PERGAMON

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Phytochemistry 62 (2003) 887–900

PHYTOCHEMISTRY

www.elsevier.com/locate/phytochem

Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles

Cornelia Wagner[a], Michael Sefkow[b], Joachim Kopka[a,*]

[a]*Department Willmitzer, Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, D-14467 Golm, Germany*
[b]*Division of Natural Product Chemistry, Institute for Chemistry, University of Potsdam, Karl-Liebknecht-Strasse 24-25, D-14467 Golm, Germany*

Received 9 September 2002; received in revised form 18 October 2002

Abstract

The non-supervised construction of a mass spectral and retention time index data base (MS/RI library) from a set of plant metabolic profiles covering major organs of potato (*Solanum tuberosum*), tobacco (*Nicotiana tabaccum*), and *Arabidopsis thaliana*, was demonstrated. Typically 300–500 mass spectral components with a signal to noise ratio $\geqslant 75$ were obtained from GC/EI-time-of-flight (TOF)-MS metabolite profiles of methoxyaminated and trimethylsilylated extracts. Profiles from non-sample controls contained approximately 100 mass spectral components. A MS/RI library of 6205 mass spectral components was accumulated and applied to automated identification of the model compounds galactonic acid, a primary metabolite, and 3-caffeoylquinic acid, a secondary metabolite. Neither MS nor RI alone were sufficient for unequivocal identification of unknown mass spectral components. However library searches with single bait mass spectra of the respective reference substance allowed clear identification by mass spectral match and RI window. Moreover, the hit lists of mass spectral searches were demonstrated to comprise candidate components of highly similar chemical nature. The search for the model compound galactonic acid allowed identification of gluconic and gulonic acid among the top scoring mass spectral components. Equally successful was the exemplary search for 3-caffeoylquinic acid, which led to the identification of quinic acid and of the positional isomers, 4-caffeoylquinic acid, 5-caffeoylquinic acid among other still non-identified conjugates of caffeic and quinic acid. All identifications were verified by co-analysis of reference substances. Finally we applied hierarchical clustering to a complete set of pair-wise mass spectral comparisons of unknown components and reference substances with known chemical structure. We demonstrated that the resulting clustering tree depicted the chemical nature of the reference substances and that most of the nearest neighbours represented either identical components, as judged by co-elution, or conformational isomers exhibiting differential retention behaviour. Unknown components could be classified automatically by grouping with the respective branches and sub-branches of the clustering tree.

Keywords: *Arabidopsis thaliana*; Brassicaceae; *Solanum tuberosum*; *Nicotiana tabacum*; Solanaceae; Caffeoylquinic acids; Chlorogenic acid; Caffeic acid; Quinic acid; Galactonic acid; Gluconic acid; Glucaric acid; Gulonic acid; Ascorbic acid; Metabolite profiling; Mass spectral library; Gas chromatography (GC); Time-of-flight mass spectrometry (TOF-MS); Retention time index (RI)

1. Introduction

In the next few years metabolome analyses will have emerged from infancy to being firmly established as the third cornerstone of functional genomics. Multi-parallel measurements of the large variety of primary and secondary metabolites ideally complement the current focus of functional genomics, namely mRNA profiling and proteomics approaches. Metabolite profiling technology will spread not only within the area of plant biotechnology (Fiehn et al., 2000a; Fiehn, 2002; Frenzel et al., 2002; Huhman and Sumner, 2002; Roessner et al., 2000, 2001a) but across all fields of biological science, because it allows direct access to general systems analyses as is exemplified by applications in drug discovery (Boros et al., 2002) or the profiling of genetic disorders (Griffin et al., 2001).

Abbreviations: EI, Electron impact ionization; GC, gas chromatography; RI, retention time index; MS, mass spectrum; TOF, time-of-flight

* Corresponding author. Tel.: +49-331-567-8262; fax +49-331-567-898262.

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

0031-9422/03/$ - see front matter © 2003 Published by Elsevier Science Ltd.
doi:10.1016/S0031-9422(02)00703-3

One of the central technology platforms of metabolic profiling technology is bench-top gas chromatography coupled to mass spectrometry. The choice of this hyphenated technology was motivated by unsurpassed combination of chromatographic separation power, selectivity, sensitivity, and dynamic range of mass detection. Moreover, both gas chromatography and electron impact ionization mass spectrometry exhibit extremely high reproducibility and are now applied to multi-parallel analyses of hundreds of biological samples (Fiehn et al., 2000a). Prior and recent applications in medical diagnostics (Duez et al., 1996; Matsumoto and Kuhara, 1996; Kuhara, 2001) as well as upcoming publications within plant science also rely on GC/MS technology.

The major application of GC/MS based metabolite profiling has been phenotypic characterization and classification of genetically altered plant samples with attempts to copy the phenotype of genetic lesions by experimental treatment of non-modified plants (Roessner et al., 2001a, b). However, approaches to utilize the wealth of mass spectral information within each single GC/MS profile have not yet been explored. The final aim of this aspect is a comprehensive and automated analysis of all mass spectral components from a metabolic profile. The importance of identification and chemical classification of the mass spectral components may best be acknowledged by analogy with functional annotation of unknown genes in transcriptome analysis. Whereas genes are characterized by nucleotide sequences, electron impact ionization mass spectra (EI-MS) have a similar import for small compounds. They represent the fingerprint of the molecular fragmentation pattern of chemical structures. Annotation and classification of genes by alignment and evaluation of sequence homology today is fully automated. In parallel the comparison of mass spectra is automated to a similar degree. The NIST (National Institute of Standards and Technology, Gaithersburg, MD, USA) released a mass spectral search program which is publicly available and represents a platform independent GC/MS analysis software. However, the mass spectral match is insufficient for unequivocal substance identification, mainly because structural isomers especially conformational isomers may produce highly similar mass spectra.

Thus a typical attempt to identify a given component from a GC/MS profile is tedious. The process starts with manual interpretation of the fragmentation pattern and is supported by mass spectral comparison with commercial libraries. Pure substance of reference compounds is subsequently required to establish co-chromatography and mass spectral identity with unknown candidate components. Unfortunately the direct route via preparation of pure fractions from GC/MS runs and subsequent unambiguous identification by NMR is currently not feasible because of the enormous difference in preparative capacity and analytical sensitivity. Thus component identification from GC/MS profiles is restricted by commercial or preparative access to pure reference compounds.

As a consequence we started to develop an efficient use of those reference compounds that are available. We currently focus on the simultaneous identification and classification of unknown components in a high number of different plant matrices. For this purpose we exploit the two characteristic substance properties provided by GC/MS, i.e., retention time index (RI) and mass spectrum. Other than stand-alone mass spectrometry hyphenated techniques like GC/MS provide the potential to separate conformational isomers prior to mass detection. Thus with the aim to group and annotate unknown compounds from multiple matrices it appears to be highly negligent not to utilize the chromatographic property of natural products.

Novel software developments promise automated calculation of RI and correct extraction—in other words deconvolution—of mass spectra from GC/MS chromatograms. In addition recently developed GC/EI-TOF-MS technology apparently exhibited high mass spectral reproducibility (Veriotti and Sacks, 2000). With these incentives we explored the potential of the combination of both technologies for qualitative analysis of GC/MS based metabolic profiles. We describe the non-supervised construction of a mass spectral and retention time index database (MS/RI library) from routine GC/EI-TOF-MS metabolite profiles of a trial selection of typical plant matrices. The composition of this mass spectral compendium is characterized and its application in substance identification and automated classification of unknown components is discussed.

2. Results and discussion

2.1. Choice of technology

GC/EI-TOF-MS technology was chosen for the purpose of this project. Other than scanning type mass spectrometers like quadrupole, ion trap, or sector instruments, time-of-flight technology combined fast acquisition rates with mass spectral integrity. Whereas in typical scanning GC/MS datasets the relative fragment abundance of mass spectra shifted from peak front to tail, mass spectra stayed unchanged when monitored by TOF-MS. This property and the option of high-resolution data aquisition were shown to be technologically beneficial with respect to deconvolution of co-eluting components and mass spectral comparison of petrochemical (Veriotti and Sacks, 2000) as well as biological profiles (Veriotti and Sacks, 2001).

