

Nonsupervised Construction and Application of Mass Spectral and Retention Time Index Libraries From Time-of-Flight Gas Chromatography–Mass Spectrometry Metabolite Profiles

Alexander Erban, Nicolas Schauer,
Alisdair R. Fernie, and Joachim Kopka

Summary

Gas chromatography–mass spectrometry (GC–MS) is routinely applied to the metabolite profiling of biological samples. Time-of-flight (TOF)–GC–MS metabolite profiling is based on highly reproducible electron impact ionization. Single chromatograms may comprise 200–1000 mass spectral components. The nature and composition of these mass spectral components depend on the choice of metabolite extraction, type of biological sample, and experimental condition. The components represent mass spectral tags (MSTs) of volatile metabolites or metabolite derivatives. Identification of MSTs is the major challenge in GC–MS metabolite profiling. We describe methods suitable for the automated construction of mass spectral and retention time index databases from large sets of TOF–GC–MS profiles. Application of these libraries for automated identification by pure reference compounds and classification of hitherto unidentified MSTs from biological sources is demonstrated.

Key Words: Metabolite profiling; electron impact ionization; time-of-flight GC–MS; mass spectral matching; retention time index; metabolite classification.

1. Introduction

One of the major challenges in gas chromatography–mass spectrometry (GC–MS)-based metabolite profiling is the identification of the multitude of hitherto

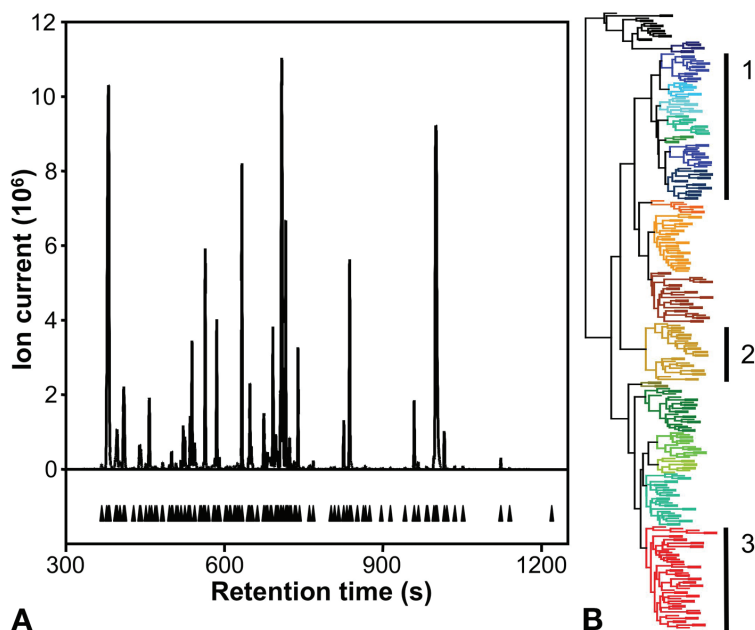


Fig. 1. Metabolite profile of an intercellular yeast extract (A). Tics below the chromatogram trace indicate positions of automated deconvolutions of mass spectra. Mass spectral components obtained from yeast metabolite profiles were clustered with mass spectra of pure reference metabolites (B). Major clusters represent (1) sugars, polyols, and polyhydroxy acids, (2) phosphorylated compounds, and (3) amino acids.

unidentified metabolic components from extracts of diverse biological samples (1,2). Automated deconvolution of single GC–MS chromatograms generates hundreds of mass spectral tags (MSTs) (Fig. 1). MSTs were previously defined as mass spectra of metabolites or metabolite derivatives (3,4), which can be unambiguously identified by mass abundance or fragment composition and chromatographic retention behavior. As a rule of thumb, less than 30–40% of the detected MSTs can currently be linked to known metabolites. Unidentified MSTs are not necessarily artifacts of the GC–MS profiling technology. These MSTs can be shown to represent metabolites by *in vivo* labeling of organisms with stable isotopes, for example, labeling of microbial cultures by U- ^{13}C -glucose (5) or feeding of $^{13}\text{CO}_2$ to photoautotrophic organisms. Thus, efforts to identify MSTs will be crucial for the further development and general applicability of GC–MS-based metabolite profiling (1–3).

Identification of MSTs is performed through two complementary approaches. The “top down” approach whereby metabolite identities are unravelled by tak-

ing, for example, a single MST of interest and establishing its structure through stepwise purification and complete structural elucidation. This approach is highly time consuming. “Top down” identification is only recommended if the biological function of the unknown MST is clearly established and if the importance of the hitherto unknown metabolite justifies the task. The second, i.e., “bottom up,” approach in which metabolites of interest to a particular researcher are analyzed by the purchase or synthesis of authentic standards is certainly less time demanding and, thus, appears to be more efficient. Identification is easily performed through standard addition experiments of pure reference compounds (6–9). Both mass spectral matching and cochromatography can routinely be established in different laboratories and can thus meet the general prerequisites of unambiguous chemical identification (3). In summary, metabolite identification can be repeated with all GC–MS equipment and is easy to cross-validate between many laboratories across the world.

We describe a largely automated method for the highly reproducible generation of MSTs and mass spectral/retention time index (MSRI) libraries. The method is designed to suppress artifacts of chemical derivatization by timed and automated in-line derivatization of metabolic extracts. Furthermore, the high degree of automation supports increased reproducibility of retention time behavior, as determined by Kovàts’ retention time indices (RI) (10). In parallel, mass spectral characteristics are quality controlled by in-built auto-tuning routines of the time-of-flight (TOF)–GC–MS system (11).

The availability of curated MSRI libraries as well as nonsupervised MSRI libraries, i.e., automated generation of MST compendia from well characterized and defined biological samples, facilitates identification of metabolites in diverse biological samples (3,12) and integrates use of commercially available mass spectral libraries (13,14), which lack retention time characteristics. The aim of this method description is to enable mass spectral library searches with single bait mass spectra of a reference substance that allow clear identification by mass spectral match and RI (Fig. 2). Moreover, the hit lists of these mass spectral searches are utilized to discover candidate component MSTs of highly similar chemical nature as compared with the bait and, thus, facilitate classification of as yet unidentified MSTs (Fig. 3).

