

Hypothesis

GC–MS libraries for the rapid identification of metabolites in complex biological samples

Nicolas Schauer^a, Dirk Steinhauser^a, Sergej Strelkov^b, Dietmar Schomburg^b, Gordon Allison^c, Thomas Moritz^d, Krister Lundgren^d, Ute Roessner-Tunali^e, Megan G. Forbes^e, Lothar Willmitzer^a, Alisdair R. Fernie^a, Joachim Kopka^{a,*}

^a Max-Planck Institute of Plant Molecular Physiology, Am Muehlenberg 1, D-14476 Golm, Germany

^b University of Köln, CUBIC – Institute of Biochemistry, Zulpicher Street 47, D-50674 Köln, Germany

^c Institute of Grassland and Environmental Research, SY23 3EB, Aberystwyth, Wales, UK

^d Umea Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-901 83 Umea, Sweden

^e Australian Centre for Plant Functional Genomics, School of Botany, University of Melbourne, Vic. 3010, Australia

Received 17 December 2004; revised 12 January 2005; accepted 19 January 2005

Available online 28 January 2005

Edited by Lukas Huber

Abstract Gas chromatography–mass spectrometry based metabolite profiling of biological samples is rapidly becoming one of the cornerstones of functional genomics and systems biology. Thus, the technology needs to be available to many laboratories and open exchange of information is required such as those achieved for transcript and protein data. The key-step in metabolite profiling is the unambiguous identification of metabolites in highly complex metabolite preparations with composite structure. Collections of mass spectra, which comprise frequently observed identified and non-identified metabolites, represent the most effective means to pool the identification efforts currently performed in many laboratories around the world. Here, we describe a platform for mass spectral and retention time index libraries that will enable this process (MSRI; www.csdb.mpg.de/gmd.html). This resource should ameliorate many of the problems that each laboratory will face both for the initial establishment of metabolome analysis and for its maintenance at a constant sample throughput.

© 2005 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: GC–MS; Gas chromatography–mass spectrometry; Mass spectral library; Metabolite profiling; Metabolomics; Retention time index

1. Introduction

In the last decade, the maturity of genomic technologies generated a vast amount of sequence data and thus allowed full insight into the finite number of genes which constitute organisms. As a consequence, biological science went through a par-

adigm-change and today focuses on unravelling gene function and regulation. With these tasks at hand, comprehensive technologies have been developed which aim at comprehensive and non-biased monitoring of gene expression, and coinciding effects on protein composition and changes in metabolism. Consequently, new fields emerged in biological science which we today call transcriptomics, proteomics and metabolomics. With increasing amount and diversity of “-omics” data, the need for standardization by the research community arises and availability of tools for a user-friendly, open access to the flood of information has become essential.

One of the first “-omics” databases, BRENDA, was developed in 1987 [1]. BRENDA is a powerful database of enzyme and metabolic information initially published as a series of books, now adapted to a relational database and accessible through the worldwide-web. BRENDA hosts about 83 000 different enzymes from 9800 different organisms and describes enzyme function, taxonomy, sequences and enzyme ligands. The Munich Information Centre for Protein Sequences (MIPS) provides databases related to protein sequences based on whole genome analysis and annotation. For example, MIPS hosts databases of *Saccharomyces cerevisiae* and *Neurospora crassa* which comprise maps of protein–protein interactions, protein localization, and information on transcription factors, cDNA libraries and gene homology [2]. Transcript profiling rapidly evolved into a worldwide accepted and generally applied laboratory tool. Subsequently, databases were designed and established, which efficiently deal with transcriptome data. The Stanford microarray database (SMD), which hosts data of over 3500 DNA-microarrays of 12 distinct organisms, including bacteria, plants and animals, was the first implementation to fulfil this aim [3].

With the full availability of the human genome sequence the need and opportunity to understand the structure and function of all proteins, beyond those with enzymatic properties, was met with respective initiatives. For example HPI, the human protein initiative, focuses on the annotation of both the human genome and proteome. As proteins are generally regarded to determine cellular function, the full exploration of the prote-

*Corresponding author.

E-mail address: kopka@mpimp-golm.mpg.de (J. Kopka).

