Journal of Chromatography B, 871 (2008) 182–190

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Retention index thresholds for compound matching in GC–MS metabolite profiling☆

Nadine Strehmel, Jan Hummel, Alexander Erban, Katrin Strassburg, Joachim Kopka*

Max Planck Institute of Molecular Plant Physiology, Department Prof. L. Willmitzer, Am Muehlenberg 1, D-14476 Potsdam-Golm, Germany

ARTICLE INFO

Article history:
Received 7 February 2008
Accepted 14 April 2008
Available online 8 May 2008

Keywords:
Retention index matching
Gas chromatography
GC–MS
Metabolite profiling
Metabolomics

ABSTRACT

The generation of retention index (RI) libraries is an expensive and time-consuming effort. Procedures for the transfer of RI properties between chromatography variants are, therefore, highly relevant for a shared use. The precision of RI determination and accuracy of RI transfer between 8 method variants employing 5%-phenyl-95%-dimethylpolysiloxane capillary columns was investigated using a series of 9 *n*-alkanes (C_{10}–C_{36}). The precision of the RI determination of 13 exemplary fatty acid methyl esters (C_8 ME–C_{30} ME) was 0.22–0.33 standard deviation (S.D.) expressed in RI units in low complexity samples. In the presence of complex biological matrices this precision may deteriorate to 0.75–1.11. Application of the previously proposed Kováts, van den Dool or 3rd–5th order polynomial regression algorithms resulted in similar precision of RI calculation. For transfer of empirical van den Dool-RI properties between the chromatography variants 3rd order regression was found to represent the minimal necessary assumption. The range of typical regression coefficients was $r^2 = 0.9988$–0.9998 and accuracy of RI prediction between chromatography variants varied between 5.1 and 19.8 (0.29–0.69%) S.D. of residual RI error, $RI_{predicted} - RI_{determined}$ ($n > 64$). Accuracy of prediction was enhanced when subsets of chemically similar compound classes were used for regression, for example organic acids and sugars exhibited 0.78 ($n = 29$) and 3.74 ($n = 37$) S.D. of residual RI error, respectively. In conclusion, we suggest use of percent RI error rather than absolute RI units for the definition of matching thresholds. Thresholds of 0.5–1.0% may apply to most transfers between chromatography variants. These thresholds will not solve all matching ambiguities in complex samples. Therefore, we recommend co-analysis of reference substances with each GC–MS profiling experiment. Composition of these defined reference mixtures may best approximate or mimic the quantitative and qualitative composition of the biological matrix under investigation.

1. Introduction

Gas chromatography hyphenated to mass spectrometry (GC–MS) is one of the most versatile and widely applied technology platforms in modern metabolomic and fluxomic studies. Post-genomic molecular physiology increasingly utilises metabolic phenotyping approaches on the quest towards systems biology [1–6]. In recent years standardisation of qualitative and quantitative aspects of these high-throughput analyses has been discussed and minimum laboratory and reporting standards were proposed [7–12]. This study aims to contribute to this ongoing process. We explored the use of retention index (RI) properties for the matching of compound identities in routine GC–MS based metabolite profiling experiments (e.g. [13,14]).

The potential of mass spectral matching to commercial libraries, such as the NIST (http://chemdata.nist.gov/mass-spc/Srch_v1.7/index.html) [15–17] and the Wiley (http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470047860.html) collections was recognised in early profiling studies. Mass spectral matching was shown to be a highly useful and necessary criterion for metabolite identification. However, mass spectral matching alone was found insufficient for non-ambiguous identification, the major obstacle being the presence of multiple structural isomers in highly complex biological samples. As a consequence, RI information based on *n*-alkanes was suggested as an additional supporting criterion for compound matching and recognition (e.g. [18,19]). Moreover, the last update of the NIST05 mass spectral library comprised empirically determined RI information and an implementation of automated RI prediction [20,21]. Subsequently, mass spectral and retention index libraries, which were

☆ This paper is part of a special volume entitled "Hyphenated Techniques for Global Metabolite Profiling", guest edited by Georgios Theodoridis and Ian D. Wilson.

* Corresponding author. Tel.: +49 331 567 8262; fax: +49 331 567 898262.
E-mail address: kopka@mpimp-golm.mpg.de (J. Kopka).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.04.042

dedicated to the analysis of the typically derivatised, methoxymated and trimethylsilylated components of routine metabolite profiling experiments, have been collected [22]. The results of these efforts were made available to the metabolite profiling community through the Golm Metabolome Database (GMD, http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html) [23].

In parallel, a software tool for the pre-processing of GC–MS based profiling experiments was developed [24]. This tool utilises both mass spectral and RI properties for compound identification. With this software tool in place and a dedicated metabolite profiling library available, which comprises contributions from multiple chromatography variants, we investigated both the precision of empirical RI determination and the potential of transferring RI properties between system variants. A previous study on the prediction of RI properties grouped information from either polar or non-polar chromatography systems. Median RI prediction errors of 65 (3.9%) and 46 (3.2%), respectively, were achieved, when chemical group contributions of compounds were considered [21]. On the other hand an initial test within GMD indicated that median predictability may be as good as ±4.81 RI units, if RI was determined using equivalent polar phases [19]. Silylation was reported to mask the functionality of substituent moieties and may allow polyfunctional compounds to revert to a "virtual hydrocarbon state" [25]. Indeed, Kováts indices of silylated compounds were predicted with a typical accuracy below 3% using in a first approximation linear regression functions which considered the atom number of the analytes [25].

Through our study we hope to contribute to the efficient sharing of RI reference libraries, such as GMD, between laboratories and present prerequisite criteria, such as empirical estimations of thresholds for retention index matching. Thus, we hope to support the ultimate goal initiated by the National Institute of Standards and Technology (NIST) to develop and utilise a general data base of chromatographic retention properties for the integrative RI and mass spectral matching of organic compounds [20]. Moreover, the chemometric efforts of predicting RI properties based on feature extractions from molecular structures currently appear to lack high accuracy. Respective predictions and training data sets may now be evaluated considering the information on the maximum possible precision of RI determination and transfer between chromatography variants which are provided through our study.

2. Experimental

2.1. Chemicals for retention time standardisation

Two compound classes are currently in use for the standardisation of retention times in routine GC–MS metabolite profiling experiments, namely *n*-alkanes (e.g. [18,26]) and *n*-alkyl fatty acid methyl esters (e.g. [27]). In this study the following reference substances were obtained from Sigma–Aldrich (Sigma–Aldrich Chemie GmbH, Munich, Germany) at highest available purity; *n*-alkanes: decane (CAS: 124-18-5), dodecane (CAS: 112-40-3), pentadecane (CAS: 629-62-9), octadecane (CAS: 593-45-3), nonadecane (CAS: 629-92-5), docosane (CAS: 629-97-0), octacosane (CAS: 630-02-4), dotriacontane (CAS: 544-85-4) and hexatriacontane (CAS: 630-06-8); *n*-alkyl fatty acid methyl esters: methyl octanoate (CAS: 111-11-5), methyl nonanoate (CAS: 1731-84-6), methyl decanoate (CAS: 110-42-9), methyl dodecanoate (CAS: 111-82-0), methyl myristate (CAS: 124-10-7), methyl palmitate (CAS: 112-39-0), methyl stearate (CAS: 112-61-8), methyl arachidate (CAS: 1120-28-1), methyl behenate (CAS: 929-77-1), methyl tetracosanoate (CAS: 2442-49-1), methyl hexacosanoate (CAS: 5802-82-4), methyl octacosanoate (CAS: 55682-92-3) and methyl melissate (CAS: 629-83-4). Alkanes were dissolved in pyridine at 0.22 mg/mL final concentration [26]. The fatty acid methyl ester (FAME) mixture was prepared separately using chloroform with final concentrations adjusted to 0.4 mg/mL or 0.8 mg/mL for liquid or solid C_8 ME–C_{30} ME standards [27]. Variations of these basic sets of retention marker mixtures are reported below (cf. Section 2.4.10).

