Technische Universität München

Lehrstuhl für Allgemeine Lebensmitteltechnologie

Safety assessment of genetically modified food – New methodologies for the analytical characterisation of rice

Thomas Frenzel

Vollständiger Abdruck der von der Fakultät
Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt
der Technischen Universität München
zur Erlangung des Akademischen Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigten Dissertation

Vorsitzender:	Univ.-Prof. Dr. Gerhard Wenzel
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Karl-Heinz Engel
	2. Univ.-Prof. Dr. Peter Schieberle
	3. Prof. Howard V. Davies, PhD, Univ. Cordoba, Spanien

Die Dissertation wurde am 13.02.2003 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 22.04.2003 angenommen.

ACKNOWLEDGEMENTS

The work presented was carried out under supervision of Prof. Dr. Karl-Heinz Engel, chair of the Institute of General Food Technology at Technical University Munich.

I would like to thank Prof. Dr. Karl-Heinz Engel for the opportunity to work on this exciting subject, for his scientific support, for many fruitful and helpful discussions, and for his guidance in completing this thesis.

I am indebted to all partners involved in the EU project SAFOTEST, especially to Prof. Dr. Angharad Gatehouse, Prof. Dr. Illimar Altosaar, and Prof. Dr. Qingyao Shu for providing the genetically modified rice material. I am grateful to have been able to work for six weeks at the Scottish Crop Research Institute. In particular my thanks to Prof. Dr. Howard Davies, Dr. Nigel Deighton, and Dr. Jim McNicol for extended discussions and sharing valuable ideas.

Thanks to Stefan Kreisz from the Institute of Brewing Technology I at Technical University Munich for providing barley material and for performing the micromalting experiments.

I am grateful to have been able to work at the Institute of General Food Technology with all my colleagues, in particular in the Novel Food Group with Dr. Gerhard Gerstner, Dr. Christine Hupfer, Iulia Cioranu, Dr. Willibald Kamm, Francisco Moreano, and Andreas Roß. Special thanks to Andreas Miller for extended and fruitful discussions. I would like to extend my thanks to Ruth Fraek, Michael Kreiter, Elke Müller, Anke Scheffler, Verena Seckinger, Richard Röhlig, Bernhard Stoiber, Markus Wutke, Donald Yewah, and Beate Zuzan for joining my enthusiasm in this exciting subject.

I just named a few here, but many thanks to all of them.

TABLE OF CONTENTS

Abbreviations

ASCII	American Standard Code for Information Interchange
B	Basmati
Bt	*Bacillus thuringiensis*
CV	Column volume
DNA	Deoxyribonucleic acid
EI	Electron impact
EPSPS	5-enol-pyruvylshikimate-3-phosphate synthase
FAME	Fatty acid methyl ester
FAO	Food and Agriculture Organization of the United Nations
FID	Flame ionisation detection
FT	Fourier-transform
GC	Gas chromatography
GM	Genetically modified
GNA	*Galanthus nivalis* agglutinin
HPLC	High performance liquid chromatography
IFBC	International Food Biotechnology Council
ILSI	International Life Sciences Institute
IP_6	Inositolhexakisphosphate
LDPE	Low-density polyethylene
LG	Long grain
MG	Medium grain
MLR	Multiple linear regression
MS	Mass spectrometry, Microsoft
MSTFA	*N*-methyl-*N*-trimethylsilyltrifluoroacetamide
MTBE	Methyl-*tert* -butylether
mu	Mass units
NIR	Near infrared
NIRA	Near infrared accessory
NIRS	Near infrared spectroscopy
NMR	Nuclear magnetic resonance
OECD	Organization for Economic Co-operation and Development
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PLS	Partial least squares
RC	Red camargue
RMSEC	Root mean squared error of calibration
RMSECV	Root mean squared error of crossvalidation
RMSEP	Root mean squared error of prediction
$RMSEP_{Est}$	Estimated root mean squared error of prediction
SD	Standard deviation
SG	Short grain
SPE	Solid phase extraction
TMS	Trimethylsilyl-
TMSIM	Trimethylsilylimidazole
WHO	World Health Organization

1 Introduction and objectives

Genetic engineering is being increasingly applied in modern agriculture. In addition to conventional breeding, recombinant DNA techniques are used to improve agronomic and nutritional traits of food crops. Genetic engineering is a targeted technique which enables the transfer of defined genes without restriction by species barriers [1].

Agricultural biotechnology has opened new avenues in the development of plants for the production of food, feed, fibre, forestry and other products. The first genetically improved crop was the Flavr Savr® tomato which was approved for sale in the United States of America in 1994 [2]. Since then, the cultivation of several transgenic crop species has grown rapidly to more than 58 million ha worldwide [3], *i.e.* approximately 5 % of the total world acreage.

For the current generation of genetically modified (GM) crops, the improvement of agronomic traits (*e.g.* herbicide-tolerance, insect-resistance) has been a major objective. Modifications have resulted in benefits mainly orientated to the farmer and the environment with only indirect benefit to consumers resulting from, for example, pesticide reduction [4]. The lack of obvious and direct benefits for the consumer has been a main point of criticism.

In addition to further improvement of agronomic characteristics, *e.g.* drought and salinity tolerance, future trends will increasingly encompass the modification of quality traits, such as the enhancement of sensory and especially nutritional properties. Some of the ongoing developments try to meet the desire of consumers for healthy or high-tech foods in developed countries. Others are intended to assist in adjusting the nutritional status of foods to the needs of consumers in developing countries [5].

The major prerequisite for the applicability of recombinant DNA techniques in food production is the safety of the resulting products. International consensus has been reached on the principles regarding evaluation of the food safety of genetically modified plants [6, 7]. The concept of "substantial equivalence" [8] is a key element of the strategies agreed upon by international bodies [9, 10]. The concept is used to identify similarities and differences between the genetically modified food and a comparator with a history of safe food use that subsequently guides the safety

assessment process [11]. It is mainly based on the comparison of parental and genetically modified plants taking into account phenotypic characteristics, potential allergenicity of gene products and metabolites, and the chemical compositions, with a special focus on essential nutrients and critical toxicants [12]. This approach has proven to be useful for the safety assessment of the current generation of genetically modified crops [13-15]. Based on comprehensive comparative investigations, presently commercialised GM crops were shown to be equivalent to traditional counterparts except for well-defined differences. The key difference to their respective comparators are the transferred proteins on which safety testing was focused.

The adequacy of the current approach which is based on targeted compositional analysis to reveal potential unintended effects of genetic modifications has been challenged [16]. An unintended effect represents a statistically significant difference in the phenotype, response, or composition of the GM plant compared with the parent grown under identical environmental conditions which goes beyond the intended effect of the target gene [17]. Considering the increasing complexity of genetic modifications expected for the future generation of GM foods, development of profiling methods for the identification and characterisation of unintended effects has been recommended [6]. These techniques are not intended to replace any existing safety assessment procedure, but aim to provide a more complete compositional characterisation of the GM organism, in order to increase the chances of detecting unintended effects due to the application of recombinant DNA techniques.

Objectives

The comparison of the chemical composition of the genetically modified plant to that of a traditionally obtained counterpart has been a key element in the safety assessment of genetically modified crops. One of the most critical points of this procedure is the selection of compounds to be analysed. "Targeted approaches" are based on the qualitative / quantitative analysis of defined constituents. Critics of the targeted approach refer to the biased character of the concept focusing on known compounds and expected / predictable changes [16]. Profiling methods may be used

to overcome these limitations. They may be applied to different classes of cellular constituents, *e.g.* the genes, the proteins, and the metabolites.

The objective of this study was to contribute to the assessment of different profiling approaches as regards their suitability to analyse low molecular weight constituents present in a food matrix. This should be performed in order to evaluate their applicability within the safety assessment framework of GM foods. Rice (*Oryza sativa* L.) was used as model crop for method development.

Near infrared spectroscopy (NIRS) should be tested as rapid non-destructive technique enabling unspecific fingerprinting of rice grain constituents. Investigations should be focused on the possibilities to extract quantitative information on a single nutritionally relevant minor compound from complex NIR spectra. Therefore, an NIRS method for determination of the phytic acid contents in brown rice grains should be developed.

In order to complement methods for targeted investigation of single constituents, a metabolite profiling methodology should be elaborated which allows non-targeted analysis of a broad spectrum of low molecular weight compounds present in rice grains. To meet the challenge of the generation of huge amounts of data, the study should also include the development of tools for automated analysis of metabolite profiling data. The techniques involved should be checked in terms of their applicability to other cereal crops, *i.e.* maize and barley.

Finally, the methodologies should be applied exemplarily to the analytical characterisation of genetically modified rice.

2 Background

2.1 Use of GM technology in agriculture

Within a rather short period of time genetic engineering has evolved from basic science to commercial applications. The techniques developed for transformations of plants were readily applied to crops, and GM plants used for food production are now a reality [3].

Starting from 1.7 million ha in 1996, the global area of transgenic crops significantly increased to 58.7 million ha in 2002 [3]. So far, the spectrum of commercialised GM crops is essentially limited to soybean, maize, cotton and canola. Ninety-nine percent of the GM crops planted exhibit herbicide-tolerance and insect resistance as target traits [3]. This "first generation" will probably be followed by a next wave of applications of recombinant DNA techniques in food production characterised by approaches to improve the nutritional properties of raw materials and the foods made thereof ("second generation") [18].

2.1.1 Agronomic traits

The application of herbicides is an indispensable part of modern agriculture to combat weed damages. Tolerance to glyphosate (Roundup Ready®) and glufosinate (BASTA®), respectively, has been conferred to major food crops using genetic engineering. Herbicide-tolerant plants are grown on 75 % of the global area of transgenic crops. Herbicide-tolerant soybean was the most dominant (62 %) transgenic crop grown in three principal countries (USA, Argentina, Canada) in 2002 [3].

The commercially second most important trait conferred to crops by genetic engineering is insect-resistance. Plants expressing δ-endotoxin from *Bacillus thuringiensis* (Bt) have been grown on 19 % of the global area of transgenic crops. This strategy has been applied to confer insect-resistance to crops such as canola, corn, cotton and potato. Alternative strategies to achieve insect-resistance to plants, which might become of practical importance in the future have been described. They include the expression of lectines [19] and enzyme inhibitors [20].

The present improvements of agronomic properties are tailored to the needs of agriculture in developed countries. Additional strategies to improve crop productivity especially in developing countries are based on increasing crop tolerance to abiotic stress as evoked by rough environmental conditions, *e.g.* alkaline soils, metal toxicity, drought, and salinity. The tolerance of rice to low iron bioavailability in alkaline soils could be improved by genetically engineering the crop to release more iron-solubilising chelators [21]. Expression of bacterial citrate synthase in roots is one of the strategies presently being developed to overcome the problem of toxicity of aluminum released by acid soils [22]. Forty percent of the irrigated land is affected by salinity. The resulting decline in photosynthesis and the increased formation of oxygen radicals limits the crop performance. Genetic engineering can be used to increase the cellular content of osmolytes [23], to express antioxidative enzymes [24], or to introduce sodium pumps [25]. Pathways involving the biosynthesis of metabolites like polyamines, proline, glycinebetaine and trehalose were shown to be related to drought resistance [26, 27].

2.1.2 Quality traits

In industrialised countries the improvement of the agronomic performance of crops does not appear as a direct benefit to the customer purchasing the food in the super market. In order to convince the consumer on the advantage of genetic engineering and thus to increase the acceptance of GM foods, obvious traits such as sensory or nutritional properties have to be improved [5]. Genetic engineering can be applied to modify macronutrients, *e.g.* starch, fat, and protein, as well as micronutrients in foods.

Modern plant biotechnology has focused very early on the genetic modification of plant carbohydrate metabolism. Changing starch composition or accumulating fructanes in transgenic crops are examples for relevant applications [28, 29]. Modification of the chain lengths and the degree of saturation of fatty acids has resulted in commercialised products, such as high laurate canola oil [30] and sunflower seed oil with high oleic acid content [31]. Various strategies have been described to improve the protein quality of foods and feeds [32]. This ranges from functional properties, *e.g.* baking quality of wheat [33], to nutritional properties, *e.g.* enhancement of the content of essential amino acids [34].

The wide range of isoprenoids found in plants and the integration of metabolic pathways of steroids, carotenoids and retinoids [35] offers the potential to influence the content of these compounds by genetic engineering. An exciting current example is the genetically engineered biosynthesis of β-carotene in the so-called "Golden Rice" [36]. In order to increase the iron content of rice, three approaches have been combined: (i) introduction of the ferritin gene from *Phaseolus vulgaris* into rice; (ii) expression of a heat-tolerant phytase from *Aspergillus fumigatus* in rice; and (iii) overexpression of cystein-rich metallothionein-like protein endogenous in rice [37].
Owing to the present wave of "functional foods" increasingly put on the market in Western countries, there is increasing interest in bioactive plant constituents. Modifying composition and distribution of high-value compounds, such as carotenoids and vitamins or flavonoids and isoflavonoids may become "gold mines" for metabolic engineering [5, 38].

2.2 Food safety assessment of transgenic crops

2.2.1 Food safety evaluation strategies

The major prerequisite for the applicability of GM technology in food production is the safety of the resulting products. Efforts to define internationally harmonised strategies for the safety evaluation of foods derived from genetically modified organisms (GMOs) began already at the very early stage of the exploitation of recombinant DNA technology in plant breeding [8-11]. This eventually resulted in various national and international regulatory frameworks concerning market introduction of genetically modified foods. In the European Union GM food is currently regulated by Regulation (EC) 258/97 on Novel Food and Novel Food Ingredients [39]. It lays down a Community procedure for the authorisation of foods derived from GMOs including the general principles and requirements regarding food safety assessment. In the European Union [12] the risk assessment strategy for GM plants and products is based upon a consideration of:

- the characteristics of the donor and recipient organisms;
- the genes inserted and expressed;
- the potential consequences of the genetic modification;

- the potential environmental impact following a deliberate release;
- the potential toxicity and allergenicity of gene products and metabolites;
- the compositional, nutritional, safety and agronomic characteristics;
- the influence of food processing on the properties of the food or feed;
- the potential for changes in dietary intake;
- the potential for long-term nutritional impact.

The principle of "substantial equivalence" elaborated by international bodies (IFBC, OECD, FAO/WHO, ILSI) in several workshops and expert consultations since the beginning of the 1990s became a key element in the safety assessment of foods derived from GMO. The concept is used to identify similarities and differences between the GM food and a comparator with a history of safe use that subsequently guides the safety assessment process [11]. Application of this principle is a starting point in the safety evaluation, rather than an endpoint of the assessment.

An important part of this strategy is the comparison of the chemical compositions between parental and genetically modified food, with a special focus on essential nutrients and critical toxicants. Key nutrients, anti-nutrients and toxicants characteristic for plant species and information on the extent of natural variability are provided in OECD consensus documents. Such documents have been prepared for soybean, oilseed rape, maize, and potato [40, 41], and others on sunflower, sugar beet, wheat, barley and rice are in preparation [12]. They provide further guidance for compositional analysis to establish the extent of compositional equivalence. This may result in three scenarios [42]:

- A GMO-derived product to be placed on the market is equivalent within the limits of the natural variability to a conventional counterpart. In that situation no further safety testing is needed. The product is considered as safe as its traditional counterpart. Defined and well-characterised ingredients isolated from GM plants, *e.g.* sugar from herbicide-tolerant sugar beet, would fall into this category.

- A GMO or a derived product to be placed on the market is equivalent to a traditional counterpart except for some well-defined differences which are due to the genetic engineering process. In that case the safety assessment can be targeted to the identified differences. The GMOs presently on the market belong

to that category. The key difference with their respective comparators is the presence of the transferred proteins on which safety testing is focused.

- A GMO or a derived product to be placed on the market differs from a traditional food in multiple and complex respects, or there is no traditional counterpart. Such a food is not unsafe *per se* but would require a comprehensive safety assessment.

The concept of "substantial equivalence" proved to be suitable for the "first generation" of GM crops [13-15]. However, as outlined in 2.1.2, the next generation of GM foods will not be limited to plants with agronomic advantages but will increasingly focus on improvements of the nutritional properties of a crop. This will go along with an increasing complexity of genetic modifications. For example, in "Golden Rice", with introduced β-carotene biosynthesis pathway into the endosperm, a transfer of genes from *Narcissus pseudonarcissus* coding for phytoene synthase and lycopene-cyclase under control of the rice glutelin promotor, as well as from *Erwinia* bacteria coding for phytoene-desaturase under the 35S promotor was performed [36]. Owing to the unavailability of a staple food as appropriate conventional comparator, identification and characterisation of effects of the genetic modification which go beyond the primary intended effects of introducing the target genes will become increasingly challenging.

2.2.2 Identification of unintended effects

The potential occurrence of "unintended effects" is one of the concerns being raised as regards the application of recombinant DNA techniques in the production of foods. "Unintended effects" represent a statistically significant difference in the phenotype, response, or composition of the GM plant compared with the parent from which it is derived, but taking the expected effect of the target gene into account. They may be explicable in terms of current knowledge of plant biology and metabolic pathway integration and interconnections ("predictable unintended effects") or may fall outside the present level of understanding ("unpredictable unintended effects") [17]. Unintended effects may be identified by analysis of the agronomic / morphological characteristics of the new plant and by extensive compositional analysis. The breeder usually performs phenotypic comparison which has been used successfully in conventional plant breeding and has led to many new varieties with virtually no

unintended negative consequences for the consumer. The strategies applied for compositional analysis may follow a targeted and a non-targeted approach.

2.2.2.1 Targeted approach

Targeted approaches for identification of unintended effects are based on the qualitative / quantitative analysis of defined constituents. This may range from proximate analysis to the investigation of major, *e.g.* amino acids, fatty acids, and minor constituents, *e.g.* minerals. Special emphasis has been put on naturally occurring toxicants and essential nutrients or anti-nutrients. The analyses also comprised compounds which, from current biological and biochemical knowledge, might be predicted to change in response to the specific genetic modification. An example for investigation of such "predictable unintended effects" is the analysis of phenylalanine, tyrosine and tryptophane in Roundup Ready® soybeans, which took into account that the EPSPS is involved in the biosynthesis of aromatic amino acids [43].

The application of a targeted approach for comparative analysis of defined constituents has been shown to be suitable for the detection of statistically significant differences between parental and GM lines which go beyond the intended effects of the genetic modification. Genetically engineered biosynthesis of a nutritionally valuable protein (soybean glycinin) in rice and potatoes resulted in an increased content of a nutritionally important (vitamin B_6) as well as of toxicologically critical compounds (glycoalkaloids). These unintended effects were not predicted and have not been explained by the authors [44, 45].

Genetic modifications targeting key steps in primary metabolism were shown to provoke profound changes in the generated GM plants. Expression of a yeast invertase in potato tubers resulted in an altered carbohydrate metabolism and also in a reduced content of glycoalkaloids. Detailed investigations indicated that the lower glycoalkaloid content corresponded to a status of advanced maturity of the genetically modified potato relative to the conventional counterpart grown under identical environmental conditions [46].

As shown for GM canola, overexpression of an enzyme in the isoprenoid metabolism (phytoene-synthase) resulted in altered levels of the metabolites downstream the target of genetic modification (phytoene) but also in metabolic perturbations

effecting other compounds (tocopherols, chlorophyll) whose biosynthesis is linked to the modified pathway [47].

As regards the so-called "Golden Rice", introduction of phytoene synthase and phytoene desaturase into the endosperm of a single transformant was expected to induce the formation of lycopene. Surprisingly, α-carotene, β-carotene, lutein and zeaxanthine were synthesised instead of lycopene. This was explained by a further conversion of the newly introduced substrate lycopene by lycopene cyclases inherently present in normal rice endosperm [36].

2.2.2.2 Profiling techniques

In order to increase the chances to detect unintended effects, profiling methods have been suggested as tools for characterisation of changes in the composition of GM plants [6, 11]. This may be of particular relevance for GM food crops with improved nutritional or health-protecting properties, obtained through the insertion of multiple genes. The non-targeted approach exploiting *Genomics*, *Proteomics* and *Metabolomics* aims at an "unbiased" profiling of potential changes in the physiology and metabolism of the modified host organism at different cellular integration levels. In a recent review the potential of these methods for studying molecular genetics and physiology of plants has been discussed [48].

Genomics. Localisation and characterisation of the place(s) of insertion are the most direct approaches to predict and identify the possible occurrence of unintended effects due to the application of recombinant DNA techniques. With increasing knowledge of plant genomes including annotations in genomic databases and understanding of the genomic code and regulation of gene expression, sequencing of the place(s) of insertion will become increasingly informative [6].

The DNA microarray technology is a powerful tool to study gene expression. The major advantage of the microarray technology over conventional gene profiling techniques is that it allows small-scale analysis of the expression of a large number of genes in a sensitive and quantitative manner [49].

Proteomics. The study of the entire set of proteins present in a cell, organism, or tissue under defined conditions, is a well-established technique enabling the next level of analysis after transcript profiling. The main approach currently applied

involves two-dimensional gel electrophoresis followed by excision of protein spots from the gel, digestion into fragments by specific proteases, analysis by mass spectrometry and subsequent computer-assisted identification using databases [50]. The use of specific detection methods, *e.g*. immunoblotting, should also allow to screen specifically for potentially adverse proteins, *e.g*. allergens or anti-nutrients.

Metabolomics. Metabolites can be regarded as the ultimate gene products. *Metabolomics* aims to provide a comprehensive picture on the entirety of metabolites present in biological systems. Main approaches used in the course of *Metabolomics* studies are based on metabolic fingerprinting as well as on metabolite profiling techniques.

Metabolic fingerprinting. This approach aims to rapidly provide a picture on the low molecular weight compounds present in raw samples. Sample preparation is frequently restricted to a simple step and chromatographic pre-separation is avoided. Metabolic fingerprinting techniques solely rely on the resolving power of methodologies such as nuclear magnetic resonance (NMR) [51], mass spectrometry (MS) [52], and Fourier-transform (near) infrared spectroscopy (FT-NIRS) [53]. The potential of these techniques lies in the ease of sample preparation, the speed with which data can be acquired, and the high degree of reproducibility.

Metabolite profiling. The most important approaches used for metabolite profiling are based on hyphenated techniques combining various analytical tools, *e.g*. gas chromatography (GC), high performance liquid chromatography (HPLC), MS, and NMR. These methods aim at detecting, resolving, identifying, and quantifying a wide range of low-molecular weight compounds in a single sample. They can be totally unbiased or targeted to metabolites in key metabolic pathways [54].

2.3 NIRS analysis of foods

2.3.1 Introduction

The increasing demand for product quality improvement and production rationalisation in the food and agricultural industries has led to the gradual substitution of time-consuming analytical methods by more rapid and environmentally friendly analytical tools. In this respect Near Infrared Spectroscopy (NIRS) has emerged over the last years as an extremely powerful tool for quality control and process monitoring [55].

NIRS has many advantages over chemical and other instrumental methods of food and beverage analysis in terms of sample preparation and sample throughput. Spectra of intact food samples may be measured in seconds by reflectance or transmittance and mathematical processing of the spectral data enables absorption information to be extracted so that constituents and quality attributes can be determined [56].