Mass spectral deconvolution and automated calculation of RI was performed by the automated mass spectral deconvolution and identification system, AMDIS (National Institute of Standards and Technology, Gaithersburg, MD, USA). AMDIS was the first software tool available, which combined automated mass spectral deconvolution and calculation of RI. Moreover, AMDIS is publicly available and compatible with almost all file formats generated by commercial GC/MS systems including widely distributed quadrupole GC/MS systems. However, recent versions of the ChromaTOF™ software (LECO, St. Joseph, MI, USA) include fully equivalent options, but this software is restricted to the GC/EI-TOF-MS file format of the company. The choice of AMDIS in the present paper neither reflects a quality assessment nor is a software comparison intended. Alternative test versions of MS/RI libraries were successfully generated by ChromaTOF™ software version 1.6 and are in use.

AMDIS exhibited mainly two types of errors, i.e., deconvolution of more than one component per peak (Tables 3 and 4) and generation of accidental components with low intensity. Both errors were influenced by scan to scan fluctuations of the base line. Setting the smoothing option to a width of four scans and removal of components with a signal to noise ratio <75 empirically reduced both errors. A statistical assessment of residual erroneous components and of the number of true components lost was impossible, because all attempts to automatically identify artefact mass spectra by common properties failed. The tested parameter comprised the mass spectral composition and all available validation parameter of peak deconvolution which were provided by the AMDIS software. The NIST98 mass spectral search program (National Institute of Standards and Technology, Gaithersburg, MD, USA) is fully integrated with both the AMDIS and the ChromaTOF™ deconvolution software and, therefore, was chosen as software platform for MS/RI libraries.

The metabolite profiles entering the present MS/RI library were from a single series of GC/MS runs. Profiles from multiple analytical series may be combined (data not shown), however, the possibility of slight accidental changes of RI or mass spectral tuning settings was avoided in the present work. Gas chromatography was optimised to cover the volatility range of *n*-dodecane to *n*-hexatriacontane within a temperature ramp of 15 min. The data acquisition rate was set to 6 spectra per second resulting in 15–20 data points across a peak. This setting confirmed with the limit recommended for accurate description of chromatographic peaks (Dyson, 1999; van Deursen et al., 2000). Split injection used in previous publications was replaced by a hot splitless mode of injection. This development resulted in approximately 10-fold reduction of required sample amount (Table 2).

2.2. RI/MS Library construction and composition

The present MS/RI library was composed of mass spectral components obtained from 21 metabolic profiling experiments. Sample information and resulting number of components were listed in Table 1. The MS/RI library held 6205 mass spectra exhibiting an average peak purity of 47.4% (Table 2). Of these mass spectra 5.6% were derived from control experiments without plant sample. These non-sample controls were interspersed within plant samples with the aim to check for chemical contamination and in order to monitor memory effects in the course of the GC/MS sequences. Only caffeoylquinic acids exhibited slight carry over into the final control experiment (data not shown). The 18 plant profiles were generated in equal numbers from either 1–3 mg or 10–18 mg fresh weight. The nine samples with increased fresh weight had less than twice the yield of components and exhibited a significant decrease in peak purity. Pre-experiments with still higher sample load showed not only progressively reduced peak purity but also resulted in artefact RI. RI shifts were caused by peak overloading (data not shown).

The plant samples were from potato (*Solanum tuberosum*), tobacco (*Nicotiana tabacum*) and *Arabidopsis thaliana*. They comprised 53.6, 21.9, and 18.9% of library components, respectively. Leaf and root organs predominated. For potato plants three further organs were included—flower, stolon, and tuber. Root and tuber samples appeared to contain less components per fresh weight than leaf. In the case of tuber samples this finding confirms earlier observations (Roessner et al., 2000, 2001a).

A comprehensive representation of the MS/RI library with regard to the distribution of component amount and RI is given in Fig. 1. This plot may be interpreted as the full chromatographic reconstruction of all mass spectral components extracted from the present 21 metabolic profiles, where RI represents the time axis and the amount value replaces signal intensity. The amount values covered a range of four orders of magnitude, from $1.05\ 10^7$ to $1.26\ 10^3$, at threshold signal to noise ratio. The average signal to noise ratio was 397 accompanied by an average amount of $3.59\ 10^5$. Coelution was apparent (Fig. 1) to such an extent that peak identification could not rely exclusively on chromatographic retention. However, RI showed high reproducibility with typical standard deviations ranging from 0.1 to 2.0 (Tables 3 and 4). Therefore we concluded that under our experimental conditions RI is a highly valid additional parameter for substance identification.

2.3. Screening for model compounds

Purification strategies applied prior to quantitative measurements typically aim at reduction of sample complexity. Clean-up procedures use selective techniques of

Table 1
Sample description of 21 metabolic profiles generated by GC/EI-TOF-MS

Sample information				Components ($S/N \geqslant 75$)
Experiment	Species	Organ	Fresh weight (mg)	Number
1135ec05	*Arabidopsis thaliana* C24	Leaf	1	282
1135ec06	*Arabidopsis thaliana* C24	Root	2	181
1135ec07	*Nicotiana tabacum* cv. SNN	Leaf	3	295
1135ec08	*Nicotiana tabacum* cv. SNN	Root	3	226
1135ec09	*Solanum tuberosum* cv. Désirée	Leaf	1	300
1135ec10	*Solanum tuberosum* cv. Désirée	Root	3	243
1135ec13	*Solanum tuberosum* cv. Désirée	Flower	3	244
1135ec12	*Solanum tuberosum* cv. Désirée	Stolon	3	221
1135ec11	*Solanum tuberosum* cv. Désirée	Tuber	1	175
				Σ = 2167
1135ec24	*Arabidopsis thaliana* C24	Leaf	16	383
1135ec25	*Arabidopsis thaliana* C24	Root	10	327
1135ec26	*Nicotiana tabacum* cv. SNN	Leaf	11	374
1135ec27	*Nicotiana tabacum* cv. SNN	Root	10	463
1135ec28	*Solanum tuberosum* cv. Désirée	Leaf	14	503
1135ec29	*Solanum tuberosum* cv. Désirée	Root	14	366
1135ec32	*Solanum tuberosum* cv. Désirée	Flower	10	440
1135ec31	*Solanum tuberosum* cv. Désirée	Stolon	14	440
1135ec30	*Solanum tuberosum* cv. Désirée	Tuber	18	395
				Σ = 3691
1135ec03	Non-sample control [a]	Initial	Empty	119
1135ec23	Non-sample control [a]	Interspersed	Empty	76
1135ec61	Non-sample control [a]	Final	Empty	152
				Σ = 347

Mass spectral components were deconvoluted by AMDIS software. Components with signal to noise ratio <75 were removed from further analysis.

[a] Non-sample control experiments represent empty containers for sampling and storage which were fully processed in parallel with regular samples from extraction to final GC/MS analysis.

Table 2
Global description of the mass spectral and retention time index database. Sample classes are characterized by number of components (Sum), composition (%), average number (AVG), and standard deviation (SD) of components per experiment

Sample class	Experiments	Components				Signal to noise ratio	Peak purity
		Sum	%	AVG[a]	S.D.[b]	AVG[a]	AVG[a]
Total	21	6205	100.0	295	120	397	47.4
Non-sample control	3	347	5.6	116	38	267	59.7
1–3 mg (FW)	9	2167	34.9	241	46	390	51.2
10–18 mg (FW)	9	3691	59.5	410	55	447	39.4
Solanum tuberosum cv. Désirée	10	3327	53.6	333	–	–	–
Arabidopsis thaliana C24	4	1173	18.9	293	–	–	–
Nicotiana tabacum cv. SNN	4	1358	21.9	340	–	–	–
Leaf[c]	6	2137	34.4	356	–	–	–
Root[d]	6	1806	29.1	301	–	–	–
Flower[e]	2	684	11.0	342	–	–	–
Stolon[f]	2	661	10.7	331	–	–	–
Tuber[g]	2	570	9.2	285	–	–	–

The general quality is assessed by average signal to noise ratio and peak purity. Experiments constituting the classes, non-sample control, 1–3 mg (FW), 10–18 mg (FW), and the classes of plant species are indicated in Table 1.

[a] Average.

[b] Standard deviation.

[c] Experiments 1135ec05, 07, 09, 24, 26, and 28.

[d] Experiments 1135ec06, 08, 10, 25, 27, and 29.

[e] Experiments 1135ec13 and 32.

[f] Experiments 1135ec12 and 31.

[g] Experiments 1135ec11 and 30.