2. Materials

2.1. Sampling and Metabolite Extraction

1. Methanol gradient grade for liquid chromatography (Merck, Darmstadt, Germany; cat. no. CAS 67-56-1).
2. Chloroform for liquid chromatography (Merck; cat. no. CAS 67-66-3).
3. Bidistilled water approx 0.055 $\mu\text{S}/\text{cm}$ (USF Deutschland GmbH, Ransbach-Baumbach, Germany; cat. no. USF 800).

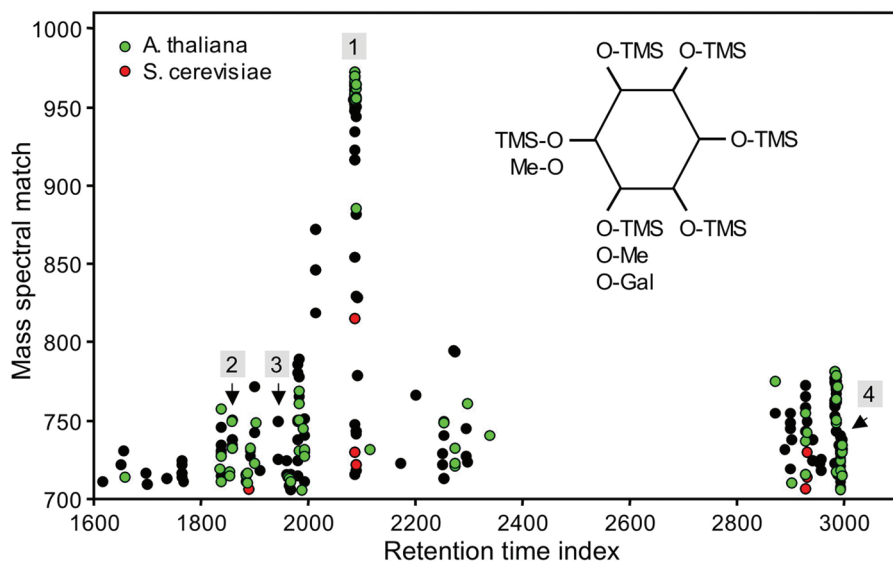


Fig. 2. Mass spectral hit list of myo-inositol (6TMS). Identified mass spectral tags were (1) myo-inositol (6TMS), (2) 3-*O*-methyl-D-chiro-inositol (5TMS), (3) 4-*O*-methyl-myoinositol (5TMS), and (4) α -D-galactopyranose-(1,3)-myo-inositol (9TMS). Occurrence in *Arabidopsis thaliana* and *Saccharomyces cerevisiae* is color coded. Mass spectral matching was performed without limits or constraints.

4. Ribitol (Sigma, Munich, Germany; cat. no. CAS 488-81-3).
5. DL-Alanine, 2,3,3,3-d₄ (Sigma; cat. no. CAS 53795-92-9).
6. D(-)-Isoascorbic acid (Sigma; cat. no. CAS 89-65-6).
7. Methyl nonadecanoate (Sigma; cat. no. CAS 1731-94-8).
8. 1.5-mL Safe-lock, tapered-bottom microvial (Eppendorf, Hamburg, Germany).
9. 2.0-mL Safe-lock, round-bottom microvial (Eppendorf).
10. Microcentrifuge 5417 (Eppendorf).
11. Oscillating ball mill MM200 (Retsch GmbH and Co. KG, Haan, Germany).
12. Teflon adaptor for 1.5- to 2.0-mL microvials (Retsch GmbH and Co. KG).
13. VA 5-mm steel balls (Th. Geyer Berlin GmbH, Berlin, Germany).
14. VR Maxi standalone vacuum concentrator with rotors R96-13 and R120-111 (Jouan Nordic, Allerød, Denmark).
15. HBP hold-back vacuum pump (Ilmvac GmbH, Ilmenau, Germany).
16. Polystat K6-1 cycling thermostat (P. Huber GmbH, Offenburg, Germany).
17. 15- and 50-mL plastic tubes with screw caps (Falcon™ Conical Centrifuge Tubes, BD Biosciences, San Jose, CA).
18. Orange silica gel (Carl Roth GmbH, Karlsruhe, Germany; cat. no. 77.1).
19. Argon 5.0 (Messer-Griesheim GmbH, Krefeld, Germany).

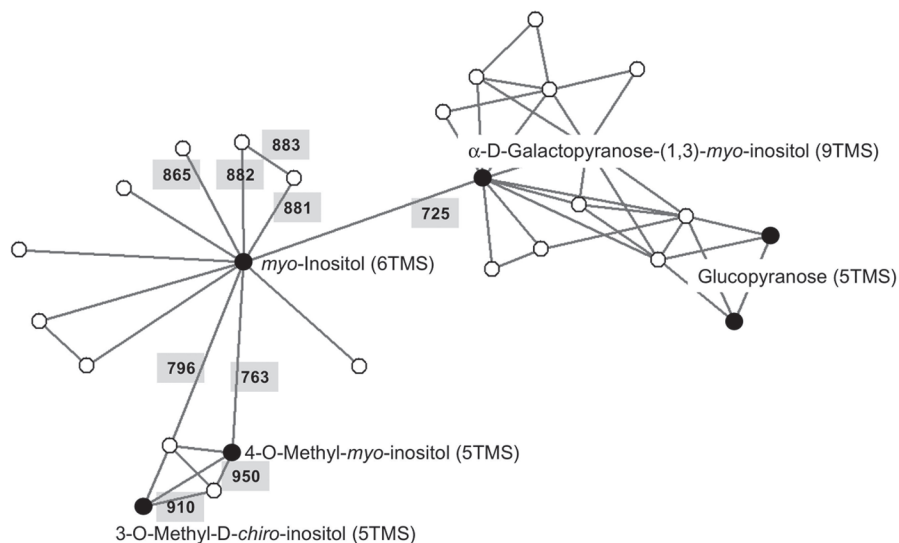


Fig. 3. Proximity map of a search for mass spectral similarity among identified and unidentified mass spectral tags from GC-MS profiles of biological sources. The search was initiated at myo-inositol (6TMS). Open circles represent hitherto unclassified mass spectra, and connecting edges represent best mass spectral match as partially indicated in shaded boxes. Mass spectral matching was performed in the mass range, m/z 85–600, with minimum abundance set to 50.