NIRS allows to obtain a comprehensive fingerprint containing qualitative and quantitative information on the constituents of a sample. After establishment of appropriate calibrations, multicomponent analyses can be performed on the basis of a single spectrum. This explains the widespread use of NIRS for screening of large sample sets in plant breeding programmes. For the same reason, NIRS might be a powerful tool for comparative compositional analysis of GM plants in the course of a safety assessment. It might replace time-consuming wet-chemical procedures owing to the possibility to simultaneously cover a broad spectrum of constituents or single compounds in a non-destructive way.

The following chapters cover basic principles of NIRS with respect to food analysis. The approaches to develop an NIRS method will be described and areas of application with special focus on cereals will be pointed out.

2.3.2 NIRS instrumentation for food analysis

Adjacent to the mid-infrared, the NIR region covers the interval between approximately 4000 and 12500 cm^{-1} (800 nm – 2500 nm). Tungsten halogen lamps are the most frequently employed near-infrared continuous spectral light sources.

Photoconductive (*e.g.* PbS) and photovoltaic (*e.g.* InGaAs) elements are commonly used as detectors for NIR light.

Up to 1978, only filter instruments designed for specific applications were commercially available. The increasing availability of monochromator instruments for routine analysis paved the way to a broader applicability of NIRS to food analysis [56]. Instruments based on Fourier-transform (FT) spectroscopy have several advantages compared to monochromator instruments in terms of signal detection efficiency and degree of accessible spectral resolution [57]. Therefore, FT spectroscopy is increasingly applied for spectroscopic analyses in the NIR region. However, due to their robustness and simplicity, filter and monochromator instruments are still widely used especially in agricultural industries.

Food samples to be analysed by NIRS are often unground and heterogeneous. Several devices like cuvettes, cells, or petri dishes made of glass or quartz are in use for sample presentation. Liquids are usually measured in transmission or in transflection. Solid materials may be measured in transmission or by diffuse reflectance. Sample spinners were developed for moving the surface of heterogeneous samples over the measuring aperture. In near infrared diffuse reflectance spectroscopy of solid material, the light is scattered at or in the sample into a wide range of angles, so that complete introduction of the light into the spectrometer is difficult. Special cells and integrating spheres are used for improving the collection efficiency of diffusely reflected light.

Figure 1 shows the measurement set-up for Fourier-transform near infrared diffuse reflectance spectroscopic analysis of solid sample material, *e.g.* rice flour. Starting from the radiation source, the light beam passes the interferometer in order to reach an iris blend, which restricts the beam divergence to the maximum acceptable for the required resolution. After passing a filter wheel, which allows reduction of the wavelength range, the beam enters the integrating sphere, where it is focused to the sample cup by means of a mirror (not shown in Figure 1). The sample cup containing solid inhomogeneous material is mounted onto a spinner, which steadily moves the surface of the sample over the measuring aperture. The NIR light reflected from the solid sample into a wide range of angles is collected by the integrating sphere and eventually focused to the detector.

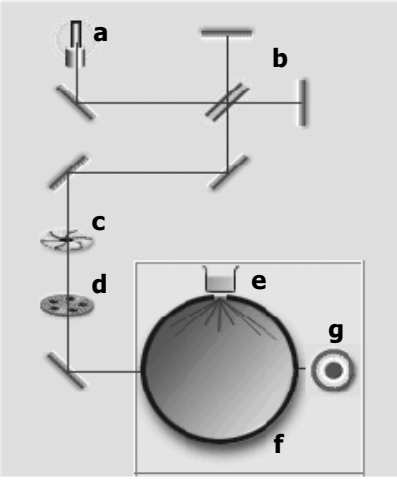

Figure 1: Measurement set-up and optical path for Fourier-transform near infrared diffuse reflectance spectroscopic analysis of solid sample material [58]. (**a**): NIR radiation source; (**b**): interferometer; (**c**) iris; (**d**): filter wheel; (**e**): sample cup with solid material mounted onto a sample spinner; (**f**) integrating sphere; (**g**): detector

2.3.3 Qualitative information in NIR spectra

NIRS is based on the absorption of light in the wavelength range of 800 nm to 2500 nm. NIR spectra of foods show broad bands which comprise overlapping absorptions corresponding to overtones and combinations involving C-H, O-H or N-H chemical bonds. Since almost all compounds found in foods contain those types of chemical bonds, NIR spectra obtained from foods include comprehensive information on their constituents.

The carbon – hydrogen (C-H) bond is the most frequent one in organic molecules. The first stretch overtone band in alkanes can be found around 1700 nm. The weaker second overtone band is expected to be around 1150 nm. The third overtone exhibits extremely weak absorption in the region of 880 nm. The relative absorptivities of the first, second and third C-H stretch overtones in the simple molecule chloroform for example are 484:22:1.

The C-H stretch overtone bands in the more complex food constituents follow similar though more sophisticated patterns due to the effect of C-C coupling of the chain units. C-H bands in wheat protein occur at 1190, 1700, 1730, and 1760 nm, those of starch, pentosanes and lipids occur in the same wavelength ranges.

Intensive overtone bands in NIR spectra of foods result from O-H (first and second overtone), and N-H (first and second overtone) bonds. Important combination bands observed in NIR spectra were assigned to the C-H, O-H, and N-H bonds in proteins,

lipids, carbohydrates. In general, absorptivities of the second and third overtones are weaker than those of the first overtone. Consequently, the most intensive absorptions in NIR spectra are observed at high wavelengths [56].

The band around 1940 nm in the spectrum of liquid water results from a combination of O-H stretch with O-H deformation vibrations. From the analytical point of view, this is the most abundant absorption in the NIR. This kind of combination band also occurs in all compounds containing hydroxyl groups. As with the O-H overtone bands, the O-H combination bands are shifted with hydrogen bonding. With increasing temperature the number of hydrogen bonds decreases which results in modified absorption characteristics.

2.3.4 Principles of quantitative analysis by NIR diffuse reflectance spectroscopy

Sample information in the near-infrared region on solid heterogeneous material is often collected by diffuse-reflectance measurements. In diffuse reflectance mode, the boundary between the sample and the surrounding medium may be considered as a series of small interfaces. If radiation transmitted through the first interface undergoes absorption, the transmitted radiation will be attenuated according to the Beer – Lambert law. The radiation could also undergo further transmittance and absorption at other interfaces. When the radiation encounters discrete particles within the sample which are much larger than the wavelength, destructive interference becomes incomplete and the radiation propagates in all directions (scattering). Particles concentrated into a thick layer give rise to multiple scattering, which results in intensification of absorption bands. This intensification arises from the radiation traversing an optical path which may be as much as eighty times the sample thickness [56].

No rigorous theory has been developed for quantitative diffuse reflectance measurements, but a number of phenomenological theories have been proposed. Equation 1 is based on the Kubelka-Munk theory and represents the mathematical model upon which the quantitative NIRS analysis of foods in diffuse reflectance mode is based:

equation 1:

$$c = k + (s/a) * log(1/R)$$

R : reflectance of the sample
s : scattering constant of the sample
a : absorptivity of the analyte
c : concentration of the analyte

The constant a is the absorptivity of the analyte as defined by the Beer – Lambert law. The actual value of s depends on a number of properties of the sample; the most important are particle size and moisture content. The art of NIRS of scattering samples lies in standardisation of these sample properties or in the selection of measurement and reference wavelengths at which s is nearly equal so that the s/a term becomes a constant. Under these conditions, equation 1 allows the establishment of a linear relationship between spectral (reflectance, R) and compositional properties (concentration, c) of a sample measured in the diffuse reflectance mode [56].

2.3.5 Development of an NIRS method for food analysis

Figure 2 summarises the steps for development of an NIRS method for quantitative analysis of a food constituent.

NIRS in food analysis is a secondary methodology and therefore depends on (classical) reference analytical methods in order to build-up calibration models.

The first step in NIRS method development is to perform a calibration experiment. This involves collecting a set of calibration samples and subjecting the samples to analysis by the reference method for the constituent of interest and by the NIR instrument. From these data the calibration method must infer a rule for predicting results in future samples.

Calibration samples should be representative for the population of samples which is desired to be analysed in the future with the instrument. It is recommended to carefully optimise spectral data acquisition in terms of reproducibility and to control the quality of spectral data as well as the reliability of reference values. Samples showing unique spectral features in comparison to the entire calibration set should be treated with care.

When selecting the calibration samples, the concentrations of the constituents of interest are another point for consideration. The range of variability should be as

large as that expected for future samples. The distribution in future samples will, in many cases, be *Gaussian* with a concentration in the centre of the range and fewer values near the limits. Instead of taking such a distribution for the calibration data set, it is considered better practice to aim for a more uniform spread of values over the whole range. This has the advantage of providing more precise estimates of the calibration constants for a given number of samples. This is likely to be beneficial in these cases, where the "unusual" samples are the ones that actually matter [56].

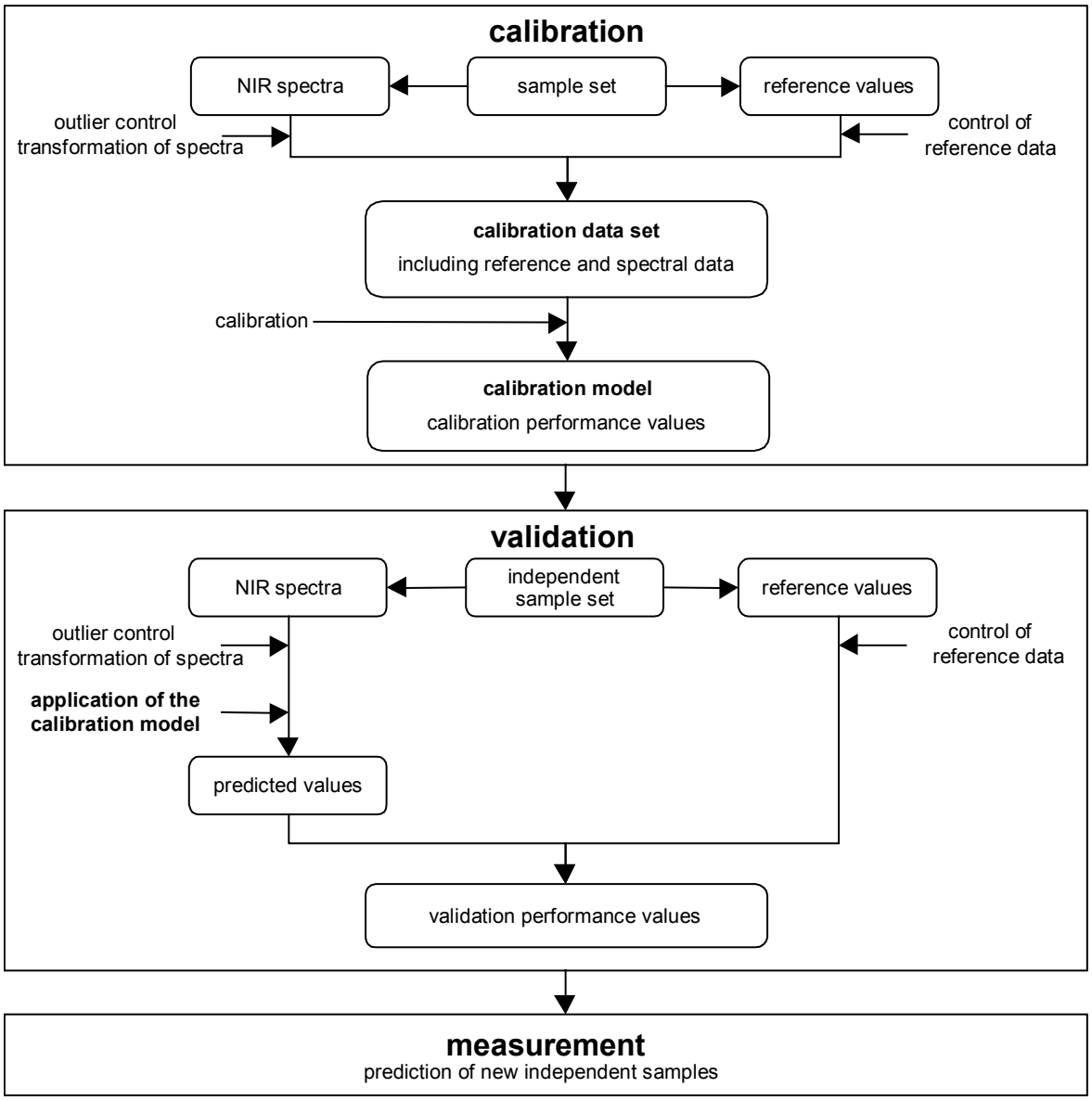

Figure 2: Steps for development of an NIRS method for quantitative analysis of a food constituent.

Spectral data may be subjected to calibration either directly or after transformation. Transformations aim to standardise NIR spectra by compensating particle size effects and reducing high correlations between spectral data at different wavelengths. This

may lead to simpler calibrations which can be interpreted more easily. As an effect of differences in particle size, spectra are shifted and rotated relatively to each other. The calculation of derivatives for standardisation is attractive because it substantially reduces particle size effects as well as high correlations between spectral data at different wavelengths. In combination with a smoothing procedure, transformed spectra are eventually obtained showing peaks rather than broad bands as observed in the original spectra. A simple standardisation method is to divide spectral data at each wavelength by a value measured at a reference wavelength.

A variety of calibration methods are in use to establish a correlation between spectral data and reference values. Relatively robust methodologies, like multiple linear regression (MLR), and full-spectrum methods, *e.g.* principle component regression (PCR) or partial least squares algorithms (PLS), can be applied to find those wavelengths whose inclusion in the calibration model results in the best prediction for the constituent of interest on the basis of spectral data.

The use of MLR for calibration results in a prediction equation as shown in equation 2.

equation 2:

$$c = a_0 + a_1 * L(\lambda_1) + a_2 * L(\lambda_2) + \ldots + a_n * L(\lambda_n)$$

c:	predicted concentration of the constituent
$L(\lambda)$:	spectral data at wavelength λ
a_0 - a_n:	regression coefficients
$a * L(\lambda)$:	term

In order to find the regression coefficients a_0 - a_n, forward stepwise regression is the most popular choice. Following this principle, the equation is built-up by consecutive addition of wavelengths until no further improvement of prediction accuracy is achieved. However, often the fit improves considerably until two or three terms have been included and then improves only slightly when further wavelengths are added. Since equations with the fewest terms are the most robust ones, it is recommended to choose a simple model rather than equations with a high number of terms.

In contrast to MLR, full-spectrum methods use the entire information in the spectra for model development. In order to reduce the quantity of data, factors are calculated from the original spectra without discarding any useful information. Factors are constructed as linear combinations of the original spectral data and

multiple linear regression on the factor scores is used to derive a prediction equation. The important difference between PCR and PLS is, that PCR constructs factors without taking into account whether they correlate with the reference data, whereas PLS constructs its factors balancing this desirable property and the aim to cover as much of the spectral variation as possible [56].

When wavelengths, number of terms and the form of the prediction equation have been chosen, the reference values of an independent sample set should be compared to those predicted on the basis of spectral data by means of the calibration equation (validation). This is done to make sure, that the model is of universal validity and does not exploit special features of the calibration data set. The criteria for choosing samples to make up a validation sample set are basically the same as those for establishing a calibration set.

On the basis of a successful validation, the developed NIRS method can be applied to predict the concentrations for the constituent of interest in new samples purely on the basis of their spectral data. Care should be taken in order to make sure, that the calibration sample set of the underlying NIRS model is still representative for the new samples to be predicted. If this is no longer the case, the calibration model should be refined by expanding the calibration sample set with a sub-set of the new samples. This requires analysis of the sub-set by the reference method and subsequently re-calculation of the prediction equation.

2.3.6 Assessment of an NIRS method

In order to assess the performance of a NIRS method, several statistical parameters are calculated which all describe the discrepancies between the specified reference values and those predicted by the NIRS model. The root mean squared error of calibration (RMSEC, often called "standard error of calibration", SEC) is calculated according to equation 3 for the calibration sample set. In an analogous way, the root mean squared error of prediction (RMSEP) is calculated for the sample set used for validation.

Optimisation of a NIRS calibration model may aim to decrease the RMSEC. However, applying this parameter alone would give a rather optimistic estimate of the model performance and provides no protection against an overfit to accidental features of the calibration sample set. Therefore a leave-one-out strategy may be used to obtain

a more realistic evaluation of the calibration model. Following this strategy, the concentration of each calibration sample is predicted on the basis of a model established without the sample to be predicted. Using this approach for assessment of an MLR calibration model, the estimated root mean squared error of prediction ($RMSEP_{Est}$) and the root mean squared error of crossvalidation (RMSECV) are calculated according to equation 3 and Table 1.

The RMSECV is considered as a very suitable statistic parameter for optimum calibration model selection [59].

equation 3:

$$\left.\begin{matrix} RMSEC \\ RMSEP_{Est} \\ RMSECV \\ RMSEP \end{matrix}\right\} = \sqrt{\frac{\sum (c_{specified} - c_{predicted})^2}{m}}$$

$RMSEC$: root mean squared error of calibration
$RMSEP_{Est}$: estimated root mean squared error of prediction
$RMSECV$: root mean squared error of crossvalidation
$RMSEP$: root mean squared error of prediction
$c_{specified}$: concentration determined by the reference method
$c_{predicted}$: concentration predicted by means of the NIRS model
m: degrees of freedom

Table 1: Procedures for calculation of NIRS performance values (MLR model)

parameter	sample set	leave-one-out*	recalculation of regression parameters	reselection of wavelengths	degrees of freedom
RMSEC	calibration	no	no	no	$n-R-1$
$RMSEP_{Est}$	calibration	yes	yes	no	n
RMSECV	calibration	yes	yes	yes	n
RMSEP	validation	no	no	no	n

n : number of samples in the calibration/validation sample set
R : number of terms in the MLR calibration model (see equation 2)
* : A leave-one-out strategy means to successively remove each individual sample from the calibration set and to analyse them on the basis of models built with the remaining samples.

The above mentioned performance values are absolute errors calculated from the differences between predicted and specified values. They are suitable to compare NIRS models. In order to assess the applicability of an NIRS method to the analysis of future samples and to replace the reference method, relative error measurements have been proposed taking into account the standard deviation or the range of concentrations observed for the samples, and the standard deviation of the reference method. Minimum performance criteria, available from the literature, have to be met

to assure that absolute prediction errors are small enough compared to the sample concentration range. Table 2 summarises performance criteria proposed in the literature as general minimum requirements for NIRS methods.

Table 2: Performance criteria proposed for NIRS methods

performance criterion			minimum requirement	literature
$\frac{(max - min)}{RMSECV}$	max:	maximum concentration observed in the calibration sample set	10	[60, 61]
	min:	minimum concentration observed in the calibration sample set		
$\frac{s}{RMSECV}$	s:	standard deviation of concentrations in the calibration sample set	3	[60]
$RMSEP_{NIRS} = k * s_{reference}$	$s_{reference}$:	standard deviation of the reference method	$k \leq 2$	[61]

Because NIRS is a secondary method, any error in the reference method is carried over into the NIRS method. Therefore, results obtained by NIRS cannot be more accurate than those obtained by the reference procedure. If systematic errors can be ruled out, the limit of accuracy can be estimated according to equation 4 [56].

equation 4:

$$RMSEP_{NIRS} > \sqrt{(s^2_{specified} + s^2_{predicted})}$$

$RMSEP_{NIRS}$: root mean squared error of prediction of the NIRS method

$s_{predicted}$: standard deviation of the NIRS method

$s_{specified}$: standard deviation of the reference method

2.3.7 Applications of NIRS in food analysis

NIRS is being widely used for quality control of raw materials, intermediate products and final products. NIRS is a non-destructive, simple and rapid analytical technique and represents an interesting alternative to time-consuming wet-chemical methods. Proximate analysis (*i.e.* determination of moisture, protein, fat, starch) still accounts for the majority of applications. Generally, NIRS is applicable to the determination of major constituents (contents > 1 %), but a few examples focusing on minor constituents, *e.g.* amino acids in cereals, soybeans and sunflower and glucosinolates in canola have been reported [61-63].

Wet-chemical analysis of amino acids is a time-consuming process involving extraction, hydrolysis, clean-up, and ion-exchange chromatography. The contents of individual amino acids in foods and feeds are often lower than 1 %. Nevertheless, NIRS calibrations were developed allowing the accurate, simultaneous and fast prediction of protein and the total contents of methionine, cystine, lysine, threonine, tryptophan, and other essential amino acids in a broad spectrum of cereals and feedstuffs. This reduced the time for actual analysis from hours to seconds. Especially for lysine and methionine, the most limiting amino acids, NIRS was shown to be superior to other methods for estimation of amino acid distribution, *e.g.* crude protein regression [61, 62].

One of the major objectives in canola breeding was to obtain cultivars low in glucosinolates. Although the contents of these anti-nutritional constituents in canola are between 0.2 % and 1.4 %, NIRS was shown to be applicable to the rapid determination of these compounds in intact rapeseeds [63].

One of the most interesting applications of NIRS in the industry is the evaluation of complex parameters related to food quality, such as taste, status of processing, or nutritional value.

Taste - related formulae that derive a taste score from the constituents have been developed. However, it is impractical for quality control to use the results of time-consuming chemical analysis. To overcome this problem, rice taste analysers based on NIRS have been developed. They are based on the fact that rice taste could be experimentally related to the balance of moisture, protein, amylose and fatty acids [55].

Whiteness, affected by the degree of milling, is related to the nutritional value of rice. It has been reported, that whiteness can be measured by NIRS. However, it does not seem to be measured directly by NIRS, but rather by determining the amount of bran remaining on the surface of the milled rice kernel. NIRS determination of the surface lipid content of milled rice has been investigated because there was a strong correlation with the degree of milling [56].

These few examples demonstrate the usefulness of NIRS for the rapid determination of a broad spectrum of parameters including those affected by more than one constituent. However, in many cases the contents of nutritionally important or toxicologically relevant constituents are close to or even below the reported detection

limit (0.5 - 1 %) of NIRS [64]. Applicability of NIRS to the determination of such minor constituents has to be proven case-by-case. The efforts might be recompensed by gaining an analytical method which is less laborious, more rapid, more environmentally friendly, and cheaper than the classical wet - chemical approach.

2.4 Metabolite profiling of foods

2.4.1 Definitions

Metabolites are the end products of cellular regulatory processes, and their levels can be regarded as the ultimate response of biological systems to genetic or environmental changes. In analogy to the terms transcriptome and proteome, the entire set of metabolites synthesised by a biological system has been named *metabolome* [65, 66]. *Metabolomics* is the science to measure the metabolome and aims to provide a comprehensive picture of the metabolites synthesised by biological systems. It tries to avoid biases against certain compound classes by chemical structure or by apparent abundance in the biological tissue [65].
Several methodologies have been proposed for metabolomics studies. Some are based on a *fingerprinting* approach where the metabolome creates a characteristic pattern that becomes itself the object of analysis [67]. *Metabolite profiling* aspires to detect, resolve, identify and quantify as many individual compounds as possible across all compound classes in an unbiased and non-targeted way [65].

Metabolite profiling applied to food analysis in terms of that definition is in an early stage of development. The objective of this chapter is to briefly cover recent applications of metabolite profiling with special emphasis on those with potential usefulness within the framework of food safety assessment.