Abbreviations: GC, gas chromatography; MS, mass spectrometry; MST, mass spectral metabolite tag; RI, retention time index; TOF, time of flight

ome will be crucial. The goal of HPI is to deliver this information in high quality to facilitate further investigations of the genomic and proteomic data [4].

Presently, a wealth of databases houses information gathered at the genomic, transcriptomic, proteomic and metabolomic levels of life, e.g. [1,5]. However, there is a significant lack of a metabolome database, capable of storing the flood of data arising from analysis of biological samples using established gas chromatography–mass spectrometry (GC–MS) techniques for metabolome [6–9] and fluxome analysis [10–12]. Most promisingly, first efforts have already been made by the plant metabolomics community to agree on conventions for data formats and the description of metabolomics experiments [13,14].

2. GC–MS based metabolome analysis: Application and key challenge

GC–MS based metabolome analysis has profound applications in discovering the mode of action of drugs or herbicides and helps unravel the effect of altered gene expression on metabolism and organism performance in biotechnological applications. The prerequisite and thus key challenge of metabolite profiling is the rapid, reliable and unambiguous identification of hundreds of metabolites in highly complex preparations, such as blood plasma, intracellular microbial extracts, or complex plant and animal samples. Identification is routinely performed by time-consuming standard addition experiments using commercially available or purified metabolite preparations. Thus, a strong need for a publicly accessible database exists, harbouring the evidence and underlying metabolite identification in complex GC–MS profiles from diverse biological sources. In addition, the non-supervised collection of as yet unidentified mass spectra of metabolites, “so-called” mass spectral metabolite tags (MSTs), will most likely be highly effective for future identification efforts and discovery of novel metabolic markers. In this report, we present a platform of mass spectral and retention time index (MSRI) libraries, generated using identical types of capillary GC columns, however, utilizing two independent GC–MS detection technologies, namely quadrupole (QUAD) GC–MS [6,7,9,15] and GC-TOF (time of flight)-MS [8,16]. In the following study, we will present three test cases which illustrate the general applicability of this library for the key processes

of GC–MS based metabolite profiling, (i) identification or preliminary classification of all MST components, which are present in any given biological sample, (ii) query for those biological samples that contain a certain metabolite, (iii) matching of metabolite identifications made on different GC–MS systems and by different laboratories.

3. Mass spectral and retention time index libraries for GC–MS

We propose public exchange and open access of mass spectral identifications from GC–MS metabolite profiles, for example, through a web-based platform of MSRI libraries (www.csbdb.mpimp-golm.mpg.de/gmd.html [17]). In addition, we provide downloadable files, which can be imported into the currently leading and widely accepted NIST02 mass spectral search program or AMDIS, the automated mass spectral deconvolution and identification system (National Institute of Standards and Technology, Gaithersburg, MD, USA) [18,19]. Both software systems are publicly available from www.chemdata.nist.gov/mass-spc/amdis/ and www.chemdata.nist.gov/mass-spc/Srch_v1.7/index.html. Our libraries are classified according to technology and degree of manual mass spectral identification that was required for the library construction. After import into NIST02, the current libraries may be fused into one or customized subsets generated. Q_MSRI and T_MSRI libraries contain MSTs, which were either generated on three identically configured quadrupole (Q_MSRI) GC–MS systems or on a single time of flight (T_MSRI) system. All systems were run with identical settings except for the temperature program and scanning rate (refer to the MSRI Library: Methods on the web). Mass spectral libraries, which exclusively comprise manually evaluated, identified or classified MSTs, are assigned to ID-libraries, indicative of supervised identifications. Libraries which were generated exclusively by automated deconvolution were assigned NS indicative of the non-supervised mode of construction. The NS-libraries may contain deconvolution errors, such as multiple mass spectra for single components, accidental deconvolutions, due to random fluctuations of background noise, or partial and mixed, in other words, chimeric mass spectra of metabolic components. In addition, detailed information on processed biological samples, source of pure reference compounds, respective collaborators and previous citations is provided. For those queries on the current mass spectral