2.2. Chemical Derivatization

1. CTC Combi PAL autosampler and PAL cycle composer software v1.5.0 (CTC Analytics AG, Zwingen, Switzerland). The chosen configuration comprises an agitator–incubator oven, a 98-sample tray for 2.0-mL vials, a 32-sample tray for 10- to 20-mL vials, three 100-mL solvent reservoirs, i.e., a syringe wash station, and a liquid version 25- μ L syringe kit mounted to the robotic autosampler arm.
2. Methoxyamination reagent: methoxyamine hydrochloride (Sigma; cat. no. CAS 593-56-6) is dissolved at 20 mg/mL in pure pyridine (Merck; cat. no. CAS 110-86-1). This reagent is prepared immediately before analysis in 1-mL aliquots and loaded into the first reagent reservoir of the CTC Combi PAL autosampler (*see Note 1*).
3. Per-silylation reagent: 1-mL vials of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA; Macherey and Nagel, Düren, Germany; cat. no. CAS 24589-78-4) is loaded into the second reagent reservoir (*see Note 1*).
4. Solvents for syringe washes were *n*-hexane (Fisher-Scientific GmbH, Schwerte, Germany; cat. no. CAS 110-54-3) and ethylacetate (Merck; cat. no. CAS 141-78-6).
5. RI standard mixture: *n*-alkanes are dissolved in pyridine (Merck; cat. no. CAS 110-86-1) at a final concentration of 0.22 mg/mL each and loaded into the agitator–incubator oven of the CTC Combi PAL autosampler (*see Note 2*). The following

substances are combined: *n*-decane (RI 1000; cat. no. CAS 124-18-5), *n*-dodecane (RI 1200; cat. no. CAS 112-40-3), *n*-pentadecane (RI 1500; cat. no. CAS 629-62-9), *n*-octadecane (RI 1800; cat. no. CAS 593-45-3), *n*-nonadecane (RI 1900; cat. no. CAS 629-92-5), *n*-docosane (RI 2200; cat. no. CAS 629-97-0), *n*-octacosane (RI 2800; cat. no. CAS 630-02-4), *n*-dotriacontane (RI 3200; cat. no. CAS 544-85-4), and *n*-hexatriacontane (RI 3600; cat. no. CAS 630-06-8). All substances were obtained from Sigma.

6. 1.1 CTVG crimp-cap vial (Chromacol, Trumbull, CT).
7. R11-Sil-r/w magnetic crimp cap (CS-Chromatography Service GmbH, Langerwehe, Germany).
8. Adjustable 11-mm crimp-cap sealer (Supelco, Munich, Germany).

2.3. TOF-GC-MS

1. Pegasus III TOF mass spectrometer (LECO Instrumente GmbH, Mönchengladbach, Germany).
2. Agilent 6890N gas chromatograph, split/splitless injector with electronic pressure control up to 150 psi (Agilent, Böblingen, Germany).
3. Conical single taper split/less liner with glass wool (Agilent), deactivation reagent (DMDCS, Restek GmbH, Bad Homburg, Germany), toluene (Sigma; cat. no. CAS 108-88-3), methanol (Merck; cat. no. CAS 67-56-1) (*see Note 3*), 7-mL glass tubes (cat. no. 23 175 11 59) with screw caps (cat. no. 29 990 12 04) (Schott, Mainz, Germany).
4. Precolumn (Restek GmbH) or VF-5ms capillary column, 30-m length, 0.25-mm inner diameter, 0.25- μ m film thickness, and a 10-m EZ-guard precolumn (Varian Inc., Lake Forest, CA).
5. Helium 5.0 carrier gas (Air Liquide, Magdeburg, Germany).

2.4. GC-MS Data Processing

1. ChromaTOF chromatography processing and mass spectral deconvolution software, v1.00, driver 1.61 (LECO Instrumente GmbH).
2. Automated mass spectral deconvolution and identification system AMDIS (National Institute of Standards and Technology [NIST], Gaithersburg, MD).
3. NIST mass spectral search and comparison software v2.0 (NIST).
4. Microsoft Office Word 2003 (Microsoft Corporation), Excel 2003 (Microsoft Corporation), software package for exploratory data analysis and statistical modelling, S-Plus 2000 standard edition release 3 (Insightful, Berlin, Germany).

3. Methods

Metabolite turnover is extremely rapid as compared to mRNA or protein turnover. Analysis of metabolite composition and changes in pool sizes, therefore, requires fast and reproducible metabolic inactivation. Samples are best shock-frozen and kept below -60°C until extraction. Maintenance of metabolic inactivation during extraction and workup procedures is essential for a robust and repeatable representation of *in situ* metabolite composition (**15**). We describe

two exemplary protocols of metabolite extraction from plant material and liquid microbial cultures. We are fully aware that the choice of metabolic inactivation and extraction protocol may influence and indeed determine the scope of metabolites that can be monitored by subsequent metabolite profiling. Variations of extraction protocols may be pursued to broaden the spectrum of metabolites that are accessible to metabolite profiling (*see Note 4*) or to perform integrated analyses of metabolome, proteome, and transcriptome (**16**) (*see Chapter 5*).

The essence of metabolite profiling is discovery of novel marker metabolites and determination of relative changes of metabolite pool sizes in comparison to reference samples (**1,2**). This approach necessitates thorough control experiments, monitoring of GC–MS system performance, and check of laboratory contaminations, which may arise from solvent and reagent impurities or leakage of vial and septum material. For these reasons, nonsample control experiments are indispensable. All chemicals and containers need to be of the highest available purity. Please consider that autoclaved material, although sterile, may nevertheless be chemically contaminated.

3.1. Experimental Design and Preparation of Samples for Metabolite Profiling

Make sure to include a set of nontreated control samples in each experiment and analysis. For a large series of analyses prepare and store a large batch of reference material. Take an additional set of control samples from this batch of reference material for each subset of analyses. Results from this reference material allow the experimenter to control for day-to-day and week-to-week variability. Thus, discovery of marker metabolites and relative changes in metabolite levels can be distinguished from accidental contaminations or changes in instrumental sensitivity.