2.4.2 Sample preparation

Sample preparation for metabolite profiling aims to efficiently extract a wide range of compounds across the different chemical classes from biological tissues. As the first step, the inherent enzymatic activity of biological samples has to be rapidly stopped. Usually, freezing in liquid nitrogen is used to achieve this effect. Other methods include lyophilisation which prevents both enzyme and transporter function or immediate addition of organic solvents and application of heat [68]. Methanol, ethanol, and water or mixtures thereof are typically used for extraction of polar metabolites [69, 70]. Methods commonly applied for total lipid extraction are based on the use of chloroform and chloroform/methanol mixtures [71, 72]. In order to enhance extraction efficiency, additional energy is put into the system, either by heat

[73] or by other techniques, such as accelerated solvent extraction [74]. Owing to the resulting co-extraction of high-molecular weight compounds, *e.g.* carbohydrates and proteins, the additional use of ultrafiltration or precipitation techniques has been reported [72, 75]. A few methodologies have been described which allow the fractionation of the total extracts obtained. They are either based on liquid-liquid distribution to separate polar and lipophilic metabolites [73] or on solid-phase extraction for separations according to the degree of polarity [72].

2.4.3 Data acquisition and analysis

The methodologies applied for analysis of the complex extracts obtained must be as comprehensive as possible and have to offer the possibility to detect, resolve, identify, and quantify a broad spectrum of compounds in a wide range of concentrations. The methods used for metabolite profiling usually combine a separation step, such as chromatography, with a universal detection technique. Gas chromatography provides high resolution separation of compounds and can be used in conjunction with a mass spectrometer (GC-MS). Both analytical techniques are highly sensitive and universal, and are capable of detecting almost any organic compound, regardless of its class or structure. Most metabolites found in plant extracts have to be converted to less polar, more volatile derivatives before they are applied to the GC column [73].

So far, GC combined with quadrupole-MS [76, 77] or time-of-flight MS [78] has been the most successful technique applied for metabolite profiling. However, the use of other measurement techniques based on coupling liquid chromatography to UV-diode array detection [54], mass spectrometry [75], and nuclear magnetic resonance [72, 79] for analysis of complex extracts of plant and environmental origin has been reported. In these cases, no derivatisation step was necessary.

In order to identify the metabolites detected, their chemical-analytical characteristics may be compared to those of reference compounds. Mass spectra libraries are available which may assist in identifying analytes on the basis of electron-impact (EI) mass spectral data obtained by GC-MS analysis [80]. Recently, reference tables comprising retention times and EI-MS data of several hundred target compounds found in potato tubers and *Arabidopsis thaliana* leaves have been published in the

public-domain [81]. Fourier-transform ion cyclotron resonance mass spectrometry, a recently developed type of ultra-high resolution mass spectrometry, allows non-targeted identification and quantification of metabolites based only on their accurate mass [82].

Metabolite profiling approaches must allow the analysis of raw data files in an unbiased way. Such files generally contain information in two or three dimensions, *i.e*. chromatographic retention time, characteristics of the physical properties of the molecules, *e.g*. mass spectra and intensities to quantify metabolite levels. Mass spectral deconvolution and identification software capable of computing purified mass spectra from the elution profile of a compound by deconvolution of the overlapping mass spectra of its neighbouring compounds has been developed. Using the deconvoluted mass spectra, peak identities are confirmed by searching mass spectral libraries [83].

Detection of relative changes in metabolite levels rather than quantification of individual compounds has been proposed as primary objective of metabolite profiling studies [68]. Multivariate data analysis, *e.g*. Principal Component Analysis (PCA) [77] and Hierarchical Cluster Analysis [84, 85] has been frequently used to group samples according to metabolite distribution. In order to find statistically significant differences between individual metabolite levels, univariate statistics, *e.g*. Student's *t* test, can be applied.

2.4.4 Recent applications

Up to the beginning of the 1990s, metabolite profiling approaches were most advanced in the area of medicine particularly in diagnostic screening for inborn errors of metabolism. In pioneering studies [86], multicomponent analyses were carried out for steroids, acids, drugs and drug metabolites. It was suggested to use the "metabolic profiles" obtained for studying drug metabolism and for evaluating the influence of drugs on metabolic pathways. More recently, GC-MS was applied to the analysis of a total of 126 urinary metabolites in order to diagnose 25 types of organic acidemia in clinical routine [76]. The earliest application of GC to plant metabolite profiling employed GC-FID/MS for analysis of barley leaves following treatment with various herbicides. The detection of changes in peak profiles following herbicide application was considered as valuable tool to locate the biochemical site of action of

novel herbicides [70]. In the late 1990s, these studies triggered the proposal to apply metabolite profiling methodologies for high-throughput large scale approaches in the field of plant functional genomics [87].

Off-line coupling of solid phase extraction to NMR was applied to the analysis of GM tomato varieties, with slow ripening characteristics through antisense RNA exogalactanase modification, and to their non-modified counterpart. ^{1}H-NMR spectra of pre-fractionated extracts revealed that α-lycopene was present in the antisense fruit at a concentration of two to four times the levels found in its parental line. This was explained by a correlation between α-lycopene levels and fruit ripening [88].

A comprehensive HPLC profiling procedure involving UV-diode array detection for plant isoprenoids has been described with applications to tomato and *Arabidopsis* [54]. It has been shown that hydrophilic interaction columns may be coupled to electrospray mass spectrometers, for structure determination (by MS/MS) and quantification of polar metabolites of plant origin [75].

The most comprehensive procedure for metabolomics based on GC-MS measurement has been described and validated using potato tubers [73] and *Arabidopsis* leaves [77]. Over 300 compounds have been quantified in *Arabidopsis* extracts (polar and non-polar fractions) but only half of them corresponded to known structures. PCA of GC-MS metabolic data has been used to demonstrate clustering of *Arabidopsis* genotypes [77]. The method was also used to characterise genotypes of potatoes modified in sucrose metabolism. The approach revealed the appearance of novel unexpected metabolites in chromatograms from transgenic tubers [84] which had not been detected before because they had not been among the targets.

Considering the recent advances of metabolite profiling methodologies in plant functional genomics, the non-targeted screening of a broad spectrum of low-molecular weight compounds has been proposed as tool for comparative compositional analysis within the safety assessment framework of GM foods [6].

3 Materials and Methods

3.1 Materials

3.1.1 Rice

3.1.1.1 Bt-rice

Rough rice grains from a genetically modified insect-resistant rice line expressing a synthetic *cry1Ab* gene from *Bacillus thuringiensis* (KMD1) and from the corresponding parental line (Xiushui 11) were delivered by Qingyao Shu, Zhejiang University, Hangzhou, China [89]. Rough rice material was obtained from GM (sample codes KMD1 (1), KMD1 (2)) and parental lines (sample codes Xiushui 11 (1), Xiushui 11 (2)) grown in two field trials in China in 2001 under identical environmental conditions.

An additional sample set consisting of brown rice kernels from rice lines Xiushui 11, KMD1 and KMD2 (sample codes X11, K11, K21) was a gift from Illimar Altosaar, University of Ottawa, Canada. KMD2 is another Bt-rice line obtained in the same way as KMD1 but by a separate transformation [90]. The three samples had been treated in the course of analysis by crushing using a Single Kernel Characterisation System (Perten, Huddinge, Sweden) and had not been stored under defined conditions.

3.1.1.2 GNA rice

Brown rice grains from a genetically modified insect-resistant rice line expressing a lectin gene from *Galanthus nivalis* (GNA) and from the corresponding parental line (ASD16) were delivered by Angharad Gatehouse, University of Newcastle, United Kingdom [91]. Brown rice material was obtained from the GM (sample code GNA) and from the parental line (sample code ASD16) grown site by site in a glasshouse trial in 2001.

3.1.1.3 Conventional rice cultivars

Rough rice and brown rice samples from the conventional cultivars *Balilla*, *Cripto*, and *Thaibonnet* were obtained from Primavera, Mühldorf, Germany. The samples were

delivered as separate batches. Their characteristics regarding origin, grain shape and year of harvest are summarised in Table 3.

Table 3: Conventionally bred rice cultivars used as study material

sample code	cultivar	grain shape*	origin	year of harvest
BA1	*Balilla*	SG	Spain	2000
BA2		SG	Italy	2000
BA3		SG	Italy	2000
BA4		SG	Italy	2000
CR1	*Cripto*	SG	France	2000
CR2		SG	Italy	2000
CR3		SG	Italy	2001
CR4		SG	Italy	2001
CR5		SG	Italy	2001
CR6		SG	Italy	2001
CR7		SG	Italy	2000
CR8		SG	Italy	2000
Thai1	*Thaibonnet*	LG	Italy	2000
Thai2		LG	Italy	2000

*: LG: long grain rice; SG: short grain rice

3.1.1.4 Commercial rice samples

Commercial brown rice samples (not parboiled) were purchased in local supermarkets (sample codes H1 – H16) or were provided by the following companies: Bayerische Reismühle, Halfing, Germany (B – B2), Rickmers Reismühle, Bremen, Germany (R1), MSC, Hamburg, Germany (E1 – E18), Primavera, Mühldorf, Germany (P1 - P8), and Huber Mühle, Mannheim, Germany (U1 – U3). Samples with sample codes J1 - J6 were a gift from M. Murata, National Food Research Institute, Tsukuba Science City, Japan. Characteristics of these commercial brown rice samples are summarised in Table 4.

Table 4: Characteristics of commercial rice samples used as study material

sample code	grain characteristics*	origin	sample code	grain characteristics*	origin
B	SG	Italy	R1	SG	USA
B1	SG	Italy	U1	LG	Italy
B2	LG	Italy	U2	LG, B	India
			U3	LG	Surinam
E1	LG	USA	H1	LG	France
E2	LG, B	India	H2	SG	France
E3	LG	Spain	H3	SG	Italy
E4	LG	Guyana	H4	LG, B	Italy
E5	LG	Spain	H5	MG	Italy
E6	LG, B	India	H6	SG	Italy
E7	LG	USA	H7	LG	Spain
E8	LG, B	India	H8	LG	unknown
E9	LG	USA	H9	SG	unknown
E10	LG	Guyana	H10	LG	unknown
E11	LG, B	India	H11	LG	Italy
E12	LG	USA	H12	LG	France
E13	LG, B	India	H13	LG	unknown
E14	LG	Guyana	H14	LG	unknown
E15	LG	Spain	H15	LG	unknown
E16	LG, B	India	H16	RC	France
E17	LG	USA			
E18	LG	Guyana			
P1	SG	Spain	J1	SG	Japan
P2	SG	France	J2	SG	Japan
P3	SG	France	J3	SG	Japan
P4	SG	Italy	J4	SG	Japan
P5	SG	Italy	J5	SG	Japan
P6	SG	Italy	J6	SG	Japan
P7	SG	Italy			
P8	SG	Italy			

*: LG: long grain rice; MG: medium grain rice; SG: short grain rice; B: Basmati rice; RC: red camargue rice;

3.1.2 Maize

Maize grains from a genetically modified insect-resistant maize breeding line expressing a synthetic *cry1Ab* gene from *Bacillus thuringiensis* (Navares) and from the corresponding parental line (Antares) were delivered by Joachim Eder, Bayerische Landesanstalt für Bodenkultur und Pflanzenbau, Freising, Germany. Maize material was obtained from the GM and the parental line grown in a field trial in Neuhof, Germany in 2000 under identical environmental conditions [92].

3.1.3 Barley

Grains of malting barley (cultivar Pasadena) were a gift from Lehrstuhl für Technologie der Brauerei I, Technical University Munich, Freising-Weihenstephan, Germany.

3.2 Methods

3.2.1 Metabolite profiling of major and minor compounds in rice grains

3.2.1.1 Preparation of standard solutions

Internal standard solution for fraction I was prepared by adding *n*-hexane (Merck, Darmstadt, Germany) solutions of undecane (1.5 mL, 1 mg/mL), hexadecane (2.5 mL, 1 mg/mL), tetracosane (4 mL, 1 mg/mL), and triacontane (4 mL, 1 mg/mL) to 10 mg of octatriacontane. Hydrocarbons were purchased from Fluka, Buchs, Switzerland.

Internal standard solution for fraction II was prepared by bringing 1 mL of a stock solution of 5α-cholestan-3β-ol (Fluka) in dichloromethane (2 mg/mL) to a final volume of 10 mL with dichloromethane.

Internal standard solution for fraction III was obtained by dissolving 20 mg of phenyl-β-D-glucopyranoside (Fluka) in 50 mL of distilled water.

Internal standard solution for fraction IV was prepared by dissolving 10 mg of p-chloro-L-phenylalanine (Fluka) in 50 mL of distilled water.

Retention time standard mix I was identical to the internal standard solution for fraction I.

Retention time standard mix II was prepared by adding 1.5 mL of *n*-hexane and *n*-hexane solutions of hexadecane (2.5 mL, 1 mg/mL), tetracosane (4 mL, 1 mg/mL), and triacontane (4 mL, 1 mg/mL) to 10 mg of octatriacontane.

3.2.1.2 Sample preparation

Rough rice was dehulled by means of a wooden rice dehuller, which was a gift from Quingyao Shu, Zhejiang University, Hangzhou, China. Dehulled rice grains (9 % - 12 % moisture) were ground with a cyclone mill (Cyclotec 1093, Foss, Höganäs, Sweden) equipped with a 500-µm screen. The flour was immediately freeze-dried for three days using a conventional freeze-drying apparatus (Delta IA, Christ, Osterode, Germany). The moisture content of the resulting material (<2 %) was determined as a loss of weight by drying at 130°C for two hours [93]. Freeze-dried flour samples were stored at -18°C in tightly closed LDPE bottles (Qualilab, Darmstadt, Germany).

3.2.1.3 Extraction procedure

Two hundred milligrams of freeze-dried rice flour were weighed into a disposable cartridge (3 mL column volume, Merck) which was sealed with plastic frits at the top and at the bottom of the flour layer. The cartridge was subsequently connected to a vacuum manifold (Supelco, Taufkirchen, Germany) equipped with valves to control the flow rate. The rice flour in the cartridge was soaked in 150 µL of methanol (Merck) for 20 min at ambient temperature. After removal of methanol by application of vacuum on top of the cartridge for 30 min, lipids were eluted with a total of 4 mL dichloromethane (Riedel de Haën, Seelze, Germany) within 20 min (lipid extract).

Traces of dichloromethane were removed from the flour under vacuum. Subsequently, polar compounds were eluted with a total of 10 mL methanol/water (80+20, v+v) within 40 minutes (polar extract).

Elutions from the cartridge proceeded by gravity flow or - if necessary - by application of weak vacuum at the bottom into 10-mL round flasks.

3.2.1.4 Fractionation and analysis of lipids

After addition of internal standard solutions for fraction I (50 µL) and fraction II (150 µL), the lipid extract was evaporated to dryness by rotary evaporation under vacuum and redissolved in 300 µL of dry methyl-*tert*-butylether (MTBE, Oxeno, Marl, Germany). To this solution, 200 µL of dry methanol and 50 µL of sodium methylate, 5.4M in methanol (Fluka), were added. After reaction for 90 min at room temperature in the dark, 1 mL of dichloromethane and 2 mL of aqueous 0.35M HCl were added. After vigorous shaking, the upper phase was discarded. The lower phase containing the transmethylated lipids was reextracted with another 2 mL of aqueous 0.35M HCl and subsequently evaporated to dryness by rotary evaporation under vacuum. After drying over phosphorus pentoxide (Fluka), transmethylated lipids were subjected to solid phase extraction (SPE).

A small amount of anhydrous sodium sulphate (Merck) was placed on top of a 500 mg silica gel SPE cartridge (LiChrolut, Merck) conditioned with one column volume (CV, 2.5 mL) of *n*-hexane. Transmethylated lipids were redissolved in 250 µL of dichloromethane and transferred to the SPE cartridge. The fatty acid methyl ester fraction (fraction I) was eluted with *n*-hexane/MTBE (100+2, v+v, 2 x 3 mL). Subsequently, the fraction of minor polar lipids (fraction II) was eluted with *n*-

hexane/MTBE (7+3, v+v, 2 CV). Elutions from the SPE cartridge proceeded by gravity flow into 10-mL round flasks. After addition of retention time standard mix I to fraction II, the eluates were evaporated to dryness by rotary evaporation under vacuum. The fatty acid methyl ester fraction was redissolved in 150 µL of dichloromethane, and 1 µL was injected into the gas chromatograph.

Fraction II containing polar lipids was redissolved in 100 µL of dry pyridine (Fluka) and 50 µL of *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide (MSTFA, Merck) was added. After flushing with argon, the flask was tightly sealed with a glass stopper and allowed to stand for 15 min at 70°C in an oil bath. The sample was then cooled to room temperature and 1 µL was analysed by gas chromatography.

3.2.1.5 Fractionation of the polar extract

After addition of the internal standard solutions for fraction III (150 µL) and fraction IV (150 µL), 2 mL of the polar extract was concentrated in a rotary film evaporator and dried over phosphorus pentoxide. Dry pyridine (100 µL) and 50 µL of trimethylsilylimidazole (TMSIM, Fluka) were added for silylation which was performed as described for fraction II. The silylated sample was diluted with 200 µL of *n*-hexane, and 200 µL of water was added for selective hydrolysis of silylated carboxyl groups. After shaking and phase separation, 100 µL of the upper phase was mixed with 50 µL of retention time standard mix I. One microliter of this mixture was injected into the gas chromatograph.

The remaining polar extract (ca. 8 mL) was concentrated in a rotary film evaporator and dried over phosphorus pentoxide. The dry residue was redissolved in 100 µL of a solution of hydroxylammoniumchloride (Merck) in pyridine (2 mg/mL). After heating for 30 min at 70°C, 50 µL of MSTFA was added and the sample was allowed to stand for another 15 min at 70°C. The silylated sample was diluted with 400 µL of *n*-hexane and 200 µL of water was added for selective hydrolysis of silylated carboxyl and amino groups. After vortexing, the mixture was allowed to stand for 5 min. The upper phase containing oximated and silylated sugars was discarded and residual sugars were re-extracted from the aqueous phase by additional treatment with 2 x 400 µL of *n*-hexane. The lower aqueous phase containing free organic and amino acids was concentrated in a rotary film evaporator and dried over phosphorus pentoxide. The residue was redissolved in 100 µL of dry acetonitrile (Merck) and

50 µL of retention time standard mix II were added. After addition of 50 µL of MSTFA, the sample was allowed to stand for 60 min at 70°C in an oil bath for silylation. It was then cooled to room temperature and 1 µL was analysed by gas chromatography.

3.2.1.6 Gas chromatography

Gas chromatography (GC) was performed on a Carlo Erba Mega HRGC 5160 (Thermo Finnigan, Austin, TX) equipped with a flame ionisation detector (FID) which was heated to 320°C. The chromatography column was a DB-1, 60 m x 0.32 mm i.d. fused silica capillary coated with a 0.25 µm film of polydimethylsiloxane (J&W Scientific, Folsom, CA). Split injection (split flow 30 mL/min) was performed at 280°C, and the column temperature was programmed from 100°C to 320°C (30 min-hold) at 4°C/min. Hydrogen was used as the carrier gas at a constant inlet pressure of 100 kPa.

FID data were acquired by means of commercially available software (ChromCard, Thermo Finnigan) using a sampling rate of 600 data points/min. Automated peak integration within the retention time range between 5 min and 50 min was performed using a peak width of 5 sec and a peak threshold of 5. For integration of peaks eluting at retention times higher than 50 min, peak threshold was set to 9, and peak width was adjusted to 15 sec. Peaks exhibiting an area lower than 3000 µV*min were not taken into account.

Quantification was performed using peak heights which had been normalised on the basis of those observed for internal standards in the respective fractions.

3.2.1.7 Gas chromatography – mass spectrometry

Gas chromatography – mass spectrometry (GC-MS) was performed on a GC 8000 Top with a Finnigan Voyager (Thermo Finnigan) as the mass selective detector. The MS interface temperature was set to 260°C. Full scan mass spectra were recorded at an electron energy of 70 eV (source temperature 200°C) within a scan range of 40-700 mu. The chromatography column was the same as used for analysis by GC-FID. Split injection (split flow 30 mL/min) was performed at 250°C, and the column temperature was programmed as described above. Helium was used as the carrier gas at a constant inlet pressure of 165 kPa.

Rice constituents were identified by comparing retention times with those of silylated and methylated reference compounds or by comparing mass spectra with entries of mass spectra libraries [80, 81].

3.2.1.8 Repeatability

To determine the repeatability (intralaboratory precision) of the fractionation procedure, five aliquots of a total extract were subjected to the described protocol and the resulting fractions were analysed by GC-FID. Repeatability of the overall metabolite profiling methodology was determined by triplicate analysis of a rice sample comprising extraction, fractionation, derivatisation, and GC-FID investigation.

3.2.1.9 Linearity

In order to determine the linear range of the overall metabolite profiling methodology, increasing amounts of rice flour (100 mg – 800 mg) were extracted, fractionated and derivatised as described above. Normalised peak heights obtained by GC-FID analysis were plotted against the amount of flour extracted. For quantitative evaluation of linearity, the correlation coefficient between normalised peak heights and amounts of extracted flour was calculated.

3.2.1.10 Recovery

For determination of recoveries, rough rice material was analysed in triplicate. Solutions containing standard compounds were spiked to the first sample at the start of the analytical procedure. The peak heights determined for the spiked compounds (h_1) were compared to those (h_2) observed for the rice fraction, which had been obtained from the second sample and spiked at the end of analysis. Spiked amounts of naturally occurring compounds were in the same ranges as those found to be endogenous in rice. In order to be able to calculate the percentages of spiked amounts recovered at the end of analysis (recovery w), the peak heights (h_3) determined by investigation of the unspiked fraction obtained from the third rice sample were taken into account. Experiments for determination of recoveries were performed in duplicate (6 samples). Recoveries were calculated according to equation 5.

equation 5:

$$w = \frac{h_1 - h_3}{h_2 - h_3} * 100$$

w: recovery [%]

h_1: peak height determined in the fraction obtained from a rice sample spiked at the start of analysis

h_2: peak height determined in the rice fraction spiked at the end of analysis

h_3: peak height determined in the unspiked fraction obtained from an unspiked rice sample

The protocols applied for the four fractions were as follows:

Fraction I : Twenty microliters of dichloromethane solutions containing tetracosane (0.83 mg/mL), squalene (Acros, Geel, Belgium, 0.25 mg/mL), and tripalmitin (Sigma, Steinheim, Germany, 8.5 mg/mL) were spiked to 200 mg of rice flour at the start of the analytical procedure. In order to add standard compounds at the end of analysis, 20 µL of dichlormethane solutions containing tetracosane (0.83 mg/mL), squalene (0.25 mg/mL), and methyl palmitate (8.5 mg/mL) were spiked to fraction I prior to GC analysis.

Fraction II : Twenty microliters of dichloromethane solutions containing 5α-cholestan-3β-ol (1.5 mg/mL), δ-tocopherol (Sigma, 0.25 mg/mL), stigmasterol (0.45 mg/mL), and palmitic acid (Frey und Lau, Henstedt-Ulzburg, Germany, 3 mg/mL) were spiked to rice flour and to fraction II prior to silylation, respectively.

Fraction III : Ten milligrams of solid sucrose and 50 µL of an aqueous stock solution of fructose, glucose, raffinose, and phenyl-β-D-glucopyranoside (5 mg/mL) were added to rice flour prior to lipid extraction. In order to spike standard compounds at the end of analysis, silylated derivatives obtained from 10 µL stock solution and 2 mg sucrose were added to fraction III prior to GC investigation.