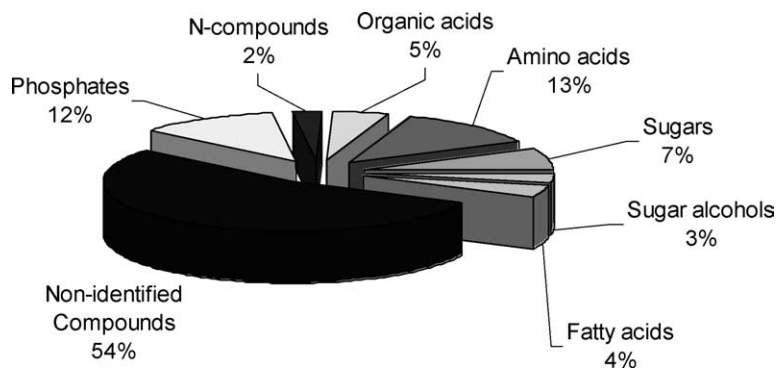


Fig. 1. The distribution of major compound classes, which were identified in GC-TOF-MS profiles in cellular extracts from *Corynebacterium glutamicum*. Note that more than 50% of the metabolites that are covered by GC–MS are currently non-identified MSTs.

Table 1
Metabolites found in preparations of blood plasma samples from domestic sheep

Class	Metabolite	
<i>Amino acids</i> ^a	2-Aminobutyric acid	
	4-Hydroxyproline	
	Alanine	
	β-Alanine	
	Glycine	
	Alanine	
	Arginine	
	Asparagine	
	Cysteine	
	Glutamic acid	
	Glutamine	
	Homoserine	
	Isoleucine	
	Leucine	
	Lysine	
	Methionine	
	Phenylalanine	
	Proline	
	Serine	
	Threonine	
	Tryptophan	
	Tyrosine	
	Valine	
	<i>N</i> -Acetylglycine	
	Ornithine	
	Pyroglutamic acid	
	<i>S</i> -Methyl-cysteine	
	<i>Organic acids</i>	2-Ketoglutaric acid
		4-Hydroxybenzoic acid
		Benzoic acid
		Citric acid
		Erythronic acid
Fumaric acid		
Gluconic acid		
Glucuronic acid		
Glutaric acid		
Glyceric acid		
Gulonic acid		
Isocitric acid		
Itaconic acid		
Malic acid		
Threonic acid		
<i>trans</i> -Sinapinic acid		
<i>Lipids</i>		9,12-(<i>cis,cis</i>)-Octadecadienoic acid
	Hexadecanoic acid	
	Octadecanoic acid	
	α-Tocopherol	
	β-Sitosterol	
	Campesterol	
	Cholesterol	
<i>Phosphates</i>	Adenosine-5-monophosphate	
	Glycerol-3-phosphate	
	<i>myo</i> -Inositol-phosphate	
	Phosphoric acid	
<i>Sugars</i> ^a	Arabinose	
	Fructose	
	Glucose	
	Raffinose	
	Ribose	
<i>N</i> -compounds	Allantoin	
	Hypoxanthine	
	Inosine	
	Thymine	

Table 1 (continued)

Class	Metabolite
<i>Alcohols</i>	Erythritol
	Glycerol
	<i>myo</i> -Inositol
	Sorbitol
	Threitol
	Xylitol
	Kaempferol

^aWith rare exceptions DL stereoisomers are not separated on the current choice of GC capillary column.

collection, which cannot be performed within NIST02, we offer a tab delimited compilation of the manually evaluated mass spectra (refer to the MSRI library: descriptions on the web) and access through web query forms. Currently, we support queries within ID-libraries, such as compound search, mass spectral search using names or mass spectra and customized library generation for subsets of mass spectra.

We previously demonstrated that both mass spectrum and retention time index are required for unequivocal metabolite identification in GC–MS profiles [16]. This feature was not available in commercial mass spectral comparison software. Therefore, the central feature of our web search forms is optional restriction of searches to RI windows and sorting of hit lists according to RI deviation and mass spectral similarity. For a shortlist of the currently implemented matching tools and queries please refer to [17].

