[53] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 2006; 78: 779-87.
[54] Baran R, Kochi H, Saito N, *et al.* MathDAMP: a package for differential analysis of metabolite profiles. BMC Bioinformatics 2006; 7: 530.
[55] Bertsch A, Gröpl C, Reinert K, Kohlbacher O. OpenMS and TOPP: open source software for LC-MS data analysis. Methods Mol Biol 2011; 696: 353-67.
[56] Katajamaa M, Miettinen J, Orešič M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 2006; 22: 634-6.
[57] Sturm M, Bertsch A, Gröpl C, *et al.* OpenMS - an open-source software framework for mass spectrometry. BMC Bioinformatics 2008; 9: 163.
[58] Brodsky L, Moussaieff A, Shahaf N, Aharoni A, Rogachev I. Evaluation of peak picking quality in LC-MS metabolomics data. Anal Chem 2010; 82: 9177-87.
[59] Melamud E, Vastag L, Rabinowitz JD. Metabolomic analysis and visualization engine for LC-MS data. Anal Chem 2010; 82: 9818-26.
[60] García-PérezI, Vallejo M, García A, Legido-Quigley C, Barbas C. Metabolic fingerprinting with capillary electrophoresis. J Chromatogr A 2008; 1204: 130-9.
[61] van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 2006; 7: 142.
[62] Field D, Sansone SA, Collis A, *et al.* Megascience. 'Omics data sharing. Science 2009; 326: 234-6.
[63] 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.
[64] Hardy NW, Taylor CF. A roadmap for the establishment of standard data exchange structures for metabolomics. Metabolomics 2007; 3: 1573-3890.
[65] Pedrioli PG, Eng JK, Hubley R, *et al.* A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotechnol 2004; 22: 1459-66.
[66] Tanaka S, Aoshima K, Miura Y, YO. 57th ASMS Conference on Mass Spectrometry and Allied Topics (American Society for Mass Spectrometry). 2009; 31 May to 04 June; PA.
[67] Sugimoto M, Hirayama A, Ishikawa T, *et al.* Differential metabolomics software for capillary electrophoresis-mass spectrometry data analysis. Metabolomics 2010; 6: 27-41.
[68] Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics 2008; 9: 504.
[69] Yu W, He Z, Liu J, Zhao H. Improving mass spectrometry peak detection using multiple peak alignment results. J Proteome Res 2008; 7: 123-9.
[70] van Nederkassel AM, Daszykowski M, Eilers PH, Heyden YV. A comparison of three algorithms for chromatograms alignment. J Chromatogr A 2006; 1118: 199-210.
[71] Nordström A, O'Maille G, Qin C, Siuzdak G. Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. Anal Chem 2006; 78: 3289-95.
[72] Nevedomskaya E, Derks R, Deelder AM, Mayboroda OA, Palmblad M. Alignment of capillary electrophoresis-mass spectrometry datasets using accurate mass information. Anal Bioanal Chem 2009; 395: 2527-33.
[73] Eilers PH. Parametric time warping. Anal Chem 2004; 76: 404-11.
[74] Robinson MD, De Souza DP, Keen WW, *et al.* A dynamic programming approach for the alignment of signal peaks in multiple gas chromatography-mass spectrometry experiments. BMC Bioinformatics 2007; 8: 419.
[75] Christin C, Hoefsloot HC, Smilde AK, Suits F, Bischoff R, Horvatovich PL. Time alignment algorithms based on selected mass traces for complex LC-MS data. J Proteome Res 2010; 9: 1483-95.
[76] Díaz R, Ibáñez M, Sancho JV, Hernández F. Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2011; 25: 355-69.
[77] Blom KF. Estimating the precision of exact mass measurements on an orthogonal time-of-flight mass spectrometer. Anal Chem 2001; 73: 715-9.
[78] Gobom J, Mueller M, Egelhofer V, Theiss D, Lehrach H, Nordhoff E. A calibration method that simplifies and improves accurate determination of peptide molecular masses by MALDI-TOF MS. Anal Chem 2002; 74: 3915-23.
[79] Turner PG, Taylor S, Goulermas JY, Hampton K. Simulation of high- and low-resolution mass spectra for assessment of calibration methods. Rapid Commun Mass Spectrom 2007; 21: 305-13.
[80] Williams DK Jr., Chadwick MA, Williams TI, Muddiman DC. Calibration laws based on multiple linear regression applied to matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry. J Mass Spectrom 2008; 43: 1659-63.
[81] Strittmatter EF, Rodriguez N, Smith RD. High mass measurement accuracy determination for proteomics using multivariate regression fitting: application to electrospray ionization time-of-flight mass spectrometry. Anal Chem 2003; 75: 460-8.
[82] Mihaleva VV, Vorst O, Maliepaard C, *et al.* Accurate mass error correction in liquid chromatography time-of-flight mass spectrometry based metabolomics. Metabolomics 2008; 4: 171-82.
[83] Scheltema RA, Kamleh A, Wildridge D, *et al.* Increasing the mass accuracy of high-resolution LC-MS data using background ions: a case study on the LTQ-Orbitrap. Proteomics 2008; 8: 4647-56.
[84] Warrack BM, Hnatyshyn S, Ott KH, *et al.* Normalization strategies for metabonomic analysis of urine samples. J Chromatogr B Analyt Technol Biomed Life Sci 2009; 877: 547-52.
[85] Schnackenberg LK, Sun J, Espandiari P, Holland RD, Hanig J, Beger RD. Metabonomics evaluations of age-related changes in the urinary compositions of male Sprague Dawley rats and effects of data normalization methods on statistical and quantitative analysis. BMC Bioinformatics 2007; 8 Suppl 7: S3.
[86] Sysi-Aho M, Katajamaa M, Yetukuri L, Oresic M. Normalization method for metabolomics data using optimal selection of multiple internal standards. BMC Bioinformatics 2007; 8: 93.
[87] Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J. Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 2004; 20: 2447-54.
[88] Wang W, Zhou H, Lin H, *et al.* Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem 2003; 75: 4818-26.
[89] Brown M, Wedge DC, Goodacre R, *et al.* Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. Bioinformatics 2011; 27: 1108-12.
[90] Hilmer JK, Bothner B. Physical signal modulation of time-of-flight mass analyzers increases precision and decreases noise. Rapid Commun Mass Spectrom 2011; 25: 795-805.
[91] Dehaven CD, Evans AM, Dai H, Lawton KA. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. J Cheminform 2010; 2: 9.
[92] Draper J, Enot DP, Parker D, *et al.*. Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'. BMC Bioinformatics 2009; 10: 227.
[93] Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics 2006; 7: 234.
[94] Wolf S, Schmidt S, Müller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 2010; 11: 148.
[95] Zhang H, Zhang D, Ray K, Zhu M. Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry. J Mass Spectrom 2009; 44: 999-1016.
[96] Shinoda K, Sugimoto M, Tomita M, Ishihama Y. Informatics for peptide retention properties in proteomic LC-MS. Proteomics 2008; 8: 787-98.
[97] Sugimoto M, Kikuchi S, Arita M, Soga T, Nishioka T, Tomita M. Large-scale prediction of cationic metabolite identity and migration time in capillary electrophoresis mass spectrometry using artificial neural networks. Anal Chem 2005; 77: 78-84.
[98] Sugimoto M, Hirayama A, Robert M, Abe S, Soga T, Tomita M. Prediction of metabolite identity from accurate mass, migration

time prediction and isotopic pattern information in CE-TOFMS data. Electrophoresis 2010; 31: 2311-8.

[99] Horai H, Arita M, Kanaya S, *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 2010; 45: 703-14.

[100] Lee R, Ptolemy AS, Niewczas L, Britz-McKibbin P. Integrative metabolomics for characterizing unknown low-abundance metabolites by capillary electrophoresis-mass spectrometry with computer simulations. Anal Chem 2007; 79: 403-15.

[101] Chalcraft KR, Lee R, Mills C, Britz-McKibbin P. Virtual quantification of metabolites by capillary electrophoresis-electrospray ionization-mass spectrometry: predicting ionization efficiency without chemical standards. Anal Chem 2009; 81: 2506-15.

[102] Parsons HM, Ekman DR, Collette TW, Viant MR. Spectral relative standard deviation: a practical benchmark in metabolomics. Analyst 2009; 134: 478-85.

[103] Bellew M, Coram M, Fitzgibbon M, *et al.* A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. Bioinformatics 2006; 22: 1902-9.

[104] Lange E, Tautenhahn R, Neumann S, Gröpl C. Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. BMC Bioinformatics 2008; 9: 375.

[105] Matsuda F, Shinbo Y, Oikawa A, *et al.* Assessment of metabolome annotation quality: a method for evaluating the false discovery rate of elemental composition searches. PLoS One 2009; 4: e7490.

[106] Fischer JD, Holliday GL, Rahman SA, Thornton JM. The structures and physicochemical properties of organic cofactors in biocatalysis. J Mol Biol 2010; 403: 803-24.

[107] Habeck C, Stern Y. Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's disease. Cell Biochem Biophys 2010; 58: 53-67.

[108] Nyamundanda G, Brennan L, Gormley IC. Probabilistic principal component analysis for metabolomic data. BMC Bioinformatics 2010; 11: 571.

[109] Do JH, Choi DK. Clustering approaches to identifying gene expression patterns from DNA microarray data. Mol Cells 2008; 25: 279-88.

[110] Čuperlović-Culf M, Belacel N, Culf AS, *et al.* NMR metabolic analysis of samples using fuzzy K-means clustering. Magn Reson Chem 2009; 47 Suppl 1: S96-104.

[111] Dougherty ER, Barrera J, Brun M, *et al.* Inference from clustering with application to gene-expression microarrays. J Comput Biol 2002; 9: 105-26.

[112] Pears MR, Cooper JD, Mitchison HM, Mortishire-Smith RJ, Pearce DA, Griffin JL. High resolution 1H NMR-based metabolomics indicates a neurotransmitter cycling deficit in cerebral tissue from a mouse model of Batten disease. J Biol Chem 2005; 280: 42508-14.

[113] Faria AV, Macedo FC Jr., Marsaioli AJ, Ferreira MM, Cendes F. Classification of brain tumor extracts by high resolution ^{1}H MRS using partial least squares discriminant analysis. Braz J Med Biol Res 2011; 44: 149-64.

[114] Wiklund S, Johansson E, Sjöström L, *et al.* Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Anal Chem 2008; 80: 115-22.

[115] Westerhuis JA, Hoefsloot HCJ, Smit S, *et al.* Assessment of PLSDA cross validation. Metabolomics 2008; 4: 81-9.

[116] Carrola J, Rocha CM, Barros AS, *et al.* Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. J Proteome Res 2011; 10: 221-30.

[117] Breiman L. Random forests. Machine learning 2001; 45: 5-32.

[118] Kuhn S, Egert B, Neumann S, Steinbeck C. Building blocks for automated elucidation of metabolites: machine learning methods for NMR prediction. BMC Bioinformatics 2008; 9: 400.

[119] Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006; 7: 3.

[120] Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006; 2: 171-96.

[121] Sylvester RJ. Combining a molecular profile with a clinical and pathological profile: biostatistical considerations. Scand J Urol Nephrol Suppl 2008; 42: 185-90.

[122] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 1995; 57: 289-300.

[123] Seifert EL, Fiehn O, Bezaire V, *et al.* Long-chain fatty acid combustion rate is associated with unique metabolite profiles in skeletal muscle mitochondria. PLoS One 2010; 5: e9834.

[124] Fawcett T. An introduction to ROC analysis. Pattern recognition letters 2006; 27: 861-74.

[125] Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. Bioinformatics 2009; 25: 3151-7.

[126] Henneges C, Bullinger D, Fux R, *et al.* Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection. BMC Cancer 2009; 9: 104.

[127] Izquierdo-García JL, Rodríguez I, Kyriazis A, *et al.* A novel R-package graphic user interface for the analysis of metabonomic profiles. BMC Bioinformatics 2009; 10: 363.

[128] Hummel J, Strehmel N, Selbig J, Walther D, Kopka J. Decision tree supported substructure prediction of metabolites from GC-MS profiles. Metabolomics 2010; 6: 322-33.

[129] Brougham DF, Ivanova G, Gottschalk M, Collins DM, Eustace AJ, O'Connor R, Havel J. Artificial neural networks for classification in metabolomic studies of whole cells using 1H nuclear magnetic resonance. J Biomed Biotechnol 2011; 2011: 158094.