Table 3
Summary of the mass spectral search list obtained with the bait mass spectrum of persilylated 3-caffeoyl quinic acid

Components Identification	Number of replicates	Retention time index		Match	
		AVG[a]	S.D.[b]	AVG[a]	S.D.[b]
3-Caffeoyl quinic acid (BP)	9	2990.3	0.7	891	96
3-Caffeoyl quinic acid[d]	24[c]	3113.7	2.0	843	132
4-Caffeoyl quinic acid (BP)	7	3009.1	1.2	783	34
5-Caffeoyl quinic acid[d]	15	3190.0	0.8	772	99
4-Caffeoyl quinic acid[d]	20	3168.3	0.9	738	84
Quinic acid	23[c]	1854.9	1.0	701	72
A1	4	3096.8	0.5	671	88
A2	4	2982.8	0.4	667	108
A3	5	2851.8	0.4	644	96
A4	4	2924.5	0.3	600	70
A5	3	2936.0	0.4	589	13
A6	17	1715.2	0.7	574	32
Caffeic acid	13	2139.0	0.4	560	24
A7	20	1809.1	1.3	559	28
A8	3	2856.2	1.5	552	57
p-Coumaric acid[d]	12	1947.0	0.6	550	27
A9	4	2587.9	0.3	543	28
A10	10	1756.4	0.4	542	5
Galactonic acid	9	1991.7	0.7	533	8
A11	9	1766.5	0.4	532	20
1-Caffeoyl quinic acid	1	3397.0	–	710	–
4-Caffeoyl quinic acid (BP)[d]	2	3138.7	–	624	–
1-Caffeoyl quinic acid (BP)	1	3296.6	–	619	–
1-Caffeoyl quinic acid (BP)	1	3152.5	–	586	–
Caffeic acid (BP)	1	1985.5	–	564	–
5-Caffeoyl quinic acid (BP)	1	3007.7	–	527	–
Ascorbic acid	1	1946.4	–	527	–
Isoascorbic acid	1	1957.4	–	521	–
o-Coumaric acid	1	1821.6	–	497	–

Twenty groups of components are shown (group size ⩽3, match factor ⩽532). Identifications were confirmed manually. Tested reference substances which were not observed in plant samples are appended (BP; by-product of chemical synthesis or derivatization).

[a] Average.

[b] Standard deviation.

[c] Deconvolution of more than one component per peak.

Table 4
Summary of the mass spectral search list obtained with the bait mass spectrum of persilylated galactonic acid

Components Identification	Number of replicates	Retention time index		Match	
		AVG[a]	S.D.[b]	AVG[a]	S.D.[b]
Galactonic acid	19	1991.6	0.7	893	79
Gluconic acid	11	1997.0	1.1	849	54
B1	17	2010.9	0.5	836	54
B2	12	1986.6	0.8	820	45
Gulonic acid[d]	8	1959.2	0.7	786	52
B3	8	1766.4	0.5	783	54
B4	4	1700.2	0.1	766	11
Ribitol[e]	21	1727.1	0.3	762	12
Mannitol	15	1927.7	0.6	761	34
B5	5	1744.8	0.2	760	33
B6	11	1919.7	1.0	754	30
Threonic acid	24[c]	1559.6	0.7	754	25
B7	11	1756.3	0.4	750	41
Glucose methoxyamine (BP)	16	1908.5	0.8	746	17
B8	8	1603.1	0.4	740	21
B9	5	1972.5	0.4	738	18
B10	8	1541.0	0.4	736	32
Galactose methoxyamine	6	1883.7	0.7	734	13
B11	13	2766.1	1.3	732	27
B12	17	2105.0	0.8	731	25
B13	11	2128.0	0.3	726	25
B14	3	2290.9	0.2	722	41
Glucose methoxyamine	17	1889.8	1.4	721	21
B15	9	2041.3	0.3	720	27
Sorbitol	1	1931.1	–	763	–
Galactitol	1	1935.5	–	753	–
Glucaric acid	1	2014.0	–	743	–
Xylitol	1	1710.0	–	732	–
Ascorbic acid	1	1946.4	–	632	–
Isoascorbic acid	1	1957.4	–	637	–

Twenty four groups of components are shown (group size ⩽3; match factor ⩽720). Identifications were confirmed manually. Tested reference substances which were not observed in plant samples are appended (BP; by-product of chemical synthesis or derivatization).

[a] Average.

[b] Standard deviation.

[c] Deconvolution of more than one component per peak.

[d] Identified as by-product of D-gulonic acid gamma-lactone.

[e] Internal standard.

extraction and enrichment. Finally a small number of target compounds which exhibit unequivocal chromatographic separation are prepared and analyzed. In contrast metabolic profiling utilizes substance specific and in the context of complex samples selective means of detection. For example selective ions after mass fragmentation (Fiehn et al., 2000a; Huhman and Sumner, 2002) or characteristic chemical shifts within NMR spectra (Fan et al., 2001) are exploited in order to resolve complex chromatograms into clear substance identifications. An example of the chromatographic separation obtained by GC/EI-TOF-MS is shown in Fig. 2. In this potato leaf matrix the persilylated 3-caffeoylquinic acid derivative was accompanied by 3 co-eluting components exhibiting the unique masses $m/z=259$ and $m/z=597$. Different to scanning GC/MS chromatograms abundant and specific TOF mass traces of the 3-caffeoylquinic acid derivative, $m/z=255$, 307, and 345 (Fuchs and Spiteller, 1996), exhibited complete overlap when scaled to maximum. Even an ion like $m/z=447$, which in the case of the 3-caffeoylquinic acid derivative had a mass spectral proportion of less than 0.1%, showed deviation only of shape but not of maximum intensity of the chromatographic trace. All mass spectra mentioned in the following are implicitly of substances which were subject to methoxyamination and trimethylsilylation.

In contrast to chemical pre-purification or selected ion processing of complex GC/EI-TOF-MS profiles we introduced selectivity to our analyses by generating subsets of MS/RI library components by exploiting mass spectral similarity with single bait mass spectra. Hit lists of mass spectral searches are routinely ordered according to mass spectral match factors. Here we present the 400 highest ranking results of mass spectral searches in analogy to Fig. 1 as two dimensional plots with one axis representing RI and the second axis used to describe the respective match factor of the library component (Figs. 3 and 4).

2.3.1. Exemplary search for 3-caffeoylquinic acid

The top ranking mass spectral components extracted from the MS/RI library by comparison with the mass spectrum of 3-caffeoylquinic acid ranged from match factor 510 to the best fit of 949. Almost the complete

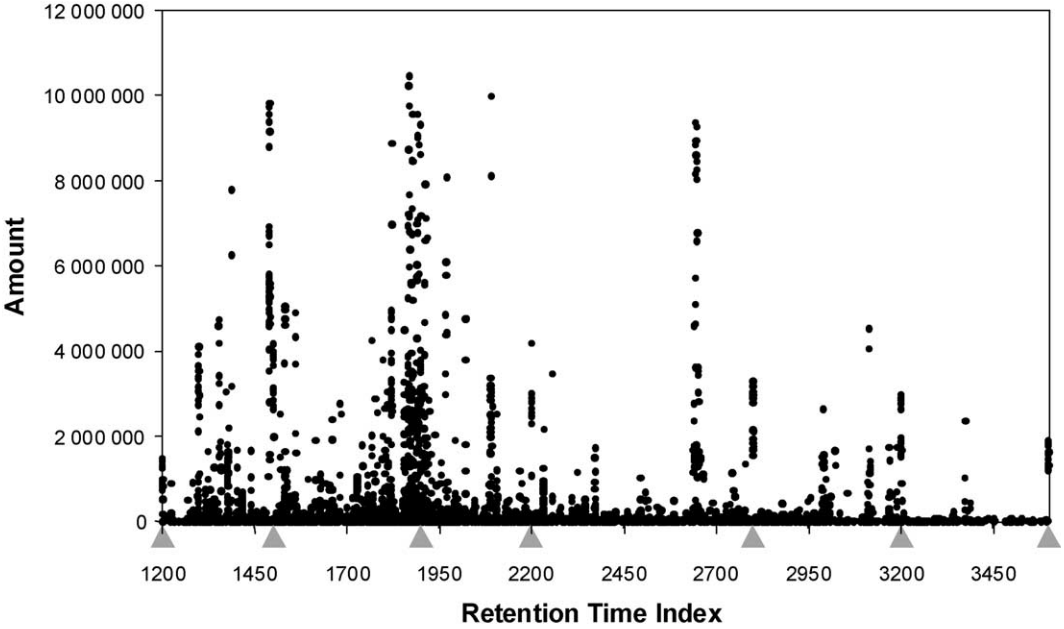

Fig. 1. Full chromatographic reconstruction of all mass spectral components comprising the mass spectral and retention time index database. Arrows indicate the chromatographic position of the retention time standards, *n*-dodecane (RI 1200), *n*-pentadecane (RI 1500), *n*-nonadecane (RI 1900), *n*-docosane (RI 2200), *n*-octacosane (RI 2800), *n*-dotriacontane (RI 3200), and *n*-hexatriacontane (RI 3600).

