

# **GC-TOF-MS basierte Analyse von niedermolekularen Primär- und Sekundärmetaboliten agrарwirtschaftlich bedeutsamer Nutzpflanzen**

Dissertation  
zur Erlangung des akademischen Grades  
„doctor rerum naturalium“  
(Dr.rer.nat.)  
im Wissenschaftsbereich  
„Analytische Biochemie“

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam

von  
Dipl. Chem. Nadine Strehmel  
geboren am 21.3.1983 in Brandenburg/Havel

Golm, den 12.10.2010

Online veröffentlicht auf dem  
Publikationsserver der Universität Potsdam:  
URL <http://opus.kobv.de/ubp/volltexte/2011/5123/>  
URN <urn:nbn:de:kobv:517-opus-51238>  
<http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-51238>

## Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und unter Verwendung keiner anderen als den von mir angegeben Quellen und Hilfsmitteln verfasst habe. Ferner erkläre ich, dass ich bisher weder an der Universität Potsdam noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Nadine Strehmel

Golm, den 12.10.2010

**A Einleitung**

A.1 Verschiedene Omik-Technologien	1
A.2 Die Wissenschaftsdisziplin der Metabolomik	2
A.3 Probleme und Lösungsansätze metabolischer Untersuchungen	6
A.4 Motivation und Zielsetzung	7

**B Materialien und Methoden**

<b>B.1 Materialien</b>	9
B.1.1 Chemikalien und Standards	9
B.1.2 Pflanzenmaterial	11
B.1.2.1 Versuch „Sortenscreening“	11
B.1.2.2 Versuch „Zuchtklone“	11
B.1.2.3 Versuch „Keimhemmung“	12
B.1.2.4 Versuch „Physiologisches Alter“	13
B.1.2.5 Versuch „Mechanische Belastung von Kartoffelknollen“	13
B.1.2.6 Versuch „Gepaarte NMR und GC-TOF-MS Analyse“	13
B.1.2.7 Etablierung der Analytik von Phytohormonen	14
B.1.3 Geräte und Hilfsmittel	14
<b>B.2 Methoden</b>	16
B.2.1 Phänotypische Untersuchungen	16
B.2.1.1 Lagerungsverlust	16
B.2.1.2 Chipseignung	17
B.2.1.3 Schwarzfleckigkeitsneigung	17
B.2.2 Sauerstoffmessung	18
B.2.3 GC-TOF-MS basierte Metabolitenanalyse	18
B.2.3.1 Probenentnahme	18
B.2.3.2 Probenvorbereitung	19
B.2.3.2.1 Extraktion polarer Substanzen	19
B.2.3.2.2 Derivatisierung polarer Substanzen	19
B.2.3.3 GC-TOF-MS Messung	19
B.2.3.4 Datenprozessierung	20
B.2.3.5 Datenauswertung	20
B.2.3.6 Datenarchivierung	20
B.2.4 Normierung der Daten	21
B.2.4.1 Statistische Auswertung	21

---

B.2.5 Bestimmung saurer Phytohormone	21
B.2.6 Klassifizierung unbekannter GC-MS Signale	22
B.2.6.1 Gepaarte Korrelationsspektroskopie	22
B.2.6.1.1 Extraktion polarer Substanzen aus Nutzpflanzen	23
B.2.6.1.2 Präparative Flüssigkeitschromatographie	23
B.2.6.1.3 NMR Messung	23
B.2.6.1.4 GC-TOF-MS Messung	23
B.2.6.2 Selektive An- und Abreicherung polarer Substanzen	24
B.2.6.2.1 Hydroxyliertes Polymer „Plexa“	24
B.2.6.2.2 Anionenaustauscher „SAX“	24
B.2.6.2.3 Kationenaustauscher „MP1“	24
B.2.6.3 Entscheidungsbäume	24
<b>C Ergebnisse und Diskussion</b>	<b>26</b>
C.1 Analytische Lösungsansätze	26
C.1.1 Retentionszeit und Retentionsindex	26
C.1.2 Der massenspektrale Vergleich	34
C.1.3 Referenzspektrenbibliothek	36
C.1.4 Standardisierung des Extraktions- und des Derivatisierungsprozesses	39
C.1.5 Langfristige Datenarchivierung	40
C.1.6 Analytik von Spurensubstanzen in biologischem Gewebe	42
C.1.6.1 Fest-Flüssig-Extraktion	43
C.1.6.2 Festphasenextraktion	43
C.1.6.3 Weitere Anreicherung	46
C.1.6.4 Detektion	46
C.1.6.4.1 Optimierung der GC-TOF-MS Analyse	47
C.1.6.4.2 Vergleich zwischen GC-TOF-MS und UPLC-FTICR-MS	49
C.1.7 Zusammenfassung und Bewertung der etablierten Analytischen Lösungsansätze	49
C.2 Physiologisch relevante Stoffwechselprodukte in der Kartoffelknolle	51
C.2.1 Indikative Biomarker	51
C.2.1.1 Mechanische Belastung von Kartoffelknollen	51
C.2.1.2 Lagerung von Kartoffelknollen	56
C.2.1.2.1 Der Einfluss der Temperatur auf das Lagerungsverhalten	61
C.2.1.2.2 Der Einfluss von Keimhemmern auf das Lagerungsverhalten	62

C.2.2 Prädiktive Biomarker	67
C.2.2.1 Beurteilung der Qualität von Kartoffeln nach der Ernte hinsichtlich Schwarzfleckigkeitsneigung und Chipseignung	68
C.2.3 Zusammenfassende Beurteilung von Biomarkern in der Matrix Kartoffelknolle	71
C.3 Klassifizierung von unbekanntem GC-MS Signalen	73
C.3.1 Entscheidungsbaumverfahren	73
C.3.1.1 Trainingsphase	74
C.3.1.2 Klassifizierungsphase	78
C.3.2 Gezielte Anreicherung mittels Festphasenextraktion	81
C.3.3 Geparate GC-MS und NMR Analyse	83
C.3.3.1 Optimierung der Fraktionierung	83
C.3.3.2 Geparate NMR und GC-TOF-MS Analyse	87
C.3.4 Bewertung der Ergebnisse zur Klassifizierung von GC-MS Signalen	91
 <b>D Zusammenfassung</b>	 93
 <b>E Literaturverzeichnis</b>	 97
 <b>Anhang I</b>	 102
 <b>Anhang II</b>	 144

---

$\alpha$	Signifikanzniveau
A	primäres <b>Amin</b>
A(1TMS)	einfach silyliertes <b>Amin</b>
A(2TMS)	zweifach silyliertes <b>Amin</b>
ABA	<b>Abscisic Acid</b> ; Abscisinsäure
Abb	<b>Abbildung</b>
Ala	<b>Alanin</b>
A-Nummer	<b>Analyt-Nummer</b>
amu	<b>Atomic Mass Unit</b> ; atomare Masseneinheit
AnEx	<b>Anion Exchanger</b> ; Anionenaustauscher
AS	alpha <b>Aminosäure</b>
AS(1TMS)	einfach silylierte alpha <b>Aminosäure</b>
AS(2TMS)	zweifach silylierte alpha <b>Aminosäure</b>
AS(3TMS)	dreifach silylierte alpha <b>Aminosäure</b>
AS	<b>Analytstruktur</b>
B	<b>Brokkoliröschen</b>
BNA	<b>Böhm-Nordkartoffel-Agrarproduktion</b>
BSA	<b>Bundessortenamt</b>
BSTFA	<i>N,O</i> - <b>Bis</b> (trimethylsilyl)trifluoroacetamid
bzw	<b>beziehungsweise</b>
CAS	<b>Chemical Abstracts Service</b>
$^{13}\text{C}$	$^{13}\text{C}$ Isotop
°C	Grad <b>Celsius</b>
CatEx	<b>Cation Exchanger</b> ; Kationenaustauscher
C18	Reversed-Phase; Umkehrphase
Andi-ms	<b>Analytical Data Interchange format for Mass Spectrometry (netCDF)</b>
CE	<b>Capillary Electrophoresis</b> ; Kapillarelektrophorese
cm	<b>Centimeter</b> ; Zentimeter
Col	<b>Columbia</b>
d	<b>day</b> ; Tag
Da	<b>Dalton</b>
E	<b>East</b> ; Ost
EI	<b>Electron Impact</b> , Elektronenstoßionisation
EIGTMS	Labor <b>EIGTMS</b>
EIMOR	Labor <b>EIMOR</b>
EIQUAD	Labor <b>EIQUAD</b>
EIROE	Labor <b>EIROE</b>
EISTR	Labor <b>EISTR</b>
EITRAP	Labor <b>EITRAP</b>
EITTMS	Labor <b>EITTMS</b>
engl	<b>englisch</b>

---

ESI	<b>E</b> lectrospray Ionisation; Elektrosprayionisierung
ETA	<b>E</b> thanolamin
eV	<b>E</b> lektronenvolt
FP	<b>F</b> alse <b>P</b> ositive; falsch positive Menge
F <sub>pr</sub>	Fisher-Snedecor Test
FN	<b>F</b> alse <b>N</b> egative; falsch negative Menge
g	<b>G</b> ramm
GA3	<b>G</b> ibberellin <b>A3</b> , Gibberellinsäure
GA4	<b>G</b> ibberellin <b>A4</b>
GA7	<b>G</b> ibberellin <b>A7</b>
GABA	<b>G</b> amma- <b>A</b> minobutyric <b>A</b> cid, 4-Aminobuttersäure
GC	<b>G</b> as <b>C</b> hromatography, Gaschromatographie
GCxGC	<b>C</b> omprehensive Two-Dimensional <b>G</b> as <b>C</b> hromatography
GC-TOF-MS	<b>G</b> as <b>C</b> hromatography - <b>T</b> ime of <b>F</b> light - <b>M</b> ass <b>S</b> pectrometry
GC-MS	<b>G</b> as <b>C</b> hromatography - <b>M</b> ass <b>S</b> pectrometry
GI	<b>G</b> leichung
Glycerol-3-P	<b>G</b> lycerol- <b>3</b> -phosphat
GMD	<b>G</b> olm <b>M</b> etabolome <b>D</b> atabase
h	<b>H</b> our; Stunde
<sup>2</sup> H	<sup>2</sup> <b>H</b> Isotop
HLB	<b>H</b> ydrophilic <b>L</b> ipophilic <b>B</b> alance
HP	<b>H</b> auptprodukt
HPLC	<b>H</b> igh <b>P</b> erformance <b>L</b> iquid <b>C</b> hromatography
Hz	<b>H</b> ertz
IAA	<b>I</b> ndole-3- <b>a</b> cetic <b>A</b> cid, Indolessigsäure
IC	<b>I</b> ndependent <b>C</b> omponent; Unabhängige Komponente
ICA	<b>I</b> ndependent <b>C</b> omponent <b>A</b> nalysis; Unabhängige Komponentenanalyse
K	<b>K</b> artoffelknolle
KEGG	<b>K</b> yoto <b>E</b> ncyclopedia of <b>G</b> enes and <b>G</b> enomes
kg	<b>K</b> ilogramm
L	<b>L</b> iter
LC	<b>L</b> iquid <b>C</b> hromatography
LC-MS	<b>L</b> iquid <b>C</b> hromatography - <b>M</b> ass <b>S</b> pectrometry
Leu	<b>L</b> euicin
Lys	<b>L</b> ysin
M	<b>M</b> olar
m	<b>M</b> eter
m	Steigung
Massbank	Metabolische Datenbank
MCC	<b>M</b> atthews <b>C</b> orrelation <b>C</b> oefficient
MDN35	Labor <b>M</b> DN35



---

MEOX	<b>methoxyaminiert</b>
mol	<b>MicroStation Development Language</b>
MeltDB	Programm für metabolische Untersuchungen
Metalign	Programm für metabolische Untersuchungen
Metmax	Programm für metabolische Untersuchungen
MHz	<b>Megahertz</b>
M-Nummer	<b>Metabolit-Nummer</b>
[M] <sup>+</sup>	<b>Molekülion</b>
mg	<b>Milligramm</b>
min	<b>Minute</b>
mL	<b>Milliliter</b>
mm	<b>Millimeter</b>
MeOX	<b>Methoxyaminhydrochlorid</b>
MS	<b>Mass Spectrometer, Massenspektrometer</b>
MS	<b>Metabolitstruktur</b>
MS-MS	Tandem-MS
<sup>15</sup> N	<sup>15</sup> N Isotop
µL	<b>Mikroliter</b>
µm	<b>Mikrometer</b>
n	Achsenabschnitt
N	<b>North; Nord</b>
NIST	<b>National Institute of Standards and Technology</b>
MSTFA	<i>N</i> - <b>methyl-N</b> -(trimethylsilyl) <b>trifluoroacetamid</b>
ng	<b>Nanogramm</b>
n <sub>L</sub>	Anzahl stark verfärbter Knollen
n <sub>M</sub>	Anzahl mäßig verfärbter Knollen
n <sub>S</sub>	Anzahl kaum verfärbter Knollen
NMR	<b>Nuclear Magnetic Resonance, Kernspinresonanz</b>
NP	<b>Nebenprodukt</b>
PCA	<b>Principal Component Analysis; Hauptkomponentenanalyse</b>
pK <sub>s</sub>	Säurestärke
PLS	<b>Partial Least Squares</b>
PLS-VIP	<b>Partial Least Squares - Variables Importance in the Projection</b>
Plexa	Festphasenextraktionskartusche
p	<b>Precision; Genauigkeit</b>
p	Signifikanzwert
PPL	Festphasenextraktionskartusche
Put	<b>Putrescin</b>
QUAD	<b>Quadrupol</b>
R	Korrelationskoeffizient
r	<b>Recall; Trefferquote</b>

---

rel	<b>relativ</b>
RI	<b>Retentionsindex</b>
rpm	<b>Rounds per Minute</b>
s	<b>Sekunde</b>
SA	<b>Salicylic Acid; Salicylsäure</b>
Ser	<b>Serin</b>
SI	<b>Schwarzfleckigkeitsindex</b>
SIM	<b>Selected Ion Monitoring</b>
S/N	<b>Signal-to-Noise Ratio; Signal zu Rausch Verhältnis</b>
SOM	<b>Self-organizing Map</b>
SPE	<b>Solid Phase Extraction; Festphasenextraktion</b>
split 50	Gaschromatographische Injektionsmethode
SQL	<b>Structured Query Language</b>
STABW	<b>Standardabweichung</b>
Suc	<b>Succinat</b>
t	<b>Tonne</b>
t	<b>Retention Time; Retentionszeit</b>
t <sub>x</sub>	Retentionszeit des früh eluierenden Alkans
t <sub>y</sub>	Retentionszeit des spät eluierenden Alkans
t <sub>Analyt</sub>	Retentionszeit Analyt
Tab	<b>Tabelle</b>
tag	<b>target; Fragmentmasse zu einer bestimmten Elutionszeit</b>
TagFinder	Programm für metabolische Untersuchungen
TargetSearcher	Programm für metabolische Untersuchungen
TCA	<b>Tricarboxylic Acid Cycle; Tricarbonsäurezyklus</b>
TMS	<b>Trimethylsilyl</b>
TN	<b>True Negative; richtig negative Menge</b>
TOF	<b>Time of Flight; Flugzeitmassenspektrometer</b>
TP	<b>True Positive; richtig positive Menge</b>
TRAP	ion <b>trap</b> ; Ionenfalle
Trp	<b>Tryptophan</b>
t-Test	Hypothesentest mit <b>t</b> -verteilter <b>Test</b> prüfgröße
txt	<b>Textdatei</b>
UPLC-FTICR-MS	<b>Ultra Performance Liquid Chromatography-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry</b>
v	<b>Volumenanteil</b>
Val	<b>Valin</b>
VANTED	<b>Visualization and Analysis of Networks containing Experimental Data</b>
Var	<b>Variety; Kultivar</b>
VAR5	Labor <b>VAR</b> ian DB-5
XCMS	Programm für metabolische Untersuchungen

xml	Extensible <b>M</b> arkup <b>L</b> anguage; erweiterbare Auszeichnungssprache
Z	<b>Z</b> eatin
zB	zum <b>B</b> eispiel

- Seite 27    Abb. 1 Verlauf der Retentionszeit innerhalb einer Messserie für die Verbindung Octadecansäuremethylester A) auf dem System VAR5 und B) dem System MDN35.
- Seite 29    Abb. 2 Standardabweichung des Retentionsindex in Bezug auf die zugegebenen Fettsäuremethylester im Hinblick auf A) das Modell von van den Dool, B) die Spline Funktion, C) eine Polynomiale Regression erster Ordnung und D) eine Polynomiale Regression fünfter Ordnung in Mix 1 (schwarz), Mix 2 (rot), „Reisblatt“ (grün) und „Bäckerhefe“ (gelb) für das System VAR5.
- Seite 31    Abb. 3 Der Verlauf des Retentionsindex im Laufe einer Messserie A) auf dem System VAR5 und B) dem System MDN35 für die übliche Betriebsdauer einer Kapillarsäule.
- Seite 32    Abb. 4 Transfer von Retentionsindices verschiedener Laboratorien im Hinblick auf A) Säulen gleicher Polarität und B) Säulen unterschiedlicher Polarität.
- Seite 33    Abb. 5 Ungenauigkeit der Retentionsindexbestimmung nach Transfer des Retentionsindex von dem VAR5 System auf das EIGTMS System über den relevanten Messbereich.
- Seite 36    Abb. 6 Vergleich der massenspektralen Vergleichsparameter A) NIST\_R; B) NIST\_Prob; C) Euklidische Distanz und D) 1-Skalarprodukt in Bezug auf alle Analyten in Bezug auf „Mix 2“.
- Seite 37    Abb. 7 Vergleich der Qualität von Durchschnittsspektren auf der Basis der mittleren Fragmentanzahl aller Analyten.
- Seite 38    Abb. 8 Einordnung GC-MS zugänglicher Metaboliten von Arabidopsis thaliana (thale cress) in den allgemeinen KEGG Referenzstoffwechselweg.
- Seite 42    Abb. 9 Datenbankschema zur Archivierung der erhobenen metabolischen und phänotypischen Daten.
- Seite 44    Abb. 10 Optimierung des A) Elutionsvolumens und des B) Anreicherungsolumens für das Sorbens Plexa zur Anreicherung und Abreicherung der Zielmoleküle ABA (○), GA3 (□) und IAA (Δ).

- Seite 46    Abb. 11 Vergleich verschiedener Sorbentien hinsichtlich des Abreicherungsgrades.
- Seite 48    Abb. 12 Optimierung des Derivatisierungsvolumens für die Matrix „Arabidopsisblatt“ nach Anwendung einer Hauptkomponentenanalyse.
- Seite 48    Abb. 13 Vergleich zweier Temperaturprogramme für die gaschromatographische Analyse anhand des detektierten Signal-zu-Rausch-Verhältnisses (S/N).
- Seite 53    Abb. 14 Unabhängige Komponentenanalyse (ICA) angewendet auf alle gemessenen Massenfragmente.
- Seite 54    Abb. 15 Relative Änderung der mit Zweifaktorieller Varianzanalyse identifizierten endogenen Metaboliten hinsichtlich der Faktoren „Zeit“ und „Mechanische Belastung“.
- Seite 56    Abb. 16 Relative Änderungen A) in der Respirationsrate und B) im Proteingehalt nach mechanischer Belastung für das Kultivar Afra und das Kultivar Milva.
- Seite 57    Abb. 17 Unabhängige Analyse der Hauptkomponenten in Hinblick auf das Physiologische Alter für die Lagersaison 2006/2007 angewendet auf alle detektierten gaschromatographischen Signale.
- Seite 59    Abb. 18 Relative Zeitliche Änderungen im Primär- und Sekundärmetabolismus während der Lagerung in den Kultivaren Arnika (gelb), Cilena (rot), Marabel (blau) und Turdus (grün) für die Lagerperiode 2007/2008 nach Anwendung der Einfaktoriellen Varianzanalyse bezüglich des Faktors „Zeit“.
- Seite 62    Abb. 19 Unabhängige Komponentenanalyse (ICA) angewendet auf alle detektierten GC-MS Signale im Hinblick auf die Lagerung von Kartoffelknollen der Kultivare Arnika (gelb), Cilena (rot), Granola (schwarz), Marabel (blau) und Turdus (grün) nach Applikation von A) Kümmelöl und B) Chlorpropham.
- Seite 63    Abb. 20 Änderungen im A) Gewichtsverlust und B) in der Keimung für die Kultivare Arnika, Cilena, Granola, Marabel und Turdus nach Behandlung mit Kümmelöl (schwarz), nach Behandlung mit Chlorpropham (rot) und ohne Behandlung bei gleicher Lagerungszeit für die Lagerperiode 2007/2008.

- Seite 64 Abb. 21 Relative zeitliche Änderungen im Primärmetabolismus während der Lagerung von Kartoffelknollen der Kultivare Arnika (gelb), Cilena (rot), Granola (schwarz), Marabel (blau) und Turdus (grün) nach Behandlung mit Kümmelöl.
- Seite 65 Abb. 22 Relative zeitliche Änderungen im Primärmetabolismus während der Lagerung von Kartoffelknollen der Kultivare Arnika (gelb), Cilena (rot), Granola (schwarz), Marabel (blau) und Turdus (grün) nach Behandlung mit Chlorpropham.
- Seite 69 Abb. 23 Allgemeiner Arbeitsfluss zur Extraktion prädiktiver Biomarker aus GCTOF-MS Profilen.
- Seite 70 Abb. 24 Relativer A) Fructose- und B) Glucosegehalt in den 20 untersuchten Kultivaren für den Standort Böhlendorf.
- Seite 71 Abb. 25 Genauigkeit in der Vorhersage der Qualität von Kartoffelchips für eine Züchtungspopulation.
- Seite 75 Abb. 26 Beurteilung der Güte der etablierten Entscheidungsbäume anhand der Trefferquote ( $r$ ), der Genauigkeit ( $p$ ) und des F-Wertes.
- Seite 79 Abb. 27 Evaluierung des Entscheidungsbaumverfahrens nach Anwendung auf eine mittlere chemische Mischung anhand der richtig positiven-, falsch positiven-, richtig negativen- und falsch negativen Menge unter Verwendung von Metabolitstrukturen.
- Seite 82 Abb. 28 Identifizierte Analyten und erwartete Unbekannte aus repräsentativen Extrakten von Kartoffelknolle (K) und Brokkoliröschen (B) nach Anreicherung an einer CatEx (Kationenaustauscher), einer AnEx (Anionenaustauscher) und Plexa Kartusche(C18).
- Seite 85 Abb. 29 Retentionszeiten ausgewählter Substanzen des Testmixes nach Optimierung der flüssigkeitschromatographischen Bedingungen.
- Seite 85 Abb. 30 Optimierte chromatographische Bedingungen zur Vorfraktionierung biologischer Extrakte an einer Cyanopropylsäule.
- Seite 86 Abb. 31 Hauptkomponentenanalyse angewendet auf alle detektierten im splitlosen Modus aufgenommenen GC-MS Signale.

- Seite 89    Abb. 32 Varianz innerhalb der gesammelten Fraktionen    nach A) split 50 GC  
Messung und B) NMR Messung innerhalb der gesammelten Fraktionen.
- Seite 91    Abb. 33 Pearsonkorrelation von ausgewählten Primärmetaboliten für biochemisch  
bedeutende Stoffklassen.

- Seite 31 Tab. 1 Präzision der Bestimmung des Retentionsindex wichtiger Analyten im Primärmetabolismus ausgedrückt als Standardabweichung ( $n = 45$ ) des Retentionsindex.
- Seite 34 Tab. 2 Transfer von Retentionsindices anderer Laboratorien auf das VAR5 Laboratorium und die daraus resultierende relative Retentionsindexabweichung.
- Seite 40 Tab. 3 Relative Standardabweichung in % verschiedener interner Standards in Bezug auf das jeweilige Experiment.
- Seite 49 Tab. 4 Nachweisgrenzen saurer Phytohormone im Hinblick auf die UPLC-FTICR- MS und die GC-TOF-MS Analyse unter der Annahme  $S/N \geq 2$ .
- Seite 61 Tab. 5 Korrelationskoeffizient ( $R$ ) und Signifikanzwert ( $p$ ) in Bezug auf den Gewichtsverlust und die Keimung nach Anwendung einer Pearsonkorrelation auf die gemessenen GC-TOF-MS Daten.
- Seite 66 Tab. 6 Signifikanzwert ( $p$ -Wert) der beiden sortenunabhängigen t-Tests (biologische Behandlung vs. keine Behandlung) der GC-MS detektierten Stoffwechselprodukte im Hinblick auf die Lagerung biologisch behandelter Knollen und unbehandelter Knollen während der Lagerperiode 2008/2009.
- Seite 76 Tab. 7 Gegenüberstellung wichtiger Klassifizierungsparameter für die Etablierung von Entscheidungsbäumen basierend auf Analytstruktur (AS) und Metabolitstruktur (MS).
- Seite 80 Tab. 8 Abschätzung der Robustheit des entwickelten Entscheidungsbaumverfahrens anhand der Klassifizierung von Hauptmetaboliten der Matrices „Reisblatt“ und „Bäckerhefe“.



Die vorliegende Dissertation wurde im Zeitraum von August 2007 bis Juli 2010 am Max-Planck-Institut für Molekulare Pflanzenphysiologie in Potsdam-Golm angefertigt. Diese Arbeit wurde durch ein BMBF Projekt und ein EU Projekt unterstützt. Für das mir entgegengebrachte Vertrauen und die Möglichkeit die Arbeit dort durchzuführen danke ich PD Dr. J. Kopka und Prof. L. Willmitzer.

Prof. L. Willmitzer und Dr. D. Walther danke ich ebenfalls für die Mitwirkung in meiner Kommission am MPIMP.

Weiterhin danke ich meinen Kooperationspartnern:

- Prof. T. van Beek (Universität Wageningen)
- Dr. P. Giavalisco (MPIMP Potsdam)
- Dr. J. Hummel (MPIMP Potsdam)
- Dr. R. Peters (Versuchsstation Dethlingen)
- Dr. U. Praeger (ATB Bornim )
- Dr. M. Steinfath (MPIMP Potsdam)
- Dr. E. Tacke (Böhm Nordkartoffel Agrarproduktion)
- Dr. J. van Dongen (MPIMP Potsdam)
- Prof. J. Vervoort (Universität Wageningen)

für das Probenmaterial, die Messungen und die zahlreichen Diskussionen,

dem GC-MS Service Team (I. Fehrle, Dipl. Ing. A. Erban) für die GC-MS Messungen,

meinen Praktikanten (M. Kuczmierczyk, K. Karunasekara und H. Zoehrer) für die Unterstützung im Labor

und der MPIMP Infrastrukturgruppe.

Weiterhin danke ich dem Büro 2.132 (A. Luedemann, K. Strassburg, L. von Malotky, J. Huege ) für die angenehme Arbeitsatmosphäre.

Mein ganz besonderer Dank gilt meiner Familie.

## **GC-TOF-MS basierte Analyse von niedermolekularen Primär- und Sekundärmetaboliten agrarwirtschaftlich bedeutsamer Nutzpflanzen**

Die Qualität von Nutzpflanzen ist von zahlreichen Einflussfaktoren wie beispielsweise Lagerbedingungen und Sorteneigenschaften abhängig. Um Qualitätsmängel zu minimieren und Absatzchancen von Nutzpflanzen zu steigern sind umfangreiche Analysen hinsichtlich ihrer stofflichen Zusammensetzung notwendig. Chromatographische Techniken gekoppelt an ein Massenspektrometer und die Kernspinresonanzspektroskopie wurden dafür bislang verwendet. In der vorliegenden Arbeit wurde ein Gaschromatograph an ein Flugzeitmassenspektrometer (GC-TOF-MS) gekoppelt, um physiologische Prozesse bzw. Eigenschaften (die Schwarzfleckigkeit, die Chipsbräunung, das Physiologische Alter und die Keimhemmung) von Nutzpflanzen aufzuklären. Als Pflanzenmodell wurde dafür die Kartoffelknolle verwendet.

Dazu wurden neue analytische Lösungsansätze entwickelt, die eine zielgerichtete Auswertung einer Vielzahl von Proben, die Etablierung einer umfangreichen Referenzspektrenbibliothek und die sichere Archivierung aller experimentellen Daten umfassen. Das Verfahren der Probenvorbereitung wurde soweit modifiziert, dass gering konzentrierte Substanzen mittels GC-TOF-MS analysiert werden können. Dadurch wurde das durch die Probenvorbereitung limitierte Substanzspektrum erweitert.

Anhand dieser Lösungsansätze wurden physiologisch relevante Stoffwechselprodukte identifiziert, welche indikativ (klassifizierend) bzw. prädiktiv (vorhersagend) für die physiologischen Prozesse sind. Für die Schwarzfleckigkeitsneigung und die Chipseignung wurde jeweils ein biochemisches Modell zur Vorhersage dieser Prozesse aufgestellt und auf eine Züchtungspopulation übertragen. Ferner wurden für die Schwarzfleckigkeit Stoffwechselprodukte des Respirationsstoffwechsels identifiziert sowie Aminosäuren, Glycerollipide und Phenylpropanoide für das Physiologische Alter als relevant erachtet. Das physiologische Altern konnte durch die Anwendung höherer Temperaturen beschleunigt werden. Durch Anwendung von Keimhemmern (Kümmelöl, Chlorpropham) wurde eine Verzögerung des physiologischen Alterns beobachtet. Die Applikation von Kümmelöl erwies sich dabei als besonders vorteilhaft. Kümmelöl behandelte Knollen wiesen im Vergleich zu unbehandelten Knollen nur Veränderungen im Aminosäure-, Zucker- und Sekundärstoffwechsel auf. Chlorpropham behandelte Knollen wiesen einen ähnlichen Stoffwechsel wie die unbehandelten Knollen auf.

Für die bislang noch nicht identifizierten Stoffwechselprodukte wurden im Rahmen dieser Arbeit das Verfahren der „gezielten An-/Abreicherung“, der „gepaarten NMR/GC-TOF-MS Analyse“ und das „Entscheidungsbaumverfahren“ entwickelt. Diese ermöglichen eine Klassifizierung von GC-MS Signalen im Hinblick auf ihre chemische Funktionalität. Das Verfahren der gekoppelten NMR/GC-TOF-MS Analyse erwies sich dabei als besonders erfolgversprechend, da es eine Aufklärung bislang unbekannter gaschromatographischer Signale ermöglicht.

In der vorliegenden Arbeit wurden neue Stoffwechselprodukte in der Kartoffelknolle identifiziert, wodurch ein wertvoller Beitrag zur Analytik der Metabolomik geleistet wurde.

## A Einleitung

In der Natur nehmen Pflanzen eine entscheidende Rolle ein. Sie sind nicht nur eine wichtige Nahrungsquelle, sondern auch ein entscheidender Sauerstoffproduzent für heterotrophe Organismen. Ohne Pflanzen gäbe es kein Leben auf unserer Erde. Obwohl mehr als 3000 Pflanzenarten genießbar sind, können 90 % aller Nahrungsmittel 20 Pflanzenarten zugeordnet werden. Die bedeutsamsten sind dabei Reis, Weizen, Gerste, Kartoffel, Tomate, Soja und Erbse. Diese werden als Nutzpflanzen bezeichnet<sup>[1]</sup>.

Sie werden nicht nur von der Nahrungsmittelindustrie genutzt, sondern auch von der Veredelungsindustrie. So wird zum Beispiel die Stärke der Kartoffelknolle nicht nur als Nahrungsmittel, sondern auch in der Papier- und Textilindustrie sowie bei der Herstellung von Zuckern verwendet. Das Ziel ist dabei in naher Zukunft noch resistenterer und ertragreichere Kulturpflanzen zu etablieren und zu kultivieren. Daher wird im Rahmen von interdisziplinären Forschungsprojekten versucht, in den Züchtungsprozess einzugreifen. Die Auswirkungen von Kältestress<sup>[2, 3]</sup>, Salzstress<sup>[4]</sup>, Schwermetallstress<sup>[5]</sup> und osmotischem Stress<sup>[6]</sup> auf die Pflanze wurden in der Vergangenheit in besonderem Maße untersucht. Dafür wurde hauptsächlich auf die Modellpflanze *Arabidopsis thaliana* zurückgegriffen. Die erworbenen Erkenntnisse sollen auf Nutzpflanzen übertragen werden, um sie auch in anderen Regionen anbauen zu können. Die Qualität von Nutzpflanzen soll möglichst vor Ort eingeschätzt werden, um gezielt in den Züchtungsprozess einzugreifen. Da hochwertige instrumentelle Analytik vor Ort oft nicht eingesetzt werden kann, wird das Ziel verfolgt, möglichst einfache analytische Verfahren zu entwickeln mit deren Hilfe die relevanten Substanzen vor Ort quantifizierbar sind. Dafür ist die Entwicklung von Schnelltests erforderlich, wie sie bereits in der Umweltanalytik eingesetzt werden. Die Zielsubstanzen werden aus großangelegten Auswahlverfahren identifiziert und verifiziert. Durch die Analyse von im Freiland angezogenen Pflanzen über einen Zeitraum von mehreren Jahren wird sichergestellt, dass die erhobenen Stoffwechselprodukte zuverlässig sind. Die Anwendung chromatographischer Trenntechniken, wie der Gaschromatographie (GC), der Flüssigkeitschromatographie (LC) und deren Kopplung an ein Massenspektrometer (MS) sowie der Kapillarelektrophorese (CE) und MS-Kopplung haben dies ermöglicht. Sie stellen einen ersten Schritt in Richtung Schnelltests dar. Die identifizierten Substanzen sollen in ferner Zukunft von Züchtern unter Verwendung geeigneter Schnelltests bestimmt werden, um die Qualität ihrer Ernte zu beurteilen.

### A.1 Verschiedene Omik-Technologien

Bis zur Etablierung dieser gekoppelten analytischen Methoden wurde der Extrakt einer Pflanze anhand von mehreren langwierigen Probenvorbereitungsschritten nasschemisch aufgereinigt und die Aktivität ausgewählter Enzyme bestimmt<sup>[7]</sup>. Die Analyse großer Probemengen blieb aufgrund des erhöhten Arbeitsaufwands aus. Die relevanten Enzyme wurden überwiegend aus Transkriptanalysen identifiziert, wodurch mit dieser Vorgehensweise keine neuen Stoffwechselwege gefunden wurden. Die Gesamtheit aller Stoffwechselwege ist jedoch für die

Regulation von biologischen Prozessen von großer Bedeutung. Daher ist die Kenntnis der Stoffwechselwege und ihrer Einflussfaktoren von außerordentlicher Relevanz für eine zielgerichtete Nutzung ausgewählter Pflanzen.

Transkriptanalysen hingegen sind aufgrund der hohen Kosten nur im begrenzten Umfang durchführbar. Weiterhin sind viele der wirtschaftlich bedeutsamen Nutzpflanzen noch nicht vollständig sequenziert, wodurch eine Analyse von Nutzpflanzen anhand der Transkriptomik nicht geeignet ist<sup>[8]</sup>.

Die seit Beginn dieses Jahrhunderts sich entwickelnde Wissenschaftsdisziplin der Proteomik umgeht diese Hürde. Sie ist allerdings auf die Analyse von Makromolekülen beschränkt und erlaubt daher lediglich Rückschlüsse auf Proteinfunktionen<sup>[9]</sup>. Änderungen im Metabolom sind jedoch im Vergleich zum Proteom viel umfassender und ermöglichen eine umfassende Charakterisierung von Nutzpflanzen und der darin ablaufenden biologischen Prozesse. Unter dem Metabolom fasst man alle Stoffwechselprodukte innerhalb einer Zelle beziehungsweise eines Gewebes zusammen. Die Umsetzung dieser wird von Enzymen reguliert.

## A.2 Die Wissenschaftsdisziplin der Metabolomik

Die Wissenschaftsdisziplin der Metabolomik wurde bislang zur Analyse pflanzlicher Extrakte<sup>[10]</sup>, Mikroorganismen<sup>[11-13]</sup> und Stoffwechselwegen in Säugetieren<sup>[14]</sup> erfolgreich angewandt. Aktuelle Untersuchungen zielen auf die Analyse einzelner Zellen ab, welche vor der Analyse zu einem repräsentativen Homogenat vereinigt werden<sup>[15]</sup>. Anhand von Metabolomik konnten Stoffwechselprodukte identifiziert und relevante Stoffwechselwege unter Verwendung von transgenen Pflanzen gezielt nachgewiesen werden<sup>[16]</sup>. Die Metabolomik beinhaltet die Analyse von Primär- und Sekundärmetaboliten, deren Größe 1000 Da nicht überschreitet und gibt Informationen zur Konzentration der detektierten Substanzen innerhalb der untersuchten Proben. Die große Herausforderung in der Analyse des Metaboloms besteht in der simultanen Erfassung und Quantifizierung möglichst aller Substanzen in einem Gewebe. Problematisch sind hierbei die verschiedenen Stoffklassen und der breite Konzentrationsbereich (pg/L bis mg/L). Die Zahl der endogenen Metaboliten im pflanzlichem Gewebe wurde von Fiehn et al. auf 200000 geschätzt<sup>[17]</sup>. Da die Menge der detektierten Metaboliten und der Konzentrationsbereich mit der Pflanze und dem zu analysierenden Organ der Pflanze variiert, können die Metaboliten bei Verwendung von nur einer analytischen Technologie nicht vollständig aufgeklärt werden. Es muss daher zu Beginn eines Experiments geklärt werden, welches Gewebe analysiert werden soll und wie hoch dessen Komplexität ist. Außerdem muss eine Hypothese zu den beteiligten Stoffwechselwegen aufgestellt werden, die dort ablaufen. Wird als Beispiel die Kartoffelknolle gewählt, so ist klar, dass diese eine sehr heterogene Struktur besitzt und nur reproduzierbar geprobt werden kann, indem eine Durchschnittsprobe aller Gewebereinheiten durch die Kartoffelknolle genommen wird<sup>[18]</sup>. Inhaltsstoffe der Kartoffelknolle sind Stärke, Proteine, Organische Säuren, Mineralstoffe, Aminosäuren, Zucker, Lipide, Polyphenole, Vitamine, Alkaloide und Nukleotide. Mit Ausnahme der Stärke, den Proteinen und den Mineralstoffen können diese Substanzen mit Hilfe der Gaschromatographie-Massenspektrometrie Kopplung

(GC-MS) analysiert werden<sup>[19]</sup>. Für andere Matrices, wie beispielsweise die Blüte von *Arabidopsis thaliana* oder die Frucht von *Vitis vinifera* ist die Flüssigkeitschromatographie gekoppelt mit der Massenspektrometrie (LC-MS) besser geeignet<sup>[20]</sup>. Diese Organe beinhalten eine Vielzahl von komplexen chemischen Verbindungen, welche mit Hilfe der GC nur schwer analysiert werden können, da derartige Verbindungen nur sehr schwer verdampfbar sind. Ein guter Übergang in die Gasphase ist jedoch für die gaschromatographische Analyse unabdingbar.

Die qualitative und quantitative Zusammensetzung eines Extraktes wird durch die Probenvorbereitung entscheidend beeinflusst. Die Extraktionseffizienz einer Substanz wird durch die Wahl des Extraktionsmittels, der Extraktionsdauer, der Temperatur und der Matrix bestimmt. Darüber hinaus kann eine komplexe Matrix chemische Reaktionen im Extrakt auslösen und Substanzen zum Teil abbauen, so dass diese schließlich in ihrer Originalstruktur nicht mehr detektierbar sind. Die Lagerung der Extrakte bei tiefen Temperaturen verlangsamt diese ungewollten Prozesse. Dazu muss die zu analysierende Probe zuvor in flüssigem Stickstoff schockgefroren und bis zur Extraktion bei -80 °C gelagert werden. Eine methanolische Extraktion (15 min) bei erhöhter Temperatur (70 °C) hat sich neben anderen Extraktionsvarianten zur Analyse polarer Substanzen bewährt<sup>[21]</sup>. Auf diese Weise wurden die Enzyme im Extrakt denaturiert und die Gleichgewichte sämtlicher ablaufender Reaktionen auf die Seite der Produkte verschoben. Ein Nachteil dieser Extraktionsmethode besteht darin, dass thermolabile Substanzen nicht analysiert beziehungsweise nur Abbauprodukte detektiert werden. Aus diesem Grund wurden weitere Extraktionsmethoden unter Einsatz von deutlich niedrigeren Temperaturen etabliert. Hierbei haben sich eine Mischung aus Wasser, Methanol und Chloroform (1 : 2.5 : 1; v/v/v), kaltes Methanol, saure sowie basische Extraktionsmittel wie Perchlorsäure, Trifluoressigsäure, Salzsäure, Kalilauge und Natronlauge für die Analyse von Primär- und Sekundärmetaboliten bewährt. Für die Analyse gering konzentrierter Substanzen wurde bislang die Festphasenextraktion (SPE) in einzelnen Laboratorien angewendet<sup>[22-24]</sup>. Der mögliche Verlust biochemisch relevanter Substanzen und der höhere experimentelle Aufwand dieser Probenvorbereitungstechnik führten dazu, dass sich die Festphasenextraktion nicht in der Hochdurchsatzanalytik der Metabolomik durchgesetzt hat. Auch die Vorteile der analytisch reineren Extrakte und die Möglichkeit einer genaueren Quantifizierung der Inhaltsstoffe haben nichts daran geändert. Dennoch konnte mit Hilfe der Festphasenextraktion die Funktionsweise ausgewählter Phytohormone, wie der Abscisinsäure (ABA), der Gibberellinsäure (GA3) und der Indoleessigsäure (IAA) in höheren Pflanzen aufgeklärt werden<sup>[25-32]</sup>.

Niedermolekulare Primär- und Sekundärmetaboliten wurden bisher mit Hilfe von GC-MS, LC-MS, CE-MS und der Kernspinresonanzspektroskopie (NMR) analysiert<sup>[33]</sup>. In Bezug auf die Massenspektrometrie wurden verschiedene Detektoren erfolgreich eingesetzt. Dazu zählen Quadrupole (QUAD)<sup>[34]</sup>, Ionenfallen (TRAP)<sup>[35]</sup> und Flugzeitmassenspektrometer (TOF)<sup>[10, 36]</sup>, welche an einen Gaschromatographen oder einen Flüssigkeitschromatographen gekoppelt wurden<sup>[37]</sup>. Nach gaschromatographischer Trennung wurden die Analyten schließlich chemisch oder durch Elektronenstoß (EI) ionisiert. Im Gegensatz dazu wurden an die flüssigkeitschromatographische Trennung eine Elektronensprayionisationseinheit (ESI) und ein hochauf-

lösendes Massenspektrometer gekoppelt<sup>[38, 39]</sup>. Jede dieser Analysentechnologien erfasst aufgrund des vorgeschalteten Probenvorbereitungsschrittes und des vorgegebenen Analysenprozesses nur ein begrenztes Substanzspektrum. Um leicht verdampfbare Analyten zu detektieren ist die gaschromatographische Analyse besonders gut geeignet. Diese können durch geeignete Derivatisierungsreagenzien, wie beispielsweise Silylierungs- oder Acetylierungsreagenzien erzeugt werden<sup>[40]</sup>. Die Verdampfbarkeit stellt für flüssigkeitschromatographische Analysen kein Ausschlusskriterium dar. Bei dieser Trenntechnik müssen die Analyten lediglich mit der mobilen Phase mischbar sein und dürfen nicht zu starke Wechselwirkungen mit der stationären Phase eingehen. Bei der NMR-Spektroskopie hingegen stellt die im Vergleich zur Massenspektrometrie geringe Sensitivität einen entscheidenden Nachteil dar. Diese kann durch den Einsatz von höheren Magnetfeldstärken, cryogenen Probenzellen und RF Techniken erhöht werden. Diese technischen Erneuerungen sind für NMR basierte metabolische Untersuchungen von Vorteil<sup>[33]</sup>.

Die gepaarte Analyse von Extrakten mit Hilfe der GC-MS, der LC-MS, der CE-MS und der NMR Spektroskopie ermöglicht eine umfassende Analyse niedermolekularer Primär- und Sekundärmetaboliten<sup>[41-44]</sup>. Die Zusammenführung der Messdaten ist bislang noch nicht automatisiert möglich, da alle Analysentechnologien unterschiedliche strukturelle Informationen liefern und ein unterschiedliches Substanzspektrum abdecken<sup>[45]</sup>.

In jüngster Zeit hat sich daher die Analysenmethode der Zweidimensionalen Gaschromatographie (GCxGC; englische Bezeichnung: Comprehensive Gas Chromatography) bewährt<sup>[46-51]</sup>. Die Chromatographie in der zweiten Dimension führt zu einer weiteren Auftrennung der Signale im Chromatogramm und somit zu zusätzlichen wertvollen Strukturinformationen. Die Verwendung dieser Technologie zur Aufnahme von metabolischen Profilen hat sich allerdings in der Routineanalyse aufgrund der großen Datenmenge und mangelnden Robustheit noch nicht durchgesetzt.

Neue analytische Lösungsansätze sind notwendig, um die umfangreichen Messinformationen innerhalb kürzester Zeit auf die biologische Fragestellung anzuwenden. Die Überlegungen beinhalten die sichere Archivierung der gewonnenen Rohdaten, die Archivierung von Strukturdaten der bisher identifizierten analytischen Signale und robuste Parameter für den Vergleich von Referenz- und Experimentmassenchromatogrammen. Beispielsweise verfügt ein GC-MS Massenchromatogramm über zwei Informationsebenen: die Retentionszeit und das Massenspektrum. Beide Parameter sind für eine eindeutige Identifizierung unabdingbar. Die Retentionszeit einer Verbindung resultiert aus den intermolekularen Wechselwirkungen mit dem Säulenmaterial, während das Massenspektrum aus der harten Elektronenstoßionisation im Massenspektrometer zugänglich ist. Die Retentionszeit ermöglicht eine eindeutige Identifizierung von Stereoisomeren, also von Verbindungen mit einem gleichen Massenspektrum, solange diese basisliniengetrennt sind. Die Elektronenstoßionisation führt zu einem äußerst robusten Fragmentierungsverhalten und ist für die Erstellung von Referenzbibliotheken geeignet<sup>[52]</sup>. Für die anderen Analysentechnologien, wie die LC-MS und die NMR Spektroskopie ergeben sich andere Identifizierungsparameter. Bei LC-MS Analysen sind dies die Retentionszeit und die exakte Masse. Bei LC-MS/MS Analysen zählt zu den

Identifizierungsparametern neben der Retentionszeit die Fragmentmasse der Tochterionen. Bei NMR Analysen spielt die chemische Verschiebung und die Kopplungskonstante der Resonanzen eine entscheidende Rolle.

Zur Speicherung der Information innerhalb einer Datenbank wurden neue Programme geschrieben. Diese können die aufgenommenen Messinformationen innerhalb einer Matrix in Form der Retentionszeit und Fragmentmasse x Chromatogramm anordnen. Ausgewählte Beispiele sind die Programme „TagFinder“<sup>[53]</sup>, „Metalign“<sup>[54]</sup>, „MeltDB“<sup>[55]</sup>, „TargetSearcher“<sup>[56]</sup>, „Metmax“<sup>[46]</sup> und „XCMS“<sup>[57]</sup>. Diese Programme werten die metabolischen Daten nach unterschiedlichen Gesichtspunkten aus. Für alle Programme ist jedoch eine umfangreiche Spektrenbibliothek erforderlich, welche anhand von kommerziellen Referenzsubstanzen verifiziert wurde. Die „**Go**lm **Metabolome Database**“ (GMD)<sup>[58, 59]</sup>, „Massbank“<sup>[60]</sup>, die vom Fiehn Laboratorium vertriebene Referenzbibliothek<sup>[61]</sup> und die vom NIST Institut (**N**ational **I**nstitute of **S**tandards and **T**echnology) erhältliche Massenspektrenbibliothek<sup>[62, 63]</sup> sind hierfür nur einige Beispiele. Besonders vorteilhaft für die Analyse des Metaboloms sind die von analytischen Laboratorien etablierten Spektrenbibliotheken, da diese von silylierten und methoxyaminierten Verbindungen ausgehen und ebenfalls „erwartete Unbekannte“ unter Angabe der jeweiligen Pflanze beinhalten. Massenchromatogramme, denen bisher noch keine chemische Struktur zugeordnet werden konnte werden als erwartete unbekannte Massenchromatogramme bezeichnet<sup>[64]</sup>. Diese wurden im Rahmen mehrfacher Wiederholungen als biologisch relevant erachtet. Damit die bekannten gaschromatographischen Signale und die erwarteten unbekanntes gaschromatographischen Signale innerhalb von mehreren analytischen Laboratorien verfolgt werden können, wurden am Max Planck Institut für Molekulare Pflanzenphysiologie A-Nummern (**A**lyt-**N**ummer) eingeführt. Eine A-Nummer leitet sich aus dem Retentionsindex eines gaschromatographischen Signals ab und ist für jede detektierte Verbindung eindeutig. Da der Retentionsindex bis auf zwei Nachkommastellen reproduzierbar bestimmbar ist, ergibt sich durch Multiplikation des Retentionsindex mit 100 ein sechsstelliger eindeutiger Bezeichner (engl. identifier)<sup>[59]</sup>. Dieser ist im Rahmen der Strukturierung metabolischer Daten ebenfalls für die entsprechenden Metaboliteneinträge eingeführt worden. Der Metabolitenbezeichner gruppiert die neuangelegten Metaboliten nicht entsprechend einer vorgegebenen Logik, sondern erhöht sich für jeden neuangelegten Metaboliten um eine Einheit. Unter einem Metaboliteneintrag werden alle natürlich vorkommenden Stereoisomere und Isotopomere einer chemischen Verbindung zusammengefasst. Ein Metaboliteneintrag kann nach Anwendung der Derivatisierung mehreren gaschromatographischen Signalen zugeordnet werden, falls dieser über mehrere reaktive Gruppen verfügt. Der Metabolitenbezeichner einschließlich der Isotopomere und Stereoisomere ist für eine chemische Verbindung eindeutig. Für eine eindeutige Identifizierung von Signalen innerhalb von großangelegten Messserien sind glaubwürdige Grenzwerte notwendig, die eine eindeutige Identifizierung von Massenchromatogrammen ermöglichen. Derzeit sind allerdings noch keine allgemeinen Grenzwerte für den Retentionsindexabgleich abgesteckt worden. In Bezug auf den massenspektralen Vergleich verfügt das Programm NIST über allgemeine Grenzwerte und individuelle Parameter, die einen Vergleich ermöglichen. Die Auswertung großangelegter Studien zur Analyse von Primär- und

Sekundärmetaboliten ist mit diesem Programm sehr zeitaufwendig, da keine Retentionsindexeinschränkungen vorgenommen werden können. Ein visueller Vergleich zwischen Referenzspektrum und Experimentspektrum ist oft erforderlich, um die wahre Identität des Signals sicherzustellen.

Weiterhin sind chemisch definierte Referenzmixturen notwendig, um die entsprechenden Signale zu verifizieren. Die Zusammensetzung dieser Referenzmixturen hängt von der biologischen Fragestellung und der zu analysierenden Matrix ab. Für die Zusammenstellung der chemischen Mischung ist entscheidend, ob eine bestimmte Verbindung oder Verbindungsklasse analysiert oder ein Stoffwechselweg näher untersucht werden soll.

### A.3 Probleme und Lösungsansätze metabolischer Untersuchungen

Um die verbleibenden unbekanntes Signale eines Experimentes hinsichtlich ihrer Funktionalität zu klassifizieren, wurden diese bisher über eine bestmögliche Übereinstimmung mit der Spektrenbibliothek abgeglichen und die Ähnlichkeit des gemessenen Signals im Vergleich zu dem Signal der Spektrenbibliothek als Maß angegeben. Auf diese Weise konnten Stereoisomere identifiziert werden, welche bislang noch nicht in der Spektrenbibliothek vertreten sind. Weiterhin konnte die Ähnlichkeit zu einer bestimmten Verbindungsklasse nachgewiesen werden<sup>[65]</sup>.

Ein anderer Ansatz besteht darin, durch die Untersuchung von kommerziell erhältlichen Chemikalien unbekannte Massensignale aufzuklären<sup>[66]</sup>. Auf diese Weise konnte die Spektrenbibliothek um weitere Massenchromatogramme erweitert und identifizierte Signale verifiziert werden. Die gezielte Aufklärung unbekannter Signale blieb jedoch auf diesem Weg aus. Des Weiteren sind neue biologisch relevante Substanzen, die noch nicht in der Datenbank vorhanden sind, auf diesem Weg nicht zugänglich. Zu den größten Ausschlusskriterien dieser Strategie zählen neben der kommerziellen Verfügbarkeit der hohe Preis analytisch reiner Standards. Obwohl die Identifizierung im Nanogrammbereich möglich ist, muss die Referenzchemikalie mindestens einmal erworben werden (kleinste Abpackung: 1mg). Zwar sind Primärmetaboliten zu einem großen Teil kommerziell erhältlich, die Analyse von Sekundärmetaboliten stellt jedoch ein großes Problem dar. Im Allgemeinen sind Sekundärmetaboliten sehr komplexe Moleküle, die durch komplizierte und aufwendige Syntheseschritte hergestellt werden können. Die Substanzen können ebenfalls aus dem pflanzlichen Extrakt isoliert und anhand von weiteren analytischen Technologien näher charakterisiert werden<sup>[67]</sup>.

Eine weitere Strategie bei der Aufklärung unbekannter Signale ist der Einsatz von multiplen Derivatisierungsstrategien und die Anzucht von  $^{13}\text{C}$ ,  $^{15}\text{N}$  und  $^2\text{H}$  markierten Pflanzen<sup>[68]</sup>. Die Kenntnis der Zahl der funktionellen Gruppen und der Elementarzusammensetzung von Fragmenten haben dazu geführt, dass einzelne Substanzen aufgeklärt werden konnten. Ein Nachteil dieser Strategie ist der hohe Aufwand und die Beschränkung auf kleine Moleküle.

Bioinformatische Arbeitsgruppen verfolgen daher das Konzept des maschinengestützten Lernens funktioneller Gruppen. Im Zuge metabolischer Untersuchungen wird eine Vielzahl von



Signalen aufgenommen, denen eine chemische Struktur zugeordnet werden kann. So konnte der Retentionsindex von Peng und van Ham für einige Verbindungen modelliert werden und die Elementzusammensetzung unbekannter Verbindungen von der Arbeitsgruppe Stein vorhergesagt werden<sup>[69-73]</sup>. Varmuza, Werther und Mistrik nutzten massenspektrale Fragmente zur Vorhersage struktureller Einheiten. Dabei erwiesen sich Neuronale Netzwerke als besonders erfolgversprechend<sup>[74]</sup>. Zur Klassifizierung von EI Spektren wurden außerdem die Hauptkomponentenanalyse (PCA)<sup>[75]</sup>, Clustering-Algorithmen und Selbstorganisierende Abbildungen (SOM) verwendet<sup>[76]</sup>. MS-MS Spektren wurden kürzlich von der Arbeitsgruppe Neumann modelliert<sup>[76]</sup>.

Obwohl die Wissenschaftsdisziplin der Metabolomik ständigen Veränderungen ausgesetzt ist, hat sie dazu beigetragen, die grüne Biotechnologie zu revolutionieren. Gängige analytische Technologien, wie die GC-MS, die LC-MS, die CE-MS und die NMR Spektroskopie haben dazu beigetragen. Durch die Analyse niedermolekularer Primär- und Sekundärmetaboliten konnten insbesondere in der Modellpflanze *Arabidopsis thaliana* Substanzen identifiziert werden, die für biologische Prozesse entscheidend sind. Diese konnten mit Hilfe von weiteren „Omik“ Technologien, wie der Analyse des Transkriptoms und des Proteoms bestätigt werden. Diese Erkenntnisse werden jetzt auf Nutzpflanzen, wie Reis, Mais, Weizen und Kartoffel übertragen.

#### A.4 Motivation und Zielsetzung

Daher sind neue analytische Ansätze notwendig, die eine schnelle Analyse einer Vielzahl von Proben und Identifizierung möglichst aller Metaboliten ermöglicht, um biochemisch relevante Stoffwechselprodukte von Nutzpflanzen zu analysieren. Die identifizierten Substanzen sind Grundlage für die Etablierung photometrischer Schnelltests und daher für die Nahrungsmittelindustrie von großem Interesse. Diese Substanzen sollen aus GC-TOF-MS Analysen gewonnen werden, welche dafür mit den entsprechenden phänotypischen Daten in Zusammenhang gebracht werden.

Im Rahmen dieser Arbeit wurde die Kartoffelknolle als Modell ausgewählt und mittels GC-TOF-MS analysiert. Die Kartoffelknolle wurde bereits sehr intensiv untersucht und wird für viele Zwecke angebaut. Ernteverluste aufgrund von Schwarzfleckigkeit oder frühzeitiger Keimung sollten daher vermieden werden. Bisher konnten keine robusten Stoffwechselprodukte in Bezug auf die Keimung und die Schwarzfleckigkeit gelagerter Kartoffelknollen identifiziert werden.

Das Ziel dieser Arbeit lag daher in der Identifizierung physiologisch relevanter Stoffwechselprodukte in der Kartoffelknolle. Weiterhin soll durch diese Arbeit nicht nur die GC-TOF-MS basierte Analyse niedermolekularer Primär- und Sekundärmetaboliten erleichtert werden, sondern auch ein Beitrag zur Klassifizierung unbekannter gaschromatographischer Signale geleistet werden. Auf diese Weise sollen bislang unbekannte GC-MS Signale hinsichtlich ihrer chemischen Funktionalität eingeordnet werden.

Ausgewählte Ergebnisse dieser Arbeit wurden in folgenden Journalen veröffentlicht und sind im Anhang II zusammengefasst.

- **Strehmel N**, Hummel J, Erban A, Strassburg K, Kopka J (2008) Retention index thresholds for compound matching in GC-MS metabolite profiling. *Journal of Chromatography B* 871: 182-190.
- **Strehmel N**, Praeger U, König C, Fehrle I, Erban A, Geyer M, Kopka J, van Dongen J (2010) Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impact. *Postharvest Biology and Technology*, 56(2): 109-116.
- Steinfath M, **Strehmel N**, Peter R, Groth D, Hummel J, Steup M, Selbig J, Kopka J, van Dongen J (2010) Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach *Plant Biotechnology Journal*, 8: 900–911.
- Hummel J, **Strehmel N**, Selbig J, Walther D, Kopka J (2010) Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics*, 6(2):322-333.
- Kempa S, Hummel J, Schwemmer T, Pietzke M, **Strehmel N**, Wienkoop S, Kopka J, Weckwerth W (2009) An automated GCxGC-TOF-MS protocol for batch-wise extraction and alignment of mass isotopomer matrixes from differential <sup>13</sup>C-labelling experiments: a case study for photoautotroph-mixotroph grown *Chlamydomonas reinhardtii* cells. *Journal of Basic Microbiology* 49(1): 82-91.

Weiterhin wurden Resultate dieser Arbeit auf Konferenzen und an Forschungsinstituten in Form von Vorträgen oder Postern vorgestellt:

- **Strehmel N**, Steinfath M, Tacke E, Hummel J, Fehrle I, van Dongen J, Geigenberger P, Kopka J, Potentielle metabolische Biomarker für die Kartoffelzüchtung, *ANAKON März 2009 Berlin*
- **Strehmel N**, Steinfath M, Hummel J, Kopka J, Assessing Potato Tuber Quality by means of GC-TOF-MS Metabolite Profiling, *Doktorandenseminar Februar 2010 Attendorn*
- Hummel J, **Strehmel N**, Selbig J, Walther D, Kopka J, The Golm Metabolome Database – a valuable resource for metabolomic studies, *Metabolomics & More März 2010 Freising*
- **Strehmel N**, Giavalisco P, Kuczmierczyk M, Erban A, Kopka J, Solid Phase Extraction (SPE) – An essential technology for enhanced metabolite profiling studies, *Metabolomics & More März 2010 Freising*
- **Strehmel N**, Metabolic Biomarkers for Potato Breeders, metanomics GmbH, eingeladener Vortrag, Februar 2010
- **Strehmel N**, Assessing Potato Tuber Quality by means of non-targeted GC-TOF-MS based Metabolite Profiling, IPB Halle, eingeladener Vortrag, April 2010

## B Materialien und Methoden

### B.1 Materialien

#### B.1.1 Chemikalien und Standards

Im Folgenden sind die im Rahmen dieser Arbeit verwendeten Chemikalien und Standards zusammen mit ihrem Hersteller aufgelistet.

- Festphasenextraktion
  - Abscisic Acid: TCI Europe, Zwijndrecht, Belgien (CAS: 21293-29-8)
  - Ameisensäure: Sigma, München, Deutschland (CAS: 64-18-6)
  - Ammoniak: Sigma, München, Deutschland (CAS: 7664-41-7)
  - Essigsäure: Sigma, München, Deutschland (CAS: 64-19-7)
  - GA4: OlChem, Olomouc, Tschechien (CAS: 468-44-0)
  - GA7: OlChem, Olomouc, Tschechien (CAS: 510-75-8)
  - Gibberellic Acid: Sigma, München, Deutschland (CAS: 77-06-5)
  - Indole-3-acetic Acid: Sigma, München, Deutschland (CAS: 87-51-4)
  - Tryptophan: Sigma, München, Deutschland (CAS: 73-22-3)
  - Zeatin: Sigma, München, Deutschland (CAS: 1637-39-4)
  
- Präparative Flüssigkeitschromatographie
  - Acetonitril: Merck, Darmstadt, Deutschland (CAS: 75-05-8)
  
- Extraktion polarer Substanzen
  - Chloroform: Merck, Darmstadt, Deutschland (CAS: 67-66-3)
  - Destilliertes Wasser (0.055µS/cm): USF 800, Ransbach-Baumbach, Deutschland (CAS: 7732-18-5 )
  - D(-)-Isoascorbinsäure: Sigma, München, Deutschland (CAS: 89-65-6)
  - Methanol: Merck, Darmstadt, Deutschland (CAS: 67-56-1)
  - Ribitol: Sigma, München, Deutschland (CAS: 488-81-3)
  
- Derivatisierung polarer Substanzen für die GC-TOF-MS Analyse
  - n-Decan: Sigma, München, Deutschland (CAS: 124-18-5)
  - n-Docosan: Sigma, München, Deutschland (CAS: 629-97-0)
  - n-Dodecan: Sigma, München, Deutschland (CAS: 112-40-3)
  - n-Dotriacontan: Sigma, München, Deutschland (CAS: 544-85-4)
  - n-Hexatriacontan: Sigma, München, Deutschland (CAS: 630-06-8)
  - *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamid (MSTFA): MACHEREY-NAGEL GmbH & Co. KG, Düren, Deutschland (CAS: 24589-78-4)
  - n-Nonadecan: Sigma, München, Deutschland (CAS: 629-92-5)

- *N,O*-Bis(trimethylsilyl)trifluoroacetamid (BSTFA): MACHEREY-NAGEL GmbH & Co. KG, Düren, Deutschland (CAS: 25561-30-2)
- n-Octacosan: Sigma, München, Deutschland (CAS: 630-02-4)
- n-Octadecan: Sigma, München, Deutschland (CAS: 593-45-3)
- n-Pentadecan: Sigma, München, Deutschland (CAS: 629-62-9)
- Methoxyaminhydrochlorid: Sigma, München, Deutschland (CAS: 593-56-6)
- Pyridin: Sigma, München, Deutschland (CAS: 110-86-1)
  
- Retentionsindex- und Retentionszeituntersuchungen
  - Methylarachidat: Sigma, München, Deutschland (CAS: 1120-28-1)
  - Methylbehenat: Sigma, München, Deutschland (CAS: 929-77-1)
  - Methyldecanoat: Sigma, München, Deutschland (CAS: 29619-64-5)
  - Methyldecanoat: Sigma, München, Deutschland (CAS: 111-82-0)
  - Methylhexacosanoat: Sigma, München, Deutschland (CAS: 5802-82-4)
  - Methylmelissat: Sigma, München, Deutschland (CAS: 629-83-4)
  - Methylmyristat: Sigma, München, Deutschland (CAS: 124-10-7)
  - Methylnonadecanoat: Sigma, München, Deutschland (CAS: 1731-94-8)
  - Methylnonanoat: Sigma, München, Deutschland (CAS: 1731-84-6)
  - Methyloctanoat: Sigma, München, Deutschland (CAS: 111-11-5)
  - Methyloctacosanoat: Sigma, München, Deutschland (CAS: 55682-92-3)
  - Methylpalmitat: Sigma, München, Deutschland (CAS: 112-39-0)
  - Methylstearat: Sigma, München, Deutschland (CAS: 112-61-8)
  - Methyltetracosanoat: Sigma, München, Deutschland (CAS: 2442-49-1)
  
- Interne Standards zur Verifizierung von Extraktions- und Derivatisierungsverlusten
  - 4-Aminobutan-d6-säure: Isotec, Illinois, USA (CAS: 14493-42-6)
  - Arginin, 13C6,15N4: Isotec, Illinois, USA (CAS: 202468-25-5)
  - DL-Alanin 2,3,3,3-d4: Sigma, München, Deutschland (CAS: 53795-92-9)
  - Ethanol-d4-amin: Isotec, Illinois, USA (CAS: 85047-08-1)
  - Glucose, 13C6: Isotec, Illinois, USA (CAS: 110187-42-3)
  - Glutamin, 13C5: Sigma, München, Deutschland (CAS: 184161-19-1)
  - Glycerol-d8: Isotec, Illinois, USA (CAS: 7325-17-9)
  - Leucin-d3: Isotec, Illinois, USA (CAS: 87828-86-2)
  - Lysin-d4-dihydrochlorid: Isotec, Illinois, USA (CAS: 284664-88-6)
  - Prolin, 13C5,15N: Isotec, Illinois, USA (CAS: 202407-23-6)
  - Putrescin-d4: Isotec, Illinois, USA (CAS: 88972-24-1)
  - Saccharose, 13C12: Isotec, Illinois, USA (CAS: 41055-68-9)
  - Serin, 13C3, 15N: Isotec, Illinois, USA (CAS: 202407-34-9)
  - Succin-d4-säure: Sigma, München, Deutschland (CAS: 14493-42-6)
  - Tryptophan-d5: Isotec, Illinois, USA (CAS: 90424-95-6)
  - Uracil, 2-13C, 15N2: Isotec, Illinois, USA (CAS: 181517-11-3)

- Valin-d8: Sigma, München, Deutschland (CAS: 349086-38-0)
- Zitronen-d4-säure: Isotec, Illinois, USA (CAS: 147664-83-3)

### B.1.2 Pflanzenmaterial

Das Material für die Versuche „Sortenscreening“ und „Zuchtklone“ von Kartoffeln wurde von der Böhm-Nordkartoffel-Agrarproduktion Ebstorf bezogen. Für die Versuche „Keimhemmung“ und „Physiologisches Alter“ wurden die Kartoffelproben von der Versuchsstation Dethlingen erhalten. Das Ziel dieser Versuchsreihen lag in der Detektion physiologisch relevanter Stoffwechselprodukte in Bezug auf die Kartoffelknolle. Die Kartoffeln wurden dafür jeweils unter Freilandbedingungen angezogen.

Das Material für den Versuch der „gepaarten NMR und GC-TOF-MS Analyse“ von *Arabidopsis thaliana* und *Brassica napus*, sowie von *Brassica oleracea* var. *italica*, *Brassica oleracea* var. *gemmifera* und *Brassica oleracea* var. *gongylodes* wurde von Kooperationspartnern des Metaphorprojektes (Großbritannien, Frankreich) bereitgestellt. Das Ziel dieser Versuchsreihe lag in der Klassifizierung unbekannter GC-TOF-MS Signale in Nutzpflanzen. Dafür wurden für diese Versuchsreihe bewusst Pflanzen der Kreuzblütlerfamilie ausgewählt, da diese weitaus mehr Metaboliten beinhalten als die Kartoffelknolle.

#### B.1.2.1 Versuch „Sortenscreening“

Die wirtschaftlich bedeutsamen Kartoffelsorten *Afra*, *Agria*, *Alliance*, *Arnika*, *Calla*, *Cilena*, *Elfe*, *Festien*, *Goldika*, *Granola*, *Kuras*, *Lady Claire*, *Marabel*, *Milva*, *Omega*, *Red Fantasy*, *Roberta*, *Tomensa*, *Topas* und *Turdus* für den Versuch „Sortenscreening“ wurden 2006 und 2007 in Ebstorf (53°01'43.62"N; 10°24'54.61"E) und in Böhlendorf (54°35'59.95"N; 13°17'46.12"E) angezogen. Die Kartoffeln wurden direkt nach der Ernte hinsichtlich ihrer Inhaltsstoffe analysiert und anschließend bei niedrigen Temperaturen gelagert (4 °C), um eine Keimung der Knollen zu verzögern. Nach einer Lagerzeit von vier Monaten wurden die Kartoffelknollen im Januar auf ihre Eignung zur Herstellung von Chips und ihre Schwarzfleckigkeitsneigung untersucht. Dies wurde in Kooperation mit Herrn Dr. Tacke von der Böhm-Nordkartoffel-Agrarproduktion (BNA), durchgeführt. Die erhaltenen Daten wurden anschließend mit den metabolischen Daten zusammengeführt, um ein Modell zur Schwarzfleckigkeit und zur Chipseignung aufzustellen.

#### B.1.2.2 Versuch „Zuchtklone“

Im Jahr 2008 wurde eine Kreuzungsnachkommenschaft von Kartoffelknollen in Ebstorf (53°01'43.62"N; 10°24'54.61"E) angebaut, die 58 Nachkommen umfasste. Diese wurden nach der Ernte bei 4 °C eingelagert und im Januar 2009 entsprechend dem Versuch „Sortenscreening“ auf ihre Eignung zur Herstellung von Kartoffelchips und auf ihre Schwarzfleckigkeitsneigung untersucht. Dieser Versuch war erforderlich, um die etablierten Modelle aus Versuch B.1.2.1 zu verifizieren.

### B.1.2.3 Versuch „Keimhemmung“

Die folgenden Versuchsreihen wurden in Kooperation mit Herrn Dr. Peters von der Versuchsstation Dethlingen durchgeführt. Hierzu wurden die sehr unterschiedlich alternden Sorten *Annabelle*, *Arnika*, *Carmona*, *Cilena*, *Festien*, *Granola*, *Jelly*, *Karlana*, *Lady Claire*, *Likaria*, *Lolita*, *Marabel*, *Opal*, *Pirol*, *Roberta*, *Satina*, *Tomensa*, *Turdus*, *Verdi* und *Vienna* für den Versuch „Keimhemmung“ 2008 in Dethlingen (52°57'06.23"N; 10°07'41.95"E) angebaut und nach ihrer Ernte von September 2008 bis April 2009 bei 5 °C eingelagert. Dabei wurden 50 % der Kartoffeln im Oktober 2008 mit einem biologischen Keimhemmer (Kümmelöl) behandelt und die anderen 50 % unbehandelt weiterhin bei 5 °C gelagert. Die gelagerten Knollen wurden im Dezember 2008 (kurz nach der Einlagerung) und im März 2009 (kurz vor der Auslagerung) bezüglich ihrer Inhaltsstoffzusammensetzung untersucht, um physiologisch relevante Stoffwechselprodukte zu identifizieren.

Die Sorten *Arnika*, *Cilena*, *Granola*, *Marabel* und *Turdus* wurden 2006 und 2007 in Dethlingen angebaut. Die Erntemenge wurde anschließend gedrittelt, wobei ein Drittel mit einem biologischen Keimhemmer (Kümmelöl) und ein weiteres Drittel mit einem chemischen Keimhemmer (Chlorpropham) behandelt wurde. Der Rest wurde als Kontrolle unbehandelt eingelagert. Die Lagerung aller Fraktionen erfolgte bei 5 °C. Die gewünschte Temperatur wurde durch den Einsatz von kühler Außenluft beziehungsweise maschineller Kühlung erreicht. Die gelagerten Knollen wurden während der Lagersaison 2006/2007 im Oktober, Dezember, Februar, April und in der darauf folgenden Saison aufgrund der veränderten Witterungsverhältnisse im November, Dezember, Februar und März auf ihren Phänotyp und ihre Inhaltsstoffzusammensetzung näher untersucht. Da der biologische Keimhemmer erst wenige Wochen nach der Einlagerung angewendet wurde, konnten die biologisch behandelten Knollen nur zu drei Zeitpunkten untersucht werden (Dezember, Februar und März).

- Biologische Keimhemmung:

Bei der biologischen Keimhemmung wurden die Kartoffelknollen unbehandelt eingelagert und jeweils ab Ende Oktober wöchentlich mit Kümmelöl behandelt. Die eingesetzte Menge an Kümmelöl variierte nach einem vom Hersteller vorgegebenen Schema mit abnehmender Konzentration von 25 mL/t Kartoffel bis 10 mL/t Kartoffel. Vier Wochen vor der Auslagerung wurde die Behandlung beendet, um eine starke Anreicherung von Kümmelöl in der Knolle zu vermeiden.

- Chemische Keimhemmung:

Bei der chemischen Keimhemmung wurden die Kartoffelknollen vor der Einlagerung einmalig mit Chlorpropham (Isopropyl-3-chlorcarbanilat) besprüht. Die Aufwandmenge betrug 60 mL/t Kartoffel, um den vorgeschriebenen Grenzwert für die Nahrungsmittelindustrie während der Auslagerung nicht zu überschreiten.

#### B.1.2.4 Versuch „Physiologisches Alter“

Die Sorten *Arnika*, *Cilena*, *Granola* und *Turdus* wurden 2006 in Dethlingen angebaut, nach der Ernte bei 5 °C unbehandelt eingelagert und im folgenden Jahr erneut ausgesät. Das erhaltene Pflanzgut wurde wiederum bei 5 °C eingelagert, kurz vor der Auslagerung auf 4 °C und 8 °C konditioniert und vier Wochen später beprobt, um Stoffwechselprodukte zu identifizieren welche das Physiologische Alter beschleunigen.

#### B.1.2.5 Versuch „Mechanische Belastung von Kartoffelknollen“

Die mechanischen Belastungsexperimente an Kartoffelknollen wurden in Kooperation mit Dr. Ulrike Praeger vom Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. durchgeführt. Dafür wurden die Sorten *Afra* und *Milva* von einer höhenverstellbaren Fallvorrichtung aus 50 cm Höhe zehnmal auf die Seite fallen gelassen. Um sicherzustellen dass jeweils die gleiche Stelle der Knolle belastet wird, wurde diese markiert. Nach dem zehnten Fall wurde die belastete und die gegenüberliegende unbelastete Knollenseite metabolischen Untersuchungen unterzogen. Dazu wurden Proben in Zeitabständen von 30 min, 6 h, 24 h und 48 h nach der mechanischen Belastung genommen und auf ihre Inhaltsstoffzusammensetzung mittels GC-TOF-MS untersucht, mit dem Ziel Änderungen im Primärmetabolismus nach mechanischer Belastung festzustellen.

#### B.1.2.6 Versuch „Gepaarte NMR und GC-TOF-MS Analyse“

Für den Versuch der „gepaarten NMR und GC-TOF-MS Analyse“ wurden *Arabidopsis thaliana* des Ökotyps Col-0 und *Brassica napus* der Gattung Drakkar unter kontrollierten Bedingungen im Gewächshaus in Golm angezogen. Von fünf Rapspflanzen der Gattung Drakkar wurde neun Wochen nach der Aussaat eine repräsentative Blattprobe entnommen, das Material zu einer Probe vereinigt und eine gepaarte Analyse durchgeführt. Dazu wurden die erhaltenen Extrakte mittels präparativer Flüssigkeitschromatographie fraktioniert und die erhaltenen Fraktionen mit Hilfe von <sup>1</sup>H-NMR und GC-TOF-MS analysiert. Schließlich wurden beide Datensätze über eine Pearson Korrelation zusammengeführt<sup>[77]</sup>. Zur Untersuchung von *Arabidopsis Thaliana* wurden mehrfach ca. 30 Pflanzen an 46-52 Tagen nach der Aussaat geerntet, die erhaltenen Blattproben zu einer repräsentativen Probe vereinigt und eine gepaarte Analyse durchgeführt. Für diese Versuchsreihe wurden mehrere *Arabidopsis*-pflanzen geerntet, um die gleiche Menge an biologischem Material für die weitere Analyse zur Verfügung zu haben.

Die weiteren Proben (Brokkoli „1852“, Rosenkohl „Bedford“, Rosenkohl „Maximus“, Weißkohl „Greyhound“, Weißkohl „Savoy Ormskirk Late“, Brokkoli „GD33 GK020779“ und Kohlrabi „Kolibri bulbous stem“) wurden von Kooperationspartnern des Metaphorprojektes erhalten und unter Freilandbedingungen angezogen. Bei diesen Nutzpflanzen wurde jeweils der essbare Teil mittels präparativer Flüssigkeitschromatographie fraktioniert und die Fraktionen anschließend mit Hilfe von <sup>1</sup>H-NMR und GC-TOF-MS analysiert.

### B.1.2.7 Etablierung der Analytik von Phytohormonen

Zur Etablierung der Analytik von Phytohormonen in pflanzlichem Gewebe wurde in flüssigem Stickstoff gemörseretes Arabidopsisblatt beziehungsweise entsprechend aufbereitete Kartoffelknolle verwendet. Die Arabidopsispflanzen wurden aus dem Gewächshaus in Golm bezogen. Keimende Kartoffelknollen der Sorten *Granola* und *Molly* wurden privat bezogen und bis zur Prozessierung entsprechend Kap. B.2.5 bei kühlen Temperaturen im Dunkeln gelagert.

### B.1.3 Geräte und Hilfsmittel

Die in dieser Arbeit verwendeten Geräte und Hilfsmittel sowie deren Hersteller sind im Folgenden aufgeführt.

- Sauerstoffmessung:
  - Sauerstoffsensor: PreSens, Regensburg, Deutschland
  - Trockenschrank: WTC Binder, Tuttlingen, Deutschland
  
- Pipettierhilfe:
  - Pipetten und Spitzen: Eppendorf, Hamburg, Deutschland
  
- Reaktionsgefäße:
  - 1.5 mL und 2 mL Eppendorfreaktionsgefäße: Eppendorf, Hamburg, Deutschland
  
- Probenvorbereitung für die GC-MS Analyse:
  - Mühle der Firma Retsch MM400: Retsch GmbH Co., Haan, Deutschland
  - VA Stahlkugeln (5 mm): Th. Geyer Berlin GmbH, Berlin, Deutschland
  - Analysenwaage: Mettler Toledo, Greifensee, Schweiz
  - Vortex Genie 2: Bender und Hobein AG Zürich, Schweiz
  - Thermoschüttler: Eppendorf Thermomixer Compact, Hamburg, Deutschland
  - Mikrozentrifuge 5417C: Eppendorf, Hamburg, Deutschland
  - Glasvial („CTVG crimp cap vial“): Chromacol, Trumbull, CT, USA
  - Vakuumkonzentrator mit Rotoren R96-13 and R120-111: Jouan Nordic, Allerød, Dänemark
  - Vakuumpumpe: Ilmvac GmbH, Ilmenau, Deutschland
  - Thermostat: P. Huber GmbH, Deutschland
  - automatische Derivatisierung: CTC Analytics AG, Zwingen, Schweiz
  
- GC-MS Messung:
  - Gaschromatograph: Agilent 6890N, Agilent Technologies, Böblingen, Deutschland
  - Quarzglas kapillarsäule: VAR5, Varian, Darmstadt, Deutschland
  - Quarzglas kapillarsäule: MDN35, MACHEREY-NAGEL GmbH & Co. KG, Düren,



Deutschland

- Flugzeitmassenspektrometer: Pegasus III, LECO Instrumente GmbH, Mönchengladbach, Deutschland
- GC-MS Datenauswertung:
  - ChromaTOF™ Software: Leco, St. Joseph, MI, USA
  - TagFinder Software: MPIMP, Golm, Deutschland
  - Microsoft SQL Server Management Studio Express (2005)
  - Metagenealsewbservice: <http://metagenealyse.mpimp-golm.mpg.de>
  - Multiexperiment Viewer (MeV 4.2)
  - R Software (R 2.8.1)
  - SigmaPlot (11.0)
- Bestimmung saurer Phytohormone und gezielte An- und Abreicherung polarer Substanzen:
  - Gefriertrocknung: Christ, Osterode, Deutschland
  - 4 °C - Zentrifuge: 5417R Centrifuge, Eppendorf, Hamburg, Deutschland
  - 1 mL Bond Elut Plexa™ Extraktionskartusche 30 mg Sorbentbettmasse: Varian, Darmstadt, Deutschland
  - 1 mL Bond Elut PPL™ Extraktionskartusche 50 mg Sorbentbettmasse: Varian, Darmstadt, Deutschland
  - 6 mL FOCUS Extraktionskartusche 50 mg Sorbentbettmasse: Varian, Darmstadt, Deutschland
  - 3 mL Strata™ C18-T Extraktionskartusche 500 mg Sorbentbettmasse: Phenomenex, Aschaffenburg, Deutschland
  - 1 mL Supel™ Select HLB Extraktionskartusche 30 mg Sorbentbettmasse: Supelco, München, Deutschland
  - 1 mL Strata™ SAX Extraktionskartusche 50 mg Sorbentbettmasse: Phenomenex, Aschaffenburg, Deutschland
  - 1 mL SPEC MP1 Extraktionskartusche 50 mg Sorbentbettmasse: Phenomenex, Aschaffenburg, Deutschland
  - VacMaster-10 (Vakuumkammer): Biotage AB; Uppsala, Schweden
  - 15 mL und 50 mL verschraubbare Plastikröhrchen („Falcon™ Conical Centrifuge Tubes“): BD Biosciences, San Jose, CA, USA
  - BEH C18 Umkehrphasensäule: Waters, Manchester, UK
- Gepaarte Korrelationsspektroskopie:
  - Ultraschallbad: Branson 2510, St. Louis, USA
  - Zentrifuge für Glasbehälter: MSE Mistral 2000, Buckinghamshire, England
  - präparative HPLC: Shimadzu, Shimadzu Europa GmbH, Duisburg
  - Gefriertrocknung: Edwards-Modulyo EF4, Crawly, Vereinigtes Königreich

- Semipräparative Aminopropylsäule (waters spherisorb S5 NH2 10 x 250 mm): Alltech GmbH Chromatographiezubehör
  - Analytische Aminopropylsäule (waters spherisorb 5um NH2 4.6 x 250 mm): Alltech GmbH Chromatographiezubehör
  - „prevail carbohydrate“ Säule ES 5u 250x4.6 mm: Alltech GmbH Chromatographiezubehör
  - 600 MHz AMX NMR Spektrometer: Bruker, Bremen, Deutschland
- Entscheidungsbäume
    - Chem Draw Software (ChemDraw Ultra 11.0.503)
    - Microsoft SQL 2008 Server Analysis Services™

## B.2 Methoden

### B.2.1 Phänotypische Untersuchungen

Die phänotypischen Untersuchungen wurden in Kooperation mit Kartoffelzüchtern durchgeführt. Dabei wurde der Lagerungsverlust einzeln gelagerter Knollen von Mitarbeitern des Arbeitskreis Peters und die Chipseignung sowie die Schwarzfleckigkeitsneigung von Mitarbeitern des Arbeitskreis Tacke bestimmt. Diese Daten wurden im Anschluss mit den metabolischen Daten in Zusammenhang gebracht, um ein entsprechendes Modell zum jeweiligen physiologischen Prozess aufzustellen. Für die Versuche der „Chipseignung“ und der „Schwarzfleckigkeitsneigung“ wurden prädiktive Modelle aufgestellt und für die Versuche des „Physiologischen Alters“ und der „Keimhemmung“ indikative Metaboliten extrahiert.