Fraction IV : Fifty microliters of an aqueous solution containing glycine (Merck, 1.1 mg/mL), glutamic acid (Merck, 1.1 mg/mL), glutamine (Merck, 0.4 mg/mL), lysine hydrochloride (Merck, 1.5 mg/mL), calcium citrate (Merck, 1.15 mg/mL) and p-chloro-L-phenylalanine (0.4 mg/mL) were spiked to the polar rice extract and 40 µL of the same solution to fraction IV prior to re-silylation.

3.2.1.11 Determination of natural variability

For calculation of the natural variability (S) for a compound in a set of samples, normalised peak heights (h) were related to their median (Med). According to equation 6, ratios calculated for the highest and the lowest normalised peak height observed for a compound in the sample set were summed to obtain S.

equation 6:

$$X_+ = (\frac{h}{Med} - 1) * 100 \quad (h > Med)$$

$$X_- = (\frac{Med}{h} - 1) * 100 \quad (h < Med)$$

$$S = max(X_+) + max(X_-)$$

h: normalised peak height

Med: median of normalised peak heights for a compound in a set of samples

$max(X_+)$: maximum value observed for X_+ in the sample set [%]

$max(X_-)$: maximum value observed for X_- in the sample set [%]

S: natural variability

3.2.2 Metabolite profiling of maize grain constituents

Sample preparation and extraction of low molecular weight compounds from freeze-dried maize flour were performed as described in 3.2.1.2 and 3.2.1.3. In contrast to rice, fractions III and IV were prepared starting from 1 mL of polar extract. All further fractionation steps, derivatisation and gas chromatographic/mass spectrometric analysis were carried out according to the procedures described in 3.2.1.4 - 3.2.1.7.

3.2.3 Metabolite profiling of germinating barley

3.2.3.1 Micromalting of barley grains

In order to obtain malted barley at different malting stages, barley grains (cultivar Pasadena) were subjected to a micromalting procedure according to a guideline published by "Mitteleuropäische Brautechnische Analysenkommission" (MEBAK) [94]. Micromalting was carried out at Technical University Munich, Institute of Brewing Technology I, Freising-Weihenstephan, Germany. Samples were taken 5 hours after the start of the micromalting process and then daily until day 7. An additional sample was taken after kilning.

3.2.3.2 Sample preparation

Grain samples at different malting stages were immediately frozen in liquid nitrogen and subsequently freeze-dried for 3 days using a commercial freeze-drying apparatus (Christ). Freeze-dried grains with shoots were ground with a cyclone mill (Cyclotec 1093, Foss) and the flour obtained was stored at –18°C in tightly sealed LDPE bottles (Qualilab).

3.2.3.3 Extraction, fractionation, and GC analysis of malted flour samples

Extraction of low molecular weight compounds from freeze-dried malted flour samples was performed as described in 3.2.1.2 and 3.2.1.3. In contrast to rice, fractions III and IV were prepared starting from 1 mL of polar extract. All further fractionation steps, derivatisation and gas chromatographic/mass spectrometric analysis were carried out according to the procedures described in 3.2.1.4 - 3.2.1.7.

3.2.4 Automated comparative analysis of metabolite profiling data

3.2.4.1 Sample preparation

Metabolite profiling chromatograms obtained from rice were used for method development. Extraction, fractionation and capillary gas chromatographic analysis were performed following the protocol described in 3.2.1. In order to obtain a spiked sample, palmitic acid (60 µg), δ-tocopherol (5 µg), 5α-cholestan-3β-ol (30 µg), and stigmasterol (9 µg) were added to 200 mg of rice flour prior to extraction. For retention time standardisation, 50 µL of a retention time standard mix (see 3.2.1.1) containing undecane (C_{11}), hexadecane (C_{16}), tetracosane (C_{24}), triacontane (C_{30}), and octatriacontane (C_{38}) was added to the rice fractions obtained from spiked and unspiked rice flour prior to GC-FID investigation.

3.2.4.2 Data acquisition and generation of files

Data acquisition (FID responses and corresponding time points) and automated peak integration were carried out as described in 3.2.1.6 using commercially available software (ChromCard, Thermo Finnigan, Austin, TX). Peak heights and corresponding retention times were stored in Microsoft (MS) Excel files. FID responses and corresponding time points were saved as ASCII files (100 time points per minute).

Chromatographic data obtained from the minor lipid fraction (fraction II) of the spiked and the unspiked rice flour served as sample chromatogram and reference chromatogram, respectively.

For further processing, *i.e.* matching and comparison of the sample to the reference chromatogram, the files were imported into a self-tailored software based on Microsoft Excel Visual Basic. The tools developed are described in general form in the following chapters.

3.2.4.3 Manual chromatogram matching

Manual chromatogram matching, *i.e.* correction of retention time differences and normalisation of detector responses, is performed by using a graphical user interface implemented in the software. After import of a sample and a reference chromatogram, *i.e.* detector responses and corresponding time points (optionally an additional list containing peak heights/areas and corresponding retention times), a baseline function is automatically calculated for the sample chromatogram. This is done by searching response minima within time intervals of the chromatogram. Subsequently, linear equations are computed describing the baseline shape between two adjacent minima.

For retention time correction, time intervals of the sample chromatogram can be shifted, linearly stretched and/or linearly compressed along the time axis until they match accurately to the reference chromatogram. Stretching and compressing is automatically restricted to time intervals exhibiting detector responses close to the baseline.

Detector response normalisation is achieved by automated baseline correction and by linear stretching/compressing of the sample chromatogram along the response axis using a pre-selectable peak as reference.

In order to be able to accurately control the matching efficiency, the software provides features for zooming into narrow time intervals of the chromatograms.

3.2.4.4 Automated retention time matching

In order to achieve automated retention time matching, input of the retention times observed for the five retention time standard compounds (C_{11}, C_{16}, C_{24}, C_{30}, C_{38}) in the sample chromatogram and in the reference chromatogram is required. By using

the software tool, retention times of all peaks in the sample chromatogram are corrected for the retention time difference observed for C_{24} (shift). Subsequently, two equations are calculated by means of quadratic regression describing a correlation between the remaining retention time differences observed for corresponding standards and the retention time distances to C_{24}. The two quadratic equations are subsequently applied to correct the retention times of all peaks in the sample chromatogram eluting earlier and later than C_{24}, respectively (stretching/compressing).

3.2.4.5 Automated standardisation of peak heights/peak areas

Standardisation of peak heights/peak areas is achieved on the basis of the peak height/peak area ratio calculated for a pre-selectable standard peak in the sample and the reference chromatogram.

3.2.4.6 Build-up of databases

Databases are built from peak-based and response-based datasets. Peak-based datasets consist of standardised heights/areas and corresponding corrected retention times of all peaks detected in a chromatogram and are obtained by either automated retention time matching or manual chromatogram matching. Response-based datasets consisting of normalised detector responses and corresponding corrected time points (100 time points per minute) are generated by manual chromatogram matching. Microsoft Excel spreadsheets are used to store the data.

3.2.4.7 Calculation of theoretical chromatographic data / chromatograms from databases

Calculation of theoretical chromatographic data from a series of pre-selected peak-based datasets is performed by combining the heights/areas of corresponding peaks for the calculation of minima, maxima, means and standard deviations. Peaks are combined if their corrected retention times are within the same retention time window. The width of the retention time window can be adjusted by setting the parameter "width".

Calculation of theoretical chromatograms from a series of pre-selected response-based datasets is performed by combining normalised detector responses at

corresponding corrected time points for the calculation of maxima, minima and means.

3.2.4.8 Generation of matrices from peak-based databases for multivariate analyses

This tool allows to transform peak-based data in a series of pre-selected datasets into a data format suitable for multivariate analyses, *e.g.* principal component analysis or discriminant analysis. The tool assigns exactly the same retention time to all corresponding peaks in the series of datasets. Peaks are considered to belong together if their corrected retention times are within the same retention time window. The width of the retention time window can be adjusted by setting the parameter "width".

The format of transformed data which are stored in MS Excel spreadsheets is illustrated in Figure 3.

	sample 1	*sample 2*	*...*	*sample n*
t_1	h_{11}	h_{21}	...	h_{n1}
t_2	h_{12}	h_{22}	...	h_{n2}
.	.	.	...	.
.	.	.	...	.
.	.	.	...	.
t_m	h_{1m}	h_{2m}	...	h_{nm}

Figure 3: General structure of matrices generated by transformation of a series of peak-based datasets for multivariate analyses. *t:* retention time, *h:* standardised peak height.

Principal component analysis using the generated matrix was performed by means of the XLSTAT add-in for MS Excel (Addinsoft, Paris, France).

3.2.4.9 Peak-based chromatogram comparison

The tailored tool developed for automated peak-based chromatogram comparison uses the standardised peak heights/areas at corresponding corrected retention times for analysis. Peak-based datasets as well as theoretical chromatographic data calculated from a series of pre-selected peak-based datasets can be defined as comparators (sample and reference). Comparison is performed peak-by-peak. Peak heights/areas are compared if their corresponding corrected retention times are within the same retention time window. The width of this retention time window can

be adjusted by setting the parameter "width". The parameter "min. peak height/area" defines a peak height/area threshold. Peaks exhibiting a height/area lower than the threshold value are not taken into account for comparison. The parameter "blanks" refers to retention time windows to be excluded from chromatogram comparison.

The magnitude of the percentage difference (w) between the normalised heights of corresponding peaks in the sample (h_S) and the reference chromatogram (h_R) was calculated according to equation 7.

equation 7:

$$w = \frac{h_S - h_R}{min(h_S; h_R)} * 100$$

w: magnitude of a percentage peak height difference [%]

h_S, h_R: normalised heights of corresponding peaks in the sample (S) and the reference (R) chromatogram

$min(h_S; h_R)$: expression is equal to h_S if $h_S < h_R$, and vice versa

Results of comparison are summarised in a graphical and a tabular report. The graphical report contains separate plots indicating percentage peak height/area differences and additional peaks as dots along the retention time axis. Differences falling below a threshold value (parameter "detect differences higher than") are not reported. A positively signed value expressing the magnitude of a percentage peak height/area difference indicates a higher peak height/area in the sample chromatogram in comparison to the counterpart in the reference and vice versa. If no counterpart is found in the reference, a dot appears in the graphic report indicating the positively signed height/area of the additional peak in the sample, and vice versa.

The tabular report lists the number and heights/areas of the additional peaks, the number and magnitudes of percentage peak height/area differences, the total number of peaks detected in the sample and the reference, and the number of peaks taken into account for comparison.

In order to detect statistically significant differences between two series of peak-based datasets, mean heights/areas and standard deviations are automatically calculated for all corresponding peaks. Subsequently, percentage differences

between the means are assessed for statistical significance by means of Student's t-test. Finally, graphical and tabular reports are generated which summarise number and magnitudes of the statistically significant differences.

3.2.4.10 Response-based chromatogram comparison

The tailored tool developed for automated response-based chromatogram comparison uses the normalised detector responses and the corresponding corrected time points for analysis. Two response-based datasets as well as theoretical chromatograms calculated from a series of pre-selected response-based datasets can be defined as comparators (sample and reference).

Selection of points to be included for comparison is based on the first derivative of the chromatograms and on the level of the normalised detector response. Points of the chromatograms exhibiting a slope lower than an adjustable parameter value (parameter "slope") are taken into account for comparison. In order to determine the first derivative of a chromatogram at each time point, mathematical functions describing its shape are calculated for adjacent narrow time intervals by means of Newton's polynomial interpolation.

The parameter "min. response" defines a minimum threshold for the response level. Ranges of the normalised chromatograms falling below this threshold (noisy ranges) are excluded from comparative analysis.

Percentage response differences are calculated in an analogous way as described for percentage peak height differences (see equation 7). Results of response-based chromatogram comparison are summarised in a graphical and a tabular report. The graphical report shows an overlay of the two comparators. Points of the chromatograms taken into account for comparison are labelled by empty circles. If a response difference has been detected, points are labelled by filled circles. Percentage response differences smaller than an adjustable parameter value (parameter "detect differences higher than") are not reported. The tabular report lists the number and magnitudes of percentage response differences. These values are positively signed if a higher response level has been detected in the sample chromatogram in comparison to the reference, and vice versa.

3.2.5 NIRS method for determination of phytic acid in brown rice

3.2.5.1 Sample pre-treatment and sample storage

Rice grains were dehulled, ground, freeze-dried and stored as described in 3.2.1.2.

3.2.5.2 Reference analytical determination of phytic acid in brown rice flour

One gram of a freeze-dried brown rice flour sample was weighed into a 50-mL oak-ridge centrifuge tube (Nalgene, Merck Eurolab, Darmstadt, Germany) and defatted over night with 3 mL of petrol ether (Merck, Darmstadt, Germany). After removal of the organic solvent by means of a pasteur pipette and by a gentle stream of air, phytic acid was extracted from defatted brown rice flour at ambient temperature for two hours with 20 mL of 2.4 % hydrochloric acid (pH = 0.6) under continuous shaking (Heto Master Shake, Merck Eurolab, Darmstadt, Germany). The aqueous flour extract obtained was centrifuged for 10 min (15,000 min^{-1}, 15°C) by means of a cooling centrifuge (ZK 401, Eppendorf, Hamburg, Germany), and the brown rice flour was re-extracted with 5 mL of water. The combined extracts were adjusted to pH = 6 with 1N and 0.2M sodium hydroxide solution (Merck, Darmstadt, Germany), and brought to a final volume of 50 mL water. In order to remove protein precipitates, the extract was centrifuged for another 10 min (15,000 min^{-1}, 15°C). The clear supernatant obtained was diluted with water (1:5) and subjected immediately to anion-exchange chromatography.

Anion-exchange chromatography was performed using disposable cartridges (3 mL), equipped with large-volume adapters and packed with 0.5 g resin (Dowex AG 1x4 200-400 mesh, Fluka, Neu-Ulm, Germany), which had been activated with 3 mL water (over night), 3 mL 1N sodium hydroxide solution (1.5 hours), and 3 mL 1N hydrochloric acid (1.5 hours) prior to use. Anion-exchange columns were washed with ca. 10 mL of water until eluates exhibited neutral pH. Re-activation of the resin was achieved by washing with 10 mL of 0.7N sodium chloride solution (1.5 mL/min), followed by 10 mL of water (1.0 mL/min). Anion-exchange columns were good for clean-up of a total of three flour extracts.

After a 10 mL aliquot of diluted brown rice extract had been loaded onto the activated anion-exchange column (1.0 mL/min), a fraction containing inorganic phosphate was eluted with 15 mL 0.1N sodium chloride solution (1.5 mL/min). The

phytic acid fraction was subsequently eluted with 15 mL 0.7N sodium chloride solution (1.5 mL/min) and collected in glass tubes, which had been specifically pre-treated in order to remove inorganic phosphate. This procedure had included consecutive rinsing with basic $KMnO_4$ solution, acidified H_2O_2 solution (3 %) and water. Afterwards, the tubes were cleaned with water only.

For derivatisation, a 3 mL aliquot of the purified phytic acid fraction was added immediately to 3 mL of 0.005N hydrochloric acid and mixed thoroughly with 1 mL of modified Wade solution which had been prepared by dissolving 0.030 g $FeCl_3 \times (H_2O)_6$ (Merck) and 0.300 g sulfosalicylic acid (Merck) in 100 mL of water. After 10 minutes, absorption at 500 nm was measured spectrophotometrically (Uvikon, Bio-Tek, Neufahrn, Germany) against water in disposable plastic cuvettes. The phytic acid content in the purified fraction, calculated as sodium phytate (c in µg/mL), was derived from the measured absorption on the basis of a linear calibration equation.

Each brown rice flour extract was analysed in triplicate and results were averaged for calculation of the phytic acid content in brown rice flour (w, g/100g) according to equation 8.

equation 8:

$$w = \frac{c * 50}{m * 10^6} * 5 * 1.5 * 0.725 * 100$$

w: phytic acid content in brown rice flour [%]

c: mean phytic acid content in three purified fractions, calculated as sodium phytate [µg/mL]

m: amount of rice flour extracted [g]

In order to establish a calibration equation, solutions of sodium phytate (Sigma, Taufkirchen, Germany) in water with increasing concentrations (15 µg/mL, 30 µg/mL, 45 µg/mL, 60 µg/mL, 75 µg/mL) were subjected to anion-exchange chromatography as described above. This resulted in calibration standards with the theoretically calculated sodium phytate concentrations 10 µg/mL, 20 µg/mL, 30 µg/mL, 40 µg/mL, and 50 µg/mL, which were analysed spectrophotometrically as described for the phytic acid fractions obtained from brown rice flour. Calibrations were performed daily. Purified calibration standards and purified phytic acid fractions obtained from rice were analysed in parallel.

Repeatability of the overall methodology was determined by four-fold investigation of a brown rice sample (one clean-up step per extract) on four days. Repeatability of the chromatographic clean-up procedure was calculated from results obtained by triplicate analysis of one rice extract.

Recovery of phytic acid from brown rice flour was determined in triplicate. Seven, ten, and thirty-seven milligrams of solid sodium phytate were spiked to one gram of a brown rice flour sample with known phytic acid content, respectively. This corresponds to an increase in phytic acid content by 0.54 %, 0.73 %, and 2.68 %, respectively. Recovery was calculated as ratio of the content determined in the spiked samples to that theoretically expected.

3.2.5.3 NIRS analysis of freeze-dried brown rice flour

NIRS analysis of freeze-dried brown rice flour was performed in diffuse reflection by means of a commercial Fourier-transform (FT) NIR spectrometer, equipped with an integrating sphere and an InGaAs detector (Spectrum One NTS, Perkin Elmer, Rodgau-Jügesheim, Germany). About 5 g of brown rice flour was placed into a petri dish which was subsequently mounted onto a continuously rotating measurement unit (NIRA sample spinner, Perkin Elmer, Rodgau-Jügesheim, Germany). Spectra were recorded from 1000 nm to 2500 nm in 0.5 nm steps (resolution: 8 cm^{-1}). One hundred spectra were averaged to obtain a mean NIR spectrum for each sample. Spectral data (background subtracted) were stored as log(1/R) values at corresponding wavelength points.

3.2.5.4 Pre-processing of spectral data

Pre-processing of raw spectral data (calculation of derivatives, multiple scatter correction, smoothing) was performed by means of commercially available software (Spectrum Quant, Spectrum Quant+, Perkin Elmer, Rodgau-Jügesheim, Germany). For export, pre-processed spectral data were saved as ASCII files.

3.2.5.5 Establishment of a calibration model

In order to establish a correlation between data obtained by NIRS and phytic acid contents determined by reference analysis, calibration datasets consisting of pre-processed spectral data and reference analytical data for phytic acid were imported

into software programmes for statistical analysis. Ridge regression was performed by means of MATLAB (The Mathworks GmbH, Unterföhring, Germany), calibration equations based on Principal Component Regression and Partial Least Squares algorithms were calculated using Spectrum Quant+ (Perkin Elmer, Rodgau-Jügesheim, Germany).

For forward stepwise Multiple Linear Regression (MLR), calibration datasets (5-point-smoothed, 25-point-first derivatives of spectral data and reference analytical data) were imported into Microsoft Excel spreadsheets and analysed by means of a self-tailored tool based on Microsoft Visual Basic Macros. Using this tool, imported data were averaged in 20-nm-segments. Additionally, ratios of averaged spectral data were calculated for all possible wavelength pairs. Subsequently, transformed data were subjected to regression analysis in order to calculate regression coefficients in calibration equations of the following form (equation 9):

equation 9:

$$c = a_0 + a_1 * L_1(\lambda) + a_2 * L_2(\lambda) + \ldots + a_n * L_3(\lambda)$$

c: phytic acid content of brown rice flour determined by reference analysis

$a_0,\ a_1, \ldots, a_n$: regression coefficients

$L_1(\lambda), L_2(\lambda), \ldots, L_n(\lambda)$: transformed spectral data or ratios of transformed spectral data

$a * L(\lambda)$: term

The calibration model was optimised by altering the number and the range of wavelength points to be taken into account for model establishment. Performance of calibration models was evaluated on the basis of the Root Mean Squared Error of Calibration (RMSEC), estimated Root Mean Squared Error of Prediction ($RMSEP_{Est}$), and Root Mean Squared Error of Crossvalidation (RMSECV), which were calculated as described in 2.3.6. For further optimisation, dependence of RMSEC, $RMSEP_{Est}$ and RMSECV on the number of terms in the calibration equation was investigated. The objective was to minimise root mean squared errors by using a minimum number of terms.

3.2.5.6 Validation of the calibration model

Validation of the optimised calibration model was performed using a series of independent datasets, which had not been taken into account for establishment of the calibration model. Phytic acid contents of validation samples were derived from pre-processed spectral data by applying the calibration equation, and the contents predicted were compared to those determined by means of reference analysis. For performance evaluation, the Root Mean Squared Error of Prediction (RMSEP) was calculated. Dependence of RMSEP on the number of terms was investigated to confirm appropriate selection of the number of terms in the model established.

3.2.5.7 Analysis of phytic acid in brown rice by means of NIRS

In order to determine the phytic acid contents in brown rice grains by means of NIRS, sample pre-treatment, acquisition of NIR spectra and pre-processing of spectral data were performed in the same way as described for calibration and validation samples. Subsequently, phytic acid contents were predicted on the basis of pre-processed spectral data by applying the calibration equation. Reproducibility of the NIRS method was determined by triplicate analysis of a brown rice flour sample.

4 Results and Discussion

4.1 Development of an NIRS method for determination of phytic acid in rice

4.1.1 Introduction

Phytic acid (*myo*-inositol [1,2,3,4,5,6] hexakisdihydrogenphosphate; IP_6) is a well-known naturally occurring anti-nutrient found in cereal grains, seeds and legumes (Figure 4). Formation of stable complexes with minerals and proteins results in decreases of their bioavailability and digestibility, respectively. Low bioavailability of iron and zinc from diets based on staple cereals and legumes that are high in phytic acid was shown to be a major factor for Fe and Zn deficiencies which are widespread in developing countries [95]. Numerous studies revealed that consumption of diets low in phytic acid may increase the absorption of these two trace minerals [96-98]. Other studies have shown that the phytic acid content of soy and wheat flour must be reduced by ≥ 90 % to achieve a two-fold increase in iron absorption [99]. Efforts have been undertaken to decrease the phytic acid content of food crops and to improve the bioavailability of minerals from IP_6 - containing diets. Using genetic screening, several low-phytic-acid mutants of maize and barley have been identified [100]. Lucca *et al.* reported on a combination of three genetic engineering approaches for improving the bioavailability and the level of iron in rice grains. This involved expression of a thermostable phytase from *Aspergillus fumigatus*, active over the pH range found in the gastrointestinal tract [101].

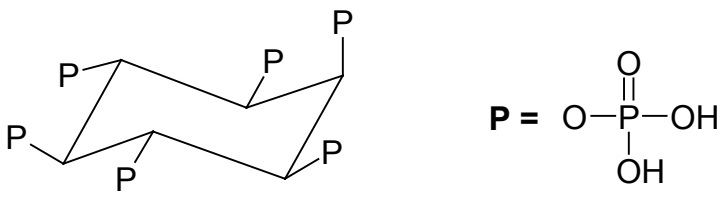

Figure 4: Structure of phytic acid [102]

Brown rice is relatively rich in phytic acid (content: ca. 1 %). Hunt *et al.* determined a zinc bioavailability of 77 % from brown rice. Milling led to a considerable decrease in phytic acid content (0.07 %) and to an improved zinc absorption (92 %) from the

resulting white rice grains [103]. Considering the role of rice as staple food and the anti-nutritional properties of phytic acid, it is important to screen rice cultivars used in human nutrition for their phytic acid contents. For the same reason, IP_6 should be taken into consideration in the course of a safety assessment of genetically modified rice.

