analytical chemistry

Article

pubs.acs.org/ac

CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets

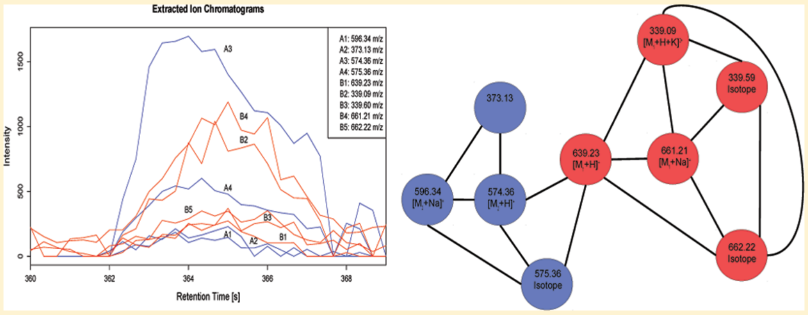
Carsten Kuhl,*[,†] Ralf Tautenhahn,[‡] Christoph Böttcher,[†] Tony R. Larson,[§] and Steffen Neumann*[,†]

[†]Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany

[‡]Department of Chemistry and Molecular Biology, Center for Metabolomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

[§]Centre for Novel Agricultural Products, Department of Biology, University of York, United Kingdom

S *Supporting Information*



ABSTRACT: Liquid chromatography coupled to mass spectrometry is routinely used for metabolomics experiments. In contrast to the fairly routine and automated data acquisition steps, subsequent compound annotation and identification require extensive manual analysis and thus form a major bottleneck in data interpretation. Here we present CAMERA, a Bioconductor package integrating algorithms to extract compound spectra, annotate isotope and adduct peaks, and propose the accurate compound mass even in highly complex data. To evaluate the algorithms, we compared the annotation of CAMERA against a manually defined annotation for a mixture of known compounds spiked into a complex matrix at different concentrations. CAMERA successfully extracted accurate masses for 89.7% and 90.3% of the annotatable compounds in positive and negative ion modes, respectively. Furthermore, we present a novel annotation approach that combines spectral information of data acquired in opposite ion modes to further improve the annotation rate. We demonstrate the utility of CAMERA in two different, easily adoptable plant metabolomics experiments, where the application of CAMERA drastically reduced the amount of manual analysis.

Mass spectrometry (MS) is one of the dominant analysis methods for metabolomics experiments. In metabolite profiling studies, a large number of complex samples are analyzed. Typically, samples are separated prior to ionization and MS-based detection, mostly chromatographically either by gas chromatography (GC) or liquid chromatography (LC). An overview of techniques and applications was given by Dunn.[1] Depending on the sample preparation method and the analyzed organism, samples contain anywhere between dozens to thousands of compounds, e.g., the estimated number of metabolites in *Escherichia coli*[2] is just above 1000, in human serum[3] above 4000, and 5000 to 25000 for higher plants.[4] The coverage within an experiment is much lower due to analytical limitations.

The typical metabolomics data processing pipeline first performs a feature detection step. The term feature describes a two-dimensional bounded signal: a chromatographic peak (retention time) and a mass spectral peak (m/z). Several software packages exist for feature detection, for example, the closed-source but freely available MetAlign,[5] or frameworks with open-source licenses, such as OpenMS,[6] MZmine,[7] and XCMS.[8] Other packages, some of them specific for LC/MS-based proteomics, have been reviewed elsewhere.[9]

Upon ionization, an individual chemical compound gives rise to one or more ion species, which can be observed in the same mass spectrum. Those ion species include isotopologue ions, fragment ions, and, in particular for electrospray ionization (ESI), adduct and cluster ions. A summary can be found in Keller et al.[10]

Received: September 23, 2011
Accepted: November 23, 2011
Published: November 23, 2011

dx.doi.org/10.1021/ac202450g | *Anal. Chem.* 2012, 84, 283−289

For biological interpretation, users are mainly interested in the compounds, rather than the redundancy of the different ion species, which induce an undesired bloat in the number of observed features, e.g., for an *Arabidopsis thaliana* seed extract Bottcher et al.[11] reported 434 features for 180 compounds. The complexity of both the downstream statistical analysis and subsequent compound identification especially in untargeted metabolite profiling experiments is unduly increased.

To address these problems, two additional processing steps are desired for LC/MS data analysis: (1) grouping all features which are derived from the same analyte, and (2) annotation of the type of ion species. The first step alone achieves both a data reduction and a first estimation of the total number of detectable compounds in a MS analysis. Such an estimate can be used for the optimization of the analytical protocol, similar to Yanes et al.[12] where the authors used the feature number as optimization criterion. Both steps together can reveal quasi-molecular ions, whose annotation is essential for further metabolite identification, such as elemental composition calculation based on accurate mass and isotope pattern or tandem MS analysis.

The authors of Brown et al.[13] have developed a workflow using the retention time, m/z-difference, and intensity correlation across samples to group related features, both reducing the number of relevant features down to 50% and matched 60% of the remaining features against the Manchester Metabolome Database (MMD). Intensity correlation across samples is also used by Alonso et al.,[14] and a data-reduction of 86% is reported.