#### B.2.1.1 Lagerungsverlust

Kartoffelknollen mit einer Gesamtmasse von ca. 7 kg wurden in luftdurchlässige Netzsäcke gefüllt, gewogen und zusammen mit Füllkartoffeln bei 5 °C in Großkisten gelagert. Die Wärme- und Feuchtigkeitsabfuhr erfolgte über Außenluft beziehungsweise über maschinell gekühlte Raumluft. Während der Lagerung wurden für den Versuch „Physiologisches Alter“ kurz vor der Einlagerung beziehungsweise Auslagerung (November und März in den Lagerperioden 2006/2007 und 2007/2008) und für den Versuch „Keimhemmung“ zu vier verschiedenen Zeitpunkten (Oktober/November, Dezember, Februar und März/April in den Lagerperioden 2006/2007 und 2007/2008; Dezember und März in der Lagerperiode 2008/2009) sechs Knollen für die Metabolitenanalyse entnommen. Die im Netzsack verbliebenen Kartoffeln wurden für die Bestimmung des Lagerungsverlustes verwendet. Dabei wurden die Probenbeutel zurückgewogen, die Knollen entkeimt und die Keime gewogen. Die nassfaulen und trockenfaulen Knollen wurden ausgelesen und gewogen. Die jeweiligen Gewichte wurden in Relation zum Ausgangsgewicht gesetzt und entsprechen den Größen "Keimung in %", "Nassfaule in %" und "Trockenfaule in %". Die Differenz zum Gesamtverlust entspricht dem

"Gewichtsverlust in %". Die Summe aller vier Verlustgruppen ergibt den Lagerverlust. Der Lagerungsverlust ist wichtig, um die Qualität gelagerter Knollen einzuschätzen. Schrumpelige, faulende und keimende Knollen werden von dem Verbraucher abgelehnt und sollten von der Industrie vermieden werden.

#### B.2.1.2 Chipseignung

Zehn Knollen eines Kultivars wurden in ca. 2 mm dicke Scheiben gerieben und solange in heißem Öl (160-180 °C) frittiert bis dies keine Blasen mehr bildete. Anschließend wurde die resultierende Bräunung der Scheiben festgestellt und entsprechend der Wageningen Farbskala klassifiziert<sup>[78]</sup>. Besonders dunklen Stücken wurde dabei ein Wert von 1 zugeordnet, hellen und gelblichen Stücken ein Wert von 9. Diese Werte ermöglichen eine Aussage zur Chipseignung einzelner Kartoffelkultivare und sind notwendig, um ein Modell zur Chipseignung aufzustellen.

#### B.2.1.3 Schwarzfleckigkeitsneigung

Die Prüfung auf Schwarzfleckigkeitsneigung erfolgte gemäß der BSA (Bundessortenamt)-Richtlinie: „Prüfung auf Schwarzfleckigkeitsneigung“ von Mitarbeitern der Arbeitsgruppe Tacke<sup>[78]</sup>. Dafür wurden 20 unbeschädigte Knollen einer Sorte in einer Trommel für 50 s und 380 rpm bei 4-6 °C belastet und nach der mechanischen Belastung 4-5 Tage bei Raumtemperatur unter Lichtausschluss gelagert. Danach wurden die Knollen entlang ihrer größten Schnittfläche geschnitten und der Anteil der verfärbten Fläche an den Knollenhälften ermittelt. Dieser wurde entsprechend den folgenden Kategorien klassifiziert:

- 0: keine Verfärbung
- L: bis zu 25 % der Oberfläche ist verfärbt (Verfärbungstiefe < 5 mm)
- M: bis zu 50 % der Oberfläche (Verfärbungstiefe < 5 mm) und/oder 25 % der Oberfläche (Verfärbungstiefe > 5 mm) sind verfärbt
- S: mehr als 50 % der Oberfläche (Verfärbungstiefe < 5 mm) und/oder 25 % der Oberfläche (Verfärbungstiefe > 5 mm) sind verfärbt

Der Schwarzfleckigkeitsindex SI ergibt sich aus diesen drei Kategorien (Gl.1)<sup>[79]</sup>.

$$SI = \frac{(n_L * 1) + (n_M * 3) + (n_S * 5)}{n_{total}} \quad (1)$$

- n Anzahl der Knollen in der entsprechenden Kategorie (L, M, S)
- total Anzahl der untersuchten Knollen je Kultivar

Dieser kann problemlos in die Boniturwerte „Anfälligkeit gegenüber Schwarzfleckigkeit“ umgerechnet werden. Diese können Werte von 1 bis 9 einnehmen. Ein Wert von 1 deutet auf eine besonders hohe Resistenz gegenüber Schwarzfleckigkeit hin, während ein Wert von 9 auf eine sehr hohe Anfälligkeit signalisiert. Analog zum Versuch der „Chipseignung“ wurden im Rahmen dieses Versuches Stoffwechselprodukte extrahiert, welche für die Schwarzfleckigkeitsneigung relevant sind und eine Vorhersage zur Schwarzfleckigkeitsneigung ermöglichen.

### B.2.2 Sauerstoffmessung

Kartoffelknollen der Sorte *Afra* wurden 30 min nach mechanischer Belastung in ein hermetisch verschließbares Glasgefäß gegeben und für 2 Tage in einen Trockenschrank bei 33 °C gestellt. Dabei wurde das Gefäß mehrmals belüftet, um die Knollen vor anoxischem Stress zu bewahren. Parallel dazu wurden unbehandelte Knollen der Sorte *Afra* bei diesen Bedingungen gelagert. In beiden Gefäßen wurde die Sauerstoffkonzentration mit einem Sauerstoffsensoren erfasst und der resultierende Sauerstoffverbrauch festgestellt, da eine erhöhte Respirationsrate von Kartoffelknollen nach mechanischer Belastung erwartet wurde.

### B.2.3 GC-TOF-MS basierte Metabolitenanalyse

Die GC-TOF-MS basierte Metabolitenanalyse ermöglicht eine quantitative Erfassung niedermolekularer Primär- und Sekundärmetaboliten und ist in Bezug auf die Analyse physiologischer Prozesse von Kartoffelknollen (Kap. B.1.2.1-Kap. B.1.2.5) die Methode der Wahl. Sie kann in folgende Arbeitsschritte untergliedert werden:

1. Probenentnahme
2. Probenvorbereitung
3. Messung
4. Datenprozessierung
5. Datenauswertung
6. Datenarchivierung.

#### B.2.3.1 Probenentnahme

Von den Kartoffelknollen wurde das proximale und distale Ende entfernt und das Mark der Knolle mitsamt Schale für die folgende Analyse verwendet. Dazu wurde das Mittlere der Knolle mittels einer Küchenreibe in dünne Kartoffelscheiben gerieben, die Scheiben zügig in flüssigem Stickstoff schockgefroren, mit einem Pistill grob vorzerkleinert und in einer Mühle der Firma Retsch unter Verwendung von 4-5 Stahlkugeln für 2 min bei 20 Hz zu einem feinen Pulver verarbeitet. Von dem Pulver wurden (50±2) mg in Eppendorfreaktionsgefäße gegeben und bis zur weiteren Behandlung bei -80 °C zwischengelagert.

### B.2.3.2 Probenvorbereitung

Für die Probenvorbereitung müssen die Inhaltsstoffe zunächst mit einem geeigneten Lösungsmittel aus der Matrix der „Kartoffelknolle“ herausgelöst und anschließend in eine Form überführt werden, so dass sie der gaschromatographischen Analyse leicht zugänglich sind. Die Aufarbeitung und Messung der Proben erfolgte entsprechend der von Erban et al. entwickelten Prozedur<sup>[36]</sup>.

#### B.2.3.2.1 Extraktion polarer Substanzen

Für die Flüssig-Flüssig Extraktion wurde ein Mix aus kaltem Methanol (-20 °C), internen Standards (siehe Anhang Tab. 1) und Methylnonadecanoat (2 mg/mL in Chloroform) in einem Verhältnis von 10:1:1 (v/v/v) hergestellt. Davon wurden jeweils 360 µL zu den entsprechenden Kartoffelhomogenaten gegeben. Die erhaltenen Suspensionen wurden gründlich durchmischt und für 15 min bei 70 °C auf dem Thermoschüttler inkubiert. Anschließend wurden die Proben auf Raumtemperatur abgekühlt und mit 200 µL Chloroform versetzt. Nach erneuter Inkubation bei 37 °C wurden 400 µL Wasser hinzupipettiert. Anschließend wurde das Gemisch für 5 min bei 14000 rpm zentrifugiert, 10 µL der oberen Phase in ein Glasvial überführt und das Lösungsmittel in der Vakuumzentrifuge für 3 h bei 4 °C entfernt.

#### B.2.3.2.2 Derivatisierung polarer Substanzen

Das eingetrocknete Extrakt wurde in 10 µL Methoxyaminhydrochlorid (20 mg/mL in Pyridin) gelöst und für 90 min bei 40 °C geschüttelt. Anschließend wurden 20 µL Mix (7:1, v/v) aus *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamid (MSTFA) und *n*-Alkanen (Decan, Dodecan, Pentadecan, Octadecan, Nonadecan, Docosan, Octacosan, Dotriacontan, Hexatriacontan; 0.22 mg/mL in Pyridin) hinzupipettiert und die Lösung für weitere 45 min bei 40 °C gehalten. Die Derivatisierung erfolgte unter Verwendung eines CTC Combi PAL Autosamplers.

#### B.2.3.3 GC-TOF-MS Messung

Es wurden 1 µL der derivatisierten Probe splitlos bei 230 °C in den Gaschromatographen injiziert und temperaturprogrammiert aufgetrennt. Dabei betrug der Heliumfluss 0.6 mL/min, die Spülzeit 1 min und der Gasstrom 20 mL/min. Die Trennung erfolgte mit Hilfe einer 5PDM VF-5ms Quarzglas kapillarsäule mit 0.25 µm Filmdicke, 30 m Länge, 0.25 mm Innendurchmesser und 10 m EZ-Vorsäule. Für die Trennung wurde der Gaschromatograph 1min isotherm bei 70 °C gehalten, danach mittels einer Temperaturrampe innerhalb von 1866 s auf 350 °C erwärmt und anschließend für 5 min bei dieser Temperatur gehalten. Es wurde im Rahmen dieser Arbeit eine sehr langsame Temperaturrampe (9 °C/min) gewählt, damit möglichst alle endogenen Metaboliten von den Kartoffelknollen quantitativ aufgetrennt werden. Daraufhin wurde der Gaschromatograph auf seine Ausgangstemperatur temperiert und die

Eluate mit Ausnahme der Eluate der Abkühlphase mit Hilfe eines nachgeschalteten Massenspektrometers analysiert. Dazu wurden die Eluate durch eine Transferline zur Ionenquelle geleitet, welche eine Temperatur von 250 °C hatte. Die erzeugten Ionen wurden bei geeigneter Detektorspannung mit einer Scangeschwindigkeit von 20 scans/s im Bereich von 70-600 amu aufgenommen. Die Wahl der hohen Scangeschwindigkeit resultierte aus der Verwendung eines Flugzeitmassenspektrometers als Analysator.

#### B.2.3.4 Datenprozessierung

Die erhaltenen Chromatogramme wurden mit Hilfe der ChromaTOF™ Software basislinienkorrigiert (Basislinie: gerade über dem Rauschen; Glättung: 20; Peakbreite: 6, S/N: 2.0) und als cdf Dateien exportiert. Diese wurden anschließend mit Hilfe der TagFinder Software in Textdateien konvertiert und in die TagFinder Software importiert. Für den Import wurde ein Grenzwert von 50 Intensitätseinheiten festgesetzt, da für das Signal-zu-Rausch Verhältnis ein Faktor von zwei gewählt wurde.

Anschließend wurde der Retentionsindex nach dem Modell von van den Dool mit Hilfe der TagFinder Software berechnet, um Retentionszeitunterschiede zwischen den Proben zu kompensieren<sup>[80]</sup>. Dazu wurde die Retentionszeit der Alkane auf der Fragmentmasse 85 aus den jeweiligen Chromatogrammen extrahiert. Diesen wurde ein Retentionsindex von 100-Kohlenstoffanzahl zugeordnet und danach der Retentionsindex der übrigen Signale berechnet<sup>[80]</sup>. Nach der Retentionsindexberechnung wurden die entsprechenden Notizen zur Probenentnahme, Probenvorbereitung und Messung an die Chromatogramme gefügt.

Schließlich wurden alle importierten Chromatogramme anhand ihrer gemeinsamen Fragmentmasse und ihres Retentionsindex in einer gemeinsamen Matrix angeordnet, in der vertikal alle Intensitäten eines Chromatogrammes und horizontal alle Intensitäten einer Masse zu einem bestimmten Retentionsindex über das gesamte Experiment aufgelistet waren. Diese Matrix wurde anschließend als „txt“- beziehungsweise „xml“ exportiert. Anschließend wurde nach Normierung (Kap. B.2.4) eine multivariate Statistik angewandt. Die multivariate Statistik deckt Abhängigkeiten zwischen einzelnen Signalen und Proben auf und ist für einen Biochemiker essentiell, um den Rahmen der Auswertung überschaubar zu halten<sup>[81]</sup>. Diese extrahierten Signale müssen im Anschluss daran identifiziert werden.

#### B.2.3.5 Datenauswertung

Für die Identifizierung der gaschromatographischen Signale wurde die generierte „tag“ Datei einem zweistufigen Vergleichsprozess unterworfen. Dazu wurde die aktuellste Version der Referenzbibliothek aus der Datenbank exportiert<sup>[82]</sup> und danach jede „Zeitgruppe“ und jedes „Cluster“ mit mehr als drei Fragmentmassen mit dieser abgeglichen<sup>[53]</sup>. Die Annotationen wurden anschließend in Form einer xml Datei gespeichert und in die Datenbank importiert.

Für den Abgleich wurde eine Retentionsindexungenauigkeit von 0.5-1 % festgelegt. Außerdem mussten mindestens drei Fragmentmassen des Experiments mit denen des Referenzspektrums

übereinstimmen. Auf diese Weise wurde die Wiederfindung kleiner aromatischer Moleküle sichergestellt, welche nur sehr wenige Fragmentmassen aufweisen. Weiterhin wurde während des Abgleichsprozesses auf die Anwesenheit charakteristischer Fragmentmassen (Anhang Tab. 2) sowie auf die Anwesenheit des  $[M]^+$  beziehungsweise  $[M-15]^+$  Signals geachtet. Die charakteristischen Fragmentmassen können bestimmten Stoffklassen zugeordnet werden. Das  $[M]^+$  beziehungsweise  $[M-15]^+$  Signal ist für eine chemische Verbindung charakteristisch. Schließlich wurden sämtliche Identifizierungen anhand eines synthetischen Standardmixes verifiziert. Dieser beinhaltet sämtliche identifizierte Metaboliten im Arabidopsisblatt (Anhang Tab. 3; Persönliche Kommunikation Dipl. Ing. Alexander Erban).

#### B.2.3.6 Datenarchivierung

Zur langfristigen Sicherung der erhobenen Daten wurden alle INNOX Experimente in die Golm Metabolome Database (GMD) importiert<sup>[59]</sup>. Der Import der GC-TOF-MS Daten erfolgte durch das Programm `gmdTagListParser`<sup>[82]</sup>. Die jeweiligen Bonituren zu den Proben wurden mit Hilfe eines SQL Skripts in der GMD hinterlegt. Die auf Frischgewicht und internen Standard normierten Messdaten wurden mit Hilfe der VANTED Software visualisiert<sup>[83]</sup>.

#### B.2.4 Normierung der Daten

Die GC-MS Daten wurden gemäß van den Berg et al. normiert<sup>[84]</sup>. Hierzu wurden die basislinienkorrigierten Daten zunächst auf das Frischgewicht und den internen Standard bezogen. Anschließend wurde der Median signalweise über das gesamte Experiment gebildet und jeder Messwert auf diesen bezogen. Im Anschluss daran wurde der natürliche Logarithmus gebildet. Fehlwerte wurden während dieser Prozedur nicht ersetzt. Diesen wurde nach Abschluss der Normierung ein Wert von 0 zugeordnet.

##### B.2.4.1 Statistische Auswertung

Die Hauptkomponentenanalyse (PCA) und die Unabhängige Komponentenanalyse (ICA) sind zwei gängige multivariate Analysenmethoden und werden zur Klassifizierung der gemessenen Proben angewendet. Sie wurden mit Hilfe von „Metagenealyse“ (<http://metagenealyse.mpimp-golm.mpg.de/>) durchgeführt. Die Einfaktorielle und Zweifaktorielle Varianzanalyse wurden mit Hilfe des Multiexperiment Viewers durchgeführt, während die Korrelationsanalysen mit der R Software und die t-Tests mit der SigmaPlot Software vorgenommen wurden.

#### B.2.5 Bestimmung saurer Phytohormone

Um saure Phytohormone in Nutzpflanzen zu bestimmen wurde der Schritt der Probenvorbereitung modifiziert (Kap. B.2.3.2). Dazu wurde von 6-8 *Arabidopsis thaliana* Col-0 Pflanzen die Rosette in flüssigem Stickstoff schockgefroren, zu einem Pool vereinigt und mit

einem Mörser und Pistill zu einem feinen Pulver homogenisiert. Von dem Homogenat wurde 1 g in ein Falconröhrchen überführt und für fünf Tage entsprechend den Standardeinstellungen gefriergetrocknet. Das erhaltene Material wurde in 12 mL eines Ameisensäure-Wassergemisches (2 % : 98 %, v/v) aufgeschlämmt und für 30 min bei 4 °C inkubiert. Anschließend wurde die Suspension für 10 min bei 4 °C zentrifugiert und die obere Phase in ein neues Falconreaktionsgefäß überführt. Diese Extraktion wurde zweimal mit dem Rückstand wiederholt. Schließlich wurden alle drei Extrakte vereinigt und 30 mL an einer 1 mL Bond Elut Plexa™ Extraktionskartusche angereichert, die zuvor mit 1 mL Methanol und 1 mL eines Ameisensäure-Wassergemisches (2 % : 98 %, v/v) gewaschen und konditioniert worden war. Dabei wurde darauf geachtet, dass die Kartusche nicht trocken läuft. Nach Anreicherung der Analysenlösung wurde die SPE Kartusche mit 1 mL einer Lösung aus 2 %-iger Ameisensäure gewaschen und für 5 min starker Unterdruck angelegt. Die anschließende Elution in ein Spitzeppendorfreaktionsgefäß erfolgte in zwei Schritten durch Zugabe von jeweils 500 µL Methanol. Der Fluss wurde so eingestellt, dass maximal zwei Tropfen innerhalb einer Minute den Kartuschenausgang passierten. Das erhaltene Eluat wurde in der Vakuumzentrifuge für 3 h bei 5 °C eingedampft.

Für die UPLC-FTICR-MS Analysen wurde der Rückstand in 100 µL einer 80 %-igen methanolisch-wässrigen Lösung (v/v) gelöst und 2 µL davon auf eine BEH C18 Umkehrphasensäule mit 1.8 µm großen Partikeln gegeben. Die aufgegebene Probemenge wurde danach durch einen linearen Lösungsmittelgradienten (**Solvent A**: 0.1 %-ige Ameisensäure in Wasser und **Solvent B**: 0.1 %-ige Ameisensäure in Methanol) unter Verwendung eines Flusses von 400 µL/min aufgetrennt. Dabei wurden **Solvent A** und **Solvent B** in einem Verhältnis von 70:30 (v/v) für 2 min durch die Säule gepumpt und im Anschluss innerhalb von 15 min die Laufmittelzusammensetzung dahingehend verändert, dass 0 % **Solvent A** und 100 % **Solvent B** durch die Säule gepumpt wurden. Das Eluat wurde durch Elektrosprayionisierung im negativen Modus im Bereich von 120-450 amu verfolgt. Die UPLC-FT-ICR-MS Messungen wurden von Herrn Dr. P. Giavalisco durchgeführt. Die erhaltenen Signal zu Rausch Verhältnisse wurden anschließend mit denen aus der GC-TOF-MS verglichen.

Für die GC-TOF-MS Analysen wurde das eingetrocknete Extrakt entsprechend Kap. B.2.3.3 behandelt. Anstelle von MSTFA wurde entsprechend Birkemeyer et al.<sup>[30]</sup> BSTFA (N,O-Bis(trimethylsilyl)trifluoroacetamid) verwendet. Weiterhin wurde für die Trennung aufgrund von Sensitivitätsproblemen ein steilerer Temperaturgradient (25 °C/min) gewählt.

## B.2.6 Klassifizierung unbekannter GC-MS Signale

### B.2.6.1 Gepaarte Korrelationsspektroskopie

Für die gepaarte Korrelationsspektroskopie wurden methanolische Extrakte der Kreuzblütlerfamilie mittels präparativer HPLC vorfraktioniert, die Fraktionen mittels NMR Spektroskopie und anschließend mittels GC-TOF-MS analysiert. Die erhaltenen Datensätze wurden schließlich über eine Pearsonkorrelation miteinander in Verbindung gebracht. Auf diese



Weise wurden zusätzliche spektroskopische Informationen für die gaschromatographischen Signale erhalten.

#### B.2.6.1.1 Extraktion polarer Substanzen aus Nutzpflanzen

Für die Extraktion wurden 0.5 g Pflanzenmaterial in ein vorgekühltes Glasröhrchen überführt, mit 1.5 mL Methanol versetzt und die erhaltene Suspension für 15 min in einem Ultraschallbad behandelt. Die erhaltene Lösung wurde im Anschluss für 5 min bei 2000 rpm zentrifugiert und die obere Phase abgenommen. Es wurde an dieser Stelle auf eine Phasentrennung verzichtet, um den Verlust potentieller Analyten möglichst gering zu halten.

#### B.2.6.1.2 Präparative Flüssigkeitschromatographie

Die Extrakte wurden mit Hilfe einer präparativen HPLC fraktioniert. Dafür wurde eine Aminopropylsäule verwendet und ein binäres Lösungsmittelgemisch (**Solvent A**: 0.1 %-ige Ameisensäure in Wasser, **Solvent B**: 0.1 %-ige Ameisensäure in Acetonitril) gewählt. Die Zusammensetzung wurde variiert, indem 70 % Acetonitril für 10 min durch die Säule geleitet wurden. Anschließend wurde innerhalb von 10 min das System auf 10 % Acetonitril äquilibriert, diese Konzentration für 40 min aufrecht erhalten und das System innerhalb von einer Minute auf die Ausgangskonzentration von 70 % Acetonitril gebracht. Dieser Zustand wurde für 9 min beibehalten bis die nächste Probe zur Fraktionierung eingespritzt wurde. Während der Fraktionierung wurden von 2 mL Extrakt jeweils 11 Fraktionen erhalten. Die ersten sieben Fraktionen umfassen den acetonitrilreichen Elutionsbereich, die weiteren drei Fraktionen den polaren Elutionsbereich und die letzte Fraktion den Konditionierungsbereich (Anhang Tab. 4).

#### B.2.6.1.3 NMR Messung

Die Extrakte wurden nach dem Fraktionierungsschritt entsprechend gefriergetrocknet und für die NMR Messungen in [D<sub>4</sub>]Methanol aufgenommen, um polare Substanzen zu analysieren. Die <sup>1</sup>H NMR Untersuchungen wurden an einem 600 MHz AMX NMR Spektrometer von Herrn Prof. Jacques Vervoort (Universität Wageningen) durchgeführt und waren notwendig, um die gepaarte NMR- und GC-TOF-MS Analyse durchzuführen.

#### B.2.6.1.4 GC-MS Messung

100 µL Extrakt aus der NMR Messung wurden in einer Vakuumzentrifuge bis zur Trockene eingedampft und wie im Kapitel B.2.3.2.2 und B.2.3.3 beschrieben analysiert. Die Derivatisierung der Proben erfolgte mit Hilfe von BSTFA und Methoxyaminhydrochlorid, da nicht mit Hilfe des Combi PAL Systems derivatisiert wurde.

### B.2.6.2 Selektive An- und Abreicherung polarer Substanzen

Für den Versuch der selektiven An- und Abreicherung wurden die Extrakte von Kartoffelknollen und Brokkoliröschen entsprechend Kap. B.2.3.2.1 hergestellt, in der Vakuumzentrifuge bis zur Trockene eingedampft und anschließend in einem geeigneten Lösungsmittel aufgenommen. Dies wurde durch das jeweilige Sorbens bestimmt (Kap. B.2.6.2.1-B.2.6.2.3). Die SPE Kartuschen wurden schließlich mit jeweils 2.9 mL Extrakt beladen.

#### B.2.6.2.1 Hydroxyliertes Polymer „Plexa“

Die eingeeengten Aliquots wurden in 0.67 %-iger Ameisensäure gelöst und die Festphasenextraktionskartusche jeweils mit 1 mL Methanol beziehungsweise 1 mL Wasser gewaschen und konditioniert. Nach Beladung wurde mit 1 mL 5 %-igem Methanol gewaschen und mit 1 mL reinem Methanol eluiert.

#### B.2.6.2.2 Anionenaustauscher „SAX“

Die eingeeengten Aliquots wurden in 0.5 %-iger methanolischer Ammoniaklösung gelöst. Die Säulen wurden entsprechend Kap. B.2.6.2.1 gewaschen und konditioniert. Nach Anreicherung wurde die Säule mit 1 ml Lösungsmittel (0.5 %-ige methanolische Ammoniaklösung) gewaschen und für die Elution 1 mL 5 %-ige methanolische Ameisensäure verwendet.

#### B.2.6.2.3 Kationenaustauscher „MP1“

Die eingeeengten Aliquots wurden in 2.9 mL einer 2 %-igen Essigsäure gelöst. Die Festphasenextraktionskartusche wurde mit 400  $\mu$ L reinem Methanol gewaschen und 400  $\mu$ L 0.1 M Essigsäure konditioniert. Anschließend wurden die Kartuschen nach Beladung mit 400  $\mu$ L 0.1 M Essigsäure sowie 200  $\mu$ L 50 %-igem Methanol gewaschen. Schließlich erfolgte die Elution zweimal mit je 500  $\mu$ L einer 10 %-igen methanolischen Ammoniaklösung.

### B.2.6.3 Entscheidungsbäume

Für die Etablierung des Entscheidungsbaumverfahrens wurden 1095 Metabolit- und 1222 Analytstrukturen anhand ihrer CAS Nummer und ihres Synonyms im SciFinder recherchiert, mit Hilfe des „CS ChemDraw Ultra Programms“ als „mdl mol“ gespeichert und in die GMD importiert. Anschließend wurde diese Information an die entsprechenden Einträge („Compound“) geknüpft und die zugehörigen Verbindungen zwischen Analyt und Metabolit beziehungsweise Spektrum und Analyt mit Hilfe des Programms GMD Editor gelegt<sup>[82]</sup>. Für das Erlernen der Entscheidungsbäume wurden nur Strukturelemente in Betracht gezogen, welche in mehr als 3 % der Fälle in den Analyt- beziehungsweise Metabolitstrukturen vorhanden waren. Die Entscheidungsbäume umfassen die folgenden Strukturelemente: 1,2 Aminoalkohol (R = H,

Alkyl, Aryl), 1,2 Diol (R = H, Alkyl, Aryl), Acetal ((R<sub>1</sub>)<sub>2</sub>(OR<sub>2</sub>)<sub>2</sub>: R<sub>1</sub> = H, Alkyl, Aryl; R<sub>2</sub> = Alkyl, Aryl), alpha Aminosäure (R = H, Alkyl, Aryl), Aldehyd (R = H, Alkyl, Aryl), Alken (R = H, Alkyl, Aryl), Alkohol (R = Alkyl), Amin (R<sub>1</sub>N(R<sub>2</sub>)<sub>2</sub>: R<sub>1</sub> = Alkyl, Aryl; R<sub>2</sub> = H, Alkyl, Aryl), Aromat (Konjugiertes System muss die Hückel Regel erfüllen.), Carbonsäure (R = H, Alkyl, Aryl), Carbonsäureamid (R = H, Alkyl, Aryl), Carbonsäureester (R<sub>1</sub>COOR<sub>2</sub>: R<sub>1</sub> = H, Alkyl, Aryl; R<sub>2</sub> = Alkyl, Aryl), sekundäres Carbonsäureamid (R<sub>1</sub>CONHR<sub>2</sub>: R<sub>1</sub> = H, Alkyl, Aryl; R<sub>2</sub> = Alkyl, Aryl), Carbonyl, Ether (R = Alkyl, Aryl), Hemiacetal ((R<sub>1</sub>)<sub>2</sub>(OR<sub>2</sub>)<sub>2</sub>: R<sub>1</sub> = H, Alkyl, Aryl; R<sub>2</sub> = Alkyl, Aryl), Heterocyclus (aromatisches System mit mindestens einem Nichtkohlenstoffatom), Hydroxy (R = Alkyl, Aryl), Keton (R = Alkyl, Aryl), Phenol, Phosphat, primäres aliphatisches Amin (R = Alkyl, Aryl), primärer Alkohol (R = Alkyl, Aryl), primäres Amin (R = Alkyl, Aryl) und sekundärer Alkohol (R = Alkyl, Aryl). Diese wurden nacheinander mittels des Microsoft SQL 2008 Server Analysis Services™ erstellt. Dafür wurden die Massenspektren in zwei Gruppen unterteilt: zum einen in eine Gruppe in der das jeweilige Massenspektrum über die betrachtete funktionelle Gruppe verfügte und in eine zweite Gruppe, in der das Massenspektrum diese funktionelle Gruppe nicht innehatte. Im Anschluss wurden zwei Trainingsdatensätze erstellt und Kriterien für die An- und Abwesenheit von funktionellen Gruppen und Strukturelementen aufgestellt<sup>[82]</sup>. Dazu wurden die Fragmentmassen 73, 74, 75, 147, 148 und 149 ausgeschlossen, da sie sich aus der Trimethylsilylfunktionalität ergeben. Weiterhin wurden die <sup>13</sup>C Isotopomere während der Trainingsphase nicht betrachtet, da sie den Datensatz extensiv vergrößern. Alle präprozessierten massenspektrenalen Informationen wurden einem Fischertest unterzogen und nur diejenigen Massenfragmente für das folgende Entscheidungsbaumverfahren in die nähere Auswahl gezogen, die mit in die 1000 bestplatzierten Fischerstatistik fielen. Für das Trainieren der Entscheidungsbäume wurde im Allgemeinen ein Abbruchkriterium von 10 Massenspektren gewählt.

## C Ergebnisse und Diskussion

### C.1 Analytische Lösungsansätze

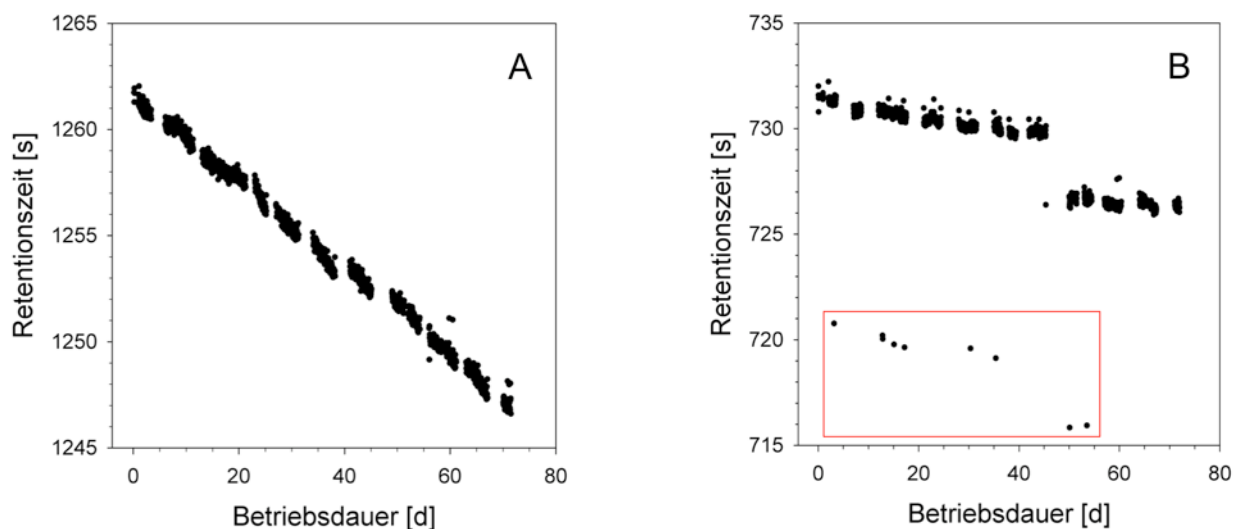
Höhere Pflanzen wurden bereits mehrfach mit Hilfe von GC-TOF-MS untersucht. Die Analyse von Nutzpflanzen mit dem Ziel robuste metabolische Substanzen für einen biologischen Prozess zu gewinnen, ist bislang noch nicht systematisch durchgeführt worden. Damit die GC-TOF-MS basierte Analyse von Nutzpflanzen und anderen biologischen Matrices innerhalb eines zeitlich überschaubaren Rahmens durchführbar ist, wurde untersucht welche Messparameter und Informationen notwendig sind, um agrarwirtschaftlich bedeutsame Nutzpflanzen mittels GC-TOF-MS zu analysieren. Dazu wurden analytische Lösungsansätze geschaffen, die eine schnelle Aufklärung biologischer Prozesse in Nutzpflanzen ermöglichen. Diese Lösungsansätze sind zum Teil in eigenen Arbeiten<sup>[85]</sup> und Publikationen als Koautor<sup>[46, 86]</sup> eingegangen. Beispiele dafür sind die Ermittlung von Grenzwerten für den Substanzabgleich, eine Korrektur und Erweiterung von Einträgen in die bereits vorhandene Spektrenbibliothek, die Archivierung von neuen experimentellen Daten in der hauseigenen metabolischen Datenbank und eine Verbesserung der Probenvorbereitung mit dem Ziel diese an die Lösung des Problems anzupassen. Darüber hinaus wurde untersucht wie robust sich die einzelnen Parameter im Verlauf einer gewöhnlichen Messserie verhalten. Dies war erforderlich, da für die Analyse biologischer Proben im Allgemeinen eine Vielzahl von Proben notwendig ist, um ein möglichst vollständiges Bild aller ablaufenden metabolischen Prozesse zu erhalten. Die untersuchten Replika sollten daher so wenig variant wie möglich sein, um Unterschiede zwischen den vorgegebenen Probengruppen nicht zu überdecken. Im Folgenden wurde die Robustheit der Retentionszeit näher untersucht.

#### C.1.1 Retentionszeit und Retentionsindex

Die Robustheit der Retentionszeit wurde an 13 verschiedenen Fettsäuremethylestern überprüft (Anhang Tab. 5). Dazu wurden die entsprechenden Retentionszeiten aus 1340 Chromatogrammen extrahiert, die der Fragmentmasse 87 zugeordnet werden können. Weiterhin wurde die Variabilität der Retentionszeit ermittelt. Für diese Versuchsreihe wurde die Fragmentmasse 87 ausgewählt, da sie ein sehr stabiles Reaktionsprodukt der McLafferty-Umlagerung ist. Die McLafferty-Umlagerung tritt bei aliphatischen C=O, C=N und S=O Verbindungen im Massenspektrometer auf.

Die folgenden Messungen wurden an zwei unabhängigen GC-TOF-MS Systemen (VAR5 und MDN35) durchgeführt. Beide finden in der nichtgerichteten Analyse des Primärmetaboloms Anwendung. Das VAR5 System<sup>[36]</sup> verwendet für die metabolischen Untersuchungen eine 5 % Phenyl-95 % Dimethylpolysiloxan kapillarsäule, einen Temperaturgradienten von 9 °C/min und eine Scangeschwindigkeit von 20 scans/s, während das MDN35 System<sup>[10]</sup> das Stoffgemisch mittels einer 35 % Phenyl-65 % Dimethylpolysiloxan kapillarsäule, einem steileren Temperaturgradienten (15 °C/min) und einer Scangeschwindigkeit von 20 scans/s auftrennt.

Abbildung 1 veranschaulicht den prinzipiellen Verlauf der Retentionszeit aller Fettsäuremethylester am Beispiel von Octadecansäuremethylester für das VAR5 System (Abb. 1A) und das MDN35 System (Abb. 1B). Der ausgewählte Elutionsbereich ist sehr komplex und für die vorliegende Fragestellung von Interesse. In diesem Bereich der Chromatographie eluieren überwiegend Zucker und organische Säuren. Daher wurde er für diesen Vergleich ausgewählt.



**Abb. 1 Verlauf der Retentionszeit innerhalb einer Messserie für die Verbindung Octadecansäuremethylester A) auf dem System VAR5 und B) dem System MDN35.** Die Ausreißer sind einer verfrühten Injektion des Autosamplers zuzuschreiben (rote Umrandung).

Die Retentionszeit nimmt innerhalb von 1340 Chromatogrammen systematisch ab. Für das VAR5 System verkürzte sich die Retentionszeit um 1.2 % (Abb. 1A) und für das MDN35 System um 0.8 % (Abb. 1B). Diese Verringerung der Retentionszeit ist für eine fundierte Biomarkerstudie unzureichend. Sie wird durch Modifikationen der stationären Phase und durch Interaktionen komplexer Analysengemische mit dem Säulenmaterial verursacht. Weiterhin wirken sich Variationen der Säulenlänge, nicht sachgemäße Injektionen durch den Autosampler und Änderungen im Fluss beziehungsweise Druck negativ auf die Stabilität der Retentionszeit aus<sup>[87, 88]</sup>. Da die Verkürzung der Retentionszeit nicht vermeidbar ist, wird allgemein nach Modellen zur Standardisierung der Retentionszeit gesucht. Dies ist auch Ziel dieser Arbeit.

In der Literatur sind bislang eine Vielzahl mathematischer Algorithmen zum Anordnen von Retentionszeiten beschrieben<sup>[89, 90]</sup>. Andere Arbeitsgruppen verfolgen das Konzept des Retentionsindex. Bei dem Konzept des Retentionsindex werden dem Stoffgemisch vor der Messung Zeitstandards zugefügt, die nicht mit dem Stoffgemisch reagieren<sup>[80, 91, 92]</sup>. Anschließend wird die Retentionszeit anhand einer linearen Interpolation entsprechend dem Modell von van den Dool in den Retentionsindex (Gl.2) umgerechnet<sup>[80]</sup>.

$$RI = 100 \cdot (y-x) \cdot \frac{t_{\text{Analyt}} - t_x}{t_y - t_x} + 100x \quad (2)$$

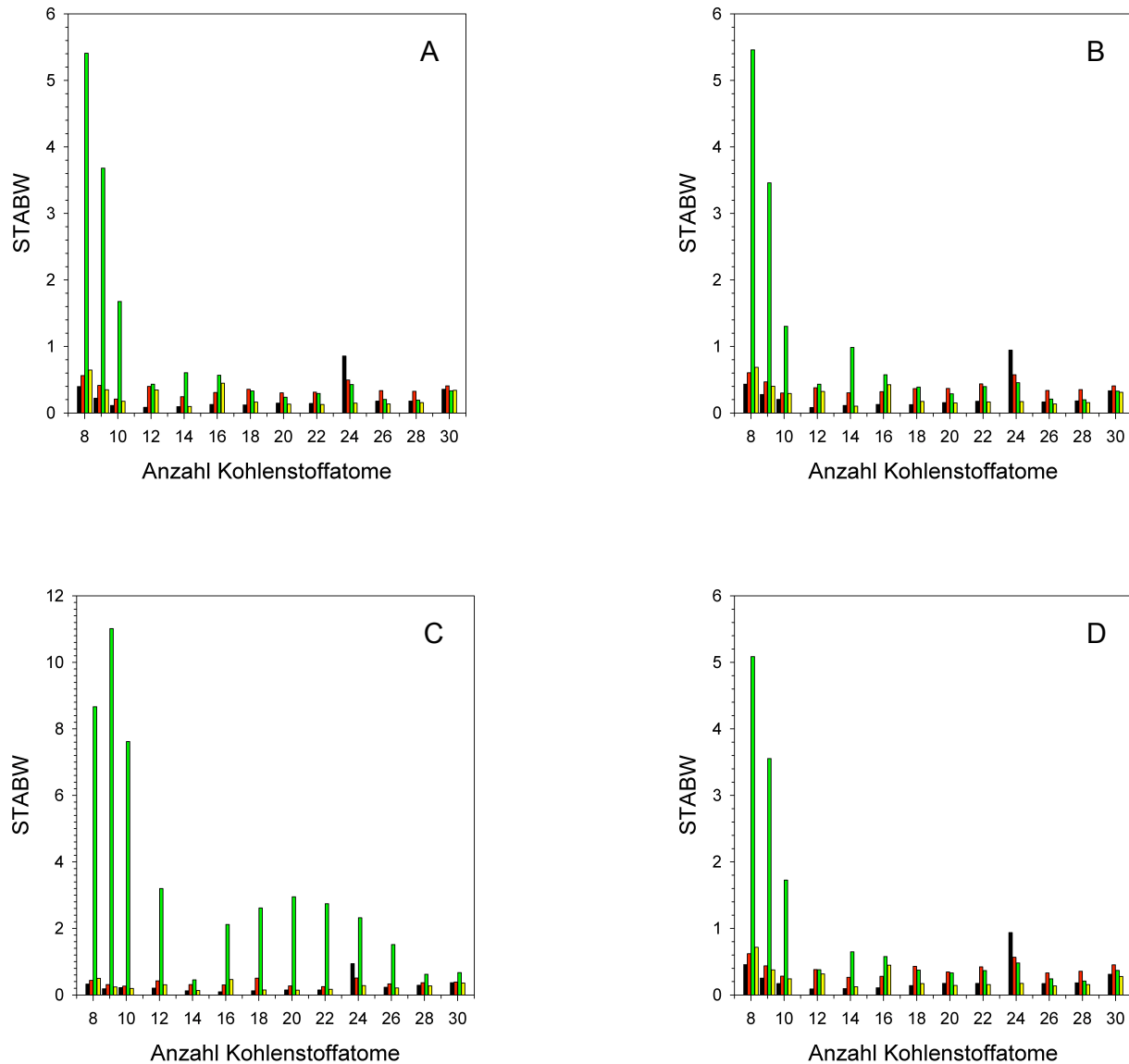
RI	Retentionsindex
x	Anzahl Kohlenstoffatome des Retentionszeitstandards vor dem Analyten
y	Anzahl Kohlenstoffatome des Retentionszeitstandards nach dem Analyten
t	Retentionszeit des Analyten beziehungsweise der umgebenden Alkane

Alternativen zum van den Dool Modell sind polynomiale Regressionen (1.-5. Ordnung) und die Spline Funktion. Sie wurden noch nicht zur Retentionsindexberechnung verwendet und könnten eine höhere Präzision liefern. Dies ist für umfangreiche Biomarkerstudien von enormer Wichtigkeit. Für die folgenden Untersuchungen wurde daher eine annähernd homologe Reihe von Fettsäuremethylestern ausgewählt und der Retentionsindex der Fettsäuremethylester anhand der Alkane berechnet (Gl.2). Im ersten Schritt wurden dafür die entsprechenden Retentionszeiten der Alkane aus den Chromatogrammen extrahiert, die der Fragmentmasse 85 zugeordnet werden können. Anschließend wurde jedem Alkan ein Retentionsindex von  $100 \cdot$  Kohlenstoffanzahl zugeordnet und für jeden Fettsäuremethylester der Retentionsindex nach Maßgabe des entsprechenden Modells berechnet. Hierzu wurde ebenfalls die Fragmentmasse 87 gewählt. Es wurde für diese Versuchsreihe eine annähernd homologe Reihe von Fettsäuremethylestern (Anhang Tab. 5) zu jeder Probe vor der Messung hinzugegeben. Diese können hinsichtlich ihrer Polarität mit den endogenen Metaboliten verglichen werden und ermöglichen so einen Vergleich zwischen verschiedenen Matrizes. Schließlich wurde in der Arbeitsgruppe von Herrn Dr. Kopka beobachtet, dass eine zehnfach konzentriertere Probe zu einer Erhöhung des Retentionsindex führt<sup>[93]</sup>. Die Alkane dienten im Rahmen dieses Versuches als Retentionszeitstandards. Alkane haben den Vorteil, dass sie nicht in polaren Extrakten vorkommen und keine Reaktionen mit den endogenen Substanzen eingehen.

Da das gesuchte Modell im Prinzip auf jedes biologische System angewendet werden soll, wurden die Untersuchungen an zwei synthetischen Matrizes (Zusammensetzung von „Mix 1“ siehe Anhang Tab. 6 und von „Mix 2“ siehe Anhang Tab. 3) und zwei biologischen Matrizes („Reisblatt“ und „Bäckerhefe“) auf dem VAR5 System und dem MDN35 System durchgeführt. Die synthetische Mischung „Mix 1“ beinhaltete 28 biologisch relevante Substanzen und die synthetische Mischung „Mix 2“ 96 Metaboliten. Dies sind physiologisch relevante Substanzen des Primärstoffwechsels. Die Extrakte wurden hinsichtlich steigender Komplexität vermessen, d.h. zuerst „Mix 1“, danach „Mix 2“ und abschließend die „Bäckerhefe“ sowie das „Reisblatt“. Nach den Messungen wurde die Standardabweichung (STABW,  $n = 45$ ) des Retentionsindex aller Fettsäuremethylester bestimmt und zur Evaluation der Robustheit des Retentionsindex herangezogen.

In den untersuchten Modellen zeichnete sich im Mittel kein Unterschied zwischen dem Modell von van den Dool (Abb. 2A) und der Spline Funktion (Abb. 2B) ab. Eine Erhöhung der

Regressionsordnung bewirkte eine höhere Robustheit für das VAR5 System (Abb. 2C und Abb. 2D).



**Abb. 2 Standardabweichung des Retentionsindex in Bezug auf die zugegebenen Fettsäuremethylester im Hinblick auf A) das Modell von van den Dool, B) die Spline Funktion, C) eine Lineare Regression und D) eine Polynomiale Regression fünfter Ordnung in Mix 1 (schwarz), Mix 2 (rot), „Reisblatt“ (grün) und „Bäckerhefe“ (gelb) für das System VAR5.**

So verringerte sich die Standardabweichung des Retentionsindex für die Matrix „Reisblatt“ im Mittel um den Faktor 3.2 für das VAR5 System (bei Erhöhung der polynomialen Ordnung von eins auf fünf). Für das MDN35 System (Anhang Abb. 1) war dieser Effekt weniger stark

ausgeprägt. Hier genügte eine lineare Regression zur Anpassung der Daten. Eine Erhöhung der Regressionsordnung bewirkte hier nur minimale Veränderungen.

Ein Vergleich der vier Matrizes miteinander zeigt, dass die beiden biologischen Matrizes (grüne und gelbe Balken) eine geringere Genauigkeit als die beiden synthetischen Matrizes (schwarze und rote Balken) aufweisen. So wurde bei den synthetischen Matrizes im Mittel eine Ungenauigkeit von 0.2-0.3 RI Einheiten festgestellt, während die biologischen Matrizes eine Ungenauigkeit von 0.7-1.3 RI Einheiten zeigten. Eine Ursache hierfür ist die erhöhte Komplexität der biologischen Extrakte. Diese weisen viel mehr Bestandteile als die synthetischen Matrizes auf. Hinzu kommt, dass diese Substanzen zum Teil koeluiieren

Darüber hinaus fällt auf, dass die Ungenauigkeiten in der Retentionsindexbestimmung über den Chromatographiebereich schwanken. So konnte der Retentionsindex im frühen Chromatographiebereich aufgrund der beginnenden Temperaturrampe (Methyloctanoat bis Methyldecanoat) nur sehr ungenau bestimmt werden, während die Genauigkeit im mittleren (Methylstearat bis Methylbehenat) und im späten Chromatographiebereich (Methyloctacosanoat bis Methylmelissat) höher war. Dieser chromatographische Effekt wurde durch die Komplexität der Matrix offensichtlich noch verstärkt. Die Matrix „Reisblatt“, welche mit mehr als 600 Dekonvolutionen sehr komplex ist, wies im Vergleich zur Matrix „Hefe“ bei allen Modellen im frühen Chromatographiebereich die geringste Präzision auf (grüne und gelbe Balken). Die hohen Standardabweichungen der Retentionsindices deuten auf diesen Effekt hin.

Das Modell von van den Dool (Abb. 2A) und die Spline Funktion (Abb. 2B) hatten die geringsten Ungenauigkeiten bei der Berechnung des Retentionsindex zur Folge. Die polynomiale Regression erster Ordnung (Abb. 2A) wies im Vergleich zu allen anderen Modellen die höchste Ungenauigkeit auf. Im Folgenden wurden sämtliche Retentionszeiten entsprechend dem Modell von van den Dool in Retentionsindices umgerechnet, was in Einklang mit der Literatur steht<sup>[80]</sup>.

Eine weitere Analyse der Standardabweichung des Retentionsindex wichtiger Primärmetaboliten, wie Citrat, Valin, Glycin, Ribitol und Glucose bestätigt, dass die Komplexität der Matrix einen wesentlichen Einfluss auf die Genauigkeit der Retentionsindexbestimmung hat (Tab. 1).

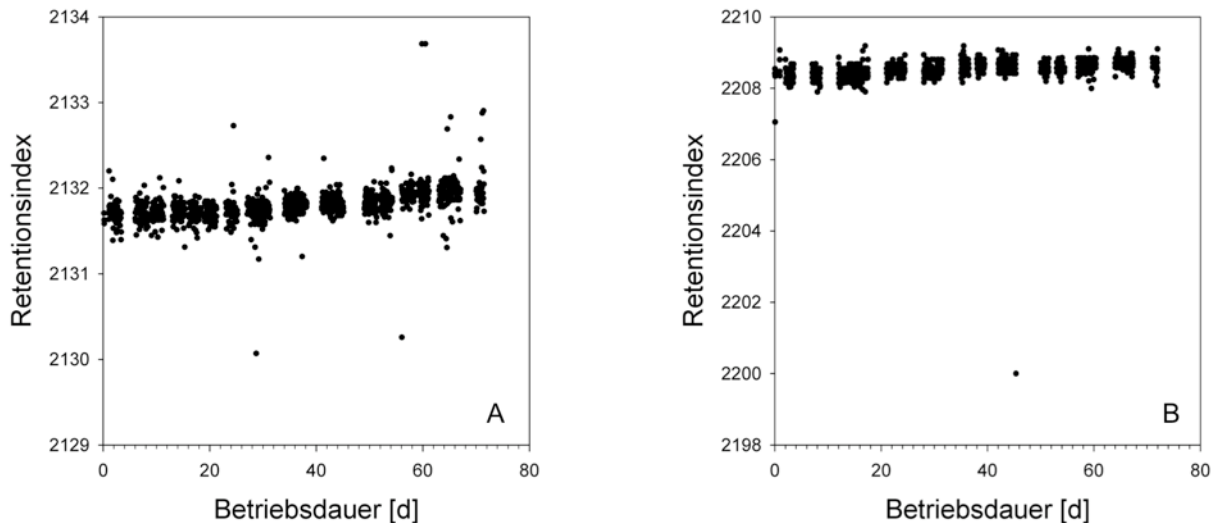
Die Matrix „Reisblatt“ wies im Vergleich zur „Bäckerhefe“ und zum „Mix 1“ mit Ausnahme von Glucose (1MeOX, 5TMS) die höchsten Ungenauigkeiten in der Retentionsindexbestimmung auf (0.8-2.5 RI Einheiten). Glucose zeigte in der Matrix „Bäckerhefe“ die höchsten Ungenauigkeiten. Das kann durch den hohen Glucosegehalt in der Bäckerhefe erklärt werden. Im Vergleich zur Matrix „Bäckerhefe“ wies die Matrix „Reisblatt“ eine 0.2 bis 7-fach höhere Retentionsindexstandardabweichung auf. Dies bestätigt die Erkenntnisse aus der Analyse der Fettsäuremethylester hinsichtlich der Komplexität der Matrizes (Abb. 2). Die Ergebnisse zeigen die Übertragbarkeit auf andere Substanzklassen.



**Tab. 1 Präzision der Bestimmung des Retentionsindex wichtiger Analyten im Primärmetabolismus ausgedrückt als Standardabweichung (n = 45) des Retentionsindex.** TMS: Trimethylsilyliert; MeOX: Methoxyaminiert; Die Zahlen in Klammern geben die Anzahl derivatisierter funktioneller Gruppen an. Die maximale Retentionsindexgenauigkeit ist durch einen roten Schriftzug hervorgehoben.

Analyt	Retentionsindex	Retentionsindex		
		„Mix 1“	„Bäckerhefe“	„Reisblatt“
Citrat (4TMS)	1803	0.37	0.43	<b>1.24</b>
Valin (2TMS)	1207	0.30	0.42	<b>2.51</b>
Glycin (3TMS)	1302	0.20	0.28	<b>1.94</b>
Ribitol (5TMS)	1712	0.54	0.53	<b>0.83</b>
Glucose (1MeOX, 5TMS)	1898	0.63	<b>4.00</b>	0.85

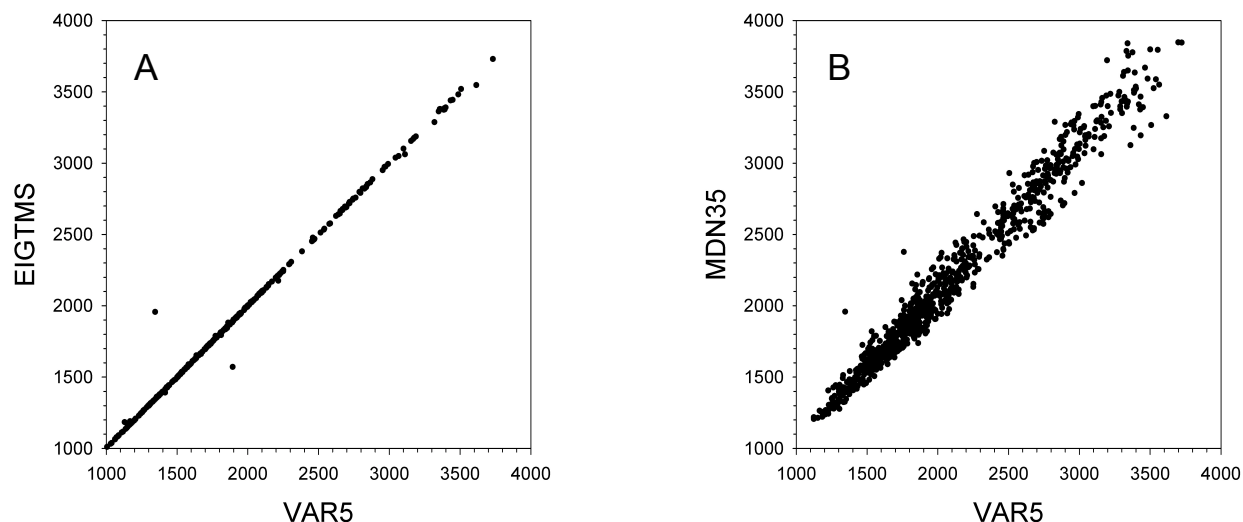
Werden die in Abb. 1A und Abb. 1B dargestellten Retentionszeiten in Retentionsindices umgerechnet, so ist der Retentionsindex innerhalb von 1340 Chromatogrammen sehr stabil (Abb. 3). Die Ausreißer für das VAR5 System sind zum Teil auf Inkonsistenzen in der Injektion des Autosamplers zurückzuführen. Durch Berechnung des Retentionsindex aus der Retentionszeit konnte eine Stabilisierung über die Betriebsdauer erreicht und damit der Einsatz dieser Methode für metabolische Studien in Nutzpflanzen ermöglicht werden.



**Abb. 3 Der Verlauf des Retentionsindex im Laufe einer Messserie A) auf dem System VAR5 und B) dem System MDN35 für die übliche Betriebsdauer einer Kapillarsäule.**

Referenzchromatogramme wurden bislang nicht von jeder kommerziell erhältlichen Chemikalie aufgenommen, da mit der Standardisierung metabolischer Daten Spektrenbibliotheken und Experimente in zunehmender Zahl im Internet veröffentlicht werden<sup>[86, 94]</sup>. Daher wurde untersucht, inwieweit Retentionsindices von verschiedenen Laboratorien vergleichbar und übertragbar sind.

Der Retentionsindex wird durch die Polarität der Säule, das Temperaturprogramm und die Wahl der internen Standards beeinflusst. Zunächst wurde der Einfluss der Polarität der Säule auf den Retentionsindex untersucht. Dafür wurden Säulen gleicher beziehungsweise unterschiedlicher Polarität verwendet (Abb. 4A und Abb. 4B).

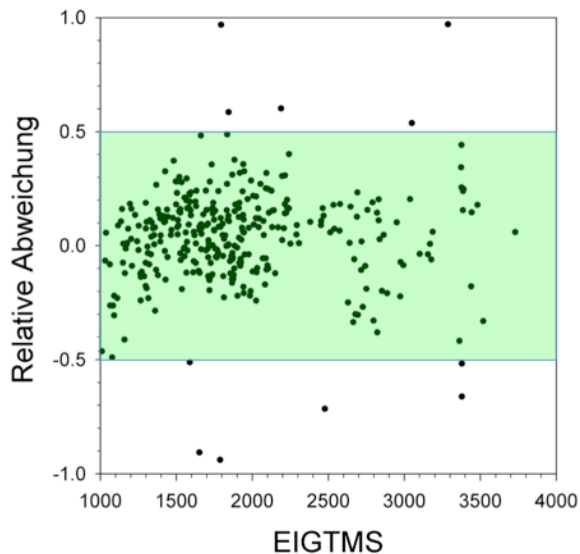


**Abb. 4 Transfer von Retentionsindices verschiedener Laboratorien im Hinblick auf A) Säulen gleicher Polarität und B) Säulen unterschiedlicher Polarität.** Die beiden Ausreißer Adenin (1TMS), Hexadecan-1-ol (1TMS) in A) sind auf Fehlidentifizierungen auf dem EIGTMS System zurückzuführen und die beiden Ausreißer 2-Amino-2-deoxyglucose-6-phosphat (6TMS) und Homoglutamin (3TMS) in B) auf Fehlidentifizierungen auf dem VAR5 System.

Bei Säulen gleicher Polarität wurde der Retentionsindex durch eine 5 %-Phenyl-95 %-Dimethylpolysiloxan-Kapillarsäule in zwei verschiedenen Laboratorien (EIGTMS<sup>[95]</sup> und VAR5<sup>[36]</sup>) bestimmt (Abb. 4A). Anhand des linearen Verlaufes ist eindeutig feststellbar, dass der Retentionsindex zwischen den beiden Laboratorien übertragbar ist. Der Retentionsindex vom VAR5 Laboratorium konnte mit Hilfe eines Polynoms dritter Ordnung auf das EIGTMS Laboratorium ( $R^2 = 0.99984$ ) übertragen werden. Ein Polynom höheren Grades ( $n > 3$ ) verringerte den durch den Transfer entstehenden Fehler nicht. Der Vergleich von zwei Säulen mit unterschiedlicher Polarität (5 %-Phenyl-95 %-Dimethylpolysiloxan/35 %-Phenyl-65 %-Dimethylpolysiloxan) zeigt zwar einen linearen Verlauf (Abb. 4B), jedoch existiert im Vergleich zu Abbildung 4A eine deutlich größere Streuung der Retentionsindices. Dies ist anhand des deutlich niedrigeren Korrelationskoeffizienten erkennbar ( $R^2 = 0.96934$ ). Die unzureichende Vergleichbarkeit ist vermutlich auf sekundäre Wechselwirkungen der Säule mit dem Analyten

zurückzuführen. Aufgrund dieser Ergebnisse wurden die folgenden Untersuchungen lediglich auf Säulen gleicher Polarität beschränkt.

Der Retentionsindex wurde zwischen den verschiedenen Laboratorien mit Hilfe eines Polynoms dritter Ordnung übertragen, wobei das Ziel verfolgt wurde höhersiedende Analyten besser anzupassen. Mit zunehmender Retentionszeit nimmt die Übertragbarkeit des Retentionsindex ab (Abb. 5). Dieses Phänomen konnte nicht mit dem höheren Siedepunkt der Analyten in Einklang gebracht werden. Die schlechtere Übertragbarkeit ist vielmehr auf die fehlende Menge an Daten in diesem Bereich zurückzuführen.



**Abb. 5 Ungenauigkeit der Retentionsindexbestimmung nach Transfer des Retentionsindex von dem VAR5 System auf das EIGTMS System über den relevanten Messbereich.** Jeder Punkt entspricht einem gaschromatographischen Signal. Der für optimal befundene Bereich wurde grün hervorgehoben.

Um dieses Ergebnis zu bestätigen, wurde der Versuch auf sechs Laboratorien (Tab. 2) erweitert. Dabei wurden die Säulen von unterschiedlichen Herstellern bezogen und unterschiedliche Temperaturprogramme zur Auftrennung der Analysenlösung verwendet. Zur Bestimmung des Retentionsindex wurden außerdem unterschiedliche n-Alkane verwendet. Im Mittel muss durch den Transfer von Retentionsindices vom VAR5 System auf ein äquivalentes System eine Ungenauigkeit von 0.5 % RI Einheiten in Kauf genommen werden (Tab. 2). Dieser Grenzwert ist insbesondere im späten Bereich der Chromatographie kritisch, da hier der absolute Fehler höher ist im Vergleich zum frühen Chromatographiebereich (Abb. 5). Hinzukommt, dass in diesem Bereich überwiegend Polysaccharide eluieren, welche ähnliche Massenspektren aufweisen und sich lediglich in ihrem Retentionsindex unterscheiden.

Der entsprechende Retentionsindex sollte daher so genau wie möglich auf dem jeweiligen Analysensystem bestimmt werden. Weiterhin ist der ermittelte Grenzwert von 0.5 % RI Einheiten für eine eindeutige Identifizierung anderer Epimere unzureichend, welche in einem

sehr schmalen Chromatographiebereich (z.B. Monosaccharide) eluieren. Die geforderte Genauigkeit kann hier durch den Retentionsindextransfer nicht mehr gewährleistet werden. Der festgelegte Grenzwert ist allerdings ausreichend für eine eindeutige Identifizierung von geometrischen Isomeren, da diese über mehrere Retentionsindexeinheiten voneinander basisliniengetrennt sind.

**Tab. 2 Transfer von Retentionsindices anderer Laboratorien auf das VAR5 Laboratorium und die daraus resultierende relative Retentionsindexabweichung.** Die Identität der einzelnen Laboratorien leitet sich von den Abkürzungen VAR5, EIGTMS, EITTMS, EIMOR, EIROE, EISTR, EIQUAD und EITRAP ab.

Laboratorium	Säulenbezeichnung	Temperaturrampe [°C/min]	Anzahl gemeinsamer Analyten	Retentionsindexabweichung [%]
VAR5 <sup>[36]</sup>	VF5	9	488	-
EIGTMS <sup>[95]</sup>	RTX5	9	348	0.38
EITTMS <sup>[95]</sup>	RTX5	15	274	0.42
EIMOR <sup>[206]</sup>	DB5	40	179	0.52
EIROE <sup>[208]</sup>	VF5	6	157	0.59
EISTR <sup>[13]</sup>	DB5	6	175	0.55
EIQUAD <sup>[207]</sup>	RTX5	5	318	0.64
EITRAP <sup>[35]</sup>	Eq5	3	65	0.52

Nachdem neue Referenzspektren hinzugefügt worden, zeigt eine Wiederholung dieser Untersuchungen die Erfordernis der Aktualisierung der Funktion zur Vorhersage des Retentionsindex. Während der Übertragung des Retentionsindex vom EIGTMS System auf das VAR5 System wurde festgestellt, dass eine lineare Regression bereits ausreichend ist. Höhere polynomiale Regressionen bewirken einen zusätzlichen Fehler in der Retentionsindexübertragung. Bis zu diesen Untersuchungen wurden 262 neue Massenchromatogramme für das VAR5 System hinzugefügt. Diese umfassen den gesamten Chromatographiebereich und sämtliche biochemische Substanzklassen. Die zusätzlichen Massenchromatogramme während der isothermen Phase bei 350 °C ermöglichen die Anpassung durch eine lineare Regression.

### C.1.2 Der massenspektrale Vergleich

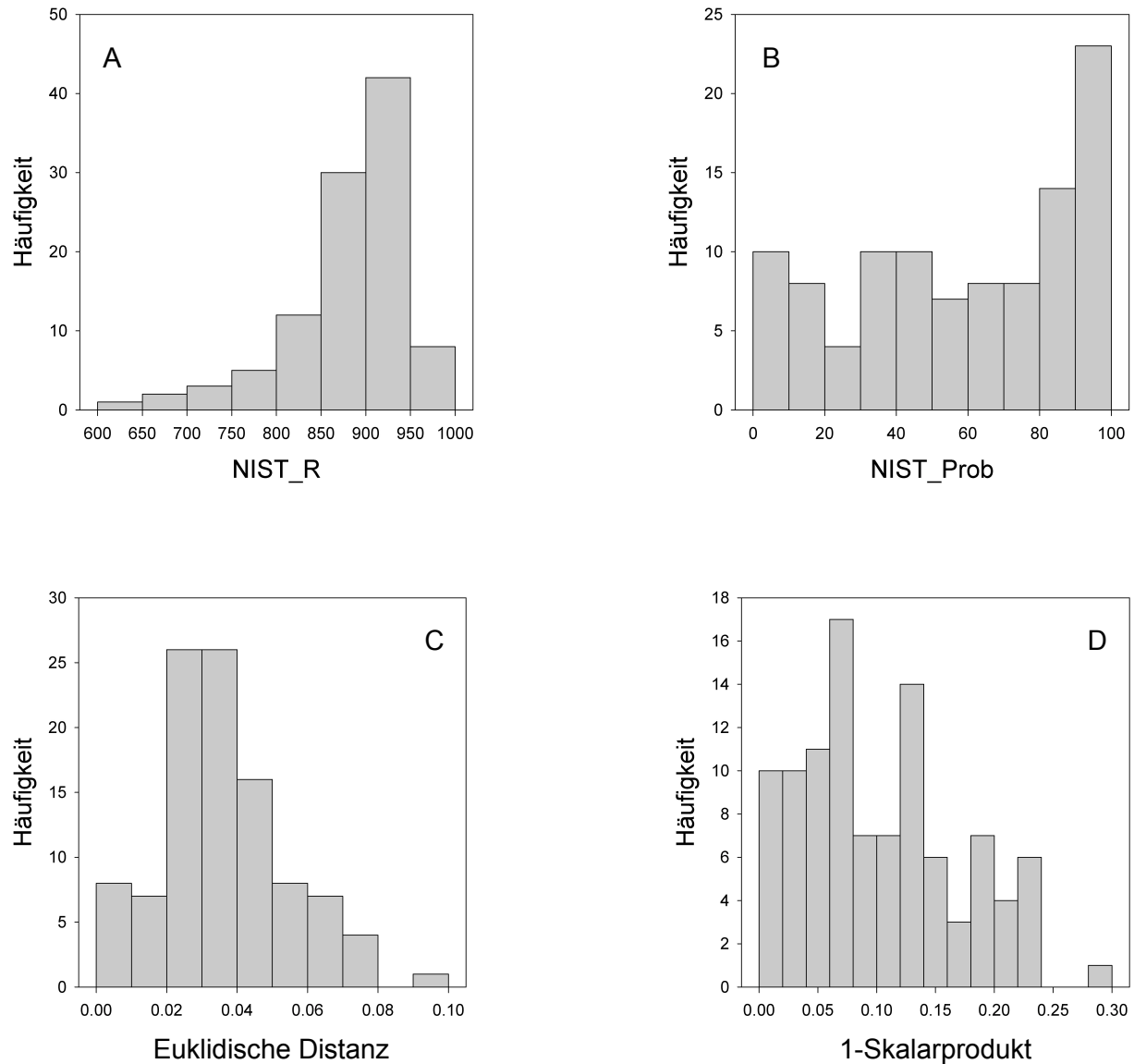
Für eine eindeutige Zuordnung eines Signals zu einer chemischen Verbindung ist nicht nur der Retentionsindex entscheidend, sondern auch das Massenspektrum. Dies muss aus dem Chromatogramm extrahiert und dekonvolviert werden. Da die Dekonvolution von

Massenspektren biologischer Proben aufgrund der hohen Komplexität zum Teil versagt, wurden für die Erstellung der Referenzspektrenbibliothek nur Chromatogramme verwendet, die aus gering konzentrierten Proben stammen<sup>[96]</sup>.

Obwohl der massenspektrale Vergleichsprozess bereits mehrfach untersucht wurde, sind bislang keine allgemeinen Kriterien publiziert worden, welche vorgeben wie ein Massenspektrum mit einem Referenzspektrum abgeglichen werden muss<sup>[97-101]</sup>. Das „National Institute of Standards and Technology“ (NIST) hat mit dem Programm „MS Search v.2.0“ für diese Fragestellung zwar einen Algorithmus entwickelt, diesen aber nicht publiziert<sup>[63]</sup>. Daher wurden mehrere gängige massenspektrale Vergleichsparameter, wie die Euklidische Distanz, das Skalarprodukt, die Hammingdistanz und die Jaccarddistanz näher untersucht und miteinander verglichen. Die euklidische Distanz quadriert alle Intensitätsdifferenzen zwischen gemessenem Spektrum und Referenzspektrum, addiert diese auf und zieht im Anschluss die Quadratwurzel. Für die Bestimmung des Skalarproduktes werden die Intensitäten aller übereinstimmenden Massenfragmente beider Spektren multipliziert und schließlich aufsummiert. Zur Bestimmung der Hammingdistanz und der Jaccarddistanz wird die Anwesenheit einzelner Massen in beiden Spektren überprüft. Die Intensität dieser nimmt hier nur eine untergeordnete Rolle ein.

Für diese Versuchsreihe dienten die von NIST ausgegebenen Parameter (R-Wert, Wahrscheinlichkeit [%]) als Bezugsgröße. Dazu wurden 119 Massenchromatogramme aus einer synthetischen Probe extrahiert, die biologisch relevante Aminosäuren, Amine, Nukleotide, Phenylpropanoide, organische Säuren und Zucker enthält (Anhang Tab. 3), und mit der aktuellen Spektrenbibliothek abgeglichen. Im Anschluss wurde die Verteilung der korrekt gefundenen Analyten bestimmt und mit der von NIST verglichen. Hierbei wurde deutlich, dass die euklidische Distanz (Abb. 6C) und das Skalarprodukt (Abb. 6D) für den massenspektralen Vergleich am besten geeignet sind. Beide wiesen zwar eine unterschiedliche Verteilung auf, hatten jedoch verglichen zu den anderen Qualitätsparametern sehr kleine Vergleichswerte (0.1-0.3). Die falsch positiven Analytvorschläge, welche zu einer ungültigen Aussage des Anwenders führen, wiesen deutlich höhere Vergleichswerte auf.

Im Gegensatz dazu erwiesen sich die Parameter Hammingdistanz und Jaccarddistanz als wenig geeignet, da sie die richtig positiven Analyten nicht anhand des Vergleichsmaßes von den falsch positiven Vorschlägen unterscheiden konnten (Anhang Abb. 2). Die Ursache liegt darin, dass bei diesen Qualitätsparametern lediglich nach der An- beziehungsweise Abwesenheit von Massen in beiden Spektren gesucht wird. Die Intensität der Signale wird dabei nicht berücksichtigt. Darüber hinaus scheint es, dass die Euklidische Distanz dem R Wert (Abb. 6A) und das Skalarprodukt der Wahrscheinlichkeit des NIST Algorithmus (Abb. 6B) ähnlich sind. Es wird daher empfohlen beide Parameter für den Vergleich zu verwenden. Dennoch ist ein visueller Vergleich notwendig, um falsche positive Einträge auszuschließen. Im nächsten Schritt wurde eine massenspektrale Spektrenbibliothek erstellt, damit die detektierten Signale der untersuchten biologischen Proben möglichst schnell einer chemischen Verbindung zugeordnet werden können.



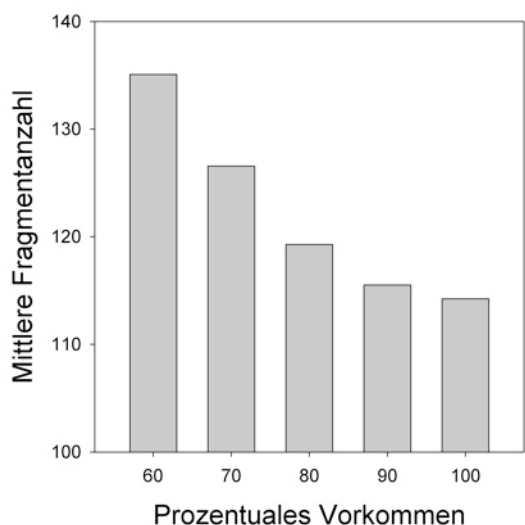
**Abb. 6 Vergleich der massenspektralen Vergleichsparameter A) NIST\_R; B) NIST\_Prob; C) Euklidische Distanz und D) 1-Skalarprodukt in Bezug auf alle Analyten in Bezug auf „Mix 2“.**

### C.1.3 Referenzspektrenbibliothek

Im Rahmen dieser Arbeit wurde die bestehende Spektrenbibliothek des Max Planck Instituts für Molekulare Pflanzenphysiologie erweitert und fehlerhafte Einträge eliminiert. Es wurde ein Algorithmus erarbeitet, der für jede chemische Verbindung durch Mittelung der vorhandenen Einzelspektren ein repräsentatives Massenspektrum berechnet. Die programmiertechnischen Arbeiten wurden von Dr. J. Hummel übernommen.

Gegenwärtig verfügt die Spektrenbibliothek über 2194 nicht-redundante Massenchromatogramme (Stand Mai 2010), denen bis zu 30 dekonvolutierte Massenspektren zu Grunde liegen. Diese stammen von biologischen Proben<sup>[58]</sup> oder von kommerziell erhältlichen Referenzsubstanzen<sup>[102]</sup>. Davon sind 843 dieser Verbindungen bislang nur in Form von Einzelspektren, 628 in Form von zwei Replikaspektren und 173 in Form von drei Replikaspektren in der Spektrenbibliothek vertreten. Die weiteren archivierten Verbindungen sind mit mehr als drei Spektren präsent. Die Zahl der zur Durchschnittsspektrenberechnung zur Verfügung stehenden Replikaspektren nimmt dennoch mit der Zahl der Analyten rapide ab (Anhang Abb. 3). Um diese Menge nicht noch durch die Wahl des Detektors einzuschränken, wurde die Berechnung nicht-redundanter Massenspektren unabhängig vom Detektortyp durchgeführt. Schließlich wurden alle Massenspektren über einen großen Massenbereich (70-600 amu) aufgenommen und auf die Aufnahme ausgewählter Massenspuren bewusst verzichtet.

Für die Berechnung nicht-redundanter Spektren wurden nur Massenfragmente in Betracht gezogen, die in 80 % der Fälle in den jeweiligen Einzelspektren der entsprechenden Verbindung vorkamen. Diese Einschränkung ist sinnvoll, da derartige Durchschnittsspektren weniger Massenfragmente aufweisen als Durchschnittsspektren mit einem geringeren Grenzwert (Abb. 7). Bei Spektren mit einem geringeren Grenzwert als 80 % wurden zu viele Massenfragmente in Betracht gezogen, welche nicht für die entsprechende Verbindung relevant sind. Das ist die Folge von Fehldekonvolutionen und koeluiierenden Substanzen. Auf die Verwendung von Durchschnittsspektren mit 100 % Vorkommenswahrscheinlichkeit wurde verzichtet, da hierbei zu viele Massenfragmente eliminiert wurden, welche in den Einzelspektren nicht detektiert wurden. Das trifft insbesondere auf den höhermolekularen Massenbereich zu.

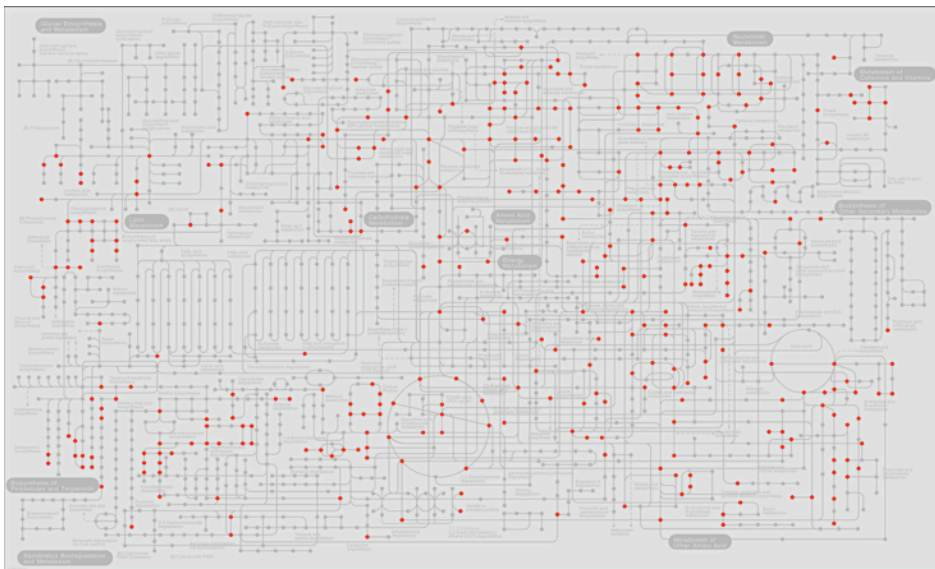


**Abb. 7 Vergleich der Qualität von Durchschnittsspektren auf der Basis der mittleren Fragmentanzahl aller Analyten.**

Im Allgemeinen konnten durch die Mittelung der Replikaspektren fehldekonvolutierte Massenfragmente und Matrixeinflüsse eliminiert werden. Zur Verbesserung der Spektrenbibliothek wird empfohlen, mit jedem Experiment weitere Massenspektren zu extrahieren und in die Massenspektrenbibliothek aufzunehmen, um so zum einen die Menge an Spektren für die Durchschnittsspektrenbibliothek zu vergrößern und zum anderen matrixabhängige Bibliotheken zu erstellen.

Auf Basis der etablierten Referenzspektrenbibliothek (Stand Mai 2010) können derzeit 65 % der gaschromatographischen Signale einer chemischen Struktur zugeordnet werden. Der verbleibende Teil (35 %) umfasst erwartete unbekannte Massenchromatogramme, welche im Rahmen früherer Messserien bereits beobachtet und inventarisiert wurden. Die bereits chemisch identifizierten Massenchromatogramme wurden aus dem Retentionsindextransfer (Kap. C.1.1) beziehungsweise durch Messen kommerzieller Referenzsubstanzen erhalten. Dabei wurden sovielerivate wie möglich mit Massenspektrum und Retentionsindex aus einem Chromatogramm extrahiert und in die Spektrenbibliothek aufgenommen. Uns war bewusst, dass womöglich nicht alle Derivate in pflanzlichen Proben detektiert werden können.

Die Massenspektrenbibliothek, welche unter folgendem Link heruntergeladen werden kann <http://gmd.mpimp-golm.mpg.de/download/>, deckt derzeit 806 Substanzen ab. Davon können 521 Substanzen für die Matrix „*Arabidopsis thaliana*“ in die Stoffwechselwege der KEGG (Kyoto Encyclopedia of Genes and Genomes) Datenbank eingeordnet werden (Abb. 8).



**Abb. 8 Einordnung GC-MS zugänglicher Metaboliten von *Arabidopsis thaliana* (thale cress) in den allgemeinen KEGG Referenzstoffwechselweg.** Die roten Punkte heben die GC-MS messbaren Stoffwechselprodukte hervor und die Striche die dazugehörigen Reaktionen. Die grauen Punkte können nicht mit der gegenwärtigen Aufarbeitungs- und Messmethode (Kapitel B.2.3) erfasst werden.

Die verbleibenden 285 Substanzen sind Verbindungen, welche noch nicht in KEGG beschrieben sind. Zum Teil handelt es sich um Laborkontaminationen, wie Alkane, Fettsäuren



und hydroxylierte Pyridinderivate. Diese gelangen während der Probenvorbereitung in die gaschromatographische Analyse. Außerdem werden während eines chromatographischen Laufes ebenfalls Siloxane detektiert. Diese stammen überwiegend von dem Säulenmaterial oder dem im Überschuss eingesetztem Silylierungsreagenz.

Entsprechend Abbildung 8 deckt die etablierte GC-MS Bibliothek und damit auch die gegenwärtige Messmethode überwiegend Aminosäuren, Fettsäuren, Amine, Alkohole, Zucker, aminierte Zucker, Zuckeralkohole, Phosphate, hydroxylierte Säuren, aromatische Verbindungen, Nukleotide, ausgewählte Lipide und kleine Sekundärmetaboliten ab. Die ausgewählte Extraktionsmethode und die GC-MS Methode erfassen diese Substanzen allerdings nicht in gleicher Qualität, da die Extraktion, die Derivatisierung und die Messung unselektiv sind. So können zum Beispiel im Extrakt chemische Reaktionen ablaufen bevor dieser prozessiert wird, da die eingetrockneten Proben bis zur online Derivatisierung auf einem ungekühlten Probensteller stehen. Dies kann eine unvollständige Derivatisierung zur Folge haben. Deshalb ist eine interne Standardisierung des Analysenprozesses erforderlich.