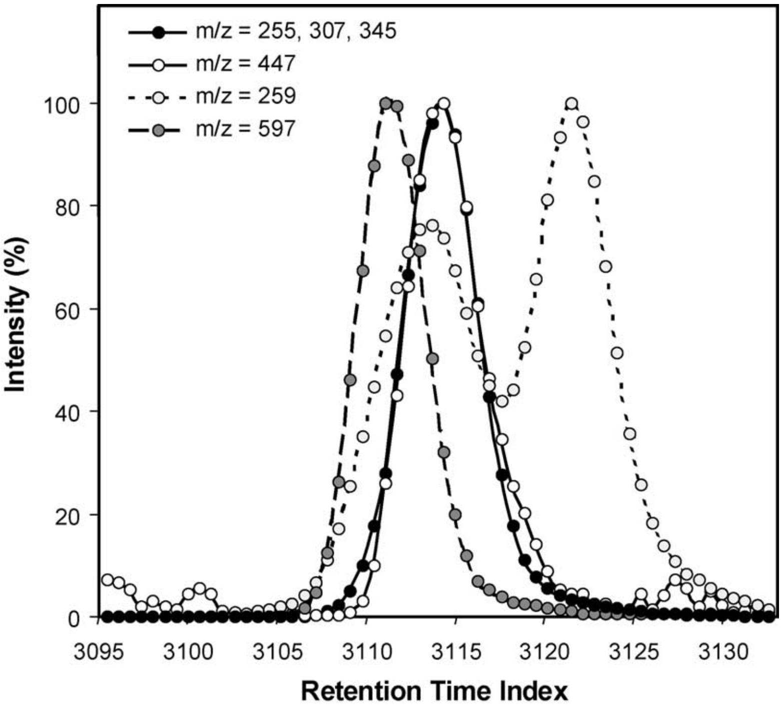

Fig. 2. Partial selected ion chromatogram of persilylated 3-caffeoylquinic acid ($m/z = 255$, 307, 345, 447) and co-eluting components ($m/z = 597$ and 259) from potato leaf. Ion intensity was scaled to maximum within the selected RI window. Different to scanning GC/MS chromatograms the single TOF mass traces of $m/z = 255$, 307, and 345 exhibited complete overlap. The trace $m/z = 447$ exemplifies a specific fragment with a mass spectral proportion $< 0.1\%$. The figure shows GC/EI-TOF-MS experiment 1135ec28 (refer to Table 1).

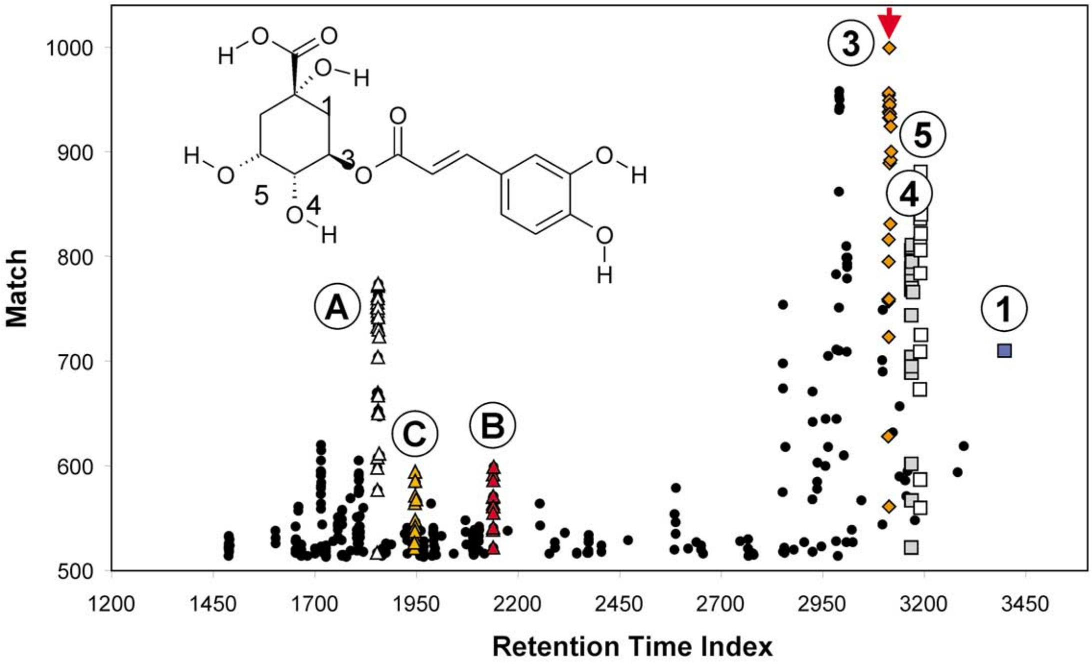

Fig. 3. Chromatographic reconstruction of the mass spectral search list of plant components and reference substances obtained with the bait mass spectrum of persilylated 3-caffeoylquinic acid (arrow). The search list contained mass spectra of the siylated derivatives of the monomers as well as positional isomers of the 3-caffeoylquinic acid conjugate. A (open triangle) quinic acid, B (red triangle) caffeic acid, C (orange triangle) *p*-coumaric acid, 1 (blue box) 1-caffeoylquinic acid, 3 (orange diamond) 3-caffeoylquinic acid, 4 (grey box) 4-caffeoylquinic acid, and 5 (open box) 5-caffeoylquinic acid. Replicate mass spectra from plant samples were confirmed manually. Silylated protons and ester bond position of the respective isomers are indicated on the inset chemical structure.

range of RI was covered. Twenty-three mass spectral components from plant samples were detected and manually confirmed to represent 3-caffeoylquinic acid. This compound was found in all samples of the solanaceous species and in root samples of *Arabidopsis thaliana*. The average match factor of this group of mass spectral components was 843±132 with RI = 3113.7±2.0 (Table 3). Redundant deconvolution of single chromatographic peaks was found to be insufficiently repressed. However, this error appeared to be irrelevant or even beneficial for the positive identification of plant components.

The complete hit list was dissected into groups of components which were separated by RI gaps >2.0. Thirty-nine of these groups had more than two members. The components of these groups were compared with reference profiles of commercially available quinic acid, caffeic acid, *o*-coumaric acid, *p*-coumaric acid, 3-caffeoylquinic acid, and synthetic preparations of the three isomers 1-caffeoylquinic acid, 4-caffeoylquinic acid, and 5-caffeoylquinic acid. Main derivatives (Fig. 3) and by-products were taken into account. Eight of the top ranking groups of components were identified to represent substances which were chemically related to the initial bait, 3-caffeoylquinic acid. In detail quinic acid, caffeic acid, *p*-coumaric acid, 3-caffeoyl-, 4-caffeoyl-, and 5-caffeoylquinic acid were found in addition to respective by-products of both commercial or synthetic preparations (Table 3). Both 4-caffeoylquinic acid and 5-caffeoylquinic acid were expected to be present in potato (Griffiths and Bain, 1997; Percival and Baird, 2000) and tobacco tissue (Koeppe et al., 1969; Baumert et al., 2001) next to the main isomer 3-caffeoylquinic acid. The deconvoluted GC/EI-TOF mass spectra of the silylated main products of 3-caffeoyl-, 4-caffeoyl-, and 5-caffeoylquinic acid were confirmed by published data including the proportions of the key ions $m/z = 307$, 345, 447 (Fuchs and Spiteller, 1996). The TOF mass detector, however, generated characteristic mass spectra with increased proportions of ions with low m/z and reduced proportions of high m/z (Supplementary file 2) as compared with mass spectra generated by high resolution sector field instruments (Fuchs and Spiteller, 1996).