Alternatively, similarity across chromatographic peak shapes allows the grouping of related features. Ipsen et al.[15] use a χ^2 test to check for exact coelution. In case of LC/MS data acquired on TOF instruments with a time-to-digital converter, the test provides p-values for the (un)certainty of coelution. The test works best with low ion counts, and the instruments' detector saturation correction had been disabled for this evaluation. ACD/IntelliXtract[16] is a commercial software solution to cluster features based on their retention time and the annotation of ion species according to a given rule table.

Both correlation across samples and peak shape analysis techniques are used in the R package ESI.[17] A fixed m/z-difference rule table is used for annotation and detection of isotopic peaks. The same approach was later used by Scheltema et al.[18] for high-resolution LC/MS data. By explicitly removing features exhibiting both similar peak shapes and intensity correlation across samples, they achieved a 60% size-reduction of the feature list.

In this paper we present the CAMERA package, which integrates multiple methods for grouping related features and uses a dynamic rule table for the annotation of ion species. We evaluate the performance of CAMERA with several validation experiments and demonstrate the analysis of two metabolomics experiments.

■ THEORY, ARCHITECTURE, AND ALGORITHMS

The analysis workflow with CAMERA is shown in Figure 1 and numbering (1−5) describes the typical workflow order. In the next paragraphs the steps are explained in more detail.

Creating Compound Spectra Based on Retention Time ①. The initial creation of compound spectra has to be fast, if dozens to hundreds of samples with thousands of features have to be processed. We select the most intense feature from the feature table not yet assigned to a compound spectrum and calculate a feature specific retention time window, typically 60% of the chromatographic peak fwhm (full width at half-maximum) around the centroid. All features within this range are then included into a new compound spectrum. This step is repeated until all features are assigned to a compound spectrum. The most intense feature usually has the highest signal-to-noise (S/N) ratio and often provides the most accurate estimate of the centroid and retention time.

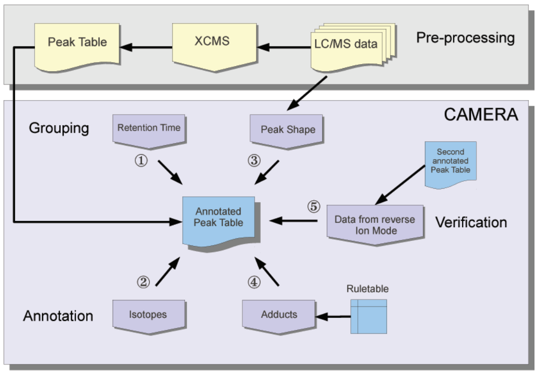


Figure 1. The CAMERA workflow for LC/MS data analysis. Raw data files are preprocessed with XCMS (upper part) and the resulting feature lists are passed to CAMERA. The feature grouping steps integrate retention time ① and chromatographic peak shape ③. Features are identified as isotopic peak ②, and adducts are annotated ④ using a dynamic rule table. Optionally, the annotation can be verified ⑤ with LC/MS data acquired in the opposite ion mode.

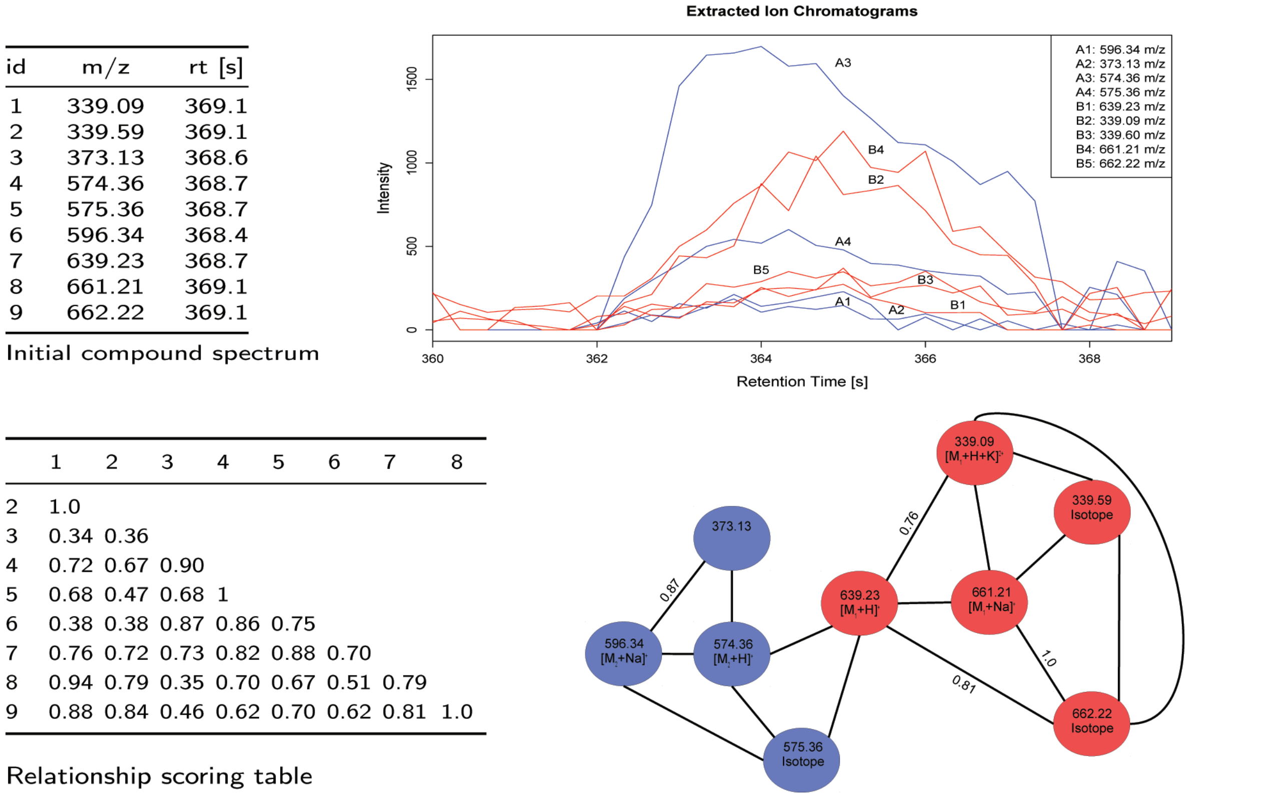
Isotopic Peak Detection and Charge State Calculation ②. The detection of isotopic patterns is required to deduce the charge states. Within each compound spectrum we calculate a pairwise m/z distance matrix and detect isotopes which exhibit a m/z-difference of $1.0033/z$[19] and also pass an additional intensity ratio check, described in detail in the Supporting Information, section S1.