The present version of the Q_MSRI_ID library contains 1166 identified or annotated MSTs, which represent 574 non-redundant compounds. Of these compounds 306 are unambiguously identified, while the residual MSTs are annotated with the best mass spectral match from a commercially available mass spectral collection (National Institute of Standards and Technology, Gaithersburg, MD, USA). The T_MSRI_ID collection has a similar size, namely 855 MSTs with 229 identifications within the set of 632 non-redundant components. The non-supervised collections comprise close to 30000 MSTs from a range of plant organs, root, leaf, tuber, stolon, flower, and fruits in different developmental stages, and suitable non-samples controls. Plant species covered are model plants, crops and related wild species, such as *Lotus japonicus*, *Arabidopsis thaliana*, *Solanum tuberosum*, *Nicotiana tabacum*, *Solanum lycopersicum*, *Solanum pennellii*, *Solanum parviflorum*, *Solanum pimpinellifolium*, *Solanum habrochaites*, *Solanum neorickii*.

4. Test cases

4.1. Test case 1: Analysis of sample composition

Non-supervised MSRI data allow screening for differences in various samples. A given biological sample can be compared to the non-supervised library. All MSTs, which match in their mass spectrum and RI, within certain thresholds, such as mass spectral match >650 and RI deviation <3.0, will be presented as possible hits, thus allowing the evaluation of whole biological samples for differences in composition with respect to mass spectral datasets from the established MSRI library. In the following, we applied the above thresholds for automated identification but still performed additional manual verification on each of the best hits.

A supervised database is a valuable tool to identify compounds with known RI and mass spectra in specific biological samples. A typical example of the metabolite composition from polar bacterial extracts demonstrates the scope of GC–MS based metabolite profiling (Fig. 1). To further illustrate the power of this tool, we have chosen the plant specific flavonol (kaempferol), a phytosterol (β -sitosterol) and vitamin E (α -tocopherol). Taking into account that sheep are herbivores, we expected to find β -sitosterol and kaempferol also in sheep plasma samples. To test our hypothesis, we have performed a MSRI library searching for these compounds in sheep blood plasma samples, resulting in mass spectral hits for β -sitosterol and kaempferol in the plasma composition (Table 1).

It is also known that tocopherol and its derivatives play an important role in the human diet and thus are important targets for novel nutrigenomics approaches [20,21]. Tocopherol is additionally widely hypothesized to be helpful in preventing diseases associated with oxidative stress. Therefore, the question arose whether this substance can be easily identified in mammal tissues? Here, we demonstrate the power of a MSRI library to search for α -tocopherol in different samples from animal, as well as plant tissues. Considering the importance of this compound for mammals, we expected to find α -tocopherol in animal samples and querying the library indeed resulted in the identification of α -tocopherol in blood plasma sample from sheep (Table 1).

4.2. Test case 2: Analysis of metabolite occurrence

A laboratory which maintains a GC–MS based metabolite profiling facility will continually need to identify metabolites. Frequently, the identity of previously non-identified MSTs will be discovered and the question will arise if these MSTs were found in previous experiments [7,16,22,23] or by other labora-

tories [8,9]. It will therefore be important to identify the type and source of sample, which showed this MST. For this purpose, non-supervised mass spectral libraries, which may hold independently repeated analyses of each type of sample, will be valuable tools. We chose chlorogenic acid, a typical secondary product of *solanaceous* species, and the ubiquitous precursor quinic acid to demonstrate the possible gain of knowledge to be retrieved from non-supervised mass spectral libraries (Table 2). Our analysis indicated the presence above detection limit of quinic acid in almost all profiles analysed, whereas caffeic acid, the second precursor of chlorogenic acid, was present above detection limit only in leaf and *L. japonicus* nodules. In agreement with expectations, chlorogenic acid and its positional isomers were found with good mass spectral match and RI deviation in *Solanum* samples.