Provide at least 6 (better 8–16) replicate samples of each experimental condition. Perform replications at the level of individual plants or cell cultures rather than repeating assays of the same sample (**6–9,15**). Pooling of samples from a set of plants or cell cultures and repeated analyses of this pool is advised when sample size is small. Analysis of sample pools, however, is less informative with respect to the underlying variability inherent to the experiment and nature of biological samples.

3.1.1. Metabolic Inactivation and Extraction of Plant Material

1. Shock-freeze plant material in liquid nitrogen and keep below -60°C throughout processing. Use precooled 2.0-mL safe-lock microvials or wrap samples in precooled aluminium foil. Store either in liquid nitrogen or at -80°C until further processing. The amount of required sample may vary depending on species and plant organ. Always perform test analyses when analyzing previously unknown

samples or novel experimental conditions. The following protocol describes a typical analysis that is optimized for 60-mg fresh weight (± 5 –10%) of dicot leaves. Monocot leaves or root material may require more material (factor 2–4), whereas storage organs, flowers, or cold-stressed material may be performed with smaller amounts (factor 0.1–0.5). The optimum sample load is best determined by adjusting the major metabolic components to the upper detection limit of GC–MS, while still avoiding peak overload (*see Subheading 3.4.*).

2. The preparation of representative aliquots from large samples, greater than 125 mg (fresh weight), requires homogenization using a precooled mortar and pestle and subsequent generation of small aliquots of the desired amount of material. Keep samples in liquid nitrogen throughout the process. Avoid condensing ice and be careful not to spill the final powder by boiling liquid nitrogen. The powder may be stored in liquid nitrogen or in a -80°C freezer using screw-cap or safe-lock vials. Be careful to evaporate residual liquid nitrogen at -80°C before caps are sealed.
3. Small samples, 5–125 mg (fresh weight), are homogenized using steel balls that fit into 2.0-mL round-bottom microvials. Sets of 5–10 microvials are mounted onto an oscillating ball mill and exposed to two 3-min bursts at 15/s frequency. Steel balls, microvials, and the mounting adaptor need to be precooled in liquid nitrogen. Homogenized samples are extracted within microvials without removal of the steel balls. Sample weight is best determined after shock-freezing. Differential weighing of frozen powder or nonhomogenized material can be performed in cooled 2.0-mL microvials before adding steel balls. Avoid high air humidity and use dry ice for cooling to obtain stable zero point calibration.
4. Take frozen 2.0-mL microvials with homogenized samples from the freezer and add 360 μL of extraction mixture (*see Notes 5 and 6*). The extraction mixture needs to be precooled to -20°C and is best degassed by bubbling argon or nitrogen gas. Use an oil filter between gas supply and high-performance liquid chromatography bubbling device. Shake samples thoroughly using a vortex mixer and keep on ice until all samples are processed.
5. Shake all samples simultaneously for 15 min at 70°C and subsequently cool to room temperature. Solvent evaporation may generate excess pressure. Vent microvials after a 1-min incubation at 70°C and reclose vials thoroughly.
6. Add 200 μL CHCl_3 , shake thoroughly using a vortex mixer, and incubate at 37°C .
7. Add 400 μL H_2O to induce phase separation, shake thoroughly using a vortex mixer, and separate liquid and solid phases in a microcentrifuge for 5 min at approx 22,000g. Addition of H_2O may be omitted for a joined analysis of the lipophilic and polar metabolic complement of the sample.
8. Take a 10- μL aliquot of the upper phase, which contains the polar metabolic complement of the sample, and transfer into a crimp cap-tapered glass vial suitable for GC–MS analyses (*see Note 7*). In the following, we describe automated analysis of 10 μL of the polar or a combined liquid extract. In case of manual processing, 1.5–2.0 mL safe-lock microvials may be used to dry, transport, and store metabolic extracts (*see Subheading 3.2.*). For analysis of the lipophilic

metabolic complement, take a 100- μ L aliquot of the lower liquid phase and process manually. The analysis of the lipophilic complement induces strong chromatographic memory effects and is not recommended for high-throughput split or splitless GC–MS injection.

9. Dry 10- μ L samples in a vacuum concentrator for a minimum of 2 h at room temperature or lyophilize larger sample volumes overnight.

3.1.2. Metabolic Inactivation and Extraction of Yeast Liquid Cultures

The major challenge in metabolite profiling of microbial cultures is the separation of intracellular metabolites from secreted metabolites and residual components of liquid growth media, the so-called footprint, while rapidly inactivating metabolism during sampling. Typically, cell suspensions are rapidly sprayed into precooled polar organic solvents, such as methanol, which dilute the media and shock-freeze the cells (17–19). We recommend growth media spiked with nonmetabolized low molecular weight compounds for the control of residual liquid medium, which is unavoidably trapped in the cellular periplasm. In the case of yeast we successfully used lactose, which cannot be utilized by yeast, at 1–10% (w/w) concentration of the major carbon source in the growth medium. We furthermore suggest use of synthetic-defined growth media (SD) instead of complex media. Complex media contain numerous compounds in high concentrations. These substances will obscure intracellular metabolites even in cases of only small medium contaminations.

1. Prepare 5-mL yeast batch cultures in SD medium and time the sampling to the late logarithmic or to the stationary growth phase ($OD_{595} = 1.8$). Follow general recommendations for yeast growth (17–19). Make sure to prepare noninoculated samples for nonsample control of the experiments. Avoid unwanted chemical contaminations of the liquid cultures. Sterilized glassware and media are devoid of microbial contaminations but might nevertheless have received chemical deposits from the autoclave. Media components may decompose while exposed to high temperatures.
2. Sample the complete culture at routine growth temperature, 28°C, by rapid decanting or use temperature equilibrated disposable pipet tips for sampling 5-mL aliquots from larger batches. Avoid slow temperature changes before sampling. Continue to agitate batch cultures until sampling. Thus, sedimentation of cells and changes in mechanical stress are circumvented.
3. Rapidly mix 5 mL medium with 20 mL precooled 60% methanol, methanol:water, 6:4 (v/v). 60% methanol is best prepared as a large batch and partitioned into 50-mL screw-cap plastic tubes, which are kept before and after sampling in a methanol/dry ice bath at approx –60°C.
4. Spin down cells no longer than 5 min at approx 3200g in a temperature-controlled centrifuge preset to –20°C.
5. Immediately after centrifugation, collect plastic tubes into the methanol/dry ice bath. Decant supernatant cautiously and perform an optional gentle rinse with a

small volume of precooled 60% methanol. During temperature adjustment the supernatant might get slightly turbid but should not freeze solid.