[130] Jansen MA, Kiwata J, Arceo J, Faull KF, Hanrahan G, Porter E. Evolving neural network optimization of cholesteryl ester separation by reversed-phase HPLC. Anal Bioanal Chem 2010; 397: 2367-74.

[131] Hanrahan G. Computational neural networks driving complex analytical problem solving. Anal Chem 2010; 82: 4307-13.

[132] Yetukuri L, Tikka J, Hollmén n J, Orešič M. Functional prediction of unidentified lipids using supervised classifiers. Metabolomics 2010; 6: 18-26.

[133] Cakir T, Hendriks MM, Westerhuis JA, Smilde AK. Metabolic network discovery through reverse engineering of metabolome data. Metabolomics 2009; 5: 318-29.

[134] Jourdan F, Cottret L, Huc L, *et al.* Use of reconstituted metabolic networks to assist in metabolomic data visualization and mining. Metabolomics 2010; 6: 312-21.

[135] Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc 2011; 6: 743-60.

[136] Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 2009; 37: W652-60.

[137] Kono N, Arakawa K, Ogawa R, *et al.* Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API. PLoS One 2009; 4: e7710.

[138] Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 2010; 38: D355-60.

[139] Caspi R, Foerster H, Fulcher CA, *et al.* The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 2008; 36: D623-31.

[140] Latendresse M, Karp PD. Web-based metabolic network visualization with a zooming user interface. BMC Bioinformatics 2011; 12: 176.

[141] Byrnes RW, Cotter D, Maer A, Li J, Nadeau D, Subramaniam S. An editor for pathway drawing and data visualization in the Biopathways Workbench. BMC Syst Biol 2009; 3: 99.

[142] Fiehn O, Kristal B, van Ommen B, *et al.* Establishing reporting standards for metabolomic and metabonomic studies: a call for participation. OMICS 2006; 10: 158-63.

[143] Sumner LW, Amberg A, Barrett D, *et al.* Proposed minimum reporting standards for chemical analysis. Metabolomics 2007; 3: 211-21.

[144] Takeda I, Stretch C, Barnaby P, *et al.* Understanding the human salivary metabolome. NMR Biomed 2009; 22: 577-84.

[145] Nishiumi S, Shinohara M, Ikeda A, *et al.* Serum metabolomics as a novel diagnostic approach for pancreatic cancer. Metabolomics 2010; 6: 518-28.

[146] Sugimoto M, Goto H, Otomo K, *et al.* Metabolomic profiles and sensory attributes of edamame under various storage duration and temperature conditions. J Agric Food Chem 2010; 58: 8418-25.

[147] Sugimoto M, Koseki T, Hirayama A, *et al.* Correlation between

sensory evaluation scores of Japanese sake and metabolome profiles. J Agric Food Chem 2010; 58: 374-83.

[148] Iino K, Sugimoto M, Soga T, Tomita M. Profiling of the charged metabolites of traditional herbal medicines using capillary electrophoresis time-of-flight mass spectrometry. Metabolomics; DOI:10.1007/s11306-011-0290-7.

[149] Kotłowska A, Sworczak K, Stepnowski P. Urine metabolomics analysis for adrenal incidentaloma activity detection and biomarker discovery. J Chromatogr B Analyt Technol Biomed Life Sci 2011; 879: 359-63.

[150] Sorlie T, Perou CM, Tibshirani R, *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001; 98: 10869-74.

[151] Enot DP, Beckmann M, Overy D, Draper J. Predicting interpretability of metabolome models based on behavior, putative identity, and biological relevance of explanatory signals. Proc Natl Acad Sci U S A 2006; 103: 14865-70.

[152] Beckmann M, Enot DP, Overy DP, Draper J. Representation, comparison, and interpretation of metabolome fingerprint data for total composition analysis and quality trait investigation in potato cultivars. J Agric Food Chem 2007; 55: 3444-51.

[153] Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 2002; 99: 6562-6.

[154] Soga T, Sugimoto M, Honma M, *et al.* Serum metabolomics reveals gamma-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. J Hepatol 2011; 55: 896-905

[155] Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. Metabolomics 2010; 6: 78-95.

[156] Kuhn T, Willighagen EL, Zielesny A, Steinbeck C. CDK-Taverna: an open workflow environment for cheminformatics. BMC Bioinformatics 2010; 11: 159.

[157] Tautenhahn R, Patti GJ, Kalisiak E, *et al.* metaXCMS: second-order analysis of untargeted metabolomics data. Anal Chem 2011; 83: 696-700.

[158] Benton HP, Wong DM, Trauger SA, Siuzdak G. XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. Anal Chem 2008; 80: 6382-9.

[159] Grigsby CC, Rizki MM, Tamburino LA, Pitsch RL, Shiyanov PA, Cool DR. Metabolite differentiation and discovery lab (MeDDL): a new tool for biomarker discovery and mass spectral visualization. Anal Chem 2010; 82: 4386-95.

[160] Gao J, Tarcea VG, Karnovsky A, *et al.* Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. Bioinformatics 2010; 26: 971-3.

[161] Hiller K, Hangebrauk J, Jager C, Spura J, Schreiber K, Schomburg D. MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. Anal Chem 2009; 81: 3429-39.

[162] Lommen A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem 2009; 81: 3079-86.

[163] Haimi P, Uphoff A, Hermansson M, Somerharju P. Software tools for analysis of mass spectrometric lipidome data. Anal Chem 2006; 78: 8324-31.

[164] Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol 2011; 696: 291-303.

[165] Kastenmüller G, Römisch-Margl W, Wägele B, Altmaier E, Suhre K. metaP-server: a web-based metabolomics data analysis tool. J Biomed Biotechnol 2011; 2011.

[166] Biswas A, Mynampati KC, Umashankar S, *et al.* MetDAT: a modular and workflow-based free online pipeline for mass spectrometry data processing, analysis and interpretation. Bioinformatics 2010; 26: 2639-40.

[167] Hoffmann N, Stoye J. ChromA: signal-based retention time alignment for chromatography-mass spectrometry data. Bioinformatics 2009; 25: 2080-1.

[168] Xia J, Wishart DS. MetPA: a web-based metabolomics tool for pathway analysis and visualization. Bioinformatics 2010; 26: 2342-4.

[169] Cottret L, Wildridge D, Vinson F, *et al.* MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. Nucleic Acids Res 2010; 38: W132-7.

[170] Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res 2010; 38: W71-7.

[171] Carroll AJ, Badger MR, Harvey MA. The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. BMC Bioinformatics 2010; 11: 376.

[172] Wang SY, Ho TJ, Kuo CH, Tseng YJ. Chromaligner: a web server for chromatogram alignment. Bioinformatics 2010; 26: 2338-9.

Received: July 17, 2011 Revised: October 25, 2011 Accepted: December 07, 2011

Published online 1 December 2011 *Nucleic Acids Research, 2012, Vol. 40, Database issue* ***D809–D814***
doi:10.1093/nar/gkr1170

MMMDB: Mouse Multiple Tissue Metabolome Database

Masahiro Sugimoto[1,2,3,*], Satsuki Ikeda[1], Kanako Niigata[1], Masaru Tomita[1,2], Hideyo Sato[4] and Tomoyoshi Soga[1,2]

[1]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017, [2]Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252-8520, [3]Graduate School of Medicine and Faculty of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan and [4]Department of Food and Applied Life Science, Faculty of Agriculture, Yamagata University, Tsuruoka, Yamagata 997-8555, Japan

Received August 23, 2011; Revised October 21, 2011; Accepted November 11, 2011

ABSTRACT

The Mouse Multiple Tissue Metabolome Database (MMMDB) provides comprehensive and quantitative metabolomic information for multiple tissues from single mice. Manually curated databases that integrate literature-based individual metabolite information have been available so far. However, data sets on the absolute concentration of a single metabolite integrated from multiple resources are often difficult to be used when different metabolomic studies are compared because the relative balance of the multiple metabolite concentrations in the metabolic pathways as a snapshot of a dynamic system is more important than the absolute concentration of a single metabolite. We developed MMMDB by performing non-targeted analyses of cerebra, cerebella, thymus, spleen, lung, liver, kidney, heart, pancreas, testis and plasma using capillary electrophoresis time-of-flight mass spectrometry and detected 428 non-redundant features from which 219 metabolites were successfully identified. Quantified concentrations of the individual metabolites and the corresponding processed raw data; for example, the electropherograms and mass spectra with their annotations, such as isotope and fragment information, are stored in the database. MMMDB is designed to normalize users' data, which can be submitted online and used to visualize overlaid electropherograms. Thus, MMMDB allows newly measured data to be compared with the other data in the database. MMMDB is available at: http://mmmdb.iab.keio.ac.jp.

INTRODUCTION

Metabolomics, the newest 'omics', is defined as the comprehensive identification and quantification of small molecules that provides a holistic view of cellular metabolism. The metabolomic network, downstream of the central dogma, transfers regulatory information from other omics data, such as genomics, transcriptomics and proteomics; thus, the metabolomic profile can be expected to directly reflect cellular phenotype (1). Metabolomic profiling has been used in biological studies in various fields, e.g. microorganism, plant, food, agricultural, pharmaceutical, clinical and medical sciences.

Nuclear magnetic resonance (NMR) (2) and mass spectrometry (MS) (3,4) are the major analytical techniques that are used in metabolomics. The relatively low sensitivity of NMR and spectral overlap limits the number and variety of metabolites that can be observed simultaneously. MS combined with a separation system prior to MS, e.g. gas chromatography (GC)–MS (5), liquid chromatography (LC)–MS (6) and capillary electrophoresis (CE)–MS (7), is currently the leading analytical platform because it provided higher selectivity and sensitivity. Because of the diverse physical and chemical properties of the metabolites, no single analytical method can comprehensively profile data sets, and each method has its own advantages and disadvantages (8). GC–MS is a well-established technology that is capable of profiling only volatile compounds and generally requires an initial derivatization procedure. LC–MS can be used to monitor a wider variety of non-volatile compounds; however, optimization of sample processing and LC column selection that depends on the target analytes is necessary. In contrast, CE–MS can monitor all charged metabolites within two (positive/negative) modes which allows for the simultaneous profiling of many pathways.

*To whom correspondence should be addressed. Tel: +81 235 29 0528; Fax: +81 235 29 0574; Email: msugi@sfc.keio.ac.jp

Metabolomics has contributed to the accumulation of knowledge about metabolites and their chemical (enzyme) reactions and this information has been stored in various databases. These databases can be classified broadly into two types: (i) databases that contain metabolic pathways based on information from the literature which has been integrated and curated manually and (ii) databases that contain raw or processed data from analytical system like, for example, mass spectrometry, that allow comparisons between stored data and users' experimental data to be made. Examples of the first type of database include the Kyoto Encyclopedia of Genes and Genomics (KEGG) that contains a large collection of metabolites, enzyme and chemical drug from various species (9), MetaCyc (MetaCyc.org) that includes experimentally verified metabolic pathways and enzyme information (10), MetaCrop (11) containing metabolites from six kinds of crops and Reactome which contains human metabolites (12). These databases visualize metabolic pathways using data from the literature. The Human Metabolome Database (HMDB) contains the endogenous metabolites in human biofluids and their quantified concentrations collected from the literature (13) and the Small Molecular Pathway Database (SMPD), which is fully linked to HMDB, provides manually curated pathways (>350) for the human metabolites (14). There are also several commercial pathway databases that contain integrated knowledge and well-studied pathways, e.g. Cell Signal pathways (www.cellsignal.com), Sigma–Aldrich pathways (http://www.sigmaaldrich.com/life-science/cell-biology/learning-center.html), and Ambion pathways (http://www.ambion.com), and ProteinLounge (http://www.proteinlounge.com/).

Examples of the second type of database include the Golm Metablome Database (GMD@CSB.DB) and FiehnLib that contain GC/MS spectra (15), METLIN that provides the mass spectra of metabolites and drugs (16), MassBank that contains mass spectra from various types of MS (17) and HMDB that contains NMR data and the mass spectra of LC–MS and GC/MS along with data comparison and search functions.

Metabolic systems intricately vary in their response to environmental change; these variations are controlled by enzyme regulation. Thus, the simultaneously observed concentrations of metabolites in many pathways are indicators that provide insights to a more holistic understanding of their biological significance. This is in contrast to the integrated profiles available in the literature. Therefore, we established MMMDB, a database that contains a large collection of metabolites in multiple tissues from single mice that were obtained using capillary electrophoresis time-of-flight MS (CE-TOFMS) (18,19).