Compound Spectrum Refinement Graph ③. Depending on the chromatographic separation, the resulting compound spectra might still encompass features of two or more closely coeluting compounds. We use a graph-based algorithm to integrate three more cues for an improved separation (see Figure 2 for an example).

First, we use the chromatographic peak shape similarity. CAMERA uses the raw data to obtain the extracted ion chromatograms (EIC) for each feature and calculates a pointwise pearson correlation of the intensities between the chromatographic peak boundaries for all pairs of features in a compound spectrum. CAMERA uses the EICs from the sample which had the most intense feature, often the one with the best S/N ratio. Alternatively, the peak shape correlation can be performed for all samples in the experiment. Second, we include the pearson correlation of intensities across all samples for each pair of features in a compound spectrum. Finally we encode the isotope relationship between two features detected in step ② as 1, and 0 otherwise. These three values are combined as shown in eq 1.

$$\mathrm{score}(x, y) = \mathrm{CAS}_{xy} + \mathrm{ISO}_{xy} + \frac{1}{N}\sum_{i=1}^{N} \mathrm{CPS}_{ixy} \tag{1}$$

The score which represents the relationship between two features x and y is the combination of the intensity correlation across samples (CAS) for these two features, the binary



id	m/z	rt [s]
1	339.09	369.1
2	339.59	369.1
3	373.13	368.6
4	574.36	368.7
5	575.36	368.7
6	596.34	368.4
7	639.23	368.7
8	661.21	369.1
9	662.22	369.1

Initial compound spectrum

	1	2	3	4	5	6	7	8
2	1.0							
3	0.34	0.36						
4	0.72	0.67	0.90					
5	0.68	0.47	0.68	1				
6	0.38	0.38	0.87	0.86	0.75			
7	0.76	0.72	0.73	0.82	0.88	0.70		
8	0.94	0.79	0.35	0.70	0.67	0.51	0.79	
9	0.88	0.84	0.46	0.62	0.70	0.62	0.81	1.0

Relationship scoring table

Figure 2. Schematic clustering of low-intensity features initially grouped by retention time into a single compound spectrum. Top left: the features, initially grouped by retention time. Top right: the EICs of all features. The labels A and B correspond to the result after graph clustering. Bottom left: the scoring matrix, used as edge weights in the graph. Bottom right: the relationship graph, where edges indicate an above-threshold score. The node labels include the ion species annotation, and the node color shows the graph separation after refinement with the LPC algorithm (A = blue, B = red).

encoded presence or absence of an isotope relationship, and the peak shape correlation (CPS_i) calculated for sample i.

In a graph, all features in a compound spectrum, which could still include features of two or more closely coeluting compounds, are represented as nodes, connected by edges with this score as edge weight. Several algorithms for graph separation have been developed, and we employ the "Highly-connected-subgraphs" (HCS[20]) from the R package RBGL or the "label propagation community" (LPC[21]) from the R package igraph. After the graph clustering, the initial compound spectrum is split into one refined compound spectrum for each subgraph. Figure 2 shows an example for a relationship graph before and after separation. Both coeluting compounds were separated completely.

Annotation of Adducts, Common Neutral Losses, and Cluster-Ions ④. For ESI, uncharged compounds are ionized through adduct formation with cations or anions or abstraction of protons. In addition, neutral losses occur leading to the formation of fragment ions. An annotation of these ion species reduces the number of features which have to be considered further in the downstream analysis. From at least two annotated ions, the molecular mass can be calculated, which is necessary to search in compound libraries or to calculate the elemental composition of the neutral compound.