Table 1
Defined mixture of 28 authenticated commercially available reference substances (DRM-mix)

Name	Reference number of the chemical abstracts service (CAS)	Final concentration (mg/L)
Benzoic acid, 4-hydroxy-	CAS: 99-96-7	9.9
Alanine, D-	CAS: 338-69-2	10.3
Caffeic acid	CAS: 331-39-5	10.3
Cholesterol	CAS: 57-88-5	10.0
Citramalic acid, D-	CAS: 626-10-8	5.0
Citric acid	CAS: 77-92-9	10.0
Fucose, L-	CAS: 2438-80-4	10.3
Glucose, alpha-D-	CAS: 492-62-6	10.0
Glutaric acid, 2-oxo-	CAS: 328-50-7	10.1
Glycine	CAS: 56-40-6	10.1
Isoleucine, DL-	CAS: 443-79-8	9.9
Lactitol	CAS: 81025-04-9	5.1
Lactose, beta-D-	CAS: 5965-66-2	5.2
Lanosterol	CAS: 79-63-0	10.5
Maltose, D-	CAS: 6363-53-7	20.8
Maltotriose	CAS: 1109-28-0	10.2
Palatinose	CAS: 13718-94-0	10.1
Panthothenic acid, D-	CAS: 137-08-6	9.9
Putrescine	CAS: 333-93-7	9.6
Pyridine, 3-hydroxy-	CAS: 109-00-2	9.9
Ribitol	CAS: 488-81-3	10.0
Ribose, D-	CAS: 50-69-1	10.3
Sorbitol, D-	CAS: 50-70-4	5.3
Sorbose, L-	CAS: 414-273-3850	9.9
Stigmasterol	CAS: 83-48-7	9.6
Threitol, DL-	CAS: 6968-16-7	10.6
Urea	CAS: 57-13-6	10.3
Valine, L-	CAS: 72-18-4	10.1

2.2. Preparations to assess the effect of matrix on retention time standardisation

2.2.1. Defined reference mixture of authenticated substances (DRM-mix)

A defined mixture of 28 authenticated reference substances referred to as the defined reference material (DRM-mix) was prepared as suggested by Fiehn and co-workers on the 2nd annual conference of the Metabolomics Society, 2006 in Boston, MA, USA (http://fiehnlab.ucdavis.edu/Boston%202006%20workshop.pdf). The substances were dissolved in chloroform–methanol–water (1:2.5:1, v/v/v) and diluted to 1 L final volume (Table 1). A defined volume of DRM, 267 μL, was dried using a VR Maxi vacuum concentrator with rotors R96-13 or R120-111 (Jouan Nordic, Allerod, Denmark) fitted to a hold-back vacuum pump (Ilmvac GmbH, Ilmenau, Germany) and subsequently subjected to routine chemical derivatisation and GC–MS profiling analysis (cf. Sections 2.3 and 2.4).

2.2.2. Defined reference material of a yeast intracellular extract (DRM-yeast)

A 1 L liquid batch of yeast, *Saccharomyes cerevisiae* strain S288C, was cultivated from a deep frozen stock using synthetically defined growth medium supplemented with yeast nitrogen base (Difco, Kansas City, MO, USA). Cells were harvested at optical density (OD_{595}) ~1.8. Subsequently intracellular metabolites were prepared as described earlier [26]. In short, 5 mL of yeast culture was

rapidly mixed with 20 mL methanol–water (6:4, v/v), which was pre-cooled to 60 °C. The metabolically inactivated cells were separated from residual growth medium and surplus methanol–water by centrifugation at <−20 °C. Cell pellets were re-suspended and extracted (1) 15 min at 70 °C using 374 μL extraction medium comprising 350 μL methanol, 12 μL internal standard 1 and 12 μL internal standard 2 (cf. below), (2) 10 min at 30 °C using 263 μL chloroform–water (188:75, v/v). The two respective supernatants after centrifugation were combined without liquid partitioning. Extracts of multiple preparations were pooled and equal 500 μL aliquots were dried by vacuum concentration for subsequent GC–MS profiling analysis (cf. Sections 2.3 and 2.4). This defined reference material is named DRM-yeast. The internal standards 1 and 2 are part of the routine metabolite preparation procedure [26], but were not required for this study. For the purpose of complete reporting, internal standard 1 contained 0.2 mg/mL ribitol (CAS: 488-81-3), 1 mg/mL 2,3,3,3-d_4-alanine (CAS: 53795-92-9), and 0.5 mg/mL D-isoascorbic acid (CAS: 89-65-6) dissolved in methanol and water, respectively. Internal standard 2 consisted of methyl nonadecanoate (CAS: 1731-94-8) dissolved at 2 mg/mL in chloroform.

2.2.3. Defined reference material of a rice leaf extract (DRM-rice)

Rice seeds, *Oryza sativa* ssp. indica, were submersed for 60 s in warm water (40 °C), transferred to Petri dishes containing wet cellulose tissue and germinated in the dark. After 2 days germinating seedlings were acclimated to the illuminated greenhouse and growth was allowed to continue to 3–5 cm seedling size. Subsequently, rice seedlings were transferred to hydroponic culture with a weekly exchange of liquid medium [28]. Four weeks after transfer complete shoot material was harvested and shock-frozen in liquid nitrogen. A pooled sample of shoot material from 25 plants was ground under liquid nitrogen to obtain a fine homogenous powder. An aliquot of 120 mg from this homogenate was extracted 15 min at 70 °C with 300 μL methanol, and 30 μL of internal standard 1 and internal standard 2, respectively. Finally 600 μL chloroform–water (1:2, v/v) was added, liquid phase partitioning performed by centrifugation. Multiple preparations of the upper polar phase were pooled and equal 40 μL aliquots dried by vacuum concentration. This defined reference material is in the following called DRM-rice.

2.3. Synthesis of analytes by chemical derivatisation

The dried materials were re-dissolved and chemically modified by 90 min agitation at 30 °C with 10 μL methoxyamine reagent, i.e. 40 mg/mL methoxyamine hydrochloride (CAS: 593-56-6) in pyridine. Then 90 μL reagent mixture, comprising *N*-methyl-*N*-trifluoroacetamide (MSTFA, CAS: 24589-78-4) trimethylsilylation reagent, *n*-alkane-mixture, and FAME-mixture (1000:16:4, v/v/v) were added and agitation continued 30 min at 37 °C. Injection volume was 1 μL of 100 μL final derivative volume [27]. In the following we define the term, analyte, to represent the products of chemical derivatisation, which are subjected to GC–MS analysis. A single compound may generate multiple analytes due to partial silylation and/or E,Z-isomers formed by methoxymation.

2.4. Chromatography variants

In this study we analysed a library collection [22,23] of analyte RI properties, which were recorded in various laboratories using essentially 8 variants of the original GC–MS based metabolite profiling method [13,14]. Besides the use of three detector technologies, namely quadrupole (Q), ion trap (TRAP) and time of flight (TOF) based mass spectral detection, which were deemed irrelevant for the present investigation, chromatography settings were modified. Specifically temperature programming, type of capillary column and choice of column manufacturer were varied. Most chromatography variants used 5%-phenyl-95%-dimethylpolysiloxane (5PDM) or equivalent capillary columns. For comparative purposes we included a variant using 35%-phenyl-65%-dimethylpolysiloxane (35PDM). In the following we describe the essential parameter selections of our own chromatography variants and present the relevant settings as extracted from publications of the other contributing laboratories.