DATABASE DESCRIPTION

Database content

CE-TOFMS was used to analyze 10 tissues, cerebra, cerebella, thymus, spleen, lung, liver, kidney, heart, pancreas and testis, and also plasma in a non-targeted manner (<1000 Da) so that all possible metabolite peaks were profiled. Wild-type mice have been backcrossed to C57BL/6J for 10 or more generations, and were subsequently bred for 2 years in specific pathogen-free animal housing facilities at Yamagata University Medical School (20). Plastic cage with sawdust on the floor were used with keeping the day lights on 12-h light/dark cycle, maintaining the temperature between 23°C and 24°C. Diet (Oriental MF, Oriental Yeast Co., Tokyo, Japan) and water were available to the mice at all times. Two of 8-week-old male mice were sacrificed between 10:00 am and 12:00 pm. Duplicated data from two mice were included in the database.

Raw data were analyzed using MasterHands software (21), which detects all possible peaks, eliminates noise (e.g. spike noise) and interprets redundant features (e.g. isotopic and adduct noise) commonly observed in ESI–MS data (22). Migration times of CE–MS electropherograms were normalized by dynamic programming-based correlation methods to generate an aligned data matrix (23). We detected 428 kinds of peaks without redundant features [on an average, 351 ± 54 peaks (SD) on each sample] and identified 219 metabolites (192 ± 20).

For details of the measurement methods, the conditions used for CE-TOFMS, and the data processing procedure see Methods in Supplementary Data. Quantified concentration for each peak were identified with matched standard compounds and recorded in the database. To eliminate systematic bias, the peak areas of the 209 unidentified peaks were normalized using the same internal standards and this data was also included in the database. An overview of the profiled data and the results of the multivariate analyses are depicted in Supplementary Figures S1 and S2 using visualization software (24).

Reproducibility of migration times in CE–MS data, i.e. the peak location along the electropherogram axis, is lower than the retention/elution times in LC–MS and GC/MS data. This makes data comparison between different runs difficult. To help address this problem, the migration times of internal standards simultaneously measured in each data were registered for each metabolite entry. Both raw and background-subtracted mass spectrum with their interpreted annotations, such as isotope, fragment, adduct ions, were also registered. Identified metabolites were linked to KEGG (9), HMDB (13), ChEBI (25) and PubChem (http://pubchem.ncbi.nlm.nih.gov/).

Web interface

A screenshot of MMMDB is shown in Figure 1. All quantified data sets can be downloaded as separate csv files for each tissue or plasma. In addition to the search and browse metabolites functions, users can upload their own data and compare it on the website with the data that is stored in the database. The distribution of the quantified data in each tissue can be visualized as a bar graph and the chemical structure is also visualized. The web interface can be customized interactively on the website.

Similar to the search in HMDB and MassBank, users can use m/z or molecular weight as a query to search

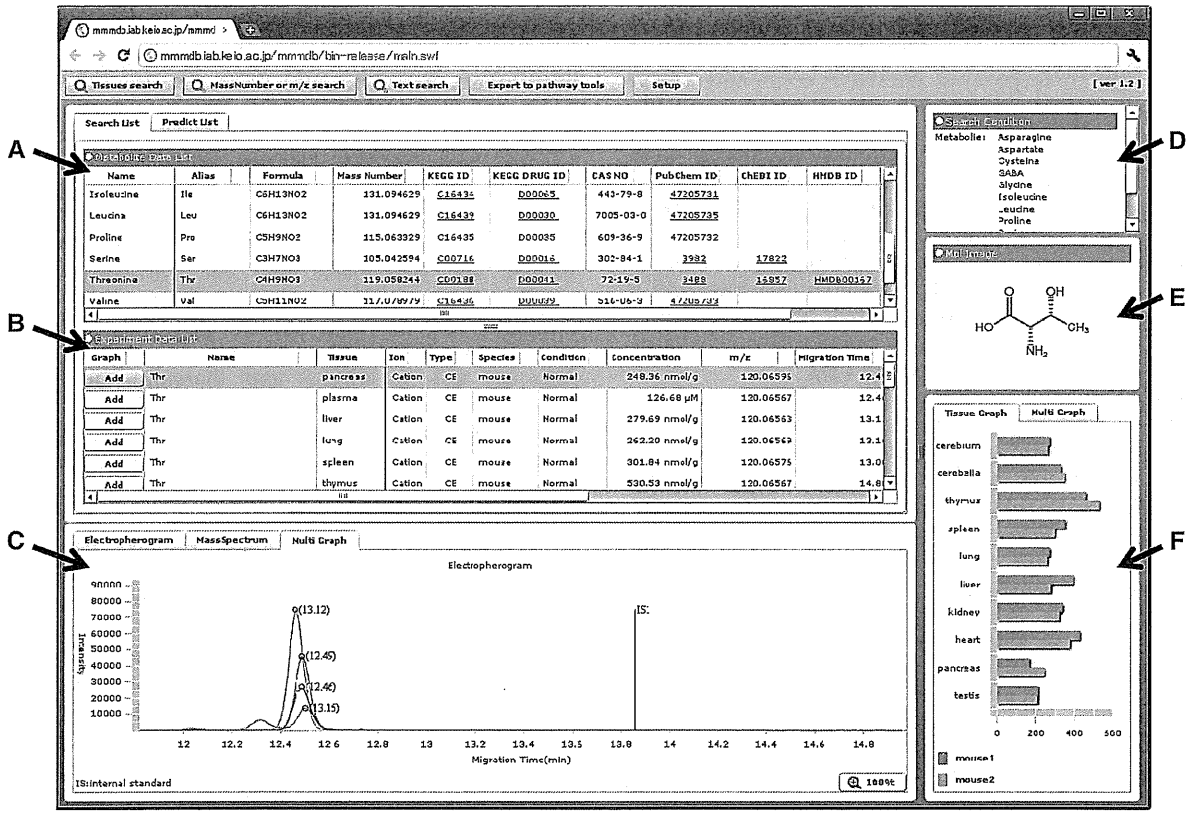

Figure 1. An MMMDB screenshot. (**A**) List of metabolites that fulfilled the search condition. (**B**) List of the registered entries of the metabolites selected from panel (A). (**C**) Overlaid electropherograms displayed by the Multi Graph option. The other options, Electropherogram and Mass Spectrum, will display a single electropherogram or the mass spectrum, respectively. (**D**) The search conditions used. (**E**) The molecular structure of the displayed metabolite. (**F**) Tissue graph showing the quantified concentrations of the selected metabolites in each tissue.

the metabolites. The search results list the possible metabolites along with redundant features, such as isotope and fragment, which may help identify the interesting peaks.

For users to be able to compare their CE–MS data with the stored data, user can upload a data file containing the electropherogram, mass spectrum and migration times of internal standards in a structured csv format. In MMMDB, the migration times of the electropherogram are corrected using a polynomial equation derived from the migration times of multiple internal standards (26), and the overlaid electropherograms are visualized. Based on the difference between m/z and migration times, the searched results are evaluated and assigned scores which indicate the matched possibility. For non-CE–MS users, the electropherogram can be omitted from the upload data, and comparison are then made based only on mass spectrometry data. These functions help identify the peaks in the users' data sets. To use the data visualization tools in the pathway mode, a part of the data sets, for example, all the metabolites in a tissue, can be exported and used as the input data file for the Pathway Projector (27) and for Vanted (28).

In addition to the data sets, web interfaces provide a data browsing navigator, a tutorial movie and document, and an example of users' data that can be compared on-line. A tool for formatting users' data is also provided.

Database implementation

The sever programs are designed as a versatile independent to operating system (OS). Client Adobe Flex 3 was used for the framework server program which was implemented on a Java platform [standard edition 6.0 (1.6.0_26)]. Although users are required to install the Adobe Flash Player, the interactive operations without reloading that are used, for example, to visualize overlaid electropherograms, are enabled and the user interface does not depend on the browser programs. We used Tomcat 6.0 for the web server and PostgreSQL 9.0.4 for the database. The server programs were implemented on CentOS 5.6 with 4 GB memory.

DISCUSSION

The principal feature of MMMDB is the comprehensive large collection of the absolute concentrations of metabolites in various pathways in multiple tissues from single mice using a single measurement platform. In contrast to

changes in a single metabolite, profiles of multiple metabolites in many pathways are important for metabolomics research. Metabolic pathway databases containing a large amount of related information are currently available; KEGG (9), HumanCyc for human specific metabolic pathways (29), BioCarta Collections visualizing well-curated pathways (http://www.biocarta.com/genes/allPathways.asp) and Ingenuity Pathways Analysis (IPA) (http://www.ingenuity.com) providing both pathway and structured literature information are examples of these. MouseCyc is a mouse specific metabolic database and a subset of MetaCyc, provides various build-in pathway visualization tools and a pathway characteristics comparison function for mouse and human (30).

Excellent tools to visualize the quantified concentrations of multiple metabolites in metabolic pathways are also available. Pathway Projector (27) has distributed various graphs of KEGG pathways using Google Maps technologies and KEGG Atlas (31). This is useful to explore the relation of multiple metabolites from large metabolic pathways using various search functions. Vanted (28) provides an editable pathway using a template that can be downloaded from several databases, such as KEGG (9) and MetaCrop (11), with several statistical analyses tools, such as, for example, a self-organizing map (SOM) that can be used to depict the metabolite relations by reducing the metabolic complexity. With these visualization tools, the profiled data sets in MMMDB provide both an overview and the characteristic relations in complicated metabolic systems.

Multivariable analyses of the different tissue profiles allowed their tissue-specific bias to be visualized. Principal component (PC) analysis (Supplementary Figure S1) showed, as expected, that along with the first PC axis, the profiles in liver and kidney were exceptionally different from those in other tissues. Plots along the third PC axis also showed that the differences between kidney and liver profiles were larger than the differences between any of the other tissue profiles. Plots of cerebra and cerebella profiles exist close together in all PC spaces, indicating that these two profiles are similar. Interestingly, compared to the other tissues, the testis profiles were closest to the cerebra and cerebella profiles. Clustering (Supplementary Figure S2) also produced consistent results revealing, for example, the similarities in the brain profiles. Metabolites categorized as amino acids (betaine and taurine), organic acids (malate, lactate, succinate and citrate), those synthesized from arginine (creatine and carnitine), and involved in nucleotide pathways (uridine), tryptophan metabolism (indole-3-acetaldehyde), glutathione metabolism (5-oxoproline) and other pathways (glycolate, 2-hydroxypentanoate, allantoin and acetohydroxamate) were consistently higher than other metabolites in both the averaged tissue profiles and in the plasma profile. The profiles in the other metabolites in plasma are more constant than those in the tissues.

MMMDB has several limitations. The observed profiles include the metabolites in glycolysis, the tricarboxylic acid cycle, the pentose phosphate pathway, the nucleotide pathway, the amino acid pathway and the urea cycle. Because CE–MS can detect only soluble and charged metabolites, molecules with other feature, for example, lipids, volatile metabolites and molecules in secondary metabolism, are not covered in the current data sets. Using LC–MS and GC/MS that are complementary to the CE–MS data would raise the coverage; however, when non-CE–MS data is included, several metabolites that are also monitored by CE-MS make it necessary to include the instrument and the correlation of the overlapped metabolite concentrations need to be evaluated for the data sets to be extended. Another limitation of MMMDB is that several metabolites in the data are not in KEGG and therefore the Pathway Projector and Vanted cannot visualize this data. The development of a data converter for other visualization tools or for the original pathway visualization function is also necessary.

As for many other databases, the development and expansion of MMMDB is on-going. Because, the profiling technique itself does not limit its possible applications (32,33), new data sets of more segmented brain tissue and of various biofluids such as urine and saliva, are being developed. In addition to data from healthy mice (the control), data from mice with various diseases (34) could also be integrated. In contrast to profile data, metabolomic data using stable isotope labeling is also useful for tracing the metabolite flux distribution (35). Ideally, the identification of all unknown peaks would make more comprehensive improve analyses possible. We also aim to replace the currently unknown data with computational estimations of the unknown peaks (36) and their quantification (37).