CAMERA uses a dynamic rule set, which is created from the combination of lists of observable ions. Each rule describes a specific ion species with the mass difference to the molecular mass, ion charge, and the number of molecules the ion species contains. All m/z-differences within a compound spectrum are matched against the dynamic rule set. Matches with the same molecular mass hypothesis (below a given relative error) are combined into hypothesis groups. If no peaks can be explained via the rules, a reliable annotation is impossible. CAMERA does not use ad-hoc heuristics such as assuming that the most intense feature in a spectrum is the $[M + H]^+$-ion. Afterward, conflicting hypothesis groups are resolved as described in the Supporting Information, section S2.

Combining Data from Opposite Ion Modes for Verification ⑤. In metabolite profiling, samples are often measured in both positive and negative ion modes to increase the metabolite coverage. Although some compounds ionize in only one mode, many compounds are detectable in both. In these cases, the complementary ions provide further evidence for the quasi-molecular ion.

CAMERA includes a novel annotation verification algorithm using compound spectra measured in both ion modes. The algorithm calculates m/z-differences for all features of corresponding compound spectra from both modes within a retention time window. These differences are matched against a second, cross-polarity rule table. If a cross-polarity rule matches, it will either (1) annotate two previously unannotated ions, e.g., $[M + H]^+$ and $[M - H]^-$, or (2) verify an existing annotation, or (3) conflict with an existing annotation. In the latter case, the existing annotation is replaced. The cross-polarity rule table should only contain common and trusted combinations because these rules can override the single-polarity annotations.

Documentation and Availability. CAMERA is implemented in R, the packages for Windows (both 32 and 64 bit), Mac OS, and Linux are available from the Bioconductor repository[22] since release 2.4 in 2009.

■ EXPERIMENTAL SECTION

Reagents and Materials. All solvents used for sample preparation and analyses were of LC/MS-grade quality (CHROMASOLV, Fluka). A list of standard compounds used for the recovery experiment including sum formulas, molar masses, PubChem IDs and suppliers can be found in the Supporting Information, section S3. L-Tryptophan-2′,4′,5′,6′,7′-d_5 (98%) was purchased from Cambridge Isotope Laboratories. *Arabidopsis thaliana* (ecotype Col-0) was grown for 6 weeks on a soil/vermiculite mixture (3/2) in a growth cabinet with 8 h light (150 μE m^{-2}s^{-1}) at 22 °C and 16 h dark at 20 °C. Seeds of *Brassica napus*, *Brassica oleracera*, and *Brassica rapa* were kindly provided by D. Strack, Department of Secondary Metabolism, Leibniz Institute of Plant Biochemistry, Halle. All other seeds were obtained from local distributors. Procedures for extraction of leaf and seed material are provided in the Supporting Information, section S4.

Feeding Experiments. Plants were sprayed with 5 mM aqueous silver nitrate solution 1 h after the beginning of the light period. After 5 h, 25 rosette leaves originating from 5 individual plants were excised at the petiole and immersed in PCR tubes containing either 200 μL of water or 200 μL of an aqueous [ring-D_5]-Trp solution (1 mM), respectively. Leaves were incubated for an additional 2 days in a growth cabinet under the same conditions as described above. Individual leaves of the same treatment were pooled, frozen in liquid nitrogen, and stored at −80 °C until analysis.

Ultraperformance Liquid Chromatography (UPLC)/ESI-Quadrupole Time-of-Flight (QTOF)MS Analysis. Chromatographic separations were performed on an Acquity UPLC system (Waters) equipped with a HSS T3 column (100 mm × 1.0 mm, particle size 1.8 μm, Waters) applying the following binary gradient at a flow rate of 150 μL min^{-1}: 0−1 min, isocratic 95% A (water/formic acid, 99.9/0.1 (v/v)), 5% B (acetonitrile/formic acid, 99.9/0.1 (v/v)); 1−16 min, linear from 5 to 95% B; 16−18 min, isocratic 95% B; 18−20 min, isocratic 5% B. The injection volume was 2.7 μL (full loop injection). Eluted compounds were detected at a spectra rate of 3 Hz from m/z 100−1000 using a MicrOTOF-Q-I hybrid quadrupole time-of-flight mass spectrometer (Bruker Daltonics) equipped with an Apollo II electrospray ion source in positive and negative ion modes. We made sure that the concentration of the samples do not lead to saturation of the MS detector system, which is known to cause shifts of m/z, and retention time centroids of the features leads to truncated chromatographic peak profiles and distorted isotopic patterns. For detailed instrument settings and acquisition of collision-induced dissociation mass spectra see the Supporting Information, section S4.

LC/MS Data Preprocessing. Processing of raw data was performed with the XCMS package.[8] For the feature detection, we used the XCMS *centWave*[23] algorithm with the following parameters: snthresh = 6, ppm = 30, peakwidth = (5,12), prefilter = (2,200). The feature alignment was performed with the standard *group.density* algorithm from XCMS with bw = 3 and mzwid = 0.015. Afterward, each data set was processed with CAMERA functions in the following order *groupFWHM*, *findIsotopes*, *groupCorr*, and *findAdducts* using standard parameters. The Supporting Information, section S9 provides runtime measurements of CAMERA.