4.3. Test case 3: GC–MS system transfer of metabolite identifications

Almost all metabolites were analysed either in different laboratories or on two GC–MS technology platforms, GC–QUAD–MS and GC–TOF–MS. The resulting information on retention time indices from both technology platforms clearly demonstrated strict linearity in a comparative analysis of both systems, provided the same type of capillary column was used (Fig. 2). Thus, RI prediction through regression appears highly feasible for different GC–MS systems, but only when identical column types are used. Nevertheless, we detected compound specific deviations from the prediction (Fig. 2). On average we observed an error of ~ 5.4 RI units, but most deviations were minor and within the expected range taking the typical reproducibility of retention time indices within one system into consideration [16], namely up to 2.0 RI units (standard deviation), depending mostly on changes in metabolite amount. In

Table 2
Occurrence of metabolites in non-supervised libraries of *A. thaliana*, *L. japonicus* and *Solanum* species

Species	Organ	Quinic acid		Caffeic acid		Chlorogenic acid		4-Caffeoylquinic acid		5-Caffeoylquinic acid	
		Match	Δ RI	Match	Δ RI	Match	Δ RI	Match	Δ RI	Match	Δ RI
<i>Arabidopsis thaliana</i> (L.) Heynh.	Leaf	649	1.9								
<i>Arabidopsis thaliana</i> (L.) Heynh.	Root	723	2.0								
<i>Lotus japonicus</i>	Root lateral										
<i>Lotus japonicus</i>	Root primary										
<i>Lotus japonicus</i>	Nodule			752	–2.4						
<i>Lotus japonicus</i>	Flower	795	0.4								
<i>Lotus japonicus</i>	Leaf developing	685	0.1								
<i>Lotus japonicus</i>	Leaf mature	565 ^a	0.2								
<i>Solanum lycopersicum</i>	Root	963	–0.2			974	–0.4				
<i>Solanum lycopersicum</i>	Leaf	854	0.5	838	–0.8	975	0.1	939	1.1	822	0.7
<i>Solanum lycopersicum</i>	Green fruit	967	1.4			970	1.6				
<i>Solanum lycopersicum</i>	Orange fruit	930	1.3			965	1.6				
<i>Solanum lycopersicum</i>	Red fruit	964	0.7			949	1.8	916	2.0		
<i>Solanum neorickii</i>	Fruit 45DAF	964	0.9			976	1.1	555 ^a	2.1		
<i>Solanum neorickii</i>	Leaf	950	–0.9	862	–1.9	974	–1.2	971	0.9	827	1.0
<i>Solanum habrochaites</i>	Fruit 45DAF	959	0.6			949	–0.4	920	1.2		
<i>Solanum habrochaites</i>	Leaf	909	0.5	825	–1.9	936	0.6	962	0.9	747	0.9
<i>Solanum parviflorum</i>	Fruit 45DAF	964	0.8			888	0.8				
<i>Solanum parviflorum</i>	Leaf	926	–1.0	856	–1.6	975	–0.2	971	1.2	802	1.8
<i>Solanum pennellii</i>	Fruit 45DAF	799	0.0			974	–0.6				
<i>Solanum pennellii</i>	Leaf	953	–0.4	848	–0.6	973	1.3	925	1.5		
<i>Solanum pimpinellifolium</i>	Fruit 45DAF	778	–0.7			977	0.9				
<i>Solanum pimpinellifolium</i>	Leaf	912	–1.6	847	–2.4	966	0.1	961	1.0	840	0.4

Presence of a metabolite is validated by best mass spectral match on a scale 0–1000 (match) and smallest deviation of retention time index (Δ RI).

^aLow mass spectral match results from mixed mass spectra with a co-eluting compound (presence of compound was manually verified).