The following steps can be downscaled according to the initial concentration of cells in suspension culture as determined by OD_{595} of diluted samples. The following volumes are as required for a 5-mL culture of $OD_{595} = 1.8$.

6. Add 374 μL extraction mixture for yeast intercellular metabolites immediately (*see Note 8*). The extraction mixture needs to be precooled to -20°C and is best degassed (*see steps 4 and 5*). At this step the cells should easily resuspend. If cells form a semi-solid viscous pellet, the temperature control was inadequate for metabolite profiles and needs to be optimized. Critical steps are centrifugation and time between decanting of the supernatant and resuspension into the extraction mixture. Slightly viscous yeast pellets may be resuspended in small droplets of ice-cold water prior to adding the extraction mixture. Metabolite profiling of these samples is not recommended.
7. Transfer resuspended samples from 50-mL plastic tubes into 7-mL screw-cap glass tubes for simultaneous extraction, 15 min at 70°C . Shake glass tubes intermittently and depressurize at least once. Allow to cool for 5 min at room temperature.
8. Add 188 μL CHCl_3 and extract 10 min at 30°C with intermittent vigorous shaking using a vortex mixer.
9. Add 75 μL of bidistilled H_2O , spin down cellular debris, and transfer a 10- μL aliquot of the combined polar and lipophilic extract into a crimp cap-tapered glass vial suitable for GC-MS analyses. Phase separation into a polar and lipophilic metabolic complement may be induced by adding 400 μL H_2O prior to centrifugation. Subsequent steps are as previously described (*see Subheading 3.1.1.*).

3.2. Storage and Transport of Metabolite Extracts

Metabolite extracts are best stored at low temperatures and under nonoxidizing conditions. If possible, long periods of storage and transport should be avoided. Samples can be transported and stored for up to 4 wk. Longer periods have not been tested.

1. After drying samples in a vacuum concentrator or lyophilization, flush the vacuum system with an inert gas, such as argon or nitrogen, instead of ambient air before removing samples.
2. Seal GC vials under inert gas using magnetic crimp caps and an adjustable crimp-cap sealer. Seal vials in plastic bags with silica gel. Combine the full number of vials comprising one experiment in single bags.
3. Transport sealed bags for short periods at room temperature otherwise on dry ice and store at -20 or -80°C .
4. Allow temperature equilibration at room temperature before opening bags for further analysis.

3.3. TOF-GC–MS Metabolite Profiling

Profiling of metabolite extracts involves a two-step chemical derivatization, which (1) substitutes carbonyl moieties through methoxyamination and (2) comprises a per-silylation prior to the GC–MS analysis of the reaction products (6–9). Samples are injected while dissolved in silylation reagent. Major sources of analytical variability are the imprecise dispensing of reagent volumes and the variable timing of the per-silylation reaction. In typical experiments, 50–100 samples are processed. Chemical derivatization was hitherto performed simultaneously on a batch of samples prior to injection. Thus the exposure time to the silylation reagent of the first and last sample within a batch differed considerably, i.e., 50–100 h in setups of 60 min per single GC–MS run. As a result, instable derivatives decomposed, side products of silylation reagents accumulated, and slow evaporation caused notable sample concentration. An optimization of the chemical reaction and GC–MS analysis was, therefore, in high demand.

We employ a CTC Combi PAL with a single syringe autosampler for automated and timed in-line derivatization, vial transport, and injection for GC–MS analysis. Vials are transported from the vial tray to positions within the agitator–incubator oven and finally back to the injection position by means of magnetic crimp caps. In short, in-line chemical derivatization requires samples to be dried within GC glass vials and sealed under nitrogen or argon. Each sample is processed in four equal time intervals of 45 min each. The first two intervals are assigned to methoxyamination (90 min), the third to per-silylation (45 min), and the fourth to a single slow or alternately two fast GC–MS runs per sample (total time <45 min). A typical TOF-GC–MS profile of a preparation of intracellular yeast metabolites is shown in **Fig. 1**.

3.3.1. In-Line Chemical Derivatization

1. The following instructions require 10 μL of metabolic extracts to be dried in 1.1 CTVG crimp-cap vials. The sealed vials are positioned on the sample tray and kept at ambient temperature (*see Note 9*).
2. Methoxyamination: the first vial is moved to position 1 of the agitator–incubator oven, which is set to constant 40°C. A 10- μL volume of methoxyamination reagent is dispensed into the vial. The vial is then agitated twice for 45 min.
3. Per-silylation: after 90 min, agitation is interrupted by dispensing 17.5 μL per-silylation reagent. Then 2.5 μL of a retention time standard mixture are added. Agitation is resumed for an additional interval of 45 min at 40°C.
4. At the end of the last interval the GC vial is moved back to the initial position on the sample tray and 1 μL is injected for GC–MS analysis (*see Note 10*). Processed vials are kept on the sample tray until discarded.

5. For automated high-throughput analysis, samples are processed in parallel with a time lag of 45 min each. Four positions of the agitator–incubator oven are used, three for derivatization of samples and one to store the retention time index standard mixture of *n*-alkanes (*see Note 2*). The most recent sample in the process is always subject to the first methoxyamination interval. Prior samples are in the second methoxyamination period, the per-silylation interval, or in the process of GC–MS analysis, respectively.
6. Syringe washes are performed between all dispensing procedures (*see Note 11*).
7. Automation using the Combi PAL autosampler can be performed with three basic programming parts. The first part primes the in-line derivatization process and ends with injecting the first sample, while the following two samples are already under derivatization. The second part comprises three methods that allow an “endless” cycle, each cycle ending with an injection. The final programming part contains methods that end in-line derivatization by processing the last samples of an analysis series and then safely shuts down the system.