■ RESULTS AND DISCUSSION

We evaluated CAMERA with several experiments. First, using standards we analyzed the performance of compound spectrum creation and success rate of molecular mass annotation. Then, we processed the output from two different experiments, where the CAMERA results were used to perform targeted profiling of phenolic choline esters and tryptophan-derived metabolites, respectively.

Evaluation on Known Compound Mixture. For the evaluation we used a mixture of 39 known compounds (short, MM39), covering a broad mass range between 161 and 822 Da and different physicochemical properties (see the Supporting Information, section S3). The mixture was measured as pure solution and spiked in different concentrations (20, 5, 1, and 0.2 μM) into methanolic extracts of *Arabidopsis thaliana* leaves to simulate a realistically complex matrix.

The first evaluation focuses on the extraction of compound spectra, which requires a data set and a gold standard of true positive and true negative cases, i.e., pairs of peaks which should or should not be part of the same compound spectrum. Because it is very tedious to manually create a gold standard of a sufficiently large number of features from different compounds which coelute, we altered the retention times in the raw data files to artificially force "coelution" for this evaluation. We used only those peaks in the compound spectra of the MM39 for which a reliable annotation exists, to rule out false positives and randomly collected peaks from the remaining file with unrelated retention times to assemble a negative set. These data sets allowed us to calculate the precision and recall for the collection of compound spectra. The default peak shape correlation threshold of 0.75 results in a recall of 0.93, with a precision of 0.48. We also analyzed the influence of different acquisition parameters (scan rates varied from 0.5 to 6 Hz). Precision and recall had a standard deviation of 0.07 and 0.03, respectively, across the different conditions, see the Supporting Information, section S5 for details, including the ROC curves.

We then evaluated how successfully CAMERA could annotate the different ion species from a compound, which is required for the calculation of the molecular mass. We created baseline values for all annotatable compounds: we define an annotatable compound as observed to produce (1) the protonated molecular ion, (2) its first isotopic peak (required to calculate the charge state), and (3) the most prominent adduct ion (observed at 20 μM). This strategy serves as the gold standard to determine the number of annotatable compounds in the MM39 measurement; 35 out of 39 compounds pass the above requirements for the 20 μM positive mode measurement. If the mixture is diluted 2 orders of magnitude to 0.2 μM, many peaks drop below the detection limit and only 10 compounds remain annotatable.

CAMERA was able to detect the correct molecular mass in 90% of all annotatable compounds in either the positive or negative mode across all concentrations. After combining results from both ionization modes, CAMERA correctly determined molecular mass for all annotatable compounds and additionally for four compounds that were not on the gold standard list. Because the manual assignment of corresponding features in both positive/negative mode data is quite cumbersome, the combined annotation promises to annotate more compounds than a human operator could do on a routine basis.

Table 1. Calculation of Molecular Mass for the MM39 Compound Mixture Analyzed by UPLC/ESI-QTOFMS in Positive and Negative Ion Mode, Either in Pure Solvent or Spiked at Different Concentrations into a *Arabidopsis thaliana* Leaf Extract[a]

	in solvent	spiked into leaf extract				overall	
	20 μM	20 μM	5 μM	1 μM	0.2 μM		
ESI(+)	32 (35)	29 (32)	24 (28)	18 (21)	10 (10)	113 (126)	89.7%
ESI(−)	15 (19)	18 (18)	15 (16)	6 (6)	2 (3)	56 (62)	90.3%
ESI(±)	36	35	28	23	15	137	

[a] The number of annotatable compounds is shown in brackets. In the combined case, the annotations of the positive ion mode are verified and augmented with the negative ion mode data.

It is remarkable that the complex leaf matrix does not have an observable negative effect on the annotation performance. Table 1 shows the results for the individual concentrations. On closer inspection, the missing molecular mass annotations have few common causes: they occurred either because the compound spectrum did not contain enough explained features or, in other cases, several hypotheses had the same precedence scores and we did not count those as successful. For some compounds the compound spectra contained only the molecular ion and fragment ions but no further adducts. In this case, the compound cannot be annotated directly, unless the neutral loss is added to the rule set. The Supporting Information, section S10 shows an overview of the frequency of annotated adducts we observed.

In the measurements with different scan rates, we found that CAMERA only missed up to two correct annotations in those cases where either an essential (albeit low abundant) feature was not found by the feature detection algorithm or features were assigned to a different compound spectrum, especially in the case of lower scan rates where chromatographic peaks were covered by only a few scans. This suggests that CAMERA can also be used for LC/MS measurements with a low scan rate, e.g., on Orbitrap instruments at high resolution.

Case Study I: Screening for Phenolic Choline Esters in Brassicaceous Seeds. In this section we use untargeted LC/MS profiles of seeds from some *Brassicacea*, and demonstrate how CAMERA can be used to perform a neutral loss screen for phenolic choline esters as a targeted analysis strategy on a TOF instrument.