Various analytical methods have been described in literature which allow the accurate determination of the phytic acid content in cereals [104]. However, most of them are laborious and time - consuming. Near infrared spectroscopy, a non-destructive technique, represents an interesting alternative to wet - chemical methodologies. NIRS is a widely used analytical technique which has been mainly applied to the analysis of major compounds, *e.g.* fat and protein [56]. So far, no NIRS method allowing the determination of phytic acid contents in rice grains has been published. Methods for NIRS analysis of phytic acid in cotton seed and feedstuffs have been reported [105, 106]. However, NIRS performance data were poor compared to those of NIRS methods for analysis of major food constituents [55, 56].

The objective of this study was to develop an NIRS method for analysis of the minor brown rice constituent phytic acid, which should be applicable to a broad spectrum of cultivars and commercial samples. In order to build an NIRS calibration model, brown rice samples of different grain shape and origin, grown in different seasons, were analysed for their phytic acid contents using a photometric reference method. NIR spectra were obtained from ground freeze-dried brown rice samples using a diffuse reflectance accessory. Subsequently, multivariate statistics were applied to establish a correlation between spectral data and results obtained by reference analysis. The calibration model was optimised to meet the challenge of the high sample variability which in general complicates the determination of a minor component via NIRS. Validation of the model was carried out using an independent set of brown rice samples. Applicability of the developed NIRS method to the characterisation of brown rice samples regarding their phytic acid content was evaluated on the basis of generally accepted NIRS performance criteria.

4.1.2 Determination of phytic acid in brown rice by the reference method

Brown rice of different grain shape (short, medium, and long grain rice) and different origin (Italy, USA, Japan, France, Spain, India, Guyana, Surinam), grown in different seasons (2000, 2001), was used as material. Samples were obtained from commercial suppliers or purchased at local supermarkets. Reference data on their phytic acid contents were obtained using the method described by Latta and Eskin [107]. It is based on the extraction of phytic acid under acidic conditions from freeze-dried rice flour and subsequent sample clean-up by means of anion-exchange chromatography. Phytic acid is quantified by an indirect photometric method which uses an iron (III) – sulfosalicylic acid complex as the colouring agent. Validation of the reference method yielded good repeatability data for the clean-up step (s = 0.05 %, n = 3) as well as for the overall methodology (s = 0.05 %, n = 4) and revealed excellent recovery (mean: 97 %, s = 1 %, n = 3) of spiked IP_6 from brown rice flour (range: 0.54 % - 2.69 %, n = 3).

Figure 5 shows the results obtained by reference analysis for 67 brown rice samples. The phytic acid contents in the brown rice samples analysed ranged from 0.85 % to 1.49 % (mean: 1.15 %). The majority of samples (45 of 67) exhibited phytic acid contents between 1.00 % and 1.30 %. Results were in agreement with data reported in literature [108].

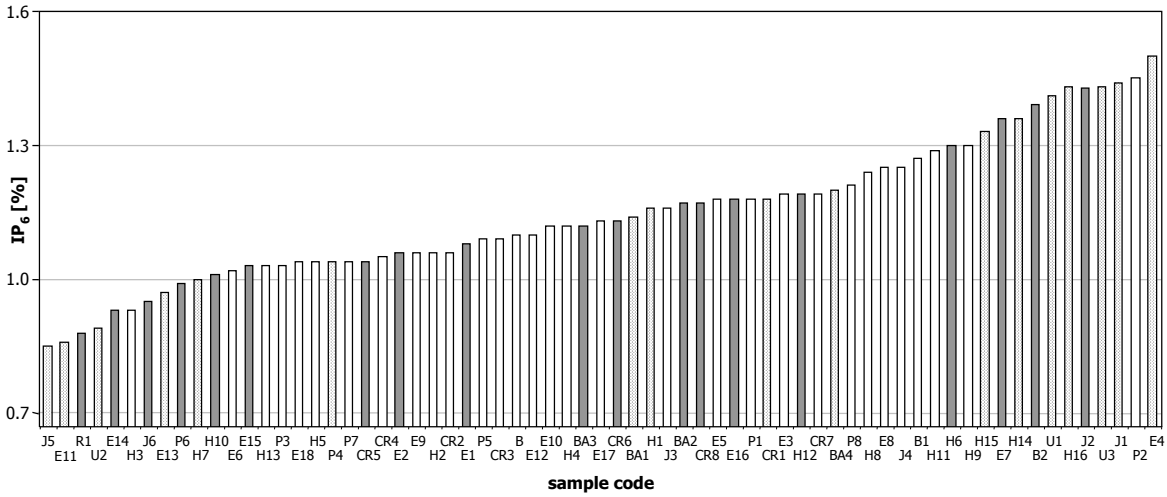

Figure 5: Phytic acid (IP_6) contents of 67 brown rice samples determined as reference data for establishment of an NIRS method for phytic acid analysis. For sample codes see chapters 3.1.1.3 and 3.1.1.4. White and patterned bars display reference data for samples assigned to the calibration set, gray bars those used for the validation sample set. White patterned bars display reference data for samples assigned to a sub-set consisting of equal proportions of samples low, medium and high in phytic acid.

4.1.3 Calibration and validation sample sets

Samples analysed by means of the reference method were divided into a set to be used for NIRS calibration and an independent set for validation of the calibration model to be developed. Reference data as well as sample characteristics were taken into consideration in order to obtain sets representative for the entire sample spectrum regarding phytic acid content, grain shape, and origin. This resulted in a validation set comprising 19 brown rice samples (see gray bars in Figure 5). Characteristics of calibration and validation samples are illustrated in Figures 6 and 7.

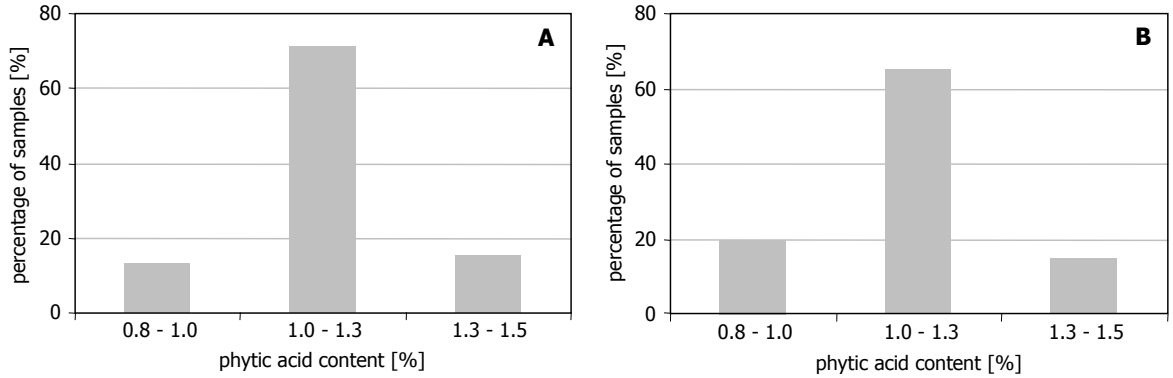

Figure 6: Percentages of brown rice samples exhibiting low (0.8 % - 1.0 %), medium (1.0 % - 1.3 %) and high (1.3 % - 1.5 %) phytic acid contents in (**A**) the calibration sample set (n = 48) and (**B**) the validation sample set (n = 19) to be used in the course of NIRS method development.

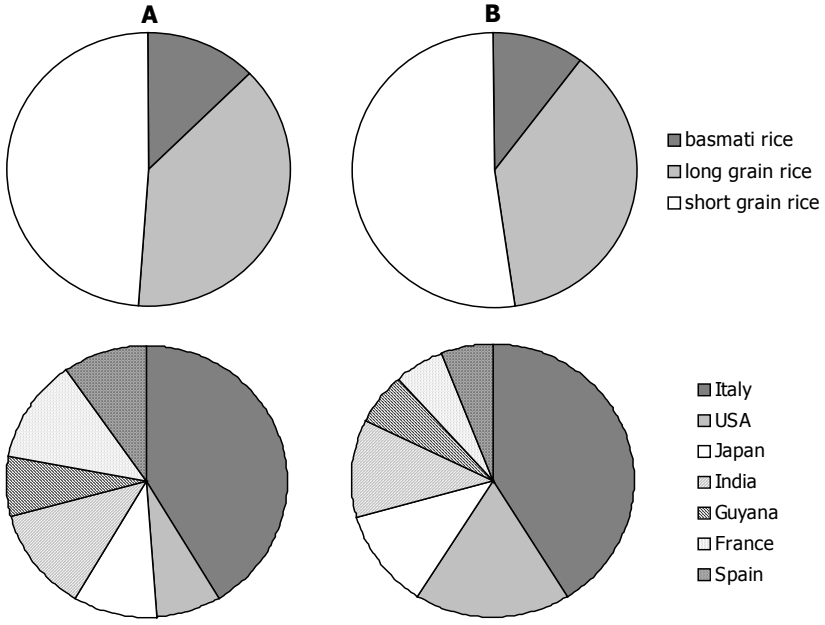

Figure 7: Distribution of brown rice samples of different grain characteristics and of different origin in (**A**) the calibration sample set (n = 48) and (**B**) the validation sample set (n = 19) to be used in the course of NIRS method development.

4.1.4 NIRS analysis of brown rice samples

NIRS analysis of brown rice was performed using ground freeze-dried samples as material. Samples were prepared by using a cyclone mill equipped with a 500 µm screen to standardise particle size, and by subsequent freeze-drying for three days. The moisture content of the resulting material determined as loss of weight by drying at 130°C for two hours was < 2%.

NIR spectra of flour samples were recorded from 1000 nm to 2500 nm in diffuse reflectance mode using a commercial FT-NIR spectrometer. In order to meet the challenge of sample inhomogeneity, flour samples were placed into a petri dish which was subsequently mounted onto a rotating measurement unit for NIRS analysis. This allowed accumulation and averaging of NIR spectra obtained from flour surface moving over the measuring aperture. Spectral data were saved as log(1/R) values at corresponding wavelength points.

NIRS measurement was optimised regarding spectral resolution and number of scans to be accumulated. The objective was to gain as much information as possible (high resolution) in a reproducible manner during a reasonable acquisition time. The standard deviation spectrum calculated from five-fold NIRS analysis of one rice sample was used for evaluation of reproducibility. Optimum results were obtained by averaging 100 spectra per sample, which had been recorded in 0.5 nm steps (resolution: 8 cm^{-1}). Setting a higher parameter value for spectral resolution (4 cm^{-1}) did not result in additional information but required double acquisition time to obtain reproducible NIR spectra. Figure 8A shows an NIR spectrum obtained from freeze-dried brown rice flour using optimised acquisition parameters (resolution: 8 cm^{-1}; 100 spectra per sample).

4.1.5 Pre-processing of original NIR spectra

Owing to differences in particle size distributions between the brown rice flour samples analysed, the resulting NIR spectra were significantly shifted and rotated relatively to each other. The objective of pre-processing was to reduce the influence of differences in particle size distributions as well as high correlations between original spectral data at different wavelengths. Following commonly applied standard procedures [56], this was achieved by calculation of derivatives from original spectra. In combination with a smoothing procedure, transformed spectra were eventually

obtained showing peaks rather than broad bands as observed in the original spectra (Figure 8B/C). These smoothed derivatives of the original spectra were used for establishment of calibration models.

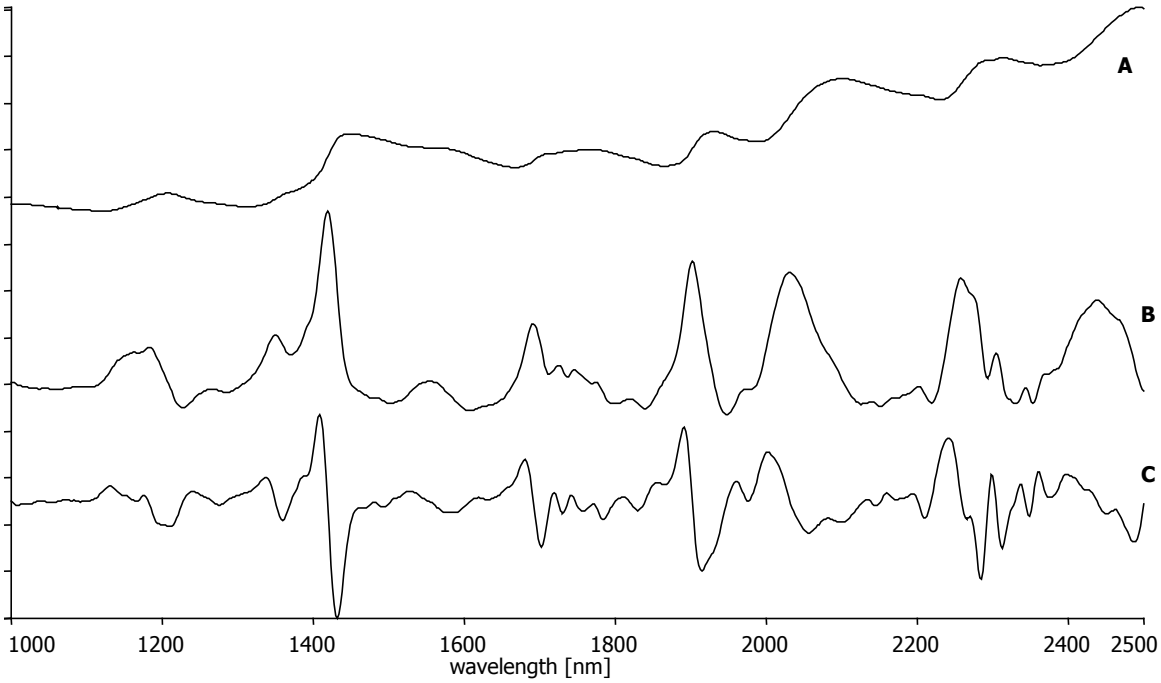

Figure 8: NIR spectra obtained from freeze-dried brown rice flour in diffuse reflection mode; **A**: original NIR spectrum; **B**: first derivative spectrum; **C**: second derivative spectrum.

4.1.6 Calibration model for NIRS determination of phytic acid in brown rice

The calibration model was developed in order to establish a correlation between spectral data and the phytic acid contents determined by reference analysis. Eventually, the calibration method must infer a rule for predicting the reference analysis results in future samples. This faces the challenge of the high sample variability regarding grain shape and origin on the one hand, and the low natural variability of the phytic acid contents in the brown rice samples analysed, on the other hand. IP_6 concentrations of samples in the calibration set differed by a factor up to 1.75. Various methods (multiple linear regression, ridge regression, principal component regression, partial least squares algorithms) were taken into consideration for model development. Considering the aim to develop a method applicable to a broad spectrum of rice samples, forward stepwise multiple linear regression (MLR), known as a robust methodology [56], was initially applied. In accordance to a previously reported method for NIRS analysis of phytic acid in cotton

seed, smoothed first derivatives of spectral data were averaged in 20 nm segments and ratios of averaged spectral data were calculated for each wavelength pair [105]. Averaged spectral data, ratios of spectral data and reference data of samples belonging to the calibration set were subsequently imported into self-tailored tools, in order to perform forward stepwise MLR. The models established were evaluated qualitatively on the basis of graphic plots, which allowed comparison of phytic acid contents predicted by means of NIRS to those determined by reference analysis. The root mean squared error of calibration (RMSEC), the estimated root mean squared error of prediction ($RMSEP_{Est}$) and the root mean squared error of crossvalidation (RMSECV) were calculated for quantitative assessment of model performance.

The use of forward stepwise MLR on the basis of the first derivative of the original spectra obtained from calibration samples (n = 48) resulted in a calibration model (model A) with an RMSEC of 0.10 % and a correlation coefficient of 0.74 (Figure 9).

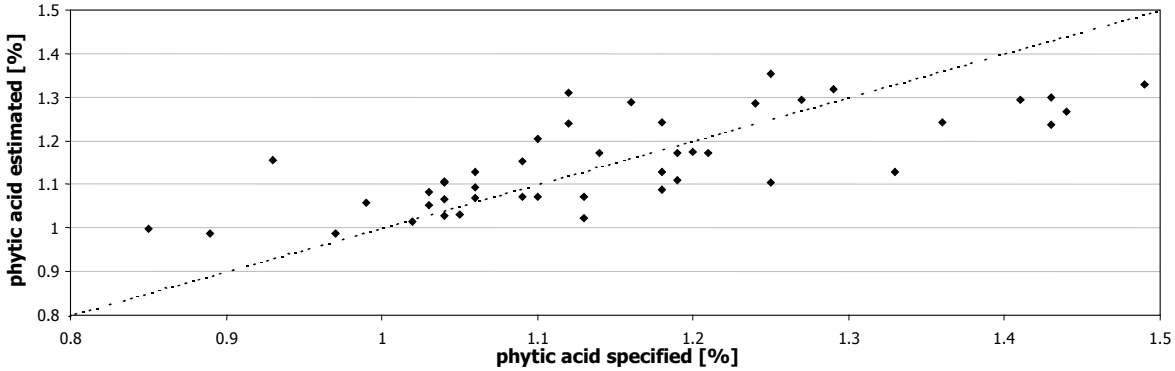

Figure 9: Evaluation of an NIRS calibration model for the determination of phytic acid in brown rice using MLR and the first derivative of spectral data obtained from the calibration sample set. Each square represents data for an individual sample. The phytic acid contents of the rice samples determined by the reference method (=specified) were plotted against the results predicted by the calibration model (=estimated). The closer the squares are located to the bisecting dashed line the better are the predictions of the phytic acid contents.

Spectral data at wavelength points between 2000 nm and 2300 nm showed the best correlation to reference data. However, due to predominance of samples exhibiting phytic acid contents between 1.00 % and 1.30 %, NIRS results for samples low (0.80 % - 1.00 %) and high (1.30 % - 1.50 %) in phytic acid were too high, and too low, respectively. Based on a principle reported in literature [56], the problem was overcome by the use of a sub-set for calibration consisting of equal proportions of samples low, medium and high in phytic acid (see white patterned bars in Figure 5).

This resulted in a calibration model (model B) with an improved RMSEC of 0.07 % and a higher correlation coefficient of 0.94 (Figure 10).

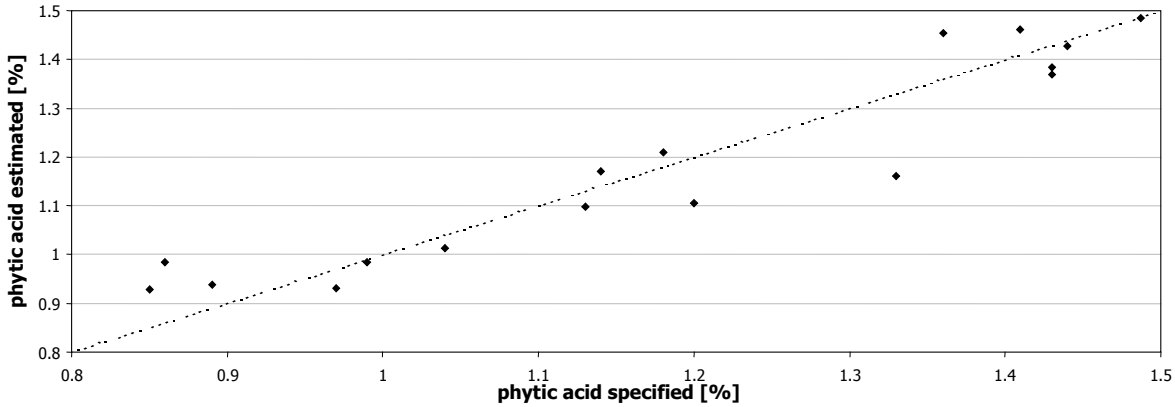

Figure 10: Evaluation of an NIRS calibration model (model B) for the determination of phytic acid in brown rice using MLR and the first derivative of spectral data obtained from a sub-set of calibration samples (n = 17). For further explanations see **Figure 9**.

Based on model B, the three rice samples X11, K11, and K21 were analysed by means of NIRS. The grains of these samples had been crushed in the course of previous investigations by means of a Single Kernel Characterisation System. Investigations by reference analysis had revealed that their phytic acid contents were lower than the lowest content observed for the calibration samples (0.85 %). This might be due to phytate cleavage catalysed by the endogenous phytase present in rice [109, 110]. The phytic acid contents predicted for the three samples on the basis of the NIRS calibration model were in excellent agreement with reference data (Figure 11). Accurate NIRS prediction of phytic acid contents outside the calibration range was seen as a first indication for the validity of the NIRS approach.

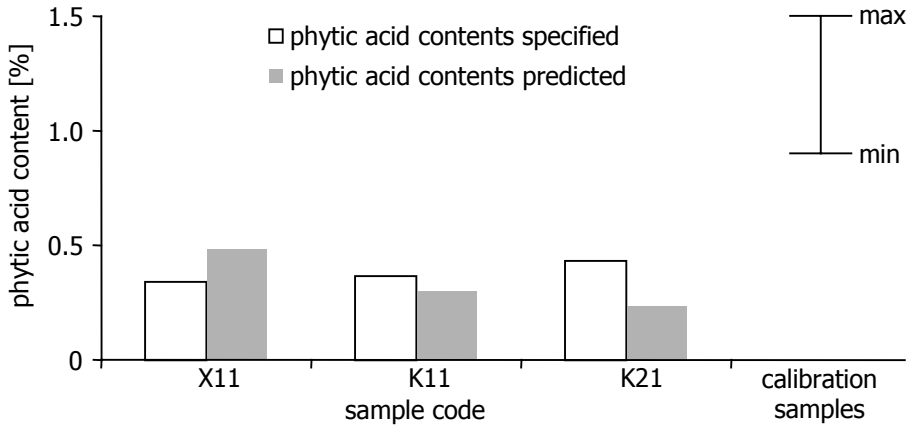

Figure 11: Phytic acid contents of low-phytate rice samples determined by means of the reference method (phytic acid contents specified) and by means of an NIRS method (phytic acid contents predicted). For sample codes see 3.1.1.1. The vertical line displays the range of phytic acid contents determined by reference analysis of the samples assigned to the NIRS calibration set.

In order to expand the calibration range of the NIRS model, one low-phytate rice sample was included into the calibration sample set. This resulted in a calibration model (model C, RMSEC = 0.07 %) showing a linear relationship (correlation coefficient: 0.97 %) between reference data and predicted data in a range from 0.3 % to 1.5 % (Figure 12).