3.3.2. TOF-GC–MS

1. Injection parameters: injection of a 1- μ L sample is performed at 230°C in splitless mode with helium carrier gas flow set to 0.6 mL/min. Purge time is 1 min at 20 mL/min flow. The flow rate is kept constant with electronic pressure control enabled. Optionally and especially recommended in cases of high metabolite concentrations, injection is performed in split mode with the split ratio adjusted 1:25. As a rule of thumb, split injection may be prone to discrimination of high-boiling metabolic components, whereas splitless injection may, in rare cases, result in peak shape artifacts for low-boiling components. These artifacts occur in few chromatograms and result in different degrees of peak splitting and shoulder formation. For suppression of this peak shape artifact either inject at decreased flow or apply a 2-min pressure pulse at 110 psi during injection. However, a robust suppression of this artifact for all biological samples can currently not be recommended.
2. Chromatography parameters: chromatography is performed using a 30-m RTX-5Sil MS capillary column with an integrated guard column. The temperature program starts in isothermal mode set to 1 min at 70°C. The isothermal step is followed by a 9°C/min ramp to 350°C. The final temperature is kept constant for 5 min. Cooling is performed as fast as instrument specifications allow. The transfer line temperature is set to 250°C and matches ion source conditions.
3. Mass spectrometer parameters: the ion source is set to maximum instrument specifications, 250°C. High-boiling metabolic components exhibit increased peak tailing at lower temperature settings. The recorded mass range is $m/z = 70$ –600 at 20 scans/s. Mass spectrometric solvent delay with filaments turned off is 6.6–7.5 min, the remaining chromatography is fully monitored with omission of cool down periods. Manual mass defect is set to 0, filament bias current is –70 V, and detector voltage is approx 1700–1850 V depending on detector age. The instrument tune is automated and performed without EPA tune compliance.

3.4. Automated Deconvolution of Mass Spectra

Automated deconvolution of MSTs from GC–MS metabolite profiles is crucial for increased accuracy of metabolite identification and detection (**Fig. 1**). Deconvolution is the process of locating MSTs, also called mass spectral components, in GC–MS chromatograms and the subsequent automated purification of the mass spectral scans at peak apex from electronic and chemical background noise and cross-contaminating fragments of coeluting compounds. Both the ChromaTOF software of LECO TOF–GC–MS systems and the technology platform-independent automated mass spectral deconvolution and identification system, AMDIS, may be used to this purpose (**3,13–14**). When using AMDIS, files are best exported in CDF file format after baseline correction within the ChromaTOF software. Large TOF–GC–MS files, such as those with fast scanning acquisition rates, may be impossible to load into AMDIS using standard desktop computers. Here, we describe the use of the ChromaTOF software for automated deconvolution and construction of MSRI libraries. MSRI libraries may contain either manually curated and selected entries of identified compounds or have the purpose to provide full automatically generated collections of mass spectra from single or multiple TOF–GC–MS profiles. This process we would like to term nonsupervised construction.

1. Chromatograms are processed by ChromaTOF software with activated baseline tracking and offset set to “just above noise,” smoothing and peak width are set to 20 and 6, respectively. The signal-to-noise threshold is set to minimum 2.0 and the number of deconvolutions is unlimited.
2. RI are generated for each individual chromatogram in two steps: first a mass spectral library search is conducted to identify all expected *n*-alkanes in each chromatogram. Then retention times of the *n*-alkanes are used for chromatogram-specific RI calculation. The mass spectral library search for *n*-alkane identification is restricted to the mass range $m/z = 80\text{--}600$ and threshold signal to noise set to 20. Further criteria for identification are expected peak height and area in total ion chromatography (TIC) mode, as well as occurrence of respective molecular ions for each *n*-alkane. The retention times of the expected and verified *n*-alkanes are transferred into a chromatogram specific retention index method and the chromatogram subsequently processed with the same settings. Overloaded peaks must be avoided or excluded in order to maintain high RI accuracy.
3. Chromatogram processing results are exported to text files. All available information for each deconvoluted peak or MST is exported including auxiliary information, such as retention time index, retention time, unique mass, total signal to noise, and full mass spectrum in absolute intensity format.
4. These text files can be imported and modified in Microsoft Excel and Word. More efficient is a customized automated programmed conversion into the MSP format for import into NIST02 and AMDIS software, which needs to add RI information for the generation of MSRI libraries. During this process auxiliary

information, such as user comments, can be tagged as synonyms. MSTs can be removed or selected by signal to noise, peak purity, peak width, or RI thresholds. Thus, data can be specifically selected for import into NIST02 software and information can be added. A typical example of an identified mass spectrum is shown in the following:

Name: EITMS_163001-101_METB_1627.14_L-Glutamic acid (3TMS)

Synon: SOURCE_CHROMATOGRAM:1185EK12_1627.1

Synon: NAME:L-Glutamic acid (3TMS)

Synon: MATCH:[834; L-Glutamic acid (3TMS)]

Synon: MPIMP-ID:163001-101-1

Synon: QM:246|363|128|348|156

Synon: ROLE:METB

Synon: METABOLITE:DL-Glutamic acid

Synon: KEGG:C00025|C00302|C00217

Synon: TECHNOLOGY:GC-TOF-MS (EITMS)|GC [M1]

Synon: RI:1627.1

Synon: RT:10.253 min

Synon: SP:Standard| Signal G-1251

Synon: DATE:2001.06.01

Comments: Kopka J, Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, D-14476 Golm, Germany

Formula: C14H33NO4Si3

MW: 363

CAS no.: 15985-07-6

DB no.: 799

Num Peaks: 151

70 11; 71 6; 72 38; 73 999; 74 102;

75 454; 76 33; 77 79; 78 6; 79 7;

80 2; 81 1; 82 5; 83 6; 84 164; ...

5. Chromatogram processing results can also be exported directly from the peak table of the ChromaTOF software to NIST02 without the need of programming skills. Deconvoluted MSTs can be either added to NIST02 user libraries or exported as MSP files. Customization of library entries within ChromaTOF software before export is highly restricted. However, NIST02 offers a full toolbox for editing mass spectral information. Thus, mass spectral libraries of manually selected, identified, and curated MSTs can be easily generated and maintained with the tools and options provided by NIST02 and ChromaTOF software.
6. Examples of annotated MSRI libraries comprising identified compounds as well as unidentified MSTs and MSRI libraries, which were fully generated in the non-supervised mode, may be found at CSB.DB (<http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html>).