Galactonic acid, ascorbic acid, and isoascorbic acid exhibited mass spectral match factors = 521–532. We could therefore empirically deduce a threshold of 500–550 at which cross-matching of mass spectral components may occur. Eleven groups of components, A1–A11, with average mass spectral match $\geqslant 532$ remained non-identified. An initial manual inspection of the fragmentation pattern and comparison with the commercial NIST98 mass spectra collection allowed no further identification. However, unknown components

A1, A2, A3, A4, A5, and A8 appeared to be similar to quinic acid or caffeoylquinic acids (refer to Section 2.4).

These results strongly support the observation that mass spectral hit lists may reveal previously unknown components from plant tissues which are chemically related to the respective bait substance.

2.3.2. *Exemplary search for galactonic acid*

In the first application of the MS/RI library caffeoylquinic acid conjugates were chosen because of their highly characteristic fragmentation pattern, which is caused by the presence of two cyclic structural moieties (Fuchs and Spiteller, 1996). In contrast galactonic acid contained only functional groups which are ubiquitous in natural products, for example, carboxyl-groups, primary, and secondary hydroxyl-groups. Furthermore, components exhibiting highly similar mass spectra co-eluted within a small RI window (inset of Fig. 4). Accordingly the range of match factors among the top 400 matching components was only 632–948. Eighteen mass spectral components representing the galatonic acid derivative from plant samples were found and manually confirmed in all but tobacco leaf samples. The average match factor of this group of mass spectral components was 893±79 and RI = 1991.6±0.7 (Table 4). In total, 36 groups of components with group size $\geqslant 3$ were found.

Among the groups of components exhibiting the best mass spectral fit we identified two isomers of galactonic acid, namely the plant products gluconic acid and gulonic acid (Table 4). In addition we found that linear sugar alcohols had high mass spectral match factors = 732–763. Of the reference substances, ribitol, xylitol, mannitol, sorbitol, and galactitol, only ribitol and mannitol were identified within the MS/RI library. Ribitol was an internal standard added to all samples, whereas as mannitol originated from plant samples. Mass spectra of glucose methoxyamine and galactose methoxyamine were also among the high ranking hits and exhibited match factors = 721–746. Methoxyamine moieties are introduced by routine chemical derivatization of carbonyl-groups or cyclic actales and ketales prior to GC/MS profile analysis. Typically a main and a by-product of methoxyamination are found. Finally we identified the plant product threonic acid with match factor = 754±25 within the hit list and detected high similarity to the reference substance glucaric acid, match factor = 743. Fifteen groups of components, B1–B15, remained non-identified. Initial comparison with the commercial NIST98 mass spectra collection allowed no further identification but the components B1, B2, and B3 exhibited match factors $\geqslant 800$ when compared with either gulonic, galactonic, or gluconic acid (see Section 2.4).

Again this case exemplified that unknown components can be identified and that isomers or chemically related compounds can be detected among the top ranking mass spectral components of our MS/RI-

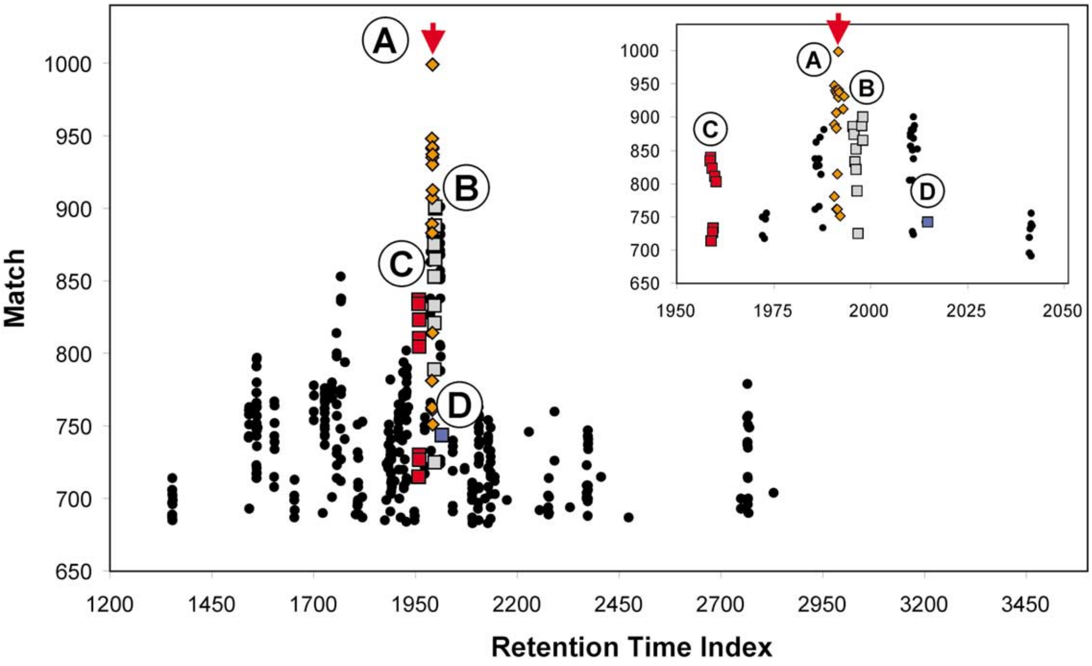

Fig. 4. Chromatographic reconstruction of the mass spectral search list of plant components and reference substances obtained with the bait mass spectrum of persilylated galactonic acid (arrow). The search list contained mass spectra of the siylated derivatives of A (orange diamonds) galactonic acid, B (grey box) gluconic acid, C (red box) gulonic acid, and D (blue box) glucaric acid. Replicate mass spectra from plant samples were confirmed manually. The inset shows RI range 1950—2050.