#### C.1.4 Standardisierung des Extraktions- und des Derivatisierungsprozesses

Ein idealer interner Standard sollte nach Möglichkeit alle Unzulänglichkeiten während der Extraktion, der Derivatisierung und der Injektion herausmitteln. Er sollte daher dem Analyten vom physikalischen Verhalten möglichst ähnlich sein. Da man jedoch im Zuge der Analyse des Metaboloms von Nutzpflanzen mehr als 100 Substanzen verschiedenster Stoffklassen simultan erfassen und quantifizieren möchte, muss man bei der Wahl des internen Standards durchaus Kompromisse eingehen. Eine mit schweren Isotopen angereicherte Matrix ist für solch eine Fragestellung immer noch die Methode der Wahl<sup>[103-105]</sup>. Dies ist allerdings bei der Matrix „Kartoffelknolle“ nur schwer möglich, da die Knolle aufgrund des hohen Stärkeanteils nur zum Teil mit schweren Isotopen angereichert werden kann. Außerdem ist der Anreicherungsgrad stark substanzabhängig und somit für eine interne Standardisierung der Matrix „Kartoffelknolle“ unzureichend<sup>[106]</sup>. Weiterhin ist es im Rahmen umfangreicher metabolischer Untersuchungen unmöglich eine große Zahl von Proben in einer überschaubaren Zeit zu markieren. Im Folgenden wurde daher überprüft, welche Substanzen für eine Standardisierung von Chromatogrammen der Matrix „Kartoffelknolle“ geeignet sind. Hierzu wurden deuterierte und <sup>13</sup>C / <sup>15</sup>N markierte Substanzen verschiedenster Stoffklassen erworben (Anhang Tab. 1) und auf ihre Tauglichkeit als interner Standard überprüft. Die Wahl der mit schweren Isotopen angereicherten Verbindungen beschränkte sich auf bedeutende Primärmetaboliten mittlerer Konzentration in der Matrix „Kartoffelknolle“<sup>[107]</sup>. Des Weiteren wurde im Rahmen dieser Versuchsreihe untersucht, ob eine stoffgruppenspezifische Normierung möglich ist und sogar eine höhere Präzision erzielt. Schließlich könnten die nachfolgenden Verbindungen zur Normierung folgender Stoffklassen herangezogen werden: Tryptophan für aromatische Verbindungen, Ethanolamin für Aminoalkohole, GABA und Succinat für Säuren, Putrescin für Polyamine, Glucose für Einfachzucker, Saccharose für Disaccharide sowie die restlichen Aminosäuren für stickstoffhaltige Säuren. Dazu wurden die relativen Standardabweichungen

aller internen Standards aus sämtlichen GC-TOF-MS Experimenten ermittelt. Diese sind für die betrachteten internen Standards in Tab. 3 zusammengefasst.

Aus Tabelle 3 ist ersichtlich, dass Succinat für eine interne Standardisierung der erzeugten GC-MS Daten am besten geeignet ist. Es wies im Vergleich zu den anderen Substanzen im Mittel die geringste relative Standardabweichung auf und wurde weiterhin in allen Experimenten wiedergefunden. Die anderen Substanzen zeigten nicht nur höhere mittlere Standardabweichungen, sondern wurden teilweise auch nicht wiedergefunden ( $^{13}\text{C}$ -Glucose und  $^{13}\text{C}$ -Saccharose), da der entsprechende endogene Metabolit zum Teil überladen war und somit die selektiven Massen der markierten Substanz diskriminierte. Daher wurden im Folgenden alle gemessenen Daten auf Succinat bezogen auch wenn hierbei lediglich carbonylhaltige Substanzen korrekt standardisiert werden. Diese Art der Standardisierung führt verglichen zu keiner Standardisierung zu einer besseren Auflösung der Daten. Dieser Effekt wurde auch bei anderen metabolischen Datensätzen beobachtet<sup>[108]</sup>.

**Tab. 3 Relative Standardabweichung in % verschiedener interner Standards in Bezug auf das jeweilige Experiment.** Ala: Alanin; ETA: Ethanolamin; GABA: 4-Aminobuttersäure; Leu: Leucin; Lys: Lysin; Put: Putrescin; Ser: Serin; Suc: Succinat; Trp: Tryptophan; Val: Valin; Die erste Zeile gibt die Nummer des entsprechenden Experimentes an, das zur Berechnung der relativen Standardabweichung herangezogen wurde.

	7031	7041	7128	8098	8108	8191	8220	9040	9071
Ala	52.9	-	74.00	67.3	72.3	36.0	25.4	51.2	-
ETA	34.1	35.4	34.3	39.2	49.2	30.3	21.2	24.5	30.1
GABA	29.14	-	22.5	51.8	55.3	31.7	19.9	28.1	31.9
Leu	36.00	-	82.4	-	24.7	31.8	17.6	19.9	30.4
Lys	31.5	49.0	29.7	62.6	73.9	39.1	29.1	26.2	48.8
Put	55.1	66.2	56.9	46.2	47.8	36.2	17.7	26.3	30.8
Ser	37.8	39.9	51.5	47.5	46.8	29.2	19.0	28.9	-
Suc	30.34	25.51	24.3	43.1	43.3	28.0	15.9	17.0	29.7
Trp	30.42	49.3	41.5	46.6	44.2	43.9	48.3	49.1	36.5
Val	41.6	30.5	-	48.6	40.5	22.8	21.3	12.4	22.7

### C.1.5 Langfristige Datenarchivierung

Die Anwendung normierter Daten auf multivariate Statistiken erfordert das Speichern der Rohdaten mit den daran geknüpften Proben- und Identifizierungsinformationen in einer Datenbank. Für die langfristige Datenarchivierung eignet sich eine Datenbank, wie zum Beispiel die GMD<sup>[82]</sup>. Hier werden die erhobenen Daten aus der GC-MS und die Boniturdaten der

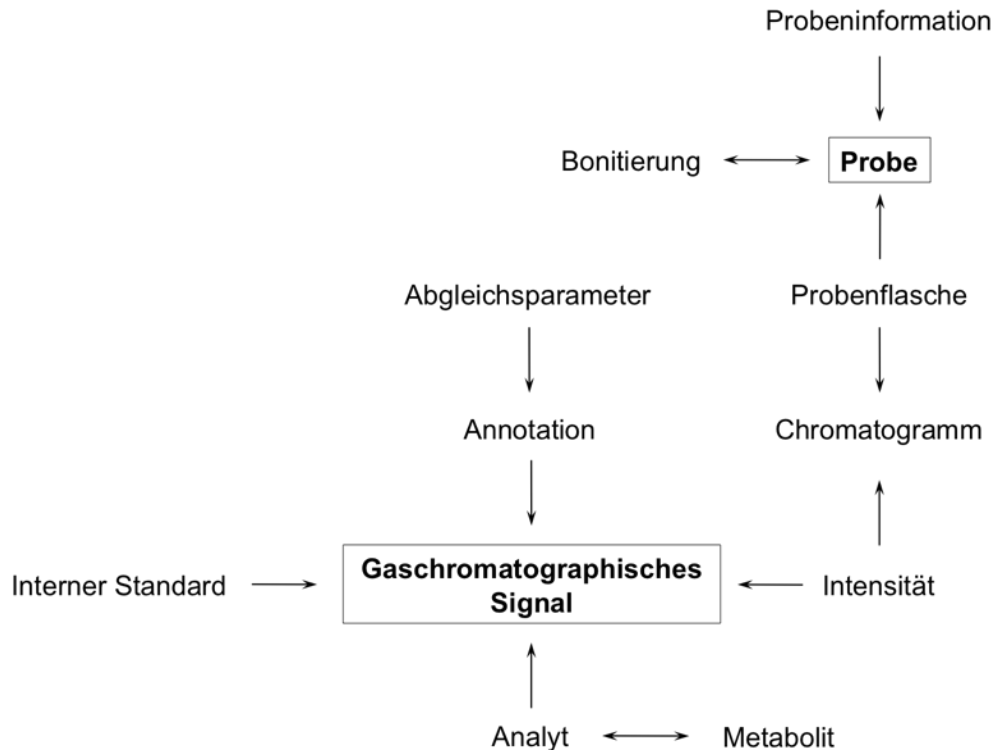
Züchter in separaten Tabellen gespeichert und im Anschluss miteinander verknüpft (Abb. 9). Dadurch können Experimente besser miteinander verglichen und physiologisch relevante Stoffwechselprodukte identifiziert werden.

Den identifizierbaren gaschromatographischen Signalen werden substanzspezifische Informationen, wie der Name des Derivates, die Summenformel, die Molmasse, die chemische Struktur und die Anzahl derivatisierter Funktionalitäten zugeordnet. Diese charakterisieren nicht nur das detektierte gaschromatographische Signal, sondern ermöglichen auch einen Vergleich mit anderen biologischen Experimenten und metabolischen Datenbanken<sup>[94, 109-114]</sup>.

Die gesammelten Informationen wurden im Folgenden auf den entsprechenden Metaboliten übertragen. Die substanzspezifischen Informationen dazu wurden in einer separaten Tabelle („Metabolitentabelle“) hinterlegt. Dies war erforderlich, da die gaschromatographischen Signale den Reaktionsprodukten aus der Methoxyaminierung und der Silylierung zuzuordnen sind. Außerdem können aus einem Metabolit zum Teil mehrere Reaktionsprodukte und daher auch mehrere gaschromatographisch aufgetrennte Signale entstehen. Die Information zum Analyten wurden in einer separaten Tabelle („Analytentabelle“) hinterlegt. Beide Tabellen beinhalten die chemische Struktur, gängige Synonyme, die CAS-Nummer und die KEGG Nummer. Die KEGG Nummer ist im Rahmen metabolischer Datenbanken ein wichtiger Identifizierungsparameter, da sie eine problemlose Suche innerhalb der KEGG Datenbank ermöglicht<sup>[115]</sup>. Weiterhin schafft sie eine Verbindung zu anderen metabolischen Datenbanken und wahrt den Bezug zur Biologie. Unter Verwendung der KEGG Nummer können die gefundenen Metaboliten in einfacher Weise in bereits aufgeklärte Stoffwechselwege eingeordnet werden<sup>[116]</sup>.

Eine Datenbank ist in der Lage neben der Archivierung von aufgenommenen GC-MS Daten und physiologischen Daten eine Archivierung von experimentellen Probeninformationen zu ermöglichen. Dies ist im Laufe der Probenvorbereitung und der Kennzeichnung von Proben erforderlich. Auf diese Weise ist einem Chromatogramm die notierte Probeninformation schnell zugänglich. Das erleichtert die Durchführung der multivariaten Statistik erheblich. Darüber hinaus können Datenmatrizes verschiedener Experimente anhand der getätigten Annotationen zusammengefügt und miteinander verglichen werden.

Diese Auswertungen sind mit Hilfe von Exceltabellen nur schwer realisierbar, da hier lediglich eine 1:1 Beziehung möglich ist und der zur Verfügung stehende Speicherplatz oft nicht ausreicht. Aus diesem Grund wurden die gemessenen Daten in Kooperation mit Dr. J. Hummel in das in Abbildung 9 dargestellte Datenbankschema integriert. Die Informationen wurden dabei so abgelegt, dass die archivierten Daten leicht auf Frischgewicht und internen Standard normierbar sind.



**Abb. 9 Datenbankschema zur Archivierung der erhobenen metabolischen und phänotypischen Daten.** Der Pfeil zeigt vom Fremdschlüssel auf den entsprechenden Primärschlüssel.

Für den Export wurden neben den normierten Intensitäten aus der Datenbank die KEGG Nummer, die CAS Nummer und die Summenformel aller identifizierten Analyten und Metaboliten ausgegeben. Sämtliche Qualitätsparameter bezüglich des Substanzvergleichs, wie die Retentionsindexabweichung vom Referenzspektrum und der massenspektrale Vergleich mit dem Referenzspektrum wurden ebenfalls in die Ausgabe ergänzt. Dies wurde von anderen Arbeitsgruppen in der Weise noch nicht realisiert.

Um das Substanzspektrum zu erweitern, wurde im Folgenden das Aufarbeitungsprotokoll soweit abgeändert, dass hoch konzentrierte Substanzen leicht abgereichert und niedrig konzentrierte Substanzen leicht angereichert werden können.

#### C.1.6 Analytik von Spurensubstanzen in biologischem Gewebe

Das vorgegebene Aufarbeitungsprotokoll (Kap. B.2.3.2.1) zielt auf die Analyse mittel- bis hoch konzentrierter Primärmetaboliten ab. Diese sind meist nicht ausschlaggebend für die Regulation von Stoffwechselfvorgängen, da sich ihr Verhalten aus der Zunahme bzw. Abnahme anderer weitaus niedrig konzentrierter Substanzen ergibt. Um dennoch halbquantitative Aussagen über derartige Regulatoren treffen zu können, wurde das Aufarbeitungsprotokoll soweit abgeändert, dass Spurensubstanzen wie z.B. Phytohormone, aus pflanzlichem Gewebe auf einfachem Weg für die Analyse zugänglich sind. Dazu wurde der Extrakt vor der Messung mittels Festphasen-

extraktion aufgearbeitet und anschließend mit Hilfe von GC-MS und LC-MS gemessen. Als Modellsubstanzen dienten Substanzen mit einer sauren Funktionalität wie beispielsweise Abscisinsäure (ABA), Indoleessigsäure (IAA) und Gibberellinsäure (GA3). Das Arabidopsis-rosettenblatt diente bei diesen Versuchsreihen als biologische Matrix.

#### C.1.6.1 Fest-Flüssig-Extraktion

Die Extraktion von Phytohormonen aus pflanzlichem Gewebe ist ein entscheidender Schritt im Zuge der Probenvorbereitung. So kann sie einerseits hinsichtlich bestimmter Substanzklassen diskriminierend wirken und andererseits die Komplexität des Analysenextraktes herabsetzen. Darüber hinaus muss die Extraktion so eingestellt sein, dass die extrahierten Substanzen aufgrund zu langer Extraktionszeit nicht weiter umgesetzt werden.

In der Literatur wurde eine Vielzahl von Methoden zur Extraktion von Phytohormonen beschrieben<sup>[27-31, 117]</sup>. Diese unterscheiden sich in der Wahl des Extraktionsmittels, (Isopropanol, 80 % Methanol, Lösungsmittelgemische), der Extraktionsdauer (0.5-12 h) sowie in der Arbeitstemperatur (-20 °C, 4 °C, 70 °C). Um die optimalen Bedingungen für die Extraktion von Phytohormonen aus pflanzlichem Gewebe zu finden, wurden bereits veröffentlichte Vorschriften getestet. Dabei konnte kein Unterschied in der Extraktionseffizienz der einzelnen Methoden festgestellt werden. Daraus folgt, dass Methanol welches bislang zur Extraktion von hochkonzentrierten Primärmetaboliten eingesetzt wurde, auch für die Extraktion von Phytohormonen geeignet ist. Es wurde mehrfach (dreimal) extrahiert und die Extrakte wurden anschließend vereinigt.

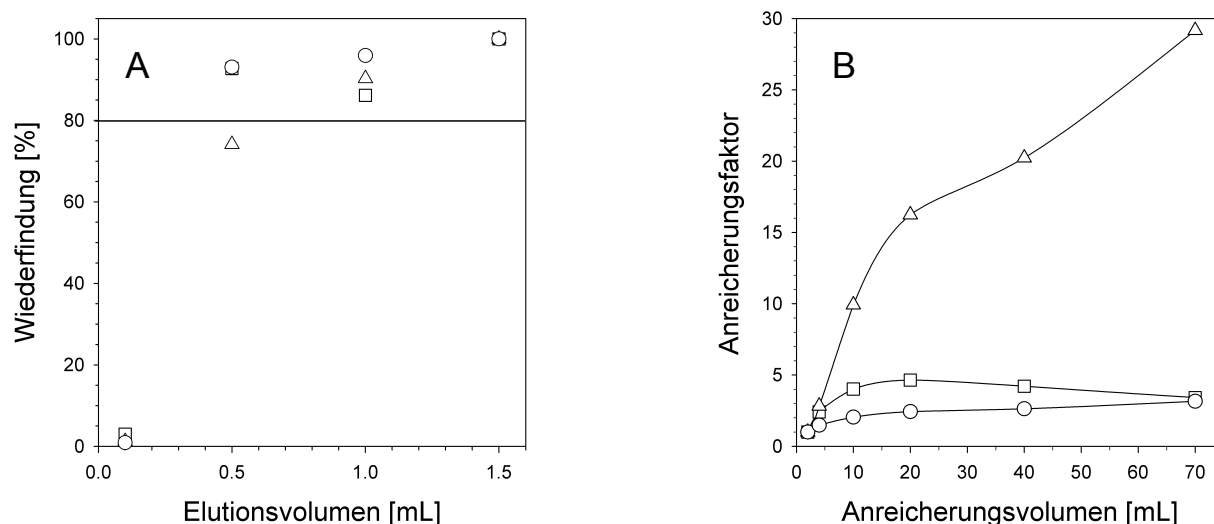
Da jedoch durch Variation der Extraktionsvorschrift kein erhöhtes Signal zu erreichen ist, muss dies entweder durch Erhöhung des Frischgewichtes oder durch einen zusätzlichen Aufkonzentrierungsschritt erreicht werden. Hierfür ist die Festphasenextraktion die Methode der Wahl. Sie setzt die Komplexität der zu analysierenden Matrix herab und reichert Zielsubstanzen um mehrere Zehnerpotenzen an<sup>[118]</sup>. Zur Etablierung der Festphasenextraktion muss als erstes eine geeignete Kartusche ausgewählt und der Beladungs-, Wasch- und Elutionsschritt optimiert werden.

#### C.1.6.2 Festphasenextraktion

Die Auswahl einer besonders geeigneten Festphase für die Anreicherung von sauren Phytohormonen und deren anschließende Abreicherung störender Koelutionen zählt zu den zentralen Fragestellungen bei der Festphasenextraktion. Dazu wurden Kartuschen unterschiedlicher chemischer Funktionalität und gleicher Geometrie untersucht (PPL, Plexa, HLB, C18; jeweils 3mL). Die HLB Kartusche besteht aus einem Polymer, welches in der Lage ist hydrophile und lipophile Wechselwirkungen mit den Analyten einzugehen. Dieses Polymer besteht aus den Monomereinheiten *N*-Vinyl-2-pyrrolidinon und Divinylbenzol. Die Sorbentien Plexa und PPL hingegen verfügen über eine hydroxylierte Polymeroberfläche und retardieren

überwiegend polare Verbindungen. Trotz unterschiedlicher chemischer Beschaffenheit sollten alle Kartuschen saure und polare Substanzen retardieren.

Die Voruntersuchungen zeigen, dass Sorbentien auf der Basis von C18 und auf der Basis von Polymeren für die Anreicherung von sauren Phytohormonen geeignet sind. Da jedoch C18 Sorbentien durch „endcapping“ immer hydrophober werden und dies zu einer chargen-abhängigen Retardierung der Zielmoleküle führt, wurden für die Optimierung des Anreicherungs volumens und des Elutionsvolumens zwei Sorbentien auf der Basis von Polymeren ausgewählt (Plexa, HLB)<sup>[119]</sup>. Beide Sorbentien retardieren die Zielmoleküle nach einem Umkehrphasenmechanismus. Dabei wurde zunächst überprüft, welches Sorbens unter gleichen Anreicherungsbedingungen den höchsten Anreicherungsfaktor erzielt. Der Anreicherungsfaktor gibt an, inwieweit die Analysenlösung während der Festphasenextraktion aufkonzentriert wurde. Da der Anreicherungsfaktor durch das Anreicherungs volumen und das Elutionsvolumen bestimmt wird, wurde zunächst untersucht, wie viel Methanol für eine quantitative Elution notwendig ist (Abb. 10A). Weiterhin wurde in einem weiteren Schritt das maximale Anreicherungs volumen (Abb. 10B) ermittelt.



**Abb. 10 Optimierung des A) Elutionsvolumens und des B) Anreicherungs volumens für das Sorbens Plexa zur Anreicherung und Abreicherung der Zielmoleküle ABA (○), GA3 (□) und IAA (△).** Die Elutionsvolumina in A) wurden nach der Elution jeweils auf 1.5 mL aufgefüllt. Zuvor wurden 1.5 mL Analysenlösung für die Anreicherung verwendet. Die horizontale Linie zeigt die erwartete Wiederfindungsrate (80 %). Die Wiederfindungsrate wurde dadurch bestimmt, dass die Signalintensität des 1.5 mL Eluats auf 100 % gesetzt wurde. Für die Ermittlung des Anreicherungsfaktors in B) wurde das Anreicherungs volumen schrittweise erhöht und die angereicherte Stoffmenge mit Hilfe von 1 mL Methanol von der Kartusche eluiert.

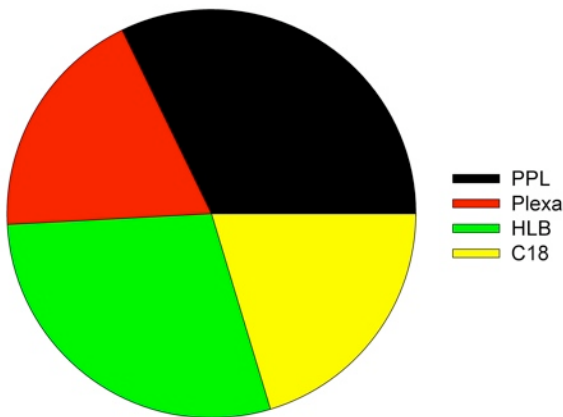
Abbildung 10A zeigt, dass 1 mL Methanol für die Elution der drei Substanzen ausreichend ist. Die Wiederfindungsrate war bei allen Substanzen größer als 80 %. Für die Elution von ABA und GA3 genügten 0.5 mL Methanol. Ein ähnlicher Trend konnte für die HLB Kartusche beobachtet werden (Anhang Abb. 4).

Da jedoch die angereicherte Stoffmenge sehr viel größer als die tatsächlich physiologisch vorkommende Stoffmenge ist, wurde im Folgenden davon ausgegangen, dass das ermittelte Elutionsvolumen für die Elution höherer Anreicherungsmengen ausreichend ist. In einer weiteren Versuchsreihe wurde untersucht, zu welchem Zeitpunkt eine Sättigung des Sorbens mit dem Analyten eintritt. Während dieser Versuchsreihe verhielten sich nicht nur die beiden Kartuschen (Plexa, HLB) sondern auch die Zielmoleküle (ABA, GA3, IAA) unterschiedlich. Der Sättigungsbereich wurde bei der Plexakartusche für die Verbindung IAA bis zu 70 mL Extrakt nicht erreicht. Die HLB Kartusche hingegen erreichte den Sättigungsbereich bereits bei 20 mL Extrakt. GA3 und ABA konnten auf beiden Kartuschen jeweils bis zu 20 mL Extrakt effizient angereichert werden. Dies zeigt, dass die Plexakartusche zusätzlich noch über Funktionalitäten verfügen muss, die mit den Analyten  $\pi$ - $\pi$  Wechselwirkungen eingehen. Diese sind bei der HLB Kartusche nicht gegeben. Die Oasis HLB Kartusche besteht laut Hersteller aus polymerisiertem *N*-Vinyl-2-pyrrolidinon und Divinylbenzol. Diese Einheiten gewähren kaum  $\pi$ - $\pi$  Wechselwirkungen. Aufgrund maximaler Anreicherung wurden im Folgenden 32 mL Extrakt für die Anreicherung an der Plexakartusche verwendet und 1 mL Methanol für die Elution. Schließlich wurde bis zu diesem Zeitpunkt noch kein Durchbruch beobachtet. Es konnte keiner der ausgewählten Analyten während der Anreicherung in der durchgeflossenen Analysenlösung detektiert werden.

Während der Festphasenextraktion sollen nicht nur die Analyten angereichert sondern auch der Extrakt von koeluiierenden Störkomponenten befreit werden. Aus diesem Grund wurden verschiedene Kartuschen im Hinblick auf eine effiziente Abreicherung der Matrix „Arabidopsisblatt“ untersucht. Darüber hinaus wurde innerhalb dieser Versuchsreihe die C18 Kartusche untersucht, da C18 repräsentativ für unpolare Sorbentien ist. Auf diese Weise wurde die Effizienz von C18-haltigen Kartuschen mit Kartuschen basierend auf Polymeren verglichen. Zur Verifizierung dieses Problems wurden die Eluate der verschiedenen Sorbentien (PPL, Plexa, HLB, C18) und ein Extrakt, welcher nicht mit Hilfe von Festphasenextraktion aufbereitet worden war, mittels GC-TOF-MS analysiert. Anschließend wurde die relative Signalintensität festgestellt und der Abreicherungsgrad bestimmt. Der Abreicherungsgrad ergibt sich aus dem Signalverhältnis zwischen SPE behandeltem Extrakt und unbehandeltem Extrakt (Abb. 11).

Es zeigte sich, dass die Plexakartusche und die C18 Kartusche die Matrix stark abreichern und zugleich auch eine Reihe von Substanzen stark anreichern. Die Plexakartusche reichert polare Substanzen, wie Succinat (Faktor: 0.5), Saccharose (Faktor: 0.7), Sitosterol (Faktor: 0.2) und Polyole (Faktor: 0.3) stark ab. Im Gegensatz dazu werden unpolare Substanzen, wie beispielsweise Fettsäuren (Faktor: 1.5), Indole (Faktor: 20) und Phytole (Faktor: 3.1) stark angereichert.

Die PPL Kartusche und die HLB Kartusche reichern die untersuchte Probe viel mehr ab als die Plexakartusche (Abb.11) und sind daher für metabolische Profilanalysen nicht geeignet. Basierend auf den vorangegangenen Untersuchungen wurde die Anreicherung saurer Phytohormone mit Hilfe des Sorbens Plexa durchgeführt.



**Abb. 11 Vergleich verschiedener Sorbentien hinsichtlich des Abreicherungsgrades.** Der Vergleich wurde auf allen GC-MS detektierbaren Substanzen durchgeführt. Diese umfassen die chemisch identifizierten gaschromatographischen Signale und die erwarteten Unbekannten.

#### C.1.6.3 Weitere Anreicherung

Um einen weiteren Anreicherungseffekt zu erzielen wurde das Eluat bis zur Trockene eingedampft, der Rückstand in Methanol (80 %-ig) gelöst und mittels UPLC-FT-ICR-MS analysiert. Dabei wurde überprüft, wie viel Methanol (80 %-ig) notwendig ist, damit bei diesem Schritt keine Analyten verloren gehen. Es wurde nachgewiesen, dass 100  $\mu$ L Methanol (80 %-ig) für die vollständige Lösung des Rückstandes notwendig sind. Nach der Festphasenextraktion konnten weniger als die Hälfte (47 %) der eingesetzten Phytohormone nicht wiedergefunden werden. Dieser Verlust wird überwiegend durch den Eintrocknungsschritt verursacht und erfordert für weitere Analysen zusätzliche Schritte zur Optimierung. Ohne den Eintrocknungsschritt könnten die Extrakte aus der Festphasenextraktion nicht mit Hilfe von Gaschromatographie gemessen werden. Der hohe Methanolgehalt würde mit dem Silylierungsmittel zu Silikonen reagieren. Außerdem würde der weitere Anreicherungsfaktor für die LC Analyse entfallen. Werden die Eluate der Festphasenextraktion (1 mL) bis zur Trockene eingengt und in 100  $\mu$ L Methanol (80 %-ig) aufgenommen, so wird trotz des hohen Substanzverlustes während der Verdampfung immerhin noch ein Anreicherungsfaktor von fünf erhalten.

#### C.1.6.4 Detektion

Bislang wurden Phytohormone überwiegend durch SIM (**S**elected **I**on **M**onitoring) detektiert, wobei das Ziel verfolgt wurde, die Sensitivität während der Detektion noch etwas zu erhöhen [27, 28, 30, 31, 117, 120-122]. Allerdings besteht bei der Analyse einer neuen Matrix die Gefahr, anstelle des Zielmoleküls eine bislang noch nicht beobachtete Koelution zu detektieren. Daher wurde die Eignung von „Full Scan Methoden“ für die Analyse von Phytohormonen untersucht. Schließlich wurden die Phytohormone während der Probenvorbereitung bereits in einem hohen Maße



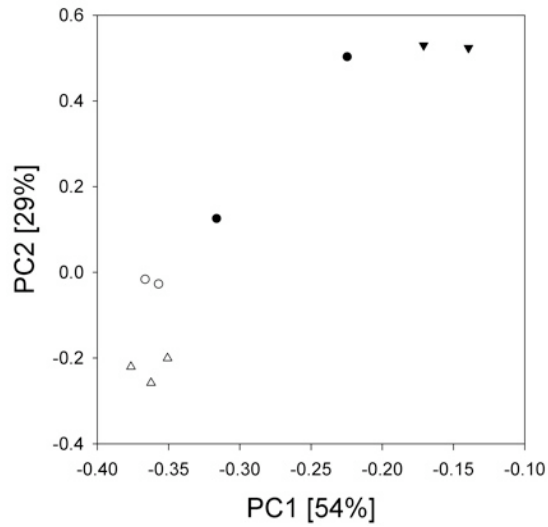
angereichert und von störenden Koelutionen getrennt. Für die Detektion der aufgearbeiteten Extrakte wurden zwei Analysetechniken ausgewählt, die am Max Planck Institut für Molekulare Pflanzenphysiologie zur Analyse des Metaboloms Anwendung finden: eine gaschromatographische Analysenmethode (GC-TOF-MS) und eine flüssigkeitschromatographische Analysenmethode (UPLC-FTICR-MS)<sup>[36, 39]</sup>. Die Ergebnisse dieser Methoden wurden anschließend miteinander verglichen. Die UPLC-FT-ICR-MS Messungen wurden von Herrn Dr. P. Giavalisco durchgeführt.

#### C.1.6.4.1 Optimierung der GC-TOF-MS Analyse

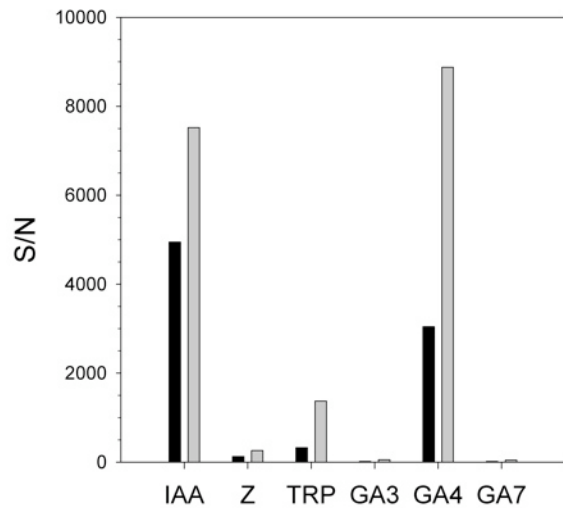
Die zeitliche Optimierung der gaschromatographischen Analyse im Zuge der Probenvorbereitung und die Erhöhung der Empfindlichkeit erfordert die Optimierung des Derivatisierungsvolumens bei einem geeigneten Temperaturprogramm. Hierbei wurde zunächst das Derivatisierungsvolumen soweit herabgesetzt bis sich das Pellet nicht mehr im Methoxyaminierungsreagenz löste. Die Temperatur (40 °C) und das Verhältnis von Methoxyaminhydrochlorid und BSTFA wurde dafür konstant gehalten. Die Hauptkomponentenanalyse (PCA) zeigt, dass das Silylierungsmittel und das Methoxyaminierungsreagenz mindestens in einem Konzentrationsverhältnis von 20 µL : 10 µL vorliegen müssen, damit eine genügend hohe Reproduzierbarkeit innerhalb einer Replikagruppe gegeben ist (Abb. 12). Ein geringeres Volumen war aufgrund einer unzureichenden und unreproduzierbaren Rücklösung nicht möglich.

Ein höheres Maß an Robustheit und Anreicherungs-effizienz kann gewährleistet werden, wenn der Extrakt in 40 µL Methoxyaminierungsreagenz und 20 µL BSTFA rückgelöst wird. Im nächsten Schritt wurde der Einfluss des Temperaturprogramms auf das Signal zu Rausch Verhältnis (S/N) untersucht. Dabei wurde erwartet, dass eine steilere Temperaturrampe ein höheres Signal zu Rausch Verhältnis zur Folge hat. Dieser Zusammenhang konnte experimentell bestätigt werden (Abb. 13). Allerdings ist dieser Effekt substanzabhängig.

Das Signal zu Rausch Verhältnis konnte für die Gibberelline bei Anwendung einer steileren Temperaturrampe um den Faktor 4 gesteigert werden, während das Zeatin und die indolhaltigen Substanzen unter den gleichen Bedingungen ein zwei- bis dreifach höheres Signal ergaben. Für die weiteren Versuche wurde das schnellere Temperaturprogramm gewählt.



**Abb. 12 Optimierung des Derivatisierungsvolumens für die Matrix „Arabidopsisblatt“ nach Anwendung einer Hauptkomponentenanalyse.** ○ 80:40; △ 40:20; ▼ 20:10; ● 10:5; (Silylierungsreagenz:Methoxyaminierungsreagenz).



**Abb. 13 Vergleich zweier Temperaturprogramme für die gaschromatographische Analyse anhand des detektierten Signal zu Rausch Verhältnisses (S/N).** schwarz: 9 °C/min; grau 25 °C/min; IAA (Indolessigsäure); Z (Zeatin); TRP (Tryptophan); GA3 (Gibberellinsäure); GA4 (Gibberellin A4); GA7 (Gibberellin A7)

## C.1.6.4.2 Vergleich zwischen GC-TOF-MS und UPLC-FTICR-MS

Um beide Detektionsmethoden (GC-TOF-MS und UPLC-FTICR-MS) miteinander zu vergleichen, wurden Verdünnungsreihen im Bereich von  $10^{-8}$ - $10^{-13}$  g Phytohormon (injizierte Menge) aufgenommen und die entsprechende Nachweisgrenze ermittelt. Hierbei wurde für das Signal zu Rausch Verhältnis ein Faktor von größer oder gleich 2 gewählt. Dabei zeigte sich, dass Moleküle mit einem geringen Molekulargewicht (IAA, SA) sensitiver mittels GC-TOF-MS erfasst werden können, während für Moleküle mit einem höheren Molekulargewicht (ABA, GA3, GA7, GA4) die empfindlichere UPLC-FTICR-MS Methode besser geeignet ist (Tab. 4). Aufgrund der Unterschiede in den Nachweisgrenzen von GC-TOF-MS und UPLC-FTICR-MS ist zu empfehlen, die Extrakte mit beiden Methoden zu untersuchen, da beide Methoden wertvolle sich gegenseitig ergänzende Informationen über die Phytohormone liefern.

Die Gaschromatographie erzielt im Allgemeinen schärfere Banden und ermöglicht eine genaue Analyse sehr komplexer Proben. Die Flüssigkeitschromatographie hingegen ermöglicht vor allem die Analyse höhermolekularer Substanzen, da der Siedepunkt dieser Substanzen nach Anwendung des Derivatisierungsschrittes für gaschromatographische Analysen zu hoch ist. Außerdem geht wertvolle Substanzmenge (50 %) während der Probenvorbereitung für die gaschromatographische Analyse verloren (Kap. C.1.6.3). Dennoch sind die resultierenden Signale der flüssigkeitschromatographischen Trennung breiter und zum Teil von Koelutionen überlagert. Dadurch wird eine Quantifizierung erschwert.

**Tab. 4 Nachweisgrenzen saurer Phytohormone im Hinblick auf die UPLC-FTICR-MS und die GC-TOF-MS Analyse unter der Annahme  $S/N \geq 2$ .** Die Angaben beziehen sich auf die injizierte Menge Phytohormon. IAA (Indolessigsäure), SA (Salicylsäure), ABA (Abscisinsäure), GA3 (Gibberellinsäure), GA7 (Gibberellin A7), GA4 (Gibberellin A4)

Metabolit	UPLC-FTICR-MS [g]	GC-TOF-MS [g]
IAA	$10^{-10}$	$10^{-12}$
SA	$10^{-10}$	$10^{-12}$
ABA	$10^{-12}$	$10^{-10}$
GA3	$10^{-12}$	$10^{-10}$
GA7	$10^{-12}$	$10^{-10}$
GA4	$10^{-12}$	$10^{-10}$

## C.1.7 Zusammenfassung und Bewertung der etablierten Analytischen Lösungsansätze

Im Rahmen dieser Untersuchungsreihen wurden wichtige analytische Lösungsansätze etabliert, die es ermöglichen GC-TOF-MS Metabolitenprofile effizienter auszuwerten. Typische

Metabolitenprofile umfassen 80-300 Chromatogramme, welche nach der Messung in einer gemeinsamen Datenmatrix angeordnet und einem stufenweisen Vergleichsprozess unterzogen werden. Dieser ist bislang nur teilweise automatisiert.

Daher wurde im Rahmen der vorliegenden Arbeit ein Grenzwert für den Retentionsindexabgleich geschaffen, der eine Auswertung von Metabolitenprofilen ermöglicht. Der Retentionsindex wurde zwischen verschiedenen Laboratorien übertragen, wodurch eine gemeinsame Nutzung metabolischer Spektrenbibliotheken und biologischer Experimente möglich wurde. Außerdem erfolgte die Integration massenspektrealer Vergleichsparameter in den automatisierten Identifizierungsprozess. Durch die Kombination von geeigneten Retentionsindexgrenzwerten und massenspektrealen Vergleichsparametern könnten in Zukunft Metabolitenprofile weitaus schneller (höherer Automatisierungsgrad) ausgewertet werden. Eine Verifizierung der Identifizierungen anhand von definierten Mixturen ist jedoch nach wie vor unabdingbar.

Damit die automatisierte Auswertung von komplexen Profildaten realisiert werden kann, wurde im Rahmen dieser Arbeit aufgezeigt, wie qualitativ hochwertige Referenzspektren aus bereits gemessenen Spektren gewonnen werden können. Der entwickelte Algorithmus erleichtert die Etablierung von umfangreichen Spektrenbibliotheken im Gegensatz zu dem von Ausloos et al. vorgeschlagenen Algorithmus und ermöglicht eine zuverlässige Identifizierung gaschromatographischer Verbindungen<sup>[123]</sup>. Die berechneten Spektren sind in der Golm Metabolome Database (<http://gmd.mpimp-golm.mpg.de/download/>) hinterlegt.

Diese beinhaltet nunmehr nicht nur Referenzspektren, sondern auch biologische Experimente. Im Rahmen dieser Arbeit wurde dazu ein Export entwickelt, welcher die gemessenen GC-TOF-MS Daten in die Stoffwechselwege der KEGG Datenbank einordnet und die Qualitätsparameter bezüglich des Substanzvergleichs mit angibt. Im Gegensatz zu anderen metabolischen Profildatenbanken können nach Anwendung des entwickelten Exportprotokolls analytische und metabolische Informationen simultan ausgegeben werden. Dadurch wird eine Verifizierung der gemessenen Signale und eine Anwendung dieser Daten für Modellierungsstudien ermöglicht. Diese sind im Zeitalter systembiologischer Ansätze erforderlich.

Durch die beschriebenen Entwicklungen hat die Metabolomik einen technischen Stand erreicht, welcher dem der anderen „-Omik“ Technologien (Proteomik und Transkriptomik) nahekommt. Die GC-TOF-MS basierte Metabolitenanalyse ermöglicht jetzt die erweiterte automatisierte Analyse ausgewählter Substanzen und biochemischer Stoffklassen.

Damit das Substanzspektrum detektierbarer Metaboliten noch erhöht werden kann, wurde im Rahmen dieser Arbeit der Schritt der Probenvorbereitung entscheidend abgeändert. Dazu wurde nicht nur eine neue Festphase verwendet, sondern der Schritt der Festphasenextraktion um einen zusätzlichen Aufkonzentrierungsschritt erweitert. Die so aufgearbeiteten Extrakte wurden mittels GC-TOF-MS und UPLC-FTICR-MS analysiert. Eine derartige Detektion („Full Scan“) und Vergleich zweier verschiedener Analysetechnologien (GC-TOF-MS vs. UPLC-FTICR-MS) ist bislang in der Literatur nicht vorgenommen worden und bildet eine solide Grundlage für weitergehende metabolische Untersuchungen.

## C.2 Physiologisch relevante Stoffwechselprodukte in der Kartoffelknolle

Die etablierten analytischen Lösungsansätze ermöglichen eine Analyse biologischer Prozesse in Nutzpflanzen. Dazu werden die Nutzpflanzen entweder unmittelbar nach ihrer Ernte oder im Verlauf eines biologischen Prozesses auf ihre stoffliche Zusammensetzung untersucht. Parallel dazu werden die Pflanzen zusätzlich bonitiert, um die Qualität der Nutzpflanzen einzuschätzen. Je nachdem ob die Nutzpflanze während des biologischen Prozesses oder mehrere Monate später hinsichtlich ihrer Qualität eingeschätzt worden ist, unterscheidet man zwischen indikativen Biomarkern<sup>[124-128]</sup> und prädiktiven Biomarkern<sup>[129]</sup>.

### C.2.1 Indikative Biomarker

Indikative Biomarker werden verwendet, um eine Knolle hinsichtlich einer vorgegebenen Eigenschaft zu klassifizieren. Dazu werden die Knollen einem gut kontrollierbaren Stress ausgesetzt und anschließend auf ihren metabolischen und physiologischen Phänotyp untersucht. Danach werden multivariate statistische Methoden auf den metabolischen Datensatz angewandt und beide Datensätze über Korrelationen zusammengeführt. Eine Voraussetzung für ein erfolgreiches Aufspüren relevanter Substanzen besteht in der Erfassung von möglichst vielen Konditionen während der Belastung. Dadurch ist eine Detektion metabolischer Flüsse möglich, wodurch eine Aussage über den biologischen Prozess abgeleitet werden kann. Im Folgenden wird an drei physiologischen Prozessen („Mechanische Belastung“, „Physiologisches Alter“, „Keimhemmung“) aufgezeigt wie indikative metabolische Biomarker an Nutzpflanzen gewonnen werden können. Als Modell dient hierfür die Kartoffelknolle. Die Ergebnisse sind zum Teil in einer eigenen Arbeit<sup>[130]</sup> und einer Publikation als Koautor eingegangen<sup>[129]</sup>.

#### C.2.1.1 Mechanische Belastung von Kartoffelknollen

Im Rahmen des Experiments „Mechanische Belastung“ wurde untersucht, inwieweit eine mechanische Belastung den Primärmetabolismus beeinflusst. Eine mechanische Belastung kann durch freien Fall oder eine Trommel simuliert werden. Dies führt zu Rissen bzw. Verfärbungen unterhalb der Belastungsstelle, welche blaugrau bis schwarz gefärbt sind<sup>[131-133]</sup>. Infolge von einer mechanischen Belastung brechen Zellmembranen auf, so dass sich das Enzym der Polyphenoloxidase im Inneren der Knolle verteilt<sup>[134]</sup>. Dort werden Tyrosin und andere Substrate zu Chinonen oxidiert, welche im nächsten Schritt zu Melaninen polymerisieren<sup>[135-137]</sup>. Melanine sind Pigmentfarbstoffe und die Ursache für die Verfärbungen unterhalb der Schale<sup>[138, 139]</sup>. Neben Tyrosin dienen ebenfalls Chlorogensäure und Kaffeesäure als Substrat<sup>[140-142]</sup>.

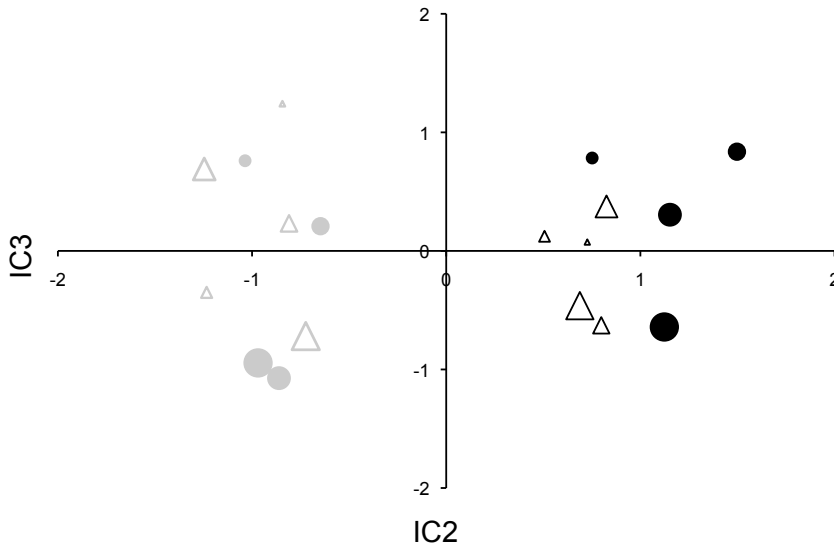
Um diesen physiologischen Prozess näher zu untersuchen wurden die Kultivare *Afra* und *Milva* ausgewählt. Diese unterscheiden sich hinsichtlich ihrer Empfindlichkeit gegenüber Schwarzfleckigkeit. Ebenfalls sind sie nicht miteinander verwandt. So weist das Kultivar *Milva*

nach den Bestimmungen der Böhm-Nordkartoffel Agrarproduktion (BNA) eine Schwarzfleckigkeitsneigung von vier auf, während dem Kultivar *Afra* eine Schwarzfleckigkeitsneigung von sieben zugeordnet werden kann<sup>[143]</sup>.

Eine Knolle kann von der Seite beziehungsweise vom apikalen Ende belastet werden. Daher wurde im ersten Schritt überprüft, an welcher Stelle der Knolle die Belastung stattfinden muss, damit der Phänotyp der Schwarzfleckigkeit in kürzester Zeit besonders stark ausgeprägt ist. Nach der mechanischen Belastung an den genannten Stellen der Knolle wurde diese im Dunkeln gelagert. Nach zwei Tagen wurde die Anzahl schwarzfleckiger Knollen gezählt und die Größe der gebildeten schwarzen Flecken registriert. Die Untersuchungen zeigen, dass eine seitliche Belastung den Phänotyp der Schwarzfleckigkeit besonders stark ausprägt. So wiesen Knollen der Sorte *Afra* nach seitlicher Belastung im Vergleich zu einer apikalen Belastung eine fünffach höhere Zahl an schwarzfleckigen Knollen auf. Darüber hinaus waren nach seitlicher Belastung die schwarzen Flecken 1.5 mal größer als nach einer apikalen Belastung aufgrund der größeren Belastungsoberfläche. Die Knollen wurden daher aus 50 cm Höhe von der Seite aus belastet. Die Belastung erfolgte zehnmals, da während einer einmaligen Belastung keine schwarzen Flecken zu beobachten waren. Nach dem zehnten Fall wurde die belastete und die gegenüberliegende unbelastete Seite mittels GC-TOF-MS auf ihre stoffliche Zusammensetzung untersucht. Darüber hinaus wurde der Aminosäure- und der Proteingehalt festgestellt, um die gefundenen Zusammenhänge besser zu verifizieren.

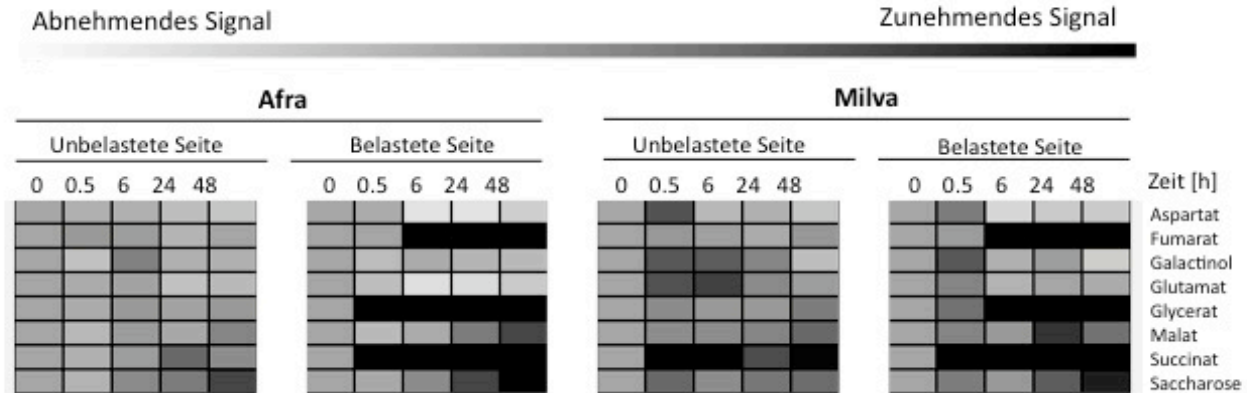
Die Unabhängige Komponentenanalyse (ICA) ist ein gängiges Verfahren zur Klassifizierung multivariater Datensätze. Sie ordnet die aufgenommenen Datenpunkte anhand einer linearen Transformationen in einem nicht-orthogonalem Koordinatensystem so an, dass die resultierenden Signalkomponenten im neuen Koordinatensystem statistisch unabhängig sind. Deshalb ist es möglich Proben ähnlicher stofflicher Zusammensetzung zu einer Gruppe zusammenzufassen und zu klassifizieren<sup>[144]</sup>. Die ICA (Abb. 14) zeigt nach vorheriger Anwendung der Hauptkomponentenanalyse (PCA) eine klare Trennung zwischen den beiden Kultivaren innerhalb der dritten Hauptkomponente (IC3) und eine deutlich geringere Trennung bezüglich der Variablen Zeit im Hinblick auf die zweite Hauptkomponente (IC2). Die übrigen Hauptkomponenten zeigten eine geringere Trennung. Die zeitliche Auftrennung war bei den belasteten Proben höher als bei den unbelasteten Proben.

Das bestätigt die Annahme, dass *Afra* empfindlicher auf eine mechanische Belastung reagiert als *Milva* und eine mechanische Belastung Änderungen im Primärmetabolismus zur Folge hat. Schließlich werden während der GC-MS Analyse überwiegend Primärmetaboliten und niedermolekulare Sekundärmetaboliten detektiert. Diese zeitlichen Änderungen waren im Kultivar *Afra* stärker ausgeprägt als im Kultivar *Milva*.



**Abb. 14 Unabhängige Komponentenanalyse (ICA) angewendet auf alle gemessenen Massenfragmente.** grau: *Milva*; schwarz: *Afra*; Punkte: mechanisch belastetes Gewebe; offene Dreiecke: unbelastetes Gewebe; Die Größe der Punkte und der offenen Dreiecke korreliert mit der Zeit.

Um die Substanzen zu detektieren, welche während der mechanischen Belastung von Interesse sind, wurde im nächsten Schritt eine Zweifaktorielle Varianzanalyse auf den vorliegenden Datensatz angewandt. Hierbei wurden die Faktoren „Zeit“ und „mechanische Belastung“ näher untersucht und für die Analyse ein Signifikanzniveau von  $\alpha = 0.05$  festgelegt (Abb. 15). Dabei wurden bezüglich des Faktors „mechanische Belastung“ die Stoffwechselprodukte Aspartat, Fumarat, Glutamat, Glycerat und Succinat als relevant erachtet und bezüglich des Faktors „Zeit“ die Stoffwechselprodukte Aspartat, Fumarat, Galactinol, Malat, Succinat und Saccharose. Fumarat, Malat und Succinat sind Stoffwechselprodukte des Respirationsstoffwechsels, Glycerat, Galactinol und Saccharose sind Vorgängerstufen dieser Substanzen. Aspartat und Glutamat werden aus Stoffwechselprodukten des Respirationsstoffwechsels gebildet. Die experimentellen Befunde aus den GC-MS Messungen weisen darauf hin, dass der Respirationsstoffwechsel und seine abzweigenden Stoffwechselwege nach der mechanischen Belastung aktiviert sind (Abb. 15). Ein Vergleich des zeitlichen Verlaufs der detektierten Stoffwechselprodukte zeigt, dass diese in zwei große Gruppen unterteilbar sind. Eine Gruppe umfasst Aspartat, Glutamat und Galactinol. Diese Stoffwechselprodukte verändern sich vorübergehend in ihrem Gehalt in den belasteten Proben. Die genannten Substanzen zeigen maximale Änderungen 0.5 h und 6 h nach der mechanischen Belastung und kehren danach aufgrund von Folgereaktionen in ihren Ausgangszustand zurück. Die andere Gruppe umfasst Succinat, Fumarat und Saccharose. Diese Stoffwechselprodukte verändern sich fortlaufend mit der Zeit. Succinat und Glycerat zeigen 0.5 h, Fumarat 6 h und Malat 24-48 h nach mechanischer Belastung einen Konzentrationsanstieg, welcher im Gegensatz zu Aspartat, Glutamat und Galactinol auf diesem Niveau konstant bleibt.



**Abb. 15** Relative Änderung der mit Zweifaktorieller Varianzanalyse identifizierten endogenen Metaboliten hinsichtlich der Faktoren „Zeit“ und „Mechanische Belastung“. Die schwarzen Rechtecke entsprechen einer relativen Änderung von fünf. Man beachte: Das Signal für Succinat ändert sich um den Faktor neun.

Da die Stoffwechselprodukte Succinat, Glycerat, Fumarat und Malat aus dem Respirationsstoffwechsel stammen beziehungsweise sich von diesem ableiten, kann geschlussfolgert werden, dass eine mechanische Belastung einen großen Einfluss auf den allgemeinen Respirationsstoffwechsel hat. Dies wurde bereits von Aepli und Keller mit Hilfe eines Infrarot-Gasanalysators beobachtet<sup>[145]</sup>. Die Änderungen im Respirationsstoffwechsel und in den abzweigenden Stoffwechselprodukten waren im Kultivar *Afra* stärker ausgeprägt als im Kultivar *Milva*. Diese Beobachtung steht im Einklang mit den Ergebnissen der ICA. Eine höhere Schwarzfleckigkeitsneigung wurde bei *Afra* beobachtet.

Die Betrachtung weiterer Stoffwechselprodukte mit charakteristischem Verlauf ist erforderlich, da viele während der Varianzanalyse aufgrund des festgelegten Signifikanzniveaus nicht erfasst werden. Während der Varianzanalyse wurden zum Teil nur geringfügige Änderungen in den unbelasteten Proben beobachtet, welche nicht direkt mit der mechanischen Belastung in Zusammenhang stehen. Deshalb wurde das Signal der belasteten Proben zum Signal der unbelasteten Proben ins Verhältnis gesetzt. Dieser Quotient wurde auf den Zustand zum Zeitpunkt  $t = 0$  (Anhang Abb. 5) bezogen, mit dem Ziel, die belasteten Proben besser miteinander zu vergleichen. Während dieser Analyse zeigen die reduzierenden Zucker Glucose und Fructose einen charakteristischen Verlauf. Ihre Gehalte nehmen 0.5 h und 48 h nach mechanischer Belastung zu und erreichen anschließend ihr Ausgangsniveau. Dieser mittels GC-TOF-MS detektierte Verlauf konnte unabhängig davon durch enzymatische photometrische Messungen bestätigt werden.

Im Rahmen dieser Versuchsreihe wurden auch starke Änderungen im Gehalt von Aconitat festgestellt. Aconitat akkumuliert 0.5-24 h nach der mechanischen Belastung und nimmt danach wieder ab. Glutamat und Aspartat hingegen weisen den entgegengesetzten Trend auf. Sie nehmen 6-24 h nach der mechanischen Belastung ab und pendeln sich danach auf ihr Ausgangsniveau ein. Dieses Verhalten kann auch für andere Stoffwechselprodukte beobachtet



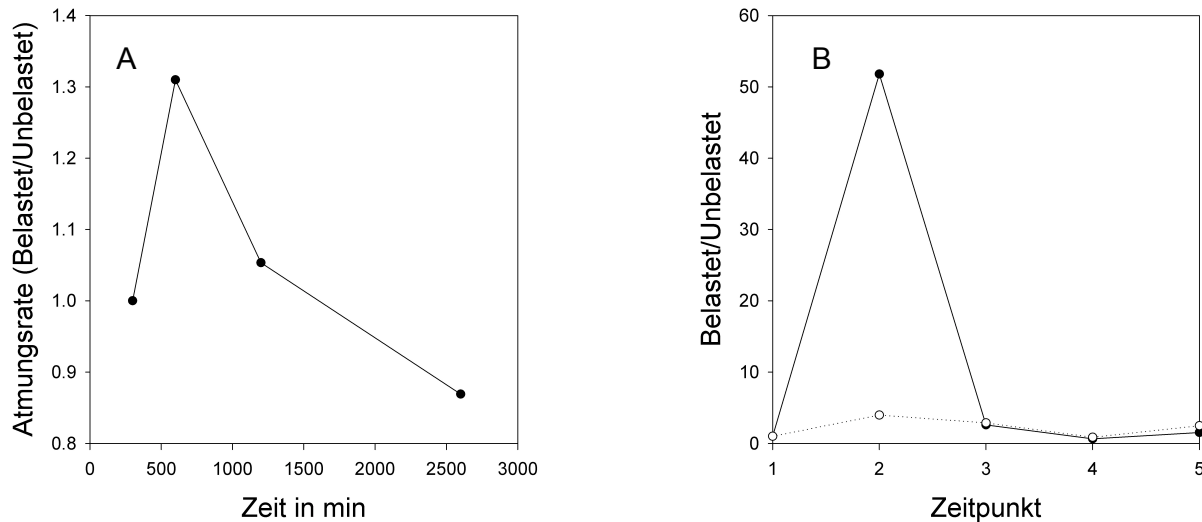
werden, welche sich vom Respirationsstoffwechsel ableiten. Lactat hingegen akkumuliert 24 h bis 48 h nach der mechanischen Belastung.

Im Gegensatz dazu ändern sich die Stoffwechselprodukte glykolytischen Ursprungs, wie Glycin, Serin, Leucin, Isoleucin, Valin und aromatische Aminosäurestoffwechselwege, wie zum Beispiel der von Tyrosin und Phenylalanin kaum. Die geringen Änderungen im Phenylpropanoidstoffwechsel stehen im Widerspruch zu den in der Literatur beschriebenen Beobachtungen und können nur damit begründet werden, dass die Knolle mehrmals belastet wurde<sup>[140, 146]</sup>. Starke mechanische Belastungen führen zu Veränderungen im Respirationsstoffwechsel. Der Phenylpropanoidstoffwechsel verändert sich nach starker mechanischer Belastung hingegen kaum.

Vergleicht man unabhängig davon die zeitliche und stoffliche Zusammensetzung der beiden Kultivare miteinander, so zeigen sich hier zum Teil deutliche Unterschiede. Das Kultivar *Afra* weist nach einer mechanischen Belastung die stärksten metabolischen Änderungen auf. Die Konzentration an Succinat und Fumarat ändert sich im Kultivar *Afra* im Vergleich zum Kultivar *Milva* stärker. *Milva* hingegen akkumuliert mehr Aconitat. Des Weiteren wurde für das Kultivar *Afra* eine stärkere Abnahme von Stoffwechselprodukten, welche sich vom Respirationsstoffwechsel ableiten, wie zum Beispiel Aspartat, Methionin und beta-Alanin beobachtet.

Die Konzentrationsabhängigkeit der nach einer mechanischen Belastung der Kartoffelknollen untersuchten Stoffwechselprodukte lässt den Schluss zu, dass nach der mechanischen Belastung bevorzugt der Respirationsstoffwechsel (Abb. 16A) und die Proteinbiosynthese (Abb. 16B) aktiv sind. Diese Parameter sind nach einer mechanischen Belastung unmittelbar erhöht, da die Knolle auf die mechanische Belastung reagiert. Nach einer bestimmten Zeit wird der Einfluss dieser Parameter geringer und gelangt in den Ausgangszustand zurück. Dieser geschilderte Verlauf wird durch die metabolischen Studien bestätigt. So konnte für das Kultivar *Afra* im Vergleich zu den unbehandelten Knollen ein 1,3-fach höherer Sauerstoffverbrauch in den behandelten Knollen nachgewiesen werden. Die Atmungsrate glich sich 100 h nach mechanischer Belastung bei den behandelten Knollen auf den Wert der unbehandelten Knollen ( $15 \text{ mg kg}^{-1} \text{ h}^{-1}$ ) an. Die späte Anpassung der Atmungsrate ist auf die nichtoptimale Messgeometrie der Sauerstoffmessung zurückzuführen (Kapitel B.2.2). Das Ziel dieser Arbeit bestand jedoch in der Verifizierung der metabolischen Untersuchungen, dies wurde mit dieser Messreihe erreicht. Daher wurde die Messgeometrie nicht weiter optimiert.

Der Verlauf der Atmungsrate konnte für das Kultivar *Milva* im Rahmen eines Folgeexperimentes nicht bestätigt werden, da diese Knollen sich innerhalb der vorgegebenen Zeit nicht verfärbten (Abb. 16A). Die Ursache dafür liegt darin, dass verschiedene Faktoren die Schwarzfleckigkeitsneigung beeinflussen<sup>[147-149]</sup>. Der TCA-Cyclus im Kultivar *Afra* war im Vergleich zum Kultivar *Milva* nach einer mechanischen Belastung erneut deutlich erhöht. Die abzweigenden Stoffwechselprodukte zeigten den erwarteten Trend.



**Abb. 16** Relative Änderungen A) in der Respirationsrate und B) im Proteingehalt nach mechanischer Belastung für das Kultivar *Afra* und das Kultivar *Milva*. *Afra* (schwarz); *Milva* (offene Kreise).

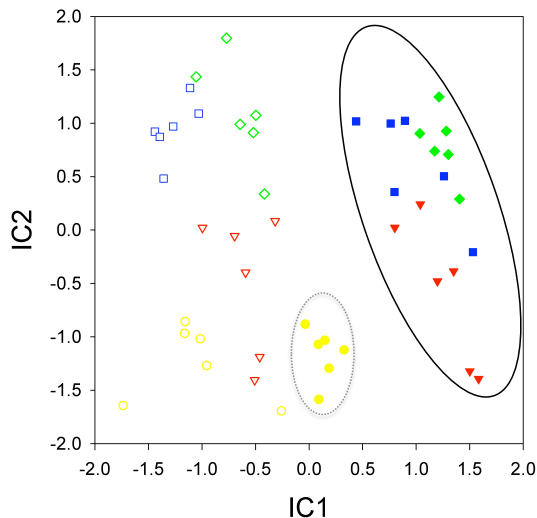
Im Kultivar *Milva* konnten im Rahmen dieses Experimentes keine Änderungen im TCA Cyclus sowie in den abzweigenden Stoffwechselwegen des TCA Cyclus beobachtet werden, da die mechanische Belastung keinen Einfluss auf den Phänotyp hatte.

Auch wenn die untersuchten Knollen der Kultivare *Afra* und *Milva* bereits über mehrere Monate gelagert wurden und im Gegensatz zu den vorangegangenen Versuchsreihen einen unterschiedlichen physiologischen Zustand aufwiesen, so bestätigt dieses Experiment eindeutig, dass die Beobachtungen zur Schwarzfleckigkeit verallgemeinert werden können, solange eine mechanische Belastung zu Verfärbungen der Kartoffelknolle führt.

### C.2.1.2 Lagerung von Kartoffelknollen

Nach der Ernte werden Kartoffelknollen in der Regel bei 4 °C im Dunkeln gelagert. In dieser Zeit sind sie zwar keinem ökologischem Stress ausgesetzt, ändern jedoch ihren Phänotyp maßgeblich. Auslöser sind qualitative und quantitative Veränderungen im Metabolom, Proteom und Transkriptom<sup>[150, 151]</sup>. Während der Lagerung verlieren die Kartoffelknollen Frischmasse und beginnen nach dem Durchlaufen der Dormanz zu keimen. Bei Anwesenheit einer ausreichenden Anzahl an Keimen werden diese erneut ausgelegt und bilden Tochterknollen aus. Dieser Zyklus, der in der Literatur auch als Physiologisches Alter beschrieben ist, wurde bereits intensiv untersucht<sup>[152-156]</sup>. Der Zusammenhang zwischen dem Phänotyp und den Stoffwechselprodukten, die mit konventionellen analytischen Verfahren identifizierbar und quantifizierbar sind geben Rückschluss auf die Qualität der gelagerten Kartoffelknollen. Sie sind daher gegenwärtig von besonderem wissenschaftlichen Interesse. Hierbei wird das Ziel verfolgt, anhand der identifizierten Stoffwechselprodukte einfache und robuste Schnelltests für die

Züchter von Kulturpflanzen zu entwickeln. Allerdings sind bisher noch keine glaubwürdigen Stoffwechselprodukte identifiziert worden, die es Züchtern ermöglichen, die Qualität von beispielsweise gelagerten Knollen mittels einfacher analytischer Verfahren abzuschätzen. Daher wird im Rahmen dieser Arbeit anhand von analytisch-identifizierbaren Produkten gezeigt, welche Stoffwechselwege während der Lagerung aktiv sind. Der Einfluss von Temperaturänderungen während der Lagerung und die Verwendung von Keimhemmern auf das Physiologische Alter der Kartoffelknollen werden untersucht. Da für einen Züchter die Phase der Dormanz entscheidend ist, wurden vier niedrig bis moderat keimende Sorten (*Arnika*, *Cilena*, *Marabel*, *Turdus*) während einer gängigen Lagerperiode auf ihre stoffliche Zusammensetzung untersucht. Die Dormanz ist von besonderer Bedeutung für die Verarbeitung von Kartoffelknollen. Keimende Knollen lassen sich nur sehr schwer weiterverarbeiten. Die ICA zeigt für das Experiment „Physiologisches Alter“, dass sich der metabolische Zustand unmittelbar vor der Einlagerung (offene Symbole) und der metabolische Zustand kurz vor der Auslagerung (gefüllte Symbole) unterscheiden (Abb. 17). Dieser Unterschied wurde innerhalb der ersten Hauptkomponente (IC1) beobachtet und war bei den Kultivaren *Cilena*, *Marabel* und *Turdus* stärker ausgeprägt (durchgezogene Ellipse) als beim Kultivar *Arnika* (gestrichelte Ellipse).



**Abb. 17 Unabhängige Analyse der Hauptkomponenten in Hinblick auf das Physiologische Alter für die Lagersaison 2006/2007 angewendet auf alle detektierten gaschromatographischen Signale.** Die größte Auftrennung wurde für die erste (IC1) und zweite Hauptkomponente (IC2) erreicht. Die offenen Symbole entsprechen den Proben unmittelbar vor der Einlagerung und die gefüllten Symbole den Proben kurz vor der Auslagerung. Die untersuchten Kultivare sind durch unterschiedliche Farben und Symbole hervorgehoben: *Arnika* (Punkt, gelb), *Cilena* (Dreieck, rot), *Marabel* (Rechteck, blau), *Turdus* (Diamant, grün).

Beim Kultivar *Arnika* war der metabolische Unterschied zwischen den beiden Probennahmezeitpunkten deutlich geringer. Dies zeigt, dass der metabolische Zustand der

Sorte *Arnika* geringeren stofflichen Änderungen unterworfen ist. Die zweite Hauptkomponente (IC2) hingegen trennte die Sorten *Arnika* und *Cilena* von den Sorten *Marabel* und *Turdus*. Die nachfolgende Saison (2007/2008) zeigte ein ähnliches Verhalten und bestätigt, dass die Erkenntnisse dieser Versuchsreihe verallgemeinert werden können.

Um die Stoffwechselprodukte zu identifizieren, welche sich während der Lagerung sowohl in Abhängigkeit von der Zeit als auch in Abhängigkeit von der Sorte ändern, wurde eine Zweifaktorielle Varianzanalyse mit dem Signifikanzniveau von  $\alpha = 0.05$  im Hinblick auf die Faktoren Zeit und Sorte auf beiden Datensätzen (Physiologisches Alter 2006/2007, Physiologisches Alter 2007/2008) durchgeführt. Hierbei wurde festgestellt, welches Kultivar sich für eine Lagerung schlechter eignet. Im Rahmen dieser statistischen Analysen wurden überwiegend Änderungen im Monosaccharid-, Saccharose- und Malatgehalt aufgrund von einer Umsetzung dieser Substanzen festgestellt. Glucose und Fructose nehmen während der Lagerung zu. Dieser Effekt ist bei der Sorte *Marabel* am stärksten und bei der Sorte *Arnika* am wenigsten ausgeprägt. Daraus kann abgeleitet werden, dass die Sorte *Marabel* nicht lange gelagert werden sollte, da ein erhöhtes Risiko zur Acrylamidbildung bei der Weiterverarbeitung zu Kartoffelchips besteht<sup>[157-159]</sup>. Die Zunahme von Monosacchariden während der Lagerung ist das Ergebnis des kälteinduzierten und des alterninduzierten Versüßens. Dieser Prozess wurde bereits ausgiebig untersucht und sollte während der Lagerung vermieden werden<sup>[160, 161]</sup>. Der Zuckergehalt vergrößerte sich im Rahmen dieser Versuchsreihen lediglich um den Faktor zwei, was als vertretbar zu bewerten ist.

Saccharose und Malat nehmen in den Kultivaren *Arnika* und *Turdus* zu und in den Kultivaren *Cilena* und *Marabel* ab. Diese Änderungen sind verglichen mit der von Glucose und Fructose deutlich geringer. Die geringe Änderung der Konzentration an Saccharose und Malat erschwert die Quantifizierung dieser Substanzen mittels einfacher photometrischer Verfahren. Die Konzentrationsänderung dieser Stoffe wurde daher von Reust et al. erst nach dem Durchlaufen der Dormanz festgestellt und auf die Regulation des Physiologischen Alters zurückgeführt<sup>[162]</sup>. Die Detektion dieser Substanzen anhand von GC-MS bestätigt, dass diese Analysentechnik viel sensitiver als die der Photometrie ist. Dafür kann Vor-Ort-Analytik durch Photometrie einfacher realisiert werden. Weiterhin kann aus diesen Untersuchungen geschlossen werden, dass alle Kultivare für eine einjährige Lagerung geeignet sind und sich ihr Profil in ähnlicher Weise ändert.

Die Anwendung der Einfaktoriellen Varianzanalyse auf den vorliegenden Datensatz zeigt Änderungen im Aminosäurestoffwechsel sowie im Ascorbat-, Caffeat- und Glycerol-3-Phosphatstoffwechsel auf, wenn ausschließlich zeitliche Änderungen betrachtet werden und sortenabhängige Effekte vernachlässigt werden (Abb. 18). Dabei folgen die Änderungen im Aminosäurestoffwechsel keinem allgemeinen Muster. Die Konzentration an Methionin nimmt während der Lagerung ab, während die Konzentration an Prolin zunimmt. Dehydroascorbat nimmt während der Lagerung ebenfalls ab, während Glycerolphosphat und Caffeat zunehmen. Die im Vergleich zu den anderen Stoffwechselprodukten starke Konzentrationsabnahme an Dehydroascorbat und Methionin sowie die starke Konzentrationszunahme an Prolin zeigen, dass der physiologische Zustand einer Knolle durch diese Substanzen bestimmt wird. Die

Abnahme von Methionin aufgrund einer Verstoffwechslung von S-Adenosylmethionin wurde bereits von Mikitzel et al. beobachtet<sup>[163]</sup>. Tudela et al. berichteten von einer starken Abnahme von Ascorbat während der Lagerung<sup>[164]</sup>. Ascorbat und Dehydroascorbat stehen miteinander in einem chemischen Gleichgewicht. Der Einfluss der anderen Substanzen während der Lagerung ist bislang noch nicht diskutiert worden.

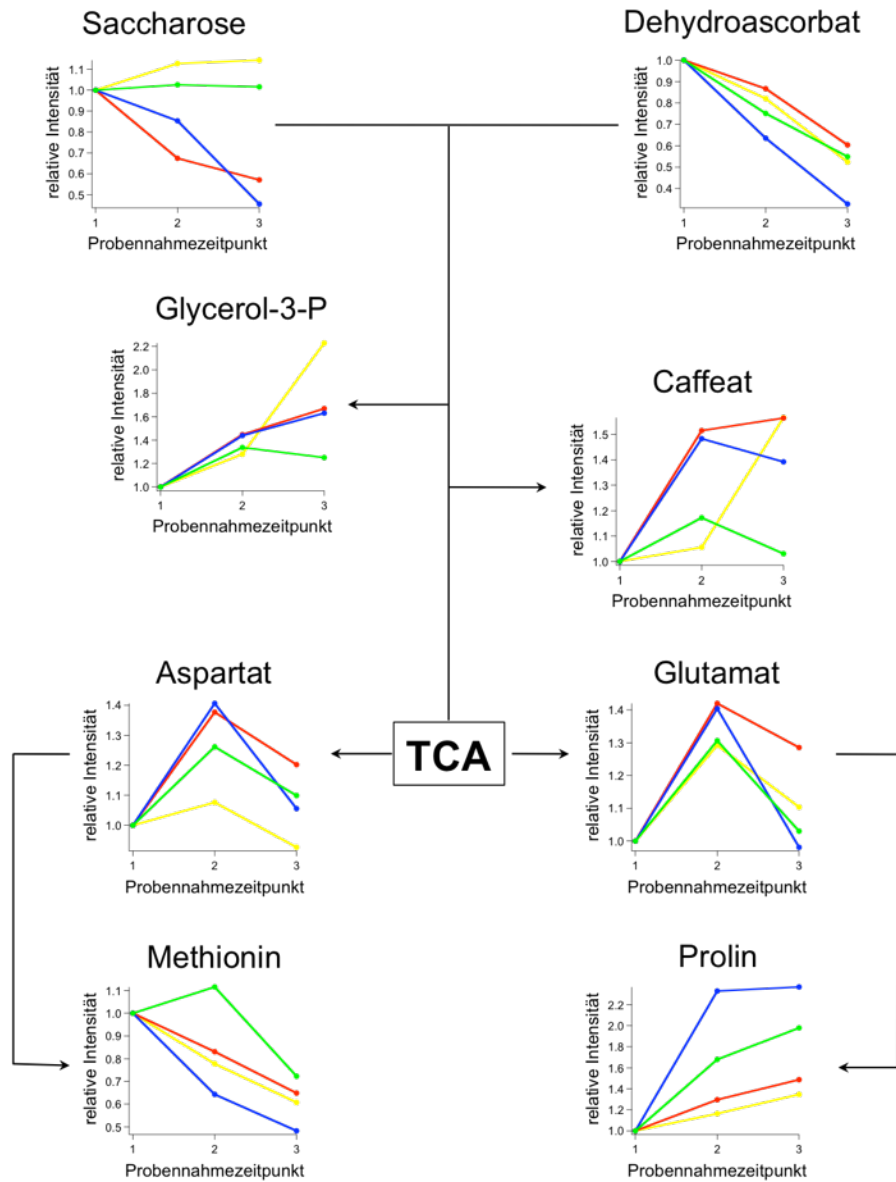


Abb. 18 Relative zeitliche Änderungen im Primär- und Sekundärmetabolismus während der Lagerung in den Kultivaren *Arnika* (gelb), *Cilena* (rot), *Marabel* (blau) und *Turdus* (grün) für die Lagerperiode 2007/2008 nach Anwendung der Einfaktorischen Varianzanalyse bezüglich des Faktors „Zeit“. Die gelagerten Kultivare wurden zu drei Zeitpunkten (November 2007, März 2008, April 2008) entnommen und ihr Primärmetabolismus mittels GC-TOF-MS untersucht.

Die Varianzanalyse ist in der Lage, statistisch relevante metabolische Unterschiede zwischen einzelnen Probengruppen festzustellen. Diese Änderungen können zum Teil nicht ausschlaggebend für den physiologischen Prozess sein und müssen daher aus einer anderen Belastung resultieren. Da die Züchter in Zukunft die Qualität ihrer Kartoffelknollen vor Ort einschätzen wollen, wurden im Folgenden Stoffwechselprodukte betrachtet, welche sich in allen vier Sorten in Relation zu einer bestimmten Eigenschaft gleichermaßen ändern. Dazu wurde der metabolische Datensatz über eine Pearsonkorrelation ( $r$ ) (Gl. 3) mit dem Lagerungsverlust korreliert<sup>[77]</sup>.

$$r = \frac{\sum_{i=1}^I (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^I (x_i - \bar{x})^2 \sum_{i=1}^I (y_i - \bar{y})^2}} \quad (3)$$

- $x_i$  Lagerungsverlust beziehungsweise Keimung für das entsprechende Kultivar
- $y_i$  Intensität des betrachteten GC-MS Signals
- $I$  Anzahl der Variablenpaare
- $\bar{x}$  Mittelwert des Lagerungsverlustes der Keimung
- $\bar{y}$  Mittelwert des betrachteten GC-MS Signals

Der Lagerungsverlust ist für Züchter ein Qualitätsparameter und setzt sich aus dem Gewichtsverlust, der Keimung, der Trockenfäule und der Nassfäule zusammen. Da diese Lagerungsprozesse im Allgemeinen zu einem unterschiedlichen Phänotyp führen, wurden diese Boniturdaten getrennt betrachtet. Dazu wurde zunächst mittels vier unabhängiger gepaarter  $t$ -Tests festgestellt, welche Bonituren sich über den betrachteten Zeitraum signifikant ändern. Für den Gewichtsverlust ( $P < 0.001$ ) und für die Keimung ( $P < 0.001$ ) wurde ein signifikanter Anstieg beobachtet. Die Parameter Nassfäule und Trockenfäule hingegen änderten sich während der betrachteten Lagerperiode kaum. Deshalb wurde eine Pearson Korrelation auf den Lagerungsverlust und die Keimung angewendet (Tab. 5). Die Parameter Nassfäule und Trockenfäule wurden hierbei nicht betrachtet. Ein linearer Zusammenhang existiert zwischen den alpha-Aminosäuren Glutamin, Glycin, Methionin sowie der gamma-Aminosäure GABA und dem Gewichtsverlust. Weiterhin wurde ein linearer Zusammenhang zwischen der Keimungsrate und dem Einfachzucker Glucose beobachtet. Die Aminosäure Methionin weist im Rahmen dieser Korrelationsstudien den niedrigsten Signifikanzwert auf. Die Aminosäuren Glutamin und Glycin besitzen den höchsten Korrelationskoeffizienten. Alle Aminosäuren nehmen mit der Lagerung ab (negativer Korrelationskoeffizient), während der Einfachzucker Glucose mit der Keimungsrate zunimmt (positiver Korrelationskoeffizient).

Aus Tabelle 5 wird ersichtlich das Aminosäuren einen entscheidenden Einfluss auf den Gewichtsverlust haben. Die gefundenen Aminosäuren leiten sich von den Aminosäuren

Glutamat und Aspartat ab und bestätigen zum Teil die Ergebnisse aus der Einfaktoriellen Varianzanalyse. Die Aminosäuren Glutamin, Glycin, Methionin und GABA nehmen mit steigendem Gewichtsverlust ab und sind Vorläuferstufen für Polyamine. Polyamine wurden bereits von Galston und Mikitzel als relevant für das Physiologische Alter diskutiert<sup>[163, 165]</sup>. Die in der vorliegenden Arbeit diskutierten Zusammenhänge im Polyaminstoffwechsel bestätigen die Erkenntnisse im Hinblick auf das Physiologische Alter. Der Einfachzucker Glucose ist für die Keimung entscheidend. Glucose korreliert, wie erwartet, positiv mit der Keimung und liefert der Knolle Energie zur Ausbildung von Keimen<sup>[166]</sup>.

**Tab. 5 Korrelationskoeffizient (R) und Signifikanzwert (p) in Bezug auf den Gewichtsverlust und die Keimung nach Anwendung einer Pearsonkorrelation auf die gemessenen GC-TOF-MS Daten.** Die Zahlen in Klammern geben die Zahl von Trimethylsilyl- (TMS) und Methoxyaminfunktionalitäten (MeOX) an. NP: Nebenprodukt.

Attribut	Detektiertes Stoffwechselprodukt	R	p
Gewichtsverlust	Glutamin (3TMS)	-0.83	$1.29 \cdot 10^{-03}$
Gewichtsverlust	Glycin (3TMS)	-0.83	$2.54 \cdot 10^{-04}$
Gewichtsverlust	Methionin (2TMS)	-0.79	$1.96 \cdot 10^{-06}$
Gewichtsverlust	4-Aminobuttersäure (3TMS)	-0.70	$2.7 \cdot 10^{-03}$
Keimung	Glucose (1MeOX, 5TMS) NP	0.71	$1.29 \cdot 10^{-03}$

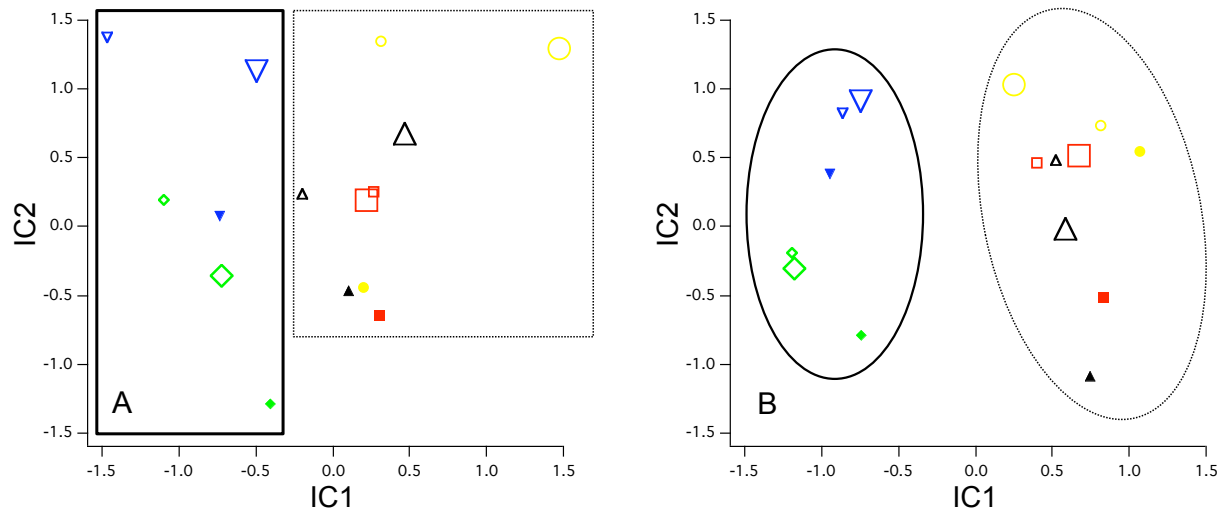
#### C.2.1.2.1 Der Einfluss der Temperatur auf das Lagerungsverhalten

Das Physiologische Alter einer Knolle wird nicht nur durch den Faktor Zeit bestimmt, sondern auch durch die Temperatur<sup>[167]</sup>. Eine höhere Temperatur beschleunigt im Allgemeinen den Alterungsprozess<sup>[160]</sup>. Im Folgenden wird der Einfluss einer Temperaturerhöhung auf die Stoffwechselwege diskutiert. Eine Temperaturerhöhung von 4 °C auf 8 °C während einer Dauer von 4 Wochen vor der Auslagerung und eine anschließende Untersuchung der stofflichen Zusammensetzung mittels GC-TOF-MS zeigte unter Verwendung der Zweifaktoriellen Varianzanalyse bezüglich der Faktoren Temperatur und Sorte keine metabolischen Unterschiede.