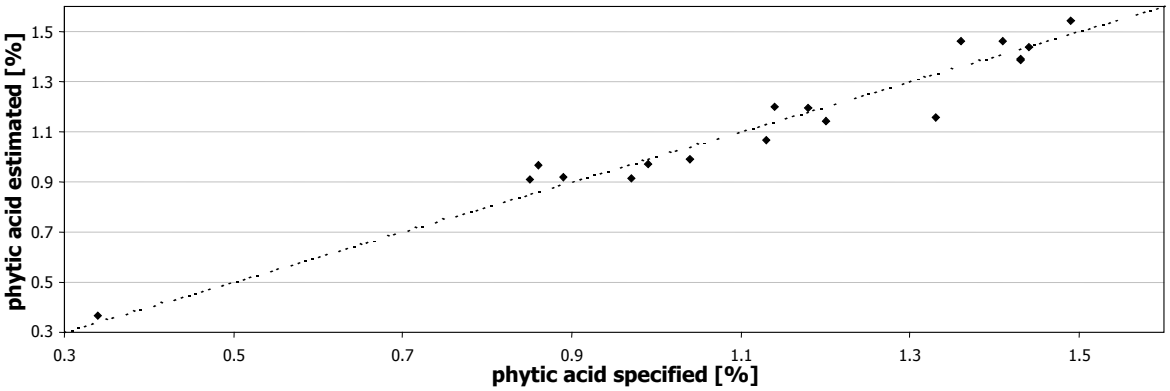

Figure 12: Evaluation of an NIRS calibration model (model C) for the determination of phytic acid in brown rice using MLR and the first derivative of spectral data obtained from a sub-set of calibration samples and from a rice sample low in phytic acid (n = 18). For further explanations see **Figure 9**.

As demonstrated in Figure 13, the refined model was further optimised by monitoring RMSEC, $RMSEP_{Est}$ and RMSECV values in dependence of the number of terms included into the calibration equation. The use of three terms resulted in an $RMSEP_{Est}$ of 0.09 %.

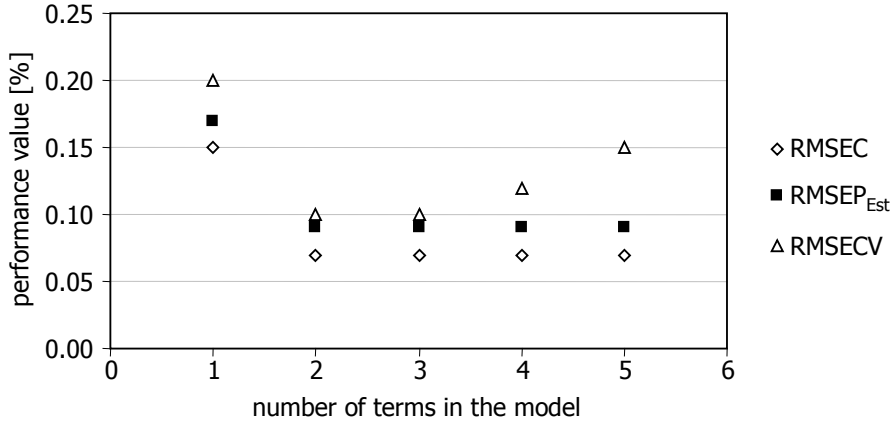

Figure 13: Evaluation of NIRS calibration equations for determination of phytic acid in brown rice using MLR and the first derivative of spectral data. Equations differ by the number of terms included. Performance data resulting from application of different equations are plotted against the number of terms included.

Crossvalidation of model C based on three terms revealed an RMSECV of 0.10 % (Figure 14).

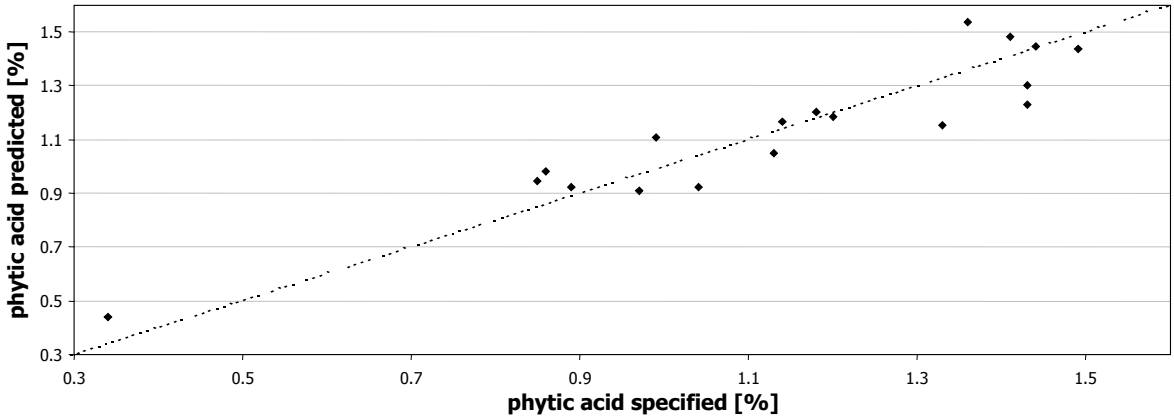

Figure 14: Crossvalidation of an NIRS calibration model (model C) for determination of phytic acid in brown rice. Each square represents data for an individual sample. The phytic acid content of each rice sample determined by the reference method (=specified) was plotted against the result predicted by a calibration model built without the respective sample (=predicted). The closer the squares are located to the bisecting dashed line the better are the predictions of the phytic acid contents.

In order to assess the performance of model C in comparison to the range/variability of the phytic acid contents in brown rice samples (standard deviation, maximum, minimum), NIRS performance criteria were calculated and compared to literature data generally accepted as minimum requirements for NIRS methods [60, 111]. Owing to the high standard deviation (0.30 %) observed for phytic acid contents within the calibration sample set used for model C, requirements reported in literature for calibration models could be fulfilled (Table 5).

Table 5: Comparison of NIRS performance criteria calculated for an NIRS calibration model for determination of phytic acid in brown rice (model C) to minimum requirements generally accepted for NIRS models [60, 111].

performance criterion			minimum requirement	model C
$\frac{(max - min)}{RMSECV}$	max:	maximum phytic acid contents observed for the calibration samples	10	10 max = 1.49 %
	min:	minimum phytic acid contents observed for the calibration samples		min = 0.34 % RMSECV = 0.10 %
$\frac{s}{RMSECV}$	s :	standard deviation of phytic acid contents calculated for calibration samples	3	3 s = 0.30 %

In order to verify the plausibility of the developed NIRS calibration model, wavelength points selected statistically for prediction of phytic acid contents in brown rice (spectral data at 2260 nm; ratio of spectral data at 2000 nm and 2220 nm; ratio of spectral data at 2000 nm and 2280 nm) were evaluated in the light of the pre-processed NIR spectrum obtained from sodium phytate (Figure 15). Occurrence of a peak in the pure compound spectrum close to the wavelength point included in the first term of the calibration model (2260 nm) supports arguments for validity of the NIRS calibration approach. Furthermore, the wavelength points selected were in excellent agreement with data report by de Boever *et al*. [106], who found strong correlations between spectral data obtained from starchy feedstuffs and phytic acid contents at 2244 nm.

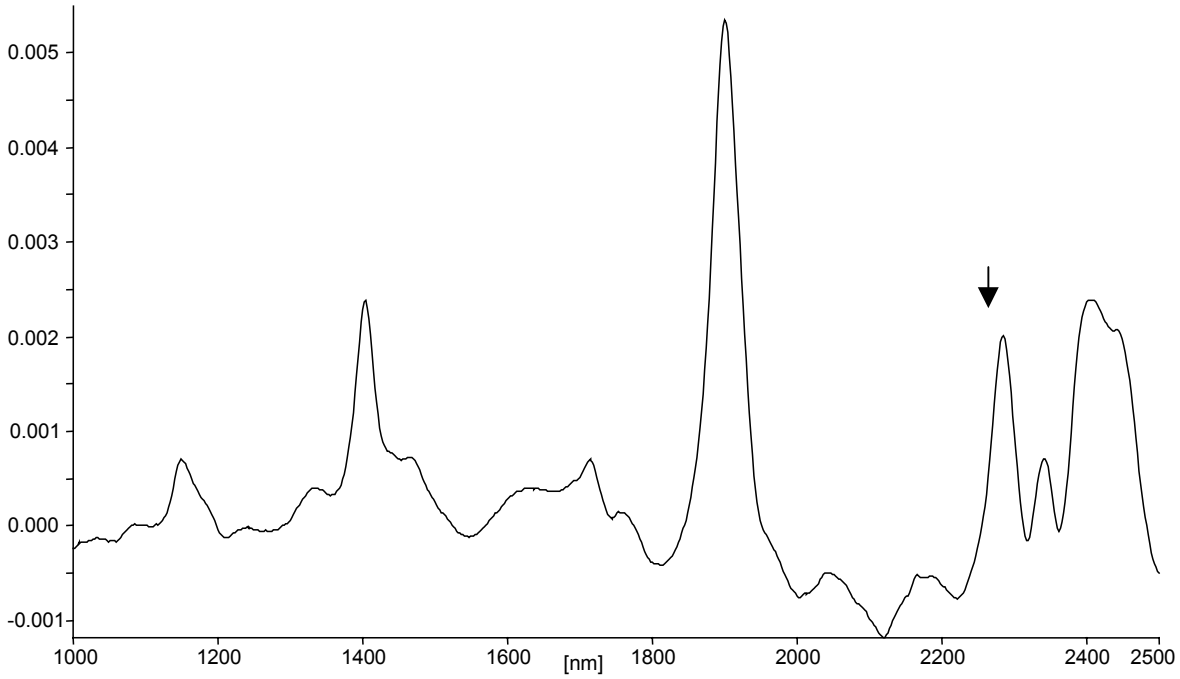

Figure 15: First derivatives of NIRS data obtained from solid sodium phytate in diffuse reflectance mode. The wavelength point, where the highest correlation between NIRS data and phytic acid contents of rice was found, is indicated by an arrow. The strong band around 1900 nm is due to the absorption of NIR light by water.

4.1.7 Validation of the calibration model

Considering the fact that generally accepted requirements for NIRS calibrations had been met on the basis of the optimised model C, this calibration model was validated. This was performed using the independent validation sample set which had been shown to be representative for the entire brown rice sample set (see 4.1.3). Validation was carried out by NIRS analysis of validation samples and

subsequent prediction of their phytic acid contents on the basis of spectral data and the calibration equation of model C. The phytic acid contents predicted were compared to those determined by reference analysis. This resulted in an root mean squared error of prediction (RMSEP) of 0.12 % (Figure 16).

Triplicate NIRS analysis of a validation sample revealed a standard deviation equal to that determined for the reference method (0.05 %).

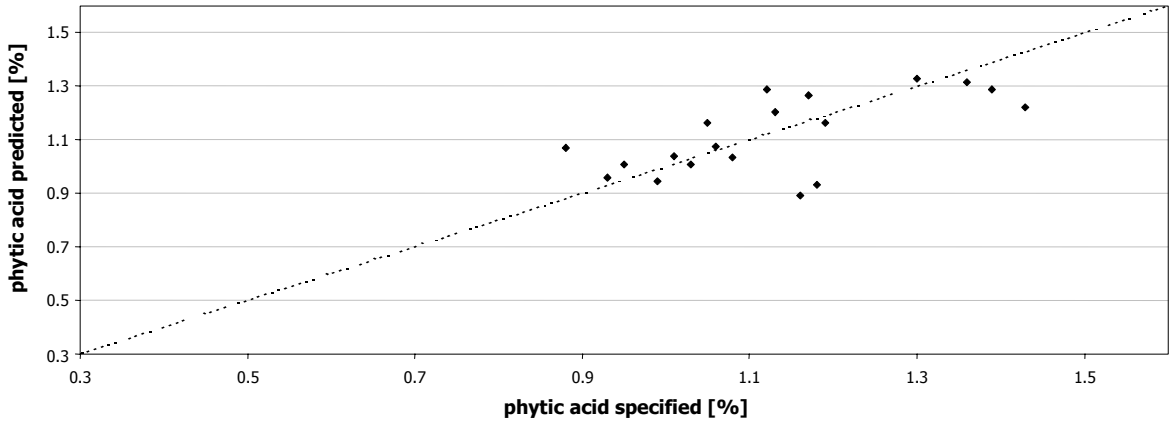

Figure 16: Validation of an NIRS calibration model (model C) for phytic acid determination in brown rice using an independent validation sample set (n = 19). Phytic acid contents predicted for validation samples on the basis of spectral data and model C are plotted against phytic acid contents determined by reference analysis.

The RMSEP achieved by applying MLR for calibration was significantly lower than RMSEPs obtained by models based on PLS algorithms (0.23 %) and ridge regression (0.20 %). Consequently, forward stepwise MLR was selected as the preferred calibration algorithm for NIRS determination of phytic acid in brown rice.

4.1.8 Method evaluation

Development of an NIRS method eventually aims to replace the reference method for routine sample analysis. Therefore performance data achieved by NIRS have to be compared to those of the reference method. This has to take into account, that NIRS as secondary method can never be more accurate than the method applied to provide reference data for calibration.

The RMSEP of the developed NIRS method (0.12 %) is higher than the lowest achievable accuracy (0.07 %), deduced from repeatabilities of the reference method (0.05 %) and of NIRS (0.05 %). According to minimum requirements described in literature [61], the RMSEP should not be higher than twice the standard deviation of the reference method (2 * 0.05 %). This criterion was nearly fulfilled. Performance

criteria for calibration models reported in literature [60, 111] could only be met on the basis of a sample set exhibiting high phytic acid variability (see 4.1.6). This indicated that the developed NIRS method is useful for rapid differentiation of brown rice samples low, medium and high in phytic acid.

MLR was chosen as calibration algorithm for model establishment. This resulted in high robustness of the method in terms of variability of sample characteristics.

Tuntawiroon *et al*. were able to increase iron bioavailibilty from 7.5 % to 22 % by decreasing the phytic acid content in a rice diet from 0.62 % to 0.12 % [98]. Hurrel *et al*. had shown, that the phytic acid content in a diet has to be reduced by ≥ 90 % in order to achieve a two-fold increase in iron absorption [99]. Taking these studies into consideration, the performance of the developed NIRS method is sufficient from a nutritional point of view. It allows to identify cultivars exhibiting low phytic acid contents and therefore improved nutritional properties.

In order to assess the progress in method development in comparison to previously reported approaches for NIRS analysis of phytic acid, the performance data achieved were evaluated in the light of those published in literature. So far, no study on applicability of NIRS to phytic acid determination in rice has been reported. As shown in Table 6, RMSEC, RMSECV and RMSEP values of the NIRS method developed were lower than data reported for phytic acid analysis in wheat by-products, cotton seed and broiler excreta. These matrices exhibit chemical compositions and phytic acid contents similar to brown rice. This indicates that progress could be achieved towards broader applicability and better performance of NIRS for phytic acid analysis.

Table 6: Performance data calculated for the NIRS method developed for determination of phytic acid in brown rice and data reported in literature for similar matrices.

	IP_6 contents [%]	RMSEC [%]	RMSECV [%]	RMSEP [%]	literature
brown rice*	0.3 – 1.5	0.07	0.10	0.12	-
feed flour	0.2 – 1.8	-**	0.35	-	[106]
gluten feed	1.1 – 2.4	-	0.28	-	[106]
cotton seed	1.6 – 3.0	0.12	-	-	[105]
broiler excreta	1.0 – 3.5	-	0.28	0.32	[112]

*: data refer to the method developed in this study; **: data not reported

4.1.9 Summary

An NIRS method for determination of phytic acid in brown rice flour was developed for the first time. A broad spectrum of rice samples was analysed by a photometric reference method as well as by Fourier-transform near infrared reflectance spectroscopy (FT-NIRS). Reference analysis revealed phytic acid contents ranging from 0.80 % to 1.50 % (mean 1.15 %, n= 67) and a predominance of samples exhibiting phytic acid contents between 1.00 % and 1.30 % (n = 45). Calibration model development was carried out using a set showing equal proportions of samples exhibiting low, medium, and high phytic acid contents. Best performance was obtained by means of forward stepwise MLR on the basis of the first derivative of spectral data. The performance data (RMSEC = 0.07 %; RMSECV = 0.10 %; RMSEP = 0.12 %) achieved were better than those reported in literature for matrices similar to rice. The developed method was assessed on the basis of generally accepted performance criteria. This led to the conclusion that the NIRS approach is applicable to the differentiation of rice samples low, medium and high in phytic acid.

4.2 Analysis of phytic acid in conventional and GM rice by means of NIRS

4.2.1 Introduction

Analysis of the chemical composition represents an important step in the safety assessment of genetically modified (GM) foods. In contrast to laborious wet - chemical methodologies, near infrared spectroscopy allows the rapid non-destructive investigation of a broad variety of matrices. Starting from raw materials, near infrared diffuse reflectance spectroscopy results in a spectrum which contains qualitative and quantitative information on a broad variety of constituents. Consequently, NIRS represents an interesting tool for rapid analytical characterisation of genetically modified as well as conventional foods. Classification of crop cultivars including GM lines using NIRS data and chemometric pattern recognition techniques has been demonstrated [113-115]. Furthermore, NIRS has been successfully applied to comparative quantitative analysis of major constituents in GM and corresponding parental lines [92]. This was achieved on the basis of calibration models, which described correlations between spectral and reference analytical data.

This study uses phytic acid as model compound in order to exemplarily demonstrate the application of NIRS to comparative analysis of a minor constituent in GM and conventional rice material. The objectives were to apply the developed NIRS method to the analysis of GM and parental rice lines and to assess the results in comparison to those obtained for conventional rice cultivars and in the light of the overall natural variability. Potential and limitations of the developed methodology should be discussed on the basis of the comparison of data predicted by NIRS and those determined by using the reference technique.

Two types of insect-resistant rice were used as GM material. Insect-resistance had been achieved by expression of a synthetic *cry1Ab* gene from *Bacillus thuringiensis* (Bt) [89] and a lectin gene from *Galanthus nivalis* (GNA) [91], respectively. Rice grains were obtained from a Bt-rice line (KMD1) and from the corresponding parental line (Xiushui 11) which had been grown under environmentally identical conditions in two field trials in China in 2001. Grains obtained from GNA rice and from the

corresponding parental line (ASD16) had been obtained from a glasshouse trial in 2001. Rice samples of the cultivars *Balilla* and *Cripto* were used as conventional material (sample codes BA1-BA4 and CR1-CR8 in Table 3). They were grown at different locations (France, Spain, Italy) and in different seasons (2000, 2001). Furthermore, a broad spectrum of commercial rice samples was available which had been provided by suppliers from Northern and Southern America, Asia, and Europe (see Table 4).

4.2.2 Phytic acid contents in conventional rice material

In order to gain insight into the intra-cultivar variability of phytic acid contents in conventional rice material, a set of rice samples comprising the cultivars *Balilla* (n = 4) and *Cripto* (n = 8) was analysed by means of NIRS. The calibration model used to derive phytic acid contents from NIRS data had been established on the basis of a sample set which included two *Balilla* and one *Cripto* sample (model C, see 4.1.6). Results were compared to the range of phytic acid contents (0.80 % - 1.50 %) determined for the set of commercial rice samples (n = 55) by reference analysis in the course of NIRS method development. The phytic acid contents predicted by NIRS for *Balilla* and *Cripto* rice grains grown at different locations and in different seasons spread from 1.00 % to 1.30 %, a range observed for 60 % (33 of 55 samples) of the commercial rice samples investigated (Figure 17). By using NIRS, no *Balilla* or *Cripto* sample exhibiting extraordinarily low (< 1.00 %) or high (> 1.30 %) phytic acid contents could be identified. The range of phytic acid contents observed for *Balilla* samples was smaller than that determined for *Cripto* (indicated by vertical lines in Figure 17). Interestingly, *Cripto* grains grown in Italy in 2001 exhibited the lowest (1.05 %) as well as the highest content (1.30 %) among *Balilla* and *Cripto* samples.

In summary, variability regarding phytic acid content determined by NIRS was higher within than between the two cultivars. Hence, no statistically significant difference in phytic acid contents between *Balilla* and *Cripto* could be detected. The results obtained were within the natural range determined for commercial rice samples.

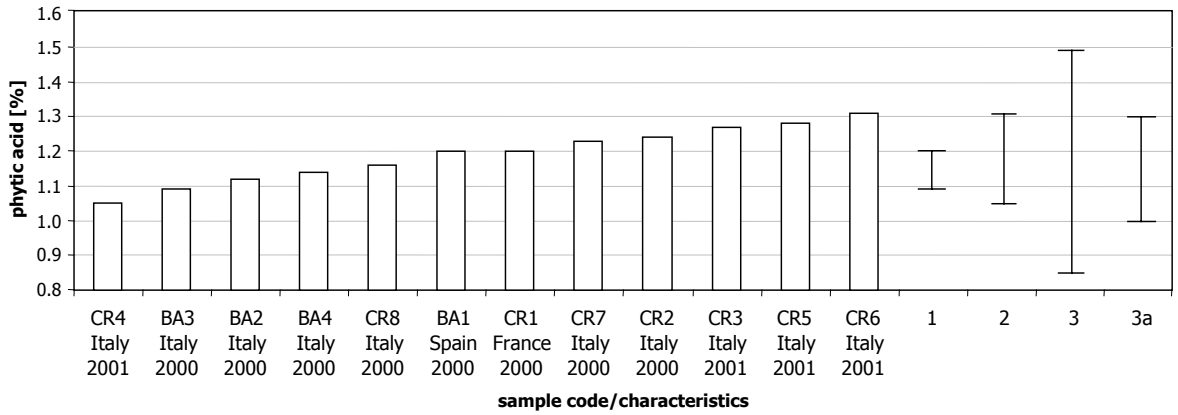

Figure 17: Phytic acid contents in conventional rice material determined by NIRS. Bars display data for rice grain samples obtained from cultivars *Balilla* (BA, N = 4) and *Cripto* (CR, N = 8) grown at different locations and in different seasons. Vertical lines display the ranges (from maximum to minimum) of phytic acid contents observed for (**1**) *Balilla* samples, (**2**) *Cripto* samples, and (**3**) commercial rice samples (natural range, N = 55). As regards the natural range, contents in 33 out of 55 samples (60 %) fell into a range between 1.00 % and 1.30 % (**3a**).

4.2.3 Phytic acid contents in genetically modified and parental rice lines

In the course of NIRS method development, rice samples with characteristics regarding cultivar and origin similar to those of genetically modified (KMD1, GNA) and parental lines (Xiushui 11, ASD16) had not been available for inclusion into the NIRS calibration model. In order to get an idea on the similarity of NIR spectra obtained from GM and parental lines to NIR spectra recorded from calibration and validation samples, a Principal Component Analysis (PCA) of their pre-processed spectral data was performed. As shown in Figure 18, scores observed for the transgenic Bt-rice lines (KMD1) were similar to those calculated for the corresponding parental lines (Xiushui 11). However, the scores were different in terms of PC 1 (sample codes KMD1 (1), Xiushui 11 (1)) and PC 2 (sample codes KMD1 (2), Xiushui 11 (2)) to those observed for calibration and validation samples. This indicated, that KMD1 and Xiushui 11 samples had to be classified as outliers exhibiting unique spectral features when compared to the calibration and validation sample sets. PC 1 scores calculated for ASD16 and GNA were higher than those observed for 88 % (59 out of 67) of the samples assigned to the calibration and validation sets.