Phenolic choline esters accumulate in considerable amounts in seeds of many plant species within the *Brassicacea*[24] family. Representatives of this compound class structurally characterized so far include substituted cinammoyl and benzoyl cholines, which are further diversified by glycosylation or oxidative coupling to monolignols. A total of 30 phenolic choline esters could be identified in seeds of the model plant *Arabidopsis thaliana* and the oil crop *Brassica napus* using LC/ESI-tandem mass spectrometry.[25] A study of the fragmentation behavior of phenolic choline esters under positive-ion electrospray-CID conditions revealed a loss of trimethylamine as the initial fragmentation step (see the Supporting Information, section S6). The formation of the corresponding fragment ion $[M - C_3H_9N]^+$ requires different collision energies depending on the compound. However, it is also inducible by in-source CID allowing a systematic screening for phenolic choline esters even by single-stage MS. For that purpose, the neutral loss detection has to be performed *in silico* after data acquisition by searching for a given *m*/*z*-difference between pairs of peaks within a set of extracted compound spectra.

We prepared extracts from seeds of 12 different *Brassicacea* species and cultivars and analyzed each extract by UPLC/ESI(+)-QTOFMS at four different in-source CID voltages (0, 30, 60, and 90 V) to induce fragmentation of a broad range of phenolic choline esters. All 48 raw data files were preprocessed with XCMS with snthresh = 5 and ppm = 20, other parameters analogous to the evaluation section. Because of the large number of chromatographically unresolved compounds eluting near the void time, compounds with a retention time below 45 s were excluded from further analysis. Afterward CAMERA was used to create the compound spectra. Each compound spectrum was then screened for peak pairs displaying a *m*/*z*-difference of 59.074 ± 0.015 corresponding to a neutral loss of trimethylamine. In addition, we included a *m*/*z*-difference of 221.126 ± 0.015, related to a successive loss of trimethylamine and anhydrohexose (162.053 Da), because $[M - C_3H_9N]^+$-type ions formed from 4-*O*-hexosylated phenolic choline esters are known to readily eliminate their hexose moiety.[25] After alignment of positively screened peak pairs, the elimination of isotopic peak pairs, and application of a reasonable intensity threshold (1000 counts), we detected a total of 90 putative choline esters. A data matrix including *m*/*z* ratios of proposed molecular ions, retention times, and intensities can be found in the Supporting Information, section S6. It should be noted, that the number of putative candidates is rapidly increasing when tolerance thresholds for *m*/*z*-differences were increased. Therefore, use of mass analyzers providing adequate resolution and mass accuracy, such as TOFMS, is mandatory for this type of screening approach in order to ensure a highly specific neutral loss detection. In order to evaluate the obtained candidate list, previously published analytical data of choline esters from seeds of *Arabidopsis thaliana* and *Brassica napus* were used for compound annotation.[25] Out of 31 choline esters described recently, we were able to retrieve 22 from our list. Seven choline ester were consistently detected across all samples. Five of them could be annotated (Table 2), including sinapoylcholine, which is known to occur as a major phenolic choline ester in seeds of numerous *Brassicacea* species.[24] Although a rigorous evaluation is not possible because the choline ester composition of analyzed seeds is unknown, recovery of the majority of compounds described in the literature demonstrates the usability of CAMERA for such a

Table 2. Five Phenolic Choline Esters Found and Annotated in All 12 Brassicaceous Seeds[a]

m/*z*	t_r [s]	NL	elemental composition	annotation
280.15	275	−59	$C_{15}H_{22}NO_4^+$	FC
310.16	279	−59	$C_{16}H_{24}NO_5^+$	SC
458.21	403	−59	$C_{25}H_{32}NO_7^+$	FC(5-8′)G
472.21	221	−221	$C_{22}H_{34}NO_{10}^+$	SC 4-*O*-Hex
476.23	303	−59	$C_{25}H_{34}NO_8^+$	FC(4-*O*-8′)G

[a] NL, neutral loss; FC, ferulolycholine; SC, sinapoylcholine; G, guaiacyl; Hex, hexose.