Die Stoffwechselprodukte Prolin, Methionin, Citrat und Saccharose wurden dagegen in allen vier Sorten (*Arnika*, *Cilena*, *Marabel*, *Turdus*) in ähnlicher Weise durch die Temperatur beeinflusst. Prolin und Citrat nahmen bei Temperaturerhöhung zu, wohingegen die Konzentration an Saccharose und Methionin mit steigender Temperatur abnahm. Diese Stoffwechselprodukte wurden bereits als essentiell für die Lagerung bei 4 °C erachtet (Kap. C. 2.1.2). Aus diesen Untersuchungen kann geschlossen werden, dass das Physiologische Alter durch den Kohlenhydratgehalt und den Aminosäuregehalt geregelt wird. Ein niedriger Kohlenhydratgehalt und ein hoher Aminosäuregehalt sind typisch für eine physiologisch junge Knolle.

C.2.1.2.2 Der Einfluss von Keimhemmern auf das Lagerungsverhalten

Das Physiologische Alter wird nicht nur durch die Anwendung niedriger Temperaturen verzögert, sondern auch durch den Einsatz von Keimhemmern<sup>[168-172]</sup>. Dabei unterscheidet man zwischen biologischen Keimhemmern, wie zum Beispiel Kümmelöl<sup>[173-177]</sup> und chemischen Keimhemmern, wie zum Beispiel Chlorpropham (Isopropyl-3-chlorcarbanilat)<sup>[178-180]</sup>. Für beide Keimhemmungsvarianten wurde eine übliche Applikation ausgewählt. Zur Untersuchung der physiologischen und metabolischen Unterschiede wurde ein Drittel der Knollen mit Kümmelöl behandelt, ein weiteres Drittel mit Chlorpropham behandelt und der Rest zur Kontrolle unbehandelt bei 4 °C eingelagert.



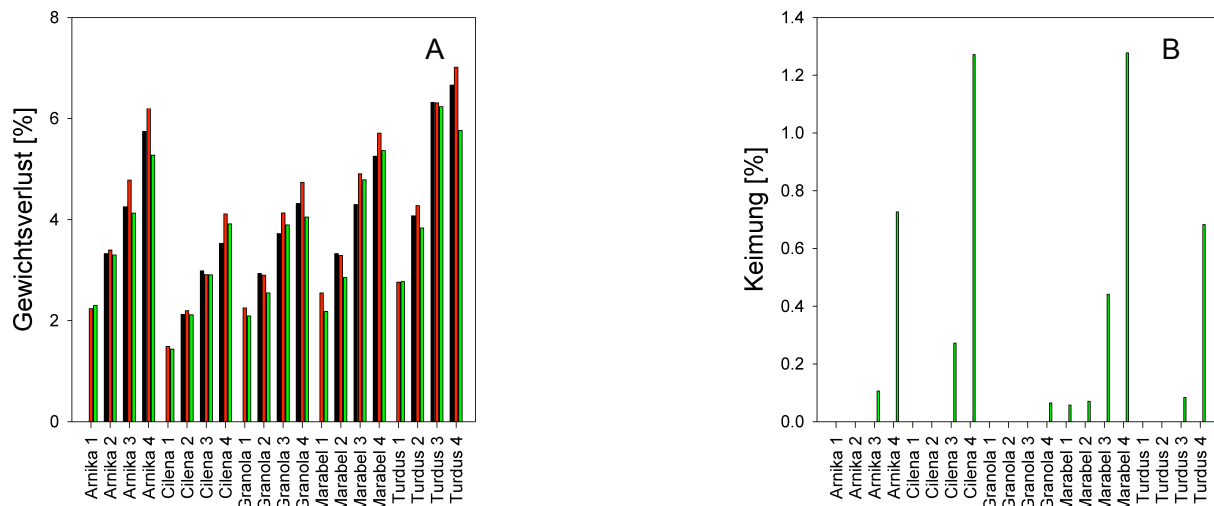
**Abb. 19 Unabhängige Komponentenanalyse (ICA) angewendet auf alle detektierten GC-MS Signale im Hinblick auf die Lagerung von Kartoffelknollen der Kultivare *Arnika* (gelb), *Cilena* (rot), *Granola* (schwarz), *Marabel* (blau) und *Turdus* (grün) nach Applikation von A) Kümmelöl und B) Chlorpropham.** Die erste (IC1) und die zweite (IC2) Hauptkomponente trennten die untersuchten Proben Gruppen. Die Knollen wurden zu drei verschiedenen Zeitpunkten (Dezember 2007, Februar 2008, April 2008) beprobt. Die Größe der Symbole korreliert mit der Zeit.

Die Unabhängige Komponentenanalyse (Abb. 19) zeigt im Hinblick auf die biologische (Abb. 19A) und die chemische Behandlung (Abb. 19B) Unterschiede zwischen dem zweiten (kleine ausgemalte Symbole) und dem dritten (mittlere offene Symbole) Probennahmezeitpunkt innerhalb der zweiten Hauptkomponente (IC2). Dies deutet auf maßgebliche Änderungen im Primärmetabolismus während dieser Phase hin. Weiterhin ist zu erkennen, dass nach einer chemischen Behandlung der Unterschied zwischen den einzelnen Sorten bestehen bleibt (durchgezogene Ellipse in Abb. 19B und gestrichelte Ellipse in Abb. 19B). Hierbei konnten die Kultivare *Marabel* und *Turdus* von den Kultivaren *Arnika*, *Cilena* und *Granola* zu jedem Zeitpunkt der Behandlung innerhalb der ersten Hauptkomponente (IC1) voneinander getrennt werden. Diese Trennung wurde auch bei den unbehandelten Knollen beobachtet. Eine derartige



Trennung war nach einer biologischen Behandlung kaum möglich (Abb. 19A), da der biologische Keimhemmer auf die ausgewählten Kultivare unterschiedlich wirkt.

Die phänotypischen Daten wie der Gewichtsverlust und die Keimung zeigen, dass Keimhemmer das Keimen von Kartoffelknollen verzögern (Abb. 20B). Beide Keimhemmer verhindern allerdings nicht den Gewichtsverlust während der Lagerung (Abb. 20A). Die Analyse der Substanzen Abscisinsäure und Salicylsäure unterstützt die physiologischen Beobachtungen. Salicylsäure nimmt in den unbehandelten Knollen während der Lagerung zu und in den behandelten Knollen nicht. Abscisinsäure hingegen ist zu einem frühen Zeitpunkt der Lagerung in den unbehandelten Knollen besonders hochreguliert und erreicht zum Ende der Lagerung das Konzentrationsniveau der behandelten Knollen. In den behandelten Knollen ist eine deutlich geringere Konzentrationsabnahme zu beobachten (Anhang Abb.6). Diese Ergebnisse stehen in Einklang mit den Beobachtungen von Biemelt et al. und Malkawi et al.<sup>[181, 182]</sup>.



**Abb. 20** Änderungen im A) Gewichtsverlust und B) in der Keimung für die Kultivare *Arnika*, *Cilena*, *Granola*, *Marabel* und *Turdus* nach Behandlung mit: Kümmelöl (schwarz), Chlorpropham (rot) und ohne Behandlung (grün) bei gleicher Lagerungszeit für die Lagerperiode 2007/2008.

Die Zweifaktorielle Varianzanalyse im Hinblick auf die Faktoren Zeit und Behandlung zeigt, dass die Behandlung mit Keimhemmern zu Unterschieden im Monosaccharid-, Polyol-, Fettsäure- und Polyaminstoffwechsel führt (Abb. 21 und Abb. 22). Fructose und Glucose steigen nach der Anwendung von Chlorpropham an, wobei das hohe Konzentrationsniveau über den gesamten Zeitraum der Untersuchung beobachtet wurde. Im Gegensatz dazu nimmt der Gehalt an Fructose und Glucose nach der Behandlung mit Kümmelöl bei den meisten Sorten zu und zwischen dem dritten und vierten Probenahmezeitpunkt wieder ab. Hohe Zuckergehalte sind, wie bereits in Kap. C.2.1.2 beschrieben, für die weiterverarbeitende Kartoffelindustrie nicht erstrebenswert und sollten vermieden werden.

Pyroglutamat und Ornithin, welche ebenfalls mit der Zweifaktoriellen Varianzanalyse erfasst wurden, zeigen im Gegensatz zu Glucose und Fructose geringere Änderungen. Pyroglutamat nimmt nach Applikation beider Keimhemmer zunächst ab und zum Ende der Lagerung wieder zu. Eine Ausnahme stellen Arnika und Turdus dar. In Bezug auf Ornithin konnte kein derartiges Muster festgestellt werden.

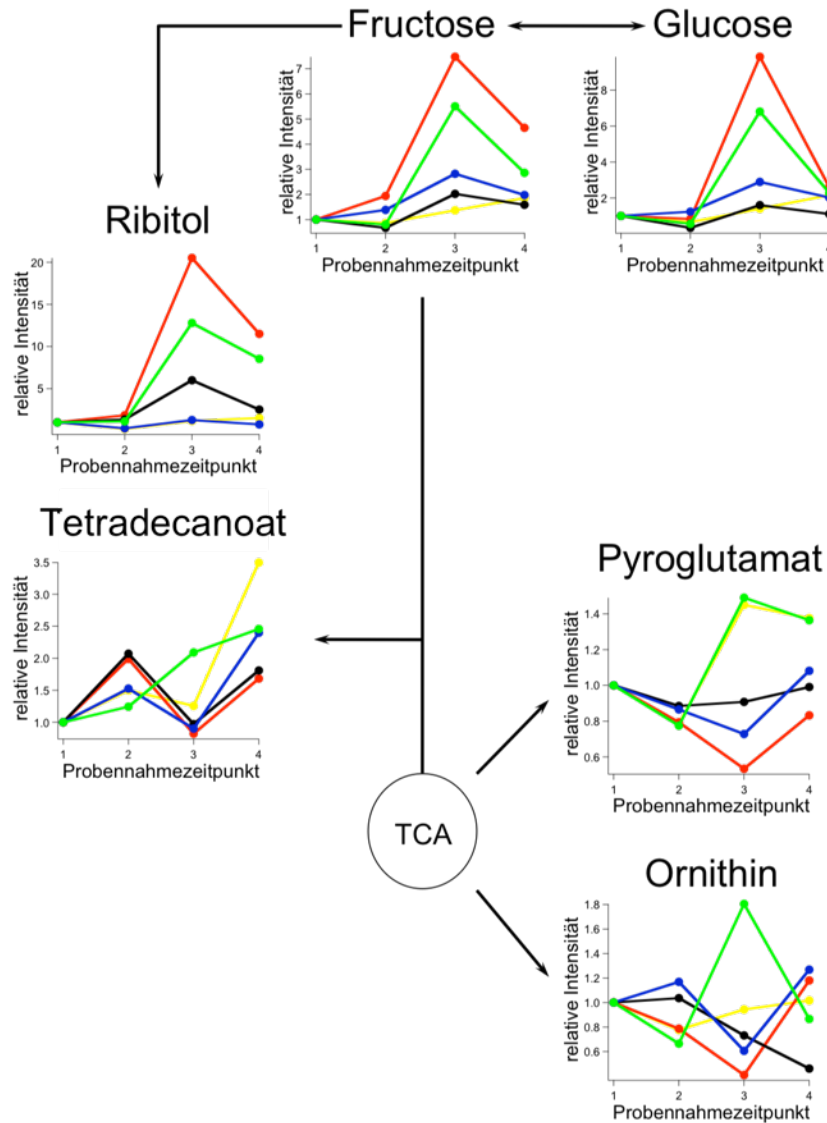
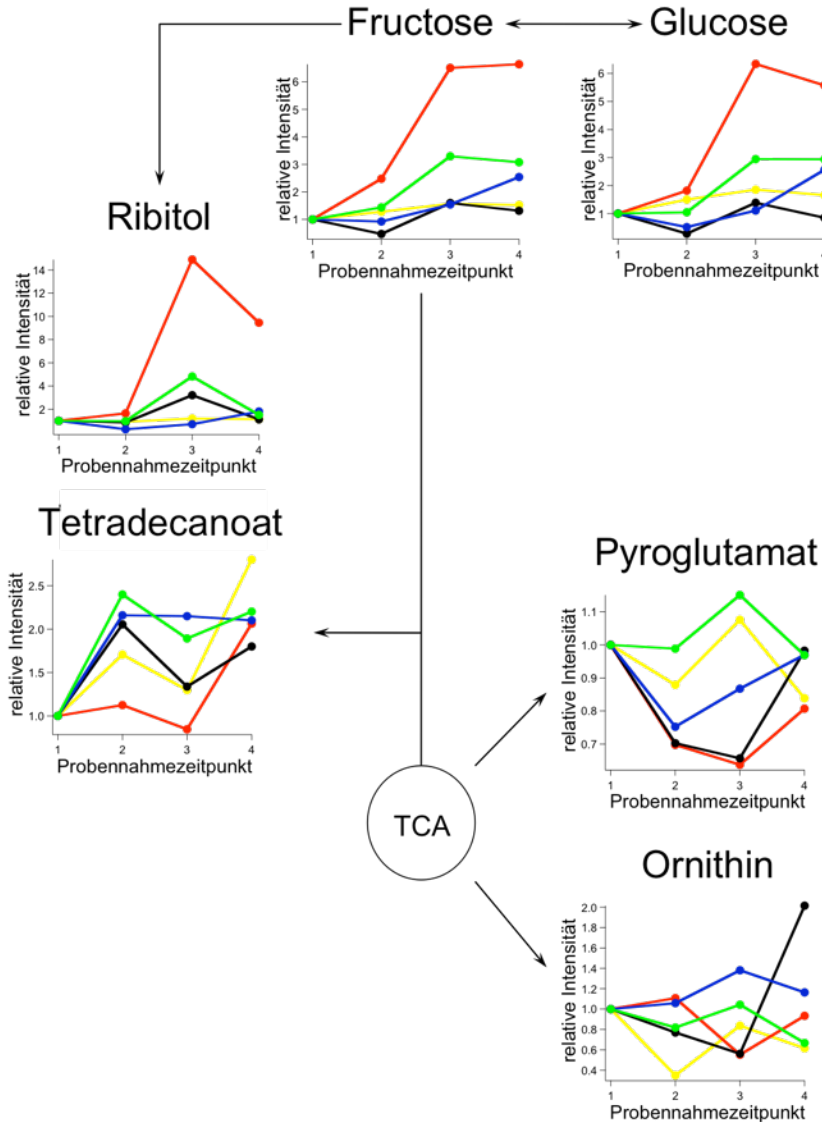


Abb. 21 Relative zeitliche Änderungen im Primärmetabolismus während der Lagerung von Kartoffelknollen der Kultivare *Arnika* (gelb), *Cilena* (rot), *Granola* (schwarz), *Marabel* (blau) und *Turdus* (grün) nach Behandlung mit Kümmelöl. Die Behandelungen wurden auf den entsprechenden Kontrollwert bezogen, so dass das Ausmaß innerhalb des Primärmetabolismus besser verglichen werden kann.



**Abb. 22** Relative zeitliche Änderungen im Primärmetabolismus während der Lagerung von Kartoffelknollen der Kultivare *Arnika* (gelb), *Cilena* (rot), *Granola* (schwarz), *Marabel* (blau) und *Turdus* (grün) nach Behandlung mit Chlorpropham. Die Behandlungen wurden auf den entsprechenden Kontrollwert bezogen, so dass das Ausmaß innerhalb des Primärmetabolismus besser verglichen werden kann.

Anders verhalten sich Ribitol und Tetradecanoat. Der Tetradecanoat-Gehalt fluktuiert in den biologisch behandelten Knollen mit Ausnahme von Arnika und Turdus. In diesen Knollen steigt der Gehalt. In den chemisch behandelten Knollen steigt der Tetradecanoat-Gehalt stetig. Ribitol wurde nach chemischer Behandlung in *Cilena* und nach biologischer Behandlung in *Cilena*, *Granola* und *Turdus* vermehrt nachgewiesen. Die Akkumulation von Polyolen und die differentielle Zunahme von Einfachzuckern und Fettsäuren nach Applikation von Kümmelöl und Chlorpropham wurde bislang noch nicht beschrieben.

Aus diesem Grund wurden für die weiteren Untersuchungen 20 unterschiedliche Kultivare ausgewählt, welche eine sehr niedrige bis mittelhohe Keimfreudigkeit zeigten. Diese wurden mit Kümmelöl behandelt und zum Vergleich unbehandelt eingelagert. Es wurde für diese Untersuchungen lediglich die biologische Keimhemmungsvariante ausgewählt, um zum einen den Probenumfang überschaubar zu halten und zum anderen zu überprüfen, ob die bisherigen Erkenntnisse auf andere Kultivare übertragen werden können. Weiterhin sind die Kartoffelzüchter vielmehr an der Applikation von natürlich vorkommenden Keimhemmern interessiert, da diese nicht kanzerogen sind und außerdem leichter verstoffwechselt werden. Aufgrund der hohen Kultivarzahl wurden die folgenden Untersuchungen an zwei ausgewählten Zeitpunkten (Dezember und März) durchgeführt. Anschließend wurden die metabolischen Unterschiede mittels GC-TOF-MS untersucht. Zwei sortenunabhängige t-Tests im Hinblick auf die beiden Gruppen, biologische Behandlung und keine Behandlung zeigten, dass der biologische Keimhemmer einen großen Einfluss auf den Aminosäurestoffwechsel, den Zuckerstoffwechsel sowie auf die Vorstufen des Sekundärstoffwechsels hat (Tab. 6). Es wurden deutliche Unterschiede im Isoleucin-, Ornithin-, Prolin-, Putrescin- und Tyrosingehalt sowie im Calystegin-, Erythritol-, Nicotinat-, Quinat- und Malatgehalt festgestellt.

**Tab. 6 Signifikanzwert (p-Wert) der beiden sortenunabhängigen t-Tests (biologische Behandlung vs. keine Behandlung) der GC-MS detektierten Stoffwechselprodukte im Hinblick auf die Lagerung biologisch behandelter Knollen und unbehandelter Knollen während der Lagerperiode 2008/2009.** Statistisch relevante Stoffwechselprodukte sind durch einen dicken Schriftzug hervorgehoben. Die Zahlen in Klammern geben die Anzahl an Silylfunktionalitäten im Derivat an.

Stoffwechselprodukte	Dezember	März
Calystegin B2 (4TMS)	0.73	<b>0.04</b>
Erythritol (4TMS)	<b>0.02</b>	0.47
Isoleucin (2TMS)	0.78	<b>6.62 · 10<sup>-03</sup></b>
Malat (3TMS)	0.70	<b>0.01</b>
Methionin (1TMS)	<b>0.03</b>	<b>0.04</b>
Nicotinat (1TMS)	0.46	<b>0.03</b>
Ornithin (3TMS)	0.53	<b>1.66 · 10<sup>-03</sup></b>
Prolin (2TMS)	0.24	<b>6.69 · 10<sup>-03</sup></b>
Putrescin (4TMS)	<b>2.18 · 10<sup>-05</sup></b>	<b>4.94 · 10<sup>-03</sup></b>
Quinat (5TMS)	<b>0.04</b>	0.48
Tyrosin (3TMS)	0.45	<b>8.61 · 10<sup>-03</sup></b>

Erythritol und Quinat waren zum ersten Zeitpunkt der Probenentnahme in Bezug auf die untersuchten Kultivare statistisch signifikant ( $\alpha = 0.05$ ) und zeigten zum zweiten Zeitpunkt der Probenentnahme keine statistisch signifikanten Unterschiede. Diese Substanzen sind daher indikativ für eine frühe zeitliche Veränderung der Lagerung.

Calystegin B2, Isoleucin, Malat, Nicotinat, Ornithin, Prolin und Tyrosin verursachen die höchsten Varianzen in Bezug auf den zweiten Zeitpunkt der Probenentnahme im März. Diese Substanzen reagierten im Vergleich zu Erythritol und Quinat sehr spät auf die biologische Behandlung und sind daher indikativ für eine späte zeitliche Änderung.

Methionin und Putrescin wiesen an beiden Zeitpunkten der Probenentnahme signifikante Unterschiede auf. Sie wurden bereits in der Versuchsreihe „Physiologisches Alter“ als relevant für die Lagerung betrachtet. Auf der Basis von stickstoffhaltigen Substanzen könnten in weiteren Versuchsreihen Schnelltests zur Evaluierung der Lagereignung etabliert werden, die ebenfalls für die Züchter von Interesse sind.

Die in den vorangegangenen Versuchsreihen detektierte Fettsäure Tetradeconoat und das Polyol Ribitol wurden im Rahmen dieser statistischen Analysen nicht detektiert. Die Ursache kann in den Unterschieden der ausgewählten Kultivare zu finden sein. In weiteren Untersuchungen sollte diese Substanzklasse gezielt anhand von weiteren zeitlichen Verläufen beobachtet werden.

### C.2.2 Prädiktive Biomarker

Prädiktive Biomarker sind im Gegensatz zu indikativen Biomarkern in der Lage eine Knolle hinsichtlich eines Phänotyps zu klassifizieren bevor dieser in Erscheinung tritt. Dazu wird der ungerichteten GC-TOF-MS Analyse ein computergestützter Modellierungsprozess nachgeschaltet. Dieser bringt die metabolischen Daten kurz nach der Ernte mit den phänotypischen Daten vier Monate nach der Ernte zusammen. Dabei wird jedem gaschromatographischen Signal ( $S/N \geq 2$ ) anhand der „**Partial Least Squares-Variables Importance in the Projection**“ (PLS-VIP) Methode ein Wert zugeordnet. Dieser gibt die Relevanz zur Vorhersage des untersuchten Phänotyps an. Während dieses Prozesses reduziert die PLS den Datensatz auf seine wesentlichen Bestandteile, während der VIP Wert eine Aussage über die jeweilige PLS Regression ermöglicht [183, 184]. Im Folgenden wurden nur gaschromatographische Signale mit einem VIP Wert größer oder gleich zwei als prädiktiv erachtet. Auf diese Weise konnten Modelle zur Vorhersage der Schwarzfleckigkeit und der Chipseignung aufgestellt werden. Diese wurden auf eine unabhängige Züchtungspopulation übertragen<sup>[129]</sup>. Die programmier- und statistisch relevanten Fragestellungen wurden in Kooperation mit Dr. M. Steinfath bearbeitet.

### C.2.2.1 Beurteilung der Qualität von Kartoffeln nach der Ernte hinsichtlich Schwarzfleckigkeitsneigung und Chipseignung

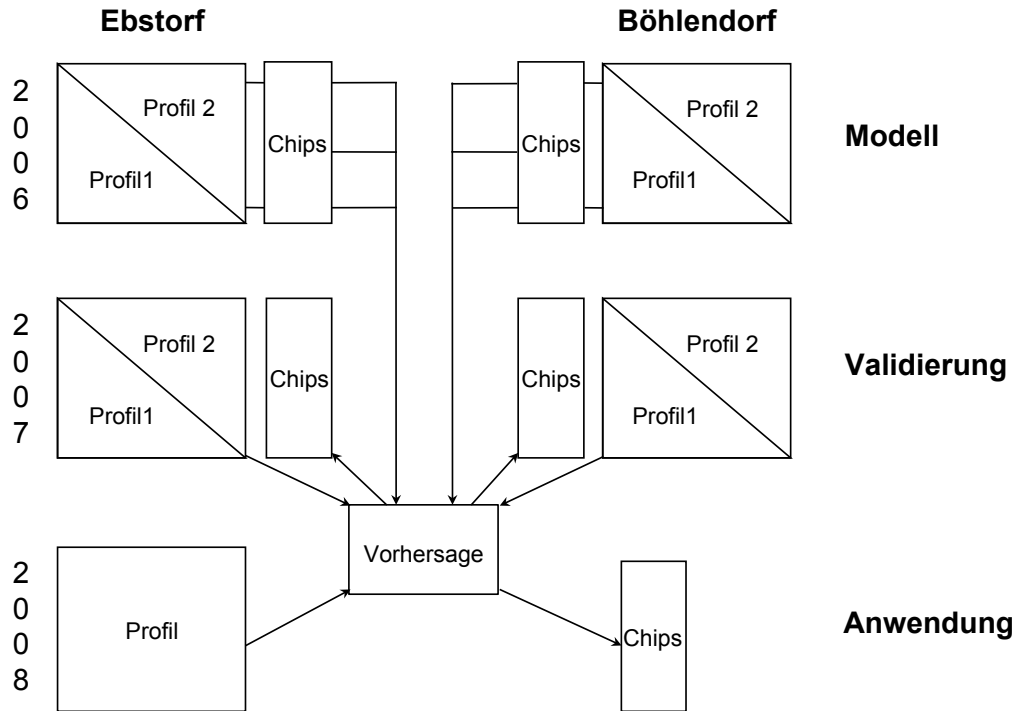
Um möglichst robuste metabolische Biomarker zur Vorhersage genießbarer Kartoffelchips zu finden, wurden 20 Kultivare über zwei Jahre an zwei unterschiedlichen Standorten in Deutschland angebaut (lehmiger Boden an der Küste vs. sandiger Boden im Inland) und die Kartoffeln so ausgewählt, dass es sich weder um Vollgeschwister noch um Eltern- oder Nachkommenskombinationen handelt. Die Grundlage für die Auswahl der Kartoffelsorten war die „Beschreibende Sortenliste“ des Bundessortenamtes<sup>[185]</sup>. Der Probensatz umfasste drei Stärke-, acht Speise-, sechs Chips- und drei Pommes Fritessorten. Durch die breite Auswahl wurde sichergestellt, dass die gefundenen Stoffwechselprodukte später auch auf andere noch nicht untersuchte Sorten übertragen werden können.

Zwei unabhängige Einfaktorielle Varianzanalysen basierend auf 20 Gruppen im Hinblick auf die beiden Standorte Ebstorf und Böhlendorf zeigen, dass der Phänotyp der Chipseignung und der Schwarzfleckigkeitsneigung hinreichend variant sind ( $p < 0.01$ ). Mit Hilfe der ungerichteten GC-TOF-MS Analyse sollten Unterschiede zwischen einzelnen Kultivaren detektierbar sein, welche auf den Phänotyp der Chipseignung und der Schwarzfleckigkeit zurückzuführen sind. Um die analytische Robustheit (relative Standardabweichung  $< 5\%$ ) zu gewährleisten wurde die GC-TOF-MS Analyse in zwei unabhängigen analytischen Läufen durchgeführt<sup>[186]</sup>.

Für die spätere Datenanalyse wurden die beiden aufgenommenen Datensätze nach Standort und Wiederholung getrennt, so dass schließlich vier Datensätze zum Aufspüren prädiktiver Biomarker zur Verfügung standen (Abb. 23). In diesen Datensätzen konnten weder beim Standort (Ebstorf und Böhlendorf) noch bei den analytischen Wiederholungen Unterschiede festgestellt werden. Deshalb wurden alle gaschromatographischen Signale als prädiktiv erachtet, welche in beiden analytischen Replikamessungen und in beiden Standorten einen VIP Wert größer oder gleich zwei besaßen.

Darüber hinaus wurden diese vier Datensätze 20 Permutationen unterworfen, wo jeweils eine Sorte aus dem Datensatz herausgenommen wurde und das Modell zur Vorhersage der Qualität von Kartoffelchips und der Empfindlichkeit gegenüber Schwarzfleckigkeit auf der Basis der verbleibenden 19 Sorten erstellt wurde. Dieses Permutationsverfahren wurde auf jeden Datensatz zwanzigmal angewendet. Hierbei wurde bei jeder Permutation eine Sorte herausgenommen und überprüft, ob die für prädiktiv befundenen Biomarker ebenfalls noch als solche detektiert werden, wenn jeweils eine Sorte aus dem Datensatz entfernt wird. Auf diese Weise konnten Korrelationen aufgespürt und eliminiert werden, die nicht mit der Chipseignung zusammenhängen.

Weiterhin wurde überprüft ob die als potentiell erachteten Substanzen in der darauf folgenden Saison (2007/2008) nach Anwendung aller beschriebenen Schritte als prädiktive Biomarker wiedergefunden werden. Dies wurde trotz der wetterbedingten Unterschiede in den Jahren 2006 (mäßig warm und verregnet) und 2007 (heiß und trocken) durchgeführt<sup>[143]</sup>.



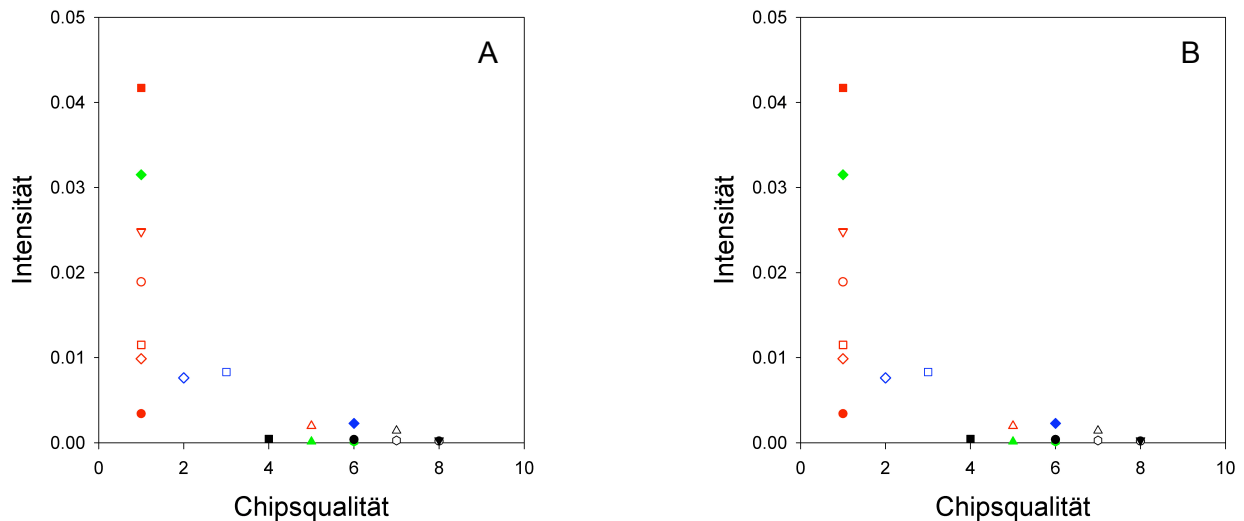
**Abb. 23 Allgemeiner Arbeitsfluss zur Extraktion prädiktiver Biomarker aus GC-TOF-MS Profilen.** Der Arbeitsfluss untergliedert sich in drei Phasen: einer Modellbildungsphase, einer Validierungsphase und einer Anwendungsphase.

Für das Erntegut der Saison 2006/2007 und 2007/2008 konnten geringe Glucose- und Fructosegehalte mit qualitativ hochwertigen Kartoffelchips notiert werden (Abb. 24). Schwarzfleckige Knollen werden durch geringe Tyrosin-, Glutamin- und Threoninegehalte sowie hohe Serin- und Valingehalte nach der Ernte hervorgerufen (Anhang Abb.7). Für die Modellbildungsphase wurde jeweils das Methoxyaminierungsnebenprodukt und das Hauptprodukt der Silylierung gewählt, da hier eine Überladung des Gaschromatographen ausgeschlossen werden kann und nur hinreichend stabile Derivate für die Analyse verwendet werden. Geringe Glucose- und Fructosegehalte wurden bereits als essentiell für die Ausprägung genießbarer Kartoffelchips beschrieben und bestätigen die Machbarkeit dieses Verfahrens <sup>[187],[188]</sup>. Sie nehmen eine zentrale Rolle in der Maillard-Reaktion ein. Hierbei werden Aminosäuren mit reduzierenden Einfachzuckern unter Hitzeeinwirkung zu kanzerogenen Verbindungen umgesetzt.

Die detektierten Aminosäuren stehen bezüglich der Schwarzfleckigkeit in keinem direkten Zusammenhang mit den Vorläuferstufen von Melaninen. Sie sind jedoch eng mit dem TCA Cyclus verknüpft und bestätigen die Ergebnisse aus Kap. C.2.1.1.

Einige Kultivare zeigten trotz geringer Glucose- und Fructosegehalte eine geringe Chipsqualität. Ähnliches wurde für den Prozess der Schwarzfleckigkeitsneigung beobachtet. Die Mehrheit der untersuchten Kultivare folgte allerdings dem beschriebenen Zusammenhang (Abb. 24 und Anhang Abb.7). Dies deutet darauf hin, dass noch weitere Stoffwechselprodukte für die Qualität

von Kartoffelchips und schwarzfleckigen Knollen verantwortlich sind. Diese konnten allerdings im Rahmen der genannten Versuchsreihen nicht erfasst werden. Die Ursache könnte darin zu suchen sein, dass zu wenig Kultivare untersucht wurden. Ein größeres Probenet in dem mehr Kultivare mit unterschiedlicher Chipseignung und Schwarzfleckigkeitsneigung vertreten sind, sollte die Qualität zukünftiger Trainingssets verbessern. Dabei sollte es möglich sein weitere Stoffwechselprodukte zu extrahieren.



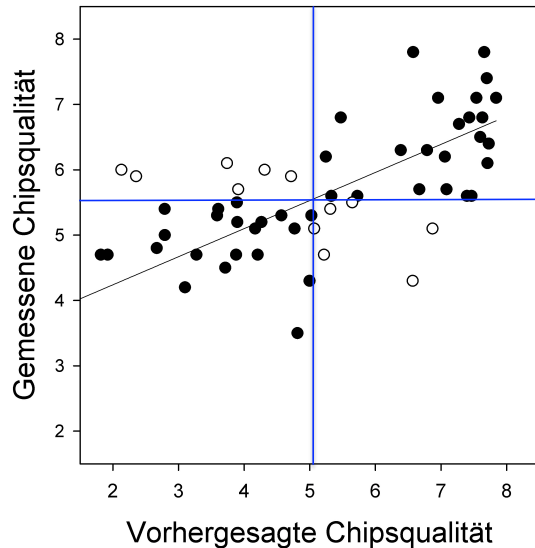
**Abb. 24 Relativer A) Fructose- und B) Glucosegehalt in den 20 untersuchten Kultivaren für den Standort Böhlendorf.** Zur besseren Übersichtlichkeit wurden die Sorten nach ihrer entsprechenden Anwendung eingefärbt: **rot** (Speisekartoffel); **grün** (Pommes Friteskartoffel); **schwarz** (Chipskartoffel); **blau** (Stärkekartoffel). ● *Cilena*; ○ *Elfe*; ▼ *Arnika*; ▼ *Marabel*; ■ *Milva*; □ *Alliance*; ◆ *Granola*; ◇ *Red Fantasy*; ▲ *Agria*; ▲ *Afra*; ● *Topas*; ○ *Lady Claire*; ● *Tomensa*; ◇ *Goldika*; ▼ *Omega*; △ *Roberta*; ■ *Calla*; □ *Turdus*; ◆ *Festien*; ◇ *Kuras*

Zur Überprüfung der Vorhersagemöglichkeit der gefundenen Stoffwechselprodukte wurden 58 Nachkommen einer Züchtungspopulation unter Verwendung des Modells untersucht, da noch nicht überprüft wurde, ob diese Modelle verallgemeinert werden können, oder nur auf die bisher untersuchten Kartoffelsorten anwendbar sind. Während dieser Analyse konnten 23 von 29 Linien im Hinblick auf die Vorhersage genießbarer Chips richtig klassifiziert werden. Ein ähnliches Verhalten zeigen Knollen mit geringerer Chipseignung (Abb. 25). Das etablierte Modell weist eine Güte von 79 % auf. Diese könnte noch erhöht werden, indem man von der F1 Generation die Knollen mit einer hohen Chipsqualität heraussortiert, diese kreuzt und anschließend deren Nachkommen, die F2 Generation, demselben Arbeitsfluss unterwirft.

In Bezug auf die Schwarzfleckigkeitsneigung konnte das Modell nicht auf die neue Züchtungspopulation übertragen werden. Aus diesem Grund wird davon ausgegangen, dass weitere noch nicht bekannte Stoffwechselprodukte die Schwarzfleckigkeitsneigung beeinflussen. Es bleibt zu klären, ob der Grenzwert für die VIP Variablen zu klein angelegt ist,



beziehungsweise das Modell nur auf bestimmte schwarzfleckige Knollen angewendet werden kann.



**Abb. 25 Genauigkeit in der Vorhersage der Qualität von Kartoffelchips für eine Züchtungspopulation.** Die offenen Kreise deuten auf falsch positive und falsch negative Klassifizierungen hin. Diese Proben liegen außerhalb des Medians der gemessenen und der vorhergesagten Chipsqualität (blaue Linien). Die Daten ergeben nach Anwendung einer linearen Regression eine hohe Pearsonkorrelation  $R^2 = 0.64$ .

Dennoch kann aus diesen Untersuchungen geschlussfolgert werden, dass mit Hilfe von PLS-VIP glaubwürdige, robuste, prädiktive Biomarker aus GC-TOF-MS Messungen gewonnen werden können. Daher sollte es möglich sein andere polygenetische Nutzpflanzen, bei denen bislang genetische Biomarkerstudien versagten, in der Zukunft auf bestimmte Phänotypen zu untersuchen. Darüber hinaus ermöglicht dieses Verfahren einen gezielten Eingriff in den Züchtungsprozess. Schließlich können die brauchbaren Knollen unmittelbar nach der Ernte weitergekreuzt werden. Ohne das etablierte Verfahren müssten die Züchter einige Monate warten, um den Phänotyp ihres Ernteguts zu evaluieren.

### C.2.3 Zusammenfassende Beurteilung von Biomarkern in der Matrix Kartoffelknolle

In der vorliegenden Arbeit wurden robuste Stoffwechselprodukte für verschiedene Belastungen während der Lagerung (mechanische Belastung, Lagerung bei unterschiedlichen Temperaturen, Lagerung nach Behandlung mit Kümmelöl und Lagerung nach Behandlung mit Chlorpropham) in der Kulturpflanze Kartoffelknolle anhand von GC-TOF-MS identifiziert.

Hierbei konnte gezeigt werden, dass mechanische Belastungen von Kartoffelknollen zu Änderungen im TCA Cyclus und den abzweigenden Aminosäurestoffwechselwegen führen. Die Änderungen klingen nach einer bestimmten Zeit ab und kehren danach in ihren

Ausgangszustand zurück. Dieser Verlauf geht mit der Entfärbung der belasteten Knollen einher und ist bislang noch nicht beschrieben worden. Die hohe Aktivität des TCA Cyclus und des Respirationsstoffwechsels sind bislang lediglich mittels einfacher analytischer Methoden festgestellt worden. Der metabolische Nachweis mittels GC-TOF-MS wurde noch nicht vorgenommen. Dabei konnte im Rahmen dieser Arbeit gezeigt werden, dass der Phenylpropanoidstoffwechsel nach einer extrem starken mechanischen Belastung nicht aktiv ist und Änderungen im TCA Cyclus sowie den abzweigenden Stoffwechselwegen deutlich überwiegen. Diese Ergebnisse unterstützen die Beobachtungen von Vertregt und stehen im Widerspruch zu denen von Stevens und Corsini<sup>[140, 146, 189]</sup>. Das führt zu der Hypothese, dass Tyrosin wichtig bei der Verfärbung von Knollen ist, jedoch dessen Konzentration allein nicht ausschlaggebend für den Prozess der Schwarzfleckigkeit ist.

Im Gegensatz dazu wird das Physiologische Alter einer Knolle durch den Gehalt von Zuckern und Aminosäuren geregelt. Die Aminosäuren Glutamin, Glycin, Methionin und GABA haben einen großen Einfluss auf den Gewichtsverlust einer Knolle. Die Keimung einer Knolle wird durch die Konzentration des reduzierenden Einfachzuckers Glucose geregelt. Die im Rahmen dieser Arbeit durchgeführten Untersuchungen bestätigen, dass Polymine maßgeblich an der Regulation des Physiologischen Alters einer Knolle beteiligt sind. Die Erkenntnisse aus der Literatur konnten im Rahmen dieser Arbeit auf andere Kultivare (*Arnika*, *Cilena*, *Marabel* und *Turdus*) übertragen werden. Darüber hinaus konnte gezeigt werden, dass die Änderungen der Vorläuferstufen von Polyaminen vielmehr relevant für das Altern einer Knolle sind und eine genauere Evaluierung des physiologischen Zustands einer Knolle ermöglichen. Dieser Zusammenhang wurde bislang noch nicht in der Literatur diskutiert.

Weiterhin wurde gezeigt, dass neben der Temperatur auch Keimhemmer wie Chlorpropham und Kümmelöl einen entscheidenden Einfluss auf den morphologischen Phänotyp und die inhaltliche Stoffzusammensetzung von gelagerten Kartoffelknollen haben. So weisen chemisch behandelte Knollen im Vergleich zu unbehandelten Knollen einen erhöhten Monosaccharid-, Polyol- und Fettsäuregehalt auf. Im Gegensatz dazu weisen biologisch behandelte Knollen überwiegend Änderungen im Aminosäuregehalt und den Vorläuferstufen des Sekundärstoffwechsels auf. Da der Aminosäuregehalt unter anderem für das Altern einer Knolle relevant ist, kann aus diesen Untersuchungen geschlossen werden, dass eine biologische Keimhemmung den Alterungsprozess einer Knolle verlangsamt. Dies ist für Kartoffelzüchter von außerordentlichem Interesse. Die Applikation von Kümmelöl (biologische Keimhemmung) könnte in Zukunft der von Chlorpropham (chemische Keimhemmung) vorgezogen werden.

Die weiterverarbeitende Kartoffelindustrie muss den Zustand einer Knolle unmittelbar nach der Ernte evaluieren, um entsprechende Maßnahmen einzuleiten. Dies ist nur mit Hilfe von umfangreichen Verfahren wie das der PLS-VIP möglich. Dieses Verfahren wurde bislang noch nicht zur Vorhersage der Qualität von Nutzpflanzen angewandt. Im Rahmen dieser Arbeit konnte erstmalig ein biochemisches Modell zur Vorhersage der qualitätsmindernden Eigenschaften von Schwarzfleckigkeit und der Chipseignung in Bezug auf die Kartoffelknolle aufgestellt werden.

### C.3 Klassifizierung von unbekanntem GC-MS Signalen

Während der Analyse der Kulturpflanze „Kartoffelknolle“ wurden eine Vielzahl von Verbindungen detektiert, denen bislang keine chemische Struktur zugeordnet werden kann. Um diese in der Zukunft zumindest hinsichtlich ihrer chemischen Funktionalität zu klassifizieren, wurden im Rahmen dieser Arbeit drei Strategien (das „Entscheidungsbaumverfahren“, die „gezielte An- und Abreicherung funktionell ähnlicher Substanzen“ sowie die „gepaarte NMR und GC-TOF-MS Analyse“) entwickelt.

#### C.3.1 Entscheidungsbaumverfahren

Das Entscheidungsbaumverfahren ist ein Klassifizierungsverfahren und wurde bislang zur Klassifizierung verschiedenster naturwissenschaftlicher Fragestellungen angewandt<sup>[190-194]</sup>. Dabei standen medizinische und pharmakologische Fragestellungen im Vordergrund. Massenchromatogramme wurden mit diesem Verfahren bislang noch nicht prozessiert, obwohl dieses Verfahren eindeutige Regeln zur Klassifizierung bietet.

Strukturaufklärungsprogramme wie zum Beispiel NIST verfügen zwar über einen Algorithmus zur Klassifizierung von Massenspektren, legen jedoch die Regeln zur Klassifizierung nicht offen und sind nicht auf der Basis von biologisch relevanten Daten entwickelt worden<sup>[63, 65]</sup>. Um diesen entscheidenden Nachteil zu überwinden, wurde im Folgenden in Zusammenarbeit mit Dr. J. Hummel ein Entscheidungsbaumverfahren zur Klassifizierung von biologisch relevanten Massenspektren entwickelt<sup>[86]</sup>. Hierbei wurden die programmiertechnischen Aufgaben von Herrn Hummel übernommen. Im Rahmen dieser Arbeit wurden die chemischen Strukturen ermittelt, in die Datenbank (GMD) importiert und ein Beitrag zur Optimierung der Entscheidungsbäume geleistet. Als Grundlage diente die bis dahin aktualisierte Massenspektrenbibliothek (Kap. C. 1.3). Diese wurde dafür unter Verwendung von stereochemisch korrekten Strukturen erweitert, obwohl die gegenwärtig genutzte gaschromatographische Methode nur anomere und epimere Verbindungen sowie geometrische Isomere anhand des Retentionsindex voneinander unterscheiden kann. Dazu wurde zu jedem mittels GC-MS messbaren Metabolit die chemische Struktur recherchiert und als „mol“ Datei archiviert. Anschließend wurde jede Struktur in kleinste organische Strukturelemente zerlegt, welche durch das Programm Checkmol vorgegeben waren (Kap B.2.6.3)<sup>[195]</sup>. Für den weiteren Prozess wurde vorausgesetzt, dass mindestens 3 % aller Spektren über die entsprechende funktionelle Gruppe verfügen mussten<sup>[86]</sup>. Auf diese Weise wurde sichergestellt, dass für alle funktionellen Gruppen und Strukturelemente genügend Beispiele vorhanden waren, in denen die funktionelle Einheit anwesend war. Für das VAR5 System wurden ebenfalls die Strukturen aller messbaren Derivate mit in der Datenbank abgelegt, solange diese aus dem jeweiligen Massenspektrum geschlossen werden konnten.

Das Entscheidungsbaumverfahren wurde im Folgenden so konzipiert, dass es in zwei Phasen unterteilt werden kann: einer Trainingsphase und einer Klassifizierungsphase. Während der Trainingsphase wurden Entscheidungsbäume für alle Strukturelemente gelernt, welche im Anschluss daran auf die unbekanntem Massenchromatogramme angewendet wurden. Daher

wurden für die Trainingsphase nur solche Massenspektren ausgewählt, die genau unter den gleichen Bedingungen entstanden sind, wie die im Anschluss Ausstehenden zur Klassifizierung.

### C.3.1.1 Trainingsphase

Im Zuge der Trainingsphase wurden alle Massenspektren inklusive Retentionsindex berücksichtigt, welche mit Hilfe von MSTFA und Methoxyaminhydrochlorid erzeugt wurden, natürlichen Ursprungs waren und an eine Metabolit- oder an eine Analytstruktur geknüpft waren. Weiterhin wurde sichergestellt, dass der Metabolit mit dem entsprechenden Analyteneintrag verknüpft war. Auf diese Weise konnten multiple Derivatisierungen leicht nachvollzogen werden. Für die Trainingsphase wurde der Datensatz schließlich in zwei Teile unterteilt: ein Teil, welcher über die entsprechende funktionelle Strukturgruppe verfügte und ein zweiter Teil, in dem die funktionelle Gruppe fehlte. Erst die Kombination aus Massenspektrum und chemischer Struktur ermöglichte die Etablierung der folgenden Entscheidungsbäume (Anhang Abb. 8 und Anhang Tab. 8). Jeder Baum repräsentiert eine funktionelle Gruppe oder ein Strukturelement und beinhaltet Massenfragmente zur Klassifizierung der An- und Abwesenheit dieser (Anhang Abb. 8).

Die Trefferquote ( $r$ ) (Gl. 4) und die Genauigkeit ( $p$ ) (Gl. 5) sowie die abgeleiteten Maße wie der  $F_{pr}$ -Wert (Gl. 6) und der Matthews Korrelationskoeffizient (MCC) (Gl. 7) waren für die weiteren Untersuchungen entscheidend zur Evaluierung der Optimierungsergebnisse.

$$r = \frac{TP}{TP + FN} \quad (4)$$

$$p = \frac{TP}{TP + FP} \quad (5)$$

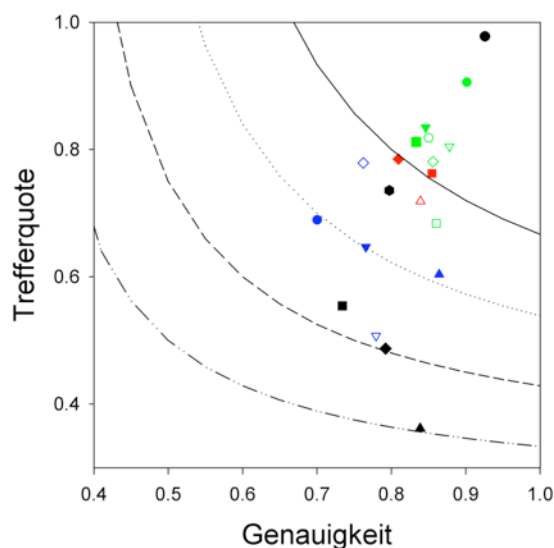
$$F_{pr} = 2 \cdot \frac{p \cdot r}{p + r} \quad (6)$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

TP richtig positive Menge  
 FP falsch positive Menge  
 TN richtig negative Menge  
 FN falsch negative Menge  
 r Trefferquote

p Genauigkeit  
 $F_{pr}$  F-Wert  
 MCC Matthews Korrelationskoeffizient

Im Hinblick auf die Optimierung der Bäume wurde zunächst der Frage nachgegangen, ob Metabolitstrukturen oder die Strukturen der Derivate für die Trainingsphase verwendet werden sollten. Aus analytischer Sicht ist es durchaus nachvollziehbar die Entscheidungsbäume auf Analytstrukturen zu etablieren, da einige Substanzklassen nach der Derivatisierung nicht nur einen unterschiedlichen Retentionsindex aufweisen, sondern auch ein unterschiedliches Massenspektrum. Letztendlich führt das Verknüpfen unterschiedlicher Massenspektren mit einer chemischen Struktur zu Ungenauigkeiten in der Vorhersage der funktionellen Strukturgruppe. Dies kommt besonders bei silylierten Verbindungen zum Vorschein. Im Gegensatz dazu ist es aus biochemischer Sicht durchaus nachvollziehbar ein besonderes Augenmerk auf die Metabolitstrukturen zu legen, da hier ein direkter Bezug zur Biologie besteht.



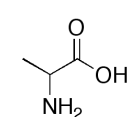
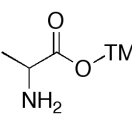
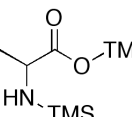
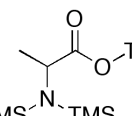
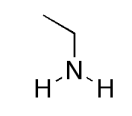
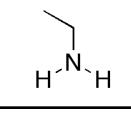
**Abb. 26 Beurteilung der Güte der etablierten Entscheidungsbäume anhand der Trefferquote ( $r$ ), der Genauigkeit ( $p$ ) und des F-Wertes.** Die sauerstoffhaltigen funktionellen Gruppen sind durch grüne Symbole, die stickstoffhaltigen funktionellen Gruppen durch rote Symbole, die carbonylhaltigen Funktionalitäten durch blaue Symbole und der restliche Teil durch schwarze Symbole hervorgehoben. 1,2 Diol ●; Acetal ▼; Alkohol ■; Hydroxy ○; Phenol □; Primärer Alkohol ▽; Sekundärer Alkohol ◇; Amin △; Primäres Amin ◆; Primäres Aliphatisches Amin ■; Aldehyd ●; Carbonsäure ◇; Carbonsäurederivate □; Carbonsäureester ▲; Carbonyl ▼; Keton ▽; Aromat ●; Alken ■; alpha Aminosäure ◆; Heterocyclus ▲; Phosphat ●; Die resultierenden F-Werte sind durch unterschiedliche Strichkonturen hervorgehoben – F = 0.9; ..... F = 0.8; - - - - F = 0.7; ——— F = 0.6.

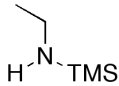
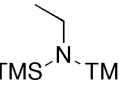
Abbildung 26 zeigt, dass zahlreiche funktionelle Gruppen und Strukturelemente durchaus anhand von Metabolitstrukturen etabliert werden können ( $F_{pr} > 0.6$ ). Wie erwartet erzielen funktionelle Gruppen und Strukturelemente mit einer Derivatisierungsstufe (1,2 Diol, Acetal,

Alkohol, Hydroxy, Phosphat, primärer Alkohol, sekundärer Alkohol) und einem charakteristischem Massenspektrum eine hohe Genauigkeit (Gl. 5). Strukturelemente mit mehreren Derivatisierungsstufen (Amine, alpha Aminosäure) und carbonylhaltige Funktionalitäten liegen zum Teil noch im festgelegten Bereich. Ihre Trefferquote (Gl. 4) und ihre Genauigkeit (Gl. 5) sind jedoch im Vergleich zu den anderen Strukturelementen deutlich geringer. Die Vorhersage von ungesättigten funktionellen Struktureinheiten wie den Einheiten Heterocyclus und Alken ist nach wie vor schwierig, da deren Massenspektren sehr verschieden sind.

Aus diesem Grund wurde an zwei funktionellen Strukturgruppen exemplarisch festgestellt, ob die Derivatstrukturen eine höhere Genauigkeit (Gl. 5) und eine höhere Trefferquote (Gl. 4) erzielen als die dazugehörige Metabolitstruktur, obwohl auf diese Weise die Klassifizierung von carbonylhaltigen Verbindungen und aromatischen Verbindungen nicht wesentlich verbessert wird. Für diese Untersuchungen wurde das Strukturelement alpha Aminosäure sowie die funktionelle Gruppe primäres Amin ausgewählt und der entsprechende Silylierungsgrad soweit erhöht bis das entsprechende Molekül vollständig silyliert war (Tab. 7).

**Tab. 7 Gegenüberstellung wichtiger Klassifizierungsparameter für die Etablierung von Entscheidungsbäumen basierend auf Analytstruktur (AS) und Metabolitstruktur (MS).** rel TP: relatives Vorkommen; p: Genauigkeit; r: Trefferquote;  $F_{pr}$ : F-Wert; MCC: Matthews Korrelationskoeffizient. Die für optimal befundenen funktionellen Gruppen sind fett und mit grüner Schrift hervorgehoben.

	Abkürzung	Struktur	rel TP [%]	p	r	$F_{pr}$	MCC
	AS	MS	51	0.8	0.51	0.62	0.59
	AS (1TMS)	AS	3	0.3	0.03	0.05	0.09
	AS (2TMS)	AS	<b>84</b>	<b>0.77</b>	<b>0.84</b>	<b>0.8</b>	<b>0.78</b>
	AS (3TMS)	AS	34	0.57	0.34	0.43	0.43
	A	MS	74	0.8	0.73	0.77	0.69
	A	AS	6	0.28	0.06	0.1	0.11

	Abkürzung	Struktur	rel TP [%]	p	r	F <sub>pr</sub>	MCC
	A (1TMS)	AS	61	0.87	0.61	0.72	0.7
	A (2TMS)	AS	72	0.99	0.72	0.84	0.84

Gemäß Tabelle 7 erscheinen die Derivatstrukturen A (2TMS) und AS (2TMS) besser für die Vorhersage geeignet zu sein als die Strukturen AS, AS (1TMS), AS (3TMS), A und A (1TMS). So erzielten die Derivatstrukturen A (2TMS) und AS (2TMS) eine deutlich höhere Präzision und Trefferquote als die entsprechenden Metabolitstrukturen (A und AS), wohingegen andere Derivatstrukturen wie AS (1TMS), AS (3TMS) und A (1TMS) eine geringere Trefferquote lieferten. Es bleibt zu klären, ob die Zahl der zur Verfügung stehenden Massenspektren oder das Fehlen charakteristischer Massenfragmente hierfür maßgeblich ist. Der Entscheidungsbaum für die funktionelle Gruppe AS (2TMS) weist die Fragmentmasse 218 und der Baum A (2TMS) die Masse 174 auf (Anhang Abb. 8). Im Gegensatz dazu konnten derartige Fragmentmassen bei den anderen funktionellen Gruppen (AS, AS (1TMS), AS (3TMS), A und A (1TMS)) nicht extrahiert werden.

Während der Trainingsphase ist nicht nur die Wahl der chemischen Struktur entscheidend, sondern auch die Parameter, die dem Baum seine endgültige Gestalt geben. Bislang wurde für das Erstellen von Entscheidungsbäumen ein Abbruchkriterium von 10 Massenspektren gewählt, mit dem Ziel die Bäume nicht zu sehr an die Trainingsdaten anzupassen. Dieses Kriterium ist jedoch nicht hinreichend, um Massenspektren im Detail voneinander zu trennen. Daher wurde untersucht, inwieweit dieses Abbruchkriterium die Leistung der Entscheidungsbäume beeinflusst. Für diese Fragestellung wurden nicht-redundante Spektren gewählt. Darüber hinaus wurde ein Grenzwert von 80 % (Kap. C.1.3) festgelegt. Unter Verwendung nicht-redundanter Spektren sollte ausgeschlossen werden, dass Replikaspektren eines Analyten mehreren Blättern zugeordnet werden. Der F<sub>pr</sub>-Wert (Gl. 6) reduzierte sich während dieser Untersuchungsreihe um 33 % und der MCC (Gl. 7) um 39 %, wenn für das Abbruchkriterium nur ein Massenspektrum zugelassen wird. Werden dieselben Untersuchungen mit dem vorher festgelegten Abbruchkriterium (10 Massenspektren) sowie auf nicht-redundanten Spektren durchgeführt, so verringert sich die Trefferquote noch zusätzlich um den Faktor 1.8 bezüglich des F<sub>pr</sub>-Wertes und um den Faktor 1.6 bezüglich des MCC-Wertes, da die Entscheidungsknoten nicht analytisch rein sind. Insofern kann aus diesen Untersuchungen die Schlussfolgerung gezogen werden, dass für den zur Verfügung stehenden Datensatz redundante Massenspektren gewählt werden sollten. Ein größerer Datensatz mit weiteren Spektren und Strukturen wie beispielsweise dem vom Fiehn Laboratorium würde zeigen, ob das Lernen von Entscheidungsbäumen auf nicht-redundanten Massenspektren einen entscheidenden Vorteil

bringt<sup>[61]</sup>. Für den Fall, dass die Entscheidung auf einen nicht-redundanten Datensatz fällt, sollte das Abbruchkriterium auf ein Massenspektrum herabgesetzt werden.

### C.3.1.2 Klassifizierungsphase

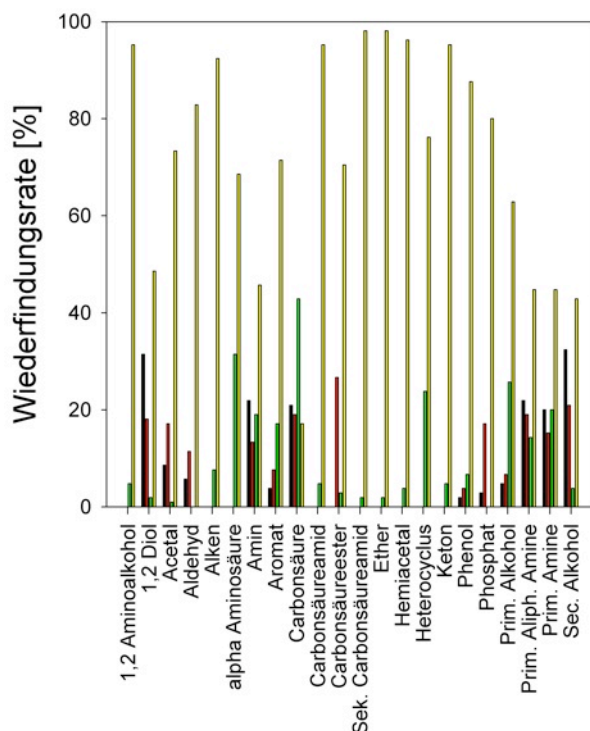
Nach Abschluss der Trainingsphase wurde jedes zu klassifizierende Massenspektrum dekonvolviert und klassifiziert. Für die Evaluierung des etablierten Verfahrens wurden zehn Chromatogramme zweier biologischer Matrizes („Reisblatt“, und „Bäckerhefe“) und ein Chromatogramm einer chemisch definierten Mischung (Anhang Tab. 3) ausgewählt.

Dabei wurden für die synthetische Mischung 105 Massenspektren dekonvolviert und klassifiziert. Hierbei konnte für keine der funktionellen Gruppen ein glaubwürdiger Grenzwert festgelegt werden (Abb. 27). Die funktionellen Gruppen 1,2 Aminoalkohol, Alken, alpha Aminosäure, Carbonsäureamid, Carbonsäureester, sekundäres Carbonsäureamid, Ether, Hemiacetal, Heterocyclus und Keton wurden während der Klassifizierungsphase nie gefunden, obwohl sie eindeutig in der chemischen Struktur vorhanden waren. Eine Ursache hierfür könnten die unterschiedlichen Massenspektren dieser funktionellen Gruppen im Trainingsdatensatz sein. Dadurch konnten in der Trainingsphase für diese funktionellen Gruppen keine optimalen Entscheidungsbäume etabliert werden. Diese Substanzklassen müssten durch weitere kommerzielle Referenzsubstanzen systematisch erweitert werden.

Die funktionellen Gruppen 1,2 Diol, Acetal, Aldehyd, sekundärer Alkohol und Phosphat (Abb. 27) konnten mit einer hinreichenden Güte klassifiziert werden. Die falsch positive Menge, welche zu einer Falschaussage des Anwenders führt, war jeweils kleiner als 5 %. Die richtig negative Menge, in der funktionelle Gruppen aufgrund des noch nicht perfekt etablierten Algorithmus nicht vorhergesagt werden, lag zwischen 10 % und 20 %. Die bessere Vorhersagekraft dieser Strukturelemente verglichen zu den bereits diskutierten kann damit begründet werden, dass diese funktionellen Gruppen ein ähnlicheres Massenspektrum nach Elektronenstoßionisation zur Folge haben.

Darüber hinaus zeigt Abbildung 27, dass die Klassifizierung in Bezug auf die Abwesenheit einer funktionellen Struktureinheit eine bessere Leistung erzielt als die Anwesenheit dieser. Die Wiederfindungsrate für die Abwesenheit einer funktionellen Gruppe war für alle betrachteten Strukturelemente viel höher als für die entsprechende Anwesenheit. Dies kann nur damit begründet werden, dass der Trainingsdatensatz zu unausgewogen ist. Dieser enthält mehr Spektren, bei denen die funktionelle Gruppe fehlt im Vergleich zu Spektren die über die entsprechende funktionelle Gruppe verfügen (Anhang Abb. 8). Die systematische Erweiterung des Referenzdatensatzes mit weiteren Massenchromatogrammen sollte die Qualität des Verfahrens schrittweise verbessern.





**Abb. 27** Evaluierung des Entscheidungsbaumverfahrens nach Anwendung auf eine mittlere chemische Mischung anhand der ■ richtig positiven-, ■ falsch positiven-, ■ richtig negativen- und ■ falsch negativen Menge unter Verwendung von Metabolitstrukturen.

Bei den biologischen Matrizes „Reisblatt“ und „Bäckerhefe“ wurde eine Vielzahl erwarteter funktioneller Gruppen und Strukturelemente in beiden Matrizes wiedergefunden (Tab. 8). Dazu wurden 10 Massenchromatogramme biologisch relevanter Hauptmetabolite (Glucose, Fructose, Threonsäure, Ribitol, Zitronensäure, Äpfelsäure, Serin und Tyrosin) dekonvolviert und anschließend klassifiziert.

Die Strukturelemente Hydroxy, Carbonyl und Aromat wurden mit einer hohen Wahrscheinlichkeit (78 %-94 %) in beiden Matrizes korrekt klassifiziert. Gleiches gilt für Phenol, sekundärer Alkohol und 1,2 Diol. Die funktionellen Gruppen Aldehyd, Keton, primärer Alkohol und Carbonsäure, sowie die derivatisierten Funktionalitäten Amin (2TMS) und alpha Aminosäure (2TMS) wurden zum Teil nicht wiedergefunden beziehungsweise wiesen bei einigen Substanzen eine Vielzahl falsch positiver Treffer auf. Dies könnte damit begründet werden, dass diese funktionellen Gruppen innerhalb des Trainingsdatensatzes sehr unterschiedliche Massenspektren aufweisen, aber auch die klassifizierten Massenspektren Koelutionen beinhalten könnten.

Daher sollten Massenspektren basierend auf Dekonvolutionen nur zur vorläufigen Struktur-erkennung verwendet werden. Zur endgültigen Strukturaufklärung sollten eine Vielzahl von Massenspektren unterschiedlicher biologischer Herkunft gesammelt und nach der Durchschnittsspektrenberechnung dem Entscheidungsbaumverfahren unterzogen werden.

Ebenfalls könnte der entscheidende Peak auf weiteren chromatographischen Systemen mit unterschiedlichen stationären Säulen gemessen werden.

**Tab. 8 Abschätzung der Robustheit des entwickelten Entscheidungsbaumverfahrens anhand der Klassifizierung von Hauptmetaboliten der Matrizes „Reisblatt“ und „Bäckerhefe“.** Die Zahlenwerte geben an mit welcher Wahrscheinlichkeit die entsprechende funktionelle Gruppe im Mittel (10 Chromatogramme) klassifiziert wurde. Eine falsch richtige Klassifizierung ist durch die Abkürzung nd (nicht detektiert) hervorgehoben und eine falsch positive Klassifizierung durch den entsprechenden Zahlenwert in roter Farbe. Die Ergebnisse der Matrix „Reisblatt“ sind durch einen fetten Schriftzug gekennzeichnet und die der Matrix „Bäckerhefe“ durch einen regulären Schriftzug.

	H y d r o x y	C a r b o n y l	A r o m a t	A l d e h y d	K e t o n	S e k A l k o h o l	P r i m A l k o h o l	D i o l	C a r b o n s ä u r e	P h e n o l	A m i n 2 T M S	A m i n o s ä u r e 2 T M S
Glucose (1MeOX, 5TMS) HP	<b>96</b> 98	<b>83</b> 83		<b>53</b> nd		<b>82</b> 84		<b>99</b> 99	<b>51</b> <b>51</b>			
Fructose (1MeOX, 5TMS) HP	<b>99</b> 99	<b>83</b> 82		<b>51</b> nd	<b>nd</b> nd	<b>82</b> 85	<b>nd</b> nd	<b>99</b> 99			<b>80</b>	
Threonsäure (4TMS)	<b>99</b> 99	<b>83</b> 83		<b>51</b>		<b>82</b> 82	<b>nd</b> nd	<b>99</b> 99	<b>nd</b> nd		<b>80</b>	
Ribitol (5TMS)	<b>99</b> 99	<b>83</b> <b>83</b>				<b>83</b> 82	<b>nd</b> nd	<b>99</b> 98				
Äpfelsäure (3TMS)	<b>99</b> 99	<b>nd</b> 82				<b>76</b> 82		<b>99</b> 99	<b>51</b> nd	<b>89</b>		
Zitronensäure (4TMS)	<b>99</b> 99	<b>nd</b> 78				<b>62</b> <b>82</b>		<b>nd</b> 98	<b>51</b> 51		<b>80</b>	
Serin (3TMS)	<b>86</b> 81	<b>83</b> 82					<b>nd</b> nd		<b>nd</b> nd			<b>nd</b> nd
Tyrosin (3TMS)	<b>91</b> 91	<b>82</b> 82	<b>94</b> 94						<b>nd</b> nd	<b>93</b> 93	<b>80</b> 80	<b>72</b> nd

Weiterhin sind umfangreiche Trenntechniken, wie die GCxGC-TOF-MS oder Aufarbeitungsverfahren, wie das der Festphasenextraktion denkbar um Koelutionen zu vermeiden. Damit in

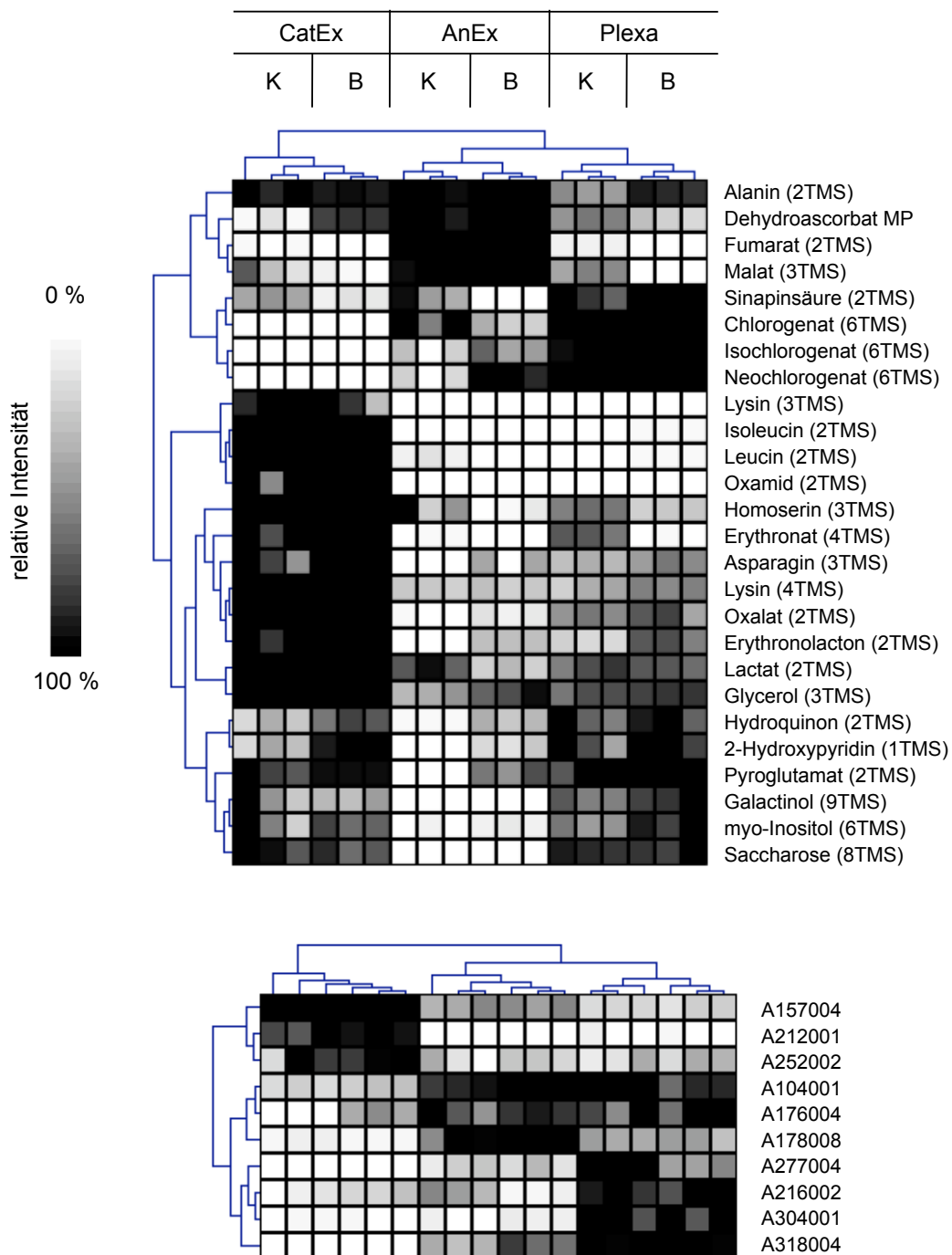
Zukunft Konnektivitäten zwischen einzelnen Atomen erkennbar sind, müssen neue funktionelle Untereinheiten definiert werden, die mehrere funktionelle Gruppen zusammenfassen und nach der Ionisation im Massenspektrometer ähnliche Massenspektren ergeben.

### C.3.2 Gezielte Anreicherung mittels Festphasenextraktion

Die Festphasenextraktion wird nicht nur angewendet, um ausgewählte Moleküle gezielt anzureichern, sondern auch um eine Matrix im Hinblick auf ihre Komplexität zu reduzieren und ausgewählte Zielmoleküle aus einer komplexen Matrix besser zu quantifizieren (Kap. C.1.6). Die Anwendung der Festphasenextraktion zur Klassifizierung unbekannter gaschromatographischer Signale ist bislang in der Literatur noch nicht beschrieben und stellt somit eine Möglichkeit dar gaschromatographische Signale zu klassifizieren. Festphasenextraktionskartuschen mit einer ähnlichen stationären Phase reichern Verbindungen ähnlicher chemischer Funktionalität an und ermöglichen eine Aussage zur Funktionalität der gaschromatographischen Signale. Um dies zu überprüfen und zu verifizieren inwieweit die Festphasenextraktion bei der Klassifizierung von unbekanntem GC-MS Signalen hilfreich ist, wurden drei SPE Kartuschen unterschiedlicher Oberflächenbelegung untersucht. Dazu wurden methanolische Extrakte aus Kartoffelknollen und Brokkoliröschen entsprechend Kap. B.2.3.2.1 hergestellt, an einer Kationenaustauscher-, einer Anionenaustauscher- und einer C18 ähnlichen Kartusche angereichert und im Anschluss mittels GC-TOF-MS untersucht.

Für diese Untersuchungen wurden bewusst zwei unterschiedliche Matrices verwendet, da die gefundenen Zusammenhänge anschließend eine bessere Verifizierung ermöglichen, obwohl die einzelnen Substanzen in beiden Matrices in unterschiedlichen Konzentrationen vorkommen. Das Ziel dieses Versuches bestand in der Auswahl von geeigneten Kartuschen für die jeweiligen Substanzklassen, der selektiven Anreicherung dieser Substanzklassen, der Reduzierung der Komplexität des Extraktes und der gaschromatographischen Analyse der polaren Metaboliten (Abb. 28).

Abbildung 28 zeigt die Anreicherung von ungesättigten und sauerstoffhaltigen Verbindungen an einer Plexakartusche. Die Plexakartusche retardierte sämtliche Derivate der Chinasäure, aromatische Verbindungen, cyclische Polyole und Disaccharide, da diese Kartusche entsprechend der Herstellerangaben die Analyten nach einem Umkehrphasenmechanismus retardierte. Umkehrphasensorbentien wurden bislang erfolgreich zur Analyse von Pestiziden aus Trinkwasser verwendet<sup>[196-201]</sup>. Im Vergleich zur Plexakartusche retardierte die Kationenaustauscherkartusche mehr Analyten. Die Kationenaustauscherkartusche zeigt die höchste Anreicherungs-effizienz von allen drei Kartuschen. Sie reichert bevorzugt Aminosäuren, Säureamide, Polyole mit einer Säurefunktionalität und stickstoffhaltige Ringsysteme an. Im Gegensatz dazu reichert die Anionenaustauscherkartusche beide Matrices am effizientesten ab. Nach der Applikation des Anionenaustauschers konnten nur wenige oxidierte Verbindungen, wie das Dehydroascorbat und Dicarbonsäuren gut wiedergefunden werden.



**Abb. 28** Identifizierte Analyten und erwartete Unbekannte aus repräsentativen Extrakten von Kartoffelknolle (K) und Brokkoliröschen (B) nach Anreicherung an einer CatEx (Kationenaustauscher), einer AnEx (Anionenaustauscher) und Plexa Kartusche(C18). Dargestellt sind diejenigen Substanzen, die mit Hilfe von einer Einfaktoriellen Varianzanalyse mit einem Signifikanzwert von  $\alpha = 0.001$  unter der Annahme von 3 Gruppen (CatEx, AnEx, Plexa) erfasst wurden. Die relative Signalintensität nimmt von weiß nach schwarz zu. (weiß: keine Signalintensität, grau: mittlere Signalintensität, schwarz: maximale Signalintensität)

Wird die Analyse auf erwartete unbekanntem Massensignale beschränkt, so zeigt sich, dass die Verbindungen A157004, A212001 und A252002 überwiegend auf einer Kationenaustauscherkartusche retardieren, während die Verbindungen A104001, A176004 und A178008 auf einer Anionenaustauscherkartusche sowie die Verbindungen A277004, A216002, A304001 und A318004 auf einer Umkehrphasenkartusche retardieren (Abb. 28). Dies ist auf ähnliche funktionelle Gruppen wie bei den bereits identifizierten Verbindungen zurückzuführen. Verbindung A157004 konnte einer Dichlorbenzoesäure zugeordnet werden, Verbindung A304001 einem Trisaccharid bestehend aus zwei Glucoseeinheiten und einer Fructoseeinheit, Verbindung A212001 einer Oxogluconsäure, Verbindung A 277004 einem Adenosinthioderivat und Verbindung A318004 einem Deoxyguanosinphosphat.

Während der Festphasenextraktion wurden nicht nur Metaboliten pflanzlicher Proben sondern ebenfalls Verunreinigungen angereichert. Verbindung A104001 ist auf ein Siloxan zurückzuführen.

### C.3.3 Gepaarte GC-MS und NMR Analyse

Eine chemische Struktur kann nur durch die Anwendung mehrerer analytischer Methoden aufgeklärt werden. So wird beispielsweise die Analysenmethode der NMR Spektroskopie in der Synthesechemie angewendet, um den Ablauf einer chemischen Reaktion zu dokumentieren. Die Anwendung von NMR Spektroskopie auf ein Multikomponentensystem wie beispielsweise auf den Extrakt einer Pflanze ist nicht so leicht möglich. Eine Ursache hierfür ist in der Komplexität der Matrix zu suchen. Diese muss zuvor herabgesetzt werden, so dass die chemischen Verschiebungen im NMR Spektrum klar voneinander getrennt sind und eindeutig einer Substanz zugeordnet werden können. Die NMR spektroskopische Analyse mehrerer Verbindungen in einem Stoffgemisch ist nur schwer möglich, da die Resonanzen oft nicht basisliniengetrennt sind und sich zum Teil überlagern. Um basisliniengetrennte NMR Resonanzen zu erhalten, muss die Komplexität des Extraktes herabgesetzt werden. Dies wurde im Rahmen dieser Arbeit durch präparative Flüssigkeitschromatographie erreicht. Diese trennt das Stoffgemisch durch die Wahl einer geeigneten Säule soweit auf, dass Substanzen ähnlicher Polarität in einer Fraktion gesammelt werden. Die erhaltenen Fraktionen können im Anschluss mit Hilfe von NMR Spektroskopie und GC-TOF-MS erfasst und korreliert werden. Die GC-TOF-MS Analyse ermöglicht nach gaschromatographischer Trennung eine Analyse von Fragmenten derivatisierter Verbindungen. Im Gegensatz dazu ermöglicht die NMR spektroskopische Analyse eine Detektion von Struktureinheiten. Die NMR Spektroskopie erlaubt Rückschlüsse über die Konnektivität einzelner Strukturelemente.

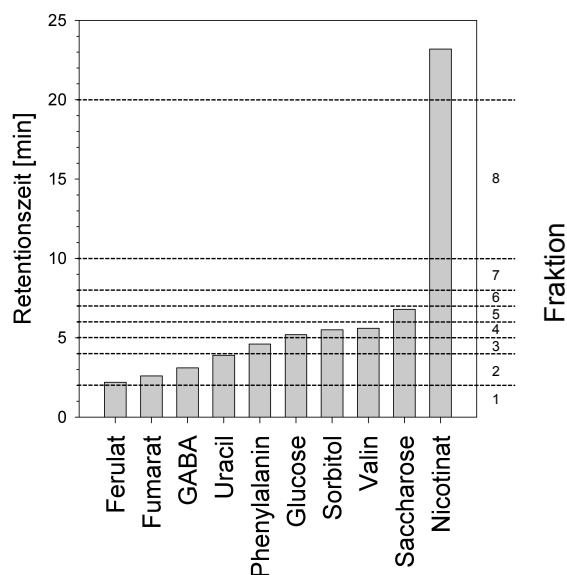
#### C.3.3.1 Optimierung der Fraktionierung

Da im Rahmen gaschromatographischer Analysen überwiegend Primärmetaboliten erfasst werden, wurde ein synthetischer Mix (Zusammensetzung siehe Anhang Tab. 7) erstellt, der zur Etablierung der Fraktionierung diente. Dieser synthetische Mix wurde so ausgewählt, dass er

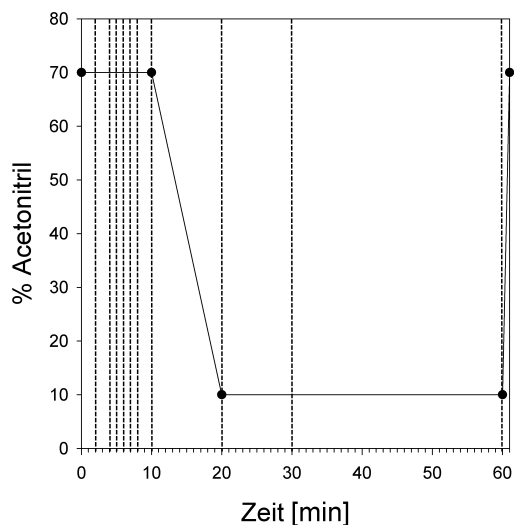
sämtliche gaschromatographisch leicht zugängliche, biologisch relevante Substanzklassen enthält. Diese Optimierungsreihe verfolgt das Ziel, die Bedingungen für den chromatographischen Trennprozess so zu wählen, dass die verwendeten Substanzen gut basisliniengetrennt sind. Während des Optimierungsprozesses wurden zwei stationäre Phasen („Prevail™ Carbohydrate ES“ und „Cyanopropyl“) sowie zwei mobile Phasen (Methanol und Acetonitril) untersucht. Der Lösungsmittelgradient wurde entsprechend Moco et al. an das gegenwärtige System angepasst<sup>[43]</sup>. Im Hinblick auf die Wahl der mobilen Phase konnte sehr schnell festgestellt werden, dass Acetonitril der bessere Eluent für die Fraktionierung ist, da dieses im Allgemeinen schärfere Signale liefert und zu einer kürzeren Elutionszeit führt. Dies ist auf die stärkere Elutionskraft von Acetonitril zurückzuführen.

Demgegenüber gestaltete sich die Wahl des Säulensystems schwieriger, da die Säulen bestimmte Stoffgruppen entweder besser oder schlechter auftrennen und dementsprechend eluieren. Epimere der Monosaccharide wurden besonders gut durch die Prevail™ Carbohydrate ES Säule aufgetrennt. Einige aromatische Verbindungen zeigten jedoch unter Verwendung der Cyanopropylsäule aufgrund der kürzeren Elutionszeit schärfere Signale. Ein längerer Chromatographiegradient wurde aufgrund der dabei auftretenden Bandenverbreiterungen nicht angestrebt. Aus diesem Grund wurde eine chromatographische Trennung basierend auf der Cyanopropylsäule etabliert, da für die vorliegende Fragestellung das Gesamtspektrum polarer Metaboliten entscheidend ist. Diese Substanzen weisen unter Verwendung der Cyanopropylsäule eine kürzere Elutionszeit auf und eluieren innerhalb eines kürzeren Zeitintervalls im Vergleich zur Prevail™ Carbohydrate ES Säule. Daraus kann die Schlussfolgerung gezogen werden, dass die Cyanopropylsäule ein größeres Substanzspektrum abdeckt. Auf die Trennung von Kohlenhydratisomeren wurde für die weiteren Analysen bewusst verzichtet. Diese werden durch die Cyanopropylsäule nicht aufgetrennt.