CONCLUSION

In summary, MMMDB contains the concentrations of a large number of metabolites simultaneously profiled using of CE-TOFMS in a non-targeted manner. The collected profiles are from 10 tissues and the plasma of single mice. Quantified values of each metabolite and the annotated mass spectra and electropherograms are also provided. A functional web interface provides various data search options and an on-line data comparison function between data uploaded by users and stored data from mass spectrum and normalized electropherograms. The database system itself was developed as a versatile system that can contain data sets from any other species or from any tissue. Thus, the next task in the development of MMMDB is to add data from other organisms and tissues, obtained under different conditions using a variety of measurement instruments.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1 and 2 and Supplementary Methods.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Shinobu Abe and Nami Abe of the Institute for Advanced Biosciences, Keio University, Japan, for their help with measurement and data analysis.

FUNDING

Funding for open access charge: Yamagata Prefectural Government and Tsuruoka City, Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Fiehn,O. (2002) Metabolomics–the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155–171.
2. Reo,N.V. (2002) NMR-based metabolomics. *Drug Chem. Toxicol.*, **25**, 375–382.
3. Aharoni,A., Ric de Vos,C.H., Verhoeven,H.A., Maliepaard,C.A., Kruppa,G., Bino,R. and Goodenowe,D.B. (2002) Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *Omics*, **6**, 217–234.
4. Castrillo,J.I., Hayes,A., Mohammed,S., Gaskell,S.J. and Oliver,S.G. (2003) An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry*, **62**, 929–937.
5. Fiehn,O., Kopka,J., Trethewey,R.N. and Willmitzer,L. (2000) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.*, **72**, 3573–3580.
6. Plumb,R., Granger,J., Stumpf,C., Wilson,I.D., Evans,J.A. and Lenz,E.M. (2003) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst*, **128**, 819–823.
7. Soga,T., Ohashi,Y., Ueno,Y., Naraoka,H., Tomita,M. and Nishioka,T. (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.*, **2**, 488–494.
8. Monton,M.R. and Soga,T. (2007) Metabolome analysis by capillary electrophoresis-mass spectrometry. *J. Chromatogr. A*, **1168**, 237–246; discussion 236.
9. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
10. Caspi,R., Altman,T., Dale,J.M., Dreher,K., Fulcher,C.A., Gilham,F., Kaipa,P., Karthikeyan,A.S., Kothari,A., Krummenacker,M. *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.
11. Grafahrend-Belau,E., Weise,S., Koschutzki,D., Scholz,U., Junker,B.H. and Schreiber,F. (2008) MetaCrop: a detailed database of crop plant metabolism. *Nucleic Acids Res.*, **36**, D954–D958.
12. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
13. Wishart,D.S., Knox,C., Guo,A.C., Eisner,R., Young,N., Gautam,B., Hau,D.D., Psychogios,N., Dong,E., Bouatra,S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**, D603–D610.
14. Frolkis,A., Knox,C., Lim,E., Jewison,T., Law,V., Hau,D.D., Liu,P., Gautam,B., Ly,S., Guo,A.C. *et al.* (2010) SMPDB: the Small Molecule Pathway Database. *Nucleic Acids Res.*, **38**, D480–D487.
15. Kopka,J., Schauer,N., Krueger,S., Birkemeyer,C., Usadel,B., Bergmuller,E., Dörmann,P., Weckwerth,W., Gibon,Y., Stitt,M. *et al.* (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, **21**, 1635–1638.
16. Smith,C.A., O'Maille,G., Want,E.J., Qin,C., Trauger,S.A., Brandon,T.R., Custodio,D.E., Abagyan,R. and Siuzdak,G. (2005) METLIN: a metabolite mass spectral database. *Ther. Drug Monit.*, **27**, 747–751.
17. Horai,H., Arita,M., Kanaya,S., Nihei,Y., Ikeda,T., Suwa,K., Ojima,Y., Tanaka,K., Tanaka,S., Aoshima,K. *et al.* (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom*, **45**, 703–714.
18. Soga,T., Baran,R., Suematsu,M., Ueno,Y., Ikeda,S., Sakurakawa,T., Kakazu,Y., Ishikawa,T., Robert,M., Nishioka,T. *et al.* (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J. Biol. Chem.*, **281**, 16768–16776.
19. Soga,T., Igarashi,K., Ito,C., Mizobuchi,K., Zimmermann,H.P. and Tomita,M. (2009) Metabolomic Profiling of Anionic Metabolites by Capillary Electrophoresis Mass Spectrometry. *Anal. Chem.*, **81**, 6165–6174.
20. Sato,H., Shiiya,A., Kimata,M., Maebara,K., Tamba,M., Sakakura,Y., Makino,N., Sugiyama,F., Yagami,K., Moriguchi,T. *et al.* (2005) Redox imbalance in cystine/glutamate transporter-deficient mice. *J. Biol. Chem.*, **280**, 37423–37429.
21. Sugimoto,M., Wong,D.T., Hirayama,A., Soga,T. and Tomita,M. (2010) Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics*, **6**, 78–95.
22. Brown,M., Dunn,W.B., Dobson,P., Patel,Y., Winder,C.L., Francis-McIntyre,S., Begley,P., Carroll,K., Broadhurst,D., Tseng,A. *et al.* (2009) Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst*, **134**, 1322–1332.
23. Baran,R., Kochi,H., Saito,N., Suematsu,M., Soga,T., Nishioka,T., Robert,M. and Tomita,M. (2006) MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics*, **7**, 530.
24. Saeed,A.I., Sharov,V., White,J., Li,J., Liang,W., Bhagabati,N., Braisted,J., Klapa,M., Currier,T., Thiagarajan,M. *et al.* (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378.
25. Degtyarenko,K., Hastings,J., de Matos,P. and Ennis,M. (2009) ChEBI: an open bioinformatics and cheminformatics resource. *Curr. Protoc. Bioinformatics*, **Chapter 14**, Unit 14.19.
26. Reijeng,J.C., Martens,J.H., Giuliani,A. and Chiari,M. (2002) Pherogram normalization in capillary electrophoresis and micellar electrokinetic chromatography analyses in cases of sample matrix-induced migration time shifts. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **770**, 45–51.
27. Kono,N., Arakawa,K., Ogawa,R., Kido,N., Oshita,K., Ikegami,K., Tamaki,S. and Tomita,M. (2009) Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API. *PLoS One*, **4**, e7710.
28. Junker,B.H., Klukas,C. and Schreiber,F. (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, **7**, 109.
29. Romero,P., Wagg,J., Green,M.L., Kaiser,D., Krummenacker,M. and Karp,P.D. (2004) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.*, **6**, R2.
30. Evsikov,A.V., Dolan,M.E., Genrich,M.P., Patek,E. and Bult,C.J. (2009) MouseCyc: a curated biochemical pathways database for the laboratory mouse. *Genome Biol.*, **10**, R84.
31. Okuda,S., Yamada,T., Hamajima,M., Itoh,M., Katayama,T., Bork,P., Goto,S. and Kanehisa,M. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.*, **36**, W423–W426.
32. Baena,B., Cifuentes,A. and Barbas,C. (2005) Analysis of carboxylic acids in biological fluids by capillary electrophoresis. *Electrophoresis*, **26**, 2622–2636.
33. Ramautar,R., Mayboroda,O.A., Somsen,G.W. and de Jong,G.J. (2011) CE-MS for metabolomics: developments and applications in the period 2008–2010. *Electrophoresis*, **32**, 52–65.

34. Garcia-Perez,I., Whitfield,P., Bartlett,A., Angulo,S., Legido-Quigley,C., Hanna-Brown,M. and Barbas,C. (2008) Metabolic fingerprinting of Schistosoma mansoni infection in mice urine with capillary electrophoresis. *Electrophoresis*, **29**, 3201–3206.
35. Fan,T.W., Lane,A.N., Higashi,R.M. and Yan,J. (2011) Stable isotope resolved metabolomics of lung cancer in a SCID mouse model. *Metabolomics*, **7**, 257–269.
36. Sugimoto,M., Hirayama,A., Robert,M., Abe,S., Soga,T. and Tomita,M. (2010) Prediction of metabolite identity from accurate mass, migration time prediction and isotopic pattern information in CE-TOFMS data. *Electrophoresis*, **31**, 2311–2318.
37. Chalcraft,K.R., Lee,R., Mills,C. and Britz-McKibbin,P. (2009) Virtual quantification of metabolites by capillary electrophoresis-electrospray ionization-mass spectrometry: predicting ionization efficiency without chemical standards. *Anal. Chem.*, **81**, 2506–2515.

Metabolomics (2012) 8:99–108
DOI 10.1007/s11306-011-0290-7

ORIGINAL ARTICLE

Profiling of the charged metabolites of traditional herbal medicines using capillary electrophoresis time-of-flight mass spectrometry

Keiko Iino · Masahiro Sugimoto · Tomoyoshi Soga · Masaru Tomita

Received: 25 November 2010 / Accepted: 14 February 2011 / Published online: 27 February 2011

Abstract The quantification of a small number of bioactive components in herbal medicines is often inadequate when attempting to elucidate a medicine's biological effects. Despite rapid advances in analytical technologies, obtaining comprehensive metabolomic profiles of herbal medicines remains difficult, due to the complexity of natural product mixtures. *Toki-Shakuyaku-San* is a Chinese medicine used widely to treat gynecological and obstetric disorders, such as infertility, dysmenorrhea, toxemia during pregnancy and neural dysfunction. It consists of *Angelica acutiloba Radix* (Toki), *Cnidium officinale Rhizoma* (Senkyu), *Paeonia lactiflora Radix* (Shakuyaku), *Atractylodes lancea Rhizoma* (Sojutsu), *Alisma orientale Rhizoma* (Takusha) and *Poria cocos Hoelen* (Bukuryo). To elucidate the composition of these herbal medicines individually, we conducted non-targeted profiling analyses of extracts of these herbs using capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS), which allows the simultaneous quantification of hundreds of charged metabolites. In total, 737 ± 183.1 (average ± SD) metabolite-derived features were observed, and of these, 119 metabolites were identified. Score plots of principal component analysis (PCA) showed a clear cluster including Shakuyaku, Bukuryo, and Sojutsu, while the other three herbs were distributed over PCA spaces. Loading plots revealed that amino acids and shikimate-derived alkaloids were the predominant metabolite constituents. Hierarchical clustering analysis revealed that few clusters overlapped in the herbal medicines tested. This report is the first demonstration of the characterization of a herbal medicine using large-scale metabolomic analysis, which is complementary to traditional quality control methods.

Keywords Capillary electrophoresis time-of-flight mass spectrometry · Herbal medicine · Charged metabolite · Metabolomic profiling

1 Introduction

Herbal medicines have been used as therapeutic agents for several thousands of years in Asian countries (Wang and Ren 2002; Wang et al. 2009). The evaluation of a herbal medicine's safety and efficacy is required before its clinical use as a complementary and alternative medicine (CAM) in western countries (Miller et al. 2004; Murray and Rubel 1992; Calixto 2000; Zhang et al. 2010). The conventional pharmacological approach to characterize herbal medicines focuses on the identification and quantification of a single or several bioactive components (Murray and Rubel 1992; Liu et al. 2008), however, the complex properties of herbal medicines often render such an approach inadequate (Wang and Ren 2002; Lao et al. 2009; Liang et al. 2009; Chan 2003). The quantitative analysis of the complete metabolite profile of herbal medicines would therefore be of significant value (Liu et al. 2008).

Electronic supplementary material The online version of this article (doi:10.1007/s11306-011-0290-7) contains supplementary material, which is available to authorized users.