2.4.1. Variant 1 (5PDM_VF5_9_TOF)

Variant 1 [26] uses helium carrier gas at 1 mL/min under constant flow control. Splitless injection at 230 °C was performed with flow transiently reduced to 0.6 mL/min into a conical, single taper liner with deactivated glass wool (Agilent Technologies, Böblingen, Germany). Purge time and flow reduction was 1 min. The 6890N gas chromatography system (Agilent Technologies) was mounted with a 5PDM VF-5 ms, 0.25 μm film thickness, 30 m × 0.25 mm fused silica capillary column (Varian, Darmstadt, Germany), which had an integrated 10 m guard column. The temperature programming comprised an initial 1 min isothermal period at 70 °C, a 9 °C/min ramp to 350 °C and a final 5 min constant heating at 350 °C. TOF-detection was performed using a Pegasus III mass spectrometry system (LECO). Mass spectral recording was set to 20 scans/s. Transfer line and ion source temperatures were set to 250 °C. The monitored mass range was *m*/*z* 70–600 amu. This range was extended to *m*/*z* 45–1000 amu for recording mass spectral tag (MST) information, namely RI and full mass spectrum, of reference compounds. Pipetting steps, automated chemical derivatisation and timed in-line injection into the GC–MS system were performed using a CTC Combi PAL autosampler and PAL cycle composer software version 1.5.0 (CTC Analytics AG, Zwingen, Switzerland).

2.4.2. Variant 2 (5PDM_RTX5_9_TOF)

Variant 2 differed only by choice of an alternative 5PDM capillary column type with equal dimensions, namely a 0.25 μm, 30 m × 0.25 mm RTX-5Sil MS with 10 m integrated guard column (Restek GmbH, Bad Homburg, Germany). The mass range was set to *m*/*z* 70–600 amu.

2.4.3. Variant 3 (5PDM_RTX5_15_TOF)

Like variant 2, variant 3 [18] had a 0.25 μm, 30 m × 0.25 mm RTX-5Sil MS capillary column with 10 m integrated guard column (Restek), but the temperature programming was altered to 2 min isothermal period at 80 °C, 15 °C/min ramp to 350 °C and 2 min at final temperature. Injection was splitless at 230 °C with a 2 min 110 psi pressure pulse at constant 1 mL/min flow rate. TOF-detection was performed using a Pegasus II mass spectrometry system (LECO). Mass spectral recording was adjusted to 6 scans/s and *m*/*z* 70–600 amu. The ion source temperature and transfer line were set to 200 °C and 250 °C.

2.4.4. Variant 4 (5PDM_DB5_40_TOF)

The method variant 4 [29] was a fast GC–TOF–MS application on a 6890N gas chromatograph (Agilent Technologies) hyphenated to a Pegasus mass spectrometry system (LECO). A 5PDM DB5-MS fused silica capillary column with 0.18 μm, 10 m × 0.18 mm dimensions (J&W Scientific, Folsom, CA, USA) was operated 2 min at 70 °C followed by a 40 °C/min ramp to 320 °C and a 1 min heating at final temperature. Injection was 1 μL at 270 °C with 1 min purge time at 20 mL/min purge flow. The transfer line and the ion source were set to 250 °C and 200 °C, respectively. The scan rate was 30 scans/s at *m*/*z* 50–800 amu.

2.4.5. Variant 5 (5PDM_VF5_6_Q)

A Trace GC ultra gas chromatograph with an AS 3000 auto sampler and a DSQ quadrupole-type mass spectrometer (ThermoFinnigan, San Jose, CA, USA) was used by variant 5 [30]. The sample was injected at 230 °C and separated on a 5PDM-type VF-5 ms 0.25 μm, 30 m × 0.25 mm fused silica capillary column (Varian), with helium at a flow rate of 1 mL/min. Temperature programming was 1 min isothermal at 70 °C, followed by 1 °C/min to 76 °C and 6 °C/min to 330 °C with 10 min final heating at 330 °C. Mass spectra were monitored with *m*/*z* 70–600 amu and 2 scans/s. The transfer line was set to 280 °C and the ion source to 250 °C.

2.4.6. Variant 6 (5PDM_DB5_6_Q)

Variant 6 [31] utilised a Trace gas chromatograph mounted with an AS 2000 auto sampler and a Trace mass spectrometer (ThermoFinnigan). Gas chromatography was performed with a 5PDM-type capillary column, namely a DB5-MS fused silica capillary column with 0.25 μm, 30 m × 0.25 mm dimensions (J&W Scientific) and helium carrier gas at 1 mL/min. Temperature programming was 1 min isothermal at 70 °C, followed by 1 °C/min to 76 °C and 6 °C/min to 325 °C with 10 min heating at 325 °C. The ion source temperature was adjusted to 220 °C. Mass spectra were recorded at 2 scans/s with *m*/*z* adjusted to 35–573 amu.

2.4.7. Variant 7 (5PDM_RTX5_5_Q)

Variant 7 [32] was performed using GC 8000 gas chromatograph coupled to a Voyager quadrupole-type mass spectrometer and an AS 2000 auto sampler (ThermoFinnigan). Gas chromatography was performed on a 0.25 μm, 30 m × 0.25 mm RTX-5Sil MS capillary column with 10 m integrated guard column (Restek) with 5 min isothermal period at 70 °C, a 5 °C/min temperature ramp to 320 °C and 1 min final heating. Sample injection was splitless at 230 °C and 1 mL/min helium carrier flow. The interface to the mass spectrometer was set to 250 °C and the ion source adjusted to 200 °C. The monitored mass range was set to *m*/*z* 40–600 amu. Mass spectra were recorded at 1.67 scans/s.

2.4.8. Variant 8 (5PDM_Eq5_3_TRAP)

The variant 8 [33] used an ion trap-type mass spectrometer, namely a PolarisQ ion trap mass spectrometer equipped with a Trace GC gas chromatograph and an AS2000 auto sampler (ThermoFinnigan). Splitless injection at 250 °C was performed with constant flow settings, 1 mL/min helium. A 5-PDM type capillary column was mounted, the Equity-5 column, 0.25 μm, 30 m × 0.25 mm (Supelco, Bellfonte, CA, USA). Chromatography settings were 3 min at 80 °C and 3 °C/min to 300 °C. The mass spectral acquisition rate was 2 scans/s with a range of *m*/*z* 50–550 amu. Transfer line and ion source temperatures were set to 250 °C and 200 °C, respectively.

2.4.9. Variant 9 (35PDM_MDN35_15_TOF)

Variant 9 represents the only chromatographic system of this study with capillary column polarity changed to a 35PDM-type [27]. The GC–TOF–MS system and basic settings were as described of variant 1. Except, the chosen capillary column, MDN-35, 0.25 μm, 30 m × 0.32 mm (Sigma–Aldrich), was operated at constant 2 mL/min helium flow starting with 2 min at 80 °C, heat ramping 15 °C/min to 330 °C and completing the cycle with 6 min at 330 °C.

2.4.10. Retention time standardisation

Method variants 1 and 9 used a combination of *n*-alkane mixture and FAME mixture for retention time standardisation and estimation of accuracy of prediction and precision of measurement and calculation (cf. Section 2.1). All other variants employed the above *n*-alkane mixture with the following variations. Variants 3–5 and 9 omitted *n*-decane because of chromatographic limitations. Variants 3, 5, and 6 lacked *n*-octadecane, whereas variant 6 had the complete set of *n*-alkanes ranging from C_{12} to C_{25}. The RIs of analytes, which were not bracketed by two retention markers, were extrapolated. Regression procedures based on all available marker compounds were applied without forcing an intercept. Alternatively calculations were performed based on the two nearest neighbours, for example the interpolation procedure according to algorithm proposed by van den Dool and Kratz [34]. Precision and accuracy were expressed in terms of standard deviation using either *n*-alkane based RI units or percent of the average and percent of expected, respectively.