Um den Referenzmix unter Verwendung der optimierten Parameter aufzutrennen, wird eine Trennzeit von 25 min benötigt (Abb. 29). Um zu verhindern, dass während der Analyse biologischer Extrakte Komponenten auf der Säule aufgrund einer zu kurzen Analysenzeit zurückbleiben, wurde die Zeit für die chromatographische Trennung der Komponenten auf 60min verlängert. Daraus ergibt sich das in Abbildung 30 dargestellte Protokoll zur Fraktionierung biologischer Proben.



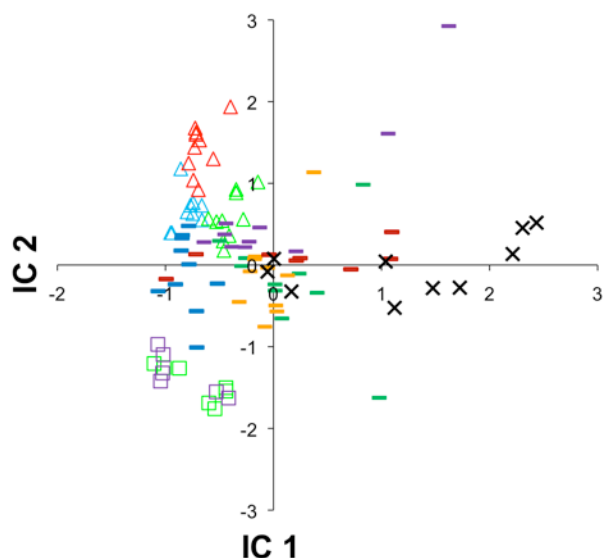
**Abb. 29** Retentionszeiten ausgewählter Substanzen des Testmixes nach Optimierung der flüssigkeitschromatographischen Bedingungen. Die entsprechenden Fraktionen sind durch gestrichelte Linien hervorgehoben.



**Abb. 30** Optimierte chromatographische Bedingungen zur Vorfraktionierung biologischer Extrakte an einer Cyanopropylsäule. Die gestrichelten senkrechten Linien geben an, ab welchem Zeitpunkt die nächste Fraktion gesammelt wurde. Es wurden 10 Fraktionen während eines chromatographischen Laufes gesammelt. Die durchgezogene Linie zeigt das Mischungsverhältnis von Acetonitril zu Wasser an.

Der Erfolg der etablierten Fraktionierung wurde beurteilt, indem der gaschromatographische Datensatz, welcher im splitlosen Modus aufgenommen wurde, mit Hilfe von einer Unabhängigen Komponentenanalyse (ICA) untersucht wurde. Hierbei wurde erwartet, dass

Fraktionen ähnlicher chemischer Zusammensetzung innerhalb einer Unabhängigen Hauptkomponente gruppierbar sind. Dieser Datensatz wurde für die folgenden Untersuchungen ausgewählt, da der gaschromatographische Datensatz, welcher mit einem Splitverhältnis von 1:50 aufgenommen wurde, maximal über 58 Signale verfügte. Während der ICA konnten die Fraktionen 1-3 (Dreiecke) von den Fraktionen 4-10 (Striche und Rechtecke) und den unfraktionierten Proben (Kreuze) getrennt werden. Weiterhin konnten die Fraktionen 9 und 10 (Rechtecke) von den Fraktionen 1-8 (Dreiecke und Striche) und den unfraktionierten Proben (Kreuze) getrennt werden. Dies zeigt, dass die präparative Flüssigkeitschromatographie die Komplexität der einzelnen Proben herabsetzt.



**Abb. 31 Hauptkomponentenanalyse angewendet auf alle detektierten im splitlosen Modus aufgenommenen GC-MS Signale.** Die zweite (PC2) und die vierte (PC4) Hauptkomponente zeigten die maximale Auftrennung. Fraktion 1  $\triangle$ ; Fraktion 2  $\triangle$ ; Fraktion 3  $\triangle$ ; Fraktion 4  $\text{---}$ ; Fraktion 5  $\text{---}$ ; Fraktion 6  $\text{---}$ ; Fraktion 7  $\text{---}$ ; Fraktion 8  $\text{---}$ ; Fraktion 9  $\square$ ; Fraktion 10  $\square$ ; unfraktioniert  $\times$

Die Ergebnisse aus der ICA wurden durch die fraktionsweise Analyse der sehr komplexen Matrix „GD33“ bestätigt. Die gesammelten Fraktionen 1-10 beinhalten jeweils Stoffklassen ähnlicher Funktionalität. Steroide und Fettsäuren wurden in der Fraktion 2, Nukleobasen in den Fraktionen 2-3, Alkohole in den Fraktionen 3-4, Einfachzucker in den Fraktionen 4-7, Zweifachzucker in den Fraktionen 6-8, aminhaltige Substanzen in den Fraktionen 6-9, Phenylpropanoide in der Fraktion 8 und polyfunktionelle organische Säuren in den Fraktionen 9-10 detektiert. Am Beispiel der einzelnen Extrakte von „GD33“ wurde gezeigt, dass auf diese Weise die Komplexität des Extraktes erheblich reduziert wurde (Anhang Abb.9) und die etablierte Fraktionierung die anschließende NMR- und GC-TOF-MS Analyse erleichtert.



### C.3.3.2 Gepaarte NMR und GC-TOF-MS Analyse

Da die Gaschromatographie bislang nur unter besonderen Umständen mit der Kernspinresonanzspektroskopie gekoppelt beziehungsweise beide Datensätze getrennt analysiert werden konnten, wurden der gaschromatographische und der kernspinresonanzspektroskopische Datensatz mit Hilfe einer Pearsonkorrelation zusammengeführt [202-204]. Somit sollten Strukturinformationen beider Analysensysteme mit Hilfe einer linearen Funktion übertragbar sein. Voraussetzung dafür ist, dass beide Systeme im linearen Bereich des Detektors arbeiten. Damit die Signale der GC-MS eindeutig den Signalen der NMR Spektroskopie zugeordnet werden können, wurde ein Grenzwert für den Korrelationskoeffizienten ( $r$ ) bestimmt. Dieser ermöglicht gaschromatographische Signale kernspinresonanzspektroskopischen Signalen zuzuordnen. Dazu wurden die identifizierten Signale aus der GC-MS anhand ihrer relativen Intensität im Massenchromatogramm in den entsprechenden NMR Spektren zurückverfolgt. Anschließend wurden diese NMR Resonanzen anhand von kommerziellen Referenzsubstanzen verifiziert und schließlich die Signalintensitäten beider Analysensysteme miteinander korreliert. Damit zwei Datensätze mit Hilfe von Korrelationen zusammengeführt werden können, muss ein ausreichendes Probenet untersucht werden, in dem eine möglichst hohe Varianz innerhalb der Probengruppen existiert. Letzteres wurde im Rahmen dieser Untersuchungsreihe dadurch sichergestellt, dass von verschiedenen Nutzpflanzen der Gattung der Kreuzblütler jeweils der essbare Teil untersucht wurde.

Die Varianz der einzelnen Matrizes innerhalb einer Fraktion wurde durch einen Boxplot abgeschätzt (Abb. 32). Ein Boxplot dient dazu, die Verteilungen der aufgenommenen Variablen miteinander zu vergleichen. Mit Hilfe des Boxplots können symmetrische Verteilungen und Ausreißer erkannt sowie Aussagen zum Ausmaß der Streuung innerhalb der Messreihe getroffen werden. Die Häufigkeitsverteilung wird anhand von fünf Kennzahlen beschrieben:

- Median
- 1. Quartil
- 3. Quartil
- unterer Whisker
- oberer Whisker

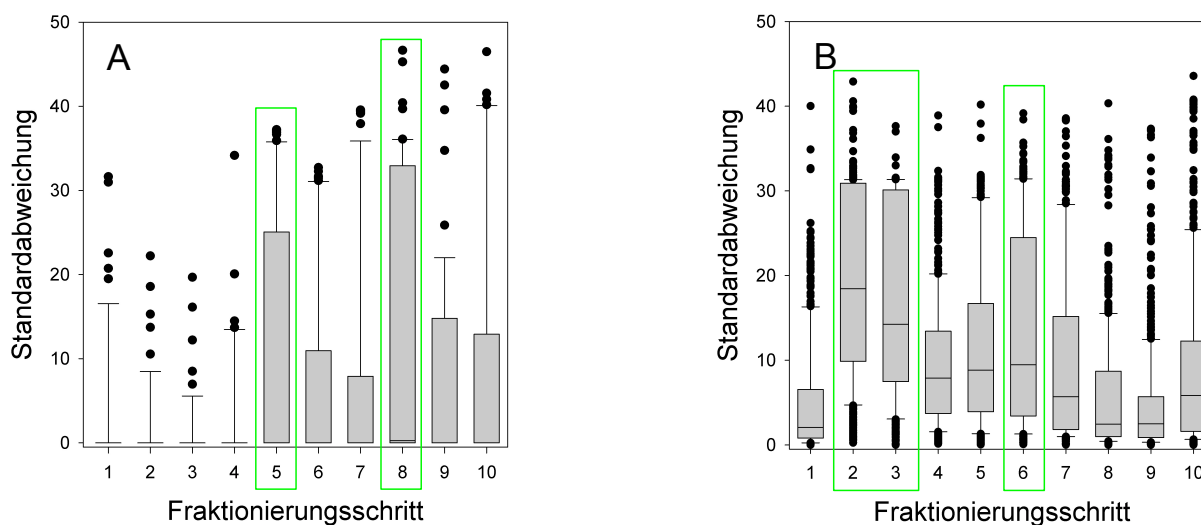
Dabei fallen 50 % der Messwerte innerhalb des 1. und 3. Quartils (eingerahmte Box, Abb. 32). Die restlichen 50 % der Werte werden durch den oberen und den unteren Whisker (oberer und unterer Balken, Abb. 32) begrenzt<sup>[77]</sup>. Sie dienen dazu Ausreißer im Experiment zu identifizieren. Damit zwei Datensätze anhand von Korrelationen zusammengeführt werden können, wird ein großer Abstand zwischen dem 1. Quartil und dem 3. Quartil (Quartilsabstand) angestrebt. Um diesen zu ermitteln wurde die Standardabweichung aller detektierten Signale ( $S/N = 4$ ) innerhalb einer Fraktion und die Verteilung der Standardabweichung für beide Analysensysteme bestimmt (Abb. 32). Wie erwartet unterscheidet sich die Varianz der detektierten Signale zwischen den einzelnen gesammelten Fraktionen (1-10) und den beiden Analysensystemen

(NMR und GC-TOF-MS). Der Quartilsabstand war innerhalb der Fraktionen 5 und 8 (grüne Umrandung, Abb. 32A) für den gaschromatographischen Datensatz am größten und innerhalb der Fraktionen 1-4 am niedrigsten. Dies zeigt, dass innerhalb der Fraktionen 5-8 die meisten gaschromatographisch messbaren Substanzen zu finden sind. Im Gegensatz dazu wies der kernspinresonanzspektroskopische Datensatz den maximalen Quartilsabstand innerhalb der Fraktion 6 (grüne Umrandung, Abb. 32B) auf sowie einen hohen Quartilsabstand innerhalb der Fraktionen 2 und 3 (grüne Umrandung, Abb. 32B). In den Fraktionen 2 und 3 eluieren überwiegend unpolare Verbindungen von der vorgeschalteten flüssigkeitschromatographischen Säule. Diese Verbindungen werden mit Hilfe der gegenwärtigen gaschromatographischen Messmethode nicht quantitativ erfasst. In weiteren Versuchsreihen müsste eine andere Derivatisierungsreaktion (Bildung von Methylestern) oder andere Trennsäule (Waxsäule, 50 %-Cyanopropylmethylpolysiloxan oder 100 % Dimethylpolysiloxan) gewählt werden.

Für die NMR Spektroskopie stellt die Polarität eines Moleküls keine Beschränkung dar. Weiterhin besteht bei kernspinresonanzspektroskopischen Untersuchungen keine Einschränkung in der Detektion höhermolekularer und geladener Moleküle. Diese werden mit Hilfe der Gaschromatographie nicht erfasst, da diese sich nur sehr schwer verdampfen lassen. Im Gegensatz zur Gaschromatographie weist die NMR Spektroskopie ihre größte Einschränkung in Bezug auf die Nachweisgrenze auf. Diese aufgeführten Unterschiede können zu Unterschieden innerhalb der Boxplots führen (Abb. 32). Allerdings kann festgestellt werden, dass beide Methoden sich ergänzen, was zu einer höheren Aussagekraft führt.

Auch wenn die Verteilung der Standardabweichung der detektierten Signale für die GC-MS und die NMR Messung innerhalb einzelner Fraktionen zum Teil verschieden ist, wurden im Folgenden beide Datensätze über eine Pearsonkorrelation zusammengeführt. Dabei wurde akzeptiert, dass einige Metaboliten aus der GC-MS Messung eventuell nicht mit denen aus der NMR Messung korreliert werden können.

Da die Kernspinresonanzspektroskopie im Gegensatz zur Massenspektrometrie insbesondere im unteren Konzentrationsbereich wenig sensitiv ist, wurde für die Korrelationsanalysen der gaschromatographische Datensatz herangezogen, der unter einem Splitverhältnis von 1: 50 aufgenommen wurde. Der andere gaschromatographische Datensatz wurde lediglich zu Vergleichszwecken herangezogen. Da die betrachteten Signale zum Teil über mehrere Fraktionen aufgrund der hohen Konzentration einzelner Analyten wiedergefunden werden konnten, wurde die Korrelation über alle gesammelten Fraktionen durchgeführt, mit dem Ziel weitere Messwerte für die Korrelation zu generieren. Dazu wurde von dem gaschromatographischen Datensatz die Signalthöhe verwendet und von dem kernspinresonanzspektroskopischen Datensatz das Integral eindeutiger Protonen manuell ermittelt.



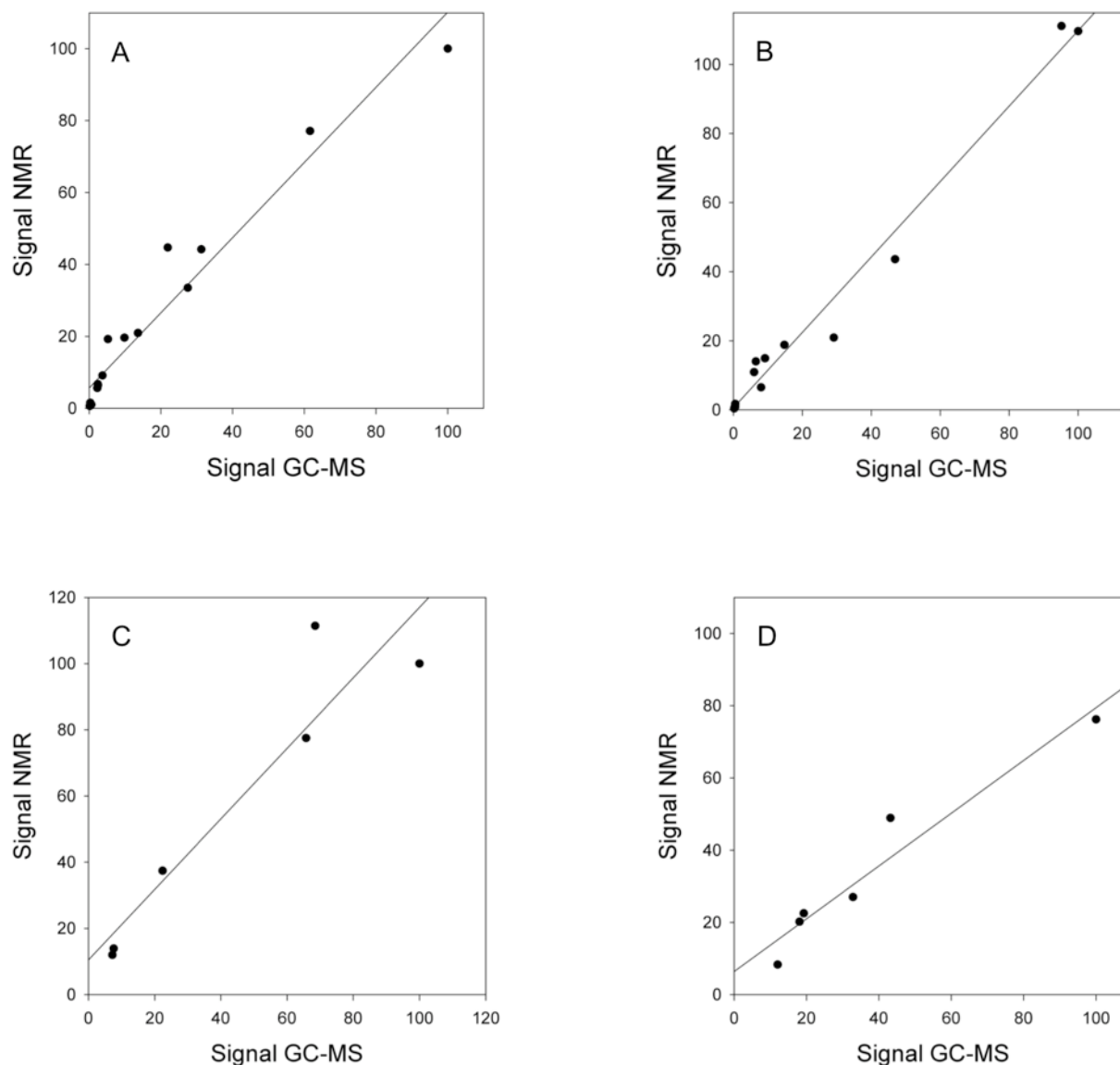
**Abb. 32 Varianz innerhalb der gesammelten Fraktionen nach A) split 50 GC Messung und B) NMR Messung innerhalb der gesammelten Fraktionen.** Die Varianz innerhalb einer Fraktion wird durch die Standardabweichung ausgedrückt.

Hierbei zeigte sich, dass die Mehrheit der polaren Verbindungen, welche innerhalb des Primärstoffwechsels eine zentrale Rolle einnehmen, miteinander korreliert werden können. Alanin, Asparagin, Aspartat, Cystin, Fructose (Abb. 33A), Glucose, Glutamat, Glutamin, Isoleucin, myo-Inositol (Abb. 33C), Prolin, Serin (Abb. 33D), Saccharose (Abb. 33B) und Valin wiesen einen sehr hohen Korrelationskoeffizienten auf ( $r^2 \geq 0.9$ ). Dieser liegt bei der gewählten Probenmenge ( $n = 4-19$ ) mit Ausnahme von Aspartat innerhalb des statistischen Vertrauensbereiches ( $p > 0.01$ ). Threonin wies mit  $r^2 = 0.8$  einen geringeren Korrelationskoeffizienten auf und fiel ebenfalls nicht mit in den statistisch signifikanten Bereich. Die Signalintensität der organischen Säuren GABA ( $r^2 = 0.5$ ), Citrat ( $r^2 = 0.5$ ) und Malat ( $r^2 = 0.1$ ) konnte nicht zwischen beiden Analysensystemen übertragen werden. Überladungseffekte und die hohe Acidität der Säuregruppen beeinflussen die Korrelation. Dennoch kann anhand dieser Untersuchungen geschlossen werden, dass für die Korrelation ein Grenzwert von 0.9 verwendet werden sollte. Dieser kann erhöht werden, indem weitere Pflanzen der Kreuzblütlerfamilie mit Hilfe des etablierten Arbeitsflusses untersucht werden. Hierbei ist es möglich die Menge falsch positiver Korrelationen zu reduzieren.

Für die eindeutige Zuordnung der Resonanzen eines NMR Spektrums zu einem gaschromatographischen Signal ist nicht nur der Korrelationskoeffizient entscheidend, sondern auch die Steigung und der Achsenabschnitt der Übertragung. Eine Steigung von eins deutet darauf hin, dass die Messinformation beider Systeme problemlos ineinander überführt werden kann. Eine doppelte Konzentrationserhöhung führt bei beiden Systemen zu einer doppelten Signalerhöhung. Eine Abweichung davon lässt den Schluss zu, dass noch weitere Substanzen an diesem Signal beteiligt sind und das Signal nicht chemisch rein ist. Aus diesem Grund wird für folgende Untersuchungen eine Steigung von eins empfohlen.

Für die genannten Substanzen konnte mit Ausnahme von Alanin ( $m = 2$ ) und Glutamat ( $m = 2.8$ ) eine Steigung von 0.7 bis 1.1 beobachtet werden. Der Achsenabschnitt dieser Substanzen variiert zwischen -7 und 24. Dieser deutlich von null verschiedene Achsenabschnitt kann nur damit begründet werden, dass jede Verbindung ihre individuelle Nachweisgrenze aufweist. Ein hoher Achsenabschnitt lässt vermuten, dass eines der Messsysteme einen „Offset“ besitzt. Dennoch sollte im Rahmen zukünftiger Analysen die Steigung und der Korrelationskoeffizient nicht den festgelegten Toleranzbereich überschreiten. Anhand des etablierten Arbeitsflusses konnte Trigonellin aufgeklärt werden. Dies ist ein Stoffwechselprodukt des Nicotinsäuremetabolismus und wurde bereits in Leguminosen detektiert<sup>[205]</sup>.

Die etablierte Analysenmethode der gekoppelten NMR- und GC-TOF-MS Analyse geht über die Klassifizierung von gaschromatographischen Signalen hinaus. Sie ermöglicht eine Aufklärung von gaschromatographischen Signalen anhand von eindeutigen Mustern der kernspinresonanzspektroskopischen Signale. Durch die Kopplung beider Messsysteme ist eine neue Qualität metabolischer Ergebnisse möglich.



**Abb. 33 Pearsonkorrelation von ausgewählten Primärmetaboliten für biochemisch bedeutende Stoffklassen.** A) Fructose  $r^2 = 0.95$ ;  $y = x + 5.7$ ; B) Saccharose  $r^2 = 0.98$ ;  $y = 1.1x + 0.6$ ; C) myo-Inositol  $r^2 = 0.98$ ;  $y = x + 10.5$ ; D) Serin  $r^2 = 0.94$ ;  $y = 1.1x + 0.6$

### C.3.4 Bewertung der Ergebnisse zur Klassifizierung von GC-MS Signalen

Bislang wurde einem unbekanntem GC-MS Signal ein eindeutiger Bezeichner zugeordnet. Dieser leitet sich aus dem Retentionsindex ab. Der Bezeichner wurde mitsamt Massenspektrum und Retentionsindex in einer Datenbank archiviert, um experimentübergreifende Aussagen zu treffen. Dabei wurde keine Hypothese über die chemische Zugehörigkeit des gaschromatographischen Signals aufgestellt. Um diesen Signalen eine chemische Funktionalität zuzuordnen, wurden im Rahmen dieser Arbeit drei Verfahren (das Entscheidungs-

baumverfahren, die Festphasenextraktion, die gepaarte NMR- und GC-TOF-MS Analyse) entwickelt, welche eine Klassifizierung von unbekanntem Massenspektren von biologischen Extrakten ermöglichen.

Das Probenvorbereitungsverfahren der Festphasenextraktion wurde bereits mehrfach zur Analyse bestimmter Verbindungsklassen angewandt. Die Anwendung dieser Technik zur Strukturaufklärung endogener Metabolite ist jedoch noch nicht vorgenommen worden. Ähnliches gilt für das Entscheidungsbaumverfahren. Es sind bislang eine Reihe von chemoinformatischen Klassifizierungsverfahren in der Literatur beschrieben. Diese wurden jedoch noch nicht auf komplexe Chromatogramme angewendet. Weiterhin legen diese Verfahren keine Regeln zur Klassifizierung von Massenspektren fest und sind auch nicht auf der Basis von biologisch relevanten Daten entwickelt worden. Analoges gilt für das Konzept der gepaarten NMR- und GC-TOF-MS Analyse.

Ein NMR- und ein GC-TOF-MS Datensatz sind noch nicht vereinigt worden, um GC-MS Signale zu klassifizieren. Beide Analysentechnologien wurden bislang erfolgreich zur Analyse einer Vielzahl von Verbindungen, zum „metabolite profiling“ und zum „metabolite fingerprinting“ angewandt. Hierbei konnten wichtige Stoffwechselprodukte identifiziert werden.

Obwohl alle drei Verfahren für die Klassifizierung von Massenchromatogrammen hilfreich sind, besitzen sie ihre Vor- und Nachteile. Das Verfahren der gepaarten NMR- und der GC-TOF-MS Analyse ist verglichen zu den anderen beiden Verfahren für die Analyse von Metabolitenprofilen am vielversprechendsten. Bei diesem Ansatz werden zwei orthogonale Analysentechnologien vereinigt und entsprechend viele Strukturinformationen gesammelt, dass selbst eine Aufklärung gaschromatographischer Signale möglich ist. Der limitierende Schritt bei diesem Verfahren besteht darin, biologisches Material mit einer hinreichenden Varianz aufzufinden und die Signale aus der NMR Messung einer chemischen Verbindung zuzuordnen. Im Gegensatz dazu basieren die anderen beiden Verfahren nur auf GC-TOF-MS Untersuchungen, so dass die Nachteile, welche bei der Interpretation von Elektronenstoßionenspektren auftreten, nicht kompensierbar sind. Die einzige Hilfestellung bei diesen Verfahren bietet die computergestützte Modellierung und die Kenntnis über die chemische Funktionalität.

## D Zusammenfassung

Das Bestreben eine effiziente und flächendeckende Ernährung der Weltbevölkerung zu gewährleisten setzt eine effektivere Verwertung von Pflanzen voraus. In diesem Zusammenhang kommt der Analyse niedermolekularer Primär- und Sekundärmetaboliten von agrarwirtschaftlich bedeutsamen Nutzpflanzen z.B. mit Hilfe von gekoppelter Gaschromatographie-Massenspektrometrie eine große Bedeutung zu. Diese Analysetechnik ermöglicht eine hohe Auftrennung des Pflanzenextraktes und liefert dadurch wertvolle Strukturinformationen der endogenen Metaboliten.

Im Rahmen dieser Arbeit wurden analytische Lösungsansätze geschaffen, die eine schnelle Auswertung einer Vielzahl von Proben ermöglichen. Dafür wurde die Methode der GC-MS ausgewählt. Die GC-MS ist in der Lage die zu analysierenden Extrakte mit hoher Präzision aufzutrennen und ein eindeutiges Fragmentierungsmuster aller im Extrakt vorkommenden Verbindungen zu liefern. Die resultierenden physikalischen Größen, wie die Retentionszeit und die Masse der Fragmentationen, ermöglichen eine Etablierung von umfangreichen Referenzspektrenbibliotheken, die dazu beitragen die gemessenen Signale aufzuklären.

Das Modell von van den Dool, welches die Retentionszeit in den Retentionsindex umrechnet, wurde erfolgreich angewendet, um die Drift in der Retentionszeit im Verlauf großangelegter Messserien zu umgehen. Dazu wurde im Rahmen dieser Arbeit untersucht, inwiefern polynomiale Regressionen 1.-5. Ordnung und die Spline Funktion zur Auswertung metabolischer Daten geeignet sind. Das Modell von van den Dool und die Spline Funktion sind für eine genaue Bestimmung des Retentionsindex geeignet. Im Rahmen dieser Arbeit konnte an vier Matrices (synthetischer Mix geringer Komplexität, synthetischer Mix mittlerer Komplexität, „Bäckerhefe“ und „Reisblatt“) erstmalig gezeigt werden, dass das Modell von van den Dool zu einer genauen Berechnung des Retentionsindex führt. Der Retentionsindex wurde für die sehr komplexe Matrix „Reisblatt“ im frühen Chromatographiebereich zu fünf RI Einheiten exakt bestimmt, während im mittleren und im späten Chromatographiebereich die Ungenauigkeit auf 0.5-1 RI Einheit reduziert werden konnte. Die weniger komplexen Matrices zeigen geringere Fehler in der Retentionsindexbestimmung.

Darüber hinaus konnte gezeigt werden, dass der Retentionsindex zwischen zwei Laboratorien unter Verwendung eines Polynoms erster Ordnung übertragbar ist, wenn beide Laboratorien Säulen gleicher Polarität verwenden. Obwohl sich die Ungenauigkeit in der Retentionsindexbestimmung durch den Übertragungsprozess auf 0.5 % RI-Einheiten vergrößert, können auf diese Weise Spektrenbibliotheken zwischen verschiedenen Laboratorien ausgetauscht werden. Weiterhin können Spektren anderer Laboratorien in den Vergleichsprozess einbezogen werden und die Spektrenbibliothek im eigenen Labor in kürzester Zeit um ein beachtliches Maß erweitert werden.

Neben dem Retentionsindex ist im Rahmen von GC-MS Untersuchungen auch der massenspektrale Vergleich entscheidend. Von der Vielzahl an Parametern für den massenspektralen Vergleich wurden im Rahmen dieser Arbeit vier Distanzmaße („Euklidische

Distanz“, „Skalarprodukt“, „Hammingdistanz“ und „Jaccarddistanz“) näher untersucht und mit den entsprechenden Parametern des Programms „NIST MS Search“ verglichen. Hierbei zeigten die Vergleichsparameter „Euklidische Distanz“ und „Skalarprodukt“ eine ähnliche Leistung, wie der vom NIST Programm ausgegebene R-Wert sowie die vom NIST Programm berechnete Wahrscheinlichkeit. Beide Vergleichsparameter diskriminierten richtig positive Treffer hinreichend von falsch positiven Treffern.

Die Identifizierung von Substanzen erfordert die Arbeit mit einer umfangreichen Spektrenbibliothek, welche ein möglichst breites Substanzspektrum abdeckt und qualitativ hochwertige Massenspektren beinhaltet. Die Qualität eines Massenspektrums wird durch die Komplexität der Matrix, die Konzentration der endogenen Metaboliten und dem Dekonvolutionsalgorithmus bestimmt. Die Etablierung von Durchschnittsspektrenbibliotheken auf der Basis von vorhandenen Massenspektren innerhalb einer Spektrenbibliothek ist für die Erzielung qualitativ hochwertiger Massenspektren die Methode der Wahl. Im Rahmen dieser Arbeit wurde gezeigt, dass mindestens 80 % aller Massenfragmente über einen Analyten detektiert werden müssen, um ein verwertbares Referenzspektrum zu berechnen und in ein Datenbankschema zu bringen, in welchem auch die phänotypischen Daten hinterlegt sind. Dieses Datenbankschema ist Teil der „Golm Metabolome Database“. Diese Vorgehensweise gestattet neben einer Normierung auf Frischgewicht und internen Standard die Einbeziehung chemoinformatischer Informationen wie der CAS Nummer und der KEGG Nummer. Erstmals erscheinen als Erweiterung der Exporttabelle die Substanzabgleichsparameter, wie die Retentionsindexabweichung und der massenspektrale Vergleich. Diese Funktionalitäten sind bisher von bestehenden Datenbanklösungen anderer Arbeitsgruppen noch nicht realisiert worden.

Da das vorgegebene Aufarbeitungsprotokoll zum „metabolite profiling“ die Analyse mittel- bis hochkonzentrierter Substanzen zum Ziel hat, wurde im Rahmen dieser Arbeit die Vorbereitung der Proben so variiert, dass vor allem Substanzen geringer Konzentration analysiert werden können. Diese sind für regulatorische Prozesse in höheren Pflanzen entscheidend. Als Modellsubstanzen wurden die sauren Phytohormone Indolelessigsäure, Abscisinsäure und Gibberellinsäure ausgewählt, die mit Hilfe einer sauren, wässrigen Extraktion (2 % Ameisensäure) und einer sich anschließenden Aufarbeitung mittels Festphasenextraktion an dem neuartigen Sorbens „Plexa“ angereichert wurden. Die aufgereinigten Extrakte wurden mit Hilfe von zwei unabhängigen Methoden (UPLC-FTICR-MS und GC-TOF-MS) analysiert. Als Matrix diente ein Extrakt blühender Arabisidopsisblätter und keimende Kartoffelknollen. Das ausgewählte Aufbereitungsverfahren reduziert nicht nur die Komplexität des gewonnenen pflanzlichen Extraktes. Es bewirkt zusätzlich einen hohen Anreicherungsfaktor, so dass eine parallele Analyse von Jasmonat, Salicylat und GA4 neben Abscisinsäure, Indolelessigsäure und Gibberellinsäure möglich ist. Diese Substanzen konnten mit dem gegenwärtigen Aufarbeitungsprotokoll nicht analysiert werden. Obwohl die identifizierten Substanzen eine unterschiedliche chemische Struktur besitzen, lassen sie sich nach dem etablierten Aufbereitungsverfahren im gleichen Extrakt nachweisen. Wie erwartet können niedermolekulare Moleküle sensitiver mittels Gaschromatographie und höhermolekulare Substanzen besser



durch Flüssigkeitschromatographie analysiert werden. Die dabei erhaltenen Ergebnisse bestätigen, dass sich beide Analysenmethoden gegenseitig ergänzen und wertvolle Informationen zur Zusammensetzung pflanzlicher Extrakte beitragen.

Unter Verwendung des etablierten Datenbankschemas und der Spektrenbibliothek konnten im Rahmen dieser Arbeit relevante Stoffwechselprodukte für die Nutzpflanze Kartoffel mittels GC-TOF-MS identifiziert werden. Dazu wurden im Einzelnen die physiologischen Prozesse der Schwarzfleckigkeit, des Physiologischen Alters und der Chipsbräunung näher betrachtet. Diese Prozesse sind von aktuellem Interesse, da sie die Qualität von Nutzpflanzen herabsetzen.

Die im Rahmen dieser Arbeit durchgeführten Untersuchungen zeigen, dass eine mechanische Belastung von Kartoffelknollen zu Schwarzfleckigkeit führt. Nach mechanischer Belastung sind Stoffwechselprodukte des TCA Cyclus (Succinat, Fumarat) und die Stoffwechselprodukte der abzweigenden Aminosäuren (Aspartat, Glutamat, Glycerat) sowie Zuckerstoffwechselwege (Saccharose, Galactinol) aktiv. Dagegen zeigen Phenylalanin und Tyrosin, welche bislang als relevant für den Prozess der Schwarzfleckigkeit beschrieben wurden, keine signifikanten Änderungen in den vorliegenden Experimenten.

Für das Physiologische Alter wurden signifikante Änderungen im Monosaccharid- (Glucose, Fructose), im Saccharose- und im Malat-Gehalt anhand einer Zweifaktoriellen Varianzanalyse (Zeit x Sorte) festgestellt. Darüber hinaus zeigt die Einfaktorielle Varianzanalyse signifikante Änderungen im Aminosäurestoffwechsel (Aspartat, Glutamat, Methionin, Prolin), in den abzweigenden Glykolysestoffwechselwegen (Glycerol-3-phosphat, Caffeat) und im Dehydroascorbatgehalt in Bezug auf den Faktor Zeit. Methionin zeigt weiterhin eine gute Korrelation mit dem Gewichtsverlust. Glutamin, Glycin und GABA zeigen ebenfalls einen hohen Korrelationskoeffizienten mit dem Gewichtsverlust. Die Konzentration dieser Stoffwechselprodukte einschließlich des Methionins nimmt während der Lagerung ab. Glucose hingegen korreliert positiv mit der Keimungsrate.

Die Anwendung von höheren Temperaturen (8 °C) beschleunigt den Alterungsprozess, während die Anwendung von Keimhemmungsmitteln den Alterungsprozess verlangsamt. Die Anwendung von Chlorpropham und Kümmelöl vor der Einlagerung von Kartoffelknollen verzögert das Austreiben neuer Keime. Jedoch hat es keinen Einfluss auf den Gewichtsverlust. Obwohl die mit Chlorpropham behandelten Knollen keine Keime aufweisen, wird bei diesen Knollen eine Akkumulation von Zuckern, Polyolen und Fettsäuren beobachtet. Diese Substanzen zeigen nach Anwendung von Kümmelöl wesentlich geringere Änderungen. Die biologische Keimhemmung hat jedoch einen großen Einfluss auf den Aminosäurestoffwechsel (Isoleucin, Methionin, Ornithin, Prolin), den Phenylpropanoidstoffwechsel (Nicotinat, Tyrosin) und den Polyaminstoffwechsel (Putrescin). Die gefundenen Stoffwechselprodukte sind indikativ für die genannten physiologischen Prozesse und können als indikative Biomarker betrachtet werden.

Im Rahmen dieser Arbeit wurden prädiktive Biomarker aus GC-MS Profilen von Kartoffelknollen extrahiert. Dazu wurde der physiologische Prozess der Chipsbräunung näher betrachtet und das Verfahren der „Partial Least Squares Variables Importance in the Projection Method“ (PLS-VIP) angewendet. Glucose und Fructose wurden dabei aus den Kartoffelknollen von zwei

Standorten (Ebstorf, Böhlendorf) und zwei Anbaujahren (2006/2007, 2007/2008) als potentielle Biomarker extrahiert. Die Anwendung des etablierten Modells auf eine neue Züchtungspopulation eines weiteren Anbaujahres bestätigt die Machbarkeit des Verfahrens (80% Vorhersagekraft).

Ferner wurden drei Verfahren entwickelt, die eine Klassifizierung von unbekanntem Signalen ermöglichen. Diese Verfahren umfassen das Entscheidungsbaumverfahren, die Festphasenextraktion zur Probenvorbereitung und die gepaarte NMR- und GC-MS Analyse. Im Rahmen des Entscheidungsbaumverfahrens wurden computergestützte Regeln festgelegt, die eine Klassifizierung unbekannter Signale hinsichtlich der An- und Abwesenheit funktioneller Gruppen ermöglichen. In einem weiteren Verfahren wurden drei verschiedene Festphasenextraktionskartuschen unterschiedlicher Funktionalität (Anionenaustauscher, Kationenaustauscher, C18 ähnliches Sorbens) hinsichtlich ihres Anreicherungsverhaltens untersucht. Hierbei konnten mit Hilfe des Kationenaustauschers Aminosäuren, Säureamide, Polyole und stickstoffhaltige Substanzen effizient angereichert werden. Die C18 ähnliche Kartusche retardiert bevorzugt ungesättigte Verbindungen. Die Anionenaustauscherkartusche hingegen konnte keine der extrahierten polaren Metaboliten in ausreichender Ausbeute anreichern. Dagegen unterstützen die Kationenaustauscherkartusche und die C18 ähnliche Kartusche die Klassifizierung einiger der unbekanntem Verbindungen hinsichtlich ihrer chemischen Zugehörigkeit durch den zusätzlichen Probenvorbereitungsschritt.

Das Verfahren der gepaarten NMR- und GC-MS Analyse greift das Konzept der präparativen Auftrennung von biologischen Extrakten auf. Durch die Anwendung einer präparativen Flüssigkeitschromatographie wird der pflanzliche Extrakt in mehrere Fraktionen (10 Fraktionen) aufgetrennt, die im nächsten Schritt mit Hilfe der NMR Spektroskopie und anschließend mit Hilfe der GC-MS analysiert wurden. Beide Datensätze wurden durch eine Pearsonkorrelation vereinigt. Die Pearsonkorrelation zeigt, dass für nahezu alle endogenen Metaboliten mit Ausnahme von Citrat, Malat und GABA, ein hoher Korrelationskoeffizient ( $r^2 = 0.9$ ) und eine Steigung nahe eins beobachtet wird. Daraus kann die Schlussfolgerung gezogen werden, dass Signale unterschiedlicher analytischer Methoden vereinigt werden können und weitere Strukturinformationen zur Klassifizierung von unbekanntem Signalen gewonnen werden können. Diese Arbeit leistet einen Beitrag zur Analytik der Metabolomik. Außerdem wird gezeigt, dass die Metabolomik in Zukunft nicht nur von Pflanzenphysiologen genutzt werden kann um physiologische Prozesse aufzuklären, sondern auch von Züchtern oder gar der Nahrungsmittelindustrie, um die Qualität des Ernteguts oder des Verkaufsproduktes einzuschätzen.

## E Literaturverzeichnis

- [1] U. Roessner, *Metabolome Analysis: An Introduction (John Wiley & Sons)* **2007**, 1, 215.
- [2] M. Q. Le, W. R. Engelsberger, D. K. Hinch, *Cryobiology* **2008**, 57, 104.
- [3] M. Korn, T. Gaertner, A. Erban, J. Kopka, J. Selbig, D. K. Hinch, *Mol. Plant* **2010**, 3, 224.
- [4] D. H. Sanchez, J. Szymanski, A. Erban, M. K. Udvardi, J. Kopka, *Plant, Cell Environ.* **2010**, 33, 468.
- [5] A. Mithofer, B. Schulze, W. Boland, *FEBS Lett.* **2004**, 566, 1.
- [6] S. Soyka, A. G. Heyer, *FEBS Lett.* **1999**, 458, 219.
- [7] J. V. Passonneau, O. H. Lowry, *Humana Press (Totowa, NJ:1993)* **1993**, 85.
- [8] S. C. Carpentier, B. Coemans, N. Podevin, K. Laukens, E. Witters, H. Matsumura, R. Terauchi, R. Swennen, B. Panis, *Physiol. Plant.* **2008**, 133, 117.
- [9] W. Weckwerth, *Annu. Plant Rev.* **2009**, 35, 258.
- [10] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A. R. Fernie, *Nat. Protoc.* **2006**, 1, 387.
- [11] M. M. Koek, B. Muilwijk, M. J. Van der Werf, T. Hankemeier, *Anal. Chem.* **2006**, 78, 1272.
- [12] G. Stephanopoulos, H. Alper, J. Moxley, *Nat. Biotechnol.* **2004**, 22, 1261.
- [13] S. Strelkov, M. von Elstermann, D. Schomburg, *Biol. Chem.* **2004**, 385, 853.
- [14] S. Dietmair, N. E. Timmins, P. P. Gray, L. K. Nielsen, J. O. Krömer, *Anal Biochem* **2010**, 404, 155.
- [15] B. Ebert, D. Zoeller, A. Erban, I. Fehrle, J. Hartmann, A. Niehl, J. Kopka, J. Fisahn, *J. Exp. Bot.* **2010**, 61, 1321.
- [16] W. Weckwerth, O. Fiehn, *Curr. Opin. Biotechnol.* **2002**, 13, 156.
- [17] O. Fiehn, *Plant Mol. Biol.* **2002**, 48, 155.
- [18] T. Shepherd, G. Dobson, S. R. Verrall, S. Conner, D. W. Griffiths, J. W. McNicol, H. V. Davies, D. Stewart, *Metabolomics* **2007**, 3, 475.
- [19] U. Roessner, L. Willmitzer, A. R. Fernie, *Plant Physiol.* **2001**, 127, 749.
- [20] A. Cuadros-Inostroza, C. Caldana, H. Redestig, M. Kusano, J. Lisec, H. Pena-Cortes, L. Willmitzer, M. A. Hannah, *BMC Bioinf.* **2009**, 10, 428.
- [21] C. Sellick, D. Knight, A. Croxford, A. Maqsood, G. Stephens, R. Goodacre, A. Dickson, *Metabolomics* **2010**, 6, 427.
- [22] M. Fraccaroli, S. Nicoletti, F. Maltese, Y. H. Choi, F. Guzzo, M. Levi, R. Verpoorte, *Biotechnol. Lett.* **2008**, 30, 2031.
- [23] T. Frenzel, A. Miller, K. H. Engel, *Cereal Chem.* **2002**, 79, 215.
- [24] A. M. Flores-Valverde, E. M. Hill, *Anal. Chem.* **2008**, 80, 8771.
- [25] A. Crozier, T. Moritz, *New Compr. Biochem.* **1999**, 33, 23.
- [26] P. I. Dobrev, L. Havlicek, M. Vagner, J. Malbeck, M. Kaminek, *J. Chromatogr., A* **2005**, 1075, 159.
- [27] X. Pan, R. Welti, X. Wang, *Phytochemistry* **2008**, 69, 1773.
- [28] S. Hou, J. Zhu, M. Ding, G. Lv, *Talanta* **2008**, 76, 798.
- [29] M. Okamoto, A. Hanada, Y. Kamiya, S. Yamaguchi, E. Nambara, *Methods Mol. Biol. (Totowa, NJ, U. S.)* **2009**, 495, 53.
- [30] C. Birkemeyer, A. Kolasa, J. Kopka, *J. Chromatogr., A* **2003**, 993, 89.
- [31] Z. Ma, L. Ge, A. S. Y. Lee, J. W. H. Yong, S. N. Tan, E. S. Ong, *Anal. Chim. Acta* **2008**, 610, 274.
- [32] X.-Q. Pan, R. Welti, X.-M. Wang, *Nat. Protoc.* **2010**, 5, 986.
- [33] H. J. Issaq, Q. N. Van, T. J. Waybright, G. M. Muschik, T. D. Veenstra, *J. Sep. Sci.* **2009**, 32, 2183.
- [34] O. Fiehn, J. Kopka, R. N. Trethewey, L. Willmitzer, *Anal. Chem.* **2000**, 72, 3573.
- [35] A. Barsch, T. Patschkowski, K. Niehaus, *Funct. Integr. Genomics* **2004**, 4, 219.
- [36] A. Erban, N. Schauer, A. R. Fernie, J. Kopka, *Methods Mol. Biol. (Totowa, NJ, U. S.)* **2007**, 358, 19.
- [37] O. Fiehn, *Trends in Analytical Chemistry* **2008**, 27, 261.
- [38] M. Glinski, W. Weckwerth, *Mass Spectrom. Rev.* **2006**, 25, 173.
- [39] P. Giavalisco, K. Kohl, J. Hummel, B. Seiwert, L. Willmitzer, *Anal. Chem.* **2009**, 81, 6546.
- [40] C. Schummer, O. Delhomme, B. M. R. Appenzeller, R. Wennig, M. Millet, *Talanta* **2009**, 77, 1473.
- [41] K. A. Aliferis, S. Jabaji, *Metabolomics* **2010**, 6, 96.
- [42] J. M. Buscher, D. Czernik, J. C. Ewald, U. Sauer, N. Zamboni, *Anal. Chem.* **2009**, 81, 2135.
- [43] S. Moco, J. Forshed, R. C. H. Vos, R. J. Bino, J. Vervoort, *Metabolomics* **2008**, 4, 202.
- [44] R. t'Kindt, K. Morreel, D. Deforce, W. Boerjan, J. Van Bocxlaer, *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2009**, 877, 3572.
- [45] T. Kind, O. Fiehn, *Bioanalysis* **2009**, 1, 1511.

- [46] S. Kempa, J. Hummel, T. Schwemmer, M. Pietzke, N. Strehmel, S. Wienkoop, J. Kopka, W. Weckwerth, *J. Basic Microbiol.* **2009**, *49*, 82.
- [47] M. F. Almstetter, I. J. Appel, M. A. Gruber, C. Lottaz, B. Timischl, R. Spang, K. Dettmer, P. J. Oefner, *Anal. Chem.* **2009**, *81*, 5731.
- [48] K. Ralston-Hooper, A. Hopf, C. Oh, X. Zhang, J. Adamec, M. S. Sepulveda, *Aquat. Toxicol.* **2008**, *88*, 48.
- [49] R. A. Shellie, *Aust. J. Chem.* **2005**, *58*, 619.
- [50] M. M. Koek, B. Muilwijk, L. L. P. van Stee, T. Hankemeier, *J. Chromatogr., A* **2008**, *1186*, 420.
- [51] M. Kusano, A. Fukushima, M. Kobayashi, N. Hayashi, P. Jonsson, T. Moritz, K. Ebana, K. Saito, *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2007**, *855*, 71.
- [52] S. E. Stein, *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770.
- [53] A. Luedemann, K. Strassburg, A. Erban, J. Kopka, *Bioinformatics* **2008**, *24*, 732.
- [54] A. Lommen, *Anal. Chem.* **2009**, *81*, 3079.
- [55] H. Neuweger, S. P. Albaum, M. Dondrup, M. Persicke, T. Watt, K. Niehaus, J. Stoye, A. Goesmann, *Bioinformatics* **2008**, *24*, 2726.
- [56] A. Cuadros-Inostroza, C. Caldana, H. Redestig, M. Kusano, J. Lisec, H. Pena-Cortes, L. Willmitzer, M. A. Hannah, *BMC Bioinf.* **2009**, *10*, 428.
- [57] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, *Anal. Chem.* **2006**, *78*, 779.
- [58] 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.* **2005**, *579*, 1332.
- [59] 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* **2005**, *21*, 1635.
- [60] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, *Journal of Mass Spectrometry* **2010**, *45*, 703.
- [61] T. Kind, G. Wohlgemuth, D. Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, O. Fiehn, *Anal. Chem.* **2009**, *81*, 10038.
- [62] I. G. Zenkevich, V. I. Babushok, P. J. Linstrom, E. White V, S. E. Stein, *J. Chromatogr., A* **2009**, *1216*, 6651.
- [63] V. I. Babushok, P. J. Linstrom, J. J. Reed, I. G. Zenkevich, R. L. Brown, W. G. Mallard, S. E. Stein, *J. Chromatogr., A* **2007**, *1157*, 414.
- [64] 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* **2007**, *3*, 211.
- [65] S. E. Stein, *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 644.
- [66] C. Boelling, *Persönliche Mitteilung* **2008**.
- [67] R. Nakabayashi, M. Kusano, M. Kobayashi, T. Tohge, K. Yonekura-Sakakibara, N. Kogure, M. Yamazaki, M. Kitajima, K. Saito, *Phytochemistry* **2009**, *70*, 1017.
- [68] D. Herebian, B. Hanisch, F.-J. Marnett, *Metabolomics* **2005**, *1*, 317.
- [69] S. E. Stein, V. I. Babushok, R. L. Brown, P. J. Linstrom, *J. Chem. Inf. Model.* **2007**, *47*, 975.
- [70] C. T. Peng, Z. C. Yang, D. Maltby, *J. Chromatogr.* **1991**, *586*, 113.
- [71] C. T. Peng, *J. Chromatogr., A* **1994**, *678*, 189.
- [72] C. T. Peng, *J. Chromatogr., A* **2000**, *903*, 117.
- [73] V. V. Mihaleva, H. A. Verhoeven, R. C. H. de Vos, R. D. Hall, R. C. H. J. van Ham, *Bioinformatics* **2009**, *25*, 787.
- [74] K. Varmuza, W. Werther, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 323.
- [75] K. Varmuza, W. Werther, D. Henneberg, B. Weimann, *Rapid Communications in Mass Spectrometry* **1990**, *4*, 159.
- [76] S. Wolf, S. Schmidt, M. Mueller-Hannemann, S. Neumann, *BMC Bioinf.* **2010**, *11*, 148.
- [77] W. Kessler, *Multivariate Datenanalyse für die Pharma-, Bio-und Prozessanalytik (Wiley-VCH)* **2006**.
- [78] Bundesortenamt, *Landbuch Verlag Hannover* **2000**.
- [79] P. Schuhmann, *Bornimer Agrartechnische Berichte* **1997**, *13*.
- [80] H. Van den Dool, P. D. Kratz, *J. Chromatogr.* **1963**, *11*, 463.

- [81] M. Steinfath, D. Groth, J. Lisec, J. Selbig, *Physiol. Plant.* **2008**, *132*, 150.
- [82] J. Hummel, *Promotion* **2010**.
- [83] B. Junker, C. Klukas, F. Schreiber, *BMC Bioinf.* **2006**, *7*, 109.
- [84] R. A. van den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde, M. J. van der Werf, *BMC Genomics* **2006**, *7*, 142.
- [85] N. Strehmel, J. Hummel, A. Erban, K. Strassburg, J. Kopka, *Journal of Chromatography B* **2008**, *871*, 182.
- [86] J. Hummel, N. Strehmel, J. Selbig, D. Walther, J. Kopka, *Metabolomics* **2010**, *6*, 322.
- [87] S. Wicar, J. Novak, N. Ruseva-Rakshieva, *Anal. Chem.* **1971**, *43*, 1945.
- [88] J. M. Vergnaud, *J. Chromatogr.* **1965**, *19*, 495.
- [89] N. Hoffmann, J. Stoye, *Bioinformatics* **2009**, *25*, 2080.
- [90] K. J. Johnson, B. W. Wright, K. H. Jarman, R. E. Synovec, *J. Chromatogr., A* **2003**, *996*, 141.
- [91] E. Kovats, *Helv. Chim. Acta* **1958**, *41*, 1915.
- [92] M. L. Lee, D. L. Vassilaros, C. M. White, *Anal. Chem.* **1979**, *51*, 768.
- [93] J. Kopka, *J. Biotechnol.* **2006**, *124*, 312.
- [94] P. Bais, S. M. Moon, K. He, R. Leitao, K. Dreher, T. Walk, Y. Sucaet, L. Barkan, G. Wohlgermuth, M. R. Roth, E. S. Wurtele, P. Dixon, O. Fiehn, B. M. Lange, V. Shulaev, L. W. Sumner, R. Welti, B. J. Nikolau, S. Y. Rhee, J. A. Dickerson, *Plant Physiol.* **2010**, *152*, 1807.
- [95] C. Wagner, M. Sefkow, J. Kopka, *Phytochemistry* **2003**, *62*, 887.
- [96] H. Lu, Y. Liang, W. B. Dunn, H. Shen, D. B. Kell, *Trends Anal. Chem.* **2008**, *27*, 215.
- [97] A. Kerber, M. Meringer, C. Ruecker, *Croat. Chem. Acta* **2006**, *79*, 449.
- [98] D. B. Stauffer, F. W. McLafferty, R. D. Ellis, D. W. Peterson, *Anal. Chem.* **1985**, *57*, 1056.
- [99] F. Gan, J.-H. Yang, Y.-Z. Liang, *Anal. Sci.* **2001**, *17*, 635.
- [100] E. L. Schymanski, M. Meringer, W. Brack, *Anal. Chem.* **2009**, *81*, 3608.
- [101] S. E. Stein, D. R. Scott, *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 859.
- [102] C. Boelling, *Persönliche Mitteilung* **2009**.
- [103] C. Birkemeyer, A. Luedemann, C. Wagner, A. Erban, J. Kopka, *Trends Biotechnol.* **2005**, *23*, 28.
- [104] A. R. Fernie, P. Geigenberger, M. Stitt, *Curr. Opin. Plant Biol.* **2005**, *8*, 174.
- [105] J. Huege, R. Sulpice, Y. Gibon, J. Lisec, K. Koehl, J. Kopka, *Phytochemistry* **2007**, *68*, 2258.
- [106] U. Roessner-Tunali, J. Liu, A. Lisse, I. Balbo, A. Perez-Melis, L. Willmitzer, A. R. Fernie, *Plant J.* **2004**, *39*, 668.
- [107] U. Roessner, C. Wagner, J. Kopka, R. N. Trethewey, L. Willmitzer, *Plant J.* **2000**, *23*, 131.
- [108] K. Strassburg, D. Walther, H. Takahashi, S. Kanaya, J. Kopka, *OMICS - J Integr Biol* **2010**, *14*, 249.
- [109] P. D. Karp, C. A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahren, S. Tsoka, N. Darzentas, V. Kunin, N. Lopez-Bigas, *Nucleic Acids Res.* **2005**, *33*, 6083.
- [110] R. Caspi, H. Foerster, C. A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S. Y. Rhee, A. G. Shearer, C. Tissier, T. C. Walk, P. Zhang, P. D. Karp, *Nucl. Acids Res.* **2008**, *36*, D623.
- [111] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn G, *Nucleic Acids Res.* **2004**, *32*, D431.
- [112] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, *Nucleic Acids Res.* **2007**, *36*, D480.
- [113] I. M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, P. D. Karp, *Nucleic Acids Res.* **2005**, *33*, D334.
- [114] P. Zhang, H. Foerster, C. P. Tissier, L. Mueller, S. Paley, P. D. Karp, S. Y. Rhee, *Plant Physiol.* **2005**, *138*, 27.
- [115] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, *Nucleic Acids Res.* **2006**, *34*, D354.
- [116] B. H. Junker, C. Klukas, F. Schreiber, *BMC Bioinf.* **2006**, *7*, 109.
- [117] F. J. Zhang, Y. J. Jin, X. Y. Xu, R. C. Lu, H. J. Chen, *Phytochem. Anal.* **2008**, *19*, 560.
- [118] J. S. Fritz, *Analytical Solid-Phase Extraction (Wiley-VCH)*, **1999**.
- [119] E. Haghedooren, E. Farkas, A. Kerner, S. Dragovic, B. Noszal, J. Hoogmartens, E. Adams, *Talanta* **2008**, *76*, 172.
- [120] R. Zhou, T. M. Squires, S. J. Ambrose, S. R. Abrams, A. R. S. Ross, A. J. Cutler, *J. Chromatogr., A* **2003**, *1010*, 75.
- [121] M. Lopez-Carbonell, M. Gabasa, O. Jauregui, *Plant Physiol. Biochem.* **2009**, *47*, 256.
- [122] P. S. Blake, J. M. Taylor, W. E. Finch-Savage, *Plant Growth Regul.* **2002**, *37*, 119.

- [123] 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.* **1999**, *10*, 287.
- [124] H. Zhou, A. B. Kantor, C. H. Becker, *Metabolome Anal.* **2005**, 137.
- [125] A. Koulman, G. A. Lane, S. J. Harrison, D. A. Volmer, *Anal Bioanal Chem* **2009**, *394*, 663.
- [126] M.P. Styczynski, J.F. Moxley, L.V. Tong, J.L. Walther, K.L. Jensen, *Anal Chem.* **2007**, *79*, 966.
- [127] A. Kaefer, T. Lingner, K. Feussner, C. Gobel, I. Feussner, P. Meinicke, *BMC Bioinf.* **2009**, *10*, 92.
- [128] E. Pohjanen, E. Thysell, J. Lindberg, I. Schuppe-Koistinen, T. Moritz, P. Jonsson, H. Antti, *Metabolomics* **2006**, *2*, 257.
- [129] M. Steinfath, N. Strehmel, R. Peters, N. Schauer, D. Groth, J. Hummel, M. Steup, J. Selbig, J. Kopka, P. Geigenberger, J. T. van Dongen, *Plant Biotechnology Journal* **2010**, *8*, 1.
- [130] N. Strehmel, U. Praeger, C. König, I. Fehrlé, A. Erban, M. Geyer, J. Kopka, J. T. van Dongen, *Postharvest Biol. Technol.* **2010**, *56*, 109.
- [131] J. C. Hughues, *Ann. Appl. Biol.* **1980**, *96*, 344.
- [132] R. Peters, *Potato Res.* **1996**, *39*, 479.
- [133] P. E. Lærke, J. Christiansena, B. Veierskovb, *Postharvest Biology and Technology* **2002**, *26*, 99.
- [134] J. C. Partington, C. Smith, G. P. Bolwell, *Planta* **1999**, *207*, 449.
- [135] J. C. Stark, D. L. Corsini, P. J. Hurley, R. B. Dwelle, *American Journal of Potato Research* **1985**, *62*, 657.
- [136] L. H. Stevens, E. Davelaar, R. M. Kolb, E. J. M. Pennings, N. P. M. Smit, *Phytochemistry* **1998**, *49*, 703.
- [137] B. B. Dean, N. Jackowiack, S. Munck, *Potato Research* **1992**, *35*, 49.
- [138] W.-L. Yang, M. A. Bernards, *Metabolomics* **2007**, *3*, 147.
- [139] L. H. Stevens, E. Davelaar, *Phytochemistry* **1996**, *42*, 941.
- [140] D. L. Corsini, J. J. Pavek, B. Dean, *Am. Potato J.* **1992**, *69*, 423.
- [141] B. B. Dean, N. Jackowiak, M. Nagle, J. Pavek, D. Corsini, *Am. Potato J.* **1993**, *70*, 201.
- [142] E. Delgado, M. I. Sulaiman, E. Pawelzik, *Potato Research* **2001**, *44*, 207.
- [143] D. Tacke, *Persönliche Mitteilung* **2009**.
- [144] M. Scholz, S. Gatzek, A. Sterling, O. Fiehn, J. Selbig, *Bioinformatics* **2004**, *20*, 2447.
- [145] A. Aeppli, E. Keller, *Potato Res.* **1980**, *23*, 25.
- [146] L. H. Stevens, E. Davelaar, *J. Agric. Food Chem.* **1997**, *45*, 4221.
- [147] A. McGarry, C. C. Hole, R. L. K. Drew, N. Parsons, *Postharvest Biol. Technol.* **1996**, *8*, 239.
- [148] A.K. Thyboa, J. Christiansenb, K. Kaacka, M. A. Petersenc, *LWT - Food Science and Technology* **2006**, *39*, 166.
- [149] I. Rogozinska, J. Hippe, K. Müller, *Potato Research* **1986**, *29*, 239.
- [150] U. Sonnewald, *Trends Plant Sci.* **2001**, *6*, 333.
- [151] P. Delaplace, M.L. Fauconnier, K.I. Sergeant, J.F. Dierick, M. Oufir, F. van der Wal, A.H.P. America, J. Renaut, J.F. Hausman, P. d. Jardin, *J. Exp. Bot.* **2009**, *60*, 1273.
- [152] P. C. Struik, W. J. M. Lommen, A. J. Haverkort, R. M. J. Storey, *Potato Research* **2007**, *50*, 205.
- [153] A. D. Pavlista, *Nebraska Potato Eyes* **2004**, *16*, 1.
- [154] P. Delaplace, Y. Brostaux, M.-L. Fauconnier, P. du Jardin, *Postharvest Biology & Technology* **2008**, *50*, 103.
- [155] A. R. Fernie, L. Willmitzer, *Plant Physiol.* **2001**, *127*, 1459.
- [156] M.M.J. Claassens, J. Verhees, L.H.W. van der Plas, A. R. v. d. Krol, D. Vreugdenhil, *Journal of Experimental Botany* **2005**, *56*, 2515.
- [157] C. Matsuura-Endo, A. Ohara-Takada, Y. Chuda, H. Ono, H. Yada, M. Yoshida, A. Kobayashi, S. Tsuda, S. Takigawa, T. Noda, H. Yamauchi, M. Mori, *Biosci., Biotechnol., Biochem.* **2006**, *70*, 1173.
- [158] A. Ohara-Takada, C. Matsuura-Endo, Y. Chuda, H. Ono, H. Yada, M. Yoshida, A. Kobayashi, S. Tsuda, S. Takigawa, T. Noda, H. Yamauchi, M. Mori, *Biosci., Biotechnol., Biochem.* **2005**, *69*, 1232.
- [159] A. Becalski, B. P.-Y. Lau, D. Lewis, S. W. Seaman, S. Hayward, M. Sahagian, M. Ramesh, Y. Leclerc, *J. Agric. Food Chem.* **2004**, *52*, 3801.
- [160] M. L. A. T. M. Hertog, L. M. M. Tijskens, P. S. Hak, *Postharvest Biol. Technol.* **1997**, *10*, 67.
- [161] R.W. Blenkinsop, R.Y. Yada, *Horticultural Reviews* **2004**, *30*, 317.
- [162] W. Reust, J. Aerny, *Potato Res.* **1985**, *28*, 251.
- [163] L. J. Mitzel, N. R. Knowles, *Plant Physiol.* **1989**, *91*, 183.
- [164] M. L. A. T. M. Hertog, L. M. M. Tijskens, P. S. Hak, *Postharvest Biology & Technology* **2002**, *26*, 75.

- [165] A. W. Galston, *BioScience* **1983**, 33, 382.
- [166] H. V. Davies, *Am. Potato J.* **1990**, 67, 815.
- [167] T. J. Johansen, J. Nilsen, *Acta Agriculturae Scandinavica, Section B - Plant Soil Science* **2004**, 54, 185.
- [168] G. E. Kleinkopf, N. A. Oberg, N. L. Olsen, *Am. J. Potato Res.* **2003**, 80, 317.
- [169] B. Daniels-Lake, R. Prange, J. Nowak, S. Asiedu, J. Walsh, *Am. J. Potato Res.* **2005**, 82, 389.
- [170] R. Prange, B. Daniels-Lake, J.-C. Jeong, M. Binns, *Am. J. Potato Res.* **2005**, 82, 123.
- [171] W. Coleman, G. Lonergan, P. Silk, *Am. J. Potato Res.* **2001**, 78, 345.
- [172] W. Kalt, R. K. Prange, B. J. Daniels-Lake, J. Walsh, P. Dean, R. Coffin, *J. Food Process. Preserv.* **1999**, 23, 71.
- [173] K. Oosterhaven, K. J. Hartmans, J. J. C. Scheffer, *Potato Res.* **1995**, 38, 219.
- [174] K. Oosterhaven, K. J. Hartmans, J. J. C. Scheffer, L. H. W. van der Plas, *Physiol. Plant.* **1995**, 93, 225.
- [175] K. Oosterhaven, B. Poolman, E. J. Smid, *Ind. Crops Prod.* **1995**, 4, 23.
- [176] K. J. Hartmans, P. Diepenhorst, W. Bakker, L. G. M. Gorris, *Ind. Crops Prod.* **1995**, 4, 3.
- [177] K. J. Hartmans, K. Oosterhaven, L. G. M. Gorris, E. J. Smid, *Med. Aromat. Plants--Ind. Profiles* **1998**, 7, 175.
- [178] R. W. Blenkinsop, L. J. Copp, R. Y. Yada, A. G. Marangoni, *Food Res. Int.* **2002**, 35, 651.
- [179] A. Jumar, K. Sieber, W. Bollmann, *Nahrung* **1968**, 12, 469.
- [180] I. Boyd, J. Dalziel, H. Duncan, *Potato Res.* **1982**, 25, 51.
- [181] S. Biemelt, M. Hajirezaei, E. Hentschel, U. Sonnewald, *Potato Res.* **2000**, 43, 371.
- [182] A. Malkawi, B. L. Jensen, A. R. Langille, *J. Plant Growth Regul.* **2007**, 26, 308.
- [183] H. Wold, *London: Academic Press* **1975**.
- [184] I.-G. Chong, C.-H. Jun, *Chemom. Intell. Lab. Syst.* **2005**, 78, 103.
- [185] Bundessortenamt, *Landbuch Verlag Hannover* **2000**, 4.
- [186] J. W. Allwood, A. Erban, S. Koning, W. B. Dunn, A. Luedemann, A. Lommen, L. Kay, R. Loescher, J. Kopka, R. Goodacre, *Metabolomics* **2009**, 5, 479.
- [187] L. E. Rodriguez-Saona, R. E. Wrolstad, *Am. Potato J.* **1997**, 74, 87.
- [188] Y. Zhang, Y. Zhang, *Crit. Rev. Food Sci. Nutr.* **2007**, 47, 521.
- [189] N. Vertregt, *Eur. Potato J.* **1968**, 11, 34.
- [190] G. Bonner, *J. Adv. Nurs.* **2001**, 35, 349.
- [191] E. Blower Paul, P. Cross Kevin, *Curr Top Med Chem* **2006**, 6, 31.
- [192] P. H. Bach, J. W. Bridges, *Arch. Toxicol. Suppl.* **1985**, 8, 173.
- [193] B. E. Butterworth, S. R. Eldridge, *Prog. Clin. Biol. Res.* **1995**, 391, 49.
- [194] S. D. Brown, A. J. Myles, *Compr. Chemom.* **2009**, 3, 541.
- [195] H. J. Feldman, M. Dumontier, S. Ling, N. Haider, C. W. V. Hogue, *FEBS Letters* **2005**, 21, 4685.
- [196] C. Bicchi, A. D'Amato, I. Tonutti, M. Taccheo Barbina, *Chromatographia* **1985**, 20, 219.
- [197] T. A. Albanis, D. G. Hela, *J. Chromatogr., A* **1995**, 707, 283.
- [198] M. L. Bao, F. Pantani, K. Barbieri, D. Burrini, O. Griffini, *Int. J. Environ. Anal. Chem.* **1996**, 64, 23.
- [199] A. B. Deger, T. J. Gremm, F. H. Frimmel, *Acta Hydrochim. Hydrobiol.* **2000**, 28, 292.
- [200] J. W. Wong, K. Zhang, K. Tech, D. G. Hayward, C. M. Makovi, A. J. Krynitsky, F. J. Schenck, K. Banerjee, S. Dasgupta, D. Brown, *2010 J. Agric. Food Chem.*, 58, 5868.
- [201] K. Vandecasteele, I. Gaus, W. Debreuck, K. Walraevens, *Anal. Chem.* **2000**, 72, 3093.
- [202] M. D. Grynbaum, D. Kreidler, J. Rehbein, A. Pura, P. Schuler, W. Schaal, H. Czesla, A. Webb, V. Schurig, K. Albert, *Anal. Chem.* **2007**, 79, 2708.
- [203] M. Kuehnle, D. Kreidler, K. Holtin, H. Czesla, P. Schuler, W. Schaal, V. Schurig, K. Albert, *Anal. Chem.* **2008**, 80, 5481.
- [204] K. Skogerson, R. Runnebaum, G. Wohlgemuth, J. de Ropp, H. Heymann, O. Fiehn, *J. Agric. Food Chem.* **2009**, 57, 6899.
- [205] Y. Yin, A. Matsui, M. Sakuta, H. Ashihara, *Phytochemistry* **2008**, 69, 2891.
- [206] J. Gullberg, P. Jonsson, A. Nordstrom, M. Sjostrom, T. Moritz, *Anal. Biochem.* **2004**, 331, 283.
- [207] G. Colebatch, G. Desbrosses, T. Ott, L. Krusell, O. Montanari, S. Kloska, J. Kopka, M. K. Udvardi, *Plant J.* **2004**, 39, 487.
- [208] U. Roessner, J. H. Patterson, M. G. Forbes, G. B. Fincher, P. Langridge, A. Bacic, *Plant Physiol.* **2006**, 142, 1087.

**Tab. 1 Zusammenfassung aller interner Standards zur Standardisierung von GC-TOF-MS Analysen.** k. A.: keine Angabe (CAS Nummer wurde im SciFinder nicht gefunden.)