K. Iino · M. Sugimoto (✉) · T. Soga · M. Tomita
Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0052, Japan
e-mail: msugi@sfc.keio.ac.jp

K. Iino · T. Soga · M. Tomita
Department of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252-8520, Japan

M. Sugimoto · T. Soga · M. Tomita
Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252-8520, Japan

Recent advances in the development of mass spectrometry (MS)-based profiling techniques have made a significant contribution to the study and quantification of the metabolome, the complete range of low molecular weight compounds of a natural sample, which is a field known as metabolomics. MS is typically used in combination with a separation system, such as liquid chromatography (LC), gas chromatography (GC) or capillary electrophoresis (CE). The selection of the separation system to be used is based on the chemical properties of the target compounds, since there is no single analytical methodology currently available capable of profiling complete metabolome. GC–MS and LC–MS are used routinely for targeted analysis (Liang et al. 2009) and are considered to be mature techniques (Weng and Jin 2002). GC–MS is limited to profiling volatile molecules, such as essential oils (Li et al. 2003; Gong et al. 2001), while LC–MS is suitable for separating and detecting more diverse molecules, such as flavonoids, glycosides, organic acids, saponins and lipids (Lao et al. 2009; Ma et al. 2007; Ren et al. 2008). However, the separation conditions used for LC should be optimized for individual molecular classes, which limits the number of metabolites simultaneously detected. CE coupled MS, in particular CE time-of-flight MS (CE-TOFMS), has been demonstrated to have high separation capability and sensitivity for the profiling of charged metabolites, including primary metabolites such as amino acids, amines, organic acids and nucleic acids (Soga et al. 2003, 2006; Monton and Soga 2007). CE alone has been used for the targeted analysis of alkaloids and flavonoids (Ganzera 2008; Lao et al. 2009; Hurtado-Fernandez et al. 2010) and used for quality control (QC) of herbal medicines (Ganzera 2008). LC-MS-based profiling showed variance attributed to harvesting region and processing protocol (Xie et al. 2008; Chan et al. 2007). Our CE–MS-based profiling method is also important for QC of herbal medicines, since charged metabolites are the dominant species dissolved in water extracts (Lao et al. 2009).

Toki-Shakuyaku-San, or *Danggui-Shaoyao-San*, is a Chinese herbal medicine used widely for the therapy of hematopoiesis and menstrual disorders, and has several functions, including smooth muscle relaxation and biofluid control (Akase et al. 2004). *Toki-Shakuyaku-San* consists of six individual herbal medicines, with each providing different pharmacological effects. These herbal medicines are Toki (*Angelica acutiloba Radix*), Senkyu (*Cnidium officinale Rhizoma*), Shakuyaku (*Paeonia lactiflora Radix*), Sojutsu (*Atractylodes lancea Rhizoma*), Takusha (*Alisma orientale Rhizoma*), and Bukuryo (*Poria cocos Hoelen*). Toki and Senkyu are used for the treatment of gynecological diseases (Lu et al. 2005; Kim et al. 2006; Yi et al. 2007; Bohrmann et al. 1967), while Takusha, Bukuryo and Sojutsu are used to improve water metabolism (Wang et al. 2004, 2008; Kitajima et al. 2003; Zhao et al. 2008), and Shakuyaku provides an anti-inflammatory effect (He et al. 2010; Ohta et al. 1993). While some of the active components, such as L-ornithine, gallic acid, paeoniflorin, ferulic acid and benzoic acid in this mixture, and their effects, have been well studied, the relationships between the complete compound profile and the pharmacological effects is not yet well understood (Hatip-Al-Khatib et al. 2004; Chen et al. 2009).

This study aimed to reveal the variance of the charged metabolites found in the six herbal medicines that constitute *Toki-Shakuyaku-San*, using CE-TOFMS. We conducted multivariate analysis to examine variations in the primary metabolites of the herbal medicine and to understand the relationships between the charged metabolites and the pharmacological effects. Most studies focus on the importance of secondary metabolites to the pharmacological effects of a herbal medicine, however we show that the profile of charged metabolites also varies among herbal medicines.

2 Materials and methods

2.1 Sample preparation

Six herbal medicines, Toki, Bukuryo, Sojutsu, Takusha, Shakuyaku and Senkyu, were obtained from Yatsume Pharmaceutical Co., Ltd. (Tokyo, Japan). Each herb was homogenized without solvent using a multi-bead shocker (Yasui Kikai Co., Osaka, Japan) at 2,500 rpm for 300 s. The crushed material (1.00 g) was dissolved in 3 ml of 50% MeOH containing 2 internal standards (200 μM each of methionine sulfone and 3-aminopyrrolidine). This solution was centrifuged at 3,000 rpm for 10 min at 4°C, and 0.2 ml of the supernatant fluid was transferred to a Millipore 5 kDa cutoff filter tube for centrifugal centrifugation (10,000 rpm × 2 h at 4°C). The filtrate was lyophilized and dissolved in 100 μl of Milli-Q (Millipore, Bedford, MA, USA) water prior to CE-TOFMS analysis.

2.2 Standard chemical compounds

Cysteine-glutathione disulphide was purchased from Apollo Scientific Ltd (Tokyo, Japan). 5-Methyl-2′-deoxycytidine was purchased from MP Biomedicals LLC. (Tokyo, Japan). γ-Glu-2AB and cystathionine were purchased from Toray Research Center (Tokyo, Japan). *N*-α-Dimethylhistidine was purchased from Bachem AG (Bubendorf, Switzerland). Isobutylamine and azetidine-2-carboxylate were purchased from Chem Service Inc. (West Chester, PA, USA). N_8-Acetylspermidine, 5-aminovalerate, *O*-acetylcarnitine,

betaine, glucosaminate, 7-methylguanine, 5-methoxy-3-indoleacetate and pyridoxamine were purchased from Fluka (Buchs, Switzerland). Isonicotinamide was purchased from Tokyo Chemical Industry (Tokyo, Japan). All other compounds were purchased from Sigma–Aldrich (St. Louis, MO, USA) or Wako (Osaka, Japan). All chemical standards were dissolved in Milli-Q water, 0.1 M HCl or 0.1 M NaOH to give 10 mM or 100 mM stock solutions. A standard mixture was prepared by diluting the stock solutions with Milli-Q water just prior to injection into the CE-TOFMS. The chemicals used were of analytical or reagent grade.

2.3 Instrument parameters

The instrumentation and measurement conditions used for CE-TOFMS were as described elsewhere (Sugimoto et al. 2010a). All CE-TOFMS experiments were performed using an Agilent CE capillary electrophoresis system (Agilent Technologies, Waldbronn, Germany), an Agilent G3250 AA LS/MSD TOF system (Agilent Technologies, Palo Alto, CA), an Agilent 1100 series binary HPLC pump which delivers sheath liquid, a G1603A Agilent CE–MS adapter kit, and a G1607A Agilent CE–ESI–MS sprayer kit (Agilent Technologies, Waldbronn, Germany). The CE–MS adapter kit used includes a capillary cassette that facilitates thermostatting of the capillary. The CE–ESI–MS sprayer kit simplifies the coupling of the CE system with the MS system, which was equipped with an electrospray ionization source. For system control and data acquisition, G2201AA Agilent ChemStation software for CE and Analyst QS software for TOFMS were used.

2.4 Measurement conditions for CE-TOFMS

Samples were separated in fused silica capillaries (50 μm i.d. × 100 cm total length) filled with 1 M formic acid (pH 1.8) as the background electrolyte. The sample solutions were injected at 50 mbar for 3 s and a voltage of 30 kV was applied. The capillary temperature was maintained at 20°C and the temperature of the sample tray was maintained below 5°C using the external coolant system. The sheath liquid, comprising of methanol/water (50% v/v) and 0.1 μM hexakis-(2,2-difluorothoxy)-phosphazene (Hexakis), was delivered at 10 μl/min. ESI–TOFMS was conducted in positive ion mode. The capillary voltage was set at 4 kV and the nitrogen gas flow rate (heater temperature 300°C) was 10 psig. In TOFMS, the fragmentor, skimmer and OCT RFV voltage were set at 75, 50 and 125 V, respectively. Automatic recalibration of each acquired spectrum was performed using the reference masses of reference standards (^{13}C isotopic ion of protonated methanol dimer $[2MeOH + H]^+$, *m/z* 66.06371, and $[Hexakis + H]^+$, *m/z* 622.02896). Mass spectra were acquired at a rate of 1.5 cycles/s over a 50–1,000 *m/z* range.

2.5 Data processing for CE-TOFMS

Raw data were analyzed with our proprietary software, MasterHands (Sugimoto et al. 2010b). Initially noise-filtering, baseline correction, peak detection and peak area integration from sliced electropherograms (*m/z* 0.02 width) were carried out. Subsequently, the accurate *m/z* value for each peak detected was calculated with Gaussian curve-fitting. A migration time normalization function was established using dynamic programming and the simplex optimization method (Soga et al. 2006). Peaks with small differences in *m/z* value (<20 ppm) and normalized migration time (<1.0 min) were treated as features. Subsequently, neutral compounds, salt ions of Na^+ and K^+, and redundant features such as fragments, adducts, isotopes, dimers and trimers were eliminated on the basis of established *m/z* differences (Brown et al. 2009). For the remaining features, metabolite identities were assigned by matching their *m/z* values and migration times with those of standard compounds. To quantify the detected metabolites, the injected volume for CE and sensitivity of MS were corrected using internal standards, then all annotated metabolites were further corrected with the same chemicals in a standard mixture to overcome different ionization patterns.

2.6 Statistical analysis

The concentration of each metabolite was divided by its average to scale the data, clustered on the basis of Euclidean distance, and visualized as a heat map representation using Mev TM4 software, version 4.6 (Dana-Farber Cancer Institute, Boston, MA) (Saeed et al. 2006). We conducted principal component analysis (PCA) using JMP software (version 8.0.2, SAS, Cary, NC). Metabolite data were mapped on the pathway map using Pathway Projector (Kono et al. 2009), a web-based pathway map application using KEGG Atlas (Kanehisa et al. 2010).

3 Results

3.1 PCA analysis of charged metabolites in herbal medicines and pharmacological effects

After removing peaks determined as non-metabolite, 737 ± 183.1 (average ± SD) peaks remained. Of these, 119 were assigned based on matched *m/z* and normalized migration times with compounds in our standard library. We conducted PCA using these annotated metabolites.

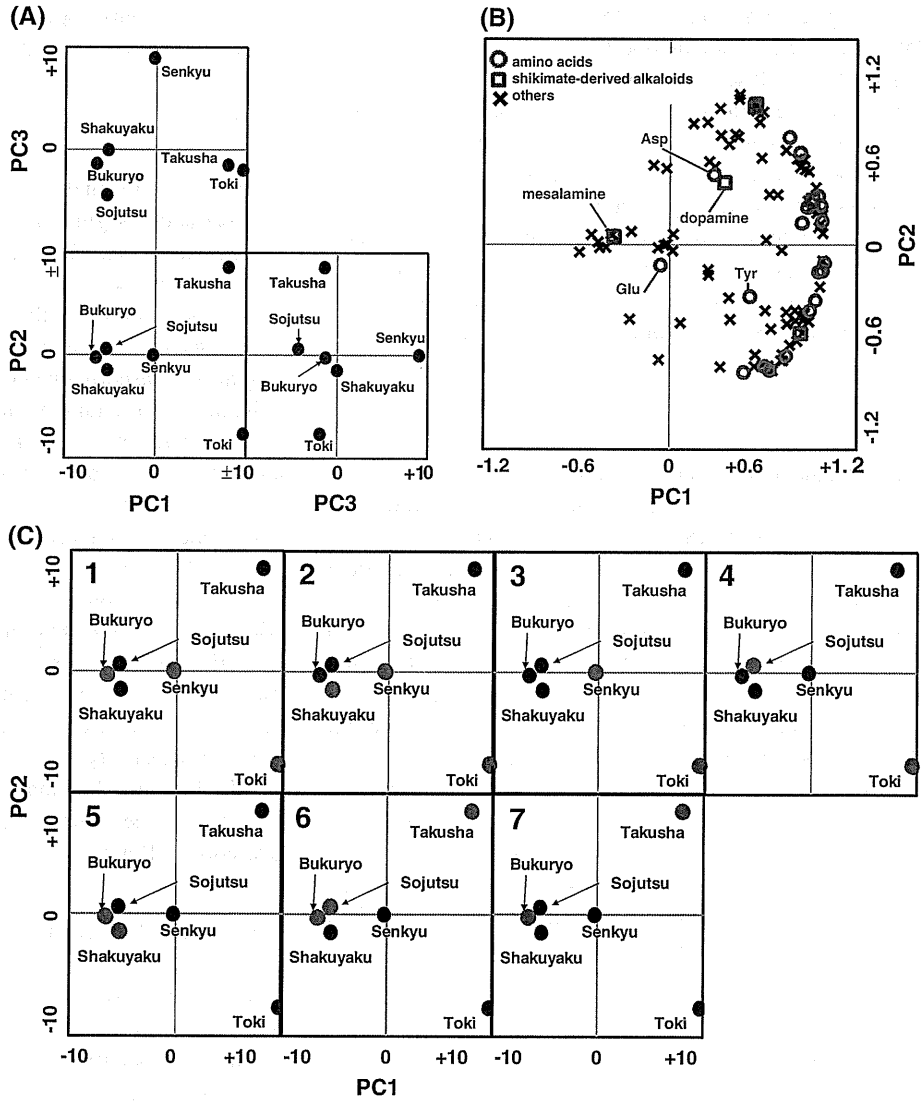

Fig. 1 PCA results of 119 annotated metabolites measured in six herbal medicines. **a** Score plots generated using all annotated metabolites. The cumulative proportions of the first, second and third PC were 43.0%, 65.9% and 83.8%, respectively. **b** Loading plots for the first and second PCs. *Red circles*, *blue squares*, and *black crosses*, represent standard amino acids, shikimate derived metabolites and other annotated metabolites, respectively. **c** Score plots for the first and second PCs. The plots colored *red* have the pharmaceutical effects of (1) tranquilization, (2) painkiller, menstruation problems (dysmenorrhea, oligomenorrhea), (3) removing blood stasis, tonic, and anemia, (4) antidiarrheal effect, (5) anti-convulsion effect (antispasmodic effect), (6) diuretic effect, and (7) oliguria, dizziness, retention of water in the stomach