3.5. Comparison and Classification of Mass Spectra

The NIST02 mass spectral search and comparison software represents the most widely accepted standard tool for analysis of mass spectra generated by GC-MS systems (13,14). Systems' manufacturers optimize automated MS tun-

ing with the aim to produce comparable mass spectra. NIST02 is mature in automation, algorithm, as well as user friendliness. However, the great challenge of identifying or at least classifying all hitherto unidentified metabolic components from GC–MS profiles of biological samples requires additional features that are not provided by NIST02. One of the most useful additional features for mass spectral comparisons is the integration of retention time index information into mass spectral comparisons. Only information on chromatographic retention will allow unambiguous identification of those stereo- and conformational isomers, which cannot be distinguished by mass spectral criteria alone (3). In addition, mass spectral classification needs to be reconsidered for those MSTs that cannot immediately be linked to a known metabolite. Here, we demonstrate first attempts to systematically deal with the challenge of identifying multiple unknown mass spectral tags from TOF–GC–MS profiles. Our present analyses are all performed using a single technology platform and a set of chromatograms that were produced on identical GC capillary columns. Transfer of our results to other technology platforms appears to be feasible but still awaits thorough investigation.

3.5.1. Clustering

Mass spectra can be directly clustered using hierarchical clustering of Euclidian distance or any other algorithm of commercially or publicly available software packages for statistical analysis. An alternative approach is clustering based on the generally accepted matching value generated by NIST02 mass spectral comparison software instead of Euclidian or other statistical distances. For “nonstandard” mass spectral distance measures and queries that incorporate RI information refer to our web pages, <http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html> (12). For the purpose of clustering a full matrix of pairwise similarity, measures of all MSTs and identified mass spectra needs to be defined through automated comparison and data export using NIST (3). We performed a combined analysis of identified and all MSTs that occur in yeast metabolite profiles (Fig. 1A). Clustering was performed as described using the S-Plus 2000 standard edition statistical software package. Clustering demonstrated the presence of major metabolite classes in TOF–GC–MS profiles, such as carbohydrates, amino acids, and organic phosphates (Fig. 1B). The mass spectrum of myo-inositol (6TMS), which we subsequently use as a test case, classifies to the sugar cluster. Most of the hitherto nonclassified MSTs sorted into clusters of identified metabolites. Thus, simple hierarchical clustering provides means to link unidentified MSTs to major metabolite classes. Some major clusters formed clear subdivisions. For example the carbohydrate cluster had disaccharide, monosaccharide, noncyclic polyol, and polyhydroxy carbonic acid branches. In total, up to 18 clear minor mass spectral clusters were found.

However, clustering might lack resolution within the terminal branches of hierarchical trees.

3.5.2. Visualization of MSRI Search Results

For the resolution of mass spectral similarity at the level of single mass spectra, the NIST02 hit lists are unsurpassed, but are lacking in visualization. The additional RI information is best shown in bi-plots with axes of RI and mass spectral match (**Fig. 2**). These plots easily accommodate auxiliary information, for example, on occurrence of MSTs in different sample types and frequency of occurrence in cases of redundant mass spectral libraries, such as nonsupervised MSRI libraries from GC–MS profiles. These visualizations allow discovery of MSTs that exhibit similarity to the bait mass spectrum. In addition, structural similarities of identified mass spectra become apparent as mass spectral similarity can be accessed. In our test case myo-inositol (6TMS) had among the top scoring identified mass spectra, methyl-substituted inositols, ononitol (5TMS) (4-*O*-methyl-myo-inositol), pinitol (5TMS) (3-*O*-methyl-D-chiro-inositol), and an inositol conjugate, galactinol (9TMS) (α -D-galacto-pyranose-[1,3]-myo-inositol). Conformational isomers of myo-inositol, such as chiro- or scyllo-inositol, rank highest but are not yet included in this and the subsequent analysis.

3.5.3. Generation of Mass Spectral Proximity Maps

Hit lists of single MSTs present good means of discovery of best matching mass spectra but do not convey an overview of similarities between many MSTs. For this purpose proximity maps are best suited (**Fig. 3**). Proximity maps visualize the journey through the “space” of mass spectral matches present within a MSRI library. The process of generating a proximity map can be manually performed by starting a mass spectral search with a mass spectrum of interest, such as myo-inositol (6TMS). The aim of this process is to discover groups of related compounds based on mass spectral similarity. The initial hit list will contain redundant mass spectra of myo-inositol (6TMS) and one best hit, which as judged by RI or already known identity, represents a different compound. We travel to this compound along the best match (865 in **Fig. 3**) and will not use this connection in the same direction again throughout the remaining journey. Instead, we perform a mass spectral search with the found best hit. In our test case, the best match of this second search was myo-inositol (6TMS). Thus, we return to myo-inositol (6TMS) and close the connection in the reverse direction as well. We then continue with the next best match of myo-inositol (6TMS) (882 in **Fig. 3**). The proximity map is subsequently generated using the same rules. The journey can be terminated after a limited number of steps, a number of visited mass spectra, or at a threshold match value.

Visualization of a proximity map can be performed using network visualization tools such as Pajek software (20). The resulting map clearly shows that myo-inositol (6TMS) has a set of 11 directly linked MSTs in our present MSRI library within a similarity range of 725 to 865. Among those we found a set of four MSTs with high “internal” similarity (910–950), which represent two methyl-substituted inositols and two putative still unidentified other methyl-inositols. Furthermore, we found an inositol conjugate, galactinol (9TMS), and a group of highly similar (match values not shown) MSTs, which form connections to glucopyranoses that are highly similar in structure to the second conjugation partner, i.e., galactopyranose, of galactinol (9TMS).