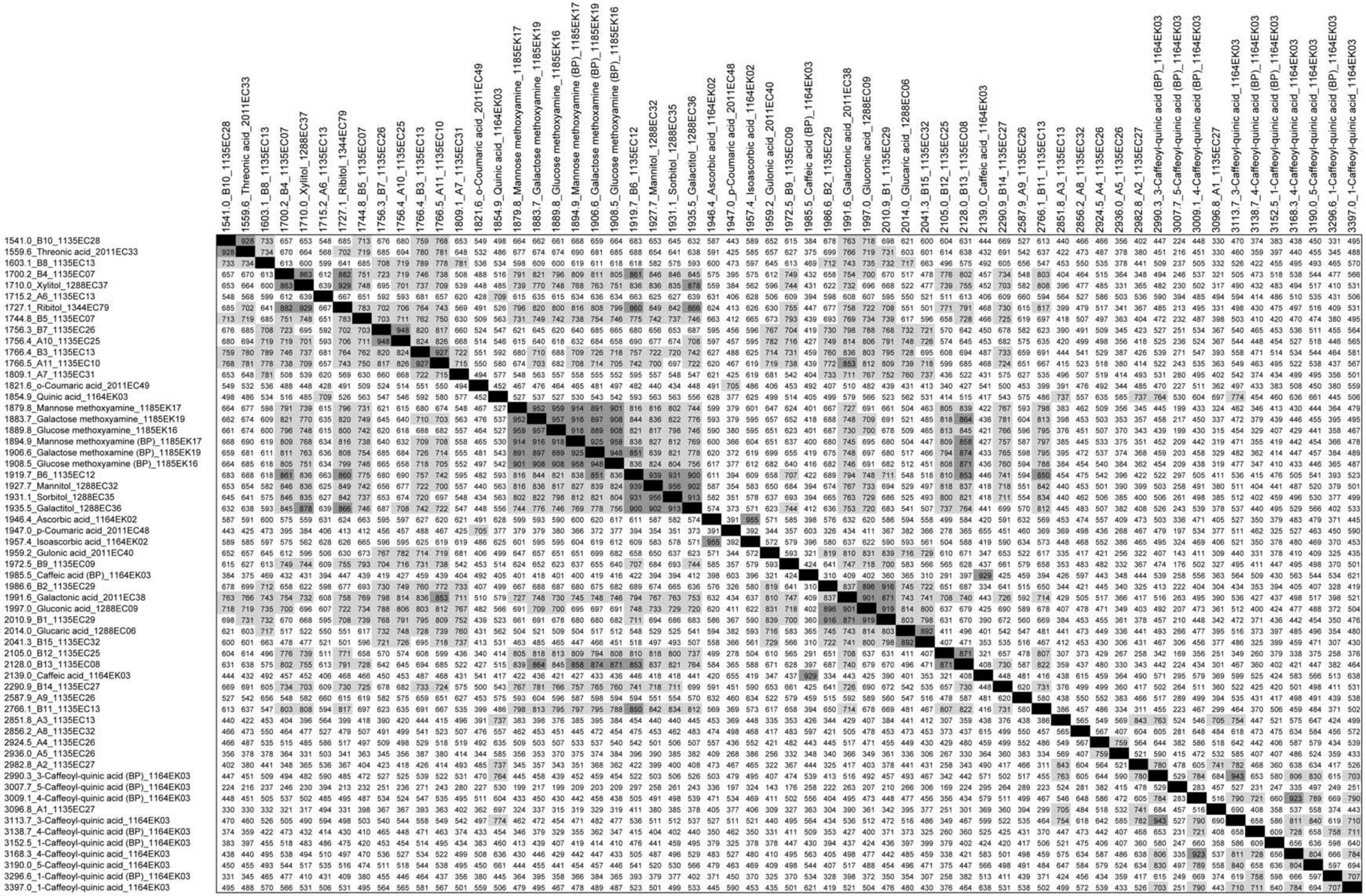

Fig. 5. Complete matrix of pair-wise mass spectral comparisons. Mass spectral components are in chromatographic order. Representative mass spectra of reference substances and unknown components from Tables 3 and 4 (refer to supplemental data 1 and 2) were compared by the NIST98 search algorithm. Match factors describe a range of 1–999, where the maximum value indicates perfect mass spectral identity. Match quality was indicated by grey shades: black, 999 (self matching); dark grey, 850–999; light grey, 700–850.

library. In this example we found in addition to other hexonic acids, $CH_3(HCOH)_4COOH$, compounds like C_5- or C_6-sugar alcohols and methoxyaminated C_6-aldoses, which all contain a straight chain pentitol group, $CH_3(HCOH)_4$-R. Moreover, we identified two matching carboxylic acids with the common structural feature, R-$(HCOH)_3COOH$.

2.4. *Classification of unknown mass spectral components*

The analysis of the MS/RI library resulted in newly identified substances present within metabolic profiles of plants, however, 11 groups of components with mass spectral similarity to 3-caffeoylquinic acid and 15 groups of components exhibiting similarity to galactonic acid

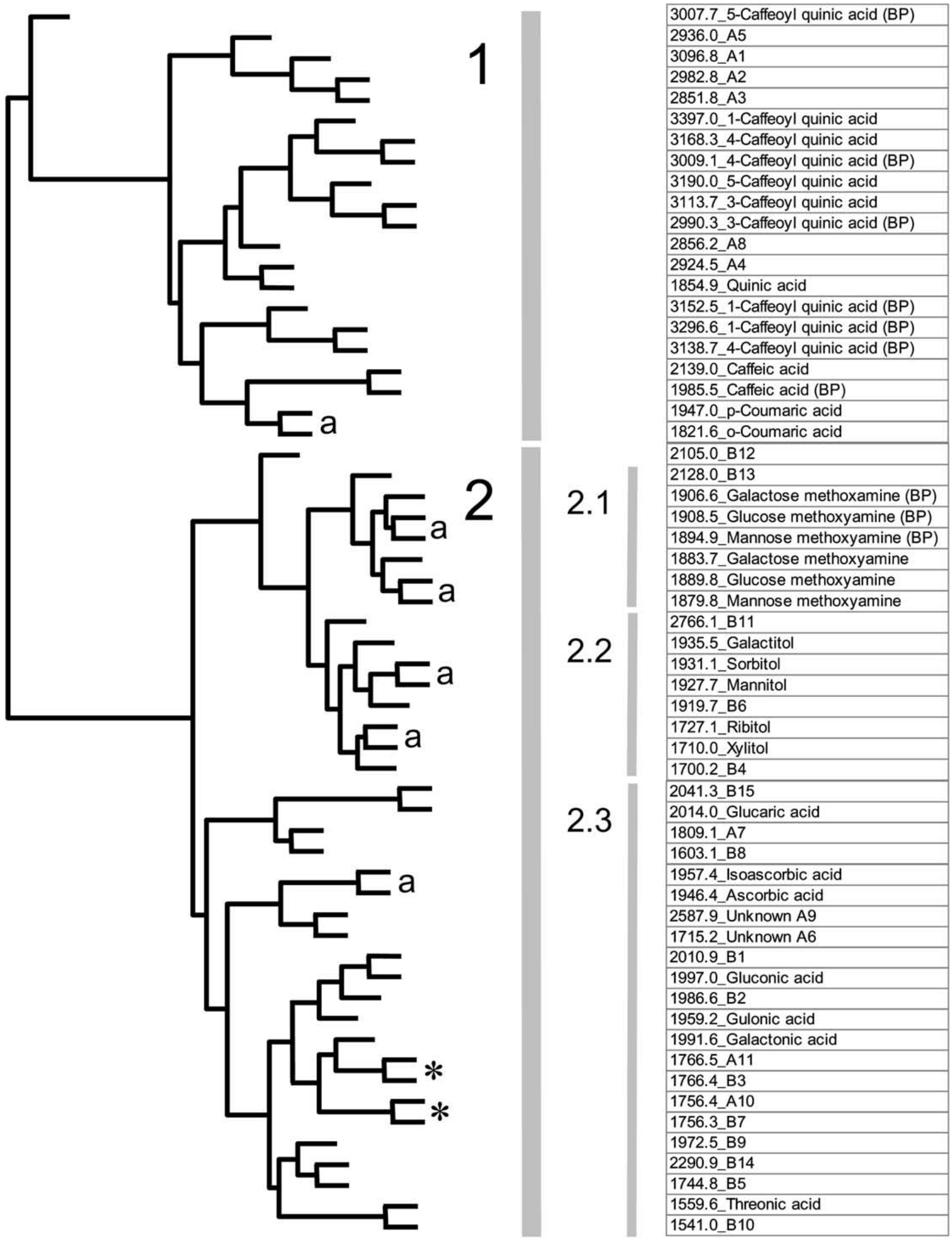

Fig. 6. Hierarchical cluster analysis of mass spectral components based on profiles of mass spectral match factors. The analysis was based on the complete matrix of match factors presented in Fig. 5 (also refer to supplemental data 1). Class 1 comprised cinnamic acid derivatives, quinic acid and respective conjugates. Class 2 was composed of carbohydrates which subgrouped into hexose methoxyamines (2.1), sugar alcohols (2.2), and poly-hydroxy-carboxylic acids (2.3). ^{a} Pairs of conformational isomers. * Redundant mass spectra of identical components.

remained non-identified. We selected representative mass spectra of all groups of unknown components and of available reference substances. The complete set of 59 mass spectra is made available as supplemental file 1 in a data format ready to be imported either into AMDIS software or into the NIST98 mass spectral search program (Appendix). A complete set of 59×59 pair-wise mass spectral comparisons was generated by the NIST98 algorithm (Fig. 5, also see supplemental data 2). Inverse comparisons had identical match factors. The comparison matrix of match factors indicated two main classes of mass spectra with partial chromatographic separation (Fig. 5). A hierarchical cluster analysis based on the comparison matrix of mass spectral match factors generated the tree of nearest neighbours presented in Fig. 6. In contrast to established procedure this classification did not only take best fits into account. We argue that best fits were clearly insufficient for the distinction of groups of redundant components (Tables 3 and 4). In contrast the cluster analysis based on comparison matrices of match factors reflected all available similarities as well as dissimilarities. Among the 20 pairs of nearest neighbours six pairs of conformational isomers and two pairs of identical but unknown components were found, for example, A11/B3, and A10/B7. Identity was confirmed manually by MS and RI comparison of all members of the respective groups from Tables 3 and 4. Furthermore, the resulting tree analysis revealed two main classes of mass spectra (Fig. 6). Class 1 comprised cinnamic acid derivatives, quinic acid and conjugates thereof. Carbohydrates constituted class 2, which was subdivided into hexose methoxyamines, subclass 2.1, sugar alcohols, subclass 2.2, and polyhydroxy-carboxylic acids, subclass 2.3, respectively.