Score plots (Fig. 1a) showed that, at the first principal component (PC), only Toki and Takusha were separated, showing highly positive scores (+9.5 and +7.9), while the others showed negative scores (<−0.2). However, at the second PC, Toki and Takusha were separated, showing strongly negative (−7.7) and positive scores (+8.6) respectively, while the others congregated around zero. At the third PC, only Senkyu was separated, showing a highly positive score (+8.9). Overall, loading plots (Fig. 1b; Supplementary Information Fig. 1) showed that most of the metabolites had similar loadings and no prominent features were observed. Standard amino acids, with the exception of aspartate, glutamate, and tyrosine (indicated with red circles), and shikimate-derived alkaloids, with the exception of mesalamine and dopamine (blue squares), were located furthest from the origin.

In the obtained profiles, 39 metabolites showed higher loading score values at the first PC (>0.8). Of these, 20 metabolites were amino acids and amino acid derived metabolites; threonine, valine, serine, tyrosine, glycine, phenylalanine, isoleucine, alanine, lysine, asparagine, arginine, citrulline, homoserine, *N*-ε-acetyllysine, *N*-acetylvaline, octopine, ornithine, *N*-acetylornithine, *N*-γ-ethylglutamine and *N*-methylalanine. Along the second PC axis, seven metabolites, 5-aminovalerate, acetylcholine, isoamylamine, *O*-acetyl carnitine, tryptamine, tyramine and anthranilate, showed high loading score values. Along the third PC axis, 2′-deoxyguanosine, 5′-deoxyadenosine, 5-methyl-2′-deoxycytidine, 5-methylthioadenosine, adenosine, ethanolamine phosphate, N_6-methyl-2′-deoxyadenosine and phosphorylcholine showed high loading score values. Of these, four compounds, 5′-deoxyadenosine,

5-methylthioadenosine, adenosine and N_6-methyl-2′-deoxyadenosine, were adenosine derived metabolites. In short, these three PCs characterized amino acids, amines and adenosine-derived metabolites, respectively. The pharmacological effects of the herbal medicines used in this study are summarized in Supplementary Information Table 1. To evaluate the relationships between these efficacies and the observed metabolite profiles of the herbal medicines, the efficacies shared by multiple herbal medicines were mapped on the score plot of PCA results (Fig. 1c). One distinct result found was the shared diuretic effect of Bukuryo and Sojutsu, which were closely located on the score plot, indicating a correlation between the observed metabolite profiles and this effect (Fig. 1c). Both Bukuryo and Sojutsu showed minus values in the first PC on score plots (Fig. 1a). The metabolites located at large minus values at the first PC on loading plots, such as β-alanine, isobutylamine, cytosine, cystathionine, nicotine and nicotinamide, may be characteristic of the PC scores of Bukuryo and Sojutsu (Supplementary Information Fig. 1). Senkyu and Toki are used for purposes including tranquilization, alleviation of pain, menstruation problems, removing blood stasis, as a tonic and for anemia, while they are located at a significant distance in PC score plots (Fig. 1c). Although, the PC values of Senkyu and Toki were not similar, the concentrations of several individual metabolites were clearly higher than in the other herbs; 3-methylhistidine, spermidine, adenosine, pipecolate, tryptophan, thiamine and uridine were high in Senkyu and Toki, while agmatine was high in Takusha, Bukuryo and Sojutsu (Supplementary Information Fig. 2).

3.2 Profiling of charged metabolites in herbal medicines

We conducted cluster analysis to assess the similarity of the charged metabolite profiles among the six herbs (Fig. 2). Those metabolites highlighted in yellow were present in one of the herbs at a concentration twofold higher than the average concentration of all the herbs tested. Metabolite concentrations in most clusters (labeled 1–3, and 6–8) were high only in one herb, with few clusters (4 and 5) showing high levels in multiple herb samples. For example, metabolites in clusters 1, 2, 3 and 6 were present at high concentrations in Bukuryo, Senkyu, Toki and Sojutsu, respectively, and the metabolites in clusters 7 and 8 were present at high concentrations in Takusha. The latter clusters contain three xanthine analogs, hypoxanthine, guanine and xanthosine. The amounts of these three metabolites are shown in Supplementary Information Fig. 3a. Metabolites in cluster 4 were present in significantly higher levels in Toki and Sojutsu than the other herbs tested. Metabolites in cluster 5 were present in higher levels in Toki, Takusha and Senkyu than in Shakuyaku,

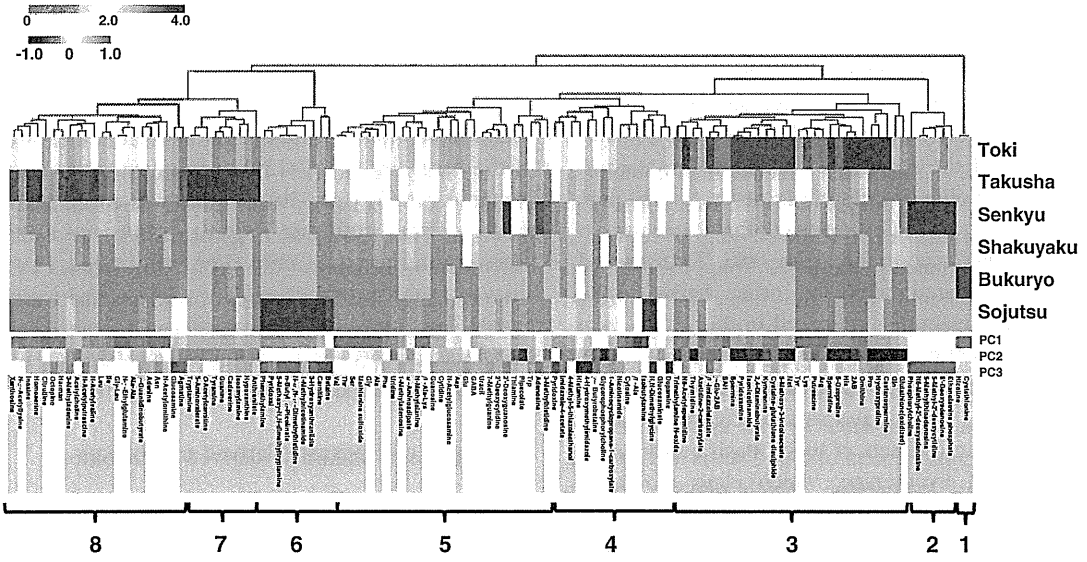

Fig. 2 Heatmap showing metabolite profiles and loading scores. *Green–red* heat map shows the quantified metabolite profiles of the six extracted herbal medicines. Each metabolite concentration shown in the heat map was divided by its average. *Blue–orange* heat map shows the loading score of the first three PCs. Metabolites highlighted in *yellow* were present in concentrations twofold higher in one herbal medicine compared with the average. See the text for numerical labels

Bukuryo and Sojutsu. A blue-orange heat map shows that the metabolites in cluster 3 were present at high loadings only in the first PC, while metabolites in clusters 7 and 8 were present at high loadings in both the first and second PC. Interestingly, metabolites in cluster 2, including five nucleic acids, ethanolaminephosphate and phosphorylcholine, showed high loading values in the third PC.

3.3 Standard amino acid concentration

Figure 3 depicts the quantification of all detected metabolites, standard amino acids, and the sum of the standard amino acids derived from the shikimate pathway. The total metabolite concentration was especially high ($>1.5 \times 10^3$ μM) in Toki, Takusha and Senkyu, while the lowest concentration (1.9×10^2 μM) was observed in Bukuryo (Fig. 3a). The sum of the standard amino acids was dominant (65–85%) except for Bukuryo (35%). The sum of tryptophan, tyrosine and phenylalanine, the shikimate derived amino acids, was notably high ($>5.9 \times 10^2$ μM) in Takusha, Toki and Senkyu (Fig. 3a). Arginine was the most dominant amino acid in Toki, Shakuyaku and Takusha (42%, 39% and 41%, respectively), and proline was present in high levels in Toki (30%) and Sojutsu (38%) (Fig. 3b).

4 Discussion

4.1 Overview of charged metabolite profiles and pharmacological effects

CE-TOFMS-based metabolomics has been used to successfully annotate mainly primary metabolites in herbal medicines, while many studies based on LC–MS or GC–MS give profiles of the secondary metabolites contained. The observed metabolite profiles of six herbs studied here showed large variances. PCA revealed that there was a particularly large variance in the herbs Toki, Takusha and Senkyu when compared with the other herbs tested (Fig. 1a). Several metabolites observed in this study have known pharmaceutical efficacies. Tryptophan is present at higher concentrations in Toki and Senkyu (Supplementary Information Fig. 2a, which have several common activities, including the improvement of blood stasis and anemia. It is a precursor of kynurenine, which plays a role in arterial vessel relaxation (Wang et al. 2010). Agmatine was present at high concentrations in Takusha, Bukuryo and Sojutsu, and is known to have a diuretic effect (Smyth and Penner 1995). Bukuryo, Sojutsu and Shakuyaku showed negative scores at the first PC and were closely plotted on the score plot (Fig. 1a) and β-alanine, isobutylamine, cytosine, cystathionine, nicotine and nicotinamide showed higher negative loading score values at the first PC (Supplementary Information Fig. 1). β-Alanine is the rate-limiting precursor of carnosine, which has an anti-inflammatory effect (Zhu et al. 2007), and single oral doses of isobutylamine have a sedative effect (Cheever et al. 1982). These metabolites therefore correlate with the clinical application of these herbs.

Clustering results, shown using a heat map, indicated that there was little similarity among the herbs tested (Fig. 2). Thus, the concentrations of many metabolites vary widely between the herbs. Izzettin Hatip-Al-Khatib et al. have shown previously that only ornithine in Toki extract is responsible for improving memory impairment (Hatip-Al-Khatib et al. 2004). Our profile revealed high concentrations of ornithine in Toki, which was consistent with their study, and in addition, other amines were also found to be present at high concentrations in Takusha. Such findings were obtained using the non-targeted profiling analysis demonstrated. Here, we discuss prominent features observed for amino acids, shikimate derived metabolites, and xanthine analogs.

4.2 Profile of amino acids

Amino acids have been reported to be the main components of Takusha extracts (Namba and Tsuda 1998), and our results are consistent with these findings; 74% of the total annotated metabolite composition was made up of standard amino acids in Takusha (Fig. 3a). However, particularly high amino acid content was not restricted to Takusha, with amino acids making up 64–85% of the composition of the other herbs tested. The only exception to this trend was Bukuryo, whose extract was composed of only 34% standard amino acids (Fig. 3a). Arginine was found to be the most dominant amino acid measured in all of the herbs tested (21–42%), again, with the exception of Bukuryo (9.4%) (Fig. 3b). Proline was found to be abundant in Toki (30%) and Sojutsu (38%), and alanine was abundant in Bukuryo (38%) (Fig. 3b). Indeed, PCA results showed high loading scores for all standard amino acids except for glutamine, asparagine and tyrosine, indicating that these amino acids may be a prominent factor useful for the characterization of these herbs (Fig. 1b).