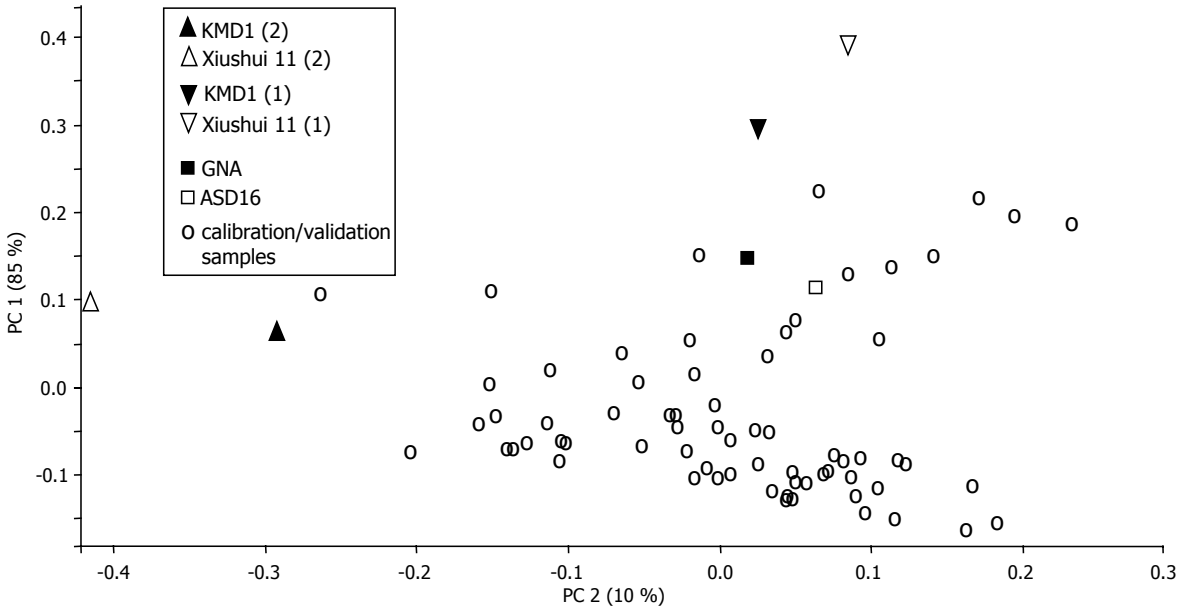

Figure 18: Principal Component Analysis of pre-processed NIRS data. Scores of principal components 1 and 2 for each sample are indicated by circles, squares and triangles. NIR spectra were obtained from calibration and validation samples used for development of an NIRS method for phytic acid analysis in brown rice (circles). Scores indicated by triangles and squares were calculated from NIRS data of GM and parental lines to be analysed using the developed NIRS method. For sample codes see 3.1.1.

Results obtained by PCA indicated that the established calibration model C might not be suitable for accurate NIRS analysis of phytic acid in genetically modified as well as in corresponding parental lines. Table 7 shows the results obtained by NIRS determination of phytic acid in the GM lines KMD1 and GNA and in the parental lines Xiushui 11 and ASD16. In agreement with the findings based on PCA, the predicted phytic acid contents were either extremely low or high; two predictions resulted even in "negative results".

Table 7: Phytic acid contents predicted by NIRS for transgenic Bt- and GNA rice (lines KMD1 and GNA, respectively) and for the corresponding parental lines (Xiushui 11 and ASD16, respectively). Phytic acid contents were derived from spectral data on the basis of calibration model C built without parental lines (see 4.1.6)

	phytic acid content predicted by NIRS [%]	
rice line	field trial 1	field trial 2
Xiushui 11	45.13	0.45
KMD1	-0.98	0.51
ASD16	0.48	-
GNA	-0.35	-

According to general experience in NIRS, accuracy of predictions for samples with unique features can be improved by inclusion of their spectral and reference data

into the calibration model. Considering the similarity of scores calculated for GM and the respective parental lines, it was decided to include one sample out of each pair into the calibration set. Consequently, parental lines were analysed by means of the reference method. Xiushui 11 exhibited phytic acid contents of 0.94 % and 1.01 % (field trials I and II, respectively). Reference analysis of ASD16 revealed a phytic acid content of 1.06 %.

Subsequently, a new calibration model was built on the basis of the calibration sample set for model C expanded by the parental lines. This led to more realistic NIRS results for parental as well as for GM lines. Phytic acid contents predicted for conventional cultivars *Balilla* and *Cripto* on the basis of the refined model were still in the same range as those observed using model C built without data from parental lines (Figure 19A).

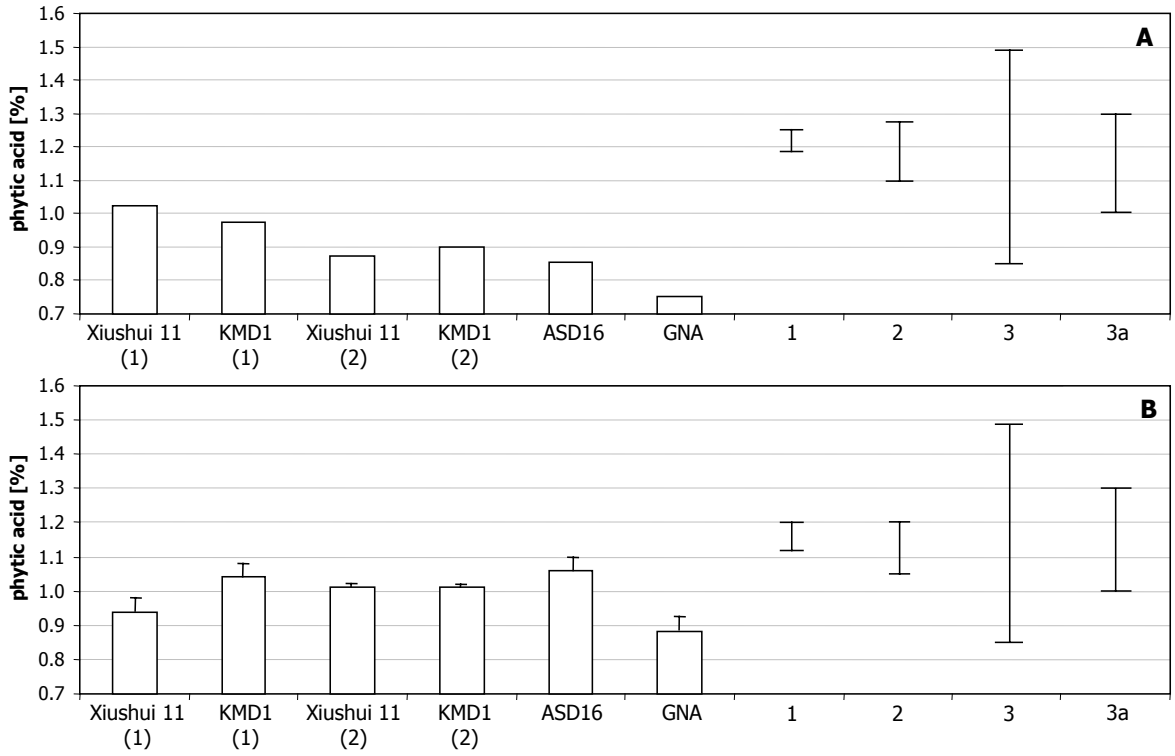

Figure 19: Phytic acid contents in GM and conventional rice material determined by NIRS and reference analysis.

A: Phytic acid contents in GM and conventional rice material determined by NIRS on the basis of a refined calibration model which includes reference and spectral data of parental lines (Xiushui 11 and ASD16). Bars display data for GM (KMD1 and GNA) and corresponding parental lines (Xiushui 11 and ASD16). For sample codes see 3.1.1. Vertical lines display the ranges (from maximum to minimum) of phytic acid contents for (**1**) *Balilla* samples (N = 4) and (2) *Cripto* samples (N = 8) determined on the basis of the refined model, and (**3**) commercial samples determined by means of reference analysis (natural range, N = 55). As regards the natural range, contents in 33 out of 55 samples (60 %) fell into a range between 1.00 % and 1.30 % (**3a**).

B: Phytic acid contents and ranges of phytic acid contents determined by reference analysis of the samples described in (**A**). Error bars display standard deviations (n = 3).

Using NIRS and the refined model, no statistically significant differences between genetically modified Bt-rice (IP_6 contents: 0.97 %; 0.90 %) and the corresponding parental lines (IP_6 contents: 1.01 %; 0.88 %) regarding phytic acid contents could be detected. Genetically modified GNA rice exhibited a lower phytic acid content (0.75 %) than its conventional counterpart (0.87 %). The magnitude of the difference was lower than the range of phytic acid contents determined for *Cripto* samples. Results obtained by NIRS analysis of GM and parental lines (range: 0.75 % - 1.01 %) were lower than those gained for the conventional cultivars *Balilla* (1.19 % - 1.25 %) and *Cripto* (range: 1.10 % - 1.27 %) and lower than phytic acid contents observed for the majority (60 %) of commercial rice samples (range: 1.00 % - 1.30 %).

The conclusions drawn on the basis of NIRS results for GM and parental lines as well as for conventional cultivars were in agreement with those drawn from the reference analytical data (Figure 19B). Phytic acid contents determined by reference analysis in GM and parental lines were higher than those predicted by means of NIRS. On the other hand, reference data for the conventional cultivars were lower than those predicted by NIRS. However, in accordance to the results obtained by NIRS, reference analysis did not reveal statistically significant differences between Bt- and corresponding parental lines regarding phytic acid content. Genetically modified GNA rice exhibited a statistically significantly lower mean phytic acid content (0.88 %) than its conventional counterpart (1.06 %). The magnitude of the difference was lower than the range of phytic acid contents determined for *Cripto* samples. Reference data of GM and parental lines were in a lower range (0.88 % - 1.06 %) than those observed for the conventional cultivars *Balilla* and *Cripto* (range: 1.05 % - 1.20 %). They were also lower than the mean phytic acid contents observed for commercial rice samples (1.15 %).

4.2.4 Conclusions

Near infrared spectroscopy was applied for rapid non-destructive determination of the phytic acid contents in genetically modified and conventional rice material. The investigations revealed low phytic acid contents for GM and corresponding parental lines in comparison to those of the majority of commercial rice samples. In contrast to the Bt-rice line, the phytic acid content determined for genetically modified GNA

rice was different to that of the conventional counterpart. Results obtained by NIRS on the basis of a calibration data set including parental lines were in good agreement with data obtained by the traditional method. The study demonstrated the applicability of NIRS for rapid non-destructive screening of GM food for differences in chemical composition to conventional counterparts. A decrease in phytic acid content is beneficial from a nutritional point of view. However, safety assessment of the GM lines used as material goes beyond the scope of the present study. Investigation of material obtained from additional field trials is required in order to prove the correlation between differences in composition and the application of recombinant DNA techniques.

Using phytic acid as model compound, the present study exemplarily shows, that NIRS allows the rapid screening of a broad spectrum of rice samples for a nutritionally important minor compound. The accuracy of NIRS determinations strongly depends on sample characteristics which might have an influence on the NIR spectrum. Predictions by means of NIRS for samples exhibiting unique spectral features, can be totally wrong. As regards applicability of NIRS within the safety assessment framework of transgenic rice, the study revealed that appropriate comparators (parental lines) must be available for calibration purposes. It was demonstrated, that even a rather simple calibration methodology like stepwise MLR allows to take unique sample characteristics into account, in order to improve the accuracy of predictions.

In the present study, the time needed for calibration and validation purposes was much higher than the time saved due to rapidness of the NIRS methodology. However, the demonstration of the general suitability of NIRS to the analysis of a minor constituent (phytic acid) in a complex matrix (rice) paves the way for the application of this technique to screen broad sample sets in the course of breeding programmes and field trials with GM crops.

4.2.5 Summary

An NIRS method was applied to the rapid non-destructive determination of phytic acid contents in genetically modified insect-resistant Bt- and GNA rice. Results were compared to those obtained from corresponding parental lines and from conventional rice samples.

The use of an NIRS calibration model built with samples which were different to GM/parental lines in terms of cultivar and origin led to false NIRS predictions. This was due to unique spectral features of the samples analysed in comparison to those used for calibration of the NIR spectrometer. Refinement of the NIRS method by inclusion of parental lines into the calibration sample set resulted in more accurate predictions which were in agreement with reference data.

No differences between genetically modified Bt-rice and corresponding parental lines regarding phytic acid contents were detected. Genetically modified GNA rice exhibited a lower phytic acid content than its conventional counterpart. The magnitude of the difference was lower than the range of phytic acid contents observed for the conventional cultivar *Cripto*. The phytic acid contents of GM as well as of parental lines were lower than those of the majority of commercial rice samples.

As regards applicability of NIRS within the safety assessment framework of transgenic rice, the study revealed that appropriate comparators (parental lines) must be available in the course of method development. This eventually allows rapid non-destructive NIRS analysis of phytic acid in genetically modified rice and the assessment of results in the light of those obtained from conventional rice samples.

4.3 Development of a metabolite profiling method for analysis of rice grains

4.3.1 Introduction

The comparison of the chemical composition of the genetically modified plant to that of a traditionally obtained counterpart has been a key element in the safety assessment of GM crops. In addition to the intended effect of genetic modification, unintended effects may result from the application of recombinant DNA techniques. The application of a targeted approach for comparative analysis of defined constituents has been shown to be suitable for the detection of statistically significant differences between parental and GM lines which go beyond the intended effects of the genetic modification [44, 45]. In order to increase the chances to detect unintended effects due to the application of recombinant DNA techniques, profiling on the metabolite level has been suggested as tool for comprehensive characterisation of the chemical compositions of GM plants [6].

The role of cereal grains as storage organs results in a complex composition characterised by extreme differences in concentrations of compounds. The spectrum of constituents is mainly determined by high amounts (up to 80 %) of polysaccharides that serve as storage substances. In addition, there are low molecular weight polar constituents belonging to various compound classes such as free sugars, sugar alcohols, organic acids, amino acids, and phenolics. The lipid content of rice is relatively low (2 - 3 %). However, even minor representatives of this class, *e.g.* phytosterols, γ-oryzanol and tocopherols, deserve attention because of their nutritional importance [108].

The objective of this study was to develop a metabolite profiling methodology using rice as model crop. The study aimed to provide a method allowing the non-targeted unbiased investigation of the low molecular weight compounds present in rice. This faces the challenge to be as comprehensive as possible and to take into account the chemical characteristics specific for different compound classes. In addition, sample preparation should ensure the suitability of the employed approach to cereals with compositions similar to rice, *e.g.* maize and barley. The technique used to investigate the complex sample extracts should allow to detect and resolve a broad spectrum of

constituents belonging to various compound classes. In order to be able to evaluate differences between metabolite profiles, the technique should also allow to obtain structural information on the compounds detected. The metabolite profiling methodology should be validated and applied to comparative analysis of cereals with known differences on the metabolite level. This should provide a positive control for the suitability of the approach to detect unintended changes in metabolite concentrations. Eventually, the methodology should be applied to the analytical characterisation of genetically modified rice.

4.3.2 Principle

A scheme of the profiling method established is shown in Figure 20. It starts with consecutive extraction of lipids and polar compounds from freeze-dried rice flour. Both extracts are characterised by the presence of major constituents (*e.g.* triglycerides and free sugars) and a spectrum of minor compounds (*e.g.* organic acids, amino acids, free fatty acids, phytosterols). Therefore, both extracts were further subdivided by using transesterification and solid phase extraction (SPE) for the lipids and by making use of the relative stability of silylated derivatives for the polar compounds. Coupled high-resolution gas chromatography (GC) to flame ionisation detection (FID) was selected as technique for routine analysis of silylated / methylated derivatives. GC coupled to mass spectrometry (MS) was employed to identify sample constituents.

4.3.3 Extraction

Rice flour was freeze-dried to standardise the water content for the subsequent extraction procedure. Several solvents were tested for extraction. The use of ethanol resulted in a complex mixture of lipids and polar compounds as described for barley [70]. However, considering the concentration differences described above, it seemed more appropriate to prepare separate extracts of polar and lipid constituents.

A method commonly applied for total lipid extraction is based on the use of a mixture of chloroform and methanol [71]. Owing to a co-extraction of non-lipids, this procedure involves an additional washing step. This often results in emulsions rendering precise phase separation difficult and resulting in potential loss of polar constituents. The consecutive use of methanol and dichloromethane proved to be

suitable to overcome these drawbacks. By applying solely dichloromethane, complete removal of the lipids from rice flour was not possible. However, disintegration of the flour matrix by pre-treatment with methanol and subsequent use of dichloromethane resulted in exhaustive extraction of lipids as demonstrated by sequential extractions. Residues of the solvent were removed from the defatted flour under vacuum.

In accordance with previously described approaches [69], a mixture of methanol and water (80+20, v+v) was used for extraction of polar compounds. Attempts to increase the polarity of the solvent by increasing the water content were not satisfactory from an operational stand point because of foaming during the concentration process under reduced pressure.

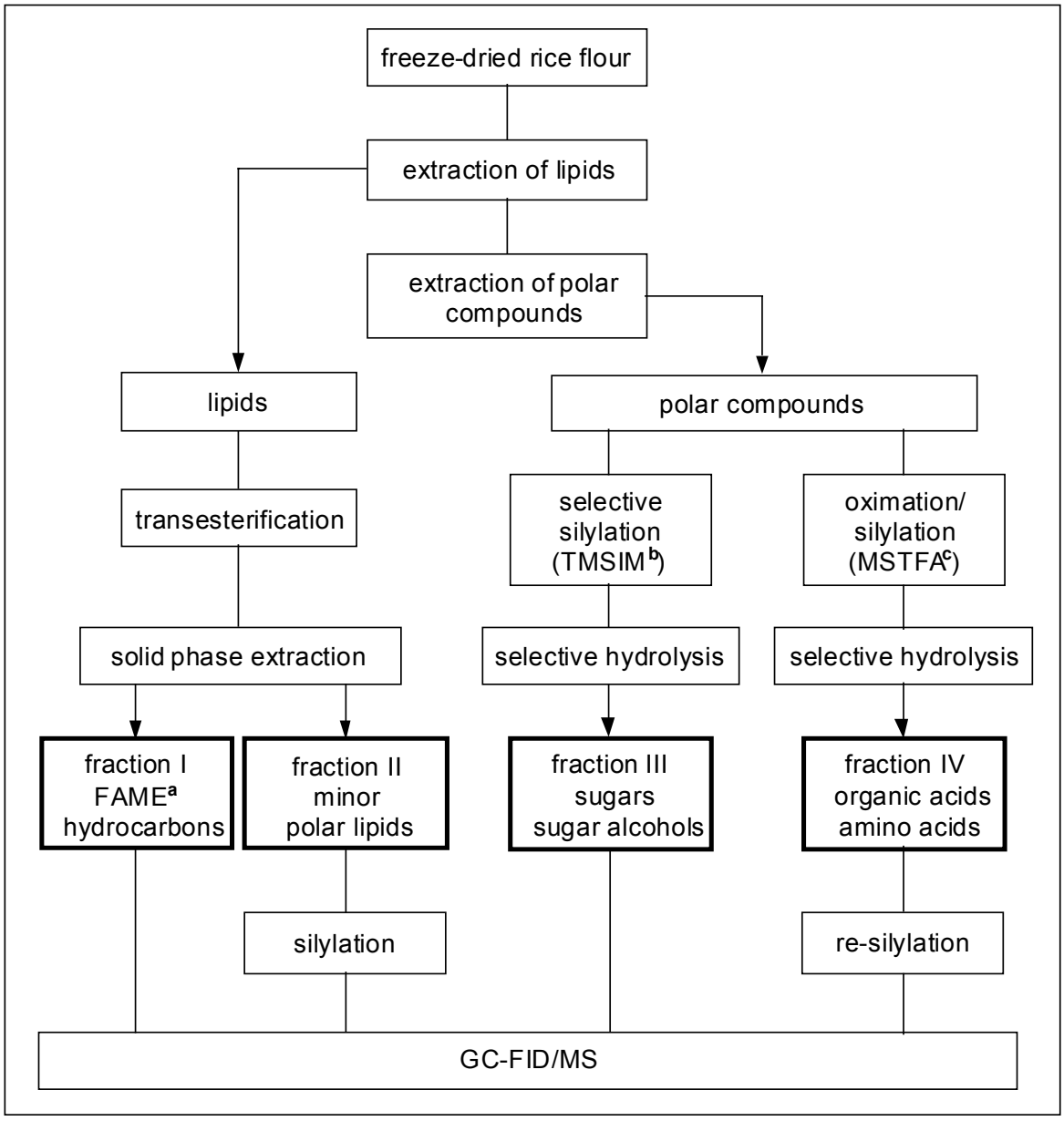

Figure 20: Metabolite profiling method. a: FAME = fatty acid methyl esters, b: TMSIM = trimethylsilylimidazole, c: MSTFA = *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide.

4.3.4 Fractionation of lipids through transmethylation and solid phase extraction

The objective of the fractionation procedure was to separate the abundant triglycerides from the minor lipids (*e.g.* sterols, tocopherols). This was achieved by transmethylation of the triglycerides and subsequent solid phase extraction on silica gel using elution solvents with increasing polarity [116, 117].

Base-catalysed transmethylation using sodium methylate in methanol as catalyst converts triglycerides and other fatty acid esters into the respective methyl esters. It can be performed under mild conditions and allows separate analysis of the metabolically important free fatty acids which are not methylated. Furthermore, it is useful for the analysis of the sum of free and esterified minor lipids, *e.g.* sterols, triterpenic alcohols.

Conditions for transesterification were optimised using the model compounds tripalmitin, cholesteryl palmitate, and γ-oryzanol. In agreement with previously described methodologies [118], transesterification of tripalmitin was accomplished within a few minutes at ambient temperature. According to [117], complete transesterification of most steryl esters can be achieved within an hour. However, the model experiment revealed that a reaction time of 90 min is required at room temperature for complete transmethylation of cholesteryl palmitate. γ-Oryzanol comprises a mixture of phytosteryl ferulates [108] known to be quite stable under alkaline conditions [119]. Consequently, only a small but highly reproducible percentage (ca. 10 %) of γ-oryzanol was cleaved within 90 min. A temperature increase to 50°C led to an accelerated cleavage of γ-oryzanol but triggered the formation of free fatty acids instead of methyl esters in the course of transesterification of tripalmitin and cholesteryl palmitate. This was not desirable because of the presence of endogenous free fatty acids in rice. Considering the reproducible process of γ-oryzanol transmethylation, a reaction time of 90 min at room temperature was chosen as optimum condition for transesterification of rice lipids.

Figure 21A shows the GC-FID chromatogram of an unfractionated lipid extract after transesterification and silylation.