Further analysis revealed that the positional isomers of caffeoylquinic acid were aggregated within a single branch of class 1. The 6 non-identified components, A1, A2, A3, A4, A5, and A8, were classified to belong to class 1 and component A4 could be put into close proximity with quinic acid. All other unknown components grouped with class 2 (Fig. 6). In detail, component B6 was suggested to represent a hexitol by proximity to galactitol, sorbitol and mannitol. Component B4 exhibited proximity to pentitols, for example, ribitol and xylitol. Component B15 was closest neighbour of glucaric acid and component B10 matched best with threonic acid. Components B1 and B2 were grouped with the hexonic acids, gluconic and gulonic acid. The component pairs A11/B3 and A10/B7 were also close to hexonic acids as was indicated by occurrence of galactonic acid within the same major branch.

3. Conclusion

Our present work clearly showed the feasibility of non-supervised construction of MS/RI libraries from automatically generated mass spectral components of metabolic profiles. We demonstrated unequivocally that RI was absolutely essential for identification of unknown metabolic components by reference substances and for the grouping of non-identified but redundant mass spectral components. We furthermore demonstrated that selectivity of analysis can be introduced via creation of subsets of mass spectral components which exhibit similarity to single reference mass spectra. Finally we introduced cluster analysis based on matrices of mass spectral match factors and clearly demonstrated applicability of this approach to the classification of unknown mass spectral components. Future efforts will focus on automated error detection of mass spectral deconvolution and the reduction of the initial signal to noise ratio threshold, which currently limits the virtual sensitivity of our MS/RI libraries and thus restricts access to the full sensitivity of GC/EI-TOF-MS (Hirsch et al., 2001; Dalluge et al., 2002). Finally we will attempt a comprehensive classification of full mass spectral libraries comprising non-identified natural products and an extended set of reference substances.

4. Experimental

4.1. Biological materials and sampling

The plant varieties used were potato (*Solanum tuberosum* cv. Désirée), tobacco (*Nicotiana tabaccum* cv. SNN), and *Arabidopsis thaliana* (L.) Heynh., ecotype C24. Potato (*Solanum tuberosum* cv. Désirée) was obtained from Saatzucht Lange AG (Bad Schwartau, Germany). All plants were cultivated on soil in growth chambers with a maximum of 120 μmol photons m^{-2} s^{-1} at leaf surface. Potato and tobacco plants were grown in 3-l pots with a 16 h-light/8 h-dark regime changing from 22 °C during the day to 18 °C at night and with relative humidity preset to constant 70%. *Arabidopsis thaliana* plants were kept in 0.125-l pots with an identical regime of illumination but changing from 60% humidity and 20 °C during the day to 75% humidity and 18 °C at night. Samples of plant organs were harvested during light periods from flowering plants. Root samples were prepared free of soil but not rinsed with water. Plant material was weighed with a precision of ±0.1 mg, immediately frozen in liquid nitrogen, and stored at −70 °C until further analysis. Sampling time before freezing did not exceed 20 s.

4.2. Reference substances and substance identification

L-Threonic acid calcium salt (CAS 70753-61-6) was purchased from Aldrich, Germany; D-(−)-isoascorbic acid (CAS 89-65-6) was from Fluka, Germany;

D-(+)-mannose (CAS 3458-28-4) was ordered from Merck, Germany; L-ascorbic acid (CAS 50-81-7), caffeic acid (CAS 331-39-5), 3-caffeoylquinic acid (CAS 327-97-9), *o*-coumaric acid (CAS 614-60-8), *p*-coumaric acid (CAS 501-98-4), D-(−)-galactonic acid gamma-lactone (CAS 2782-07-2), D-galactonic acid hemicalcium salt (CAS 6622-52-2), D-(+)-galactose (CAS 59-23-4), D-(+)-glucose (CAS 50-99-7), D-glucaric acid monopotassium salt (CAS 576-42-1), D-gluconic acid sodium salt (CAS 527-07-1), D-(+)-gluconic acid delta-lactone (CAS 90-80-2), D-gulonic acid gamma-lactone (CAS 6322-07-2), D-(−)-quinic acid (CAS 77-95-2), ribitol (CAS 488-81-3), and xylitol (CAS 87-99-0), were purchased from Sigma-Aldrich, Germany; D-galactitol (CAS 608-66-2), D-mannitol (CAS 69-65-8), and D-sorbitol (CAS 50-70-4), were from Supelco, Germany.

Solvents and reagents for extraction and derivatization were as follows: methoxyamine hydrochloride was purchased from Sigma-Aldrich, Germany; *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA) was from Macherey & Nagel, Germany; pyridine, methanol, and chloroform, all HPLC-grade, were supplied by J.T. Baker (Philipsburg, NJ).

Gulonic acid was not commercially available. This substance was identified as by-product of D-gulonic acid gamma-lactone which was subjected to the methoxyamination and persilylation protocol used for metabolic profiling (see later) and was incubated at room temperature for at least 2 days before GC/MS analysis. In support of this means of identification galactonic acid and gluconic acid were found within profiles generated in parallel from D-(−)-galactonic acid gamma-lactone and D-(+)-gluconic acid delta-lactone, respectively (data not shown).

The isomers of chlorogenic acid (3-caffeoylquinic acid), 4-caffeoylquinic acid, 5-caffeoylquinic acid, and 1-caffeoylquinic acid were synthesized as described earlier (Sefkow, 2001; Sefkow et al., 2001). The final purities of the preparations were at least 90% except for 1-caffeoylquinic acid. By-products occurring within these preparations were used for comparison and classification of components observed within profiles of natural products. Chemical abstracts system (CAS) registry numbers of the reference substances are provided if available.

4.3. GC/EI-TOF-MS metabolite profiling

Extraction, liquid partitioning, concentration to dryness, and methoxyamination of carbonyl-moieties followed by derivatizing acidic protons with *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA) prior to GC/EI-TOF-MS analysis was performed essentially as described (Fiehn et al., 2000b). For this study only the polar, methanol/water (1:1; v:v) soluble material of the initial extract was analyzed. Highly lipophilic components were removed by liquid partitioning into chloroform and discarded. The initial protocol was downscaled by a factor of 3.3 in order to accommodate sample loads of 1–60 mg fresh weight. The final reagent volume was 120 μl of pyridine/MSTFA (1:2; v:v). A mixture of the retention time standards, *n*-dodecane (RI 1200), *n*-pentadecane (RI 1500), *n*-nonadecane (RI 1900), *n*-docosane (RI 2200), *n*-octacosane (RI 2800), *n*-dotriacontane (RI 3200), and *n*-hexatriacontane (RI 3600) was included in the final reagent volume. Non-sample control experiments were performed with empty containers used for sampling and storage of plant material. The empty containers were fully processed in parallel with regular samples from extraction to final GC/MS analysis. Non-sample controls were run at the start, initial control, and end, final control, of a series of GC/MS analyses as well as interspersed. Thus initial contamination and GC/MS memory effect were monitored.

A GC 6890 (Agilent Technologies, Palo Alto, CA, USA) was operated under electronic pressure control and equipped with a split/splitless capillary inlet. Injection was 1 μl in the splitless mode with a 2 min pulse at 110 psi and injection temperature set to 230 °C. The capillary column used was a 30 m×0.25 mm inner diameter Rtx-5Sil MS with integrated guard column and a 0.25 μm film (Restek GmbH, Bad Homburg, Germany). Helium was used as carrier gas with constant flow at 1 ml min^{-1}. The temperature program was 2 min at 80 °C followed by a 15 min ramp to 350 °C and final heating for 2 min at 350 °C. The transferline to the mass spectrometer was set to 250 °C.

The time-of-flight (TOF) mass spectrometer was a Pegasus II MS system (Leco, St. Joseph, MI, USA) with an electron impact ionization (EI) source set to 200 °C. Mass spectra were monitored with an acquisition rate of 6 spectra s^{-1} in the mass range $m/z = 70$–600. Tuning and all other settings of the mass spectrometer were according to manufacturer's recommendations.

4.4. Data processing

Chromatograms were acquired with ChromaTOF™ software (LECO, St. Joseph, MI, USA). Initial processing, namely baseline substraction, smoothing, and export of the processed chromatograms into a *.cdf file interchange format were performed within the ChromaTOF™ software. Retention index calibration and mass spectral deconvolution were performed by AMDIS (Automated Mass Spectral Deconvolution and Identification System, National Institute of Standards and Technology, Gaithersburg, MD, USA) with the settings: adjacent peak subtraction = 2, medium resolution, high sensitivity, and high shape requirements. Those of the resulting components which had at least 0.001% of the total signal of the respective chromatogram were exported into a *.msp library file. The component names within the *.msp files were edited in order to include sample information which was not automatically exported from AMDIS. The edited component identifier included retention time, retention time index, signal to noise ratio, amount of component, experiment identifier, species, organ, and amount of fresh weight. Finally all edited *.msp files were imported into a combined mass spectral library within the NIST98 mass spectral search program (National Institute of Standards and Technology, Gaithersburg, MD, USA). Mass spectral comparisons were performed in the normal NIST98 identity mode without pre-search or other constraints. Due to the NIST98 software the resulting mass spectral hit lists were restricted to the top 400 best fitting mass spectral components. The hit lists were processed and evaluated by the software package for exploratory data analysis and statistical modelling, S-Plus 2000 standard edition release 3 (Insightful, Berlin, Germany).