Substanz	CAS- Nummer	Lösungsmittel- gemisch	Mischungs- verhältnis [v/v/v]	Konzen- tration [ng/mL]
Glycerol-d8	7325-17-9	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
Ethanol-d4-amin	85047-08-1	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
Valin-d8	203784-63-8	H <sub>2</sub> O:MeOH	2:1	0.2
Leucin-d3	259225-39-3	H <sub>2</sub> O:MeOH	3:1	1.0
13C5,15N-Prolin	k.A.	H <sub>2</sub> O:MeOH	2:1	1.0
Lysin-d4	284664-88-6	H <sub>2</sub> O:MeOH	2:1	1.0
4-Aminobutan-d6-säure	14493-42-6	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
13C6,15N4-Arginin	202468-25-5	H <sub>2</sub> O:MeOH	5:1	1.0
Bernstein-d4-säure	14493-42-6	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
13C12-Saccharose	41055-68-9	H <sub>2</sub> O:MeOH	2:1	0.2
13C5-Glutamin	184161-19-1	H <sub>2</sub> O:MeOH	4:1	1.0
Putrescin-d4	88972-24-1	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
13C3,15N-Serin	202407-34-9	H <sub>2</sub> O:MeOH	4:1	0.2
Tryptophan-d5	90424-95-6	H <sub>2</sub> O:MeOH	6:1	1.0
Zitronen-d4-säure	147664-83-3	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
13C6-Glucose	110187-42-3	MeOH:H <sub>2</sub> O:CH <sub>3</sub> Cl	2.5 : 1 : 1	0.2
Isoascorbinsäure	89-65-6	H <sub>2</sub> O	-	0.5
d4-Alanin	53795-92-9	H <sub>2</sub> O	-	1.0
2-13C, 15N2-Uracil	181517-11-3	MeOH	-	0.2



**Tab. 2 Übersicht über sämtliche verwendete charakteristische Fragmentmassen und Retentionsindices der identifizierten Analyten und der erwarteten unbekanntem Signale.** Die Angaben in Klammern ermöglichen eine Ableitung der Derivatstruktur. TMS: Trimethylsilylfunktionalität; MEOX: Methoxyaminfunktionalität; NP: Nebenprodukt; HP: Hauptprodukt

Analytname	Be- zeichner	RI	Fragmentmassen
1-Kestose (11TMS)	A340003	3361.28	169; 217; 361; 437
2-Deoxygalactose (1MEOX) (4TMS) NP	A178013	1779.11	142; 174; 205
2-Deoxyglucose (1MEOX) (4TMS) NP	A178016	1775.82	142; 174; 205
2-Deoxyglucose (1MEOX) (4TMS) HP	A178014	1778.12	105; 174; 205; 276; 307
2-Ethylhexansäure (1TMS)	A116005	1155.41	117; 129; 145; 160; 201
2-Hydroxypyridin (1TMS)	A105003	1031.31	97; 122; 136; 152; 167
2-Methylfumarsäure (2TMS)	A140010	1397.06	231
2-Methyläpfelsäure (3TMS)	A148001	1464.98	203; 247
2-Oxoglutarsäure (1MEOX) (2TMS) NP	A154007	1544.66	186; 198; 229; 288; 304
2-Oxoglutarsäure (1MEOX) (2TMS) HP	A158004	1572.85	186; 198; 229; 288; 304
2-Piperidinecarbonsäure (1TMS)	A127012	1271.16	84; 103; 158; 186
3-Caffeoylchinasäure (6TMS)	A299001	2981.68	219; 255; 324; 345
3-Caffeoylchinasäure (6TMS)	A311001	3099.59	219; 255; 324; 345; 397
3-Cyanoalanin (2TMS)	A138005	1371.75	130; 141; 202; 215; 243
3-Hydroxypyridin (1TMS)	A114003	1136.97	92; 122; 136; 152; 167
4-Aminobutansäure (2TMS)	A133002	1297.39	102; 115; 232
4-Aminobutansäure (3TMS)	A153003	1527.46	100; 174; 216; 246; 304
4-Caffeoylchinasäure (6TMS)	A317001	3154.54	219; 255; 307; 324
4-Hydroxybenzoessäure (2TMS)	A164003	1633.29	193; 223; 267
4-Hydroxybutansäure (2TMS)	A126002	1230.32	117
4-Hydroxycimtsäure (2TMS)	A195001	1947.53	179; 219; 249; 293; 308
5-Caffeoylchinasäure (6TMS)	A319001	3177.08	219; 255; 307; 345
6-Aminocaprinsäure (3TMS)	A174006	1725.06	96; 100; 174; 242; 332
Aconitsäure (3TMS)	A176002	1741.1	211; 215; 229; 285; 375

Analytname	Be- zeichner	RI	Fragmentmassen
Adenin (2TMS)	A188005	1873.97	165; 192; 237; 264; 279
Adipo-2,6-lactam (2TMS)	A151013	1513.09	170; 197; 244; 272; 287
Alanin (2TMS)	A110001	1087.41	100; 116; 190; 218
Allantoin (3TMS)	A209001	2068.66	189; 243; 259; 359
Arginin [-NH3] (3TMS)	A183001	1816.31	142; 157; 256
Ascorbat (4TMS)	A195002	1937.86	117; 303; 332; 449
Asparagin (2TMS)	A161001	1599.53	116; 130; 159; 186; 261
Asparagin (3TMS)	A168001	1666.44	116; 159; 188; 231; 258
Asparagin (4TMS) NP1	A164007	1622.63	188; 216; 231
Asparagin [-H2O] (2TMS)	A151002	1500.35	100; 115; 128; 201; 243
Asparaginsäure (2TMS)	A144003	1422.39	117; 130; 160; 202; 245
Asparaginsäure (3TMS)	A152002	1511.8	202; 218; 232; 306; 334
Azelainsäure (2TMS)	A181007	1790.39	117; 129; 201; 317
Benzoessäure (1TMS)	A128003	1250.39	105; 135; 179; 194
beta-Alanin (3TMS)	A144001	1424.39	100; 160; 174; 248; 290
Butyro-1,4-lactam (1TMS)	A114004	1140.45	100; 112; 142; 157
Kaffeessäure (3TMS)	A199001	1977.47	191; 219; 396
Kaffeessäure (3TMS)	A214001	2135.55	219; 307; 381; 396
Calystegin B2 (1MEOX) (4TMS) HP	A181006	1798.43	167; 257; 288; 371; 461
Calystegin B2 (4TMS)	A198005	1968.78	217; 244; 259; 284
Calystegin B2 (5TMS)	A213006	2128.17	229; 316; 331
Calystegin B4 (4TMS)	A198010	1968.78	217; 244; 259; 284
Calystegin A3 (1MEOX) (3TMS) HP	A176008	1742.65	167; 185; 216; 257
Carbodiimid (2TMS)	A100005	951.59	141; 155; 171; 186
Cellobiose (8TMS)	A278010	2768.05	191; 204; 217; 361
Zitronensäure (4TMS)	A182004	1803.92	183; 211; 257; 273; 375

Analytname	Be- zeichner	RI	Fragmentmassen
Cystein (3TMS)	A156002	1549.81	100; 116; 218; 220
Cystin (4TMS)	A241002	2290.87	100; 146; 218; 266
Cytidin (4TMS)	A272005	2697.58	168; 223; 245; 256
Cytosin (2TMS)	A153005	1526.63	240; 254
Decan	A100003	1000	85; 99; 113; 142
Decansäure (1TMS)	A147004	1456.08	117; 145; 229
Dehydroascorbat (2MEOX) HP	A185002	1838.71	157; 173; 231; 245; 316
Diethylenglycol (2TMS)	A127003	1235.42	101; 103; 117; 191; 235
Docosan	A220001	2200	85; 99; 113; 310
Dodecan	A120001	1200	85; 99; 113; 170
Dodecansäure (1TMS)	A166003	1654.1	117; 145; 257
Dotriacontan	A320001	3200	85; 99; 113
Eicosansäure (1TMS)	A245001	2453.82	117; 145
Erythritol (4TMS)	A150002	1493.27	217
Erythronsäure (4TMS)	A154001	1529.19	205; 220; 292
Ethanolamin (3TMS)	A128002	1260.25	86; 100; 174; 188; 262
Ferulasäure (2TMS)	A192002	1919.19	249; 308; 323
Ferulasäure (2TMS)	A210001	2097.85	249; 293; 308; 323; 338
Fructose (1MEOX) (5TMS) NP	A188004	1863.13	217; 277; 307; 335; 364
Fructose (1MEOX) (5TMS) HP	A187002	1853.97	217; 277; 307; 335; 364
Fructose-6-phosphat (1MEOX) (6TMS) HP	A232002	2292.48	315; 357
Fucose (1MEOX) (4TMS) HP	A173002	1716.3	117; 160
Fumarsäure (2TMS)	A137001	1346.94	115; 143; 217; 245
Schleimsäure (6TMS)	A204001	2030.47	292; 305; 333
Galactinol (9TMS)	A299002	2966.29	169; 191; 204; 305; 433
Galactose (1MEOX) (5TMS) HP	A188001	1876.46	160

Analytname	Be- zeichner	RI	Fragmentmassen
Gentiobiose (1MEOX) (8TMS) NP	A285003	2814.35	204
Gentiobiose (1MEOX) (8TMS) HP	A283011	2788.87	204; 480
Gluconsäure (6TMS)	A200001	1984.78	157; 292; 305; 319; 333
Glucopyranose (5TMS)	A189005	1879.95	191; 204
Glucopyranose (5TMS)	A198006	1959.24	117; 129; 191; 204; 217
Glucopyranose [-H <sub>2</sub> O] (4TMS)	A177004	1755.75	217; 232
Glucose (1MEOX) (5TMS) NP	A191001	1898.54	160; 229; 305; 319; 343
Glucose (1MEOX) (5TMS) HP	A189002	1878.29	160; 229; 305; 319; 343
Glucose-6-phosphat (1MEOX) (6TMS) NP	A235002	2328.97	160; 299; 357
Glucose-6-phosphat (1MEOX) (6TMS) HP	A233002	2307.34	160; 299; 357; 387; 471
Glucose-6-phosphat (6TMS)	A239001	2370.02	204; 387
beta-1,6-Anhydroglucose (3TMS)	A172001	1704.48	204; 217; 243; 333
Glutaminsäure (2TMS)	A154002	1528.2	84; 158; 174; 230; 276
Glutaminsäure (3TMS)	A163001	1614.59	128; 156; 246; 348; 363
Glutamin (3TMS)	A178001	1766.95	156; 203; 245; 347; 362
Glutamin (4TMS)	A174008	1720.58	156; 203; 227; 317
Glutamin [-H <sub>2</sub> O] (2TMS) HP	A155001	1538.9	139; 155; 229; 257; 272
Glutamin [-H <sub>2</sub> O] (3TMS) HP	A160005	1590.97	154; 211; 227; 301; 344
Glutamin [-H <sub>2</sub> O] (2TMS) NP	A148002	1469.96	139; 155; 167; 229; 257
Glycerinsäure (3TMS)	A135003	1320.33	133; 189; 205; 292; 307
Glycerol (3TMS)	A129003	1262.29	205; 293
Glycerol-3-phosphat (4TMS)	A177002	1750.51	211; 299; 315; 357
Glycin (2TMS)	A114001	1109.07	102; 176; 204
Glycin (3TMS)	A133001	1302.28	86; 100; 174; 248; 276
Glykolsäure (2TMS)	A106002	1061.49	161; 177; 205
Guanin (3TMS)	A214002	2132.12	99; 352; 367

Analytname	Be- zeichner	RI	Fragmentmassen
Hexadecansäure (1TMS)	A205001	2045.44	117; 145; 201; 313; 328
Hexadecanol (1TMS)	A193005	1958.69	97; 111; 299
Hexadecensäure (1TMS)	A202001	2021.64	145
Hexatriacontan	A360001	3600	85; 99; 113
Histidin (3TMS)	A192006	1914.53	154; 218; 254; 356
Homoserin (2TMS)	A137002	1355.33	103; 130; 146; 158
Homoserin (3TMS)	A146001	1441.73	128; 218
Hydroxylamin (3TMS)	A110002	1104.75	119; 133; 146; 161; 249
Isoascorbinsäure (4TMS)	A196002	1947.97	117; 303; 332; 449
Isoleucin (1TMS)	A119002	1174.98	86; 130; 146; 170; 188
Isoleucin (2TMS)	A132002	1286.54	102; 158; 218; 232; 260
Itaconsäure (2TMS)	A135004	1339.09	133; 215; 230; 259
Milchsäure (2TMS)	A105001	1044.68	117; 133; 191; 219
Leucin (1TMS)	A118001	1151.06	86; 130; 146; 170; 188
Leucin (2TMS)	A129002	1264.43	102; 158; 232; 260
Lysin (3TMS)	A170007	1703.69	156; 230; 258; 362
Lysin (3TMS)	A186002	1849.23	156; 174; 200; 258; 362
Lysin (4TMS)	A192003	1913.61	156; 174; 230; 317; 434
Lyxose (1MEOX) (4TMS) HP	A165006	1650.19	189; 307
Äpfelsäure (3TMS)	A149001	1479.34	217; 233; 245; 307; 335
Maltitol (9TMS)	A284001	2811.28	204; 345; 361
Maltose (1MEOX) (8TMS) NP	A277002	2745.49	160; 204; 271; 319; 361
Maltose (1MEOX) (8TMS) HP	A274001	2719.8	160; 204; 319
Maltotriose (1MEOX) (11TMS) NP	A358001	3543.33	169; 204; 217; 271; 361
Melibiose (1MEOX) (8TMS) HP	A287003	2837.47	160; 204; 361
Methionin (1TMS)	A142007	1416.63	104; 130; 131; 178; 221

Analytname	Be- zeichner	RI	Fragmentmassen
Methionin (2TMS)	A152001	1516.11	128; 176; 202; 250; 293
myo-Inositol (6TMS)	A209002	2080.27	191; 265; 305; 318; 507
myo-Inositol-1-phosphat (7TMS)	A243001	2413.92	217; 299; 315; 318; 387
myo-Inositol-2-phosphat (7TMS)	A243003	2413.92	217; 299; 315; 318; 387
myo-Inositol-2-phosphat (7TMS)	A247001	2465.39	217; 299; 315; 318; 387
Siloxan	A104001	1020.15	117; 191; 207; 295
Siloxan	A113001	1116.99	267; 355
Kohlensäure (1MEOX) (2TMS)	A114002	1129.12	133; 160; 220; 235
Siloxan	A123002	1208.56	249; 265; 281; 369
Siloxan	A130003	1285.35	207; 325; 341; 429
Diethanolamin (2TMS)	A131001	1282.13	144; 146
Siloxan	A131003	1294.51	207; 325; 341; 429
NA	A134006	1341.44	100; 143; 171; 186
Methylfuroesäure (1TMS)	A135002	1324.59	109; 111; 183
Lumichrom (2MEOX)	A136002	1333.77	134; 174; 184; 285
NA	A139006	1361.68	83; 113; 139; 157; 185
NA	A139007	1382.96	168; 240; 254
NA	A141005	1386.95	103; 133; 191
NA	A142003	1401.97	103; 144; 189; 203; 234
NA	A142005	1407.89	86; 100; 174
NA	A143002	1411.36	100; 102; 154; 156; 174
NA	A143003	1421.19	100; 102; 154; 156; 174
NA	A144006	1424.59	86; 155; 174
Asparaginsäure (2TMS)	A144007	1430.84	117; 130; 160; 202; 245
NA	A145005	1436.41	86; 168; 174
NA	A147001	1459.41	130; 156; 174; 227

Analytname	Be- zeichner	RI	Fragmentmassen
NA	A147011	1467.04	103; 117; 173; 262
NA	A155004	1544.38	117; 129
Dichlorobenzoesäure (1TMS)	A157004	1560.35	109; 173; 175; 247; 249
Asparagin	A157008	1558.47	162; 218; 261
Asparagin	A157012	1582.22	162; 218; 261
Weinsäure (4TMS)	A160001	1591.93	219; 292
Dihydroxyphenol (3TMS)	A161002	1599.61	239; 342
NA	A163004	1619.87	113; 174; 203; 230
NA	A164004	1631.1	98; 129
3,4,5 Trihydroxypentansäure (4TMS)	A165004	1639.17	217; 231; 245
NA	A166002	1643.84	117; 204; 217; 256; 291
NA	A167004	1659.51	275
Glucose-1,6-anhydro (3TMS)	A168015	1683.93	204; 217; 243; 333
Xylitol (5TMS)	A170001	1689.66	307; 319
NA	A171003	1693.23	155; 171; 244; 275
NA	A172005	1703.42	131; 156; 173; 229; 275
NA	A174005	1723.88	163; 227; 276; 302; 392
NA	A174015	1735.71	89; 133; 191;
NA	A175008	1740.64	227; 276; 302; 392
NA	A175010	1744.06	347
Ribonsäure (5TMS)	A176001	1743.81	217; 292; 307; 333
Fructosederivat	A180002	1789.79	117; 205; 217; 257; 437
Fructosederivat	A181004	1797.54	117; 217; 230; 257; 437
Fructosederivat	A181011	1798.21	117; 217; 230; 257; 437
Fructosederivat	A182009	1806.54	437
Dehydroascorbinsäure (2MEOX) NP	A188008	1867.53	245

Analytname	Be- zeichner	RI	Fragmentmassen
NA	A191007	1895.95	117; 189; 273; 363
NA	A196004	1947.09	217; 243; 271; 332; 361
NA	A196022	1959.08	205; 229
NA	A197001	1963.95	169; 217; 332
Glucopyranose (5TMS)	A197003	1958.74	117; 129; 191; 204; 217
Gluconsäure (6TMS)	A199003	1980.48	157; 292; 305; 319; 333
NA	A200003	1993.19	133
NA	A201002	2004.37	217; 292; 305; 333
NA	A203003	2014.28	103; 157; 204; 233; 319
NA	A208008	2076.85	117
NA	A211001	2098.25	157; 205; 229; 262; 319
NA	A212004	2107.21	245; 257; 292; 347
NA	A217004	2161.61	204
NA	A217010	2162.48	117; 230; 364
NA	A221004	2197.69	211; 227; 299
Glycerolgalactopyranosid (6TMS)	A231002	2298.49	129; 204; 217
NA	A237001	2361.47	204; 217; 292; 305; 375
NA	A239008	2388.38	204; 217; 292; 305; 375
NA	A245002	2438.23	245; 360; 430
NA	A249001	2475.4	156; 217; 446
NA	A250001	2484.6	117; 204; 217; 321
NA	A252002	2502.8	156; 217; 343; 415; 446
NA	A254002	2530.01	204; 219
NA	A257001	2556.82	104; 121; 167; 279
Glucopyranosylglucopyranosid (8TMS)	A262001	2609.79	169; 191; 217; 331; 361
NA	A271003	2706.57	243; 271



Analytname	Be- zeichner	RI	Fragmentmassen
NA	A272010	2715.88	169; 217; 243; 271; 361
Glucosylglucose (1MEOX) (8TMS)	A273003	2714.64	160; 169; 204; 319; 361
Maltose (1MEOX) (8TMS)	A274004	2727.82	169; 204; 217; 243; 361
NA	A275004	2736.16	169; 217; 243; 361
NA	A276001	2748.75	160; 204; 243; 305
NA	A278002	2766.73	169; 204; 217; 243; 361
Stearinsäure 2,3-dihydroxypropylester (2TMS)	A278004	2774.85	129; 203
NA	A279001	2774.27	204; 243; 305; 361
NA	A279002	2779.33	129
NA	A281001	2798.45	204; 217; 243
Isomaltose (1MEOX) (8TMS)	A286002	2846.18	160; 204; 361
NA	A292006	2920.84	204; 217; 343; 433; 479
Galactinol (9TMS)	A293001	2911.72	129; 169; 204; 217; 361
NA	A294002	2925.78	169; 191; 204
NA	A294006	2941.14	191
Galactinol (9TMS)	A300001	2987.64	169; 191; 204; 305; 433
Galactinol (9TMS)	A301005	3005.15	191; 204; 305; 361; 433
NA	A308003	3063.41	217
NA	A311002	3099.24	204; 217; 243; 337
NA	A331004	3296.63	204; 217
NA	A334002	3332.44	169; 204; 217; 361
NA	A354003	3534.42	191; 204; 217; 271; 361
Nicotinsäure (1TMS)	A133004	1301.8	78; 106; 136; 180; 195
Nonadecan	A190001	1900	85; 99; 113; 268
Nonadecansäuremethylester	A223004	2233.19	143
Nonansäure (1TMS)	A138003	1360.69	117; 129; 145; 215

Analytname	Be- zeichner	RI	Fragmentmassen
Octacosan	A280001	2800	85; 99; 113
Octadecadiensäure, 9,12-(Z,Z)- (1TMS)	A221003	2209.91	117; 129
Octadecan	A180008	1800	85; 99; 113; 254
Octadecansäure (1TMS)	A225002	2243.53	117; 129; 145; 341; 356
Octadecatriensäure, 6,9,12-(Z,Z,Z)- (1TMS)	A219008	2190.76	129;
Ornithin (3TMS)	A162001	1613.15	142; 204; 216; 243; 348
Ornithin (3TMS)	A176006	1750.77	142; 174; 186; 244; 348
Oxalsäure (2TMS)	A113002	1118.32	133; 175; 190; 219
Pentadecan	A150001	1500	85; 99; 113; 212
Phenylalanin (1TMS)	A157001	1556.05	91; 120; 130; 146; 204
Phenylalanin (2TMS)	A164001	1629.63	91; 192; 218; 266; 294
Phosphat (3TMS)	A129001	1261.98	211; 225; 283; 299; 314
Pyridin-2-carbonsäure (1TMS)	A136003	1328.67	106; 136; 180
Prolin (1TMS)	A117006	1176.03	103; 172
Prolin (2TMS)	A132003	1295.77	142;
Prolin [+CO <sub>2</sub> ] (2TMS)	A159001	1583.69	142; 186; 216
Putrescin (3TMS)	A151005	1506.15	142; 174; 304
Putrescin (4TMS)	A175002	1736.7	100; 174; 200; 214; 361
5-Oxoprolin (1TMS)	A151009	1508.58	84; 140; 158; 186;
5-Oxoprolin (2TMS)	A153002	1521.73	140; 156; 230; 258; 273
Chinasäure (5TMS)	A185001	1842.69	255; 334; 345; 419; 537
Raffinose (11TMS)	A337002	3350.64	204; 217; 361; 437; 451
Ribitol (5TMS)	A173001	1712.74	205; 217; 307; 319; 422
Ribonsäure (5TMS)	A177001	1751.51	103; 217; 292; 307; 333
Ribulose (1MEOX) (4TMS) HP	A166012	1667.1	263;
Zuckersäure (6TMS)	A201001	2000.79	292; 305; 333; 373

Analytname	Be- zeichner	RI	Fragmentmassen
Saccharose (8TMS)	A264001	2621.76	157; 319; 361; 437; 451
Serin (2TMS)	A128001	1253.25	116; 132; 159; 219; 234
Serin (3TMS)	A138001	1352.2	100; 204; 218; 278; 306
Siloxan	A147007	1461.04	281; 327; 341; 415
Siloxan	A168009	1686.33	207; 221; 295; 341; 369
Siloxan	A177007	1772.25	267; 281; 355; 429
Siloxan	A183010	1829.05	207; 221; 295; 341; 369
Siloxan	A197008	1975.86	207; 221; 295; 341; 369
Siloxan	A211012	2114.13	207; 221; 295; 341; 369
Siloxan	A212009	2117.39	207; 221; 295; 341; 369
Siloxan	A219007	2188.57	221; 267; 281; 355; 429
Siloxan	A225011	2256.16	207; 221; 295; 341; 369
Siloxan	A233004	2332.33	221; 281; 355; 429; 503
Siloxan	A240010	2399.32	207; 221; 295; 341; 369
Siloxan	A247006	2467.67	221; 281; 355; 429; 503
Siloxan	A253006	2534.88	207; 221; 295; 341; 369
Siloxan	A259009	2596.41	221; 281; 355; 429; 503
Siloxan	A266005	2661.38	207; 221; 295
Siloxan	A271010	2716.38	221; 281; 355; 429; 503
Siloxan	A278009	2780.83	207; 221; 295; 341; 369
Siloxan	A291007	2911.31	207; 221; 295; 369
Siloxan	A295011	2959.8	221; 281; 355; 429; 503
Siloxan	A307006	3079.47	221; 281; 355; 429; 503
Siloxan	A318007	3192.09	221; 281; 355; 429; 503
Sorbitol (6TMS)	A193001	1919.41	157; 307; 319

Analytname	Be- zeichner	RI	Fragmentmassen
Spermidin (4TMS)	A220002	2193.66	160; 174; 200
Spermidin (5TMS)	A226002	2251.32	116; 144; 156; 174
Bernsteinsäure (2TMS)	A134001	1310.9	129; 172; 247
Weinsäure (4TMS)	A164006	1628.94	219; 292
Tetradecansäure (1TMS)	A185004	1845.77	117; 285
Threitol (4TMS)	A149002	1485.23	205; 217
Threonsäure (4TMS)	A156001	1545.94	117; 205; 220; 292; 319
Threonin (2TMS)	A132001	1290.45	117; 130; 146; 219; 248
Threonin (3TMS)	A140001	1377.23	117; 218; 219; 291; 320
Threose (1MEOX) (3TMS) NP	A146005	1427.5	117; 161; 205; 233
Threose (1MEOX) (3TMS) HP	A147008	1445.9	117; 161; 205; 233
Trehalose-, alpha,alpha (8TMS)	A274002	2726.3	169; 191; 243; 331; 361
Tridecansäure	A176012	1749.4	117; 145; 201; 271
Tryptophan (2TMS)	A223002	2214.69	202; 218; 291; 303
Tryptophan (3TMS)	A223001	2217.55	202; 218; 291; 303
Tyramin (3TMS)	A191004	1913.02	174;
Tyrosin (2TMS)	A189006	1887.23	179; 310
Tyrosin (3TMS)	A194002	1934.12	100; 179; 218; 280; 354
Uracil (2TMS)	A136001	1335.82	99; 113; 126; 241; 255
Harnsäure (4TMS)	A211006	2094.62	456; 441; 382
Valin (1TMS)	A109001	1081.68	130; 146; 156; 174
Valin (2TMS)	A122001	1207.1	100; 144; 156; 218; 246
Vanillinsäure (2TMS)	A178004	1766.24	223; 267; 297
Xylitol (5TMS)	A171001	1694.65	217
Xylulose (1MEOX) (4TMS) HP	A168008	1665.55	173; 205; 263

**Tab. 3 Qualitative und quantitative Zusammensetzung von „Mix 2“ zur Evaluierung der Robustheit des Retentionsindex und zur Verifizierung der Identifizierungen**

Metabolit	CAS Nummer	Konzentration [ng/μL]
1-O-beta-Methylgalactopyranosid	1824-94-8	266.5
1-O-beta-Methylglucopyranosid	709-50-2	266.5
1,3-Dihydroxyaceton	96-26-4	26.7
1,6-Anhydroglucose	498-07-7	26.7
2-Amino-2-deoxyglucose	3416-24-8	266.5
2-Methyläpfelsäure	6236-09-5	22.3
2-Methylfumarsäure	498-24-8	22.3
2-Oxoglutarsäure	328-50-7	1841.4
3-Hydroxypyridin	109-00-2	222.9
4-Hydroxybutansäure	591-81-1	22.3
6 Oxopiperidincarbonsäure	3770-22-7	2822.6
6-Aminocaprinsäure	60-32-2	656.4
Adenin	73-24-5	183.2
Alanin	56-41-7	656.4
alpha,alpha- Trehalose	99-20-7	196.9
Anthranilsäure	118-92-3	22.3
Äpfelsäure	97-67-6	222.9
Arginin	74-79-3	2822.6
Asparagin	70-47-3	656.4
Asparaginsäure	56-84-8	656.4
Azelainsäure	123-99-9	1841.4
Benzoessäure	65-85-0	8.9
Bernsteinsäure	110-15-6	22.3
beta- Alanin	107-95-9	65.6

<b>Metabolit</b>	<b>CAS Nummer</b>	<b>Konzentration [ng/μL]</b>
Brenztraubensäure	127-17-3	222.9
Chinasäure	77-95-2	22.3
Chlorogensäure	202650-88-2	55.8
cis- Aconitsäure	585-84-2	22.3
Cystein	3374-22-9	656.4
Cytosin	65-46-3	1831.8
Ethanolamin	141-43-5	183.2
Fructose	57-48-7	266.5
Fructose-6-phosphat	643-13-0	26.7
Fucose	7724-73-4	26.7
Fumarsäure	110-17-8	22.3
GABA	56-12-2	656.4
Glucarsäure	25525-21-7	266.5
Gluconsäure	526-95-4	26.7
Gluconsäure-1,4-lacton		266.5
Gluconsäure-6-phosphat	921-62-0	266.5
Glucose	50-99-7	266.5
Glucose-6-phosphat	56-73-5	266.5
Glucuronsäure	6556-12-3	26.7
Glutamin	56-85-9	656.4
Glutaminsäure	56-86-0	656.4
Glyceraldehyd	56-82-6	266.5
Glycerinsäure	6000-40-4	10.7
Glycerol	56-81-5	26.7
Glycin	56-40-6	656.4
Guanin	73-40-5	183.2

<b>Metabolit</b>	<b>CAS Nummer</b>	<b>Konzentration [ng/μL]</b>
Harnsäure	69-93-2	183.2
Histidin	4998-57-6	2822.6
Homoserin	672-15-1	65.6
Leucin	61-90-5	65.6
Lysin	56-87-1	656.4
Malonsäure	141-82-2	22.3
Maltitol	585-88-6	98.5
Maltose	69-79-4	196.9
Maltotriose	1109-28-0	344.6
Melibiose	585-99-9	344.6
Methionin	63-68-3	65.6
Milchsäure	50-21-5	8.9
myo-Inositol	87-89-8	399.8
myo-Inositol-1-phosphat	573-35-3	7595.6
Nicotinamid	98-92-0	22.3
Nicotinsäure	59-67-6	22.3
Orotsäure	65-86-1	22.3
Oxalsäure	144-62-7	222.9
Phenylalanin	63-91-2	65.6
Prolin	147-85-3	65.6
Putrescin	110-60-1	183.2
Raffinose	512-69-6	196.9
Ribitol	488-81-3	10.7
Ribonsäure	17812-24-7	26.7
Ribose	50-69-1	26.7
Ribose-5-phosphat	4300-28-1	26.7

<b>Metabolit</b>	<b>CAS Nummer</b>	<b>Konzentration [ng/<math>\mu</math>L]</b>
Ribulose	5556-48-9	266.5
Saccharose	57-50-1	196.9
Salicylsäure	69-72-7	22.3
Serin	56-45-1	65.6
Shikimisäure	138-59-0	22.3
Sorbitol	50-70-4	26.7
Spermidin	124-20-9	183.2
Threitol	7493-90-5	10.7
Threonin	72-19-5	65.6
Threonsäure	3909-12-4	266.5
Threose	29884-64-8	2665.1
trans- Kaffeesäure	501-16-6	55.8
trans-4-Hydroxyzimtsäure	501-98-4	55.8
trans-Ferulasäure	537-98-4	55.8
trans-Sinapinsäure	7362-37-0	22.3
Tryptophan	73-22-3	65.6
Tyrosin	60-18-4	65.6
Uracil	66-22-8	183.2
Valin	72-18-4	65.6
Zitronensäure	77-92-9	222.9



**Tab. 4 Lösungsmittelgradient zur Fraktionierung von Brassica Spezies für die Versuchsreihe der gepaarten NMR und GC-TOF-MS Analyse**

<b>Fraktion</b>	<b>Zeit [min]</b>	<b>Acetonitrilgehalt [%]</b>	<b>Wassergehalt [%]</b>
A	0-2	70	30
B	2-4	70	30
C	4-5	70	30
D	5-6	70	30
E	6-7	70	30
F	7-8	70	30
G	8-10	70	30
H	10-20	10	90
I	20-30	10	90
K	30-60	10	90
L	61-70	70	30

**Tab. 5 Übersicht über sämtliche Fettsäuremethylester zur Verifizierung des Retentionsindexmodells.** Der Retentionsindex wurde auf der Basis von Alkanen berechnet.

Substanz	Retentionsindex	CAS-Nummer	Endkonzentration [ng/μL]
Methyldecansäure	1327.08	110-42-9	800
Methyldocosansäure	2548.56	929-77-1	400
Methyldodecansäure	1526.31	111-82-0	800
Methyleicosansäure	2342.06	1120-28-1	400
Methylhexacosansäure	2941.17	5802-82-4	400
Methylmyristinsäure	1729.57	124-10-7	800
Methylnonansäure	1225.27	1731-84-6	800
Methyloctacosansäure	3140.52	55682-92-3	400
Methyloctansäure	1120.62	111-11-5	800
Methylpalmitinsäure	1928.52	112-39-0	800
Methylstearinsäure	2131.68	112-61-8	400
Methyltetracosansäure	2739.86	2442-49-1	400
Methyltriacontansäure	3343.88	629-83-4	400

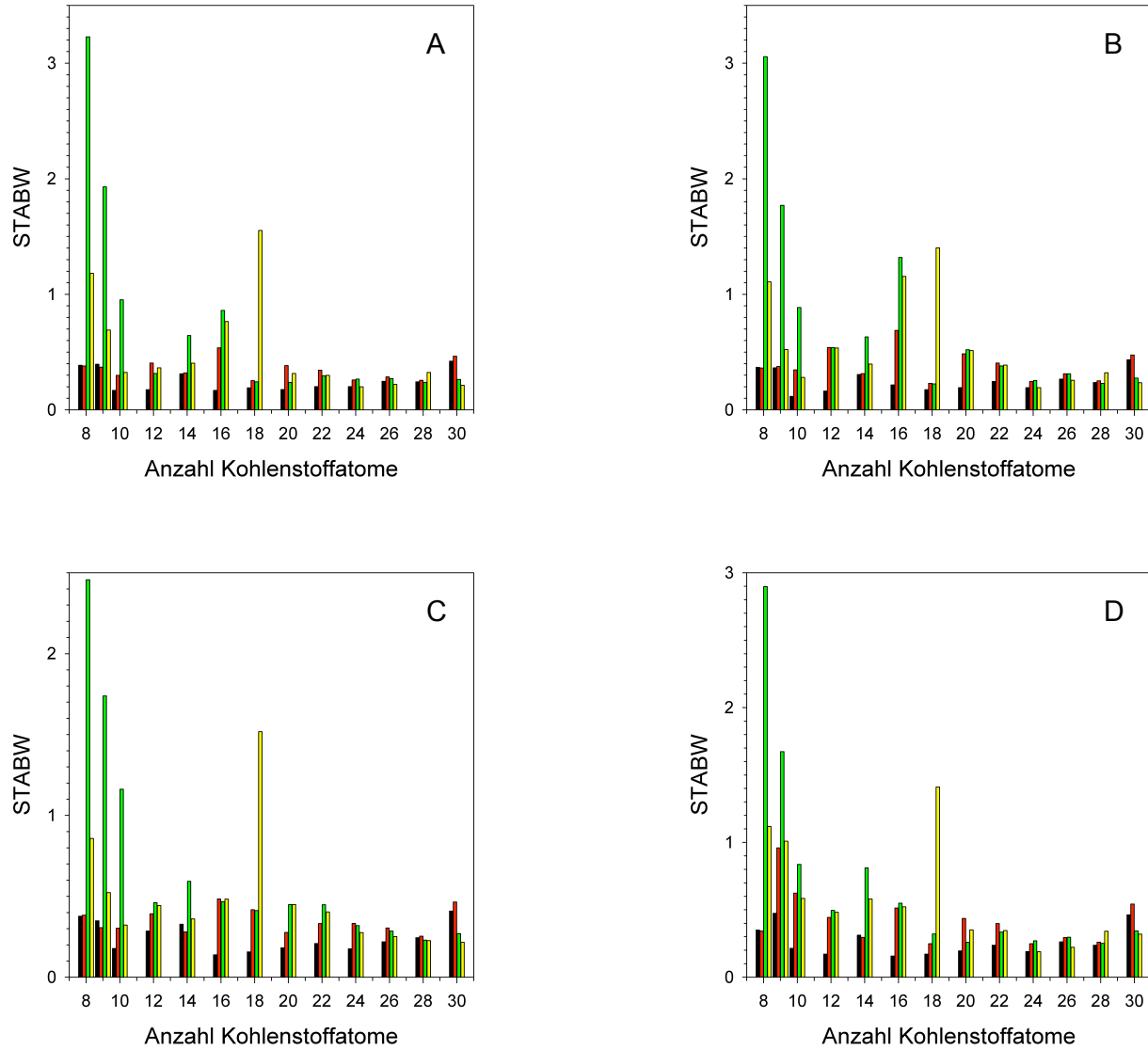
**Tab. 6 Qualitative Zusammensetzung von „Mix 1“ zur Evaluierung der Robustheit des Retentionsindex**

<b>Substanz</b>	<b>CAS Nummer</b>	<b>Endkonzentration [mg/L]</b>
2-Methyläpfelsäure	626-10-8	5
2-Oxoglutarsäure	328-50-7	10.1
3-Hydroxypyridin	109-00-2	9.9
4-Hydroxybenzoesäure	99-96-7	9.9
Alanin	338-69-2	10.3
Cholesterol	57-88-5	10
Fucose	2438-80-4	10.3
Glucose	492-62-6	10
Glycin	56-40-6	10.1
Isoleucin	443-79-8	9.9
Kaffeensäure	331-39-5	10.3
Lactitol	81025-04-9	5.1
Lactose	5965-66-2	5.2
Lanosterol	79-63-0	10.5
Maltose	6363-53-7	20.8
Maltotriose	1109-28-0	10.2
Palatinose	13718-94-0	10.1
Pantothensäure	137-08-6	9.9
Putrescin	333-93-7	9.6
Ribitol	488-81-3	10
Ribose	50-69-1	10.3
Sorbitol	50-70-4	5.3
Sorbose	414-273-3850	9.9
Stigmasterol	83-48-7	9.6
Zitronensäure	77-92-9	10

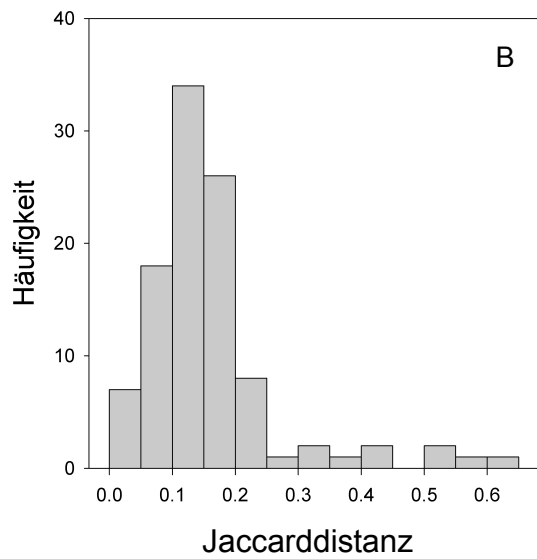
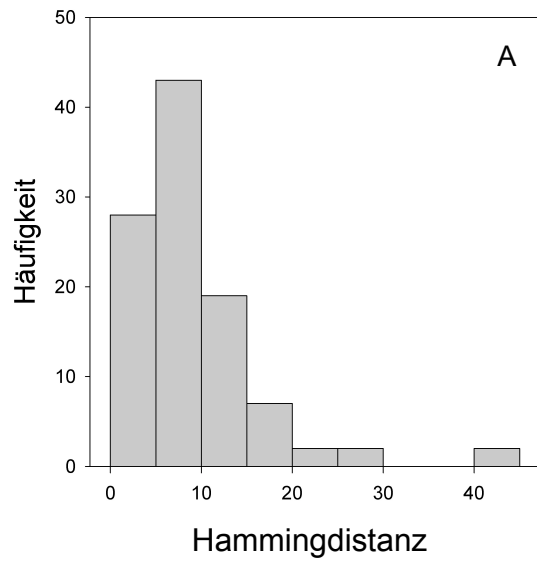
Tab. 7 Zusammensetzung des Mixes zur Etablierung der Vorfraktionierung der Brassica Spezies.

<b>Substanz</b>	<b>CAS Nummer</b>
Erythritol	149-32-6
Fructose	57-48-7
Fumarsäure	110-17-8
GABA	56-12-2
Galactose	59-23-4
Glucose	50-99-7
Mannose	3458-28-4
Nicotinsäure	59-67-6
Phenylalanin	63-91-2
Sorbitol	50-70-4
trans-Ferulasäure	537-98-4
Uracil	66-22-8
Valin	72-18-4
Zitronensäure	77-92-9

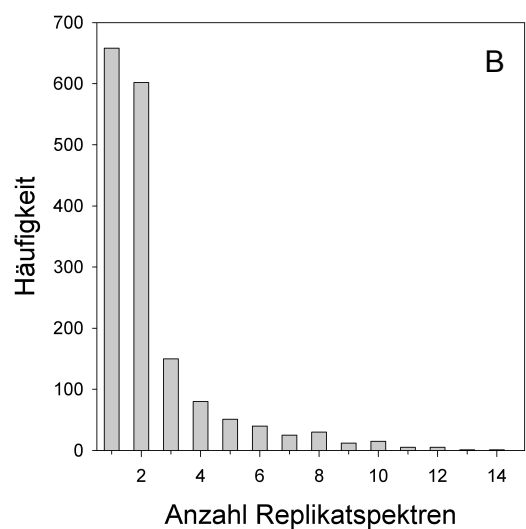
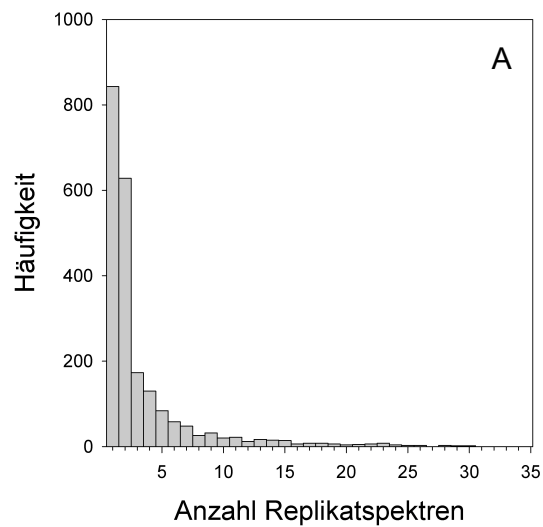
**Abb. 1 Standardabweichung des Retentionsindex in Bezug auf die zugegebenen Fettsäuremethylester im Hinblick auf a) van den Dool, b) die Spline Funktion, c) lineare Regression und d) Polynomiale Regression fünfter Ordnung in Mix 1 (schwarz), Mix 2 (rot), Reisblatt (grün) und Bäckerhefe (gelb) für das System MDN35**



**Abb. 2 Vergleich der massenspektralen Abgleichsparameter A) Hammingdistanz und B) Jaccarddistanz in Bezug auf alle identifizierten Analyten einer mittelkomplexen Probe („Mix 2“)**



**Abb. 3 Anzahl von Replikaspektren zur Durchschnittsspektrenberechnung für das VAR5 System innerhalb der GoIm Metabolome Database. A) bezüglich aller Spektren und B) bezüglich der TOF Spektren**



**Abb. 4 Optimierung des A) Elutionsvolumens und des B) Anreicherungsvolumens für das Sorbens HLB für die Zielmoleküle ABA (○), GA3 (□) und IAA (△). Die Elutionsvolumina wurden jeweils auf die gleiche Konzentration angepasst: 1.5 mL Anreicherungsvolumen und 1.5 mL Elutionsvolumen. Die horizontale Linie zeigt die erwartete Wiederfindungsrate an.**

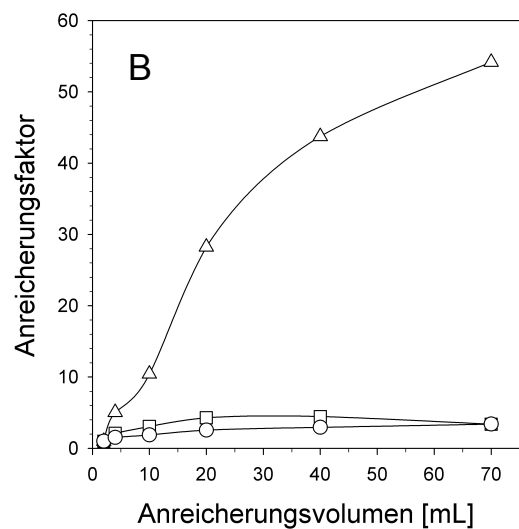
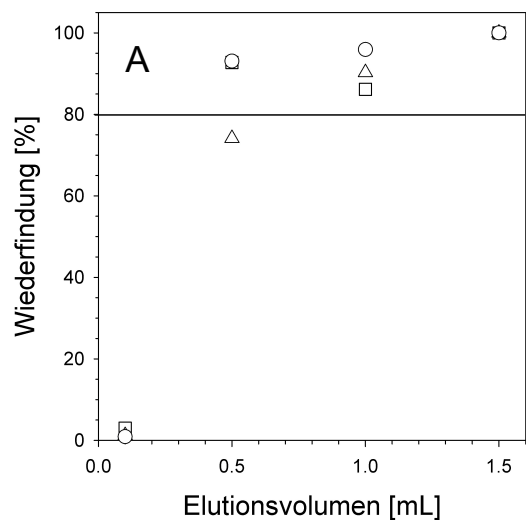




Abb. 5 Änderungen im Primärmetabolismus nach mechanischer Belastung im Kultivar Afra (blau) und im Kultivar Milva (orange). Die signifikanten metabolischen Änderungen bezüglich der Faktoren mechanische Belastung und Zeit sind jeweils durch eine graue Box hervorgehoben.

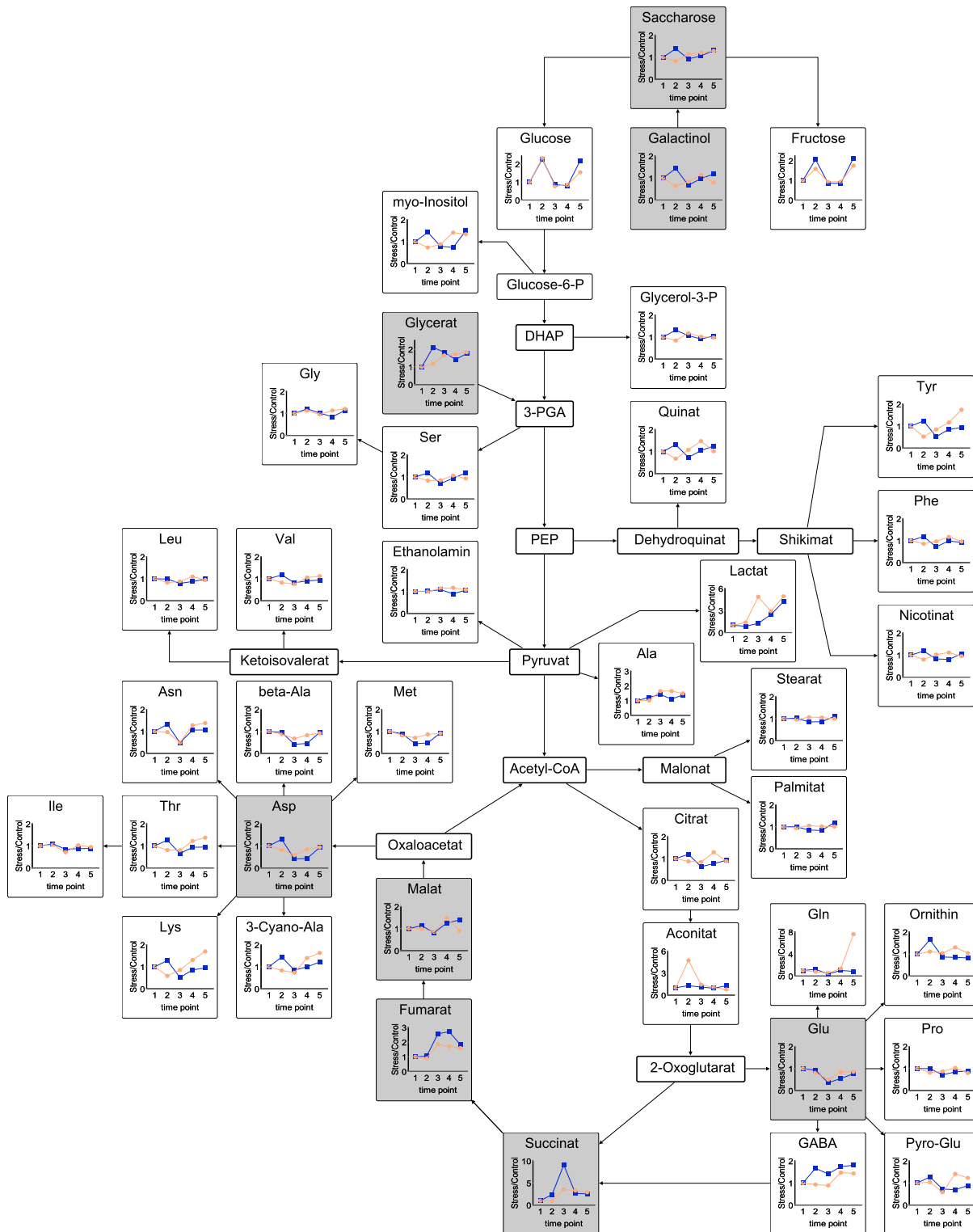
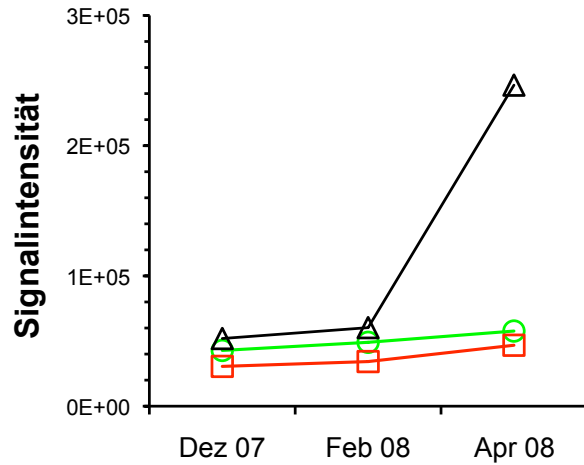


Abb. 6 Relative zeitliche Änderung der Phytohormone SA (A) und ABA (B) während der Lagerung von Kartoffelknollen des Kultivars *Marabel* nach Behandlung mit Kümmelöl (Kreise), nach chemischer Behandlung (Vierecke) und nach keiner Behandlung (Dreiecke).

A)



B)

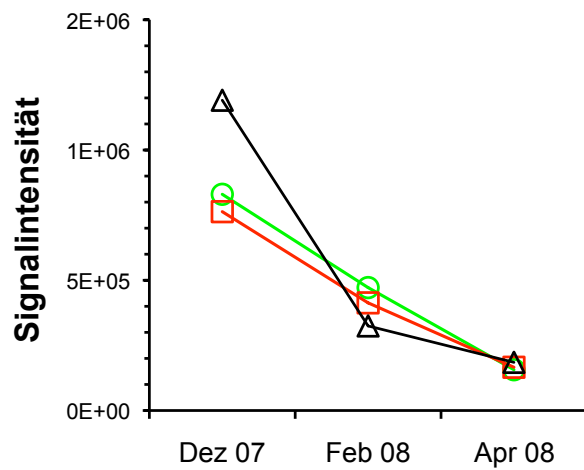
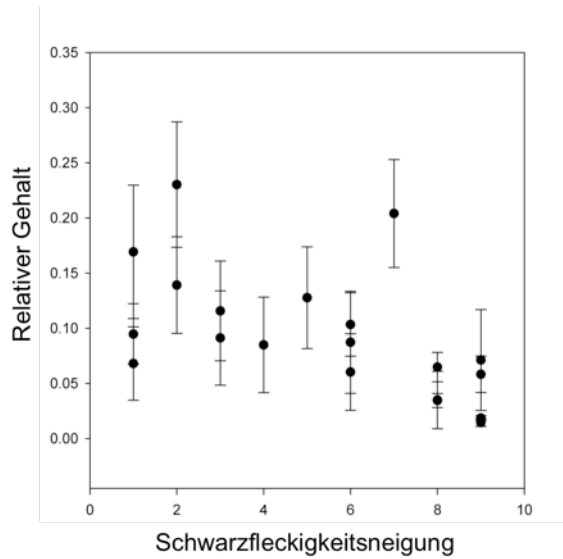
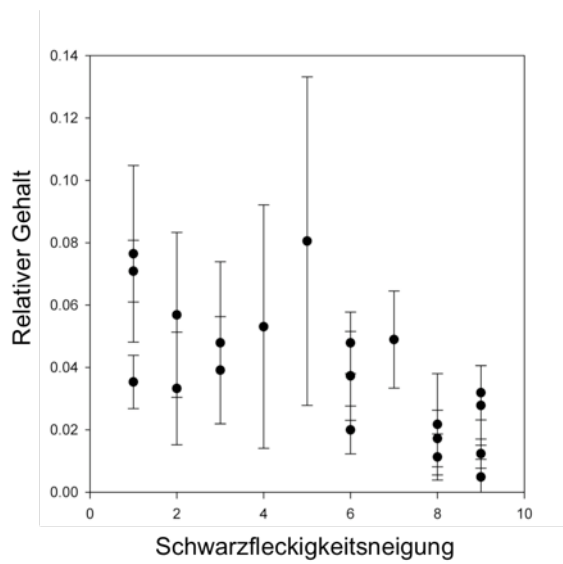


Abb. 7 Relativer A) Tyrosin-, B) Glutamin-, C) Threonin-, D) Serin- und E) Valingehalt in den 20 untersuchten Kultivaren für den Standort Böhlendorf.

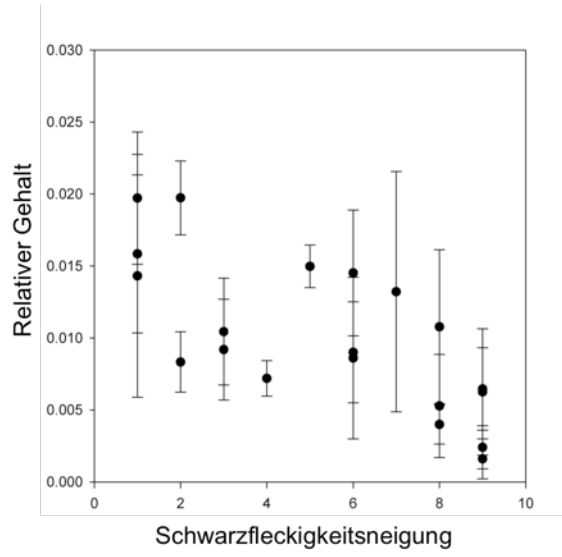
A)



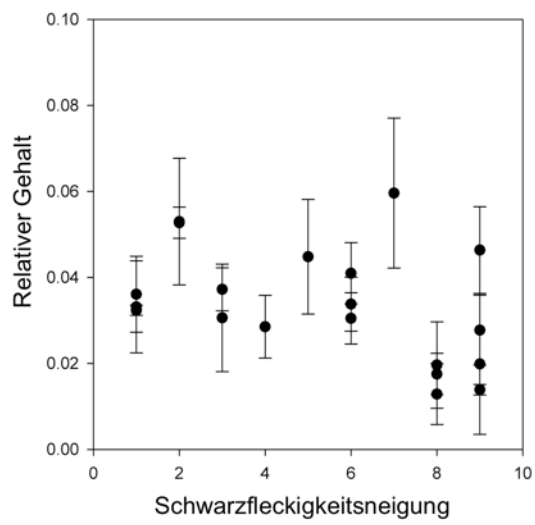
B)



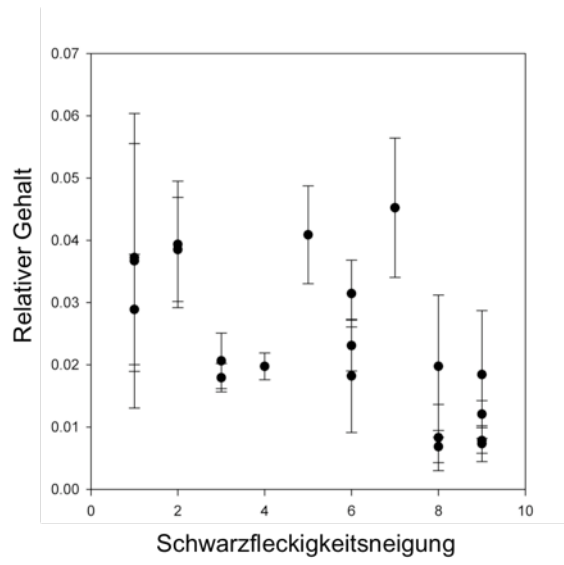
C)



D)



E)



**Abb. 8 Entscheidungsbäume zur Klassifizierung von Massenspektren für das VAR 5 System basierend auf Metabolitstrukturen.** Das Symbol X in einer grünen Box zeigt an, dass die entsprechende Fragmentmasse während der Trainingsphase extrahiert wurde.

Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim. Alkoh.	Prim. Aliph. Amin	Prim. Amin	Sec. Alkohol	
70															X						
71			X							X								X	X		
72										X	X										
76					X					X	X							X			
77																		X			
78								X													
82								X													
83								X					X								
85							X														
86						X								X							
87											X										
88													X								
89			X			X	X		X	X	X								X	X	
92								X									X				

Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureder	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim Alkohol	Prim Aliph Amin	Prim Amin	Sec Alkohol		
93							X																
94												X											
95					X									X									
97									X														
98			X					X							X	X							
99						X	X													X	X		
100			X		X	X	X								X					X	X		
101										X													
102																			X	X			
103	X		X							X	X				X				X				X
104			X									X			X								
105							X							X									
109										X													
112			X												X								X
113										X													
115													X										

Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim. Alkoh.	Prim. Aliph. Amin	Prim. Amin	Sec. Alkohol
117	X																			X
119					X													X		
123	X				X															
124													X							
125										X										
126								X												
127							X													
128						X												X		
129					X			X							X					X
130																		X	X	
132				X						X										
133						X	X	X			X		X							
143							X					X						X	X	
145																			X	
146	X			X			X											X	X	
150						X														



Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim. Alkoh.	Prim. Aliph. Amin	Prim. Amin	Sec. Alkohol
152																	X			
155					X															
156											X									
160			X						X	X	X									
163																			X	
167									X						X					
169										X			X		X					
171					X														X	
172								X	X											
173			X																	
174			X				X												X	
176	X																			X
179								X							X		X			
183			X																	
185				X																
190														X						

Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureder	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim Alkhol	Prim Aliph Amin	Prim Amin	Sec Alkohol	
191															X	X						X
193															X							
197										X												
202								X	X													
204	X		X					X														
205									X	X					X			X				
206				X							X											
211									X					X								
213								X								X						
217	X		X	X					X	X			X		X							X
218						X					X											
220						X																
221															X							
222							X															
231			X																			
233																					X	

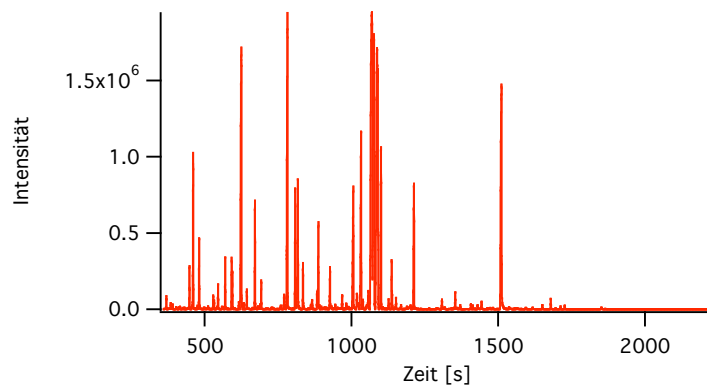
Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureder	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim Alkhol	Prim Aliph Amin	Prim Amin	Sec Alkohol	
238								X			X			X								
239								X														
240														X								
243																		X				
247											X											
250					X																	
254								X														X
271									X					X								
273	X																					
277				X								X										
279																X						
280								X								X						
284											X											
292										X	X											
293																						X
295															X							

Fragment	Diol	Acetal	Alkohol	Aldehyd	Alken	Aminosäure	Amin	Aromat	Carbonyl	Carbonsäure	Carbonsäureder	Carbonsäureester	Hemiacetal	Heterocyclus	Hydroxy	Phenol	Phosphat	Prim Alkohol	Prim Aliph Amin	Prim Amin	Sec Alkohol
299											X						X				
304									X												
305			X																		
307												X									
308					X				X	X											
309					X																
314									X												
316																	X				
319													X	X							
320																		X			
321			X																		X
324												X									
331	X																				X
334												X									
335																					X
357			X																		

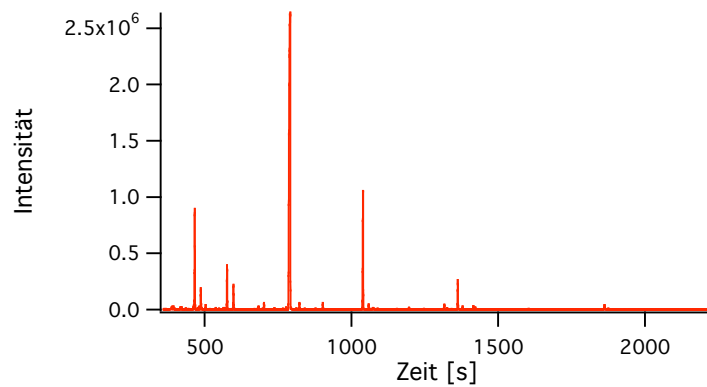
F r a g m e n t	D i o l	A c e t a l	A l k o h o l	A l d e h y d	A l k e n	A m i n o s ä u r e d	A m i n	A r o m a t	C a r b o n y l	C a r b o n s ä u r e	C a r b o n s ä u r e d e r	C a r b o n s ä u r e e s t e r	H e m i a c e t a l	H e t e r o c y c l u s	H y d r o x y	P h e n o l	P h o s p h a t	P r i m A l k o h o l	P r i m A l i p h A m i n	P r i m A m i n	S e c A l k o h o l	
361		X																				
372																						X
373	X																					
388																						X

**Abb. 9 Komplexität einzelner Fraktionen am Beispiel der Matrix „GD33“ nach präparativer Fraktionierung.**

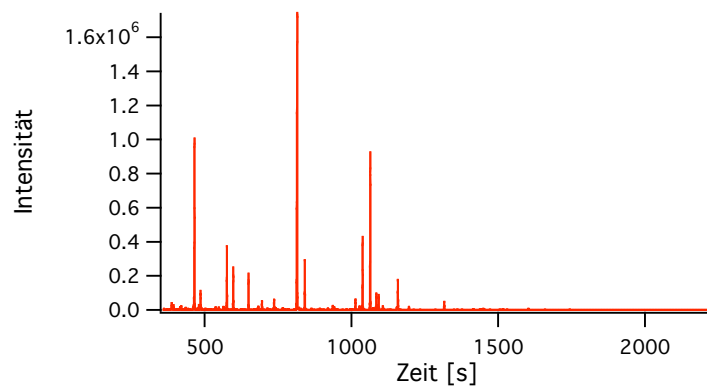
**- keine Fraktionierung**

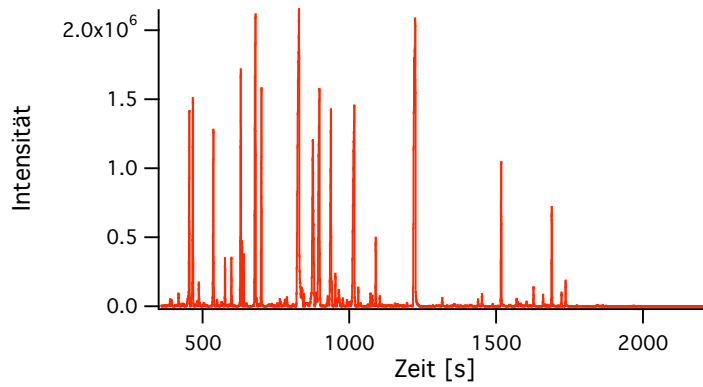
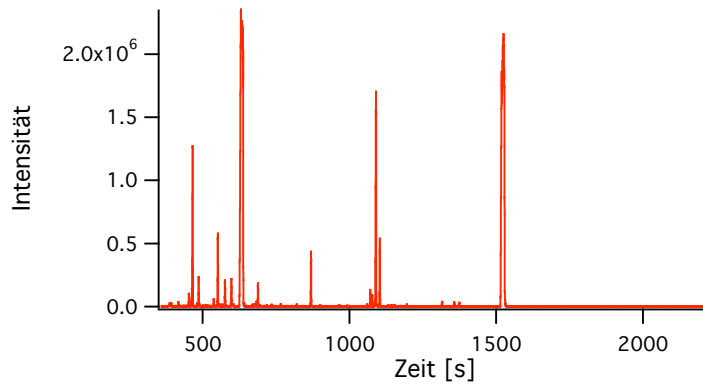
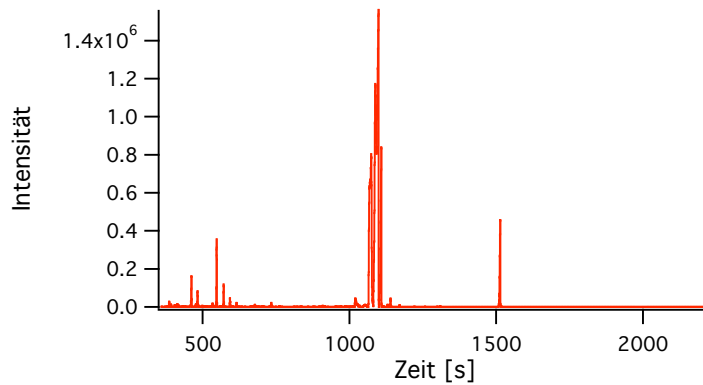


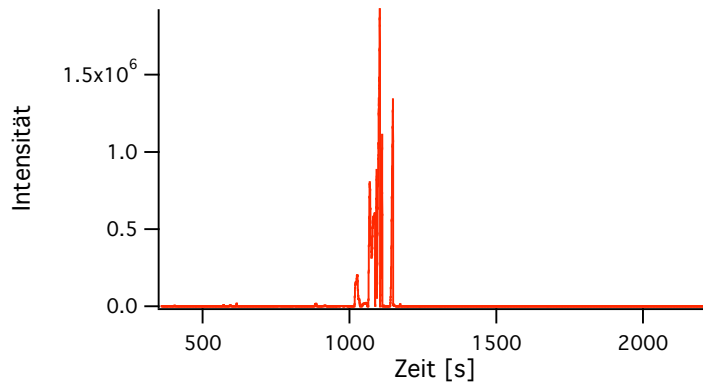
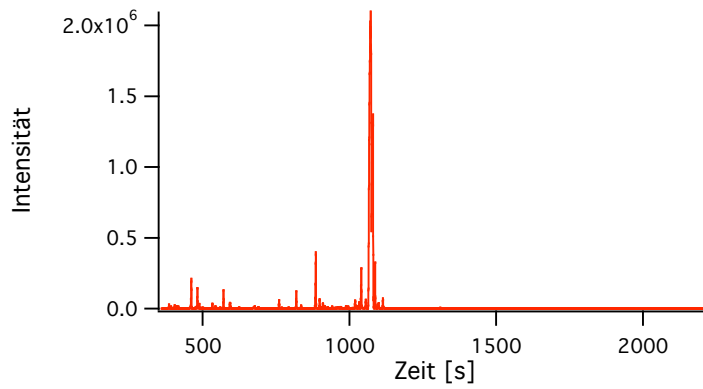
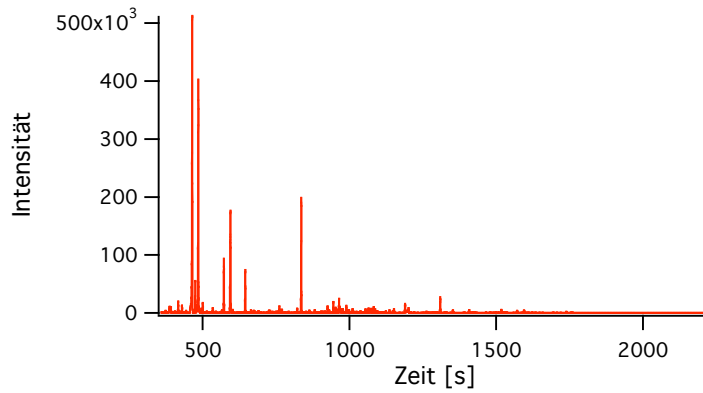
**- Fraktion K**



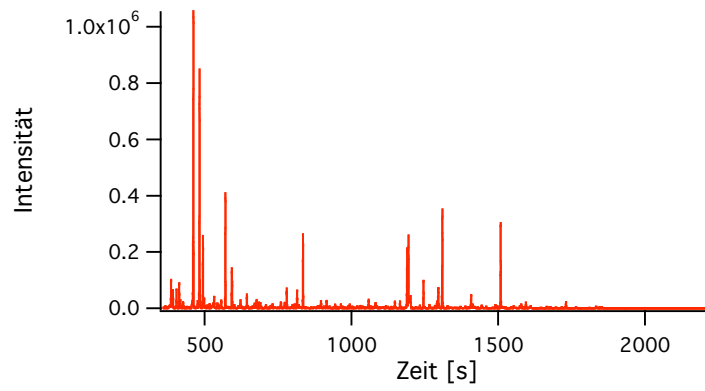
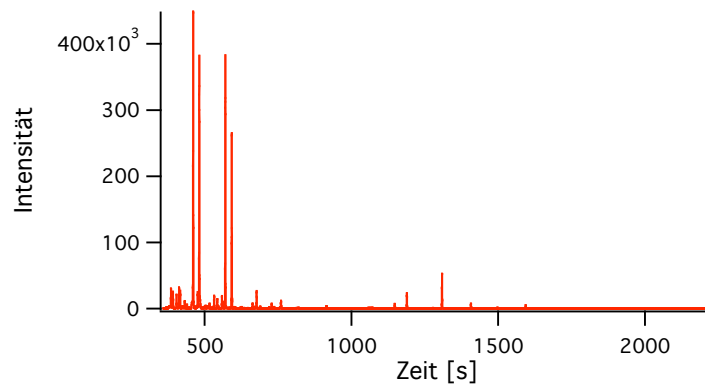
**- Fraktion I**



**Fraktion H****- Fraktion G****- Fraktion F**

**- Fraktion E****- Fraktion D****- Fraktion C**



**- Fraktion B****- Fraktion A**

Ausgewählte Ergebnisse dieser Arbeit wurden in folgenden Journalen veröffentlicht und sind im Folgenden zusammengefasst.

- Seite 145

**Strehmel N**, Hummel J, Erban A, Strassburg K, Kopka J (2008) Retention index thresholds for compound matching in GC-MS metabolite profiling. *Journal of Chromatography B* 871: 182-190.

- Seite 154

**Strehmel N**, Praeger U, König C, Fehrle I, Erban A, Geyer M, Kopka J, van Dongen J (2010) Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impact. *Postharvest Biology and Technology*, 56(2): 109-116.

- Seite 162

Steinfath M, **Strehmel N**, Peter R, Groth D, Hummel J, Steup M, Selbig J, Kopka J, van Dongen J (2010) Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach *Plant Biotechnology Journal*, 8: 900–911.

- Seite 174

Hummel J, **Strehmel N**, Selbig J, Walther D, Kopka J (2010) Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics*, 6(2):322-333.

- Seite 186

Kempa S, Hummel J, Schwemmer T, Pietzke M, **Strehmel N**, Wienkoop S, Kopka J, Weckwerth W (2009) An automated GCxGC-TOF-MS protocol for batch-wise extraction and alignment of mass isotopomer matrixes from differential <sup>13</sup>C-labelling experiments: a case study for photoautotroph-mixotroph grown *Chlamydomonas reinhardtii* cells. *Journal of Basic Microbiology* 49(1): 82-91.



Contents lists available at ScienceDirect

## Journal of Chromatography B

journal homepage: [www.elsevier.com/locate/chromb](http://www.elsevier.com/locate/chromb)Retention index thresholds for compound matching in GC–MS metabolite profiling<sup>☆</sup>

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<sub>10</sub>–C<sub>36</sub>). The precision of the RI determination of 13 exemplary fatty acid methyl esters (C<sub>8</sub> ME–C<sub>30</sub> 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,  $RI_{\text{predicted}} - RI_{\text{determined}}$  ( $n > 64$ ). 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.

© 2008 Elsevier B.V. All rights reserved.

## 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 match-

ing 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

<sup>☆</sup> 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](mailto:kopka@mpimp-golm.mpg.de) (J. Kopka).

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  $\pm 4.81$  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-

**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
Panthenic 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

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<sub>8</sub> ME–C<sub>30</sub> ME standards [27]. Variations of these basic sets of retention marker mixtures are reported below (cf. Section 2.4.10).

### 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  $\mu$ L, was dried using a VR Maxi vacuum concentrator with rotors R96-13 or R120-111 (Jouan Nordic, Allerød, 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, *Saccharomyces 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<sub>595</sub>)  $\sim$ 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 <math>-20\text{ }^\circ\text{C}</math>. Cell pellets were re-suspended and extracted (1) 15 min at 70 °C using 374  $\mu\text{L}$  extraction medium comprising 350  $\mu\text{L}$  methanol, 12  $\mu\text{L}$  internal standard 1 and 12  $\mu\text{L}$  internal standard 2 (cf. below), (2) 10 min at 30 °C using 263  $\mu\text{L}$  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  $\mu\text{L}$  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- $\text{d}_4$ -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  $\mu\text{L}$  methanol, and 30  $\mu\text{L}$  of internal standard 1 and internal standard 2, respectively. Finally 600  $\mu\text{L}$  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  $\mu\text{L}$  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  $\mu\text{L}$  methoxyamine reagent, i.e. 40 mg/mL methoxyamine hydrochloride (CAS: 593-56-6) in pyridine. Then 90  $\mu\text{L}$  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  $\mu\text{L}$  of 100  $\mu\text{L}$  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  $\mu\text{m}$  film thickness, 30 m  $\times$  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  $\mu\text{m}$ , 30 m  $\times$  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  $\mu\text{m}$ , 30 m  $\times$  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  $\mu\text{m}$ , 10 m  $\times$  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  $\mu\text{L}$  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, Bellefonte, 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<sub>12</sub> to C<sub>25</sub>. 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.

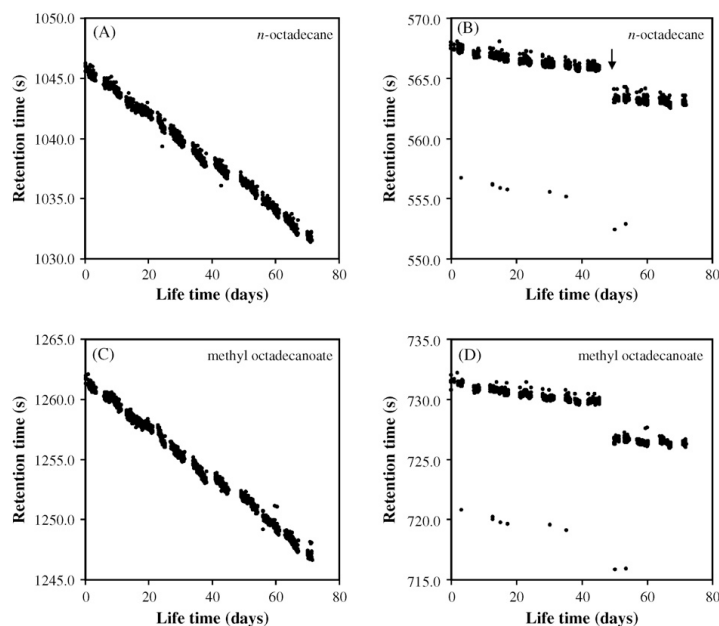
### 3. Results and discussion

#### 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<sup>a</sup>

Method of calculation	Precision of RI determination (average standard deviation)							
	Average of C <sub>10</sub> –C <sub>36</sub> <i>n</i> -alkanes				Average of C <sub>8</sub> –C <sub>30</sub> Fatty Acid Methyl esters			
	Low complexity	DRM-mix	DRM-yeast	DRM-rice	Low complexity	DRM-mix	DRM-yeast	DRM-rice
<b>Chromatography variant 1</b>								
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	<b>0.1060</b>	<b>0.0957</b>	<b>0.1056</b>	<b>0.5721</b>	<b>0.2225</b>	<b>0.2494</b>	<b>0.2520</b>	<b>1.3407</b>
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					<b>0.2128</b>	<b>0.2326</b>	<b>0.2510</b>	<b>1.1070</b>
<b>Chromatography variant 9</b>								
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	<b>0.0975</b>	<b>0.0782</b>	<b>0.2446</b>	<b>0.2957</b>	<b>0.2513</b>	<b>0.2416</b>	<b>0.4892</b>	<b>0.7296</b>
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					<b>0.3278</b>	<b>0.2527</b>	<b>0.5269</b>	<b>0.7490</b>

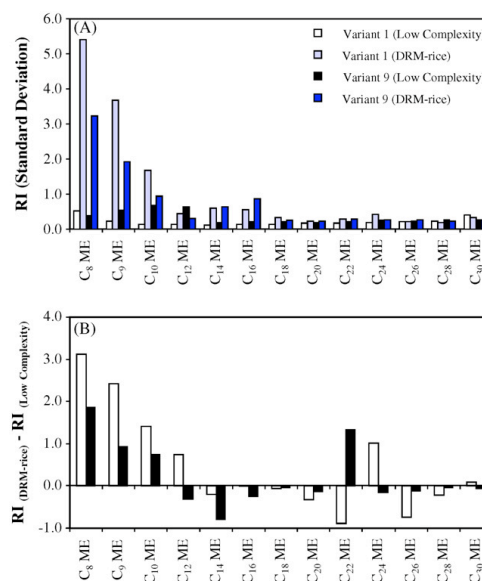
<sup>a</sup> Regression and interpolation methods were applied to calculate retention indices of *n*-alkanes and fatty acid methyl esters spiked into routine GC-TOF-MS metabolite profiles. Two chromatography variants, 1 or 9 (cf. Fig. 1), are compared. Samples had either low complexity, namely single 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<sub>8</sub> ME–C<sub>9</sub> ME region (Fig. 2A). The increased RI (S.D.) coincided with delayed retention of C<sub>8</sub> ME–C<sub>9</sub> ME. Thus, early eluting FAMES had the strongest impact on overall RI accuracy in the presence of biological matrix with average  $|RI_{(DRM-rice)} - RI_{(low\ complexity)}|$  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.



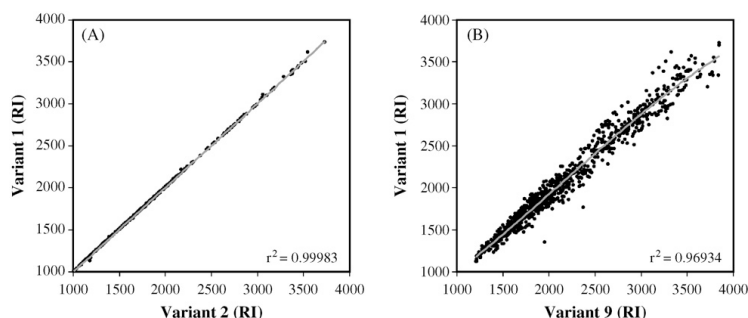
**Fig. 2.** The matrix effect negatively affecting RI precision and accuracy of methyl esters (ME) is dependent on chromatographic region. (A) Precision of RI determination was calculated as the standard deviation of RI (*n* > 40). (B) Accuracy of RI determination was estimated by comparison of RI measured in the presence of a complex biological matrix compared to a low complexity chemical background; the average  $|RI_{(DRM-rice)} - RI_{(low\ complexity)}|$  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<sub>8</sub> ME and C<sub>9</sub> ME.



**Table 3**Transfer of retention index (RI) properties between chromatography variants of the Golm Metabolome Database (GMD)<sup>a</sup>

	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
Acquisition rate (scans/s)	20	20	6	30	2	2	1.67	2
Mode of detection	TOF	TOF	TOF	TOF	Q	Q	Q	TRAP
<b>A</b>	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>B</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	–
<b>C</b>	Standard deviation ( $RI_{\text{predicted}} - RI_{\text{determined}}$ )							
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	–
<b>D</b>	Standard deviation ( $RI_{\text{predicted}} - RI_{\text{determined}}$ [% of $RI_{\text{determined}}$ ])							
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	–

<sup>a</sup> 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,  $RI_{\text{predicted}} - RI_{\text{determined}}$ , and (D) accuracy of prediction as characterised by standard deviation of residual errors expressed as percent of  $RI_{\text{determined}}$ . 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

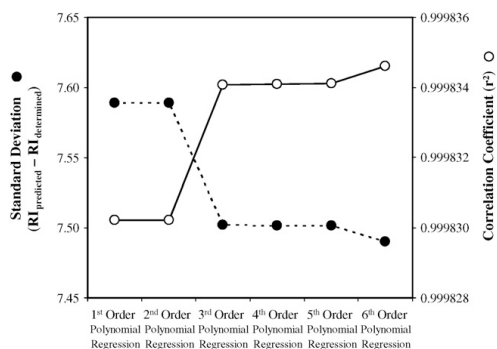
Precision of RI determination from our study was similar to the ~1 RI unit precision reported of a set of 250 volatile analytes [36]. We used the above DRM-mix (Table 1) to estimate, if results of RI precision, which were based on FAME, may be transferred to those compound classes which are relevant 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  $RI_{\text{predicted}} - RI_{\text{determined}}$  (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.

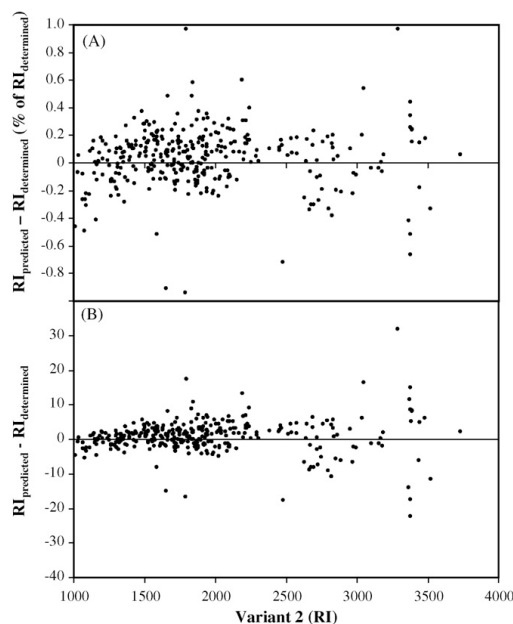
Following the principle of making the minimal required number of assumptions we decided for a 3rd order regression model and investigated all possible pair wise predictions between chromatography variants (Table 3B–D). Regression coefficients ranged from 0.99880 to 0.99987. The standard deviation of 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 trans-



**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  $RI_{\text{predicted}} - RI_{\text{determined}}$ . 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.

fer 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.

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 ( $(RI_{\text{predicted}} - RI_{\text{determined}}) \times RI_{\text{determined}}^{-1} \times 100$ , overall standard deviation, 0.39% ( $n = 348$ ), (B) absolute RI deviation,  $RI_{\text{predicted}} - RI_{\text{determined}}$ , 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 Quant-PRO 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. Reilly, 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. Steinhäuser, 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. Steinhäuser, *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. Liseč, N. Schauer, J. Kopka, L. Willmitzer, 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. Udvardy, *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.



Contents lists available at ScienceDirect

## Postharvest Biology and Technology

journal homepage: [www.elsevier.com/locate/postharvbio](http://www.elsevier.com/locate/postharvbio)Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impactNadine Strehmel<sup>a,1</sup>, Ulrike Praeger<sup>b,1</sup>, Claudia König<sup>b,2</sup>, Ines Fehrle<sup>a</sup>, Alexander Erban<sup>a</sup>, Martin Geyer<sup>b</sup>, Joachim Kopka<sup>a</sup>, Joost T. van Dongen<sup>a,\*</sup><sup>a</sup> Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany<sup>b</sup> Leibniz-Institute for Agricultural Engineering Potsdam-Bornim, Max-Eyth-Allee 100, 14469 Potsdam, Germany

## ARTICLE INFO

Article history:  
Received 13 October 2009  
Accepted 13 December 2009Keywords:  
Blackspot bruising  
GC–TOF–MS  
Mechanical stress  
Metabolite profiling  
Postharvest damage  
*Solanum tuberosum*  
Primary metabolism

## ABSTRACT

An important feature relevant to potato (*Solanum tuberosum*) tuber quality management is the susceptibility to blackspot bruising caused by melanin production after exposure to severe mechanical stress. We analyzed the previously neglected nature and sequence of responses in primary metabolism and compared the cultivars 'Afra' and 'Milva' that differ in tissue elasticity and susceptibility to blackspot bruising. Mechanical stress application was highly standardized and differential formation of bruising damage and occurrence of small cracks in the outer tuber layers were demonstrated. Concomitant GC–TOF–MS based metabolite profiling revealed characteristic changes in central metabolism, namely in intermediates of the TCA cycle and linked parts of amino acid metabolism, with the clear exception of phenylalanine and tyrosine, the immediate precursor of melanin biosynthesis. We conclude that the initial metabolic responses and the initiation of blackspot formation are distinct, and that melanin biosynthesis is not driven by early precursor accumulation through activation of the aromatic amino acid biosynthesis pathway.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

Mechanical impact on potato tubers during harvest and subsequent handling not only causes external damage such as cracking and scuffing of the skin (Hughes, 1980), but can also result in a bluish-grey to black internal discoloration of the tuber tissue. This so-called blackspot bruising is hardly visible from the outside as it is located about 2 mm beneath the surface of the tuber in the region of the vascular ring. Blackspot bruising typically occurs within 1–3 d after mechanical impact (Burton, 1989; Molema, 1999) and is considered a significant impediment for the utilization of tuber harvests throughout the world as it severely influences consumer acceptance (Peters, 1996).

Several parameters and tissue features are known to be associated with the susceptibility of potato tubers to bruising damage, such as temperature, tissue toughness, turgor pressure, starch content, potassium content, location of the impact on the tuber, tuber age and storage duration (Kunkel and Gardner, 1959; Aeppli and Keller, 1979; Brook, 1996; McGarry et al., 1996; Peters, 1996; Bajema et al., 1998; McNabney et al., 1999; Van Canneyt et al.,

2006). Some of these parameters are known to relate to mechanical tuber properties like the transmission of impact energy and the tendency of the tissue to crack upon collision (McGarry et al., 1996). Therefore, failure properties after dynamic compression of potato tuber tissue can be used to characterize differences between potato cultivars with regard to bruising sensitivity and mechanical impact (Bajema et al., 1998). Furthermore, the severity of damage appears to depend on physiological and biochemical properties such as differences in specific metabolite levels or enzyme activities, such as the chemical oxidation reactions and intermediates that are directly involved in the internal discoloration.

Upon mechanical impact, cell membranes have been shown to be damaged and the enzyme polyphenol oxidase (PPO) and phenolic substrates that are normally separated by different subcellular compartmentation come into contact now. Furthermore, PPO activity exhibits a subcellular redistribution response 12 h after mechanical impact without apparent changes in the corresponding mRNA levels (Partington et al., 1999). The amino acid tyrosine is regarded as the most important substrate for PPO-catalyzed oxidative conversion. This reaction leads to the production of quinones, which subsequently react via a non-enzymatic reaction to the black pigment melanin (Corsini et al., 1992; Dean et al., 1993). High tyrosine content is known to correlate with a high susceptibility to internal discoloration. Other phenolic compounds such as chlorogenic acid or caffeic acid are known to be relevant but their contribution to blackspot formation is less clear (Dean et al., 1993;

\* Corresponding author. Tel.: +49 331 567 8353; fax: +49 331 567 8701.  
E-mail address: [dongen@mpimp-golm.mpg.de](mailto:dongen@mpimp-golm.mpg.de) (J.T. van Dongen).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Present address: Metanomics GmbH, Tegeler Weg 33, 10589 Berlin, Germany.

McGarry et al., 1996; Laerke et al., 2002). For example, the formation of iron-chlorogenic acid chelates has been suggested to contribute to the internal discoloration of potato tuber tissue (Putz, 1995).

In contrast to what is well known about the chemical reactions directly involved in blackspot formation, only little is known about the involvement of central metabolism, especially primary carbohydrate metabolism such as the respiratory metabolic pathways consisting of glycolysis and the subsequent tricarboxylic acid (TCA) cycle. Amino acid metabolism, and the time lag between mechanical impact and the onset of discoloration have also been largely neglected until now. The initial mechanical stress signal is expected to be linked to later biochemical responses via intermediate metabolic processes as earlier studies which were the incentive for this investigation demonstrated that respiratory activity increased significantly after mechanical stress application to potato tubers (Aeppli and Keller, 1980). Indeed, a correlation of this effect with formation of blackspot bruising also became apparent in our study.

Additional knowledge was generated by previous analyses of transcriptional regulation. This evidence points towards an intermediate metabolic reprogramming. Specifically, the transcriptional induction of ubiquitin, an indicator of protein degradation, and of phenylalanine ammonia-lyase, the key step of phenylpropanoid biosynthesis including chlorogenic acid production, have been linked to the mechanical impact response of potato tubers (Rickey and Belknap, 1991).

In the present study, we investigated the biochemical changes in primary metabolism using the metabolomic toolbox of GC-TOF-MS based metabolite profiling for time course analyses. For this, a well established profiling method (Erban et al., 2007) was applied which is dedicated to a comprehensive analysis of intermediates of primary carbon and nitrogen metabolism. Only limited information about secondary metabolism can be obtained by using the method applied here as most compounds of secondary metabolism have a far too large molecular weight to be analyzed and are not properly derivatized.

The study targets a time window after mechanical impact but before any visible damage and discoloration of the potato tuber tissue occurred. Care was taken to perform a differential tissue analysis by applying highly standardized mechanical impact and by comparing the mechanically stressed part of each tuber to an equivalent part of the same tuber, which was not exposed to the direct collision event. The observed changes in central metabolism were associated with the susceptibility and the extent of subsequent blackspot formation of two potato cultivars. These potato cultivars had been chosen according to their differential physical tuber properties, which are demonstrated to affect the mechanical transmission of the force of impact.

## 2. Materials and methods

### 2.1. Plant material

Potato (*Solanum tuberosum* cv. 'Afra' and cv. 'Milva') tubers were manually harvested in September and October 2007 from a field used for commercial potato tuber production near Niedergörsdorf (51°59'N, 13°0'S) in the region of southern Brandenburg, Germany. Harvested tubers were stored at 4 °C without a sprouting inhibitor. In a previous study (Praeger et al., 2010) physiological properties of the tubers were monitored several times during the storage period starting from harvest in autumn 2007 until spring 2008. These characteristics of the plant material can be summarized as follows: the starch content of cv. 'Afra' changed from  $17.0 \pm 2.2\%$  to  $18.9 \pm 1.0\%$  of the fresh weight and for tubers of cv. 'Milva' starch content

changed from  $14.7 \pm 1.7\%$  to  $16.4 \pm 1.1\%$  during the storage period. The potassium content of cv. 'Afra' was 572 mgK/100 mg fresh mass and cv. 'Milva' 458 mgK/100 mg fresh mass. Stored tubers of cv. 'Milva' showed in winter and spring higher turgor pressure, a higher elastic modulus, higher stiffness and were more susceptible to superficial cracking after mechanical impact than tubers of cv. 'Afra'.

### 2.2. Determination of susceptibility to blackspot bruise and cracking

Previous studies (data not shown) had shown that the susceptibility to blackspot bruising could be highly variable during the first months of storage. Therefore, only tubers with a mass of  $120 \pm 30$  g were selected after 8 months of storage (May 2008) for this study. One day prior to the susceptibility test they were transferred to 12 °C and subsequently washed and weighed.

To determine the susceptibility to mechanical impact tubers were dropped once from 50 cm onto a steel plate, either on the apical end of the tuber or on its side. For this, a special fall apparatus, previously described by Geyer et al. (2009), was used that was equipped with an impact force sensor to record the impact force versus time (sample rate of 10,000 Hz). Half an hour after impact, tubers were transferred to an incubation chamber with 33 °C and 95% relative humidity until blackspot bruise formation was scored. These environmental conditions, which are common practise for testing blackspot bruising (Laerke et al., 2002) were chosen to accelerate blackspot bruising after mechanical load.

To analyze blackspot formation, tubers were cut longitudinal through the middle of the impact side and a photo of the cut surface was made. The area of blackspot bruise in each tuber was determined using the software Optimas 6.0 (BioScan Inc., Bothell Washington, USA) as described by Wormanns et al. (2000). The percentage of the amount of tubers ( $n=20$ ) with bruises and the average area of the blackspots was calculated. Additionally, occurrence of superficial cracks (depth < 2 mm) of the tuber surface was scored, and the percentage of the amount of these cracked tubers was determined.

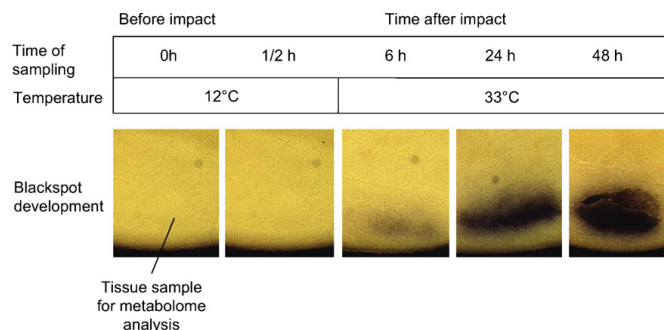
### 2.3. Tissue sampling for metabolite analysis

Metabolite profiling was performed on tubers ( $n=6$ ) after mechanical loading using the same procedure as described before in Section 2.2. Only tubers used for metabolite profiling were dropped 10 times on the same spot at the longitudinal side of the tuber, thus guaranteeing that all tuber tissue sampled prior to the first occurrence of blackspot bruising was primed to develop bruising damage. As metabolite profiling destroys the tuber sample we had to ensure that any metabolic differences between samples can be reliably related to a damage that will ultimately cause blackspot formation.

For each tuber, two samples were taken: one exactly at the spot where the tuber hit the steel plate, and a second control sample opposite to the impacted side. Samples were taken as a time course which started immediately before the first impact  $t=0$ , followed by sampling at 0.5, 6, 24 and 48 h after impact. Fig. 1 shows the covered timeframe and representative bruising damage. Tissue samples were taken with a core of approximately 10 mm deep and a diameter of 10 mm. The tuber skin was removed, and the rest of the sample immediately sliced into liquid nitrogen. Storage was at  $-80$  °C until further processing.