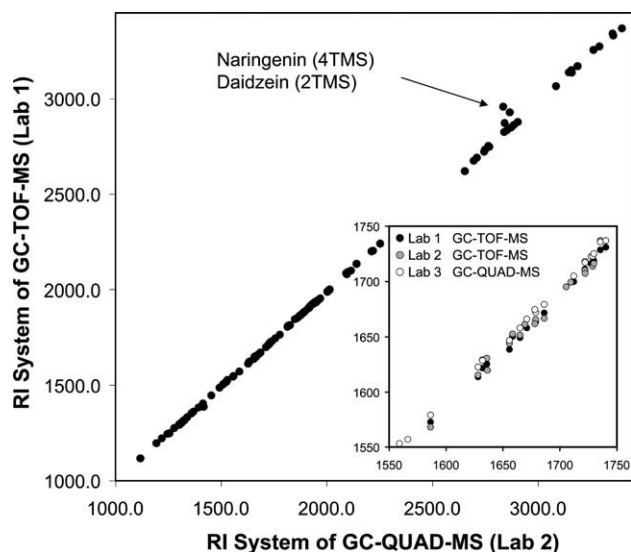


Fig. 2. Comparative analysis of retention time indices (RI), which were determined in parallel on two GC-MS technology platforms in different laboratories, demonstrates that knowledge of one RI system allows good prediction of RI in a second system. Retention time indices were determined on TOF and QUAD GC-MS systems. GC capillary columns were of identical build. Lab 2 operates TOF and QUAD GC-MS systems.

addition, typical metabolite classes, such as sugars, fatty acids or amino acids [6–9,15,16], mostly exhibited common positive or negative trends of deviation. Therefore, RI information obtained from one technology platform will allow good prediction of retention time indices, if reference compounds are already mapped on both systems. Use and implementation of RI systems for different GC column types is ongoing effort in our laboratories (data not shown) but RI prediction will require other methods than regression, because the elution-sequence of compounds is known to change.

5. Conclusions

The hypothesis and test cases described here present, for the first time, a comprehensive MSRI library database covering MSTs of GC-MS metabolite profiles from mammals, corynebacteria and major plant species. It includes in total more than 2000 fully evaluated mass spectral data sets obtained using two distinct technology platforms with 1089 non-redundant and 360 identified MSTs. The database is designed to be continuously extended by additional accessory information as it becomes available. We demonstrated the use of this MSRI library to screen biological samples for known compounds and showed the appliance of the non-supervised library for screening samples for known or recently identified mass spectra.

This library is constantly being updated with every new biological sample and application run in-house. Because even slight changes in GC-MS settings, such as carrier flow, temperature ramp, and dimension/make of capillary columns, induce shifts in retention behaviour of substances, GC-MS systems need to be recalibrated after each change. As there is currently no solution – other than recalibration – addressing the problem of RI shifts using different GC-MS machines we would like to offer to the biological and metabolite profiling commu-

nity to perform qualitative analysis of any biological sample using our currently running protocols.

In addition to offering this service to the community, we believe that the data presented here demonstrate three general applications of such libraries, which will help to advance the field. (i) The composition of still non-characterized biological samples, for example blood plasma, or microbial extracts (data not shown) can be screened for identified constituents, and tentative best matching compounds. (ii) Occurrence of identified metabolites can be analysed in a large range of biological samples, such as different plant organs or species. For this purpose, we provide libraries comprising samples from tomato, related wild type species, and other *Solanacea*, collections of different organs of *L. japonicus*, *A. thaliana*, and preparations from microbial species. (iii) Subsequent analysis of samples on two different GC-MS systems facilitates transfer of identifications made on the first system to the second. We present data on identifications, which were made in-parallel on QUAD GC-MS and GC-TOF-MS systems in different laboratories worldwide. This is therefore the first validation that metabolite profiling, when carried out with appropriate care, can yield comparable results between laboratories. Given the number of independent laboratories involved in this study, we believe that it offers similar reassurance as provided to the microarray community by the multi-laboratory Affimetrix microbial gene expression study. We are convinced that the effort described here will be useful on several levels. Not only will it meet a recently expressed demand within the metabolomics community [13,24], which was already apparent in earliest metabolomics applications in clinical diagnostics [25], but it will also aid laboratories entering the field of metabolomics.

Acknowledgement: We thank Jim Vale and co-workers at the IGER Research Farm at Bronydd Mawr for the kind provision of sheep blood samples.