4. Notes

1. Reagents are stored in 1-mL crimp-cap sample glass vials. These vials contain excess reagent but are replaced after 24 h in order to avoid aging and accumulation of contaminations.
2. The retention time index standard mixture contains high molecular weight *n*-alkanes, which tend to precipitate at low ambient temperature. The *n*-alkane mixture is best prepared at elevated temperature and during use is kept at 40°C within the heated agitator.
3. Deactivation of the glass insert liners reduces the number of cleaning cycles, which are required after liner exchange and increases column lifetime. For glass liner deactivation, dissolve 20 mL of DMDCS in 400 mL toluene and treat liners for 15 min in this solution. Then rinse twice with toluene and, finally, keep liners 15 min in methanol and rinse clean with methanol. Liners are dried, heated, and stored under inert gas and in sealed-glass tubes.
4. We describe the analysis of polar methanol and chloroform-soluble metabolites without and in combination with the lipid metabolite complement. Major additional variants are selective enrichment of acidic or basic compounds, permutations of temperature and extraction time for improved coverage of labile compounds, and application of other solvents for selective extraction. Descriptions of alternate extraction protocols may be found elsewhere within this book.
5. The internal standard premixture for the analysis of polar compounds contains ribitol, 2,3,3,3-*d*₄-DL-alanine and D(-)-isoascorbic acid. Each component is prepared separately at 10 mg/mL in bidistilled water except for ribitol, which is dissolved in methanol. These stock solutions are combined into 50 mL bidistilled water and, thus, diluted to 0.02, 0.10, and 0.05 mg/mL final concentration, respectively. Diluted stocks can be stored at -20°C for a limited time. The internal standard solution for the analysis of the lipophilic metabolic complement needs to be freshly prepared and contains 2 mg/mL nonadecanoic acid methyl ester in chloroform. The internal standard premixtures can be extended to contain any set of stable isotope-labeled or synthetic internal standards.
6. The extraction mixture for plant material contains 300 parts methanol, 30 parts internal standard premixture for the polar metabolic complement (*see* **Note 5**),

and 30 parts of the internal standard premixture for the lipophilic metabolic complement.

7. Back-up samples for in-line or manual derivatization can easily be generated by preparing additional aliquots from the surplus extracts and subsequent vacuum concentration. Note that rotors R96-13 and R120-111 require customized adaptors to accommodate tapered GC vials. Disposable 10-mL pipet tips, which are cut down to fit, may serve the same purpose.
8. The extraction mixture for yeast intercellular metabolites contains 350 parts methanol, 12 part internal standard premixture for the polar metabolic complement (*see Note 5*), and 12 parts internal standard premixture for the lipophilic metabolic complement.
9. The reagent volumes of the in-line derivatization steps are adjusted to 10- μ L sample volume. Increased sample volumes may not be fully redissolved in the 10- μ L volume of methoxyamination reagent and result in nonmethoxyaminated but subsequently per-silylated side products, such as silylated hexopyranoses. The source of these side products is residual dried extract that sticks to the walls of the GC vials. These dried residues are not accessible through high-intensity shaking by the CTC agitator-incubator oven, but do not present a problem during manual agitation. For automated processing of extract, the aliquot volumes must not exceed 10 μ L and need to be deposited at the bottom of the vial before vacuum centrifugation.
10. For continuous operation the GC-MS program needs to last less than 45 min. It is essential to either operate the GC-MS system under constant ambient temperature or check that increased ambient temperature owing to seasonal changes does not unexpectedly prolong the GC cycle time resulting in extended cooling times.
11. For the complete process of in-line derivatization a single syringe is used. This setup puts high demands on syringe cleanliness and mechanical performance. We mount a 25- μ L syringe for best mechanical robustness of plunger and needle. Reagent and sample cross-contaminations may occur with inadequate wash protocols. Major contaminants from microbial and plant extracts are disaccharides, such as sucrose and trehalose, or lipids and chlorophyll. When permanently present at high concentrations, these compounds are best removed by sequential treatment with polar and apolar solvents. The type of syringe cleaning cycle is best adjusted to the subsequent syringe task. We use hexane immediately before transferring MSTFA reagent and discard each first draw from the MSTFA reagent reservoirs taking care not to contaminate the reagents. Syringes are cleaned by maximum volume draws from the ethylacetate and *n*-hexane reservoirs.

Acknowledgments

The authors would like to thank Professor Lothar Wilmitzer and Dr. Oliver Fiehn, Max-Planck Institute of Molecular Plant Physiology (MPI-MP), Potsdam, Germany, and Professor Le Tran Binh, Institute of Biotechnology (IBT), Hanoi, Vietnam, for valuable advice, encouragement, and discussions. This work was

supported by the Max-Planck society and the Bundesministerium für Bildung und Forschung (BMBF), grant PTJ-BIO/0312854.

References

1. Bino, R. J. Hall, R. D. Fiehn, O., et al. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **9**, 418–425.
2. Fernie, A. R., Trethewey, R. N., Krotzky, A. J., and Willmitzer, L. (2004) Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.* **5**, 763–769.
3. 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. *Phytochem.* **62**, 887–900.
4. Colebatch, G., Desbrosses, G., Ott, T., et al. (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* **39**, 487–512.
5. Birkemeyer, C., Luedemann, A., Wagner, C., Erban, A., and Kopka, J. (2005) Metabolome analysis: the potential of *in vivo* labeling with stable isotopes for metabolite profiling. *Trends Biotechnol.* **23**, 28–33.
6. 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.
7. 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.
8. Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.
9. Roessner, U., Luedemann, A., Brust, D., et al. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11–29.
10. Kovàts, E. S. (1958) Gas-chromatographische charakterisierung organischer verbindungen: Teil 1. Retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. *Helv. Chim. Acta* **41**, 1915–1932.
11. van Deursen, M. M., Beens, J., Janssen, H. -G., Leclercq, P. A., and Cramers, C. A. (2000) Evaluation of time-of-flight mass spectrometric detection for fast gas chromatography. *J. Chromatogr. A* **878**, 205–213.
12. Kopka, J., Schauer, N., Krueger, S., et al. (2005) GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics* **21**, 1635–1638.
13. Ausloos, P., Clifton, C. L., Lias, S. G., et al. (1999) The critical evaluation of a comprehensive mass spectral library. *J. Am. Soc. Mass Spectrom.* **10**, 287–299.
14. 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.

15. 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.
16. Weckwerth, W., Wenzel, K., and Fiehn, O. (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4**, 78–83.
17. De Koning, W. and van Dam, K. (1992) A method for the determination of changes in glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal. Biochem.* **204**, 118–123.
18. Gonzalez, B., Francois, J., and Renaud, M. (1997) A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* **13**, 1347–1355.
19. 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.
20. Batagelj, V. and Mrvar, A. (1998) Pajek: program for large network analysis. *Connections* **21**, 47–57.

Metabolomics

Methods and Protocols

Weckwerth, W. (Ed.)

2007, 312 p. 79 illus., 8 illus. in color., Hardcover

ISBN: 978-1-58829-561-3

A product of Humana Press