### 2.4. Metabolite profiling

Frozen tuber slices were ground to a fine powder using a ball mill (MM400, Retsch GmbH, Haan, Germany) cooled with liquid



**Fig. 1.** The time resolved sampling scheme related to the development of blackspot bruising. Representative examples are shown. The site of the tissue sample core used for GC–TOF–MS based metabolome analysis is indicated. For the differential analysis of the mechanical impact effects an equivalent tissue core was taken from the opposite non-impacted side of each tuber.

nitrogen. Samples of  $50 \pm 2$  mg were aliquoted and metabolites extracted in 300  $\mu$ L methanol. Additionally, 30  $\mu$ L nonadecanoic acid methyl ester (2 mg/mL in chloroform) and 30  $\mu$ L mix of labeled internal standards were added in a ratio of 10:1:1 (v/v/v) (for details of components see [Supplementary Table 1](#)). Subsequently, samples were shaken at 70 °C for 15 min and adjusted to room temperature before adding 200  $\mu$ L chloroform and 400  $\mu$ L double-distilled water. After thoroughly shaking, the samples were centrifuged, 10  $\mu$ L of the polar phase dried down in a vacuum concentrator and stored at –20 °C until further processing. To improve the volatility of metabolites such as amino acids, nucleic acids, sugars, polyamines and polyols, we applied a methoxymation and silylation step before analysis. For derivatization, dried samples were redissolved in 10  $\mu$ L methoxyamine hydrochloride (20 mg/mL in pyridine) and kept at 40 °C for 90 min. Then, 16.6  $\mu$ L mix of N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) and n-alkanes (decane, dodecane, pentadecane, octadecane, nonadecane, docosane, octacosane, dotriacontane, hexatriacontane; 0.22 mg/mL in pyridine; alkanes were needed for chromatographic alignment) was added to each sample and incubated for another 45 min at 40 °C. Finally, 1  $\mu$ L of each sample was injected at 230 °C in a splitless mode at constant helium carrier gas flow into the temperature program driven 6890N Gas Chromatography System (Agilent Technologies, Böblingen) being mounted with a 5PDM VF-5 ms fused silica capillary column, 0.25  $\mu$ m film thickness, 30 m length, 0.25 mm inner diameter and a 10 m EZ-guard precolumn (Varian, Darmstadt, Germany). 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. The adjacent TOF-detection was performed via a Pegasus III mass spectrometry system (LECO) with 20 scans/s in the range of 70–600 amu. Derivatization and analysis steps were conducted in an automatic manner e.g. pipetting steps, automated chemical derivatization 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) as described earlier by Erban et al. (2007).

### 2.5. Data analysis and statistics

Data analysis was performed with the help of the TagFinder software (Luedemann et al., 2008). In short, peak intensities above 50 arbitrary ion current units were imported, aligned by the retention index model of van den Dool and grouped according to their common retention time and mass spectral features. The relevant

mass spectral features were identified with the help of our current in-house mass spectral and retention index reference library (Kopka et al., 2005). Hereby, we used hits with a maximum relative retention index deviation of 0.5–1% RI units of the reference spectrum (Strehmel et al., 2008) and hits matching in at least three mass fragments.

For statistics, metabolic peak intensities were base line corrected and normalized to fresh weight and the sum of internal standards. Then, the data were mean-centered per analyte, log 10-transformed and submitted to statistical analysis such as two-way analysis of variance (2-way ANOVA) and independent component analysis (ICA). The first procedure was performed using the TIGR Multiple experiment viewer software (TMEV.31) (Saeed et al., 2003) whereas the latter one was carried out with the help of the MetaGeneAlyse web service (Scholz et al., 2004). Pathway based visualization of differential metabolic responses was performed using the VANTED software (Junker et al., 2006). To compare cultivar effects and mechanical impact force at the first and the tenth drop of the tubers, data of the impact force measurements were analyzed by analysis of variance (ANOVA) with the GLM procedure (SAS version 9.1, 2002) using the Scheffe-test (\* $P < 0.05$ ).

### 2.6. Respiration measurements

The rate of respiration by intact tubers was determined by measuring the rate of oxygen consumption. Tubers (cv. 'Afra') with a total mass of approximately 2 kg were prepared in the same manner as described in Section 2.2 and were dropped 10 times from 50 cm height at the longitudinal side of the tuber onto a steel plate. Half an hour after impact, they were placed in a glass vessel (volume: 5 L, temperature: 33 °C), which could be hermetically closed while determining the rate of oxygen consumption continuously during 48 h from the headspace using an optical oxygen sensor (PreSens, Regensburg, Germany). As control treatment, tubers were handled similarly, except that they were not dropped. Depending on the velocity of O<sub>2</sub>-decline, the vessel was opened regularly to avoid inhibition of respiration due to anoxic conditions.

## 3. Results

### 3.1. Susceptibility to blackspot bruising and cracking

A tuber drop-test was carried out to determine the susceptibility to blackspot bruising and cracking of 'Afra' and 'Milva' potatoes (Table 1). Both cultivars were more susceptible to blackspot bruising

**Table 1**

Percentage of the amount of tubers ( $n=20$ ) with cracking, blackspot bruising and average area of blackspots in damaged tubers determined 48 h after a single drop towards the apical end or towards the side from 50 cm height onto a steel plate.

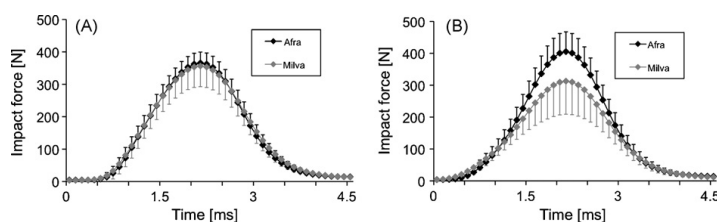
	Afra	Milva
<b>Blackspot bruising</b>		
Affected tubers (% of total tubers)		
Apical end	10	0
Side	55	15
Average area (cm <sup>2</sup> )		
Apical end	0.62 ± 0.04	0
Side	0.96 ± 0.27	1.00 ± 0.19
<b>Cracking</b>		
Affected tubers (% of total tubers)		
Apical end	0	20
Side	0	10

ing when tubers were dropped on their side as compared to the apical end. This was not only apparent from the percentage of tubers that developed blackspots as was determined two days after impact, but also by the size of the resulting blackspots. 'Afra' tubers were obviously much more susceptible to blackspot bruising than those of 'Milva'. However, a clear negative correlation was observed between the susceptibility to blackspot bruising and the occurrence of small cracks at the surface of the tuber at the collision spot (Table 1).

Mechanical properties, like tissue stiffness and elasticity, affect the transfer of the impact energy to the tuber. In order to analyze the physical variation between tubers of the two potato cultivars, the impact force was measured for ten repetitive drops. The area under an impact force versus time curve relates directly to the drop height (Brook, 1996), but since we always dropped tubers from the same height of 50 cm, the shape of the impact force versus time curve depended exclusively on physical properties of the tubers. Soft objects show a flat progression of impact force whereas the force will have a sharp peak of increase and decline when the object is firm. The average force versus time curves ( $n=24$ ) of the first impact of tubers from the two cultivars were very similar. However, the force curves of the tenth fall showed a significantly increased maximum force for 'Afra' tubers in contrast to 'Milva' tubers. The later curve profile was more flattened (Fig. 2). Apparently, the mechanical stiffness of collided tissue of 'Afra' tubers increased by repetitive impact in contrast to 'Milva' tubers.

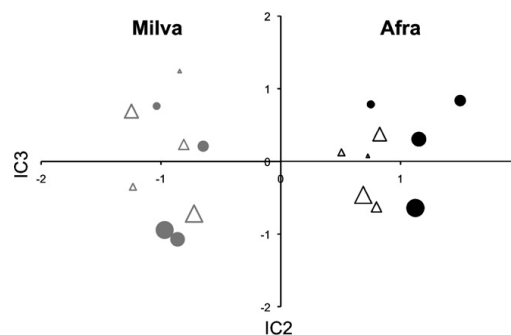
### 3.2. Global metabolic fingerprinting of tuber tissue after mechanical impact

Potato tuber tissue samples were collected following a time series ranging from shortly before the tuber drop-test ( $t=0$ ), until 48 h after. The spot at which the tuber hit the platform was marked for subsequent tissue sampling at various time points after impact.



**Fig. 2.** Average impact force-to-time curves of potato tubers of the cv. 'Afra' (black) and the cv. 'Milva' (grey) ( $n=24$ ) from the first impact (A) and the last impact (B) of ten repetitive falls towards the side from 50 cm height on a steel plate.

Please cite this article in press as: Strehmel, N., et al., Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impact. Postharvest Biol. Technol. (2010), doi:10.1016/j.postharvbio.2009.12.008



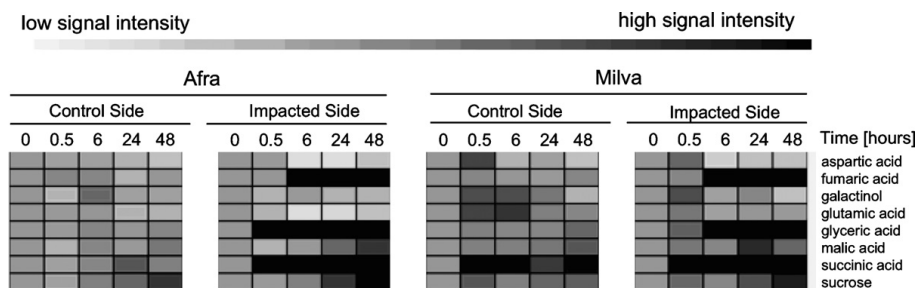
**Fig. 3.** Independent component analysis of metabolic fingerprints during blackspot development. The experimental design was best revealed by the scores plot of the second and third independent component. The full circles indicate the mechanically impacted samples, cv. Afra (black) and cv. Milva (grey), while the open triangles represent the paired non-impacted control samples. The increasing item size corresponds to the time ( $t=0, 0.5, 6, 24$  and 48 h) after mechanical impact.

For each time point, control samples were taken from the same tuber but opposite to the site of impact. These samples were used to perform a differential metabolic profile of the impact responses within each tuber using routine GC-TOF-MS technology (e.g. Erban et al., 2007). An overview of all determined metabolites, representing a large part of primary metabolism is given in Supplementary Table II.

In order to gain a first insight into the global metabolic variation caused by the differential impact responses, a principal component analysis was performed. The first five principal components covering 59% of total variance in the data set were subjected to an independent component analysis for the detection of bimodal sample distributions within the data set. The two cultivars are clearly separated by means of the third independent component (IC3), whereas a trend in time is revealed by IC2 (Fig. 3). In general, the early metabolic responses appear to be distinct from the late metabolite phenotypes. This temporal differentiation of the metabolic phenotype appeared to be more distinctive in profiles of mechanically impacted tissue as compared to the control samples of the same tubers. The metabolic response to the mechanical impact was more obvious in the cultivar 'Afra' that is more susceptible to blackspot bruising as compared to the cultivar 'Milva'. This initial global fingerprinting analysis indicates that mechanical impact disrupts metabolism in a cultivar specific and time dependent manner.

### 3.3. Identification of key metabolic changes

Further detailed analysis was performed to understand which specific metabolites are involved in the observed temporal pat-



**Fig. 4.** Heat map showing pool size changes of key metabolites selected by a 2-way ANOVA testing for variance attributed to mechanical impact or time. The grey scale encodes the magnitude pool size changes relative to  $t=0$  of each cultivar (cv. 'Afra' and cv. 'Milva'). Note that the visualization has maximum fold change = 5 in order to obtain optimal resolution of the minor changes (cf. Supplementary Table II).

terns of the responses to mechanical impact. Key metabolites of both cultivars were selected from the full set of monitored GC-TOF-MS analytes using 2-way ANOVA and application of the significance threshold,  $^*P \leq 0.05$ . The two factors "mechanical impact" and "time" were analyzed. Aspartate, fumarate, glutamate, glycerate, and succinate passed the significance threshold for the factor "mechanical impact" while aspartate, fumarate, galactinol, malate, succinate, and sucrose were found to have significant variation with respect to the factor "time". The ANOVA-positive metabolites appeared to follow two general temporal patterns (Fig. 4). This heat map display of the fold-changes in metabolite pools relative to  $t=0$  demonstrated an early pattern of transient responses at 0.5–6 h including aspartate, glutamate and galactinol. These metabolites accumulated rapidly and returned to initial levels or below. A second set of metabolites exhibited persistent changes. These metabolic changes were stacked in time. Glycerate and succinate started to change at 0.5 h followed by fumarate at 6 h and malate as well as sucrose at 24–48 h.

The strongest differential responses of the key metabolites were caused by the mechanical impact. Minor metabolic changes were also detectable in the non-impacted tissue, with the exception of succinate that was strongly affected in non-impacted tissue of cultivar 'Milva'. Differences between cultivars were quantitative rather than qualitative. As a rule the impacted tissue of the more susceptible cultivar 'Afra' appeared to respond in a more extreme manner than the 'Milva' cultivar and the response was earlier in one case, i.e. glycerate (Fig. 4).

#### 3.4. In-depth analysis of cultivar specific differential metabolic responses

The strongest key changes after mechanical impact were observed in metabolites of the central respiratory pathway. As a consequence we performed a detailed assessment of all metabolites covered by GC-TOF-MS profiling that take part in primary carbon metabolism, the TCA cycle and linked amino acid biosynthesis pathways (Fig. 5, underlying data can be found in the Supplementary Table II). Because the non-impacted part of tubers showed mostly minor metabolic responses that were not directly linked to changes in the mechanically damaged site, we decided to perform a differential analysis for each tuber. We calculated the metabolite response ratio of the mechanically stressed versus the non-impacted part of each tuber and then normalized to  $t=0$ . The resulting differential plots (Fig. 5) allowed clear comparison of metabolic time courses between the blackspot susceptible cultivar 'Afra' and the less susceptible cultivar 'Milva' with focus on the specific responses of the impact site.

The differential analysis of metabolic changes demonstrated that most changes which occurred early at 0.5 h were not specific to the site of impact; for example, absolute and differential changes of malate and glutamate can be compared (Figs. 4 and 5). However, glucose and fructose had a biphasic accumulation, which peaked at 0.5 and 48 h. In contrast, the pool sizes of sucrose and most metabolites branching off from glycolytic precursors were relatively stable. This included not only glycine, serine, and the branched chain amino acids, leucine and valine, but also and importantly the aromatic amino acid biosynthesis pathway, represented in our analysis by quinate, tyrosine and phenylalanine (Fig. 5). Strong differential increases were observed for three intermediates of the TCA cycle, namely aconitate, succinate and fumarate, which peaked sequentially between 0.5 and 24 h, whereas metabolites of pathways diverging from the TCA cycle showed a tendency to decrease transiently at 6–24 h (cf. Glu and Asp in Fig. 5). Finally, lactate exhibited a differential accumulation in the late stages, 24–48 h, of the response to mechanical impact.

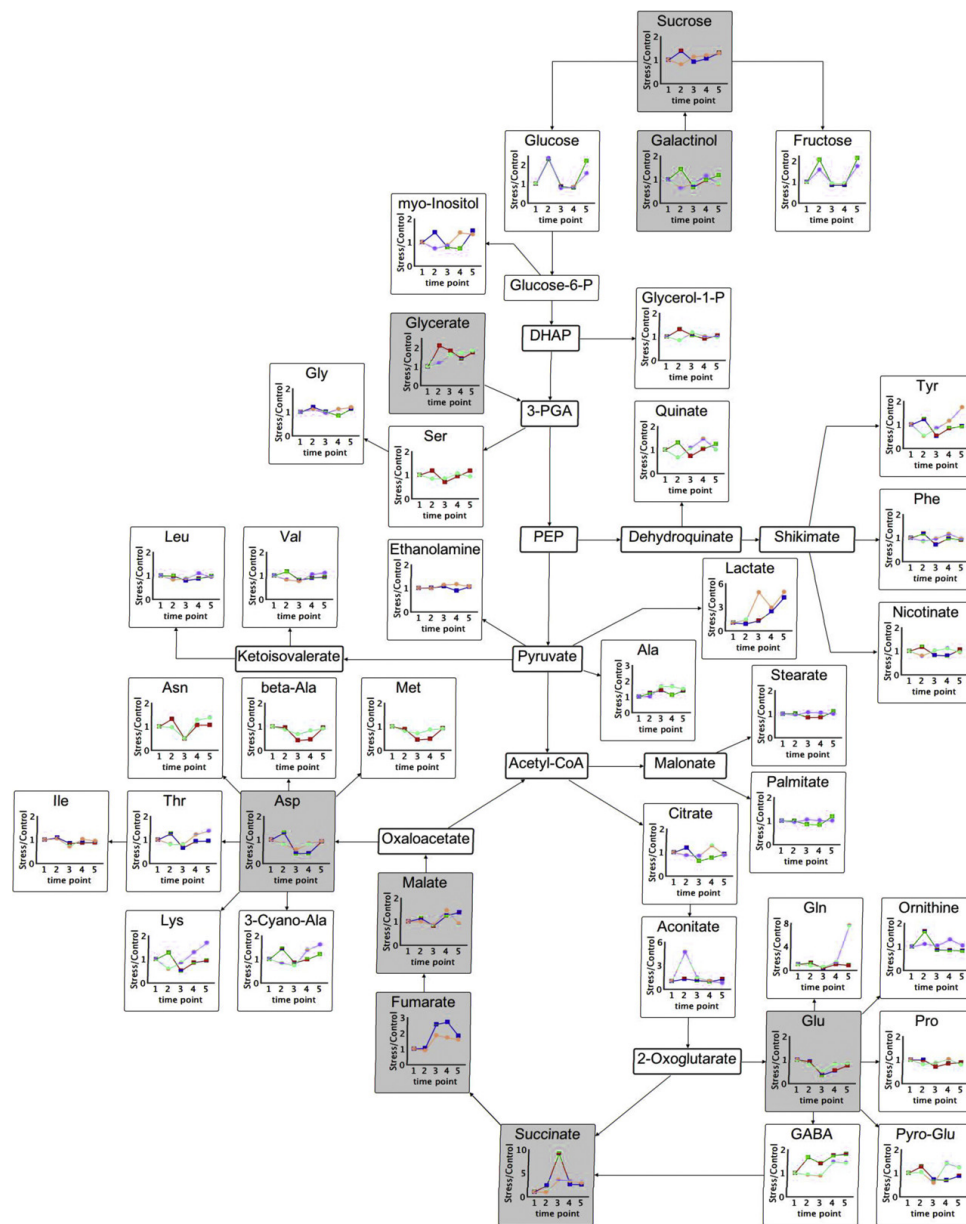
When comparing the blackspot susceptible cultivar 'Afra' to the cultivar 'Milva' only few differing features were apparent, which did not include the aromatic acid biosynthesis pathway. Cultivar 'Afra' differed most by extreme accumulation of succinate and fumarate compared to the lack of aconitate accumulation. In addition 'Afra' had stronger pool size oscillations: (1) the transient decrease at 6–24 h of metabolite pools linked to the TCA cycle was more extreme, e.g. aspartate, methionine and beta-alanine; and (2) sucrose, galactinol and *myo*-inositol showed a tendency of inverse accumulation/depletion 0.5 h after mechanical stress (Fig. 5).

In summary, the tendency of more extreme metabolic responses in cultivar 'Afra' appeared to be associated with the lower frequency of cracks in the surface tissue layers and the higher susceptibility to blackspot bruising as compared to the cultivar 'Milva'.

#### 3.5. Changes in respiratory activity after mechanical impact

As 'Afra' showed clear changes in respiratory metabolism after the mechanical impact, e.g. succinate, fumarate and lactate, we measured the rate of oxygen consumption of impacted tubers in order to verify if metabolic changes in cells are indeed reflected by changes in respiration activity. In control tubers that were not exposed to mechanical stress, the  $O_2$ -consumption rate was about  $15 \text{ mg kg}^{-1} \text{ h}^{-1}$  throughout the entire time course of the experiment. Mechanical impact induced a nearly 1.3-fold increase as compared to the control tubers kept at the same temperature ( $33^\circ \text{C}$ ), resulting in a respiration rate of  $20 \text{ mg kg}^{-1} \text{ h}^{-1}$  during the first 12 h. After 24 h the  $O_2$ -consumption of the impacted tubers approximated the  $O_2$ -consumption of the tubers without mechanical impact again ( $15 \text{ mg kg}^{-1} \text{ h}^{-1}$ ).





**Fig. 5.** Differential metabolic changes induced after mechanical stress in the blackspot susceptible cv. 'Afra' (blue) and the cv. Milva (orange). Note that time courses are displayed after calculation of differential response ratios between the impacted part and the paired non-impacted part of each tuber followed by normalization of each time course to  $t=0$ . The traces of metabolites highlighted by grey underlay are the key metabolites selected by 2-way ANOVA (cf. Fig. 4). The full data set can be found in Supplementary Table II.

#### 4. Discussion

An immense variability exists between potato cultivars in the susceptibility to potato tuber damage like blackspot bruising. The

chemical reactions that lead to the discoloration of tuber tissue after impact are well described, and it is known that metabolic precursors of primary metabolism are involved in melanin biosynthesis. However, hardly anything is known about how primary

Please cite this article in press as: Strehmel, N., et al., Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impact. Postharvest Biol. Technol. (2010), doi:10.1016/j.postharvbio.2009.12.008

metabolism of the potato tuber reacts to mechanical impact specifically during the interval before blackspot bruising becomes visible. In this study, we investigated the biochemical changes of potato tuber tissue after mechanical impact by analyzing two potato cultivars, 'Afra' and 'Milva', which are characterized by different susceptibilities to bruising.

#### 4.1. Bruising susceptibility and tissue properties of the tubers

'Afra' and 'Milva' tubers were grown under identical conditions, and manually harvested with care to prevent any bias towards the occurrence of blackspot bruising caused by the preceding harvesting procedure. A clear difference in susceptibility to blackspot bruising was observed between the cultivars, which correlated well with differential mechanical properties of the tuber tissue. Drop-tests revealed that 'Milva' tubers were clearly less susceptible to blackspot bruising than those of 'Afra' (Table 1). However, the skin of 'Milva' tubers cracked more easily than that of 'Afra' tubers. Indeed, the change of the force-to-time curves during ten successive falls indicated differential tissue properties of the two cultivars (Fig. 2). In 'Afra' tubers the maximum force increased with subsequent impacts, which is probably due to enhanced tissue stiffness. This effect might be caused by an increased amount of rebound energy and decreased absorbed energy after repetitive impacts on the tuber tissue (Hyde et al., 1993). In contrast, 'Milva' tubers showed only a slightly decreased maximum force of impact after ten recurrent falls. This result is likely explained by the occurrence of small cracks that were observed at the surface of the impacted tissue. The cracking is thought to increase the compliance of the tissue thus dampening the impact force.

The decreased impact force as measured in 'Milva' tubers can explain the reduced susceptibility to blackspot bruising of this cultivar. First of all, the damaging power of the physical impact and the energy directly transmitted to the internal cell layers is reduced when the impact force of a fall diminishes. Furthermore, the cracks will mainly damage the tissue structure of a tuber, whereas the membrane structure of the cells will get less impaired. Summarizing, it can be concluded that the cultivars are characterized by different physical properties, resulting in distinct energy transmission profiles of a pre-determined physical impact. These different physical properties are suggested to contribute to the differential susceptibility to blackspot bruising between the two cultivars.

#### 4.2. Mechanical impact and subsequent biochemical changes

Given the differential mechanical and blackspot bruising properties, metabolic differences between the cultivars were expected. When all information obtained from the metabolic fingerprinting was considered using an independent component analysis (Fig. 3), all tuber samples were clearly distinguishable according to cultivar. Furthermore, a clear time course behaviour was revealed which was, with few exceptions (Fig. 4), limited to the mechanically impacted samples. In contrast, the paired non-impacted samples did not exhibit a temporal pattern. In short, the physical impact had a clearly limited local effect on the tuber tissue which exhibited a dynamic progression over time.

The search for key metabolites involved in the mechanical stress response resulted in eight metabolic compounds that were identified to vary significantly ( $*P \leq 0.05$ ) following stress application (Fig. 4). Eight of these key compounds were identified, namely the primary metabolites were aspartate, fumarate, galactinol, glutamate, glycerate, malate, succinate and sucrose. As all key metabolites were major intermediates of primary and carbon metabolism, our study proves a clear link of mechanical stress to the remodelling of central metabolism. A similar result was found

after severe wounding of potato tissue. A clear shift of metabolites that belong to pathways of primary metabolism was found to be induced (Yang and Bernards, 2007), however these metabolic changes were mainly correlated with wound healing reactions like the production of suberin.

An in-depth analysis of central metabolic pathways was performed using the time courses of differential response ratios between the impacted versus the non-impacted part of each tuber. This approach provided a more detailed overview of changes and allowed better comparison of the two potato cultivars (Fig. 5). This analysis showed an early response of the TCA cycle intermediates, succinate and fumarate, possibly fuelled by early accumulation of fructose and glucose. Transient reduction of metabolite pools of biosynthesis pathways that branch off from the TCA cycle supported the conclusion that carbon sources are temporarily recruited to respiration rather than to amino acid biosynthesis. The early induction of respiratory processes was clearly supported by enhanced oxygen consumption directly after mechanical impact. Similar observations have been described previously for several other potato varieties (Aeppli and Keller, 1980) as well as for various crop products (Meinl, 1972; Oberbarnscheidt et al., 1996; Mempel and Geyer, 1999). Our oxygen consumption data confirm a concurrent activation of respiration and the TCA cycle. This activation apparently meets the increased energy demand after mechanical collision, however at the expense of mostly Asp and Glu metabolism. In the later phase, namely 24–48 h after mechanical impact, high lactate accumulation may indicate an increasing importance of substrate chain regeneration of reduction equivalents while respiration appears to return to initial levels.

To our surprise metabolite pools that are fuelled by glycolytic intermediates do not show clear and significant changes, even though glucose and fructose went through a biphasic accumulation peaking at 0.5 and 48 h after mechanical stress application. The same observation holds true for the monitored intermediates of the aromatic amino acid pathway. Interestingly, the level of tyrosine, which is a well-known primary substrate for the chemical reaction leading to melanine formation and phenylalanine the starting point of phenylpropanoid biosynthesis including the caffeine moiety of chlorogenic acid, does not change significantly after mechanical impact in both cultivars. Even more importantly the absolute level of tyrosine before impact was higher in the non-sensitive cultivar 'Milva'. As the tyrosine and phenylalanine pools neither increase prior to blackspot formation nor decrease concomitantly we can exclude mechanisms of blackspot susceptibility which are mediated by precursor accumulation or limitation. Instead, at least for the comparison of the cultivars we need to assume a differential flux regulation, which does not affect early intermediate pools. Nevertheless, several earlier studies showed a clear positive correlation between the endogenous level of free tyrosine (i.e. tyrosine that is not incorporated into protein) also determined in our study and the susceptibility to blackspot bruising (Corsini et al., 1992; M. Steinfath, pers. comm.), whereas the level of free tyrosine correlated negatively to the level of protein-bound tyrosine (Corsini et al., 1992). From these data it was concluded that free tyrosine increases when the rate of protein synthesis is low, which would lead to an increased susceptibility to blackspot bruising. This hypothesis might indeed be true to correlate bruising susceptibility with differences in free tyrosine from dormant non-stressed tubers, but as discussed above the tyrosine and also the phenylalanine pools are not directly affected by the physical stress applied prior to the formation of blackspot discoloration. This leads us to the assumption that physical parameters such as turgor and stiffness may be more relevant to blackspot susceptibility for the cultivars investigated in this study.

Please cite this article in press as: Strehmel, N., et al., Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impact. *Postharvest Biol. Technol.* (2010), doi:10.1016/j.postharvbio.2009.12.008

## 5. Conclusions

Our study demonstrated that upon collision of potato tubers, specific metabolic changes are induced. These changes predominantly include intermediates of primary metabolism. Especially respiratory metabolism is increased, including the up-regulation of the levels of specific TCA cycle intermediates. Concomitantly, amino acids that are derived from TCA cycle precursors are depleted. The amplitude of the metabolic changes depended on the cultivar being investigated. It appeared that identical collision events have distinct stress potential due to cultivar specific physical properties of the tuber. When tissue does not crack upon collision, the impact energy is more strongly distributed throughout a larger area of the tissue, which results in more severe changes in primary metabolism as well as an increased susceptibility to blackspot bruising.

## Acknowledgements

The authors thank Anke Langer and Hildegard Hopp for help with the preparation of some of the material. This research obtained financial support from the German Federal Ministry of Education and Research (BMBF, QUANTPRO INNOX "Innovative diagnostic tools to optimize potato breeding: systemic analysis of cellular processes and their relation to plant internal oxygen concentrations" PTJ-BIO/Wir/0101-31P4295).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.postharvbio.2009.12.008.

## References

- Aeppli, A., Keller, E.R., 1979. Einfluß des Standortes auf die Blauempfindlichkeit der Kartoffelknollen. Z. Acker- und Pflanzenbau 148, 115–130.
- Aeppli, A., Keller, E.R., 1980. Beziehung zwischen Respiration und Blaufleckigkeit von Kartoffeln nach einer mechanischen Behandlung der Knollen. Potato Res. 23, 25–32.
- Bajema, R.W., Hyde, G.M., Baritelle, A.L., 1998. Turgor and temperature effects on dynamic failure properties of potato tuber tissue. Trans. ASAE 41, 741–746.
- Brook, R.C. (Ed.), 1996. Potato Bruising. How and Why Emphasizing Blackspot Bruise. Running Water Publishing, Haslett, MI.
- Burton, W.G. (Ed.), 1989. The Potato, 3rd edition. Longman Scientific and Technical, New York.
- Corsini, D.L., Pavek, J.J., Dean, B., 1992. Differences in free and protein-bound tyrosine among potato genotypes and the relationship to internal blackspot resistance. Am. Potato J. 69, 423–435.
- Dean, B.B., Jackowiak, N., Nagle, M., Pavek, J., Corsini, D., 1993. Blackspot pigment development of resistant and susceptible *Solanum tuberosum* L. genotypes at harvest and during storage measured by three methods of evaluation. Am. Potato J. 70, 201–217.
- Erban, A., Schauer, N., Fernie, A.R., Kopka, J., 2007. Non-supervised construction and application of mass spectral and retention time index libraries from time-of-flight gas chromatography-mass spectrometry metabolite profiles. In: Weckwerth, W. (Ed.), *Metabolomics: Methods and Protocols*. Humana Press, Totowa, pp. 19–38.
- Geyer, M.O., Praeger, U., König, C., Graf, A., Truppel, I., Schlüter, O., Herold, B., 2009. Measuring behaviour of an acceleration measuring unit implanted in potatoes. Trans. ASABE 52, 1267–1274.
- Hughes, J.C., 1980. Role of tuber properties in determining susceptibility of potatoes to damage. Ann. Appl. Biol. 96, 344–345.
- Hyde, G.M., Bajema, R.W., Zhang, W., 1993. Measurement of impact damage threshold in fruits and vegetables. In: *Proceedings, IV International Symposium on Fruit, Nut, and Vegetable Production Engineering*, Valencia-Zaragoza, Spain, pp. 1–8.
- Junker, B.H., Klukas, C., Schreiber, F., 2006. VANTED: a system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics 7, 109.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmueller, E., Doermann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R., Steinhauser, D., 2005. GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21, 1635–1638.
- Kunkel, R., Gardner, W.H., 1959. Black spot of russet burbank potatoes. Proc. Am. Soc. Hortic. Sci. 73, 436–444.
- Laerke, P.E., Christiansen, J., Veierskov, B., 2002. Colour of blackspot bruises in potato tubers during growth and storage compared to their discolouration potential. Postharvest Biol. Technol. 26, 99–111.
- Luedemann, A., Strassburg, K., Erban, A., Kopka, J., 2008. TagFinder for the quantitative analysis of gas chromatography-mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics 24, 732–737.
- McGarry, A., Hole, C.C., Drew, R.L.K., Parsons, N., 1996. Internal damage in potato tubers: a critical review. Postharvest Biol. Technol. 8, 239–258.
- McNabney, M., Dean, B.B., Bajema, R.W., Hyde, G.M., 1999. The effect of potassium deficiency on chemical, biochemical and physical factors commonly associated with blackspot development in potato tubers. Am. J. Potato Res. 75, 53–60.
- Meinl, G., 1972. Untersuchungen über die Respirationsintensität von Kartoffelknollen als Indikator für Schädigungen durch Ernte- und Aufbereitungsvorfahren. Arch. Acker- u. Pflanzenbau u. Bodenkd. 16, 21–30.
- Mempel, H., Geyer, M., 1999. Influence of mechanical stresses on the respiration activity of carrots. Gartenbauwissenschaft 64, 118–125.
- Molema, G.J., 1999. Mechanical force and subcutaneous tissue discoloration in potato. Ph.D. Thesis, Wageningen, The Netherlands.
- Oberbarnscheidt, B., Geyer, M., Herold, B., 1996. Einfluss mechanischer Belastungen auf Atmung und Masseverluste bei Speisewiebeln. Bornimer Agrartechnische Berichte 8, 139–159.
- Partington, J.C., Smith, C., Bolwell, G.P., 1999. Changes in the location of polyphenol oxidase in potato (*Solanum tuberosum* L.) tuber during cell death in response to impact injury: comparison with wound tissue. Planta 207, 449–460.
- Peters, R., 1996. Damage of potato tubers, a review. Potato Res. 39, 479–484.
- Praeger, U., Herppich, W.B., König, C., Herold, B., Geyer, M., 2010. Changes of water status, elastic properties and blackspot incidence during storage of potato tubers. J. Appl. Bot. Food Qual. 83, 1–8.
- Putz, B., 1995. Derzeitiger Wissensstand zu Blau- und Schwarzfleckigkeit bzw. Beschädigungen. Kartoffelbau 46, 284–286.
- Rickey, T.M., Belknap, W.R., 1991. Comparison of the expression of several stress-responsive genes in potato-tubers. Plant Mol. Biol. 16, 1009–1018.
- Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J., 2003. TM4: A free, open-source system for microarray data management and analysis. BioTechniques 34, 374–376.
- Scholz, M., Gatzek, S., Sterling, A., Fiehn, O., Selbig, J., 2004. Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 20, 2447–2454.
- Strehmel, N., Hummel, J., Erban, A., Strassburg, K., Kopka, J., 2008. Retention index thresholds for compound matching in GC-MS metabolite profiling. J. Chromatogr. B 871, 182–190.
- Van Canneyt, T., Dierickx, W., Verschoore, R., Ramon, H., Sonck, B., 2006. Effect of pre-load, vibration frequency, temperature and specific gravity of potato tissue on visco-elastic vibration damping and complex modulus properties. Biosyst. Eng. 94, 415–427.
- Wormanns, G., Hoffmann, T., Jacobs, A., 2000. Bildanalyse zur Bestimmung der Schwarzfleckigkeit bei Kartoffeln. Landtechnik 55, 278–279.
- Yang, W.L., Bernards, M.A., 2007. Metabolite profiling of potato (*Solanum tuberosum* L.) tubers during wound-induced suberization. Metabolomics 3, 147–159.

Please cite this article in press as: Strehmel, N., et al., Time course effects on primary metabolism of potato (*Solanum tuberosum*) tuber tissue after mechanical impact. Postharvest Biol. Technol. (2010), doi:10.1016/j.postharvbio.2009.12.008

## Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach

Matthias Steinfath<sup>1,\*</sup>, Nadine Strehmel<sup>1</sup>, Rolf Peters<sup>2</sup>, Nicolas Schauer<sup>3</sup>, Detlef Groth<sup>4</sup>, Jan Hummel<sup>1</sup>, Martin Steup<sup>4</sup>, Joachim Selbig<sup>4</sup>, Joachim Kopka<sup>1</sup>, Peter Geigenberger<sup>5,†</sup> and Joost T. van Dongen<sup>1,†</sup>

<sup>1</sup>Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

<sup>2</sup>Versuchsstation Dethlingen, Munster, Germany

<sup>3</sup>Metabolomic Discoveries, Potsdam-Golm, Germany

<sup>4</sup>Departments of Bioinformatics or Plant Physiology, Institute for Biochemistry and Biology, University of Potsdam, Potsdam-Golm, Germany

<sup>5</sup>Department Biology I, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany

Received 30 September 2009;

revised 11 February 2010;

accepted 13 February 2010.

\*Correspondence (fax +49 331 567 8136;

emails steinfath@uni-potsdam.de,

geigenberger@bio.lmu.de,

dongen@mpimp-golm.mpg.de)

†These authors contributed equally to this work.

### Summary

Biomarkers are used to predict phenotypical properties before these features become apparent and, therefore, are valuable tools for both fundamental and applied research. Diagnostic biomarkers have been discovered in medicine many decades ago and are now commonly applied. While this is routine in the field of medicine, it is of surprise that in agriculture this approach has never been investigated. Up to now, the prediction of phenotypes in plants was based on growing plants and assaying the organs of interest in a time intensive process. For the first time, we demonstrate in this study the application of metabolomics to predict agronomic important phenotypes of a crop plant that was grown in different environments. Our procedure consists of established techniques to screen untargeted for a large amount of metabolites in parallel, in combination with machine learning methods. By using this combination of metabolomics and biomathematical tools metabolites were identified that can be used as biomarkers to improve the prediction of traits. The predictive metabolites can be selected and used subsequently to develop fast, targeted and low-cost diagnostic biomarker assays that can be implemented in breeding programs or quality assessment analysis. The identified metabolic biomarkers allow for the prediction of crop product quality. Furthermore, marker-assisted selection can benefit from the discovery of metabolic biomarkers when other molecular markers come to its limitation. The described marker selection method was developed for potato tubers, but is generally applicable to any crop and trait as it functions independently of genomic information.

**Keywords:** biomarker, metabolite profiles, potato tuber, feature selection, phenotyping, gas chromatography-time of flight- mass spectrometry.

### Introduction

Biomarkers predict phenotypical properties before these features become apparent and, therefore, are valuable tools for both fundamental and applied research in medical (Geifman-Holtzman and Ober Berman, 2008; Jacobsen *et al.*, 2008) and agricultural science (Meyer *et al.*, 2007; Collard and Mackill, 2008). The predictive power of biomarkers should not be impaired by environmental variation, and the biomarkers should be applicable to populations other than those from which the markers were

originally selected. Conventional biomarkers, like AFLPs, SSRs or other polymorphisms are commonly applied in modern breeding. They provide a powerful tool to identify specific lines that are likely to manifest positive traits, thereby accelerating the otherwise cumbersome process of quality assessment and progeny selection. Indeed, marker-assisted selection (MAS) has been proven to be very successful for many diploid crops, but the application of genetic markers is still problematic for species with a complex polyploid genome which is common in many crop plants. Another restraint of genetic markers is that they are

not directly applicable for highly polygenic traits, to predict epistatic or environmental effects. To overcome these limitations of genetic markers, we developed and validated an efficient screening method that can be used to identify metabolic markers predictive of phenotypical features independently of environmental variations or of the availability of genomic information. While the investment in the technology currently used for this screening is still expensive, the power of the proposed procedure lies in the parallel selection of potential biomarkers for several different traits using a great number of samples. Of hundreds to thousands compounds, the screening method identifies small sets of metabolites, each predictive for a distinct trait. Thereby, the procedure enables on the one hand the development of targeted diagnostic assays for metabolite assisted breeding and selection. On the other hand, it could lead to a targeted investigation of the biochemical mechanism underlying the expression of the trait (Strehmel *et al.*, 2010). While the latter will be especially important for fundamental research, both results of the screening procedure will allow the breeders to improve their crops in a faster and cheaper manner thus providing better products in a shorter time.

Semi-quantitative metabolomic data were used to build a predictive model describing the development of quantitative phenotypic properties that only become apparent several months after metabolite sampling. The screening procedure is untargeted and can therefore also be used to find biomarkers for traits with yet unknown underlying biochemical mechanisms. In addition, the metabolic biomarkers were successfully applied to guide the selection of the segregating progeny of crossbreeding populations that exhibit a desired phenotype.

Our biomarker screening tool was developed on potato (*Solanum tuberosum*) tubers, but the method has no limitation to any species or application purpose. To establish and validate the method, we tested two independent, commercially important features for which the biochemical reactions and their chemical intermediates are already well described. The method, however, can be employed to identify biomarkers for any other complex trait as well, such as yield, disease resistances or stress tolerance. The first trait that was tested was the susceptibility of potato tubers to black spot bruising, which is the undesired formation of dark-blue to blackish melanin spots below the peel of the tuber after being exposed to mechanical pressure (Laerke *et al.*, 2002). This occurs when tubers collide and reduces the commercial value of the tubers drastically. Second, we analysed the commercially important brown-

ing of potato chips during frying as caused by the Maillard reaction (Zhang and Zhang, 2007). Not only does the degree of browning determine the attractiveness of the industrial end product, the chemical reaction also leads to the accumulation of carcinogenic acrylamide (Zhang and Zhang, 2007) which should be reduced to a minimum. Hereafter, this trait will be referred to as chip quality.

## Results

### Phenotypic properties: dependency on environment and genotype

Screening for metabolic biomarkers was performed on 20 different potato cultivars that were grown at two locations in northern Germany, Böhlendorf and Ebstorf. These locations differ in soil quality (loamy versus sandy soil) and in climate (coastal versus inland). Plant samples and phenotype information were collected during 2 years (2006 and 2007) which were characterized by very different weather conditions (extremely warm and dry in 2006 as compared to moderate weather in 2007). Tuber tissue samples for metabolome analysis were collected directly after harvesting and phenotypic characterization of the material was performed after 4 months while the tubers were stored under conventional conditions.

For each of the cultivars, the traits 'susceptibility to black spot bruising' and 'chip quality' were determined in three replicates, which resulted in 60 data points for each location and year. Using analysis of variance (ANOVA), the proportion of the variance within the traits that could be explained by cultivar variation was estimated. For the susceptibility to black spot bruising, the  $R^2$ -value varied between 0.9 and 0.93, whereas for chip quality this value varied between 0.82 and 0.97. The calculated  $P$ -value according to the  $F$ -test as well as the Kruskal-Wallis test describing the dependence of the traits was below  $10^{-4}$  in all cases. From these calculations, it was concluded that the datasets exhibit the required phenotypic variance to serve as training and validation sets to develop a phenotype prediction model.

Trait qualities correlated well. Spearman rank correlation coefficient varied between 0.7 and 0.92 between the cultivars grown in various years and locations (Table 1), indicating a generally high reproducibility of trait expression. However, within a single variety, the susceptibility to black spot bruising appeared to vary strongly between environments (Figure 1). Because of the evenly distributed quality values of the 2006 experiments,

902 Matthias Steinfath et al.

**Table 1** Spearman correlation comparing trait qualities between the various locations and years. Vectors consisting of the median characteristic of all cultivars at each location and in each year were correlated. Panel A shows the comparison for susceptibility to black spot bruising, and panel B indicates the values obtained for chip quality

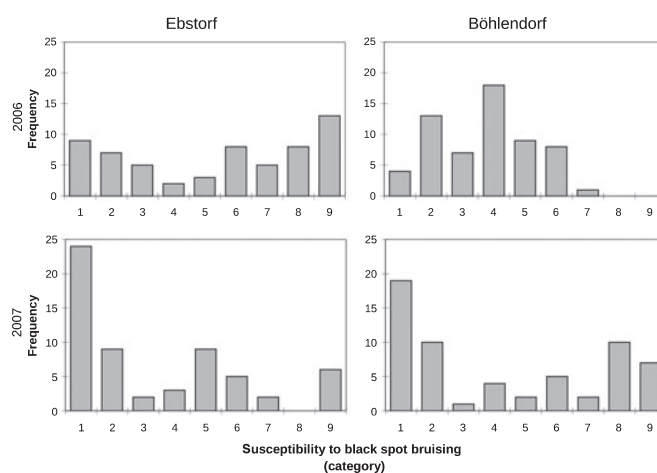
		2006		2007	
		Ebstorf	Böhlendorf	Ebstorf	Böhlendorf
A	2006	Ebstorf			
		Böhlendorf	0.79		
	2007	Ebstorf		0.79	
		Böhlendorf	0.89	0.70	0.89
B	2006	Ebstorf			
		Böhlendorf	0.65		
	2007	Ebstorf		0.55	
		Böhlendorf	0.59	0.69	0.75

this dataset was particularly suitable as training set. In 2007, a shift was observed to lower susceptibility when compared to 2006. Therefore, the 2007 dataset was used as validation set. The variation among years illustrates how environmental conditions can affect the development of a phenotype. For our purposes, this strong variation was most optimal, as a successful phenotype prediction under highly variable conditions validates the method to be robust against environmental changes. For

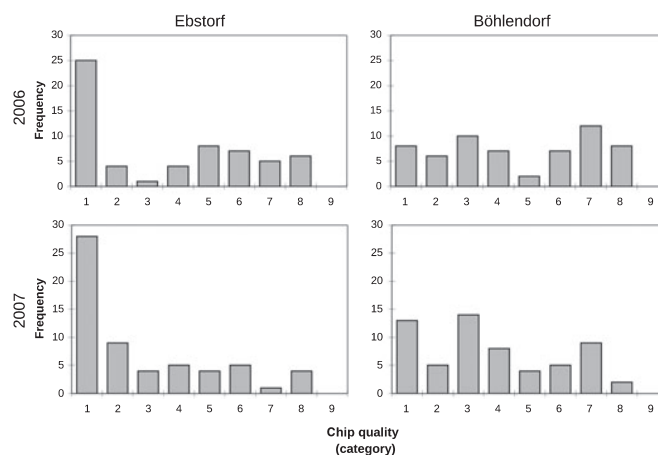
chip quality, no parameter shift was detected between growing seasons (Figure 2).

#### Selection of putative biomarkers

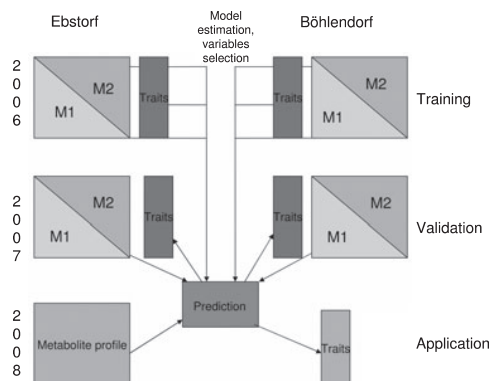
The selection of potential metabolic biomarkers from our gas chromatography–time of flight–mass spectrometry (GC–TOF–MS) derived metabolic profiles (Erban *et al.*, 2007) was based on a machine learning procedure called feature or variable selection (Guyon *et al.*, 2006). Those variables that were regarded as potential biomarkers for a given trait were represented within a GC–TOF–MS metabolite profile by their relative mass fragment intensities at corresponding retention-time windows. These will be referred to as mass features hereafter. The first step of the marker selection procedure was to assign a numerical value which indicates the importance of a mass feature for the prediction of the trait. For this purpose, we used the partial least squares (PLS) (Wold, 1975) – variables importance in the projection (VIP) method (Chong and Jun, 2005). PLS is a dimensionality reduction method necessary for datasets for which the number of variables exceeds the number of samples. The VIP value gives an estimate of the contribution of a given predictor to a PLS regression. Mass features with a VIP value (Perez-Enciso and Tenenhaus, 2003) above 2 were regarded as potential biomarkers. For each trait, the intersection of potential biomarkers was determined by comparing the profile



**Figure 1** Frequency distribution of the susceptibility scores for black spot bruising in different years and locations. Each histogram represents a single field experiment, with 60 samples (i.e. three replications for each of the 20 cultivars). The trait validation scores range from 1 to 9 (see methods part). From the figure it becomes clear that the susceptibility score distribution varies for each location and year.



**Figure 2** Frequency distribution of chip quality obtained from different years and locations. Each histogram represents a single field experiment with 60 samples (three replications for each of the 20 cultivars). The chip quality score ranges from 1 to 9 (see methods part). Note that the Ebstorf 2007 field exhibited a narrow phenotypic variation with more than 30 of the 60 samples below score 2. Furthermore it becomes clear from this figure that the location had a stronger influence on the chip quality score distribution than the year of analysis.



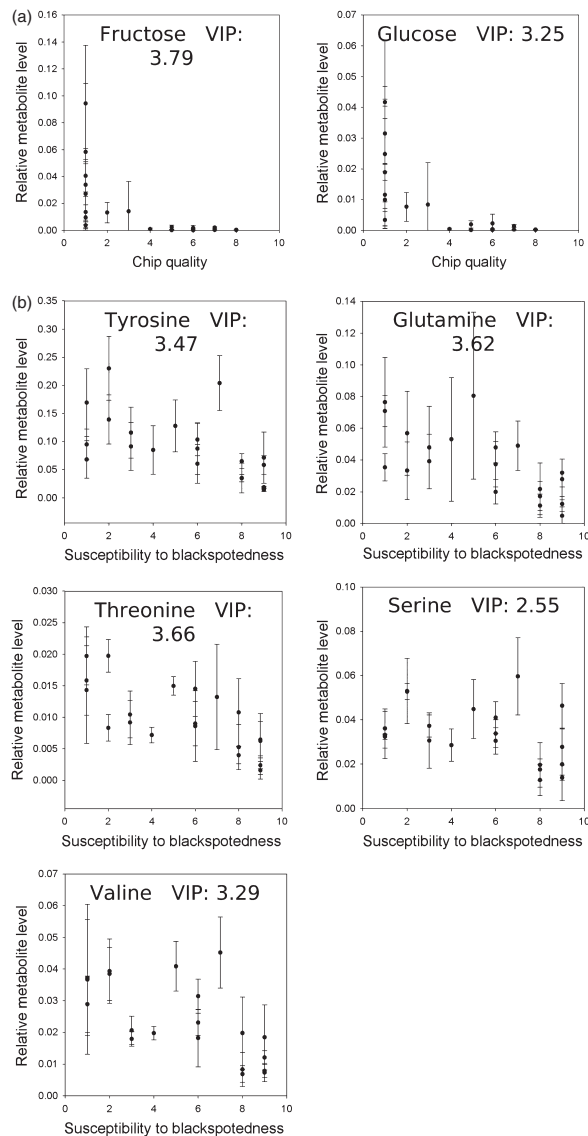
**Figure 3** Overview of experimental design and predictive marker discovery: For each year (2006, 2007) and location (Böhlendorf, Ebstorf) metabolite profiles were measured using 20 cultivars. In each year measurements were performed at two locations. The GC-TOF-MS profiling analysis was repeated twice using an additional set of replicated tubers from the same harvest (M1, M2). Metabolites were selected as putative biomarkers, if their VIP value was higher than 2 at both locations and both replicate analyses M1 and M2 of the year 2006 (variables selection). For 2008 the GC-TOF-MS measurements were carried out once using 58 novel breeding lines. The putative biomarkers were applied to the 2008 data using the models trained on the 2006 training sets.

results obtained from two different locations. Furthermore, two complete analytical replicated GC-TOF-MS experiments were performed using a second set of tuber

samples from the same harvests. The selected mass features of both measurements were identified and only those metabolites that were detected in both analytical replicate measurements were considered as potential biomarkers. Figure 3 visualizes the procedure and the resulting datasets schematically.

The robustness of the marker selection was improved by subjecting the dataset to an iterative variation of the underlying data. For this purpose, 20 alterations of each dataset were generated by each time excluding the data of a different single cultivar. The metabolites tyrosine, threonine, valine, serine and glutamine were determined to be robust and relevant in all of the subsets for the susceptibility to black spot bruising. For chip quality, glucose and fructose were detected as potential biomarkers in all subsets. The correlation for the material obtained in 2006 from Ebstorf between the measured levels of these selected metabolites and the corresponding trait qualification is visualized in Figure 4. While good chip quality always requires low fructose and glucose levels, there are a few cultivars which exhibit low chip quality despite of low fructose and glucose concentrations. With respect to black spot bruising, we observed a negative trend of metabolite concentrations with rising susceptibility. For some metabolites in specific potato varieties, outliers were observed but as the predictive model is built on the combination of various metabolic markers the impact of these exceptional cases is secondary. A full overview of the

904 Matthias Steinfath et al.



**Figure 4** Levels of selected biomarker metabolites in relation to the corresponding trait quality. Fructose and glucose were selected as their levels corresponded with chip quality (a). Serine, valine, tyrosine, threonine, and glutamine are biomarkers for the susceptibility to blackspot bruising (b). The data depicted in these graphs represent the metabolite levels as indicated by their relative  $m/z$  intensities as normalized sample fresh weights. These data are not rank based normalized in contrast to those used for the prediction model. Here, only those analytes were used, the spectroscopic properties of which are described in SI table 3. The data were obtained from tubers harvested in 2006 at the locations Ebstorf for the first technical replicate. Dots indicate the mean value over three biological replicates, error bars indicate the corresponding standard deviation. A full overview of all metabolites that were identified throughout the various experiments is given in SI table 1. The importance of each metabolite for the prediction of the corresponding trait is indicated by the VIP value on the top of each plot.



levels of all metabolites that were identified throughout the various experiments is given in Table S1.

#### Validation of the predictive power of the marker metabolites

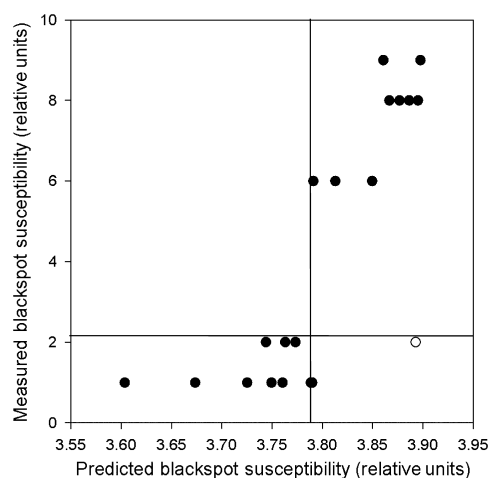
The predictive power of the selected potential biomarkers was tested by building a statistical model that correlates metabolic and phenotypic information obtained from the first environment, to predict phenotypic properties from the metabolic data obtained from the second environment. The selected metabolites were used to estimate a PLS regression model for both locations and technical replicates of the first year (2006). By doing so, we could apply four models to four validation sets. As a measure of the predictive power of these models, we took the Spearman rank correlation between the measured and the estimated traits of the validation sets. These correlations ranged between 0.53 and 0.82 (Table 2, panel A) for the susceptibility to black spot bruising, and for chip property these values varied between 0.66 and 0.75 (Table 2, panel B). From these high-correlation values, it was concluded that the prediction model is able to give a reliable indication of the quality of a predicted trait.

**Table 2** Determination of the predictive power of the selection of robust metabolic biomarkers. Prediction models as derived from the correlation between metabolites and the trait were trained on four datasets obtained in 2006 (trainings set) and applied to the four datasets as obtained in 2007 (validation set). The table presents Spearman rank correlation factors between predicted and measured values for the corresponding combination of training and validation set. Panel A shows the predictive power for the susceptibility to black spot bruising, and panel B lists the prediction of chip quality. The analytical repetitions of the GC-TOF-MS measurements with independent biological replicate tubers of the same harvests are indicated by M1 and M2, respectively

	Training set (2006)	Validation set (2007)			
		Ebsterf		Böhlendorf	
		M1	M2	M1	M2
<b>A</b>					
Ebsterf	M1	0.68	0.74	0.74	0.77
	M2	0.68	0.72	0.78	0.82
Böhlendorf	M1	0.69	0.73	0.66	0.82
	M2	0.71	0.64	0.53	0.79
<b>B</b>					
Ebsterf	M1	0.74	0.70	0.72	0.74
	M2	0.75	0.70	0.72	0.74
Böhlendorf	M1	0.71	0.66	0.70	0.71
	M2	0.74	0.70	0.71	0.75

#### Permutation tests for validating the statistical significance of the predictive power

A proper prediction model requires a true correlation between the selected biomarkers and the predicted trait quality rather than being biased by coincidental correlation between traits across the environments or years. This requirement was tested for our model by comparing our selection of potential biomarkers with results obtained from a random dataset in which our four trait vectors were subjected to identical permutations. By doing so, the correlation between traits and metabolite profiles was destroyed, whereas the correlation within a set of metabolites and between traits across location and year was conserved. As a statistical test, we used the rank correlation between estimated and measured traits averaged over all four models. For each of the 1000 permutations, we tested the average correlation which was either lower than what was calculated for the real data or no potential markers could be detected. Therefore, it was concluded that our procedure yields biomarkers that can be employed to truly predict the investigated traits.

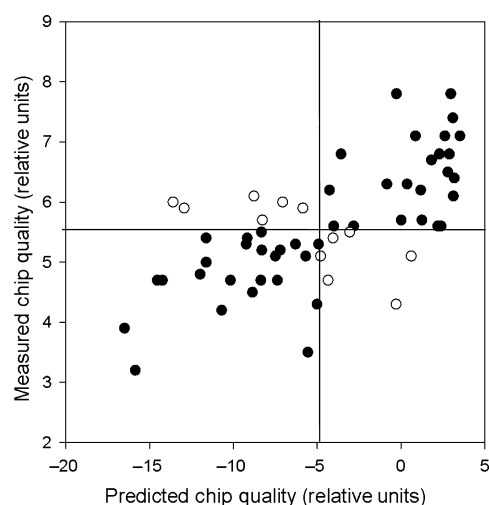


**Figure 5** Correlation between predicted and measured susceptibility to black spot bruising. The vertical and horizontal grey lines are thresholds separating high and low quality cultivars according to the predicted (vertical) and measured (horizontal) susceptibility score of black spot bruising. Cultivars for which the prediction of the trait was confirmed by the real measurements are indicated by filled dots, whereas the phenotype of the cultivar indicated by a empty dot was not properly predicted. The training set of the depicted analysis was the Böhlendorf location in 2006 and analytical replication M1, applied to the test set, Böhlendorf 2007 (M2). All 20 cultivars are displayed.

Figure 5 shows the correlation between the real values of the phenotypic properties of the Böhlendorf material obtained in 2007 and the estimated quality values as determined using the prediction model based on data from 2006. The data show that ten of the eleven cultivars that are least susceptible to black spot bruising were also predicted to have low susceptibility. Furthermore, from the ten cultivars that were predicted to have high susceptibility to black spot bruising only one was actually resistant.

#### Phenotype prediction of genotypes of a segregating breeding population

A second successful application of the metabolic biomarkers is illustrated by the ability to predict the chip quality of 58 individual lines obtained from a segregating crossbreeding population (Figure 6). For this analysis, two different potato varieties were crossed, and the offspring was propagated through cloning, thereby providing enough plant material to perform a thorough phenotype- and metabolic analysis on each of the individual offspring lines. All lines were grown simultaneously in Ebstorf,



**Figure 6** Correlation between the predicted chip quality and the measured score values of 58 different lines obtained from a single crossbreeding population. Symbol indications are as explained for Fig 1. The data shows that the phenotypes of a major proportion of the lines that were tested could be properly predicted. The training set originated from the data obtained the first technical repeat (M1) from Ebstorf in 2006.

Germany, during the growing season of 2008. Similar as for the test populations, tuber samples were taken directly after harvesting, whereas chip quality was determined for each line individually during tuber storage. The statistical models were trained on the four datasets from 2006 and applied to the metabolome profiles of the siblings of the crossbreeding population grown in 2008. Only those metabolites were considered that had been determined to represent robust candidate biomarkers as described previously. The Spearman rank correlation between estimated and measured chip quality properties of the various lines of the population ranged between 0.65 ( $P$ -value  $2.3 \times 10^{-8}$ ) and 0.68 ( $P$ -value  $5.4 \times 10^{-9}$ ), which is similar to the correlations obtained when the same cultivars were tested as those for which the model was build initially. Figure 6 shows an example: among the 29 lines of the breeding progeny that were predicted to have best chip quality properties, 23 lines (79%) truly belong to this class. Similar correlations were obtained when only fructose (0.67) or glucose (0.67) was used. The prediction of the susceptibility to black spot bruising within the crossbreeding population was not statistically significant. The results for chip quality show that the prediction model is able to ameliorate the selection efficiency of a breeder who aims to select appropriate lines from a crossbreeding population with considerable variation of the trait and a cumulative improvement is expected when the metabolic biomarker tool is applied on subsequent generations.

#### Discussion

An efficient comprehensive and untargeted procedure is described to screen for metabolic biomarkers that can be employed for predicting crop product quality several months in advance. The screening led to the identification of various metabolites that were shown valid for use as diagnostic biomarkers in postharvest quality trait identification of potato tubers (Figure 5). Furthermore, it was shown that these biomarkers can be applied to select progeny with a desired trait from a segregating crossbreeding population (Figure 6). This shows that metabolic biomarkers can be applied to genotypes other than those for which the biomarkers were selected. Thus, one of the objectives – the detection a small number of markers, which could be used for a targeted assay – was achieved: Two metabolites predictive for chip quality were identified of hundreds of compounds and were shown to predict the phenotype of individual lines of a segregating crossbreeding population.

As metabolite levels are particularly susceptible to environmental changes, it was important to identify markers, which are robust in a wide range of soil and climate conditions. The experimental design guaranteed that the robustness against such changes of the potential markers could be tested. Biomarkers could be identified in this experiment and thus validates the described method. The statistical significance of the procedure could be demonstrated by permutation tests. The application of the procedure was shown to be useful to select individuals from a segregating crossing-population having quality parameters that are significantly favourable as compared to a random selection. Especially in the early stage of the selection process, where the abundance of independent lines is high, but the number of individuals per line is low, such predictive tool will help to significantly improve the selection efficiency of a breeder.

Despite the predictive power of the biomarkers that were selected, the current marker selection could not guarantee failure-free selection. A few cultivars and crosses that had relatively low levels of fructose and glucose after harvesting appeared to develop low chip quality after 4 months (Figure 6). Apparently, these samples are classified incorrect by our prediction method suggesting that at least for those genotypes more factors are coresponsible for the discoloration than only those identified by our screening tool. This problem might occur when the amount of such 'outlier' genotypes is too small in the training set. To solve this problem, the prediction model can be made more robust and improved by increasing the sample number and sample variation of the training set by including more cultivars, crosses or environmental conditions. Nevertheless, the low amount of false-positive or false-negative predictions indicates that the requirements can be met for selecting proper biomarkers that predict the quality of a crop harvest well before products are sold or processed.

To further test the reliability of our screening method, we choose to establish the biomarker selection tool on two different commercially relevant traits for which already considerable biochemical knowledge existed. This allowed us to validate the potential biomarkers that were selected by comparing them with the existing information about the biochemical nature of the features. For chip quality, the identification of fructose and glucose as metabolic biomarker confirmed the successful application of the metabolic biomarker screening tool as both reducing sugars are well described to be involved in the Maillard reaction (Rodríguez-Zaona and Wrolstad, 1997; Zhang and Zhang, 2007).

Also tyrosine as a potential indicator for susceptibility to black spot bruising was discussed in the literature before. The dark blue-blackish colorization of potato tuber tissue that occurs during black spot bruising is caused by an enzymatic oxidation reaction of phenolic compounds, like tyrosine, into melanin which is catalysed by polyphenol oxidases (Partington *et al.*, 1999). Indeed, Stevens and Davelaar (1997) observed that the black spot potential, which they defined as the maximum possible degree of dark pigment formation of the tissue, is positively correlated with the tyrosine content. However, the black spot susceptibility which Stevens and Davelaar determined by size and colour intensity of the discoloured spot on a tuber after impact did not correlate with tyrosine content. Also another study using 115 potato varieties failed to determine a direct correlation between free tyrosine content and black spot susceptibility (Vertregt, 1968). On the other hand, Corsini *et al.* (1992) described a positive correlation between the content of free tyrosine (i.e. tyrosine that is not incorporated into protein) and the susceptibility to black spot bruising of 12 varieties throughout various growing seasons. Putting all these literature data together leads to the conclusion that tyrosine is an important factor in the formation of black spot discoloration, but its concentration alone is not a good indicator for the susceptibility to black spot bruising.

The great advantage of our biomarker search tool is the broad screening approach that is being used, which allows not to rely on the use of individual markers only, but combines several metabolic markers simultaneously. Not only does this improve the reliability of the markers, it can also lead to the discovery of novel components, not yet investigated within the context of the trait being investigated, as was shown here for the susceptibility to black spot bruising. Thereby, another goal of our screening procedure was achieved. The discovery of novel chemical components can help to guide further research that is directly targeted to improve the understanding of biochemical reactions underlying complex traits.

For the purpose of this untargeted screening, we used a well-established GC-TOF-MS metabolome analysis technique that requires specific technical knowledge and experience that might not be present in every breeding or processing company. Therefore, once a selection of metabolic biomarkers has been made, the next step would be to implement these biomarkers in agricultural or agronomical praxis by developing fast, targeted and low cost diagnostic dipstick biomarker assays. These assays could then

readily be used in the preselection process of segregating crossing populations. This is especially relevant when phenotyping is expensive and time consuming for postharvest quality traits.

The described method to screen for biomarkers is generally applicable to all crops and traits, and not necessarily to postharvest quality traits only. The results obtained here for potato properties encourage the application of the procedure to other crops and other traits, including those for which the underlying biochemical mechanisms are yet unknown. This can become especially important when the application of modern breeding tools involving MAS to characterize genetic variation between plants becomes limited. Although such genetic markers are robust and cheap tools for many traits and crops, it comes to limitations when breeding polyploidic crops for strongly environmentally controlled traits. In such cases, the combination of genetic and metabolic markers could be further investigated as promising approach for traits that have complex molecular causes (Gärtner *et al.*, 2009). However, metabolite-assisted breeding will offer new opportunities especially for crops for which hardly any genomic information is available or when the complexity of the genetic mechanisms underlying the trait is too high. The screening tool that we described and validated here improves the identification of metabolic biomarkers. This will help to increase the potential of metabolic biomarkers to become a readily employed tool to advance crop breeding.

## Experimental procedures and statistical methods

### Plant cultivation and phenotype description

To obtain a large variation in growth conditions between the various plant cultures that were used for marker selection in this study, two different fields were chosen on which 20 different commercially available potato cultivars were grown. The first field was located near the village Ebstorf, located in the middle of Germany (53°01'43.62"N; 10°24'54.61"E). This area belongs to the optimal regions for high yield potato cultivation. The soil is comprised of loess. The second field was close to the village Böhlendorf (54°35'59.95"N; 13°17'46.12"E) which is located only 35 km away from the Baltic Sea coast line in the North of Germany. The soil in this area is composed of clay. This area is generally not valued for high yield, but because of the strong coastal influence with relatively strong winds, the degree of infection and insect plagues is considerably lower. Plants were grown during two subsequent years that were characterized with highly different climate and weather conditions. During the growing

season in 2006, the months June and July were extremely warm and dry. In Ebstorf, the median maximum temperature during the summer months was 30 °C. In Böhlendorf, the median of the maximum temperature during the summer months of 2006 was 25 °C. The total amount of precipitation during the entire growing period was (231 mm rain plus 50 mm artificial watering in Ebstorf, whereas in Böhlendorf a natural amount of precipitation of 350 mm was measured). During the growing season of 2007, the median  $T_{max}$  was 24 and 21 °C for Ebstorf and Böhlendorf, respectively. The amount of rainfall during the entire growing season was 370 and 426 mm in Ebstorf and Böhlendorf, respectively. Further climate details can be found at the following web resource (<http://www.wetteronline.de> (05/04/2009)).

The cultivars used for the analysis were Afra, Agria, Alliance, Arnika, Calla, Cilena, Elfe, Festien, Goldika, Granola, Kuras, Lady Clair, Marabel, Milva, Omega, Red Fantasy, Roberta, Tomensa, Topas and Turdus. The material was cultivated, harvested and stored according to common agricultural practice. Each cultivar was grown on six independent plots that were randomly located on the field. Each plot contained 20 plants from which all tubers were harvested. Congruent to standard agricultural procedures, only tubers that were 5–8 cm in diameter were selected for further analysis. Sampling of tuber material for metabolome profiling was carried out 1 week after harvesting in September, and before the material was transferred into the final store. The proximal and distal part of the tuber were discarded and the middle part of individual tubers including the outer cell layers (epidermis and cork cells) were sliced, snap frozen in liquid nitrogen, ground to fine powder and stored at –80 °C until further sample processing (Hoehewarter *et al.*, 2008). Four months after tubers were harvested and metabolite samples were taken, the stored material was subjected to phenotyping tests for several agronomical important features. To investigate the susceptibility to black spot bruising, approximately 8 kg of tubers were centrifuged at 380 rpm in a trommel (Type KS12; Mafago Maschinenfabrik GmbH, Bad Gottleuba, Germany) for 50 s at 4–6 °C. Subsequently, the tubers were stored for 4–5 days at room temperature, and the degree of black spot bruising was scored visually for 20 tubers that were cut in halves and the gradation of black spot formation was scored using a standardized scoring table. The different categories are as follows: 0, no spot formation; L, up to 25% of the surface is coloured, but not deeper than 5 mm from the skin; M, up to half of the tuber surface is coloured, but not deeper than 5 mm and/or <25% of the surface has coloured spots that reach more than 5 mm to the inside of the tuber; S, more than half of the surface is covered with coloured spots that are not deeper than 5 mm and/or up to half of the surface is coloured more than 5 mm towards the inside of the tuber. Indexing of the probe categories was done for each triplicate sample according to the following formula:  $(n_L \times 1) + (n_M \times 3) + (n_S \times 5) / n_{(total)}$  in which  $n$  stands for the number of tubers present in the respective categories. According to these indices, the cultivars were assigned into nine classes, labelled with values from 1 to 9. The lowest class index stands for tubers that are not susceptible for black spot bruising, whereas a high index number indicates a high susceptibility.

The second phenotype test on chip quality was performed 4 months after the initial samples for metabolite profiling were taken. Per sample, single slices of 1.2–1.6 mm thick were taken

from 10 tubers and fried in a 12-L oil pan at 170 °C until the rest moisture of the tissue was approximately 2% as indicated by the oil to stop forming bubbles during frying. The browning of each individual slice was evaluated and scored according to the European standard colour scale for fried potatoes developed by the Institute for storage and processing of agricultural products (Instituut voor bewaring en verwerking van landbouwprodukten, Wageningen -IVBL-, Wageningen, the Netherlands) which contains a numbered series of nine photographs of differently coloured fried tuber slice (Ludwig, 1978). The lowest value (1) indicates a dark brown colour, whereas the highest value (9) corresponds to the highest quality (light and yellowish).

#### Sample preparation and metabolite profiling by GC-TOF-MS

The following experiments were carried out according to a previously established routine protocol for GC-TOF-MS-based metabolite profiling (Erban *et al.*, 2007). Aliquots of 50 mg deep frozen and ground potato material were subjected to 300 µL methanol extraction, with 30 µL nonadecanoic acid methylester (2 mg/mL in chloroform), 30 µL mix of internal standards (cf. Table S2) and incubated 15 min at 70 °C. Then 600 µL chloroform-water (1 : 2, v/v) was added, phase separation performed and 10 µL of the resulting upper polar phase reduced in vacuo. The subsequent derivatization was conducted in two steps, i.e. first 10 µL methoxyamine hydrochloride (40 mg/mL in pyridine) agitated 90 min at 30 °C for 90 min and second by adding 90 µL reagent mixture, comprising N-methyl-N-trifluoroacetamide and a n-alkane-mixture (decane, dodecane, pentadecane, octadecane, nonadecane, docosane, octacosane, dotriacontane and hexatriacontane (1000 : 16, v/v)) with agitation at 37 °C for 30 min. One micro litre of the derivatized sample was injected into an Agilent 6890N24 gas chromatograph and analysed as described by Erban *et al.* (2007) with the help of a Pegasus III time of flight mass spectrometer (Leco, St. Joseph, MI, USA).

#### Data preprocessing and storage of data

The resulting NetCDF files were base line corrected using Chroma-Tof software (Leco) and imported into the TagFinder software (Luedemann *et al.*, 2008) for data preprocessing and metabolite annotation and identification. The minimum intensity threshold was set to 50 arbitrary ion current units. The chromatographic alignment was conducted via the retention index model of Van den Dool and Kratz (1963). Repeatedly occurring GC-TOF-MS mass features across all constituent profiles of an experimental dataset, i.e. ion-impact ionized mass fragments, were grouped into time groups of mass fragments coeluting within the same retention index window. Mass fragments of compounds coeluting within the same time group were partitioned into clusters by correlation of ion current measurements across all profiles constituting an experimental dataset. The mass features, which were subjected to the variables selection, were either those cluster in one time group which could be annotated or which contain the maximal number of mass fragments. The resulting data matrix containing the sum intensity of each mass spectral feature, i.e.

the selected cluster of the time groups, and the corresponding sample information such as cultivar, location and year was then imported into the Golm Metabolome Database (GMD) (Kopka *et al.*, 2005). In the second step, the metabolite annotations of time group and cluster identity were added using the TagFinder software and attached to GMD upload of the respective experiments. Finally the responses (ion current measures) of the quantitative internal standards (Table S2) and sample fresh weight data were deposited for the automated calculation of relative pool size changes among the samples of an experimental dataset using the GMD database backend.

#### Rank based normalization

For each mass fragment ranks over all cultivars or lines in one environment (location, year) were determined. For each cluster, the sum of the ranks of the mass fragments was used as intensity value of the cluster.

#### Identification of potential metabolic biomarkers

The mass features, namely those clusters or time groups found to be important for the prediction of a trait, were identified via a TagFinder supported matching to the currently build of the GMD library comprising 1192 metabolites and 2073 chemical derivatives, each represented as a nonredundant mass spectrum with a characteristic retention index. To facilitate the retention index matching process, we used retention index windows of 0.5–1.0% retention index units which represent the previously validated retention index thresholds (Strehmel *et al.*, 2008). The mass spectral matching procedure was manually supervised and required least three matching mass fragments to be present within the predicted retention index windows (Table S3). The reported metabolic markers were additionally confirmed by standard addition of a multiplexed chemical mixture.

#### Regression, dimensionality reduction and variables selection

Principal component analysis (PCA) is a widely used unsupervised dimensionality reduction method. It seeks a combination of descriptive features, which maximize the variance covered by the variance ranked principal components. However, the PCA is performed without consideration of the sample trait or class properties under investigation.

PLS (Wold, 1975; Erikson *et al.*, 2001; Barker and Rayens, 2003) method offers a so-called supervised alternative to PCA. PLS identifies combinations of the original predictor variables with the maximum covariance with the response, i.e. the predicted property. These orthogonal combinations replace the original data matrix and are used in multivariate ordinary least squares regression to predict the response.

The variables selection method employed in our investigation was based on PLS. However, the contribution of each variable to the response is measured by the VIP method (Perez-Enciso and Tenenhaus, 2003; Chong and Jun, 2005). This procedure is defined by

the variable's importance in the projection of the data in such a multidimensional space of latent variables which corresponds best to the response. The original describing variables are sorted in decreasing order according to their VIP value. To decide which variables are considered to be important a cut off criterion is needed. We take a fixed threshold of  $VIP = 2$  throughout the whole procedure. As the sum of the squares of the VIPs for all variables is equal to the number of variables (Perez-Enciso and Tenenhaus, 2003), the threshold is four times higher than the average squares. Perez-Enciso and Tenenhaus (2003) applied this threshold successfully in case of a large number of variables on microarray data.

### Acknowledgements

The research described in this manuscript was funded by a grant of the German Federal Ministry of Education and Research (BMBF): QuantPro-InnOx 'Innovative diagnostic tools to optimize potato breeding: systemic analysis of cellular processes and their relation to plant internal oxygen concentrations' (PTJ-BIO/Wir/0101-31P4295). Silke Ulrich, Ines Fehrle, Anke Langer and Hildegard Hopp are acknowledged for their technical assistance. Eckard Tacke (BIOPLANT, Biotechnologisches Forschungslabor GmbH) and Hans-Reinhard Hofferbert (Böhm-Nordkartoffel-Agrarproduktion OHG) provided plant material.

### References

- Barker, M. and Rayens, W. (2003) Partial least squares for discrimination. *J. Chemometr.*, **17**, 166–173.
- Chong, I.-G. and Jun, C.-H. (2005) Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.*, **78**, 103–112.
- Collard, B.C. and Mackill, D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **363**, 557–572.
- Corsini, D.L., Pavek, J.J. and Dean, B. (1992) Differences in free and protein-bound tyrosine among potato genotypes and the relationship to internal blackspot. *Am. Potato J.*, **69**, 423–435.
- Erban, A., Schauer, N., Fernie, A.R. and Kopka, J. (2007) Metabolomics: methods and protocols. In *Nonsupervised Construction and Application of Mass Spectral and Retention Time Index Libraries from Time-of-Flight Gas Chromatography-Mass Spectrometry Metabolite Profiles*, (Weckwerth, W., ed.), pp. 19–38. Totowa: Humana Press.
- Erikson, L., Johanson, E. and Kettaneh-Wold, N. (2001) *Multi- and Megavariate Data Analysis; Principle and Applications*. Umea, Sweden: Umetrics Academy.
- Gärtner, T., Steinfath, M., Andorf, S., Lisec, J., Meyer, R.C., Altmann, T., Willmitzer, L. and Selbig, J. (2009) Improved heterosis prediction by combining information on DNA- and metabolic markers. *PLoS ONE*, **4**, e5220.
- Geifman-Holtzman, O. and Ober Berman, J. (2008) Prenatal diagnosis: update on invasive versus noninvasive fetal diagnostic testing from maternal blood. *J. Expert. Rev. Mol. Diagn.*, **8**, 727–751.
- Guyon, I., Gunn, S., Nikravesh, M. and Zadeh, L.A. (eds) (2006) *Feature extraction. Foundation and Application*. Berlin Heidelberg: Springer.
- Hoehenwarter, W., van Dongen, J.T., Wienkoop, S., Steinfath, M., Hummel, J., Erban, A., Sulpice, R., Regierer, B., Kopka, J., Geigenberger, P. and Weckwerth, W. (2008) A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. *Proteomics*, **8**, 4214–4225.
- Jacobsen, M., Mattow, J., Reipsilber, D. and Kaufmann, S.H. (2008) Novel strategies to identify biomarkers in tuberculosis. *Biol. Chem.*, **389**, 487–495.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A.R. and Steinhauser, D. (2005) gmd@csb.db: the golm metabolome database. *Bioinformatics*, **21**, 1635–1638.
- Laerke, P.E., Christiansen, J. and Veierskov, B. (2002) Colour of blackspot bruises in potato tubers during growth and storage compared to their discolouration potential. *Postharvest Biol. Technol.*, **26**, 99–111.
- Ludwig, J.W. (1978) *Het meten en beïnvloeden van de chipskleur*. Wageningen, the Netherlands: IBVL-publicatie 315, IBVL.
- Luedemann, A., Strassburg, K., Erban, A. and Kopka, J. (2008) TagFinder for the quantitative analysis of gas chromatography—mass spectrometry (GC–MS)-based metabolite profiling experiments. *Bioinformatics*, **24**, 732–737.
- Meyer, R.C., Steinfath, M., Lisec, J., Becher, M., Witucka-Wall, H., Törjék, O., Fiehn, O., Eckardt, A., Willmitzer, L., Selbig, J. and Altmann, T. (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, **104**, 4759–4764.
- Partington, J.C., Smith, C. and Bolwell, G.P. (1999) Changes in the location of polyphenol oxidase in potato (*Solanum tuberosum* L.) tuber during cell death in response to impact injury: comparison with wound tissue. *Planta*, **207**, 449–460.
- Perez-Enciso, M. and Tenenhaus, M. (2003) Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum. Genet.*, **112**, 581–592.
- Rodriguez-Zaona, L.E. and Wrolstad, R.E. (1997) Influence of potato composition on chip color quality. *Am. Potato J.*, **74**, 87–106.
- Stevens, L.H. and Davelaar, E. (1997) Biochemical potential of potato tubers to synthesize blackspot pigments in relation to their actual blackspot susceptibility. *J. Agric. Food Chem.*, **45**, 4221–4226.
- Strehmel, N., Hummel, J., Erban, A., Strassburg, K. and Kopka, J. (2008) Retention index thresholds for compound matching in GC-MS metabolite profiling. *J. Chromatogr. B*, **871**, 182–190.
- Strehmel, N., Praeger, U., König, C., Fehrle, I., Erban, A., Geyer, M., Kopka, J. and van Dongen, J.T. 2010. Mechanical impact has characteristic time course effects on primary metabolism of differentially susceptible potato (*Solanum tuberosum*) tuber tissue. *Postharvest Biol. Technol.*, **56**, 109–116.
- Van den Dool, H. and Kratz, P.D. (1963) A generalization of retention index system including linear temperature programmed dgas-liquid partition chromatography. *J. Chromatogr.*, **11**, 463–471.

- Vertregt, N. (1968), Relation between black spot and composition of the potato tuber. *Eur. Potato J.*, **11m**, 34–44
- Wold, H. (1975) *Soft Modelling by Latent Variables*. London: Academic Press.
- Zhang, Y. and Zhang, Y. (2007) Formation and reduction of acrylamide in maillard reaction: a review based on the current state of knowledge. *Crit. Rev. Food Sci. Nutr.*, **47**(5), 521–542.

### Supporting information

Additional Supporting information may be found in the online version of this article:

**Table S1** Full overview of the levels of all metabolites that were identified throughout the various experiments.

**Table S2** List of the internal standards to calibrate the retention times of a metabolite profile obtained from the GC-TOF-MS analysis.

**Table S3** Spectroscopic properties of the identified derivatives.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## Decision tree supported substructure prediction of metabolites from GC-MS profiles

Jan Hummel · Nadine Strehmel · Joachim Selbig · Dirk Walther · Joachim Kopka

Received: 30 November 2009 / Accepted: 25 January 2010 / Published online: 16 February 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Gas chromatography coupled to mass spectrometry (GC-MS) is one of the most widespread routine technologies applied to the large scale screening and discovery of novel metabolic biomarkers. However, currently the majority of mass spectral tags (MSTs) remains unidentified due to the lack of authenticated pure reference substances required for compound identification by GC-MS. Here, we accessed the information on reference compounds stored in the Golm Metabolome Database (GMD) to apply supervised machine learning approaches to the classification and identification of unidentified MSTs without relying on library searches. Non-annotated MSTs with mass spectral and retention index (RI) information together with data of already identified metabolites and reference substances have been archived in the GMD. Structural feature extraction was applied to sub-divide the metabolite space contained in the GMD and to define the prediction target classes. Decision tree (DT)-based prediction of the most frequent substructures based on mass spectral features and RI information is demonstrated to result in highly sensitive and specific detections of sub-structures contained in the compounds. The underlying set of DTs can be inspected by the user and are made available for batch processing via SOAP (Simple Object Access Protocol)-based web services. The GMD mass spectral library with the integrated DTs is freely

accessible for non-commercial use at <http://gmd.mpimp-golm.mpg.de/>. All matching and structure search functionalities are available as SOAP-based web services. A XML + HTTP interface, which follows Representational State Transfer (REST) principles, facilitates read-only access to data base entities.