2.5. Retention time retrieval, calculations and statistical procedures

The retention times from method variants 4–8 were retrieved at local chromatographic peak apices. Compound identity was manually confirmed by mass spectral match. Variants 1, 2, 3 and 9 were automatically deconvoluted [27] and mass spectra matched to a reference library through ChromaTof software (LECO). Peak lists of non-normalised mass spectra were exported and processed by TagFinder software [24]. Retention times were retrieved from these peak list files using the retention index calculation tool of the TagFinder software searching for retention times at local abundance maxima of compound characteristic mass fragments, such as *m*/*z* 71, 85, 99, 113 amu of *n*-alkanes and *m*/*z* 74, 87, 101, 143 amu of FAMEs and respective molecular masses. TagFinder has only van den Dool calculation of RIs implemented.

A Microsoft SQL Server 2005® was used as the relational database backend for storage and management of the mass spectral and chromatographic retention library information. Algorithms for RI processing were implemented using the Common Language Runtime (CLR.net), the C# programming language and Microsoft Visual Studio 2005®. Retention indices of analytes were computed using user-defined functions (UDF) of the database and T-SQL to access retention times of analytes and both the retention times and retention index definitions of the *n*-alkane or FAME marker compounds. Exploratory data visualisation was performed using Microsoft Excel software.

3. Results and discussion

3.1. Precision of empirical RI determination

Information on the empirical precision and accuracy of RI determination is prerequisite for the evaluation of RI projection methods, which aim to utilise existing RI libraries, such as provided by GMD, http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html, for the transfer of retention properties between chromatographic method variants. Ultimately, the threshold settings for compound matching will depend both on the achievable exactness of determination and projection.

3.1.1. Retention time drift

Retention time drift is one of the main obstacles to the utilisation of chromatographic compound properties for chemical identification purposes. Variations of capillary column length and artefacts of injection timing (Fig. 1) or slight changes in flow and pressure settings may strongly affect observable retention times. Moreover, capillary columns for gas chromatography have a limited life time which is limited by a slow continuous retention drift caused by gradual loss and modification of the stationary phase. These altered column properties may speed up chromatography significantly in

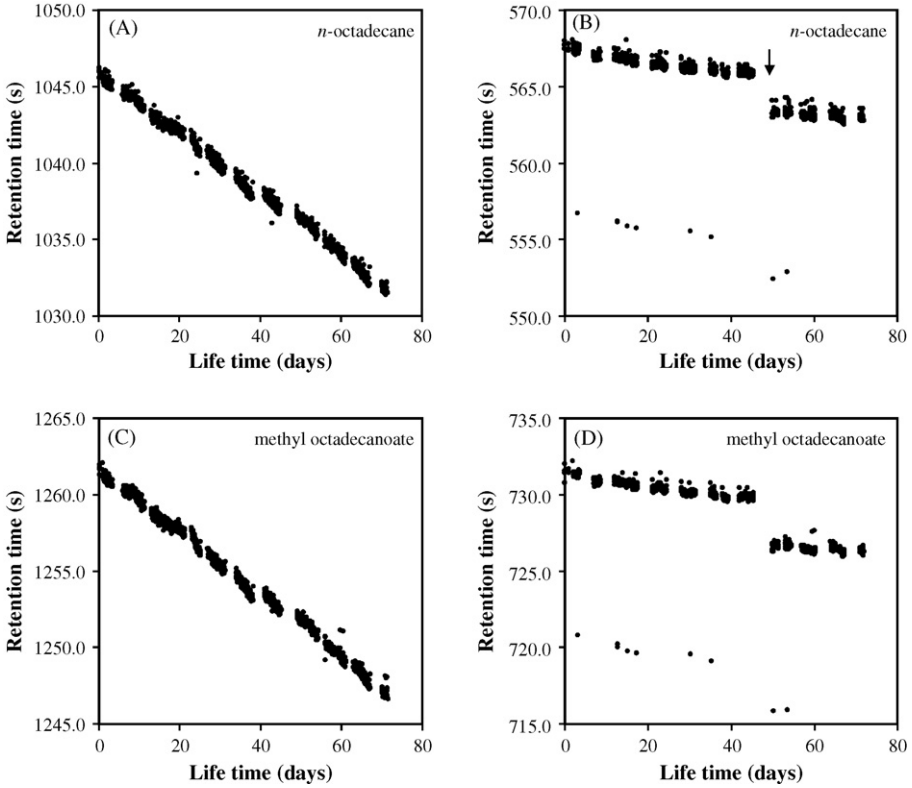

Fig. 1. Retention time drift of *n*-octadecane and methyl stearate under routine operating conditions of metabolite profiling analyses. Exemplary sequences of ~1340 chromatograms from variant 1 (A and C) and variant 9 (B and D) are displayed. Variants differ in column polarity, namely 5%-phenyl-95%-dimethylpolysiloxane versus 35%-phenyl-65%-dimethylpolysiloxane equivalent stationary phases, and temperature ramp, 9 °C/min versus 15 °C/min. Note that the shortening of capillary columns (arrow) may become necessary during routine maintenance of both method variants. The retention time outliers (cf. B and D) are artefacts caused by early injection. This artefact is not specific of variant 9. Both columns were conditioned by reagent injections prior to this selection of chromatographic runs. Total life time of both capillary columns may exceed the shown period.

the course of 1000 sample injections (Fig. 1). This drift depends on temperature ramping and column stability among other factors.

In view of these factors, we decided to apply the traditional concept of chemical retention time standardisation instead of relying on mathematical chromatography alignment procedures. In this study we used an *n*-alkane based RI system, where RI is defined as the number of carbon atoms multiplied by 100, and measurement performed by internal retention standardisation of each single chromatographic run. We selected variants 1 and 9 for investigations of the precision and accuracy of empirical RI determination, because the mass spectral scanning rate of these variants had sufficiently high resolution, namely an average of 0.086 RI units/scan and 0.136 RI units/scan, respectively, as determined through the distance by number of scans between dodecane and hexatriacontane peak apices. In comparison, other chromatographic variants constituting GMD, such as variant 7, have considerably lower chromatographic resolution, e.g. 0.679 RI units/scan.

3.1.2. Precision of retention index calculations from low complexity profiles

In the following we will demonstrate the influence of calculation methods and sample matrix on RI measurements. We partitioned the chromatographic runs of our study (Fig. 1) into sets comprising low complexity profiles, profiles containing a matrix of defined authenticated reference substances, DRM-mix, and profiles containing either intracellular metabolites of a microbial matrix, DRM-yeast, or a highly complex plant reference sample, DRM-rice. Low complexity profiles represented non-sample control runs or included a single reference substance, and thus typical mapping experiments to obtain reference RI properties. Retention times of spiked *n*-alkanes and FAMEs were retrieved from each chromatogram and RIs of FAMEs calculated using either interpolation methods, namely van den Dool [34], Kováts [35] and spline algorithms, or polynomial regression models using 1st–7th order and exponential fitting (Table 2). In the following, the precision of empirical RI determination is expressed in terms of RI standard deviation with independent replication >40 and accuracy estimated by difference of RI determined in the presence of a complex biological matrix compared to low complexity samples.

Precision of alkane RI determinations approximated chromatographic resolution in low complexity samples, when regression with increasing order was employed. However, average precision of FAME RIs remained limited to 0.22–0.33 RI units, using Van den Dool and spline interpolation or 3rd–5th order regression; exponential fitting was found to be non-optimal. These observations were made for both chromatography variants (Table 2).