References

- [1] Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and Schomburg, D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 32, Database issue: D431–433.
- [2] Mewes, H.W., Amid, C., Arnold, R., Frishman, D., Guldener, U., Mannhaupt, G., Munsterkotter, M., Pagel, P., Strack, N., Stumpflen, V., Warfsmann, J. and Ruepp, A. (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* 32, Database issue: D41–44.
- [3] Gollub, J., Ball, C.A., Binkley, G., Demeter, J., Finkelstein, D.B., Hebert, J.M., Hernandez-Boussard, T., Jin, H., Kaloper, M., Matese, J.C., Schroeder, M., Brown, P.O., Botstein, D. and Sherlock, G. (2003) The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res.* 31, 94–96.
- [4] Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, R., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U., von Mering, C., Roechert, B., Poux, S., Jung, E., Mersch, H., Kersey, P., Lappe, M., Li, Y.X., Zeng, R., Rana, D., Nikolski, M., Husi, H., Brun, C., Shanker, K., Grant, S.G.N., Sander, C., Bork, P., Zhu, W.M., Pandey, A., Brazma, A., Jacq, B., Vidal, M., Sherman, D., Legrain, P., Cesareni, G., Xenarios, L., Eisenberg, D., Steipe, B., Hogue, C. and Apweiler, R. (2004) The HUPO PSI's Molecular Interaction format – a community standard for the representation of protein interaction data. *Nat. Biotechnol.* 22, 177–183.
- [5] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acid Res.* 32, D277–D280.
- [6] Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N. and Willmitzer, L. (2000) Simultaneous analysis of metabolites in

- potato tuber by gas chromatography–mass spectrometry. *Plant J.* 23, 131–142.
- [7] Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A.R. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13, 11–29.
- [8] Gullberg, J., Jonsson, P., Nordstrom, A., Sjoström, M. and Moritz, T. (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal. Biochem.* 331, 283–295.
- [9] Strelkov, S., von Elstermann, M. and Schomburg, D. (2004) Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. *Biol. Chem.* 385, 853–861.
- [10] Fischer, E. and Sauer, U. (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC–MS. *Eur. J. Biochem.* 270, 880–891.
- [11] Sauer, U. (2004) High-throughput phenomics: experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.* 15, 58–63.
- [12] Kromer, J.O., Sorgenfrei, O., Klopprogge, K., Heinzle, E. and Wittmann, C. (2004) In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. *J. Bacteriol.* 186, 1769–1784.
- [13] Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S. and Sumner, L.W. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 9, 418–425.
- [14] Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R., Kopka, J., Lane, G.A., Lange, B.M., Liu, J.R., Mendes, P., Nikolau, B.J., Oliver, S.G., Paton, N.W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, J., Sumner, L.W., Wang, T., Walsh, S., Wurtele, E.S. and Kell, D.B. (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nat. Biotechnol.* 22, 1601–1606.
- [15] Fiehn, O., Kopka, J., Doermann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161.
- [16] Wagner, C., Sefkow, M. and Kopka, J. (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry* 62, 887–900.
- [17] Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Doermann, P., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R. and Steinhauser, D. (2005) GMD@CSB.DB: The Golm Metabolome Database, Bioinformatics advance access (published on December 21, 2004, doi:10.1093/bioinformatics/bti236).
- [18] Ausloos, P., Clifton, C.L., Lias, S.G., Mikaya, A.I., Stein, S.E., Tchekhovskoi, D.V., Sparkman, O.D., Zaikin, V. and Zhu, D. (1999) The critical evaluation of a comprehensive mass spectral library. *J. Am. Soc. Mass Spectrom.* 10, 287–299.
- [19] Stein, S.E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* 10, 770–781.
- [20] Müller, M. and Kersten, S. (2003) Nutrigenomics: goals and strategies. *Nat. Rev. Genet.* 4, 315–322.
- [21] Davis, C.D. and Milner, J. (2004) Frontiers in nutrigenomics, proteomics, metabolomics and cancer prevention. *Mutation Res.* 551, 51–64.
- [22] 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.
- [23] Kopka, J., Fernie, A.R., Weckwerth, W., Gibon, Y. and Stitt, M. (2004) Metabolite profiling in plant biology: platforms and destinations. *Genome Biol.* 5, 109–117.
- [24] Fernie, A.R., Trethewey, R.N., Krotzky, A.J. and Willmitzer, L. (2004) Innovation – metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell. Biol.* 5, 763–769.
- [25] Jellum, E. (1977) Profiling of human-body fluids in healthy and diseased states using gas-chromatography and mass-spectrometry, with special reference to organic-acids. *J. Chromatogr. B* 143, 427–462.