4.3 Profile of shikimate derived metabolites

The shikimate-derived metabolites observed in our profiles are known to be bioactive, for example dopamine and kynurenine act as signal molecules in the central and peripheral nervous systems. Metabolites contained in the Toki extract are known to bind to a receptor of the central nervous system (Liao et al. 1995) and a liquid extract of Toki containing alkaloids improved retention memory in

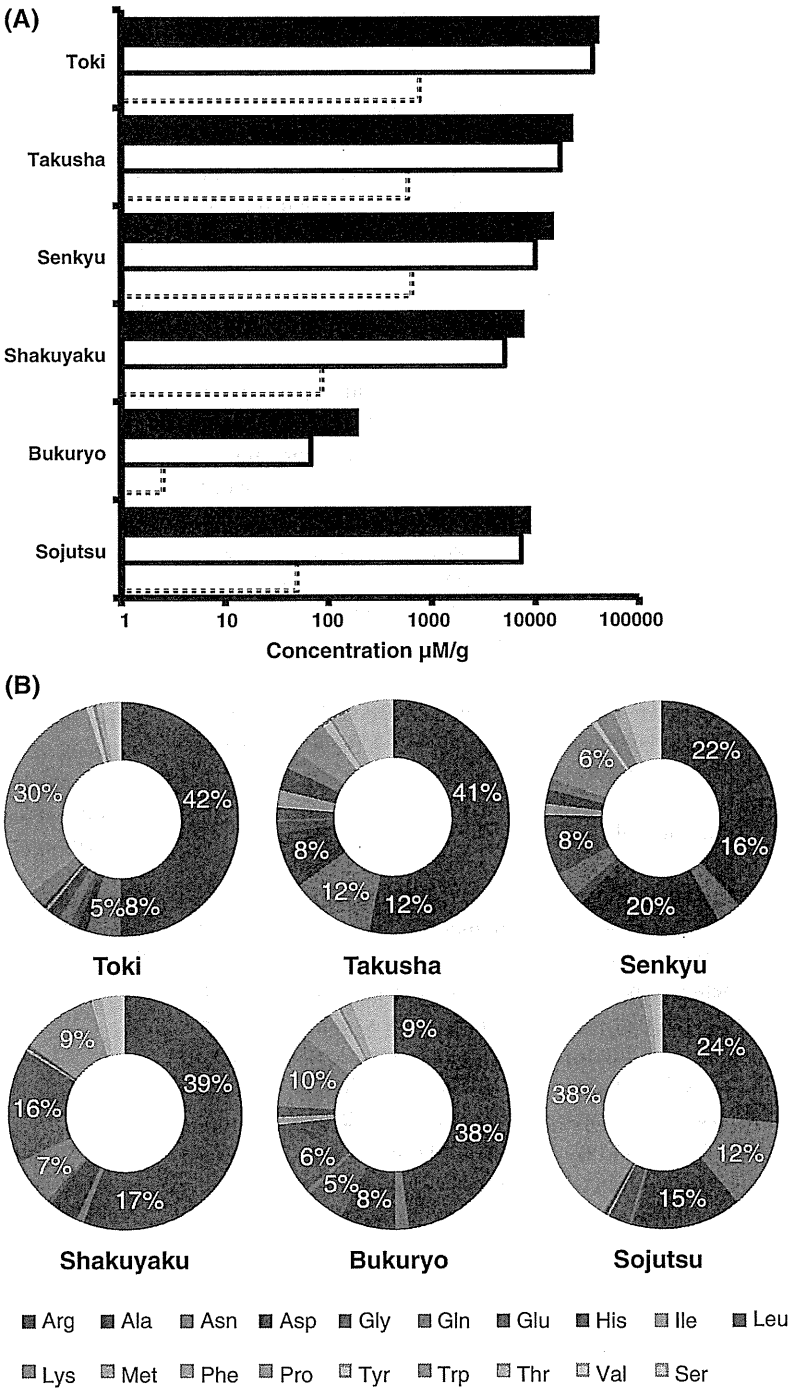

Fig. 3 Comparison of the compositions of the herbal medicines tested. **a** Total concentration of the 119 annotated metabolites (*black bar*), total concentration of standard amino acids (*white-solid bar*), and summed concentration of tryptophan, tyrosine and phenylalanine (*white-dashed bar*) are shown. **b** Standard amino acid composition of each of the six herbal medicines

rats (Hatip-Al-Khatib et al. 2004). Our profiles showed the presence of seven shikimate-derived metabolites and three precursor amino acids (tryptophan, tyrosine and phenylalanine). The total concentration of these three amino acids was higher in Toki, Senkyu and Takusha (Fig. 3a). The concentrations of the shikimate-derived metabolites detected, including kynurenine, harman, tryptamine, tyramine and anthranilate, were also high in Toki, Senkyu and Takusha (Supplementary Information Fig. 3b). These results suggest a correlation between the concentrations of shikimate-derived metabolites and those of the precursor amino acids.

Toki and Senkyu have been reported to contain ferulic acid, which is derived from tyrosine or phenylalanine

(Lu et al. 2005; Yi et al. 2007). Although, ferulic acid was not observed in our profiles, the high concentrations of the precursor amino acids of ferulic acid suggest that ferulic acid may be abundant in Toki and Senkyu. In conclusion, we have found some degree of quantitative interaction between bioactive compounds and their precursor amino acids, though there were some shikimate-derived metabolites, such as dopamine, 3-hydroxyanthranilate, which did not follow this trend.

4.4 Profiles of xanthine analogs

Xanthine analogs are known to have a wide range of biological targets. Examples include the adenosine receptor, where xanthines act as antagonists, and calcium release ryanodine-sensitive channels, where they act to sensitize the channels to calcium activation (Daly 2007). Xanthines are known to have several therapeutic benefits, including analgesic and diuretic effects (Daly 2007). Takusha was found to contain the xanthine analogs guanine (EC 3.5.4.3), hypoxanthine (EC 3.2.2.1) and xanthosine (EC 1.17.3.2 or EC 1.17.1.4) at levels more than ten, four and two times higher than the other herbs tested, respectively. Xanthine analogs may be expected to participate in the analgesic and diuretic effects of Takusha.

4.5 Limitation of this study

In the profiles we obtained, the total concentration of all annotated metabolites was higher in Toki, Takusha, and Senkyu than in the other herbs tested (Fig. 3a). The complex metabolome of these three herbal medicines, including shikimate-derived metabolites and xanthine analogs, may be responsible for the therapeutic functions associated with the herbal medicines, such as analgesia and sedation, since structurally similar compounds are known to have similar functions by receptor binding (Wambach and Casals-Stenzel 1983; Monteith et al. 1996). Further analyses are necessary to compare metabolite extraction methods, with methanol useful to achieve high recovery rates of metabolites (Ren et al. 2008; Sato et al. 2004) and water useful when considering dosage (Liang et al. 2009). More intensive studies on the individual metabolites may give insight into the effect of chemical structure on these therapeutic effects. Further study is needed to elucidate the relationship between the many polar low molecular weight compounds found in these herbal medicines and the herb's therapeutic functions, particularly their interactions with the nervous system. The quantified profiles obtained should also be validated using different samples of these herbal medicines, for example samples grown at different production sites, harvested at different times, and different batches, in order to establish the natural variability of the metabolites. Profiles of negatively charged metabolites and those obtained using other "omics" platforms should also be integrated to allow for more comprehensive analyses.

5 Concluding remarks

In this study, we have conducted a non-targeted analysis of the polar metabolite profiles of the six herbal medicines which compose *Toki-Shakuyaku-San*. Overall, there were few common characteristics observed among the herbs tested, and while the total amino acids content was consistently high among the herbs, individual amino acid content showed large variation. Notably, high concentrations of shikimate-derived metabolites were observed in Toki, Takusha and Senkyu, which helps to explain the herb's biological effects, such as sedation, analgesia and diuretic effects. Multivariable analysis revealed a high correlation between the concentrations of shikimate-derived metabolites and the concentrations of their precursor amino acids, tryptophan, tyrosine and phenylalanine, in the herbs tested. The xanthine analogs guanine, hypoxanthine and xanthosine also showed a high correlation. Analysis of the multiple pharmacological effects of these herbal medicines and comparison with the quantitative compound datasets described here would be valuable in the evaluation of the quality and efficacy of herbal medicines.

Acknowledgments This work was supported by research funds from the Yamagata Prefectural Government and the city of Tsuruoka. We thank Dr. Kazuko Otomo for technical assistance, and Wanjun Kong and Guo Jing for fruitful discussions.

References

Akase, T., Onodera, S., Matsushita, R., & Tashiro, S. (2004). A comparative study of laboratory parameters and symptoms effected by Toki-Shakuyaku-San and an iron preparation in rats with iron-deficiency anemia. *Biological and Pharmaceutical Bulletin, 27*, 871–878.

Bohrmann, H., Stahl, E., & Mitsuhashi, H. (1967). Studies of the constituents of umbelliferae plants. 8. Chromatographic studies on the constituents of *Cnidium officinale* Makino. *Chemical and Pharmaceutical Bulletin, 15*, 1606–1608.

Brown, M., Dunn, W. B., Dobson, P., et al. (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst, 134*, 1322–1332.

Calixto, J. B. (2000). Efficacy, safety, quality control, marketing and regulatory guidelines for herbal medicines (phytotherapeutic agents). *Brazilian Journal of Medical and Biological Research, 33*, 179–189.

Chan, K. (2003). Some aspects of toxic contaminants in herbal medicines. *Chemosphere, 52*, 1361–1371.

Chan, E. C., Yap, S. L., Lau, A. J., Leow, P. C., Toh, D. F., & Koh, H. L. (2007). Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed

Panax notoginseng. *Rapid Communications in Mass Spectrometry, 21*, 519–528.

Cheever, K. L., Richards, D. E., & Plotnick, H. B. (1982). The acute oral toxicity of isomeric monobutylamines in the adult male and female rat. *Toxicology and Applied Pharmacology, 63*, 150–152.

Chen, L., Qi, J., Chang, Y. X., Zhu, D., & Yu, B. (2009). Identification and determination of the major constituents in Traditional Chinese Medicinal formula Danggui-Shaoyao-San by HPLC-DAD-ESI-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis, 50*, 127–137.

Daly, J. W. (2007). Caffeine analogs: Biomedical impact. *Cellular and Molecular Life Sciences, 64*, 2153–2169.

Ganzera, M. (2008). Quality control of herbal medicines by capillary electrophoresis: Potential, requirements and applications. *Electrophoresis, 29*, 3489–3503.

Gong, F., Liang, Y. Z., Cui, H., Chau, F. T., & Chan, B. T. (2001). Determination of volatile components in peptic powder by gas chromatography-mass spectrometry and chemometric resolution. *Journal of Chromatography A, 909*, 237–247.

Hatip-Al-Khatib, I., Egashira, N., Mishima, K., et al. (2004). Determination of the effectiveness of components of the herbal medicine Toki-Shakuyaku-San and fractions of *Angelica acutiloba* in improving the scopolamine-induced impairment of rat's spatial cognition in eight-armed radial maze test. *Journal of Pharmacological Sciences, 96*, 33–41.

He, C. N., Peng, Y., Xu, L. J., et al. (2010). Three new oligostilbenes from the seeds of Paeonia suffruticosa. *Chemical and Pharmaceutical Bulletin, 58*, 843–847.

Hurtado-Fernandez, E., Gomez-Romero, M., Carrasco-Pancorbo, A., & Fernandez-Gutierrez, A. (2010). Application and potential of capillary electroseparation methods to determine antioxidant phenolic compounds from plant food material. *Journal of Pharmaceutical and Biomedical Analysis, 53*, 1130–1160.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research, 38*, D355–D360.

Kim, M. R., Abd El-Aty, A. M., Choi, J. H., Lee, K. B., & Shim, J. H. (2006). Identification of volatile components in Angelica species using supercritical-CO_2 fluid extraction and solid phase microextraction coupled to gas chromatography-mass spectrometry. *Biomedical Chromatography, 20*, 1267–1273.

Kitajima, J., Kamoshita, A., Ishikawa, T., et al. (2003). Glycosides of Atractylodes lancea. *Chemical and Pharmaceutical Bulletin, 51*, 673–678.

Kono, N., Arakawa, K., Ogawa, R., et al. (2009). Pathway projector: Web-based zoomable pathway browser using KEGG atlas and Google Maps API. *PLoS One, 4*, e7710.