3.1.3. Precision of retention index calculations from high complexity profiles

We selected sample types with increasing chemical complexity, namely DRM-mix < DRM-yeast < DRM-rice, to estimate the impact of matrix composition on RI determinations. The DRM-mix of 28 reference substances (Table 2) and varying mixtures of 20–25 substances (data not shown) did not affect RI determinations. However, both biological matrices had a negative effect. Average RI (S.D.) may increase to 0.75–1.11, depending on sample type and chromatography variant (Table 2).

Table 2
Influence of matrix composition on the precision of RI determination[a]

Method of calculation	Precision of RI determination (average standard deviation)							
	Average of C_{10}-C_{36} *n*-alkanes				Average of C_8-C_{30} Fatty Acid Methylesters			
	Low complexity	DRM-mix	DRM-yeast	DRM-rice	Low complexity	DRM-mix	DRM-yeast	DRM-rice
Chromatography variant 1								
1st Order polynomial regression	0.1788	0.1734	0.1839	2.6142	0.2274	0.2622	0.2633	3.5782
2nd Order polynomial regression	0.1525	0.1354	0.1546	0.8669	0.2342	0.2585	0.2633	1.5822
3rd Order polynomial regression	**0.1060**	**0.0957**	**0.1056**	**0.5721**	**0.2225**	**0.2494**	**0.2520**	**1.3407**
4th Order polynomial regression	0.0662	0.0469	0.0618	0.3339	0.2209	0.2487	0.2631	1.1161
5th Order polynomial regression	0.0407	0.0295	0.0410	0.1390	0.2234	0.2508	0.2640	1.1035
6th Order polynomial regression	0.0149	0.0101	0.0135	0.0449	0.2222	0.2475	0.2687	1.8134
Exponential regression	0.4098	0.4050	0.4066	5.3849	0.3582	0.4114	0.3576	5.1470
Spline interpolation					0.2326	0.2553	0.2689	1.1139
Kováts interpolation					0.2252	0.2402	0.2640	1.1555
Van den Dool interpolation					**0.2128**	**0.2326**	**0.2510**	**1.1070**
Chromatography variant 9								
1st Order polynomial regression	0.1997	0.1551	0.3390	0.3958	0.2425	0.2499	0.4866	0.7146
2nd Order polynomial regression	0.1408	0.1054	0.2744	0.3285	0.2425	0.2465	0.4901	0.7496
3rd Order polynomial regression	**0.0975**	**0.0782**	**0.2446**	**0.2957**	**0.2513**	**0.2416**	**0.4892**	**0.7296**
4th Order polynomial regression	0.0651	0.0587	0.2059	0.2341	0.2674	0.2493	0.5332	0.6965
5th Order polynomial regression	0.0305	0.0293	0.1776	0.2004	0.2921	0.2641	0.5750	0.7180
6th Order polynomial regression	< 0.0001	0.0001	0.0002	0.0002	0.5288	0.4584	2.2751	2.5882
Exponential regression	0.3328	0.3364	0.4787	0.5515	0.3143	0.3496	0.4936	0.7779
Spline interpolation					0.2764	0.2515	0.5618	0.7995
Kováts interpolation					0.5809	0.2769	0.5893	0.8377
Van den Dool interpolation					**0.3278**	**0.2527**	**0.5269**	**0.7490**

[a] Regression and interpolation methods were applied to calculate retention indices of *n*-alkanes and fatty acid methyl esters spiked into routine GC–TOF–MS metabolite profiles. Two chromatography variants, 1 or 9 (cf. Fig. 1), are compared. Samples had either low complexity, namely single reference substances and non-sample controls, or comprised complex defined reference material (DRM) of 28 reference substances (DRM-mix, $n = 45$), intracellular extracts of yeast (DRM-yeast, $n = 45$) and of rice (DRM-rice, $n = 41$). Van den Dool interpolation and 3rd order regression are highlighted by bold font.

All interpolation methods, namely spline, Kováts and van den Dool, and most regression algorithms were equally sensitive to these matrix effects. However, exponential fit, 1st, 2nd, 6th and higher order regression models did not properly reflect the impact of matrix on retention shifts.

In the following we selected van den Dool interpolation to investigate the source of reduced RI precision (Fig. 2). The strongest matrix effects were observed in early parts of the temperature programming of both investigated chromatography variants, namely the C_8 ME–C_9 ME region (Fig. 2A). The increased RI (S.D.) coincided with delayed retention of C_8 ME–C_9 ME. Thus, early eluting FAMEs had the strongest impact on overall RI accuracy in the presence of biological matrix with average $|RI_{(DRM\text{-}rice)} - RI_{(low\ complexity)}|$ equal to 1.0 (variant 1) and 0.5 (variant 9) RI units, respectively. The 3rd order regression algorithm was tested in parallel and exhibited highly similar results (data not shown). Therefore, we concluded that van den Dool interpolation and 3rd order regression are equivalent calculation approaches with respect to RI precision and accuracy.

3.1.4. Comparison of variant 1 and 9

The comparison of chromatography variants 1 and 9 demonstrated enhanced retention time stability of variant 9 (Fig. 1). The reduced retention drift appears to propagate into slightly enhanced RI precision (Fig. 2A) and accuracy (Fig. 2B). The cause of the improved retention behaviour of variant 9 was not further investigated and was deemed beyond the scope of this study. Both, the reduced duration of exposure to high temperatures per analysis cycle and possibly the altered stability of the capillary column may contribute. Moreover, the impact of matrix on RI performance may change with the biological object under investigation. For example, variant 9 performed better in the presence of DRM-rice, whereas variant 1 appeared to exhibit improved results with DRM-yeast.

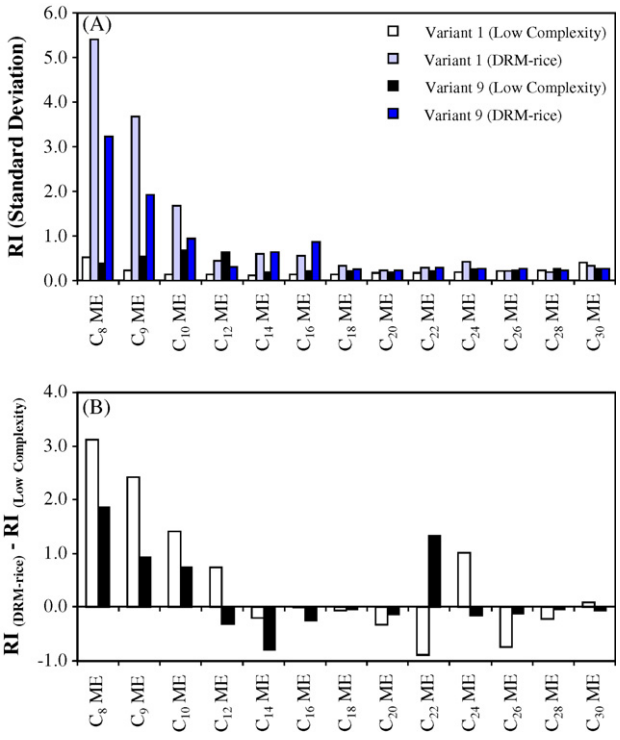

Fig. 2. The matrix effect negatively affecting RI precision and accuracy of methyl esters (ME) is dependent on chromatographic region. (A) Precision of RI determination was calculated as the standard deviation of RI ($n > 40$). (B) Accuracy of RI determination was estimated by comparison of RI measured in the presence of a complex biological matrix compared to a low complexity chemical background; the average $|RI_{(DRM\text{-}rice)} - RI_{(low\ complexity)}|$ was 1.0 (variant 1) and 0.5 (variant 9) RI units. Note that the strongest matrix effects occur at the start of the temperature program, e.g. C_8 ME and C_9 ME.