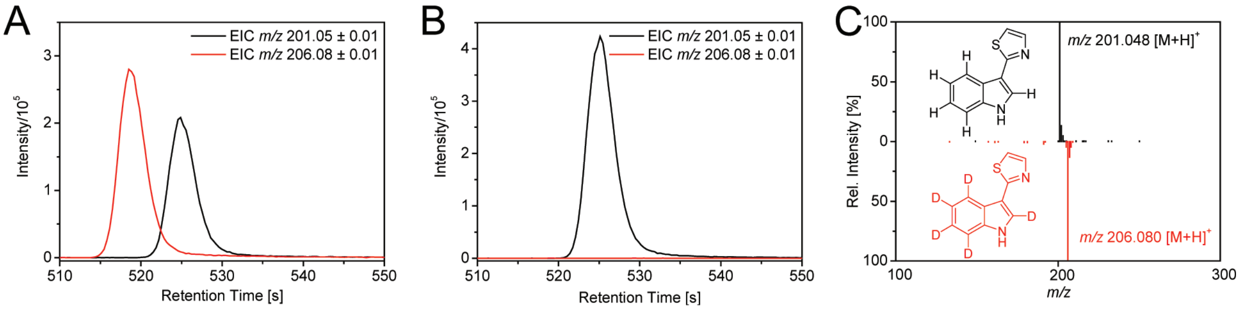


Figure 3. Identification of the phytoalexin camalexin as Trp-derived metabolite in silver nitrate-treated /*Arabidopsis thaliana*/ leaves using [ring-D_5]-Trp as the isotope-labeled tracer. Extracted ion chromatograms (EICs) corresponding to the protonoted molecular ions of camalexin (black) and D_5 -camalexin (red) obtained from UPLC/ESI(+)-QTOFMS analyses of extracts of [ring-D_5]-Trp-fed leaves (A) and control leaves (B). Extracted compound spectra of camalexin and its isotopologue are shown in the right picture (C).

screening approach. An additional advantage of this approach compared to triple quadrupole MS-based neutral loss scanning techniques is that any number of neutral losses can be simultaneously detected after data acquisition, allowing screening for a broad range of compound classes.

Case Study II: Identification of Trp-Derived Metabolites from *Arabidopsis thaliana* after [ring-D_5]-Trp Feeding. In vivo administration of isotope-labeled substrates combined with mass spectrometry-based analysis represents a powerful tool to investigate biochemical pathways. The detection of an isotope-labeled substrate incorporated into a known metabolite allows one to deduce a biosynthetic relationship between the fed precursor and the metabolite under study. Nontargeted screening for metabolites and their isotopologues after partial isotope-labeling of an endogenous precursor pool has been applied to explore unknown biosynthetic pathways and to discover novel intermediates and products.[26]

To demonstrate the applicability of the CAMERA package for such an analytical approach, the metabolic fate of the aromatic amino acid Trp was studied in the model plant *Arabidopsis thaliana* using [ring-D_5]-Trp as the isotope-labeled tracer. In *Arabidopsis*, Trp represents an important precursor for a variety of secondary metabolites including the phytoanticipin indol-3-ylmethyl glucosinolate and the phytoalexin camalexin (3-thiazol-2′-yl-indole).

Arabidopsis leaves were sprayed with silver nitrate to induce expression of Trp-metabolizing enzymes, detached from plants and fed with [ring-D_5]-Trp or water as the control. Methanolic extracts of label-fed and control leaves were analyzed in duplicate by UPLC/ESI-QTOF-MS in the positive and negative ion modes. In order to identify Trp-derived metabolites, the raw data was processed with XCMS and CAMERA to extract compound spectra and annotate isotopic peaks within these spectra. Afterward, deisotoped compound spectra extracted from data sets of label-fed leaves were screened for feature pairs that exhibit an *m*/*z*-difference of 5.031, reflecting the exchange of five hydrogen atoms by deuterium. Since deuterium labeling can slightly shift retention times, we searched for these feature pairs between compound spectra within a sliding retention time window of 8 s. For this purpose, we created a dedicated script using CAMERA functionality for the positive/negative polarity combination. We also included the *m*/*z*-difference of 4.025 because indole ring hydroxylation (a frequently observed transformation in Trp metabolism in *Arabidopsis*) results in a loss of one of the five deuterium labels. The retention time for Trp-candidates was restricted between 45 and 600 s. All features related to unlabeled Trp-metabolites have to be detectable in both label-fed and control samples whereas the labeled ones in label-fed samples only, see Figure 3. After those filtering steps, 46 putative Trp-derived metabolites could be identified in the positive ion mode and 34 in the negative ion mode. Corresponding candidate lists including compound annotation can be found in the Supporting Information, section S7.

To verify the obtained candidate lists, tandem mass spectra of quasimolecular ions of putative pairs of nonlabeled and labeled metabolites were acquired and compared (Supporting Information, section S8). Because of low peak intensities or low incorporation rates, only 19 candidate pairs could be rigorously verified following this strategy. Together with literature data, a total of 23 Trp-metabolites could be identified, of which 20 were already known from the literature. This case study clearly demonstrates applicability of CAMERA for such a screening approach, even in case of a retention time shifts when using deuterium labels.

■ CONCLUSIONS

The CAMERA package is designed to postprocess XCMS feature lists and to collect all features related to a compound into a compound spectrum. For this, a set of algorithms has been implemented in CAMERA, such as the fast retention time-based grouping but also a novel, graph-based algorithm to integrate the peak shape analysis, isotopic information, and intensity correlation across samples. The automatic sample selection avoids poor results if compounds have a low intensity (or are absent) in some samples. The ion species annotation uses a dynamic rule set and a new strategy to combine spectral information from samples measured in the positive and negative ion modes, resulting in both more and more reliable ion species annotation. We evaluated the reliability of the molecular mass calculation and found a 90% success rate for MM39 in different concentrations, both pure and after spiking the mixture at various concentrations into a complex *Arabidopsis thaliana* leaf extract.

