.0 Proposed Minimum Metadata for Sample Preparation

- Sample harvesting protocol
  - Harvesting method, time & 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.
- Extraction solvent and volume per quantity tissue
  - 1ml MeOH per 6mg lyophilized tissue
- Extract concentration, and resuspension processes
  - Dried under nitrogen, resuspended in H2O or pyridine,
- Sample enrichment (if relevant)
  - SPE (column, sorbent, manufacturer)
  - Desalting, MWCO etc.
- Derivatization
  - OMS/TMS (temperatures & duration)

4.1 Proposed Minimum Metadata Relative to Chromatography

- Chromatography Instrument
  - Manufacturer, model number, software package and version number or date,
- Auto-injector
  - Injector model/type, software version, method name, injection volume, wash cycles (volumes); solvent,
- Separation column and pre/guard column
  - manufacturer, product #, stationary media composition (support and coating e.g. silica C8 etc) & physical parameters (i.e. coating thickness for GC/MS & particle size and pore size for LC/MS), internal diameter, length, column temperature, flow rate
- Separation parameters
  - Method name, injector temperature, split or splitless mode & ratio, mobile phase compositions, mobile phase flow rates, thermal/solvent/solute gradient profiles,
- 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
  -
- Data acquisition
  - SOP Protocol name, date, operator, and publication reference (can be journal or website URL but should be publicly accessible and link should be stable.
  -

4.2 Proposed Minimum Metadata Relative to Mass Spectrometry

- Instrument
  - manufacturer, model #, operational software name & version
- 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
- 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)
- Quality Control
  - Tune, sensitivity, mass accuracy, and resolution
- Data acquisition
  - SOP Protocol name, date, operator, data acquisition rate

4.3 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)

4.4 Proposed Minimum Metadata Relative to Nuclear Magnetic Resonance

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

  VT control, pulsed field gradients (z or x,y,z) and max gradient strength; no. shims, no channels;

  Probe type (e.g. 10 mm ^{31}P, 5 mm HCN coldprobe, 3 mm flow-probe, etc.), solution or solid-state, automation or manual operation, autotune or manual tune. LC-NMR: sample handler, injection volumes, wash cycles
- 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)
- 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, For solvent suppression, technique, excitation maximum and bandwidth should be stated.

  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 t_2 and number of increments in t_1, acquisition times for t_2 and t_1 ; phase sensitive or magnitude detection;

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

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

- 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.
- 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 and conditions, correction made for susceptibility effects.

  Shift referencing method for indirect dimension in 2D experiments (direct or indirect based on γ ratios).

  For quantation, state method used (spiking with substance x at relative concentration y; intensity normalized to reference line- correction for saturation effects- T_1 values measured?). Any relaxation agents added (type, amount). For direct X-detection (esp. ^{13}C or ^{31}P), correction for NOEs as well as saturation? 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 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 F2 and number of increments in F1, acquisition times for F2 and F1, apodization parameters for F2 and F1, linear prediction for F2 and F1, zero-filling parameters for F2 and F1; decoupling mode, bandwidth and power; spin lock field strength (in Hz) and duration (in sec) (→ water suppression method, macromolecule suppression method (e.g.spin-echo sequence))
- 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).

4.5 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%) and 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.

Data analysis: method for determining positional and fractional labeling, standard error of the estimates; estimated isotope recovery in observable fractions (and fraction of total isotope supplied)

4.6 Proposed Minimum Metadata Relative to Capillary Electrophoresis ?

4.7 Proposed Minimum Metadata Relative to Electrochemical Detection ?

4.8 Proposed Minimum Metadata Relative to Infrared Spectroscopy ?

4.9 Proposed Minimum Metadata for Data Export ?