Table 3
Transfer of retention index (RI) properties between chromatography variants of the Golm Metabolome Database (GMD)[a]

	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5	Variant 6	Variant 7	Variant 8
Column brand	VF5	RTX5	RTX5	DB5	VF5	DB5	RTX5	Eq5
Temperature ramp (°C/min)	9	9	15	40	6	6	5	3
Gas chromatograph	6890N	6890N	6890N	6890N	Trace GC ultra	Trace GC	GC 8000	Trace GC
Aquisition rate (scans/s)	20	20	6	30	2	2	1.67	2
Mode of detection	TOF	TOF	TOF	TOF	Q	Q	Q	TRAP
A	Number of paired analytes							
Variant 1	488	348	274	179	157	175	318	65
Variant 2		931	623	209	226	244	518	96
Variant 3			964	184	206	192	437	93
Variant 4				299	151	127	197	71
Variant 5					264	154	224	77
Variant 6						324	190	82
Variant 7							961	96
Variant 8								103
B	Correlation coefficient (r^2)							
Variant 1	–	0.99984	0.99971	0.99945	0.99914	0.99951	0.99926	0.99956
Variant 2	0.99983	–	0.99979	0.99951	0.99910	0.99961	0.99932	0.99962
Variant 3	0.99971	0.99979	–	0.99950	0.99973	0.99977	0.99977	0.99945
Variant 4	0.99938	0.99945	0.99943	–	0.99880	0.99983	0.99886	0.99976
Variant 5	0.99916	0.99913	0.99974	0.99904	–	0.99987	0.99974	0.99962
Variant 6	0.99951	0.99961	0.99977	0.99983	0.99987	–	0.99981	0.99975
Variant 7	0.99925	0.99931	0.99977	0.99910	0.99974	0.99981	–	0.99978
Variant 8	0.99956	0.99962	0.99945	0.99976	0.99963	0.99975	0.99978	–
C	Standard deviation ($RI_{predicted} - RI_{determined}$)							
Variant 1	–	7.42	9.21	12.71	16.05	9.54	16.04	9.55
Variant 2	7.50	–	8.00	12.01	17.13	9.33	15.00	8.20
Variant 3	9.30	8.10	–	12.12	9.18	7.83	9.29	9.58
Variant 4	13.64	12.89	13.08	–	19.81	5.69	19.58	7.36
Variant 5	15.93	16.82	9.14	17.56	–	5.11	9.36	9.22
Variant 6	9.56	9.37	7.73	5.66	5.11	–	6.75	6.83
Variant 7	16.19	15.04	9.27	17.17	9.28	6.70	–	6.80
Variant 8	9.07	7.82	9.10	6.98	8.72	6.50	6.53	–
D	Standard deviation ($RI_{predicted} - RI_{determined}$ [% of $RI_{determined}$])							
Variant 1	–	0.38	0.42	0.52	0.59	0.55	0.64	0.52
Variant 2	0.39	–	0.42	0.47	0.61	0.64	0.60	0.44
Variant 3	0.43	0.44	–	0.54	0.46	0.51	0.43	0.57
Variant 4	0.53	0.48	0.54	–	0.66	0.35	0.69	0.40
Variant 5	0.59	0.62	0.45	0.62	–	0.29	0.40	0.46
Variant 6	0.56	0.67	0.50	0.35	0.29	–	0.37	0.38
Variant 7	0.64	0.60	0.42	0.64	0.40	0.36	–	0.36
Variant 8	0.50	0.43	0.55	0.38	0.44	0.37	0.35	–

[a] All included method variants were based on 5%-phenyl-95%-dimethylpolysiloxane or equivalent stationary phases and were operated at 1 mL/min constant helium flow. Column brand, temperature programming and mass spectral detection varied as indicated. (A) Number of paired analytes, which were used for 3rd order polynomial regression, (B) regression coefficients, r^2, (C) accuracy of prediction as characterised by standard deviation of residual errors, $RI_{predicted} - RI_{determined}$, and (D) accuracy of prediction as characterised by standard deviation of residual errors expressed as percent of $RI_{determined}$. Note that resulting matrices B and C are not exactly symmetrical; horizontal variants were used to predict RIs of the variants listed vertically.

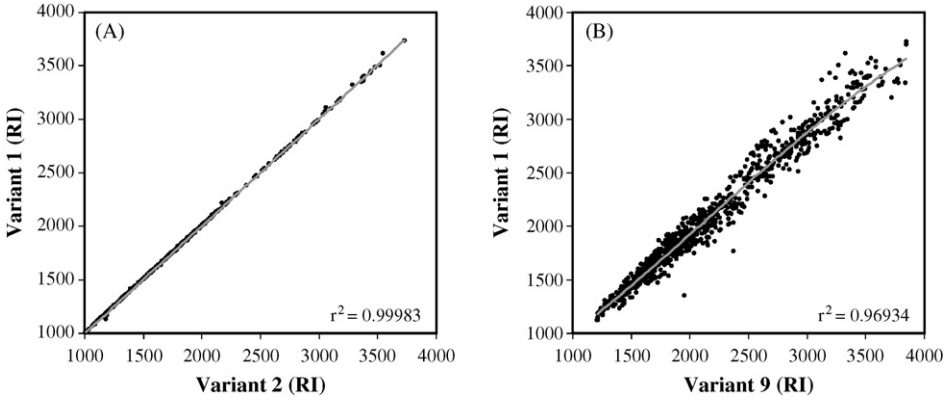

Fig. 3. Comparison of correlation of RI systems using 3rd order polynomial regression. (A) Variant 2 and variant 1 have equivalent 5%-phenyl-95%-dimethylpolysiloxane stationary phases. (B) In contrast, variant 9 utilises a 35%-phenyl-65%-dimethylpolysiloxane stationary phase. The fitted functions and correlation coefficients, r^2, are shown.

3.1.5. Check of biochemically relevant analytes

Precision of RI determination from our study was similar to the ~1 RI unit precision reported of a set of 250 volatile analytes [36]. We used the above DRM-mix (Table 1) to estimate, if results of RI precision, which were based on FAME, may be transferred to those compound classes which are relevant for routine metabolite profiling experiments. In the following, the chosen exemplary analytes are listed with RI (S.D.) of low complexity samples and – in square brackets – the respective precisions determined in the presence of DRM-yeast followed by DRM-rice. We analysed RI (S.D.) of citric acid (4TMS), 0.37 [0.43; 1.24], valine (2TMS), 0.30 [0.42; 2.51], glycine (3TMS), 0.20 [0.28; 1.94], ribitol (5TMS), 0.54 [0.53; 0.83], as well as glucose (1MEOX) (4TMS), 0.63 [4.00; 0.85]. In conclusion we found the results obtained with our FAME mixture to be representative. Precisions were influenced by matrix rather than by nature of chemical compound.

3.2. Transfer of RI properties between chromatography variants

The GMD, http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html, collection of mass spectral tags, a combination of mass spectral and RI library, was initially compiled with the intention of minimum redundancy and maximum complementation of diverse reference substances. For shared use of these data, the means of transferring RI information between the chromatography variants constituting GMD became essential. With increasing numbers of entries a substantial portion of redundancy became available (Table 3A). Sets of 65–623 paired analytes between chromatography variants allowed statistically sound investigation of regression models for RI prediction and estimation of residual error.