Lao, Y. M., Jiang, J. G., & Yan, L. (2009). Application of metabonomic analytical techniques in the modernization and toxicology research of traditional Chinese medicine. *British Journal of Pharmacology, 157*, 1128–1141.

Li, X. N., Cui, H., Song, Y. Q., Liang, Y. Z., & Chau, F. T. (2003). Analysis of volatile fractions of *Schisandra chinensis* (Turcz.) Baill. using GC–MS and chemometric resolution. *Phytochemical Analysis, 14*, 23–33.

Liang, X. M., Jin, Y., Wang, Y. P., Jin, G. W., Fu, Q., & Xiao, Y. S. (2009). Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *Journal of Chromatography A, 1216*, 2033–2044.

Liao, J. F., Jan, Y. M., Huang, S. Y., Wang, H. H., Yu, L. L., & Chen, C. F. (1995). Evaluation with receptor binding assay on the water extracts of ten CNS-active Chinese herbal drugs. *Proceedings of the National Science Council, Republic of China. Part B, Life Sciences, 19*, 151–158.

Liu, S., Yi, L. Z., & Liang, Y. Z. (2008). Traditional Chinese medicine and separation science. *Journal of Separation Science, 31*, 2113–2137.

Lu, G. H., Chan, K., Liang, Y. Z., et al. (2005). Development of high-performance liquid chromatographic fingerprints for distinguishing Chinese Angelica from related umbelliferae herbs. *Journal of Chromatography A, 1073*, 383–392.

Ma, C. M., Winsor, L., & Daneshtalab, M. (2007). Quantification of spiroether isomers and herniarin of different parts of *Matricaria matricarioides* and flowers of *Chamaemelum nobile*. *Phytochemical Analysis, 18*, 42–49.

Miller, F. G., Emanuel, E. J., Rosenstein, D. L., & Straus, S. E. (2004). Ethical issues concerning research in complementary and alternative medicine. *JAMA, 291*, 599–604.

Monteith, D. K., Emmerling, M. R., Garvin, J., & Theiss, J. C. (1996). Cytotoxicity study of tacrine, structurally and pharmacologically related compounds using rat hepatocytes. *Drug and Chemical Toxicology, 19*, 71–84.

Monton, M. R., & Soga, T. (2007). Metabolome analysis by capillary electrophoresis–mass spectrometry. *Journal of Chromatography A, 1168*, 237–246. discussion 236.

Murray, R. H., & Rubel, A. J. (1992). Physicians and healers—Unwitting partners in health care. *New England Journal of Medicine, 326*, 61–64.

Namba, T., & Tsuda, Y. (1998). *Shoyakugakugairon* (3rd ed.). Tokyo: Nankodo.

Ohta, H., Ni, J. W., Matsumoto, K., Watanabe, H., & Shimizu, M. (1993). Peony and its major constituent, paeoniflorin, improve radial maze performance impaired by scopolamine in rats. *Pharmacology, Biochemistry and Behavior, 45*, 719–723.

Ren, M. T., Chen, J., Song, Y., Sheng, L. S., Li, P., & Qi, L. W. (2008). Identification and quantification of 32 bioactive compounds in *Lonicera* species by high performance liquid chromatography coupled with time-of-flight mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis, 48*, 1351–1360.

Saeed, A. I., Bhagabati, N. K., Braisted, J. C., et al. (2006). TM4 microarray software suite. *Methods in Enzymology, 411*, 134–193.

Sato, S., Soga, T., Nishioka, T., & Tomita, M. (2004). Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *The Plant Journal, 40*, 151–163.

Smyth, D. D., & Penner, S. B. (1995). Renal I1-imidazoline receptor-selective compounds mediate natriuresis in the rat. *Journal of Cardiovascular Pharmacology, 26*(Suppl. 2), S63–S67.

Soga, T., Baran, R., Suematsu, M., et al. (2006). Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *The Journal of Biological Chemistry, 281*, 16768–16776.

Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., & Nishioka, T. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research, 2*, 488–494.

Sugimoto, M., Koseki, T., Hirayama, A., et al. (2010a). Correlation between sensory evaluation scores of Japanese sake and metabolome profiles. *Journal of Agriculture and Food Chemistry, 58*, 374–383.

Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T., & Tomita, M. (2010b). Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics, 6*, 78–95.

Wambach, G., & Casals-Stenzel, J. (1983). Structure-activity relationship of new steroidal aldosterone antagonists. Comparison of the affinity for mineralocorticoid receptors in vitro and the

antialdosterone activity in vivo. *Biochemical Pharmacology, 32*, 1479–1485.

Wang, H. X., Liu, C. M., Liu, Q., & Gao, K. (2008). Three types of sesquiterpenes from rhizomes of Atractylodes lancea. *Phytochemistry, 69*, 2088–2094.

Wang, Y., Liu, H., Mckenzie, G., et al. (2010). Kynurenine is an endothelium-derived relaxing factor produced during inflammation. *Nature Medicine, 16*, 279–285.

Wang, Z. G., & Ren, J. (2002). Current status and future direction of Chinese herbal medicine. *Trends in Pharmacological Sciences, 23*, 347–348.

Wang, J., Van Der Heijden, R., Spruit, S., et al. (2009). Quality and safety of Chinese herbal medicines guided by a systems biology perspective. *Journal of Ethnopharmacology, 126*, 31–41.

Wang, Y., Zhang, M., Ruan, D., et al. (2004). Chemical components and molecular mass of six polysaccharides isolated from the sclerotium of Poria cocos. *Carbohydrate Research, 339*, 327–334.

Weng, Q., & Jin, W. (2002). Carbon fiber bundle–Au–Hg dual-electrode detection for capillary electrophoresis. *Journal of Chromatography A, 971*, 217–223.

Xie, G., Plumb, R., Su, M., et al. (2008). Ultra-performance LC/TOF MS analysis of medicinal Panax herbs for metabolomic research. *Journal of Separation Science, 31*, 1015–1026.

Yi, T., Leung, K. S., Lu, G. H., & Zhang, H. (2007). Comparative analysis of *Ligusticum chuanxiong* and related umbelliferous medicinal plants by high performance liquid chromatography–electrospray ionization mass spectrometry. *Planta Medica, 73*, 392–398.

Zhang, A., Sun, H., Wang, Z., Sun, W., Wang, P., & Wang, X. (2010). Metabolomics: Towards understanding traditional Chinese medicine. *Planta Medica, 76*, 2026–2035.

Zhao, M., Xu, L. J., & Che, C. T. (2008). Alisolide, alisols O and P from the rhizome of *Alisma orientale*. *Phytochemistry, 69*, 527–532.

Zhu, Y. Y., Zhu-Ge, Z. B., Wu, D. C., et al. (2007). Carnosine inhibits pentylenetetrazol-induced seizures by histaminergic mechanisms in histidine decarboxylase knock-out mice. *Neuroscience Letters, 416*, 211–216.

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Serum metabolomics reveals γ-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease

Tomoyoshi Soga[1,*], Masahiro Sugimoto[1], Masashi Honma[2], Masayo Mori[1], Kaori Igarashi[1], Kasumi Kashikura[1], Satsuki Ikeda[1], Akiyoshi Hirayama[1], Takehito Yamamoto[2], Haruhiko Yoshida[3], Motoyuki Otsuka[3], Shoji Tsuji[4], Yutaka Yatomi[4], Tadayuki Sakuragawa[5], Hisayoshi Watanabe[6], Kouei Nihei[7], Takafumi Saito[6], Sumio Kawata[6], Hiroshi Suzuki[2], Masaru Tomita[1], Makoto Suematsu[5]

[1]*Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0052, Japan;* [2]*Department of Pharmacy, The University of Tokyo Hospital, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan;* [3]*Department of Gastroenterology, Faculty of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan;* [4]*Department of Neurology, Division of Neuroscience, Graduate School of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan;* [5]*Department of Biochemistry, JST ERATO Suematsu Gas Biology Project, School of Medicine, Keio University, Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan;* [6]*Department of Gastroenterology, Yamagata University School of Medicine, Yamagata 990-9585, Japan;* [7]*Department of Surgery, Shonai Hospital, 4-20 Izumi-cho, Tsuruoka 997-8515, Japan*

Background & Aims: We applied a metabolome profiling approach to serum samples obtained from patients with different liver diseases, to discover noninvasive and reliable biomarkers for rapid-screening diagnosis of liver diseases.

Methods: Using capillary electrophoresis and liquid chromatography mass spectrometry, we analyzed low molecular weight metabolites in a total of 248 serum samples obtained from patients with nine types of liver disease and healthy controls.

Results: We found that γ-glutamyl dipeptides, which were biosynthesized through a reaction with γ-glutamylcysteine synthetase, were indicative of the production of reduced glutathione, and that measurement of their levels could distinguish among different liver diseases. Multiple logistic regression models facilitated the discrimination between specific and other liver diseases and yielded high areas under receiver-operating characteristic curves. The area under the curve values in training and independent validation data were 0.952 and 0.967 in healthy controls, 0.817 and 0.849 in drug-induced liver injury, 0.754 and 0.763 in asymptomatic hepatitis B virus infection, 0.820 and 0.762 in chronic hepatitis B, 0.972 and 0.895 in hepatitis C with persistently normal alanine transaminase, 0.917 and 0.707 in chronic hepatitis C, 0.803 and 0.993 in cirrhosis type C, and 0.762 and 0.803 in hepatocellular carcinoma, respectively. Several γ-glutamyl dipeptides also manifested potential for differentiating between nonalcoholic steatohepatitis and simple steatosis.

Conclusions: γ-Glutamyl dipeptides are novel biomarkers for liver diseases, and varying levels of individual or groups of these peptides have the power to discriminate among different forms of hepatic disease.

Keywords: γ-Glutamyl dipeptides; Metabolomics; Biomarker; Capillary electrophoresis mass spectrometry; Oxidative stress; Glutathione; Hepatocellular carcinoma; Nonalcoholic steatohepatitis; Hepatitis C virus.

Received 15 July 2010; received in revised form 17 January 2011; accepted 24 January 2011; available online 18 February 2011

* Corresponding author. Address: Institute for Advanced Biosciences, Keio University, 246-2 Mizukami, Kakuganji, Tsuruoka, Yamagata 997-0052, Japan. Tel.: +81 235 29 0528; fax: +81 0235 29 0574.

E-mail address: soga@sfc.keio.ac.jp (T. Soga).

Abbreviations: HCC, hepatocellular carcinoma; AST, aspartate transaminase; ALT, alanine transaminase; γ-GTP, γ-glutamyl transpeptidase; CT, computed tomography; NAFLD, nonalcoholic fatty liver disease; SS, simple steatosis; NASH, nonalcoholic steatohepatitis; CE-TOFMS, capillary electrophoresis time-of-flight mass spectrometry; GSH, reduced glutathione; GC, gastric cancer; GCS, γ-glutamylcysteine synthetase; C, healthy control; DI, drug-induced liver injury; AHB, asymptomatic hepatitis B virus infection; CHB, chronic hepatitis B; CNALT, hepatitis C with persistently normal alanine transaminase; CHC, chronic hepatitis C; CIR, cirrhosis type C; HBs, hepatitis B surface; HBV, hepatitis B virus; HCV, hepatitis C virus; AFP, α-fetoprotein; PIVKA, protein induced by vitamin K antagonist; LC–MS/MS, liquid chromatography–electrospray tandem mass spectrometry; MLR, multiple logistic regression; APAP, acetaminophen; GS, glutathione synthetase; BSO, buthionine sulfoximine; DEM, diethylmaleate; ROS, reactive oxygen species.

Introduction

Acute or chronic viral hepatitis affects populations around the world, and the disease often progresses from chronic hepatitis and cirrhosis to hepatocellular carcinoma (HCC) [1]. Accurate diagnosis at earlier stages is necessary for improved therapeutic outcome. However, the diagnostic procedures are laborious and not risk-free. Patients with suspected liver damage are initially subjected to liver function tests that include the assessment of aspartate transaminase (AST), alanine transaminase (ALT), and γ-glutamyl transpeptidase (γ-GTP) serum levels. If these levels are abnormal, patients are then subjected to diagnostic imaging, such as ultrasound and computed tomography (CT), and assays to determine the presence of antibodies against hepatitis virus. Finally, a liver biopsy may be recommended to evaluate the severity of inflammation or fibrosis and to confirm the indications for antiviral therapy.

Recently, nonalcoholic fatty liver disease (NAFLD) has become the most common liver disease in western countries. It