Data processing with the aim to classify unknown mass spectral components was performed as follows. A mass spectral library of 59 representative identified and unknown components was compiled from GC/EI-TOF-MS profiles of reference substances and plant matrices. Each of the single components was compared separately with the complete library. The 59 resulting mass spectral hit lists were combined into a single 59×59 comparison matrix (Fig. 5) using the match factor as sole measure of mass spectral similarity. In addition to the mass spectral match factor the NIST98 software reported a reverse match factor and a probability measure of mass spectral identity. We compared the respective values of initial and inverse mass spectral comparisons of all pairs of mass spectra. Only the match factor exhibited identical values for all bi-directional, pair-wise comparisons. The reverse match factor and probability measure were disregarded for further analysis. The hierarchical cluster analysis of the comparison matrix was agglomerative using the euclidian dissimilarity measure and average linkage.

Acknowledgements

The authors thank Dr. Alisdair Fernie and Professor Lothar Willmitzer, Max-Planck-Institute of Molecular Plant Physiology, Golm, Germany, for valuable discussions, editorial suggestions and support of our work.

Appendix. Supplementary data

Supplementary file 1

Data sheet constituting Fig. 5 and used as basis for the hierarchical cluster analysis of mass spectra which is presented in Fig. 6.

The data set describes the complete matrix of pair-wise mass spectral comparisons generated by the NIST98 search algorithm. Mass spectral components are in chromatographic order on both axes. Match factors describe a range of 1–999, where the maximum value indicates perfect mass spectral identity. Match quality was indicated by colour: black, 999 (self matching); orange, 850–999; yellow, 700 - 850. Representative mass spectra of reference substances and unknown components from Tables 3 and 4 may be downloaded from supplemental file 2.

Supplementary file 2

The datafile, MS_RI.msp[1] contains representative GC/EI-TOF-MS mass spectra of all reference substances and unknown components mentioned within Tables 3 and 4.

The spectrum name was designed to allow sorting according to retention time index, e.g. 1766.4_B3_1135EC13 or 3113.7_3-Caffeoylquinic acid_1164EK03. The name of the mass spectrum contains three types of information separated by (_). The first position denotes retention time index, the second position indicates name of reference substance[2] or unknown component (refer to Tables 3 and 4), third position encodes the experiment identifier.[3]

References

Baumert, A., Mock, H.-P., Schmidt, J., Herbers, K., Sonnewald, U., Strack, D., 2001. Patterns of phenylpropanoids in non-inoculated and potato virus Y-inoculated leaves of transgenic tobacco plants expressing yeast-derived invertase. Phytochemistry 56, 535–541.

Boros, L.G., Cascante, M., Lee, W.N.P., 2002. Metabolic profiling of cell growth and death in cancer: applications in drug discovery. Drug Discovery Today 7, 364–372.

Dalluge, J., Vreuls, R.J.J., Beens, J., Brinkman, U.A.T., 2002. Opimization and characterization of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection (GC×GC-TOF-MS). Journal of Separation Science 25, 201–214.

Duez, P., Kumps, A., Mardens, Y., 1996. GC-MS profiling of urinary organic acids evaluated as a quantitative method. Clinical Chemistry 42, 1609–1615.

Dyson, N., 1999. Peak distortion, data sampling errors and the integrator in the measurement of very narrow chromatographic peaks. Journal of Chromatography A 842, 321–340.

---

[1] The file format *.msp can be imported into NIST98 mass spectral comparison software (to be downloaded from http://chemdata.nist.gov/mass-spc/Srch_v1.7/index.html or AMDIS software (to be downloaded from http://chemdata.nist.gov/mass-spc/amdis/).

[2] By-products observed in preparations of reference substances were marked (BP).

[3] Experiments representing metabolic profiles of plant samples were listed in Table 1. Experiments with reference substances were indicated within the mass spectral name but were not further mentioned in this work.

Fan, T.W.-M., Lane, A.N., Shenker, M., Bartley, J.P., Crowley, D., Higashi, R.M., 2001. Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS. Phytochemistry 57, 209–221.

Fiehn, O., 2002. Metabolomics—the link between genotypes and phenotypes. Plant Molecular Biology 48, 155–171.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000a. Metabolite profiling for plant functional genomics. Nature Biotechnology 18, 1157–1161.

Fiehn, O., Kopka, J., Trethewey, R.N., Willmitzer, L., 2000b. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Analytical Chemistry 72, 3573–3580.

Frenzel, T., Miller, A., Engel, K.H., 2002. Metabolite profiling—a fractionation method for analysis of major and minor compounds in rice grains. Cereal Chemistry 79, 215–221.

Fuchs, C., Spiteller, G., 1996. Rapid and easy identification of isomers of coumaroyl- and caffeoyl-D-quinic acid by gas chromatography mass spectrometry. Journal of Mass Spectrometry 31, 602–608.

Griffin, J.L., Williams, H.J., Sang, E., Clarke, K., Rae, C., Nicholson, J.K., 2001. Metabolic profiling of genetic disorders: a multitissue H-1 nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic disorders. Analytical Biochemistry 293, 16–21.

Griffiths, D.W., Bain, H., 1997. Photo-induced changes in the concentrations of individual chlorogenic acid isomers in potato (*Solanum tuberosum*) tubers and their complexation with ferric ions. Potato Research 40, 307–315.

Hirsch, R., Ternes, T.A., Bobeldijk, I., Weck, R.A., 2001. Determination of environmentally relevant compounds using fast GC/TOF-MS. Chimia 55, 19–22.

Huhman, D.V., Sumner, L.W., 2002. Metabolic profiling of saponins in *Medicago sativa* and *Medicago trunculata* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59, 347–360.

Koeppe, D.E., Rohrbaugh, L.M., Wender, S.H., 1969. The effect of varying u.v. intensities on the concentration of scopolin and caffeoylquinic acids in tobacco and sunflower. Phytochemistry 8, 889–896.

Kuhara, T., 2001. Diagnosis of inborn errors of metabolism using filter paper urine, urease treatment, isotope dilution and gas chromatography–mass spectrometry. Journal of Chromatography B 758, 3–25.

Matsumoto, I., Kuhara, T., 1996. A new chemical diagnostic method for inborn errors of metabolism by mass spectrometry—rapid, practical, and simultaneous urinary metabolites analysis. Mass Spectrometry Reviews 15, 43–57.

Percival, G.C., Baird, L., 2000. Influence of storage upon light-induced chlorogenic acid accumulation in potato tubers (*Solanum tuberosum* L.). Journal of Agricultural and Food Chemistry 48, 2476–2482.

Sefkow, M., 2001. First efficient synthesis of chlorogenic acid. European Journal of Organic Chemistry 6, 1137–1141.

Sefkow, M., Kelling, A., Schilde, U., 2001. First efficient syntheses of 1-, 4-, and 5-caffeoylquinic acid. European Journal of Organic Chemistry 14, 2735–2742.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., Willmitzer, L., 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. Plant Journal 23, 131–142.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., Fernie, A.R., 2001a. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13, 11–29.

Roessner, U., Willmitzer, L., Fernie, A.R., 2001b. High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. Plant Physiology 127, 749–764.

van Deursen, M.M., Beens, J., Janssen, H.-G., Leclercq, P.A., Cramers, C.A., 2000. Evaluation of time-of-flight mass spectrometric detection for fast gas chromatography. Journal of Chromatography A 878, 205–213.

Veriotti, T., Sacks, R., 2000. High speed GC/MS of gasoline-range hydrocarbon compounds using a pressure-tuneable column ensemble and time-of-flight detection. Analytical Chemistry 72, 3063–3069.

Veriotti, T., Sacks, R., 2001. High-speed GC and GC/ time-of-flight MS of lemon and lime oil samples. Analytical Chemistry 73, 4395–4402.