A correlation analysis of the retention index systems, which were measured using 5%-phenyl-95%-dimethylpolysiloxane or equivalent stationary phases, demonstrated high apparently linear correlation (Fig. 3A). This high linearity strongly contrasted with the expected low linear fit, when the RI system of variant 1 was fitted to variant 9, which utilises a 35%-phenyl-65%-dimethylpolysiloxane stationary phase (Fig. 3B). A detailed analysis of polynomial regression models applied among variants 1–8 revealed that in most cases 3rd order polynomial regression was sufficient to obtain a small increase of fit. Correlation coefficients, r^2, were improved at and beyond 3rd order regression, as was the standard deviation of the residual error, which was determined by $RI_{predicted} - RI_{determined}$ (Fig. 4). Two factors contributed to the magnitude of this residual error. (1) The residual error appeared to be a function of RI and thus may be proportional to the boiling point of analytes (Fig. 5B). (2) Single analytes may exhibit abnormally high deviations (Fig. 5). This abnormal behaviour could not be linked to a single type of analyte or the influence of specific chemical groups as determinant chemical features (data not shown). However, the amount of analyte was demonstrated earlier to have an impact on RI behaviour [19]. In this study we decided to keep the amount of substance constant for the purpose of RI mapping (cf. Sections 2.1 and 2.2) and attribute abnormally high deviations to non-documented quantitative experimental errors of previous method variants. Erroneous analyte assignments had been eliminated earlier.

Following the principle of making the minimal required number of assumptions we decided for a 3rd order regression model and investigated all possible pair wise predictions between chromatography variants (Table 3B–D). Regression coefficients ranged from 0.99880 to 0.99987. The standard deviation of residual errors varied from 5.1 to 19.8 RI units, equivalent to 0.29–0.69%. The margin of error was, thus, similar to the robustness reported of non-derivatised volatile analytes [36]. In comparison standard deviation of residual error was 108 (4.44%) RI units, when the 35-PDM variant 9 was used to predict variant 1 (Fig. 3B). The accuracy of RI transfer between chromatography variants appeared not to be subject to general systematic factors, except that variants with a shallow temperature ramp, namely variants 5–8, appeared to match better among each other. A similar observation was made with variants 1–3. In contrast to the apparent trend, the fast GC application of variant 4 had, however, best agreement with variants 6 and 8.

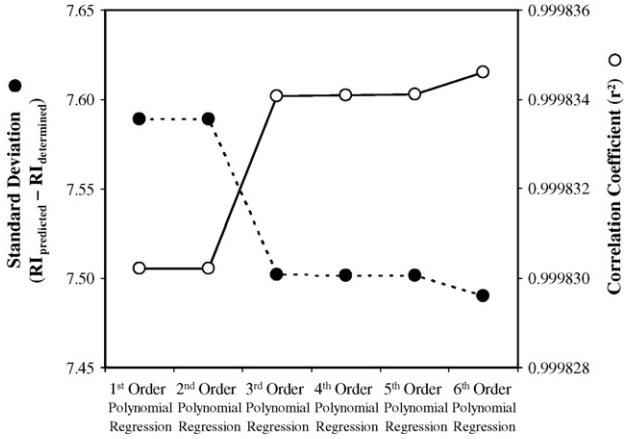

Fig. 4. Correlation coefficients, r^2 (open circle), and accuracy of prediction as determined by the standard deviation of the residual error (closed circle). Residual error of paired analytes was determined by $RI_{predicted} - RI_{determined}$. The projection of variant 2 onto variant 1 was subjected to permutation of 1st–6th order polynomial regression. 3rd order regression was found to represent the minimum required assumption for optimal prediction.

Finally we investigated, if grouping by chemical nature of analytes may improve accuracy of RI transfer between chromatography

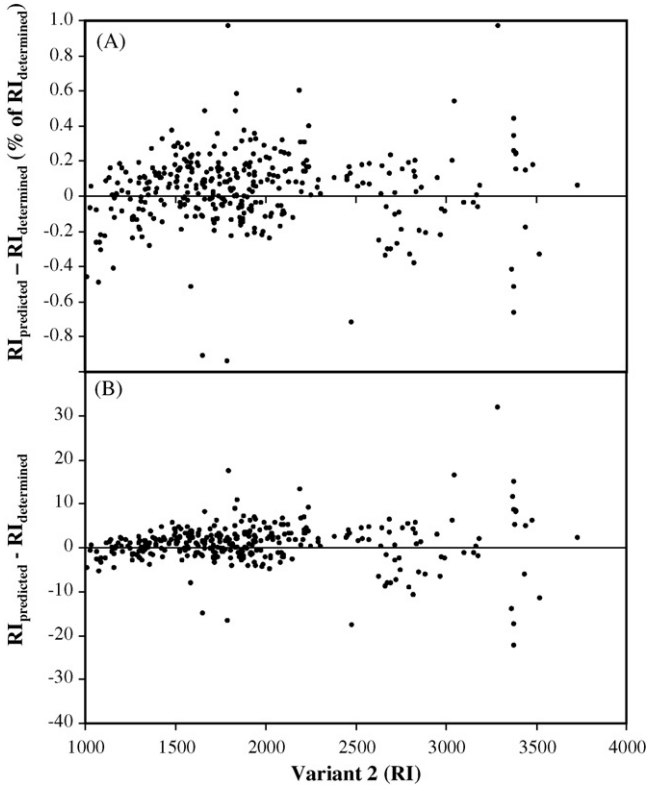

Fig. 5. Residual error of 3rd order polynomial regression using RI information of variant 2 to predict variant 1. (A) Percent of RI deviation $(RI_{predicted} - RI_{determined}) \times RI_{determined}^{-1} \times 100$, overall standard deviation, 0.39% ($n = 348$), (B) absolute RI deviation, $RI_{predicted} - RI_{determined}$, overall standard deviation 7.50 RI units (cf. Fig. 3A).

variants. This analysis was based on the report of Stein and co-authors on the use of the so-called group contributions to estimate Kováts RIs [21]. In the following we report examples taken from the projection of variant 2 onto variant 1 which had overall 7.42 standard deviation (S.D.) residual error, equivalent to 0.39% relative standard deviation and $r^2 = 0.99984$ with $n = 348$ paired analytes (Fig. 5). The subset of 29 hydroxy-, di- and tricarboxylic acids had 0.78 (S.D.) and 0.999989 (r^2). Moreover, a combination of all 37 paired sugars resulted in 3.74 (S.D.) with 0.999968 (r^2) and the set of 12 polyols and primary alcohols had 1.57 (S.D.) with $r^2 = 0.999985$. On the other hand a set of 12 compounds with purine, pyrimidine and indole *N*-heterocycles exhibited no improvement, e.g. 7.14 (S.D.) and $r^2 = 0.999285$.

4. Conclusions

We demonstrated that equal precision of RI determination could be obtained by previously reported interpolation methods [34,35] as well as spline and 3rd–5th order polynomial regression procedures. These findings held true in the presence of defined matrices and highly complex extracts from yeast cells and rice plants. For RI calculations within GMD [22,23] and the TagFinder software [24] we implemented the conventional van den Dool algorithm to best agree with earlier reports.

For transfer of RI information between chromatography variants using identical polarity of the stationary phase we selected a 3rd order regression model and implemented a respective projection procedure within GMD. This transfer procedure will provide mass spectra from GMD with RI predictions of those compounds which do not have experimentally verified variant RIs.

Moreover, we clearly demonstrated three possible levels of selecting RI thresholds. (1) In the presence of low complexity samples a threshold of 0.25 RI units may be applicable. This threshold, however, strictly applies only to controlled amounts of standards and analytes (cf. Section 2.1) (2) In the presence of highly complex samples the threshold must be set at least one order of magnitude higher. In the early chromatographic region analytes may exceed the respective threshold of approximately 2–3 RI units. (3) When projections from other chromatography variants are used the thresholds may be inferred from the standard deviations of residual errors (Table 3). As demonstrated by Fig. 5 thresholds may best be set as percent error of the expected absolute RI, for example to 0.5–1.0% (cf. Table 3D and Fig. 5).