Finally, we performed two experiments, demonstrating advanced analyses which can be performed with CAMERA. The first case study essentially performed a neutral loss screen for putative phenolic choline esters using multiple in-source voltages to induce fragmentation. In total, 90 putative choline esters were detected. The second case study demonstrated the search for mass differences as a result of [ring-D_5]-Trp feeding in *Arabidopsis thaliana* leaves. CAMERA was used to detect pairs of isotopologue features indicating 46 Trp-derived metabolites. In addition to 20 already known compounds, 3 new ones were found and verified with tandem MS. Both studies can easily be adopted to other compound classes and metabolites. The CAMERA packages for Windows, Mac OS, and Linux, manuals, and tutorials are freely available from the Bioconductor repository and its mirrors under the open source GPL license.

■ ASSOCIATED CONTENT

S Supporting Information

Experimental procedures and characterization data for all new compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ckuhl@ipb-halle.de (C.K.); sneumann@ipb-halle.de (S.N.).

■ ACKNOWLEDGMENTS

This work was supported by the California Institute of Regenerative Medicine (Grant TR1-01219), the National Institutes of Health (Grants R24 EY017540-04, P30 MH062261-10, and P01 DA026146-02), and the Department of Energy (Grants FG02-07ER64325 and DE-AC0205CH11231).

■ REFERENCES

(1) Dunn, W. B. *Phys. Biol.* **2008**, *5*, 011001.
(2) Feist, A. M.; Henry, C. S.; Reed, J. L.; Krummenacker, M.; Joyce, A. R.; Karp, P. D.; Broadbelt, L. J.; Hatzimanikatis, V.; Palsson, B. *Mol. Syst. Biol.* **2007**, *3*, 121.
(3) Psychogios, N.; et al. *PLoS One* **2011**, *6*, e16957.
(4) Trethewey, R. N. *Curr. Opin. Plant Biol.* **2004**, *7*, 196−201.
(5) Tikunov, Y.; Lommen, A.; Vos, C. d.; Verhoeven, H.; Bino, R.; Hall, R.; Bovy, A. *Plant Physiol.* **2005**, *139*, 1125−37.
(6) Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. *BMC Bioinf.* **2008**, *9*, 163.
(7) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.
(8) Smith, C.; Want, E.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.
(9) Katajamaa, M.; Oresic, M. *J. Chromatogr., A* **2007**, *1158*, 318−328.
(10) Keller, B. O.; Sui, J.; Young, A. B.; Whittal, R. M. *Anal. Chim. Acta* **2008**, *627*, 71−81.
(11) Böttcher, C.; von Roepenack-Lahaye, E.; Schmidt, J.; Schmotz, C.; Neumann, S.; Scheel, D.; Clemens, S. *Plant Physiol.* **2008**, *147*, 2107−2120.
(12) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2011**, *83*, 2152−2161.
(13) Brown, M.; Wedge, D. C.; Goodacre, R.; Kell, D. B.; Baker, P. N.; Kenny, L. C.; Mamas, M. A.; Neyses, L.; Dunn, W. B. *Bioinformatics* **2011**, *27*, 1108−1112.
(14) Alonso, A.; JuliÃ, A.; Beltran, A.; Vinaixa, M.; Díaz, M.; Ibañez, L.; Correig, X.; Marsal, S. *Bioinformatics* **2011**, *27*, 1339−1340.
(15) Ipsen, A.; Want, E. J.; Lindon, J. C.; Ebbels, T. M. D. *Anal. Chem.* **2010**, *82*, 1766−1778.
(16) ACD/IntelliXtract, Advanced Chemistry Development, Inc. www.acdlabs.com/intellixtract, 2007.
(17) Tautenhahn, R.; Böttcher, C.; Neumann, S. *Lect. Notes Comput. Sci.* **2007**, *4414*, 371−380.
(18) Scheltema, R.; Decuypere, S.; Dujardin, J.; Watson, D.; Jansen, R.; Breitling, R. *Bioanalysis* **2009**, *1*, 1551−1557.
(19) Yergey, J. A. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52*, 337−349.
(20) Hartuv, E.; Shamir, R. *Inf. Process. Lett.* **2000**, *76*, 175 −181.
(21) Raghavan, U. N.; Albert, R.; Kumara, S. *Phys. Rev. E: Stat. Nonlinear, Soft Matter Phys.* **2007**, *76*, 036106.
(22) Gentleman; Rossini; Dudoit; Hornik. The Bioconductor FAQ. http://www.bioconductor.org,2003.
(23) Tautenhahn, R.; Böttcher, C.; Neumann, S. *BMC Bioinf.* **2008**, *9*, 504.
(24) Bouchereau, A.; Hamelin, J.; Lamour, I.; Renard, M.; Larher, F. *Phytochemistry* **1991**, *30*, 1873−1881.
(25) Böttcher, C.; von Roepenack-Lahaye, E.; Schmidt, J.; Clemens, S.; Scheel, D. *J. Mass Spectrom.* **2009**, *44*, 466−476.
(26) Feldberg, L.; Venger, I.; Malitsky, S.; Rogachev, I.; Aharoni, A. *Anal. Chem.* **2009**, *81*, 9257−9266.