**Keywords** Metabolic markers · Gas chromatography (GC) · Mass spectrometry (MS) · GC-MS · Mass spectral classification · Mass spectral profiling · Metabolite fingerprinting · Metabolite profiling · Metabolomics · Metabonomics · Decision trees

### 1 Introduction

The identification of the high number of as of yet unidentified metabolic components from GC-MS profiling experiments poses a major challenge in metabolite profiling. Two factors contribute to the high complexity of typical GC-(TOF)-MS experiments. Firstly, as GC-MS inherently requires volatile analytes, metabolites of interest need to be chemically modified, for example by methoxyamination and silylation reagents (Kopka 2006, Lisec et al. 2006). Thus, more than one single analyte per metabolite may be generated and thorough chemical interpretations of observed analytes with respect to their mass spectral and retention index (RI) properties are required. Secondly, a compound library comparison as the most straightforward approach for identifying components from GC-MS analyses relies on the availability of authenticated pure reference substances. Currently, metabolite identification is only possible by a time consuming, manually supervised matching of both the RI information and the reference mass spectra stored in dedicated libraries such as the Golm Metabolome Database

J. Hummel · N. Strehmel · D. Walther · J. Kopka (✉)  
Department Prof. L. Willmitzer, Max Planck Institute of  
Molecular Plant Physiology, Am Muehlenberg 1, 14476  
Potsdam-Golm, Germany  
e-mail: kopka@mpimp-golm.mpg.de

J. Selbig  
Institute for Biochemistry and Biology, University of Potsdam,  
Karl-Liebknecht-Strasse 24-25, Haus 20, 14476 Potsdam-Golm,  
Germany



(GMD) (Kopka et al. 2005, Schauer et al. 2005). Therefore, the lack of chemically synthesized reference substances and of purified bio-molecules may be seen as the current bottleneck of comprehensive compound identification as identification is only possible if the detected compound is present in the library of reference compounds.

In order to provisionally accommodate unidentified compounds, the GMD archives respective Mass Spectral Tags (MSTs). MSTs are defined to represent the combination of chemo-physical properties, namely the mass fragmentation pattern linked to the chromatographic RI information (Wagner et al. 2003). In addition, the GMD also comprises a large compendium of identified compound entries. These entries represent known metabolite structures and are linked to the source (vendor etc.) information of the respective reference substances. Thus, the GMD may represent an ideal resource for the application of supervised machine learning algorithms for compound classification as a means for an automated annotation of MSTs. The GMD compendium may thus be used to enhance the chemical identification process of novel metabolic components discovered by GC-(TOF)-MS based metabolomic screening studies.

Currently, most novel GC-MS based analytical signals remain unidentified as there is no reference substance available. Of the total of up to 1,000 MSTs observed in typical studies, only 50–150 metabolites can be identified. The determination of the chemical sum formula associated with molecule peak and electron-impact induced fragment peaks may be feasible. The unambiguous mass spectral interpretation, however, is in many cases only possible for small molecules bearing a single functional group (Varmuza and Werther 1996). For this reason, hit list based mass spectra similarity matching has evolved as a highly successful tool for the routine assessment of mass spectra (Halket et al. 2005), and large commercial mass spectral libraries, such as NIST08 ([http://chemdata.nist.gov/mass-spec/Srch\\_v1.7/index.html](http://chemdata.nist.gov/mass-spec/Srch_v1.7/index.html)) have been developed. The employed matching approaches use various similarity scores, which were developed in conjunction with specifics of the gas chromatographic and mass spectrometric technology (Crawford and Morrison 1968, Stein 1999, Stein and Scott 1994). However, a reliable automation of mass spectral matching has not been accomplished so far. Tools utilising RI information for the matching process adapted for the processing of large metabolite profiling experiments, such as TagFinder (Luedemann et al. 2008), recommend manually supervised compound identification.

Beyond similarity-based matching of mass fragmentation patterns, algorithmic approaches that relate mass spectral features to structural properties have been pursued. For example, several low resolution mass spectral classifiers have been reported to yield promising results. Werther et al. (1994) successfully applied diverse computational

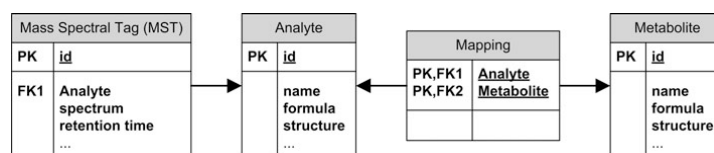
classification techniques to the recognition of simple structural moieties. These authors tested the prediction of 10 structural characteristics and found neural networks to be superior compared to k-nearest neighbour (KNN) classification, linear discriminant analysis, or principal component models. Furthermore, Varmuza and Werther (1996) presented an enhanced approach based on random sampling of training mass spectra according to predefined spectral features and the subsequent application of multivariate classification methods or neural networks. Approximately 160 spectral classifiers were developed that cover a significant portion of organic chemistry, but they may only be partially applicable to bio-molecules, such as the primary and secondary metabolites. These spectral classifiers are now part of the NIST software. Since version 2.0, the software package Mass Frontier™ supports the classification of mass spectra (<http://www.highchem.com/new-features/>), utilizing three classification methods: Principal Component Analysis (PCA), Fuzzy Clustering, and Self-Organizing Maps (SOM) (<http://www.highchem.com/mass-spectra-classification/>) (Steiner et al. 2002). Further enhanced approaches attempt to include isomer prediction from given sum formulas (Varmuza 2001).

Progress has also been made for the prediction of the presence or absence of substructures based on mass spectral features. Mass spectral classifiers were implemented using linear discriminant analysis, LDA, and partial least squares discriminant analysis, PLS-DA (Yoshida et al. 2001), or by selection of characteristic mass fragment combinations (Tang et al. 2003). In view of these successful approaches, we developed a decision tree (DT) based classifier and a web-based interface dedicated to the specific metabolomic needs of GC-MS-based profiling. In contrast to prior general organic chemistry-based efforts, we focus on metabolites and substructures of metabolic origin. Furthermore, we combine the chromatographic RI information (Strehmel et al. 2008) with mass spectral features for classification and substructure prediction. Most importantly, we chose the DT algorithm to solve the classification problem. This algorithm is employed for the recognition of patterns in mass fragmentation spectra that distinguish classes of compounds which either contain or lack a specific predefined chemical moiety. The DTs are made available via the augmented web-interface of the GMD as well as web-services to assist in the annotation of metabolomics data sets.

## 2 Methods

### 2.1 Integration of metabolite structures into the GMD

The GMD uses a Microsoft SQL Server 2008™ as the relational database backend for relating the mass spectrum



**Fig. 1** Excerpt of the GMD scheme. MSTs (mass spectral tags, i.e. repeatedly observed mass spectra with retention behaviour) are linked to analytes via experiments and a supervised annotation process.

Likewise, analytes are mapped to metabolites. Structural information has been added to both types of compounds, the metabolites and their respective analytes

and retention behaviour to an analyte, i.e. the chemically modified compound, which is mapped to represent a metabolite (Fig. 1) (Hummel et al. 2008). Both analyte and metabolite have the properties of a chemical compound and are linked to structures archived as *.mol*-files and InChi<sup>TM</sup> codes (<http://www.iupac.org/inchi/>). A typical metabolite has one to two analytes, which are generated by the chemical derivatization process inherent to the GC-MS profiling technique. Each analyte has multiple technological versions of MSTs. These replicate mass spectra and RIs are empirically determined using different mass spectral technologies, e.g. time of flight, quadrupole or ion trap based mass detectors, and variations of gas chromatographic systems (Strehmel et al. 2008).

In the current GMD release, 6,187 mass spectra are available representing 2,444 analytes and 1,535 metabolites. It should be noted that the GMD compendium is biased towards GC-MS accessible, stable, primary metabolites. Therefore, the structural moieties of the metabolite classes, amino acids, organic acids, fatty acids, fatty alcohols, sugars, sugar alcohols and respective conjugates dominate. Structural annotations are in most cases stereochemically correct, even though routine GC-MS profiling (Lisec et al. 2006, Wagner et al. 2003) allows only the differentiation of anomeric, epimeric structures and E/Z-geometric isomers.

## 2.2 Decision tree supported substructure prediction

A supervised machine learning approach using a pattern recognition algorithm was chosen to infer correlations between the sub-structure properties of known compounds and the properties of respective MSTs. For every considered functional group, we classify MSTs to belong to either the functional group containing or non-containing groups. Thus, we perform a binary classification. The DT method was applied, because multiple parameter types, categorical and numerical, can be integrated and no assumptions about numerical parameter distributions and about the nature of discriminating functions, e.g. linear, non-linear, multimodal, are required. Secondly, in contrast to the NIST Mass Spectrum Interpreter software for substructure

analysis, rules comprised of single feature decisions are returned, which describe the criteria of the mass spectral classification process and are suitable for interpretation by a GC-MS expert.

The *checkmol* program (Feldman et al. 2005) was executed to automatically extract the 21 most abundant structural features, e.g. substructures or chemical moieties, from the metabolite structures of the GMD (>3% occurrence). Subsequently, we created DTs using mass spectral and retention properties as predictors of the structural features.

The mass spectra used to train the DT algorithm were electron impact spectra of methoxyaminated and trimethylsilylated reference compounds with known structures, natural isotopomer composition, and documented reference compound sources.

It should be noted that—in the current release—we use the MSTs of chemically derivatized metabolites for the analysis of structural features present in non-derivatized metabolite structures. As the DT algorithm supports this approach, we reasoned that the biologist and GMD user's interest lies more on the metabolite structure rather than the methoxyaminated and trimethylsilylated compounds inherently required for GC-MS based metabolite profiling.

### 2.2.1 MST subsetting and RI assignment

DT training was performed separately for each considered structural feature. For this procedure, the mass spectral compendium of the GMD was divided into those mass spectra associated with metabolites containing the respective structural feature and those in which the structural feature was absent. DT training was performed with and without using the RI information linked to each MST. In order to use the RI information, a subset of training data with empirically determined RIs was created. The supported RI models are based on standardization by 9 *n*-alkanes (C10–C36) and either a 5%-phenyl-95%-dimethylpolysiloxane capillary column, in short VAR5, or a 35%-phenyl-65%-dimethylpolysiloxane column (MDN35, Lisec et al. 2006). RI information of 8 variant VAR5 chromatography methods was converted according to Strehmel et al. (2008).

### 2.2.2 Mass spectral features

A mass spectrum can be considered a point in an  $n$ -dimensional mass space with  $n$  representing the number of individual masses/charge fragment ratios as coordinate axes with associated values corresponding to fragment abundances. In order to characterise mass spectra in terms of those properties, which are potentially relevant for the structural distinction of chemical substance classes, additional spectral features have been proposed, e.g. weighted abundance of single masses, intensities of single masses normalised to the local ion current, averaged intensities of mass intervals, logarithmic transformations, modulo-14 summation, autocorrelation properties, so-called spectrum type features, and characteristic peak series features (Varmuza 2001, Xu et al. 2003).

As DT methods allow the combined use of diverse properties, we extracted three types of spectral features in addition to the above mentioned RI information. (1) Logarithmic intensities of nominal masses, as proposed previously (cf. above), were used in the mass range  $m/z$  70–600. However, only 525 spectral features (“*intensity-ig*”) were allowed after exclusion of ubiquitous mass fragments typically generated from compounds carrying a trimethylsilyl-moiety, namely the fragments at  $m/z$  73, 74, 75, 147, 148, and 149. (2) For an improved feature construction with better discriminative potential (Kotsiantis et al. 2006), we implemented the full set of logarithmic pair-wise intensity ratios, thus, adding 137,550 spectral features to the DT training input space (“*ratio lg*”). (3) In addition to these features, we used a mass distance measure to represent the mass losses of typical electron impact induced fragmentation reactions. Mass distances caused by the naturally occurring  $^{13}\text{C}$ -isotopic patterns were excluded. This spectral processing provided 524 additional spectral features (“*peak distance*”) per MST. In detail, only those mass fragments associated with a local intensity maximum were considered for peak distance calculation, whereas flanking mass fragments with descending intensities at  $(m/z) - 1(2, 3, \dots)$  or  $(m/z) + 1(2, 3, \dots)$  were removed before calculating the peak distance matrix.

In summary, MSTs were pre-processed to obtain those spectral features best representing the probability that a specific fragment is generated from a given compound (*intensity-ig*, *ratio lg*) and the mass differences between fragments indicative of the typical cleavage reactions of chemical moieties. Both types of information were used to train DTs with or without the use of RI information.

### 2.2.3 Decision tree generation

Using the Microsoft SQL 2008 Server Analysis Services™, DTs were trained for selected single structural

features. Because an SQL Server table is limited to 1,024 columns, the predictor variables had to be pre-selected. We used the Fisher ratio,  $F_r$ , for ranking the variables, with

$$F_r = \frac{(m_1 - m_2)^2}{v_1 + v_2} \quad (1)$$

with  $m_1$  and  $m_2$  equal to the means, and  $v_1$  and  $v_2$  representing the variances of classes 1 and 2 (functional group present/absent), respectively (Duda and Hart 1973; Varmuza and Werther 1996).

In total, 138,599 Fisher scores were computed for each functional group for the evaluation of the respective discriminative power of all available mass spectral features. The 1,000 highest scoring spectral features were chosen for each prediction task. When multiple feature types were used for DT training, features of each type were selected in equal proportions.

The final training set submitted to the DT algorithm comprised 1,004 columns with 1,000 columns of the pre-selected best scoring mass spectral features. Two columns containing the optional RI-related information from the VAR5- and MDN35-RI systems were added, while one column contained the present or absent call of the structural feature under investigation. The fourth column comprised the primary key reference to the respective mass spectrum entry within the GMD. Three DT training procedures were performed, generating a DT without RI information, and two DTs with RI information of either the VAR5 or the MDN35 chromatographic systems. For DTs with RI information, only those MSTs with available RI information were considered.

The DT models including selected features, transformations and other pre-processing details were saved to the server for subsequent application to user submitted MSTs of unknown structure. In the current build, the DT algorithm of the Microsoft SQL Server Analysis Services™ was parameterized according to default recommendations (cf. Table 1). The minimum node support was set to 10 spectra (tree expansion break off criterion).

### 2.2.4 Performance measures

The prediction performance was assessed by the precision ( $p$ ) and recall ( $r$ ) measures, with

$$p = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

and

$$r = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

TP, FP, TN, and FN define true positive, false positive, true negative and false negative predictions, respectively.

**Table 1** Microsoft SQL Server Analysis Services™ DT Algorithm default recommendations (cf. <http://msdn.microsoft.com/en-us/library/cc645868.aspx>)

Property	Value	Description
Hold out Max Percent	0	All available data are used as training data
Hold out Max Cases	0	All available data are used as training data
Split Method	Both (binary and complete)	Determines the method that is used to split the node
Score Method	Bayesian dirichlet equivalent to uniform prior	Specifies the method to calculate the split score
Minimum Support	10	Break off criterion of tree extension. All nodes of the resulting tree have at least 10 spectra
Maximum Input Attributes	255	Defines the number of input attributes that the algorithm can handle before it invokes feature selection
Maximum Output Attributes	255	Defines the number of output attributes that the algorithm can handle before it invokes feature selection
Complexity Penalty	Not specified	This parameter defines the likelihood of a split as a function of the number of attributes. The default is 0.5 (1–9 attributes) decreasing the likelihood, 0.9 (10–99 attributes) and 0.99 (>100 attributes) causing an increased likelihood

For a combined characterization of precision and recall, the  $F_{pr}$ -value with

$$F_{pr} = 2 * \frac{p * r}{p + r} \quad (4)$$

was computed as a frequently used performance measure in the field of information retrieval (van Rijsbergen 1979).  $F_{pr}$ -values of 1 indicate optimal, while values approaching zero correspond to minimal prediction performance.

Matthews correlation coefficient (MCC) is commonly used for the assessment of binary classifications and was shown to be robust with regard to imbalanced class distributions (Matthews 1975). MCC can be computed from the contingency table according to Eq. 5:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

MCC values range from  $-1$  (perfect inverse prediction) to  $+1$  (perfect prediction). A coefficient of 0 represents an average random prediction. The error rate obtained in cross validation (CV),  $Er_{CV}$ , is computed as

$$Er_{CV} = \frac{FP + FN}{TP + FP + TN + FN} \quad (6)$$

and should approach zero with increasing DT quality.

### 3 Results and discussion

We developed DT-based substructure prediction as a potentially powerful tool box for the structural characterisation of the numerous non-identified MSTs that are encountered in routine GC-MS based metabolite profiles. To enable evidence based substructure prediction we utilized the rich resource of mass spectra and RI information of authenticated reference compounds from the GMD. Application of supervised machine learning approaches required updating of the GMD with structure information of the contained metabolites. This added information now allows binary partitioning of the known metabolites into training data that either contain or do not contain the assessed substructure. The application of supervised machine learning algorithms, such as the DT classification, now supports the in silico characterisation of yet non-identified MSTs that are frequently recognised as relevant marker molecules by non-targeted metabolite profiling. In order to define and compare the performance of the chosen DT-classification approaches for the potential users of the GMD web site and the offered web services (cf. Sect. 3.2) we report in the following the implemented CV procedures of the provided DTs, the respective feature usage and finally assess typical application cases.

### 3.1 Evaluation of the decision tree performance

For the characterization and comparison of DT performance, we implemented measures based on the subtotal TP, FN, FP, and TN prediction of the CV contingency table. These measures allow the assessment of alternative DTs for identical substructures, e.g. DTs with or without use of RI information. Also the quality of DTs was made comparable between different substructure predictions as two general classification errors exist, (1) deficiencies resulting from imperfect MST training data, and (2) deficiencies due to imbalanced training data. Both potential errors lead to over-fitted DTs and may compromise substructure predictions.

#### 3.1.1 Cross validation of decision trees

For the reasons stated above a 50-fold CV by iterative exclusion of randomly chosen MSTs was routinely implemented. As an alternative we explored a heuristic validation process by excluding all technological replicate MSTs of single analytes (data not shown). As GMD will provide a steadily increasing number of replicate mass spectra and RIs of each analyte that is expected to improve DT-based substructure prediction we decided to use random choice of MSTs for CV. Our implemented procedure characterises DT performance and enables the calculation of overall error estimates for each substructure tree. CV results are displayed along with the respective DT information and visualization (Fig. 2). In order to demonstrate the implementation of the chosen CV process, exemplary results of a ten-fold CV are shown for the DT-classification which assesses the presence or absence of an amine moiety (Table 2).

Using a Precision-Recall-Plot (Fig. 3), we evaluated precision and recall in relation to an  $F_{pr}$ -measure threshold = 0.65 using the most recent set of DTs which includes the RI information based on standardization by 9 *n*-alkanes (C10–C36) and a 5%-phenyl-95%-dimethylpolysiloxane capillary column, in short VAR5 (Strehmel et al. 2008).

Precision and recall measures were high for almost all tested substructures. Only the three DTs targeting the prediction of *heterocycle*, *hemiacetal*, and *carboxylic acid ester* substructures resulted in slightly inferior DT-classifications. These substructures are, hence, excluded from for public use and will be targeted by future efforts to enhance the DT applied algorithm and by extending the GMD compendium by suitable training data. Since the hard to classify substructures were also those with the smallest number of available training data, we expect that extended data resources within GMD will immediately result in improved performance of substructure prediction.

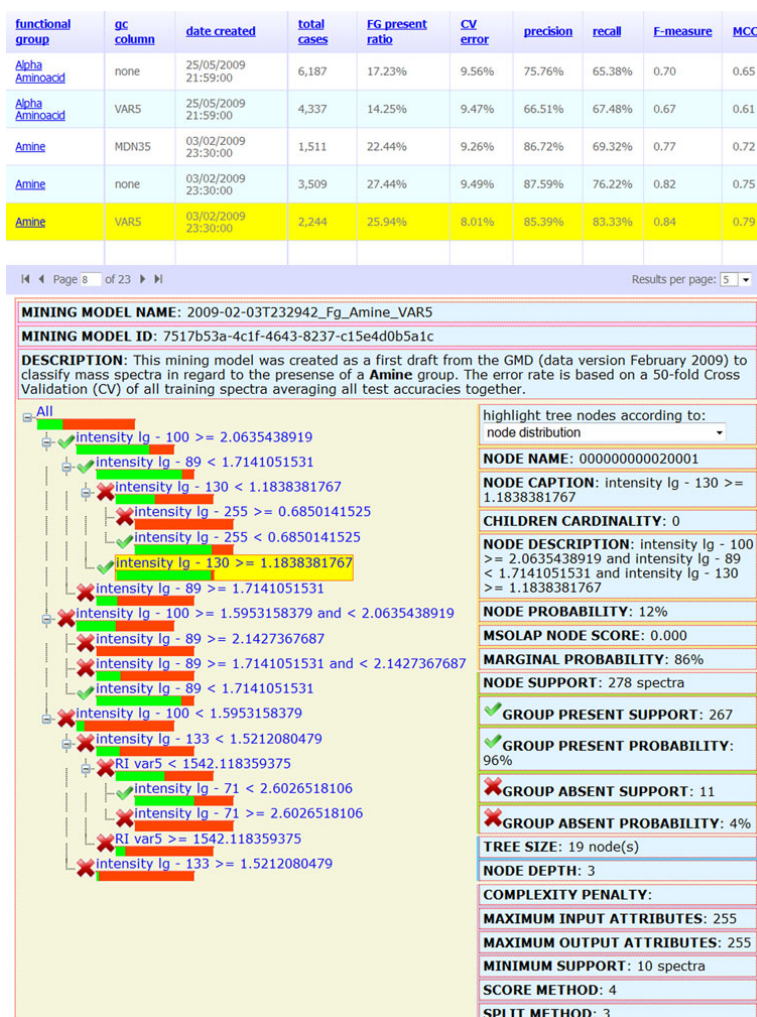
#### 3.1.2 Analysis of MST feature usage

Table 3 summarizes the MST feature usage within the current set of DTs compiled from the May 2009 GMD version including the *Var5* RI information as input variables. All DTs were generated independently, i.e. both data processing and feature pre-selection was performed separately for each substructure prior to DT training. As expected, the features incorporated in the DT vary considerably between the different predicted substructures. The usage of characteristic mass spectral fragments ( $m/z$ ) agrees with the chemical nature and hierarchy of the investigated biochemical moieties. The repeated use of characteristic mass fragments for similar substructures is apparent. For example, the fragment  $m/z = 99$  is consistently used for all amine-like substructures,  $m/z = 103$  in the case of alcohol-like substructures and  $m/z = 160$  for carbonyl-like substructures. These mass fragments may be termed canonical masses which result from fragmentation reactions that are typical of compounds belonging to certain chemical classes. For example, the mass fragment  $m/z = 160$  represents methoxyaminated aldehyde moieties, which are characteristic for reducing aldose-sugars. The fragment  $m/z = 103$  is an abundant and typical cleavage product of trimethylsilylated primary alcohols, such as non-cyclic sugar and polyol molecules. While  $m/z = 103$  represents the cation  $[\text{CH}_2\text{O}(\text{TMS})^+]$  and  $m/z = 160$  the cation  $[\text{C}=\text{NOCH}_3-\text{CH}_2\text{O}(\text{TMS})^+]$ , the source and usage of  $m/z = 99$  for the prediction of amines is not yet fully understood.

Considering the available choice of the numerous preselected 1,001 variables (a maximum of 1,000 spectral features plus one optional RI information), the DT classification uses only a comparatively small and specific number (<14) of selected features per DT (cf. summary row\*<sup>1</sup>) in Table 3). This small number and the frequent choice of specific features represent an additional safeguard against the risk of DT over-fitting. The chemical analysis and interpretation of feature usage and the analysis of the surprising absence of “*ratio-Ig*” criteria from the current DTs is in progress but was deemed beyond the scope of this study.

#### 3.1.3 Application cases

In order to characterize the potential, but also the caveats of substructure predictions using the DT algorithm provided by GMD, we performed typical application cases. New or non-identified metabolites will—in most cases—be discovered as automatically deconvoluted mass spectra from profiles of complex biological samples. Automated mass spectral deconvolution represents the typical solution to the task of analysing GC-MS profiles of highly complex



**Fig. 2** Two synchronized web controls facilitate the access to all currently available DTs and sub-structure predictions provided by the GMD interface. The *upper table* describes the investigated functional group, with optional use of the RI system, and records the date of DT generation, as well as the number and proportion of MSTs linked to present calls of the substructure. DTs are characterised by a 50-fold CV error, with corresponding precision, recall, *F*-measure and MCC values (cf. Sect. 2.2.4). The table can be sorted by activating the selected column header. The full list of all available DTs is accessible through a paging function at the left bottom of the display. An exemplary DT trained to classify mass spectra with regard to the

presence or absence of the amine substructure using the VAR5 RI information and 582 mass spectra representing metabolites with an amine moiety contrasted by 1,662 MSTs of metabolites lacking this moiety. The DT is depicted on the *left hand side*. Standard SQL Server Analysis Services Decision Tree properties of the activated tree node (indicated by *underlay*) are displayed to the right. *Light (green) bars* indicate the proportion of MSTs linked to a compound with amine moiety; *dark (red)* indicate the proportion of MSTs linked to compounds lacking this moiety. Note that we use the terms *Decision Tree* and *Mining Model* to indicate the data structures according to the SQL Server Analysis Services terminology

biological matrices; here we use the term matrix to refer to the sum of all monitored metabolites from a biological sample. Under such conditions the available deconvolution

algorithms may remove specific mass fragments either because of low compound abundance, resulting in a low signal-to-noise ratio (S/N), or due to chromatographic peak

**Table 2** Contingency table of a 10-fold CV, evaluating both performance and robustness of the DT model for the prediction of a amine sub-group generated using the February 3, 2009 version of GMD and the RI information of the VAR5 chromatographic system, in short the DT model named '2009-02-03T232942\_Fg\_Amine\_VAR5'

Observed	Partition index	Predicted				Total	
		Present		Absent			
Present	1	True positives (TP)	47	(21%)	False negatives (FN)	11	(5%)
	2		44	(20%)		14	(6%)
	3		50	(22%)		9	(4%)
	4		38	(17%)		21	(9%)
	5		39	(17%)		19	(8%)
	6		45	(20%)		13	(6%)
	7		40	(18%)		18	(8%)
	8		43	(19%)		15	(7%)
	9		52	(23%)		6	(3%)
	10		43	(19%)		15	(7%)
		Subtotal		441		141	582
	AVG		44.10	(20%)	14.10	(6%)	
	STDEV		4.58	(2%)	4.61	(2%)	
Absent	1	False positives (FP)	8	(4%)	True negatives (TN)	158	(71%)
	2		10	(4%)		157	(70%)
	3		7	(3%)		160	(71%)
	4		1	(0%)		165	(73%)
	5		2	(1%)		164	(73%)
	6		4	(2%)		162	(72%)
	7		5	(2%)		161	(72%)
	8		1	(0%)		165	(74%)
	9		10	(4%)		156	(70%)
	10		7	(3%)		159	(71%)
		Subtotal		55		1,607	1,662
	AVG		5.50	(2%)	160.70	(72%)	
	STDEV		3.44	(2%)	3.27	(1%)	
Total			496		1,748	2,244	

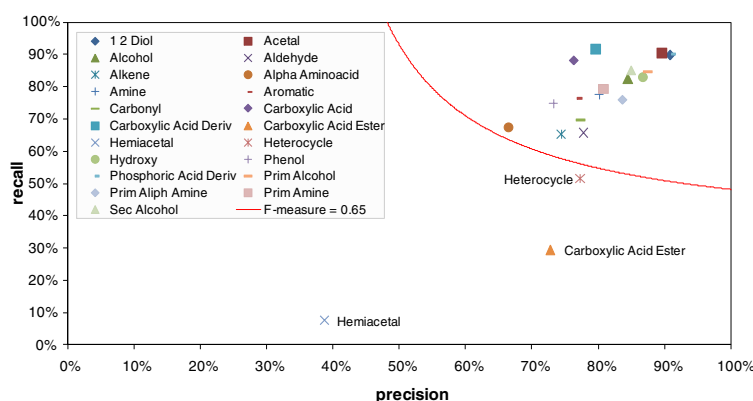
Rows 3–12 and 16–25 represent the 10 iterative partitions. Absolute numbers (column D, G) and relative proportions (%; column E, H) of TP, FN, FP and TN classifications are reported. Results of each partition add up to 100%. Subtotals, averages (AVG) and standard deviations (STDEV) are calculated over all sequential partitions. Note that high proportions of TP and TN and low STDEV (<2%) indicate high DT prediction performance

shape artefacts. Alternatively, mass fragments belonging to chromatographically co-eluting compound(s) may be incorrectly added.

Because of the ease of automated deconvolution, the experimental scientist might be tempted to base an initial structural elucidation attempt on such potentially compromised MSTs. To demonstrate the risk of such an approach, we selected four compounds derived from two complex plant matrices, namely potato tuber and rice leaf. The compounds glucose (1MeOX 5TMS), citric acid (4TMS), valine (2TMS) and putrescine (4TMS) were chosen to represent carbohydrates, organic acids, amino acids, and amines as typical metabolite classes. These metabolites

contain most of the frequently occurring metabolite substructures, which have been targeted by our DT classification approach. Figure 4 shows the metabolite structures before chemical derivatization by methoximation (carbonyl modifying) and trimethylsilylation (substituting protons bound to heteroatoms).

Our application cases convincingly demonstrate that many expected substructures are recognized with high reproducibility (Table 4). Nevertheless, clear differences with regard to the biological source or concerning single automated deconvolutions from identical sources exist. For example, the primary alcohol substructure is only recognized in part of the deconvolutions representing glucose and



**Fig. 3** Precision-Recall-plot for a set of DTs including the *Var5* RI information of the May 2009 version of the GMD. We introduced an  $F_{pr}$ -measure threshold of 0.65 indicated by a line as a minimal performance criterion. DTs not passing this threshold, namely the

DTs for the prediction of the *carboxylic acid ester*, *hemiacetal* and *heterocycle* substructures were disabled for public use. The nomenclature of the chemical substructures is according to the definitions made by the *checkmol* program (Feldman et al. 2005)

fructose, whereas the more generalized substructures containing OH-moieties are diagnosed with high repeatability. In addition, the alpha-amino acid and the more general amine-substructure were recognized with varying reproducibility, in some cases only in a small part of the automated deconvolutions.

In conclusion, automated mass spectral deconvolutions should only be considered with care and for a preliminary substructure assessment. We recommend the use of manually curated mass spectra and a statistical evaluation of multiple high quality mass spectra best obtained from multiple biological sources or from at least two different chromatographic systems. Specialized laboratories may avoid many GC co-elution artefacts by application of a two dimensional (GCxGC)-TOF-MS system.

### 3.2 Availability, GMD web site and web services

All DTs developed as part of this study have been made freely available to academic users for spectra-based compound annotation at <http://gmd.mpimp-golm.mpg.de/>. For automated batch processing, the developed platform-independent Simple Object Access Protocol (SOAP) based web service endpoint *wsPrediction* provides public access to the functionality presented here.

## 4 Concluding remarks

The mass spectrum and RI compendium of the GMD has been used as a training data set for a supervised machine learning approach using a DT algorithm for the

classification of MSTs and the retrieval of human-interpretable classification rules. The new GMD frontend provides a rich set of substructure classification models comprising mass spectral patterns and optional extensions including RI information, which group MSTs with common substructures. The offered DTs are provided as an extension to the conventional hit list based mass spectral matching approach and can be used to support the interpretation of MSTs from known metabolites and also facilitate the classification of those MSTs, which are not yet identified. The classification tools of the GMD frontend can be updated with the continuously growing set of GMD entries. The success of updating efforts can be assessed using DT cross validation (CV) parameters, such as precision, recall,  $F_{pr}$ -measure, MCC, and the CV error, which have been implemented in the course of this project to compare DT performance and to reject weak prediction models. Hence, this new web interface and application may contribute to the evidence-based classification of non-identified MSTs and follows the general recommendations of the metabolomics standards initiative for reporting standards for chemical analysis (Sumner et al. 2007).

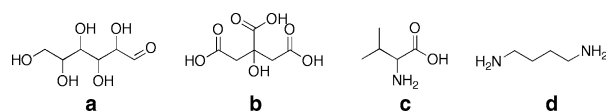
The DTs presented in this work depend on the continued curation and enhancement of the GMD content. Specifically, residual deconvolution errors will be removed, the spectral quality improved and the number of high quality replicate spectra for existing MSTs extended. Most importantly, new metabolites will be added to the GMD compendium. As a consequence, these efforts will necessitate an updating scheme for the DT substructure predictions and evaluation of DT performance will become a frequent use case.



**Table 3** Summary of MST feature usage over a set of DTs predicting several chemically similar substructures

spectral feature m/z	alcohol like					amine like				carbonyl like				diverse					Count feat. usage			
	Alcohol	Prim Alcohol	Sec Alcohol	1,2 Diol	Hydroxy	Amine	Prim Amine	Prim Aliph Amine	Alpha Aminoacid	Carbonyl	Aldehyde	Carboxylic Acid	Carboxylic Acid Deriv	Carboxylic Acid Ester	Hemiacetal	Heterocycle	Alkene	Aromatic		Phenol	Phosphoric Acid Deriv	Acetal
70	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
71	x	-	-	-	-	-	x	x	-	-	-	x	-	-	-	-	-	-	-	-	-	4
72	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	2
76	-	x	-	-	-	-	-	-	-	-	-	x	x	-	-	-	x	-	-	-	-	4
77	x	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	1
78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	1
82	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	1
83	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	x	-	-	-	2
85	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
86	-	-	-	-	x	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	2
87	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	1
88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
89	-	-	-	-	-	-	x	x	x	x	x	x	x	x	-	x	-	-	-	-	-	8
92	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	2
93	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
94	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	1
95	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	2
97	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	1
98	x	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	4
99	-	-	-	-	-	-	x	x	x	x	-	-	-	-	-	-	-	-	-	-	-	4
100	x	-	-	-	x	-	x	x	x	x	-	-	-	-	-	-	x	-	-	-	-	7
101	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	1
102	-	x	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	2
103	x	x	x	x	x	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	7
104	x	-	-	-	x	-	-	-	-	-	-	x	x	x	-	-	-	-	-	-	-	3
105	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
109	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	1
112	x	-	x	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
113	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	1
115	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	1
117	-	-	x	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
119	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-	-	2
123	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	-	-	2
124	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
125	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	1
126	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
127	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	x	-	-	1
128	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	-	-	-	-	2
129	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	x	x	-	4
130	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-	-	-	-	-	2
132	-	-	-	-	-	-	-	-	-	-	x	-	x	-	-	-	-	-	-	-	-	2
133	-	-	-	-	-	-	x	-	x	-	-	-	x	-	-	-	-	-	-	-	-	5
143	x	x	x	-	-	-	x	x	x	-	-	-	-	-	x	-	-	-	-	-	-	4
145	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	1
146	-	-	-	-	x	-	x	x	x	-	-	x	-	-	-	-	-	-	-	-	-	5
150	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	1
152	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
155	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	1
156	-	-	-	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	1
160	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4
*1	6	6	4	4	6	10	8	10	7	3	4	9	6	6	2	6	7	8	3	0	0	
*2	35	26	29	24	35	30	29	27	24	19	23	41	38	24	6	27	21	42	20	7	10	
RI VAR5	2	-	2	2	-	-	-	2	3	2	-	4	2	-	2	2	4	2	4	-	-	
m.-diff	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	
ratio lg	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

“x” marks the use of a feature in a DT. The intensity lg features are limited in this example to  $m/z < 161$ . The full mass range is used however with a preference of small  $m/z$ . Row\*<sup>1</sup> lists the number of distinct intensity lg features in each DT. Row\*<sup>2</sup> shows the number of DT nodes using intensity lg allowing for repetitive occurrence of a feature in single DTs. The columns are sorted according to similarity of substructure classes. The final column lists the frequency of the “intensity lg” feature usage across all considered DTs and may indicate potential canonical mass fragments. DTs for the sub-groups *Phosphoric Acid Deriv.* and *Acetal* use only intensity lg features greater than  $m/z = 160$ . Thus, corresponding values in row \*<sup>1</sup> are “0”



**Fig. 4** Exemplary metabolites and typical substructures retrieved by the *checkmol* program (Feldman et al. 2005). **a** glucose: primary alcohol, secondary alcohol, 1,2 diol, aldehyde, hydroxyl, carbonyl, alcohol; **b** citric acid: carboxylic acid, hydroxyl, alcohol; **c** valine: alpha-amino acid, amine, carboxylic acid, hydroxyl, carbonyl, alcohol; **d** putrescine: primary aliphatic amine, amine

**Table 4** Reproducibility of substructure predictions using automatically deconvoluted MSTs from routine metabolite profiles of complex biological matrices (selected examples)

Metabolite	Substructure	Potato tuber										Rice leaf										
		Deconvolution										Deconvolution										
		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	
Glucose	Hydroxy	100	100	100	100	100	100	100	100	100	99	100	100	100	100	100	100	100	100	100	100	
	Alcohol	99	99	99	99	99	99	99	99	99	–	99	99	99	99	99	99	99	99	99	99	
	Prim. alcohol	–	99	99	99	99	99	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
	Sec. alcohol	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	
	1,2 Diol	99	99	99	99	99	99	99	99	–	99	99	99	99	99	99	99	79	79	79	99	79
	Carbonyl	–	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	–	–	–	–	–
Citric acid	Hydroxy	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	
	Alcohol	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	
	Sec. alcohol	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	99	
	Carbonyl	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	
Valine	Carboxylic acid	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
	Alpha aminoacid	97	97	97	–	–	–	–	–	97	–	–	–	97	–	–	–	–	–	–	–	–
	Amine	–	–	–	–	–	–	–	–	91	–	–	–	–	–	–	–	–	–	–	–	–
	Hydroxy	–	–	–	–	–	–	–	–	83	–	83	83	83	83	83	100	100	100	100	100	100
	Carbonyl	–	–	–	–	–	–	–	–	91	–	–	–	–	–	–	91	91	91	91	91	91
	Carboxylic acid	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	90	90	90	90	90	90
Putrescine	Carboxylic acid deriv.	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	90	90	90	90	90	90
	Prim. aliph. amine	86	86	86	94	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86
	Prim. amine	–	–	94	94	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	Amine	–	–	97	91	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Automated deconvolutions of each metabolite were generated from 10 independently processed GC-TOF-MS profiles of potato tubers and rice leaves. The listed values represent confidence (probability) of identification within the used DT nodes (“–” indicates non-recognized expected substructures)

Furthermore, the extension towards DT analyses of those substructures, which are underrepresented in the current GMD dataset, appears to be an attractive goal. Finally, given the availability of multiple DTs for the prediction of one particular substructure, the application of DT forests may be worthwhile.

**Acknowledgements** The authors acknowledge the long standing support and encouragement by Prof. L. Willmitzer, Prof. M. Stitt and Prof. R. Bock, Max Planck Institute of Molecular Plant Physiology (MPI-MP), Am Muehlenberg 1, D-14476 Potsdam-Golm, Germany. The authors thank Dr. D. Steinhauser, Dr. A. R. Fernie, A. Erban, I. Fehrle, J. Hannemann and M. Kuczmierczyk for the generation of metabolite structures and the interactive discussions during project realization.

**Funding** 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: Systematic analysis of cellular processes and their relation to plant internal oxygen concentrations”, FKZ 0313813A, and the European META-PHOR project, FOOD-CT-2006-036220.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

Crawford, L. R., & Morrison, J. D. (1968). Computer methods in analytical mass spectrometry. Identification of an unknown compound in a catalog. *Analytical Chemistry*, 40, 1464–1469.

- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: Wiley.
- Feldman, H. J., Dumontier, M., Ling, S., Haider, N., & Hogue, C. W. V. (2005). CO: A chemical ontology for identification of functional groups and semantic comparison of small molecules. *FEBS Letters*, 579, 4685–4691.
- Halket, J. M., Waterman, D., Przyborowska, A. M., Patel, R. K. P., Fraser, P. D., & Bramley, P. M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of Experimental Botany*, 56, 219–243.
- Hummel, J., Selbig, J., Walthert, D., & Kopka, J. (2008). The Golm Metabolome Database: A database for GC-MS based metabolite profiling. In J. Nielsen & M. Jewett (Eds.), *Metabolomics a powerful tool in systems biology. Topics in current genetics Vol. 18* (pp. 75–96). Berlin, Heidelberg, New York: Springer.
- Kopka, J. (2006). Current challenges and developments in GC-MS based metabolite profiling technology. *Journal of Biotechnology*, 124, 312–322.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al. (2005). GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics*, 21, 1635–1638.
- Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data preprocessing for supervised learning. *International Journal of Computer Science*, 1, 111–117.
- Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., & Fernie, A. R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Protocols*, 1, 387–396.
- Luedemann, A., Strassburg, K., Erban, A., & Kopka, J. (2008). TagFinder for the quantitative analysis of gas chromatography-mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics*, 24, 732–737.
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica Biophysica Acta*, 405, 442–451.
- Schauer, N., Steinhäuser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., et al. (2005). GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Letters*, 579, 1332–1337.
- Stein, S. E. (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, 10, 770–781.
- Stein, S. E., & Scott, D. R. (1994). Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry*, 5, 859–866.
- Steiner, F. M., Schlick-Steiner, B. C., Nikiforov, A., Kalb, R., & Mistrik, R. (2002). Cuticular hydrocarbons of Tetramorium ants from central Europe: Analysis of GC-MS data with self-organizing maps (SOM) and implications for systematics. *Journal of Chemical Ecology*, 28, 2569–2584.
- Strehmel, N., Hummel, J., Erban, A., Strassburg, K., & Kopka, J. (2008). Retention index thresholds for compound matching in GC-MS metabolite profiling. *Journal of Chromatography B*, 871, 182–190.
- Sumner, L., Amberg, A., Barrett, D., Beale, M., Beger, R., Daykin, C., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3, 211–221.
- Tang, Y., Liang, Y., & Fang, K. T. (2003). Data mining in chemometrics: Sub-structures learning via peak combinations searching in mass spectra. *Journal of Data Science*, 1, 481–496.
- van Rijsbergen, C. J. (1979). *Information retrieval*. Newton, MA: Butterworth-Heinemann.
- Varmuza, K. (2001). From MS data via chemometrics to chemical structure information. Informatics and mass spectrometry. In *13th Sanibel conference on mass spectrometry*. American Society for Mass Spectrometry, Sanibel Island, FL, USA, pp. 1–11.
- Varmuza, K., & Werther, W. (1996). Mass spectral classifiers for supporting systematic structure elucidation. *Journal of Chemical Information and Computer Sciences*, 36, 323–333.
- Wagner, C., Sefkow, M., & Kopka, J. (2003). Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry*, 62, 887–900.
- Werther, W., Löhninger, H., Stancl, F., & Varmuza, K. (1994). Classification of mass spectra: A comparison of yes/no classification methods for the recognition of simple structural properties. *Chemometrics and Intelligent Laboratory Systems*, 22, 63–76.
- Xu, C. J., He, P., & Liang, Y. Z. (2003). Building an honest tree for mass spectra classification based on prior logarithm normal distribution. *Journal of Data Science*, 1, 497–509.
- Yoshida, H., Leardi, R., Funatsu, K., & Varmuza, K. (2001). Feature selection by genetic algorithms for mass spectral classifiers. *Analytica Chimica Acta*, 446, 483–492.

## Research Paper

**An automated GCxGC-TOF-MS protocol for batch-wise extraction and alignment of mass isotopomer matrixes from differential  $^{13}\text{C}$ -labelling experiments: a case study for photoautotrophic-mixotrophic grown *Chlamydomonas reinhardtii* cells**

Stefan Kempa<sup>1</sup>, Jan Hummel<sup>2</sup>, Thorsten Schwemmer<sup>1</sup>, Matthias Pietzke<sup>1</sup>, Nadine Strehmel<sup>2</sup>, Stefanie Wienkoop<sup>1,3</sup>, Joachim Kopka<sup>2</sup> and Wolfram Weckwerth<sup>1,3</sup>

<sup>1</sup> GoFORSYS, University of Potsdam, Germany

<sup>2</sup> Max-Planck-Institute of Molecular Plant Physiology, Potsdam, Germany

<sup>3</sup> Department Molecular Systems Biology, University of Vienna, Austria

Supporting Information for this article is available from the author on the WWW under [http://www.wiley-vch.de/contents/jc2248/2008/200800337\\_s.pdf](http://www.wiley-vch.de/contents/jc2248/2008/200800337_s.pdf)

Two dimensional gas chromatography coupled to time-of-flight mass spectrometry (GCxGC-TOF-MS) is a promising technique to overcome limits of complex metabolome analysis using one dimensional GC-TOF-MS. Especially at the stage of data export and data mining, however, convenient procedures to cope with the complexity of GCxGC-TOF-MS data are still in development. Here, we present a high sample throughput protocol exploiting first and second retention index for spectral library search and subsequent construction of a high dimensional data matrix useful for statistical analysis. The method was applied to the analysis of  $^{13}\text{C}$ -labelling experiments in the unicellular green alga *Chlamydomonas reinhardtii*. We developed a rapid sampling and extraction procedure for *Chlamydomonas reinhardtii* laboratory strain (CC503), a cell wall deficient mutant. By testing all published quenching protocols we observed dramatic metabolite leakage rates for certain metabolites. To circumvent metabolite leakage, samples were directly quenched and analyzed without separation of the medium. The growth medium was adapted to this rapid sampling protocol to avoid interference with GCxGC-TOF-MS analysis. To analyse batches of samples a new software tool, MetMax, was implemented which extracts the isotopomer matrix from stable isotope labelling experiments together with the first and second retention index (RI1 and RI2). To exploit RI1 and RI2 for metabolite identification we used the Golm metabolome database (GMD [1]) with RI1/RI2-reference spectra and new search algorithms. Using those techniques we analysed the dynamics of  $^{13}\text{CO}_2$  and  $^{13}\text{C}$ -acetate uptake in *Chlamydomonas reinhardtii* cells in two different steady states namely photoautotrophic and mixotrophic growth conditions.

**Keywords:** Metabolomics / Genome sequence / Metabolic labelling / Metabolic flux analysis / Plant systems biology

Received: October 28, 2008; accepted: November 21, 2008

DOI 10.1002/jobm.200800337

## Introduction

*Chlamydomonas reinhardtii* is a green alga studied for decades as a model system in photo-synthesis research. In 2007 the genome sequence was published and *Chlamydomonas reinhardtii* developed into a model organism for plant systems biology called the “green yeast”

**Correspondence:** Wolfram Weckwerth, University of Vienna, Althanstr. 14, 1090 Vienna, Austria  
**E-mail:** wolfram.weckwerth@univie.ac.at  
**Phone:** +43-1-4277-577 00  
**Fax:** +43-1-4277-9 577  
**Mobile:** +43-(0)-664-60277 577 00

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim



www.jbm-journal.com

[2]. *C. reinhardtii* is target for studies on fundamental processes of photosynthesis [3], circadian rhythms [4], starch synthesis [5] and biotechnological studies, as it produces lipids and hydrogen under certain conditions [6, 7].

Recently, we complemented *in silico* genome annotation with metabolomics and proteomics data revealing new putative genes and pathways [8]. This work demonstrated the power of high-throughput techniques like metabolomics and proteomics for the analysis of model organisms and newly sequenced organisms, however, is only the very first step to assess the dynamics of the metabolite and protein compartment in the living cell [9]. Here, further technical method development is needed.

Metabolomic techniques aim to identify and quantify the genome-wide set of metabolites – the metabolome – in an organism, cell-type, tissue or biofluid. It becomes immediately clear that not a single technology is able to cope with the complexity of a total metabolome because of the diversity of chemical structures and their heterogeneous physico-chemical properties [10]. In recent years several key technologies emerged to carry out metabolomic analysis such as NMR-fingerprinting, liquid chromatography coupled to high resolution and high mass accuracy mass spectrometry (LC/MS) and gas chromatography coupled to mass spectrometry (GC/MS) (for an overview see [11]). Here, classical GC/MS techniques developed into one of the standard-operation-procedures (SOP's) for non-targeted rapid metabolomic screening in biological samples [12]. A refinement of metabolomic analysis using a GC/MS technique was achieved with the application of new hyphenated techniques such as the coupling of gas chromatography to time-of-flight mass spectrometry (GC-TOF-MS) [9, 13–15]. However, besides the inherent limitations of the technique – chemical derivatization is necessary to make the metabolites volatile and compatible with gas chromatography, restriction to a lower molecular mass range (up to 600 m/z) – GC-TOF-MS metabolomics is further limited in several aspects such as coelution of individual compounds in the complex sample and structural elucidation of the metabolites. Identification of chemical structures from a complex metabolomic sample is carried out by using a spectral reference library, e.g. the NIST or GMD library [1, 16, 17]. The process of mass spectral library matching and identification of metabolites depends on the quality of the deconvoluted mass spectra from GC-TOF-MS raw-data. The better the gas chromatographic separation and resolution of coeluting compounds in front of the mass spectrometer the better is the generation of indi-

vidual metabolite mass spectra and subsequently the library search. Two-dimensional gas chromatography coupled to TOF-MS (GCxGC-TOF-MS) analysis introduces such a further gas chromatographic separation phase and is therefore an extension of one-dimensional GC-TOF-MS metabolome analysis. GCxGC-TOF-MS overcomes limitations of one dimensional GC-TOF-MS by providing better separations of coeluting peaks, a second retention time/index and enriched and cleaner mass spectra for metabolite identification via library search. However, despite these clear advantages the analysis of multitudes of samples is not yet automated, thus strongly restricted with respect to high throughput analysis of samples in contrast to one dimensional GC-TOF-MS. To circumvent these limitations we present a convenient GCxGC-TOF-MS protocol together with a novel algorithm MetMax for batchwise data processing. The protocol is used to analyse  $^{13}\text{C}$ -labelling pattern in *Chlamydomonas reinhardtii* cells. *C. reinhardtii* is an unicellular photosynthetic organism and it can switch from photoautotrophic to mixotrophic growth where – dependent on the growth medium – acetate can function as the main organic carbon source. Acetate is incorporated via the glyoxylic acid cycle into malate. Under pure photoautotrophic conditions inorganic carbon dioxide is fixed by ribulose 1,5-bisphosphate carboxylase-oxygenase (RUBISCO) via the Calvin-Benson cycle. These processes happen in different compartments of the cell. The Calvin-Benson cycle is located within the chloroplast and the acetate fixation most probably within the mitochondria. But finally both processes provide carbon skeleton and energy for growth and reproduction. In contrast to higher plants *C. reinhardtii* is better accessible for stable isotope labelling experiments and metabolic flux analysis. Metabolic flux analyses are complementary to metabolomics approaches which measure metabolite steady state or transitory levels. With flux analysis it is possible to investigate different modes of carbon fixation in *C. reinhardtii* and to identify putative new pathways [18, 19]. Thus, convenient protocols are urgently needed to carry out metabolic flux analysis in biological systems using stable isotope metabolic labelling and classical metabolic profiling techniques. Here, we propose a convenient protocol for flux analysis using GCxGC-TOF-MS analysis of *C. reinhardtii* after  $^{13}\text{C}$ -stable isotope labelling and sub-sequent data processing using a novel algorithm MetMax.

## Materials and methods

### <sup>13</sup>C-labelling of *Chlamydomonas reinhardtii* cells, sample extraction

#### and GCxGC-TOF-MS analysis

CC503 was cultivated at 21 °C under 100  $\mu\text{E}/\text{m}^2 \cdot \text{s}$  white light (Osram fluora) on an orbital shaker (110 rpm; INFORS HT Multitron) in a 12 h light 12 h dark cycle in 250 ml Erlenmeyer flasks in a medium containing HEPES buffer as described in May *et al.* [8]. Experiments were performed at an optical density of 0.8 OD750. For stable isotope labelling studies 10 mM <sup>13</sup>C sodium acetate (Sigma) or <sup>13</sup>C sodium hydrogen carbonate (Sigma) were added at the indicated time points (see Fig. 3A). For GCxGC-TOF-MS analysis 0.4 ml *Chlamydomonas reinhardtii* culture were analysed as described in May *et al.* [8], except that second column separation conditions were modified. The temperature gradient of the GC oven was 9 °C/min and of the secondary oven 7 °C/min leading to a modified elution time of compounds from the secondary column. Data were analysed using Leco Chromatof<sup>®</sup> software, and the alignment and export of complete isotopomer mass spectra and fragment intensities including M0, M1, M2, etc. was carried out with MetMax (see below).

#### MetMax algorithm

MetMax (metabolite matrix builder) was implemented as a Windows Forms Application using the Net 3.5 framework, the C# programming language, and Microsoft Visual Studio 2008<sup>®</sup>. Different operating modes were implemented, summarized in the following sub paragraphs:

**RI analysis** mode allows the extraction and subsequent alignment of different retention indexes for i) the 1<sup>st</sup> chromatographic dimension: Kováts, van den Dool, polynomial fitting of 3<sup>rd</sup> order, and ii) the 2<sup>nd</sup> chromatographic dimension: van den Dool, relative retention time (RRT) in comparison to the alkanes.

As proposed by Strehmel *et al.* [20] the 'van den Dool' algorithm was used to calculate the retention index based on the alkane retention times for the first dimension. For the second chromatographic dimension, in addition to an adapted van den Dool index, a ratio between measured 2<sup>nd</sup> dimension retention time and predicted 2<sup>nd</sup> dimension retention time was implemented. Here, a Spline function adapted from [21] was used to predict a virtual alkane 2<sup>nd</sup> dimension retention time from the 1<sup>st</sup> dimension retention time.

Both approaches for the characterisation of the 2<sup>nd</sup> retention behaviour were found to perform equal in

terms of the chromatographic resolution. Hereby, the analysis is dependent on the number of expected wrap-arounds per analyte, the column dead time, and column length respectively.

**Isotope concentrator** mode is used to elucidate the isotopic composition of metabolites after isotope labelling experiments. An alignment of intensities, extracted from the mass spectrum, matching isotopic peaks of the metabolites is created by considering a list of expected isotope masses per metabolite.

**Library search** mode uses the 1<sup>st</sup> dimension van den Dool retention index, the 2<sup>nd</sup> dimension alkane based relative retention time (RRT), and a dotproduct based spectra match score, using basepeak normalised spectra, to find potential library hits within the Golm Metabolite Database (GMD) for all mass spectra constituting the chromatograms in question.

Prior to the processing with MetMax, GCxGC-TOF-MS raw chromatograms were deconvoluted and converted into text format using the LECO software. MatMax creates a matrix, where the columns represent experiments and the rows an analyte/metabolite. With some mouse-clicks one can select different output styles: retention times, retention indices, intensity or area of a given quant mass. Dependent on the chosen operating mode, a matrix containing result identifiers (metabolite names, metabolite names together with isotope masses, etc.) in the row headers, analysis identifier in the column headers and results (retention times, retention indices, intensities of peaks, matching analyte spectra, etc.) in the cells, is created.

Reading the given data files in succession and processing the value within the cell corresponding to the current file and observed operation mode depending result, the DataTable representing the result matrix is assembled step-by-step. A test-version of the program will be made available upon request.

#### Statistical analysis

For independent component analysis (ICA) an in house Matlab script was used [22]. The covariance of the data was first analysed by PCA giving a restricted set of principal components (PCs) covering 95% of variance. ICA was then applied to these new components and new independent components were ranked by the kurtosis measure. The contributions of each metabolite to an independent component can be obtained by combining the transformation matrix *W* of PCA with the transformation matrix *V* of ICA to a direct transformation  $U = W \cdot V$ . The elements of the *i*-th vector in *U* represent the individual contributions – the loadings – to the *i*-th independent component (IC<sub>*i*</sub>). For more details see

Scholz & Selbig [22]. The ICA algorithm used in this study is CuBICA4.

## Results

### A novel algorithm MetMax for batchwise processing of GCxGC-TOF-MS analyses using first and second retention index and mass spectral library search for metabolite identification

One of the bottlenecks for the application of GCxGC-TOF-MS in metabolomics experiments with high sample numbers and large numbers of analytes is sample and data processing including metabolite identification and sample versus metabolite alignment. Therefore new workflows for data analysis need to be established. We have chosen a strategy where the deconvolution and peak finding algorithms of the Leco Chromatof® software are used for raw-data export. Subsequently individual peaks are aligned and quantitative mass tags are exported using the new designed software tool MetMax (metabolite matrix builder, see methods). This tool comprises several possibilities of data analysis and export (see Materials and Methods and Fig. 1).

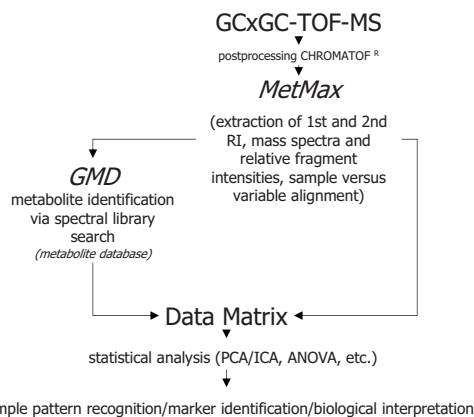
Essentially, a first and a second retention index are calculated from the raw-data and used for library search and metabolite identification in an in house metabolite mass spectral reference library called GMD [1].

The spectra and 1<sup>st</sup> and 2<sup>nd</sup> retention index information of reference compounds are stored in the GMD database and used to generate a 2D library. The program MetMax is able to use the imported spectra and retention index information for identification search in GMD and subsequent sample versus metabolite alignment to generate a data matrix for convenient statistical analysis.

Based on this data export, stable isotope labelling experiments can be analysed using the export and alignment of the mass isotopomer spectrum (M0-Mn) and related fragment ion intensities.

### A modified quenching and extraction procedure for *Chlamydomonas reinhardtii* cells coping with cell leakage

Starting with an extraction protocol established by Bölling and Fiehn [23], we analysed key factors influencing metabolite concentration and cell leakage: altered methanol concentration, homogenization of the pellet, and suspension time cells remaining in the quenching solution (at  $-20\text{ }^{\circ}\text{C}$ ) on internal metabolite concentration and cell leakage.

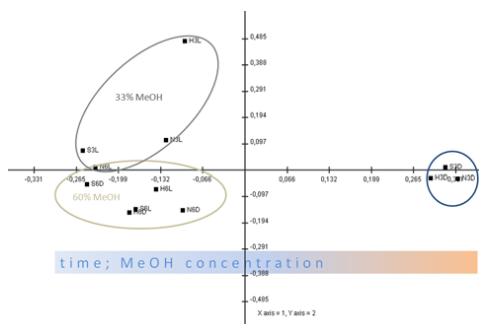


sample pattern recognition/marker identification/biological interpretation

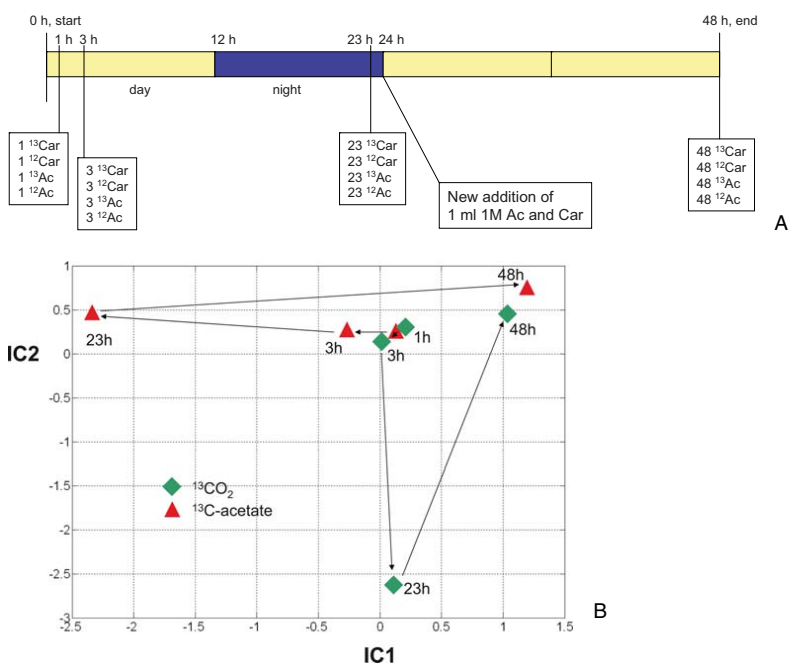
**Figure 1.** Strategy for batchwise raw-data processing and metabolite identification from GCxGC-TOF-MS data combining Leco Chromatof® software and a novel algorithm called MetMax. Raw-data are postprocessed with the chromatof software. Deconvoluted mass spectra are exported with a first and a second retention index (RI1 and RI2) and imported into library search with the GMD database [1]. The algorithm MetMax uses the total information of sample-metabolite information to perform a data alignment. A data matrix is generated where the samples and the metabolites are aligned. This data matrix can be used for statistical analysis and the identification of relevant biomarkers using multivariate statistics [9]. The algorithm is also able to export total mass spectra and isotopomer intensities which were used in the present study to calculate  $^{13}\text{C}$ -stable isotope incorporation into metabolites.

Using a laboratory strain CC503 the experiments revealed that the extraction procedure (sonication, homogenization or no further treatment) caused no remarkable metabolic differences, whereas an increase of methanol concentration and the time cells remain in the quenching solution have a strong influence on metabolite concentration indicating strong cell leakage. A large number of metabolites decreased remarkably, e.g., small amino acids such as glycine were found reduced by up to 80%. Thus, a quenching procedure with subsequent separation of cells from the medium introduces uncontrolled metabolic changes (see Fig. 2). We tested filtering *Chlamydomonas reinhardtii* cells to separate cells from the medium and obtained approximately  $0.3 \times 10^6$  cells/cm<sup>2</sup> (0.2 mg cells) which is a strong limitation of biomass for analysis. Furthermore, protein profiles were drastically changed by this procedure (personal communication Stefanie Wienkoop). Many of the observed phenomena might be caused by the cell wall deficiency of the strain used, CC503.

To overcome these problems, we decided to harvest, quench and analyse without separating cells and medium. To achieve this we had to adjust the composition



**Figure 2.** PCA plot of different extraction methods of *Chlamydomonas reinhardtii* cells. Different quenching was performed with variations of the MeOH concentration and variations of the time the cells were remaining in the quenching solution. Quenching of *Chlamydomonas reinhardtii* culture: using 32.5% MeOH (3) and 62.5% MeOH (6), directly (D) or later (L; 30 min) processed; homogenised (h), sonicated (s), no treatment (n); each point represents 3 replicates. Each quenching/extraction method has an influence on the metabolite composition if cells and growth medium are separated. It is only possible to compare results when identical quenching/extraction/separation procedures are used taking varying metabolite leakages and recoveries into account. Therefore we decided to quench/extract the total sample without separation of cells and growth medium (see Materials and Methods for extraction process and Results for further details).

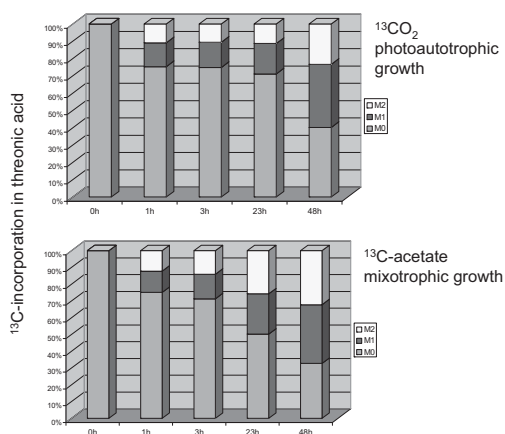


**Figure 3.** (A) Design of the  $^{13}\text{C}$ -labelling experiment. A timecourse was measured and the metabolite fraction was analysed by GCxGC-TOF-MS. (B) Differential dynamics of metabolite labelling pattern analysed by unsupervised sample pattern recognition technique (PCA/ICA). First, the raw-data were processed as shown in Fig. 1 with MetMAX. Subsequently, the percentage of  $^{13}\text{C}$ -incorporation was calculated from the mass isotopomer distribution. The resulting total data matrix of  $^{13}\text{C}$ -metabolite incorporation was loaded into multivariate statistical analysis and the sample pattern was visualised. Strongest differences between  $^{13}\text{CO}_2$  and  $^{13}\text{C}$ -acetate appear after 23 h labelling in a day/night rhythm. An additional  $^{13}\text{C}$ -labelling pulse under continuous light shifts the cells in the same direction indicating a metabolic dilution effect.

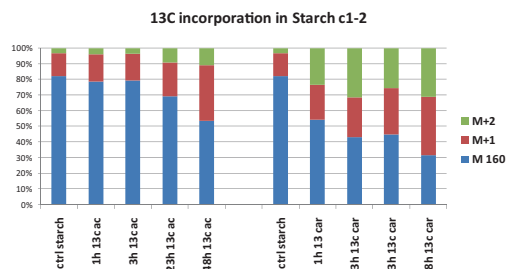


of the growth medium to be compatible with subsequent GC-TOF-MS analysis. Essentially, the Tris-buffer of conventional TAP growth medium was exchanged against HEPES buffer (see methods). Substitution of TRIS against HEPES was beneficial because of the interference of high concentrations of TRIS with the GC/MS analysis. High concentrations of TRIS disturb by overloading the chromatography. This problem was not observed with HEPES buffer. The reproducibility of the method was tested and found comparable to other 'in house' standard procedures for the analysis of plant metabolomes (e.g. Arabidopsis). The reproducibility of the new method was tested by repetitive extraction and preparation of the same sample material and found to be below 20% coefficient of variation (supplemental Table S1).

**GCxGC-TOF-MS analysis of *Chlamydomonas reinhardtii* cells grown under photoautotrophic/mixotrophic conditions using  $^{13}\text{C}$   $\text{CO}_2$  and  $^{13}\text{C}$  acetate**  
*Chlamydomonas reinhardtii* cells were grown under photoautotrophic ( $^{13}\text{CO}_2$ ) and mixotrophic ( $^{13}\text{C}$ -acetate) conditions (see Fig. 3). Cells were harvested after different time points and metabolites were analyzed using GCxGC-TOF-MS. The isotopomer matrix was analysed



**Figure 4.** Differences in  $^{13}\text{C}$ -incorporation in threonic acid depending on photoautotrophic growth or mixotrophic growth. The isotopic masses from the indicated time points were extracted from the mass spectra using MetMax. Due to the pattern of  $^{13}\text{C}$  accumulation threonic acid was identified as a discriminatory marker in the PCA/ICA analysis (see Fig. 3B). There is a significant difference in the intensities of the mass isotopomer distribution at the time point 23 h in contrast to the later time point of 48 h.



**Figure 5.**  $^{13}\text{C}$ -incorporation into starch shown mass fragment 160 (including M+1 and M+2 isotopic masses) representing the C1 and C2 residue of glucose.

by the MetMax software and dynamics of the  $^{13}\text{C}$  incorporation in the distinct metabolite pools inspected.

In Fig. 3B the complete mass isotopomer matrix was analysed by principal/independent component analysis (PCA/ICA). The visualization of the sample pattern reveals different trajectories of  $^{13}\text{C}$ -labelling dynamics under these different growth conditions with the time point 23 h as most discriminatory. One of the metabolites which showed differences in the labelling pattern at this time point and identified in the loadings of the PCA/ICA plot is threonic acid (see Fig. 4). Interestingly, after further 24 hours of labelling under continuous light the differences between the photoautotrophic and mixotrophic growth seem to diminish indicating a metabolic dilution effect.

#### Starch $^{13}\text{C}$ -labelling course

The incorporation of  $^{13}\text{C}$  into starch was monitored within the glucose moieties after starch hydrolysis by GC-TOF-MS (Fig. 5). There is a clear difference observed between photoautotrophic and mixotrophic growth. The kinetics of starch-labelling is much faster in the photoautotrophic grown cells. For detailed analysis of carbon flow under photoautotrophic and mixotrophic growth conditions metabolic models have to be established.

#### Discussion

##### *Chlamydomonas reinhardtii* quenching and extraction

For analyses of cellular metabolite levels it is necessary to stop the metabolic activity of the cells within milliseconds. To apply fast quenching of metabolism cold or heat shock inactivation is largely used. Injection of *Chlamydomonas reinhardtii* culture into cold MeOH

(aprox.  $-50\text{ }^{\circ}\text{C}$ ) is an approved method to stop metabolism. In our experiments we found the cell leakage rate very high thereby disturbing any metabolite profiling approach. Our experiment demonstrated that *Chlamydomonas reinhardtii* cells display a high leakage rate when subjected to a MeOH solution. The leakage is increasing at higher MeOH concentration and shows also a time dependency – longer exposure leads to higher metabolite loss (Fig. 2). Boelling and co-worker have shown that wild type *Chlamydomonas reinhardtii* strains exhibit the half leakage rate compared to cell wall deficient strains that were used in our studies. The high TRIS concentrations in the commonly used TAP medium prohibit GC/MS based metabolomics studies (see methods). The use of HEPES buffer enabled us to use the *Chlamydomonas reinhardtii* cells and medium for metabolomics studies. Including also the growth medium into metabolomic analyses the sum of intra and extra cellular metabolite pools is measured. However, our data demonstrate, that the observed metabolites display a fast incorporation of stable isotope labelled carbon sources. Thus we conclude that a major part of the metabolite pools is inside the cells and is accessible for cellular metabolism. This method was tested for its reproducibility and we found for the large number of metabolites a technical variance of 10–20% which is comparable to other published GC/MS based studies (supplemental Table 1).

#### GCxGC-TOF-MS

The GCxGC-TOF-MS technology provides high resolution, high sensitivity and cleaner deconvoluted mass spectra. Thus it is the method of choice for stable isotope labelling experiments using ultracomplex samples from biological origin. The drawback is still the absence of good algorithms and software to analyze GCxGC-TOF-MS chromatograms batchwise. A major aim was to generate a typical sample versus metabolite data matrix based on the export of specific – or all – mass fragments from the deconvoluted mass spectra. Thus we established a new software tool (MetMax; metabolite matrix builder) which can perform those actions. MetMax imports the text export from Leco Chromatof<sup>®</sup> including information about compound name, quant mass, area, retention time on first and second dimension, quantification result and the deconvoluted mass spectra. However, during stable isotope labelling experiments the mass spectra composition is changing dramatically and therefore a library-search is much more difficult or even impossible. Thus, a manual postannotation is obligatory.

#### Data processing using MetMax

MetMax can align the found and annotated peaks within all chromatograms and automatically export the intensities of the selected masses. Therefore a list of compounds and the selected masses and range of isotope masses was defined and was imported in MetMax – alternatively 1D or 2D data can be analyzed by MetMax. Additionally MetMax provides the possibility to import deconvoluted chromatograms including the annotation of the retention time markers (alkanes). MetMax then automatically calculates the retention indices for first and second dimension respectively the relative retention time on the second dimension (Table 1). The retention index information from first and second dimension as well as the mass spectra are compared to library information. MetMax aligns found compounds in all samples and quantifies the identified metabolites according to a list of quant masses stored in the GMD.

#### Initial $^{13}\text{C}$ -stable isotope labeling experiments with *Chlamydomonas reinhardtii* cells

Stable isotope labelling experiments were performed using  $^{13}\text{C}$  sodium hydrogencarbonate to analyse photoautotrophic carbon flow photoautotrophic and  $^{13}\text{C}$  sodium acetate for mixotrophic growth. Thus the main routes of carbon assimilation are compared within the same organism. Our data show that the different pathways of carbon assimilation under photoautotrophic and mixotrophic growth are reflected in the isotopomer abundance of the monitored metabolites. The PCA/ICA analysis shows that the label distribution within the mass spectra of selected metabolites differs mainly at the 23 h time point, meaning that the route of carbon differs upon photoautotrophic and mixotrophic growth according to the different pathways of  $^{13}\text{C}$ -incorporation. Photoautotrophic carbon assimilation proceeds via the Calvin-Benson cycle and mixotrophic growth mainly by glyoxylic acid cycle and subsequent TCA cycle. At the 48 h time point the label distribution is more comparable for both growth conditions indicating a dilution effect of the  $^{13}\text{C}$ -incorporation. One metabolic marker could be identified as threonic acid (see Fig. 4). The mass isotopomer distribution is clearly different at the time point 23 h ( $^{13}\text{C}$ -incorporation via  $^{13}\text{CO}_2$  metabolic labelling after 23 h: M0 50%, M1 23%, M2 27%;  $^{13}\text{C}$ -incorporation via  $^{13}\text{C}$ -acetate metabolic labeling after 23 h: M0 71%, M1 17%, M2 12%) and more similar at time point 48 h ( $^{13}\text{C}$ -incorporation via  $^{13}\text{CO}_2$  metabolic labeling after 48 h: M0 33%, M1 35%, M2 32%;  $^{13}\text{C}$ -incorporation via  $^{13}\text{C}$ -acetate metabolic labeling after 48 h: M0 40%, M1 36%, M2

**Table 1.** 1D and 2D Retention time and retention indices of reference compounds using a GCxGC-TOF-MS protocol.

Compound	1D RI	2D RT	2D FRT	2D RI
Aconitic acid, cis- (3TMS)	1742.2	2.95	1.147804	1901.96
Adenine (2TMS)	1884.8	4.79	1.711064	2795.92
Alanine (2TMS)	1023.55	1.9	1.147631	1223.68
Alanine (3TMS)	1360.9	2.04	0.985312	1334.21
Alanine, beta- (3TMS)	1426.8	2.11	0.979339	1389.47
Anthranilic acid (1TMS)	1510.6	3.88	1.717054	2424.49
Anthranilic acid (2TMS)	1618	3.34	1.39681	2204.08
Aspartic acid (2TMS)	1427.4	2.66	1.234618	1792.86
Benzocic acid, (1TMS)	1266.7	2.66	1.363198	1792.86
Butyro-1,4-lactam (1TMS)	1179.5	2.63	1.426985	1771.43
c10	1000	1.62	1	1000
c12	1200	1.87	1	1200
c15	1500	2.25	1	1500
c18	1800	2.82	1	1800
c22	2200	3.33	1	2200
c28	2800	4.8	1	2800
Caffeic acid, cis- (3TMS)	2101.2	3.9	1.2673	2432.65
Caproic acid, 6-amino- (3TMS)	1725	2.55	1.002446	1714.29
Carbodiimide (2TMS)	964.34	1.62	1.028234	1000
Cinnamic acid, 4-hydroxy-, trans- (2TMS)	1951.8	4.01	1.406374	2477.55
Citric acid (4TMS)	1800.1	2.75	1.029963	1800
Dodecanoic acid (1TMS)	1650.3	2.67	1.096443	1800
Erythritol (4TMS)	1477.6	1.87	0.841924	1200
Erythrose (1MEOX) (3TMS) MP	1441.4	2.09	0.961616	1373.68
Ethanolamine (3TMS)	1258.4	1.8	0.926773	1144
Ferulic acid, trans- (2TMS)	2057.7	4.76	1.599038	2783.67
Fructose (1MEOX) (5TMS) MP	1853.8	2.23	0.807772	1484.21
Fumaric acid (2TMS)	1346.5	2.2	1.072166	1460.53
Fumaric acid, 2-methyl- (2TMS)	1397.8	2.32	1.095921	1550
Galacturonic acid (1MEOX) (5TMS) MP	1919.4	2.51	0.885239	1685.71
Glucaric acid-1,4-lactone (4TMS)	1940.4	3.14	1.103033	2050.98
Gluconic acid-1,5-lactone (4TMS)	1872.8	2.86	1.026042	1831.37
Glucopyranose, D- (5TMS)	1959.1	2.38	0.833488	1592.86
Glucopyranoside, 1-O-methyl-, alpha- (4TMS)	1863.6	2.73	0.983982	1800
Glucose (1MEOX) (5TMS) BP	1899.2	2.31	0.819149	1542.86
Glucose (1MEOX) (5TMS) MP	1881	2.28	0.814452	1521.43
Glucose, 1,6-anhydro, beta- (3TMS)	1700.4	2.84	1.13334	1815.69
Glucose, 2-deoxy- (1MEOX) (4TMS) BP	1774.6	2.34	0.891207	1564.29
Glucuronic acid (1MEOX) (5TMS) BP	1939.4	2.53	0.888749	1700
Glutamic acid (2TMS)	1550.7	2.83	1.225775	1807.84
Glutaric acid, 2-oxo- (1MEOX) (2TMS) BP	1542.6	2.99	1.300665	1933.33
Glutaric acid, 2-oxo- (1MEOX) (2TMS) MP	1568	3.28	1.408453	2160.78
Glyceraldehyde (1MEOX) (2TMS) BP	1214.7	1.94	1.02754	1255.26
Glycerol (3TMS)	1265.6	1.74	0.891716	1096
Glycine (2TMS)_1109.07	1098.4	1.83	1.033063	1168
Glycine (3TMS)	1309.7	1.97	0.982066	1278.95
Hexadecanoic acid (1TMS)	2050.3	3.28	1.128942	2160.78
Homoserine (2TMS)	1361.3	2.35	1.135041	1571.43
Hydrogen sulfide (2TMS)	992.84	1.6	0.993108	984
Lactic acid, DL- (2TMS)	1085.6	1.65	0.955514	1024
Leucine (1TMS)	1178.8	2.04	1.106863	1334.21
Lysine (3TMS)	1881.9	2.83	1.010921	1807.84
Malic acid (3TMS)_1479.83	1470.9	2.32	1.049052	1550
Malic acid, 2-methyl- (3TMS)	1463.4	2.23	1.012733	1484.21
Malonic acid (2TMS)	1214.9	2.1	1.112286	1381.58
Maltitol	2810.1	3.08	0.660466	2003.92
Maltose BP	2749.1	3.23	0.617311	2121.57
Maltose MP	2713.2	3.18	0.584028	2082.35
Mannitol?	1918.7	2.07	0.730058	1357.89
Melibiose	2825.5	3.59	0.794954	2306.12

24%). Threonic acid is suggested to be an intermediate of ascorbate metabolism [24, 25]. However, the exact pathway of  $^{13}\text{C}$ -incorporation into threonic acid in *C. reinhardtii* remains speculative and further experiments have to be carried out to determine the exact fate of threonic acid and its connection to  $\text{CO}_2$ -fixation.

Also the incorporation of stable isotopes within the starch reflects strongly the different routes of carbon assimilation. Starch is faster labelled in  $^{13}\text{CO}_2$  grown cells compared to  $^{13}\text{C}$ -acetate growth. Thus, the next level will be to resolve each individual metabolite with respect to kinetics of  $^{13}\text{C}$ -incorporation and to analyse the obtained data in metabolic flux models, which have to be established. The workflow for GCxGC-TOF-MS stable isotope labelling studies presented in this study in combination with metabolic flux-modelling will be a major resource to extend classical methods like measuring metabolite average levels.

### Acknowledgements

Financial support was provided by Bundesministerium für Bildung und Forschung (BMBF): a grant for „FORSYS – Forschungszentren Systembiologie in Deutschland“ (<http://www.goforsys.de/>).

### References

- [1] Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A. R. and Steinhauser, D., 2005. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, **21**, 1635–1638.
- [2] Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H. *et al.*, 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245–250.
- [3] Depege, N., Bellafiore, S. and Rochaix, J.D., 2003. Role of chloroplast protein kinase Stt7 in LHClI phosphorylation and state transition in *Chlamydomonas*. *Science*, **299**, 1572–1575.
- [4] Mittag, M., 2001. Circadian rhythms in microalgae. *Int. Rev. Cytol.*, **206**, 213–247.
- [5] Ball, S.G., 2002. The intricate pathway of starch biosynthesis and degradation in the monocellular alga *Chlamydomonas reinhardtii*. *Australian J. Chem.*, **55**, 49–59.
- [6] Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. and Darzins, A., 2008. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant Journal*, **54**, 621–639.
- [7] Tsygankov, A.A., Kosourov, S.N., Tolstygina, I.V., Ghirardi, M.L. and Seibert, M., 2006. Hydrogen production by sulfur-deprived *Chlamydomonas reinhardtii* under photoautotrophic conditions. *International J. Hydrogen Energy*, **31**, 1574–1584.
- [8] May, P., Wienkoop, S., Kempa, S., Usadel, B., Christian, N., Rupprecht, J., Weiss, J., Recuenco-Munoz, L., Ebenhoh, O., Weckwerth, W. and Walther, D., 2008. Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. *Genetics*, **179**, 157–166.
- [9] Weckwerth, W., 2008. Integration of metabolomics and proteomics in molecular plant physiology – coping with the complexity by data-dimensionality reduction. *Physiologia Plantarum*, **132**, 176–189.
- [10] Weckwerth, W., 2003. Metabolomics in systems biology. *Annu. Rev. Plant Biol.*, **54**, 669–689.
- [11] Weckwerth, W. (ed.), 2006. *Metabolomics: Methods and Protocols*. Humana Press.
- [12] Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. *Nature Biotechnology*, **18**, 1157–1161.
- [13] Weckwerth, W., Tolstikov, V. and Fiehn, O., 2001. Metabolomic characterization of transgenic potato plants using GC/TOF and LC/MS analysis reveals silent metabolic phenotypes. *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*, American Society of Mass Spectrometry, Chicago, 1–2.
- [14] Weckwerth, W., Wenzel, K. and Fiehn, O., 2004. Process for the integrated extraction identification, and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*, **4**, 78–83.
- [15] Weckwerth, W., Loureiro, M. E., Wenzel, K. and Fiehn, O., 2004. Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl. Acad. Sci. USA*, **101**, 7809–7814.
- [16] Stein, S.E., 1999. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Amer. Soc. Mass Spectrometry*, **10**, 770–781.
- [17] Stein, S.E. and Scott, D.R., 1994. Optimization and testing of mass-spectral library search algorithms for compound identification. *J. Amer. Soc. Mass Spectrometry*, **5**, 859–866.
- [18] Schwender, J., Goffman, F., Ohlrogge, J.B. and Shachar-Hill, Y., 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, **432**, 779–782.
- [19] Sauer, U., 2006. Metabolic networks in motion: C-13-based flux analysis. *Molecular Systems Biology*, **10**, 62–72.
- [20] Strehmel, N., Hummel, J., Erban, A., Strassburg, K. and Kopka, J., 2008. Retention index thresholds for compound matching in GC-MS metabolite profiling. *J. Chromatogr. B*, **876**, 10–18.