In conclusion, accuracy of RI prediction may be much improved compared to earlier reports, when strictly equivalent stationary phases are exclusively considered. However, the estimated thresholds remain in part too broad for an unambiguous identification of isomers, especially in the presence of complex biological matrix. Therefore, we recommend for routine profiling analyses the co-analysis of defined mixtures of reference substances with each single GC–MS metabolite profiling experiment. These reference mixtures should be adjusted in quantitative and qualitative composition to the respective biological matrix under investigations. Mixtures may comprise (1) sets of authenticated reference substances which should cover the range of expected metabolite classes and (2) should contain selected isomers of those compound classes which cannot be distinguished by mass spectrometry alone.

Acknowledgements

This work was supported by the Max Planck Society, the QuantPro program of the Bundesministerium für Bildung und Forschung (BMBF), sub-project "InnOx–Innovative diagnostic tools to optimise potato breeding: Systemic analysis of cellular processes and their relation to plant internal oxygen concentrations", FKZ 0313813A, and the European META-PHOR project, FOOD-CT-2006-036220. The authors acknowledge the long standing support and encouragement by Prof. L. Willmitzer, Max Planck Institute of Molecular Plant Physiology (MPI-MP), Am Muehlenberg 1, D-14476 Potsdam-Golm, Germany.

References

[1] J.K. Nicholson, J.C. Lindon, E. Holmes, Xenobiotica 29 (1999) 1181.
[2] R.N. Trethewey, A.J. Krotzky, L. Willmitzer, Curr. Opin. Plant Biol. 2 (1999) 83.
[3] L.W. Sumner, P. Mendes, R.A. Dixon, Phytochemistry 62 (2003) 817.
[4] R.J. Bino, R.D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B.J. Nikolau, P. Mendes, U. Roessner-Tunali, M.H. Beale, R.N. Trethewey, B.M. Lange, E.S. Wurtele, L.W. Sumner, Plant Sci. 9 (2004) 418.
[5] G. Stephanopoulos, H. Alper, J. Moxley, Nat. Biotechnol. 22 (2004) 1261.
[6] J. Nielsen, S.G. Oliver, Trends Biotechnol. 23 (2005) 544.
[7] H. Jenkins, N. Hardy, M. Beckmann, J. Draper, A.R. Smith, J. Taylor, O. Fiehn, R. Goodacre, R.J. Bino, R.D. Hall, J. Kopka, G.A. Lane, B.M. Lange, J.R. Liu, P. Mendes, B.J. Nikolau, S.G. Oliver, N.W. Paton, S. Rhee, U. Roessner-Tunali, K. Saito, J. Smedsgaard, L.W. Sumner, T. Wang, S. Walsh, E.S. Wurtele, D.B. Kell, Nat. Biotechnol. 22 (2004) 1601.
[8] J.C. Lindon, H.C. Keun, T.M.D. Ebbels, J.M.T. Pearce, E. Holmes, J.K. Nicholson, Pharmacogenomics 6 (2005) (2005) 691.
[9] A.L. Castle, O. Fiehn, R. Kaddurah-Daouk, J.C. Lindon, Brief. Bioinform. 7 (2006) 159.
[10] O. Fiehn, B. Kristal, B. Van Ommen, L.W. Sumner, S.A. Sansone, C. Taylor, N. Hardy, R. Kaddurah-Daouk, Omics J. Intergrat. Biol. 10 (2006) 158.
[11] O. Fiehn, D. Robertson, J. Griffin, M. van der Werf, B.J. Nikolau, N. Morrison, L.W. Sumner, R. Goodacre, N.W. Hardy, C. Taylor, J. Fostel, B. Kristal, R. Kaddurah-Daouk, P. Mendes, B. van Ommen, J.C. Lindon, S.A. Sansone, Metabolomics 3 (2007) 175.
[12] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Metabolomics 3 (2007) 211.
[13] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. Trethewey, L. Willmitzer, Nat. Biotechnol. 18 (2000) 1157.
[14] U. Roessner, C. Wagner, J. Kopka, R.N. Trethewey, L. Willmitzer, Plant J. 23 (2000) 131.
[15] P. Ausloos, C.L. Clifton, S.G. Lias, A.I. Mikaya, S.E. Stein, D.V. Tchekhovskoi, O.D. Sparkman, V. Zaikin, D. Zhu, J. Am. Soc. Mass Spectrom. 10 (1999) 287.
[16] J.M. Halket, D. Waterman, A.M. Przyborowska, R.K.P. Patel, P.D. Fraser, P.M. Bramley, J. Exp. Bot. 56 (2005) 219.
[17] S.E. Stein, J. Am. Soc. Mass Spectrom. 10 (1999) 770.
[18] C. Wagner, M. Sefkow, J. Kopka, Phytochemistry 62 (2003) 887.
[19] J. Kopka, J. Biotechnol. 124 (2006) 312.
[20] V.I. Babushok, P.J. Linstrom, J.J. Reed, I.G. Zenkevich, R.L. Brown, W.G. Mallard, S.E. Stein, J. Chromatogr. A 1157 (2007) 414.
[21] S.E. Stein, V.I. Babushok, R.L. Brown, P.J. Linstrom, J. Chem. Inf. Model. 47 (2007) 975.
[22] N. Schauer, D. Steinhauser, S. Strelkov, D. Schomburg, G. Allison, T. Moritz, K. Lundgren, U. Roessner-Tunali, M.G. Forbes, L. Willmitzer, A.R. Fernie, J. Kopka, FEBS Lett. 579 (2005) 1332.
[23] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmueller, P. Doermann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A.R. Fernie, D. Steinhauser, Bioinformatics 21 (2005) 1635.
[24] A. Luedemann, K. Strassburg, A. Erban, J. Kopka, Bioinformatics 24 (2008) 732.
[25] C.T. Peng, Z.C. Yang, D. Maltby, J. Chromatogr. 586 (1991) 113.
[26] A. Erban, N. Schauer, A.R. Fernie, J. Kopka, in: W. Weckwerth (Ed.), Metabolomics: Methods and Protocols, Humana Press, Totowa, 2007, pp. 19–38.
[27] J. Lisec, N. Schauer, J. Kopka, L. Wilmitzer, A. Fernie, Nat. Protoc. 1 (2006) 387.
[28] X. Yang, V. Romheld, H. Marschner, Plant Soil 164 (1994) 1.
[29] J. Gullberg, P. Jonsson, A. Nordström, M. Sjöström, T. Moritz, Anal. Biochem. 331 (2004) 283.
[30] U. Roessner, J. Patterson, M.G. Forbes, G. Fincher, P. Langridge, A. Bacic, Plant Physiol. 142 (2006) 1087.
[31] S. Strelkov, M. von Elstermann, D. Schomburg, Biol. Chem. 385 (2004) 853.
[32] G. Colebatch, G.G. Desbrosses, T. Ott, L. Krusell, O. Montanari, S. Kloska, J. Kopka, M.K. Udvardi, Plant J. 39 (2004) 487.
[33] A. Barsch, T. Patschkowski, K. Niehaus, Funct. Integr. Genomics 4 (2004) 219.
[34] H. van den Dool, P.D. Kratz, J. Chromatogr. 11 (1963) 463.
[35] E. Kováts, Helv. Chim. Acta 41 (1958) 1915.
[36] F. Bianchi, M. Careri, A. Mangia, M. Musci, J. Sep. Sci. 30 (2007) 563.