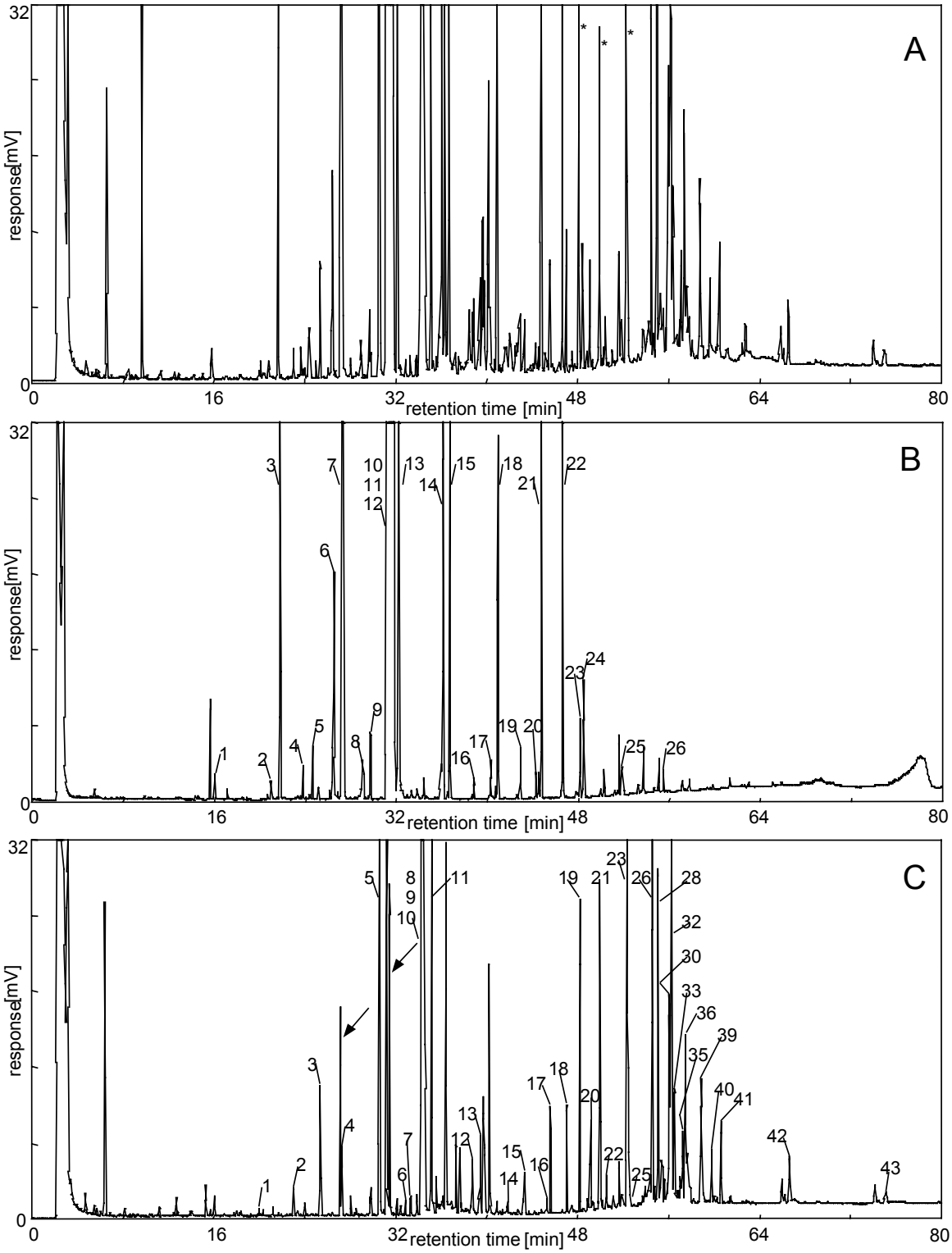

Figure 21: Gas chromatography using flame ionisation detection of a rough rice lipid extract spiked with α-, γ-, and δ-tocopherol. For conditions see 3.2.1.6. **A**: Total lipid extract after transesterification and silylation. Spiked tocopherols are indicated by asterisks. **B**: fraction I (numbers in **Table 8**). **C**: fraction II (numbers in **Table 9**). Residual fatty acid methyl esters are indicated by arrows.

Solid phase extraction of the transmethylated lipid extract using n-hexane/MTBE (100+2, v+v) results in a fraction consisting mainly of fatty acid methyl esters (FAME) derived from ester bound fatty acids and hydrocarbons (Figure 21B).

Compounds identified in this fraction by means of GC-MS using an apolar stationary phase are listed in Table 8. The isolation of the fatty acid esters in this separate fraction also opens the possibility for further analysis, *e.g.* according to degree of unsaturation, on a more polar column.

Table 8: Compounds identified by gas chromatography - mass spectrometry in fraction I obtained from a rough rice sample.

no.	compound	identification[a]	no.	compound	identification[a]
fatty acid methyl esters			14	C20:1 (cis 11)	1,2
1	C12:0	1,2	15	C20:0	1,2
2	C14:1	3	16	C21:0	1
3	C14:0	1,2	17	C22:1 (cis 13)	3
4	C15:1	3	18	C22:0	1,2
5	C15:0	1	19	C23:0	1
6	C16:1 (cis 9)	1,2	20	C24:1	1
7	C16:0	1,2	21	C24:0	1,2
8	C17:1	3	24	C26:0	1
9	C17:0	3	25	C28:0	1
10	C18:3 (cis 9,12,15)	1,2	26	C30:0	1
11	C18:2 (cis 9,12)	1,2	**hydrocarbons**		
12	C18:1 (cis 9)	1,2	22	squalene	1,2
13	C18:0	1,2	23	C29:0	1

[a]: Identification: **1**: according to NIST mass spectra library [80]; **2**: according to retention times of reference compounds; **3**: according to metabolite mass spectra library of Max Planck Institute of Molecular Plant Physiology [81].

By using hexane/MTBE (7+3, v+v) as elution solvent, a fraction containing the more polar lipids was obtained (Figure 21C). Table 9 lists representatives of the various classes (*e.g.* sterols, tocopherols, free fatty acids, waxy alcohols) identified by means of GC-MS analysis of the silylated (MSTFA) fraction. The amount of fatty acid methyl esters not retained in the previous fraction was below 2 %.

In the course of the development of the method it became obvious that separation of tocopherols and fatty acid methyl esters is a critical step in the SPE procedure. Tocopherols are occurring naturally in rice only at low levels. Therefore the transmethylated lipid extract was spiked with α-, γ-, and δ-tocopherol prior to fractionation (see asterisks in Figure 21A). After column chromatography (Figure 21C), 100 % (γ- and δ-tocopherol) and 95 % (α-tocopherol) of the added compounds could be recovered in fraction II. This clear-cut procedure was confirmed

by silylation of fraction I. Except of about 5% α-tocopherol, the resulting chromatogram showed no difference to the one shown in Figure 21B.

Table 9: Compounds identified by gas chromatography - mass spectrometry in fraction II obtained from a rough rice sample.

no	compound[a]	identification[b]	no	compound[a]	identification[b]
free fatty acids			**sterols and triterpenic alcohols**		
1	C12:0	1,2,5	25	cholesterol	1,2
3	C14:0	1,2	26	campesterol	2,3
5	C16:0	1,2,5	27	campestanol	6
6	C17:0	1	28	stigmasterol	2,3
8	C18:3 (cis 9,12,15)	1,2,5	29	Δ^7-campestenol	6
9	C18:2 (cis 9, 12)	1,2,5	31	obtusifoliol	7
10	C18:1 (cis 9)	1,2,5	32	β-sitosterol	2,3
11	C18:0	1,2,5	33	sitostanol	6
12	C20:1	1	34	Δ^5-avenasterol	7
13	C20:0	1	35	gramisterol	7
15	C22:0	1	36	cycloartenol	6,7
16	C23:0	3	37	Δ^7-stigmastenol	6,7
18	C24:0	4	38	Δ^7-avenasterol	7
22	C26:0	4	39	24-methylencycloartanol	6,7
alcohols			40	citrostadienol	7
7	C18:0	3	**phenolic compounds**		
14	C22:0	3	2	methyl p-hydroxycinnamate	1
17	C24:0	4	4	methyl ferulate	1,2
20	C26:0	4			
24	C28:0	1			
30	C30:0	4			
41	C32:0	4			
42	C34:0	4			
43	C36:0	4			
tocopherols					
19	δ-tocopherol	2			
21	γ-tocopherol	2			
23	α-tocopherol	2,3			

[a]: Metabolites identified as trimethylsilylated derivatives; [b]: Identification: **1**: according to NIST mass spectra library [80]; **2**: according to retention times of reference compounds; **3**: according to metabolite mass spectra library of Max Planck Institute of Molecular Plant Physiology [81]; **4**: based on mass spectral data; **5**: comparison of mass spectral data to those of reference compounds; **6**: comparison of mass spectral data to those reported by Xu and Godber [125]; **7**: comparison of mass spectral data to those reported by Kamal-Eldin [126].

4.3.5 Fractionation of the polar extract through differential derivatisation

Fractionation of the polar extract was performed in order to separate free sugars present as major polar low molecular weight compounds from minor constituents, such as organic acids and amino acids. Since silylating reagents are known to strongly differ in their suitability for different compound classes [120], fractionation would allow an appropriate choice of reagents to ensure complete derivatisation.

The fractionation of the polar extract is based on the relative stability of the R-$Si(CH_3)_3$ group of sugars/polyols and amino acids/organic acids to aqueous hydrolysis, a principle developed for the analysis of amino acids in human blood [121]. Mixing the silylated polar extract with n-hexane/water results in the formation of two phases. The silylated sugars and polyols which are relatively stable to hydrolysis are enriched in the hexane layer. The aqueous phase contains the free amino acids/organic acids which are readily released from their unstable trimethylsilyl derivatives.

In view of the above-mentioned differences in suitability of silylating reagents, several reagents and combinations thereof were tested. Figure 22 shows a gas chromatogram of the total polar extract obtained from rice flour after derivatisation with MSTFA. Sucrose and raffinose are not completely persilylated resulting in multiple peaks for these compounds.

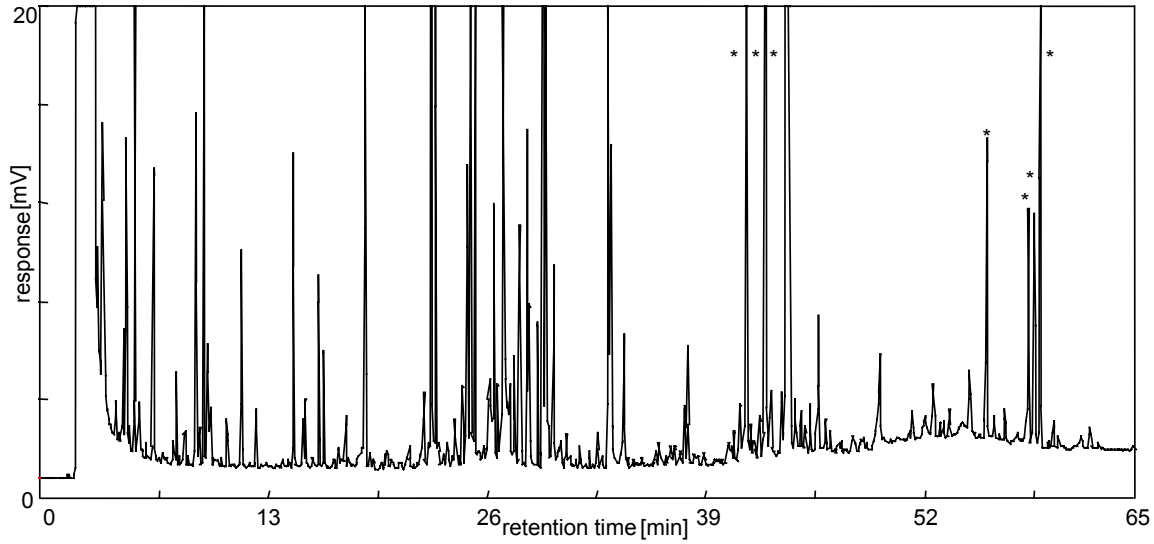

Figure 22: Gas chromatography using flame ionisation detection of a total methanol-water extract of a rough rice sample after silylation with MSTFA. For conditions see 3.2.1.6. Peaks of incompletely silylated sugars are indicated by asterisks.

Trimethylsilylimidazole (TMSIM) is a more powerful silylating agent for sugars [120]. It persilylates mono-, di-, and trisaccharides quickly under mild conditions. Silylation of the polar rice extract with TMSIM and subsequent treatment with a hexane/water mixture resulted in a fraction containing only persilylated sugars and polyols (Figure 23). Metabolites identified in a GC-MS study are listed in Table 10.

Table 10: Compounds identified by gas chromatography - mass spectrometry in fraction III obtained from a rough rice sample.

no	compound[a]	identification[b]	no	compound[a]	identification[b]
sugars			**sugar alcohols**		
4	arabinose	1,2,3	1	glycerol	1,2
5	arabinose		2	threitol	1
7	fructose	1,2,3	3	erythritol	1
8	fructose		6	xylitol	1
9	fructose		13	inositol	1,2
10	galactose	1,2,3			
12	galactose				
11	glucose	1,2,3			
13	glucose				
15	sucrose	1,2			
16	raffinose	1,2,3			

[a]: Metabolites identified as persilylated derivatives; [b]: Identification: **1**: according to NIST mass spectra library [80]; **2**: according to retention times of reference compounds; **3**: comparison of mass spectral data to those of a reference compound.

According to West and Moskowitz [122] persilylation of sugars and polyols can be achieved by means of TMSIM within 15 min at room temperature using pyridine as solvent. However, in order to dissolve constituents of polar rice extract in pyridine, heating at 70°C was required. Silylation of reducing sugars by means of TMSIM and subsequent gas chromatographic analysis on an unpolar column led to multiple peaks resulting from persilylated anomers which were readily separated. A common procedure for gas chromatographic analysis of sugars involves an additional oximation step prior to silylation. However, no clear advantage was seen in the use of oxime TMS derivatives for analysis, since for a mixture of reducing sugars (fructose, glucose, galactose) at least two peaks per sugar were observed (diastereomeres) which were not completely resolved. These findings were in agreement with those reported by Nikolov and Reilly [123].

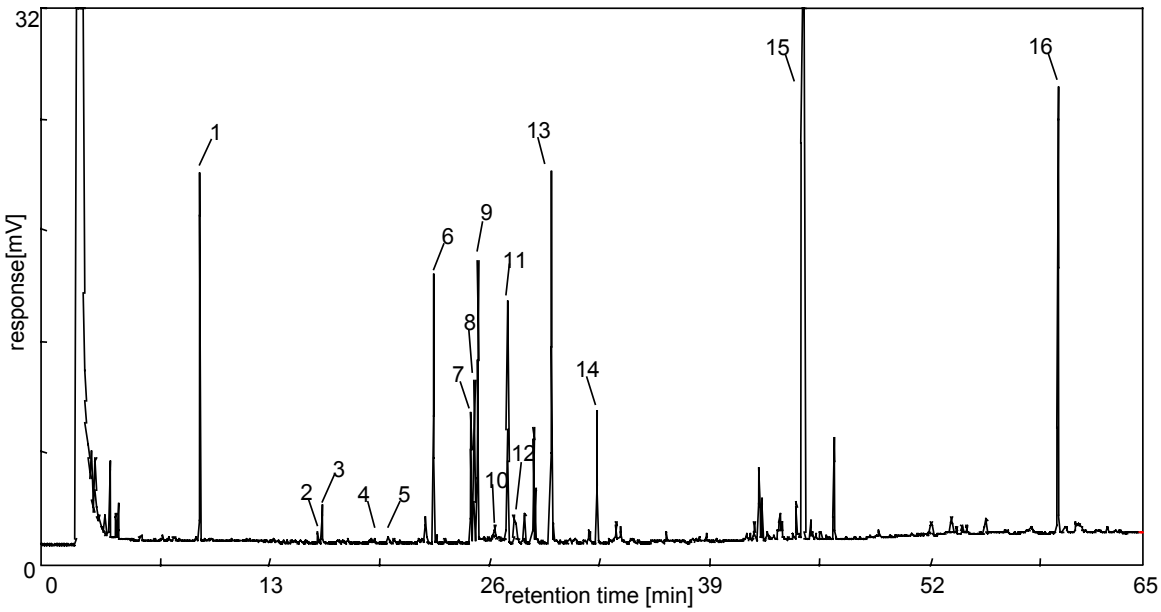

Figure 23: Gas chromatography using flame ionisation detection of fraction III. For numbers see **Table 10**. For conditions see 3.2.1.6.

The amino function of amino acids cannot be silylated using TMSIM [120]. Its presence even inhibited derivatisation of amino acids with MSTFA. Therefore, the analysis of organic acids/amino acids has to start from a separate aliquot of the polar extract. In order to protect α-ketoacids from enolisation and decarboxylation prior to trimethylsilylation an additional oximation step was necessary. A model study with α-ketoglutarate (oximated with hydroxylamine, silylated with MSTFA) demonstrated the above-described fractionation strategy to be also applicable to oximated derivatives. Addition of water has no effect on the R_1R_2C=N-O-Si$(CH_3)_3$ group but hydrolyses the R-C(O)-O-Si$(CH_3)_3$ group resulting in an enrichment of the compound in the aqueous layer. Application of this strategy to the polar extract from rice flour and re-silylation of the organic acids and amino acids enriched in the aqueous layer (fraction IV) resulted in the chromatogram shown in Figure 24. Metabolites identified in this fraction are summarised in Table 11. Sugars as major constituents of the polar extract are almost completely separated from organic acids and amino acids. In turn, no silylated amino acids were detected in the respective hexane layer.

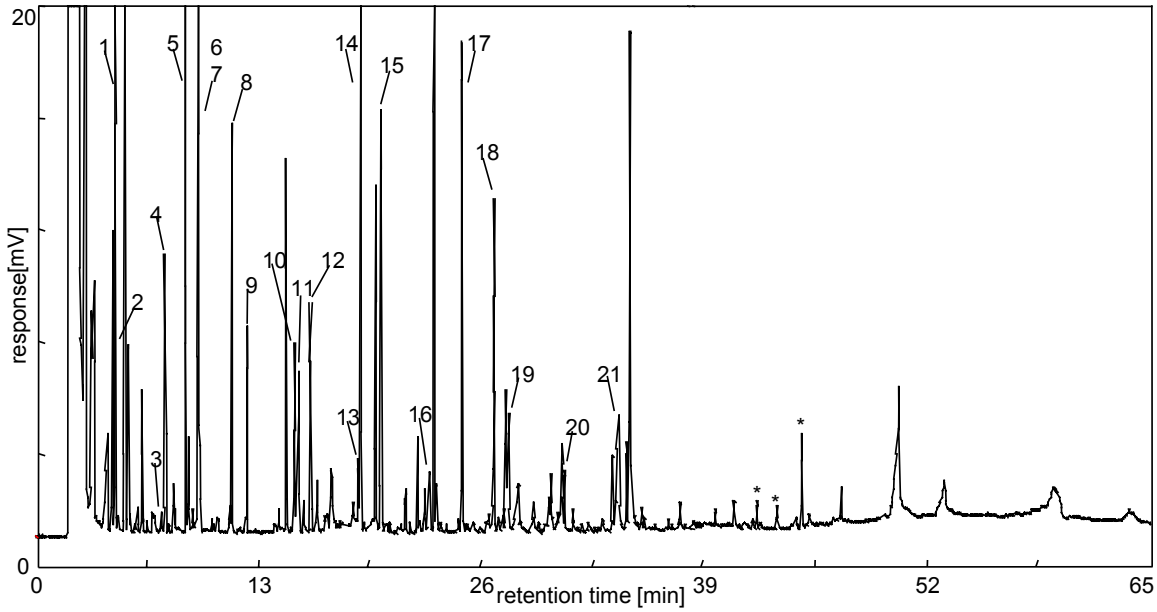

Figure 24: Gas chromatography using flame ionisation detection of fraction IV (four-fold enriched). For numbers see **Table 11**. For conditions see 3.2.1.6. Residual sugars are indicated by asterisks.

Table 11: Compounds identified by gas chromatography - mass spectrometry in fraction IV obtained from a rough rice sample.

no	compound	identification[a]	no	compound	identification[a]
amino acids and amines			**organic acids[c]**		
1	alanine (2TMS)[b]	1	5	phosphoric acid	1
2	glycin (2TMS)	1,2	7	succinic acid	1,2
3	urea (2TMS)	1	10	malic acid	1,2
4	valine (2TMS)	1	17	citric acid	1,2
6	proline (2TMS)	1,2	20	ferulic acid	1,2
8	serine (2TMS)	1			
9	threonine (2TMS)	1			
11	pyroglutamic acid (2TMS)	1			
12	aspartic acid (2TMS)	1			
13	phenylalanine (2TMS)	1			
14	glutamic acid (3TMS)	2,3			
15	asparagine (3TMS)	1			
16	glutamine (3TMS)	1			
18	allantoin (5TMS)	3			
19	tyrosin (3TMS)	1,2			
21	tryptophane (3TMS)	1			

[a]: Identification: **1**: according to NIST mass spectra library [80]; **2**: according to retention times of reference compounds; **3**: according to metabolite mass spectra library of Max Planck Institute of Molecular Plant Physiology [81]; [b]: number of trimethylsilyl (TMS) groups in parentheses; [c]: metabolites identified as persilylated derivatives.

Taking advantage of the fractionation procedure, a solvent (acetonitrile, [124]) different from that used for fraction III (pyridine) was applied in the course of re-silylation of constituents in fraction IV. Silylation of amino acids by means of MSTFA resulted in the formation of multiple derivatives differing in the number of trimethylsilyl (TMS) groups per compound. An increase in reaction temperature led to preferred formation of derivatives with a high number of TMS groups. However, complete persilylation of amino acids such as glutamine and lysine was not achieved. Therefore, conditions for trimethylsilylation of amino acids were optimised towards reproducible formation of the most abundant derivative yielded under mild conditions. Optimisation was performed by gas chromatographic analysis of amino acid mixtures and fractions obtained from rice which had been derivatised under varying conditions. Heating at 70°C for 15 min resulted in the formation of one predominant trimethylsilyl derivative from alanine, glycin, valine, proline, serine, threonine, pyroglutamic acid, aspartic acid, phenylalanine, and glutamic acid. Amounts did not change significantly after further heating for 45 min. For lysine, a maximum response was observed after 60 min. Consequently, heating for 60 min at 70°C was chosen as optimum condition for trimethylsilylation of rice constituents in fraction IV.

4.3.6 Quantification

Metabolite profiling studies aim to detect any relative changes in metabolite abundances in comparative experiments [65]. Therefore, comparison of entire profiles rather than absolute quantification of individual metabolites is of importance. For quantitative comparison of metabolite profiling chromatograms, normalised peak heights were used. Peak heights were preferred to peak areas since automated determination of heights is less error-prone than that of areas. For narrow peaks as observed in gas chromatography, peak heights correlate well to peak areas. Normalisation of peak heights in chromatograms obtained from rice was achieved by means of internal standards present in each fraction.

The methyl ester of margaric acid (C17:0) is frequently used as internal standard for quantification of fatty acid methyl esters as identified in fraction I. However, as shown in Table 8, small amounts of esterified margaric acid were found to occur

naturally in rice. Therefore, tetracosane was used instead as internal standard. It was not detected in fatty acid methyl ester fractions (fraction I) obtained from rice.
In accordance with a study reported by Toivo *et al.* [127], 5α-cholestan-3β-ol (dihydrocholesterol) was used as internal standard for fraction II which contains phytosterols as abundant constituents.
Based on a study described by Starke *et al.* [128], phenyl-β-D-glucopyranosid was applied to normalise peak heights in chromatograms obtained from fraction III which contains sugars as major constituents.
In fraction IV organic acids and amino acids are enriched. Consequently, an amino acid which does not occur naturally in rice (p-chloro-L-phenylalanine) was selected as internal standard.
Tetracosane / dihydrocholesterol and phenyl-β-D-glucopyranosid / p-chloro-L-phenylalanine were added to total lipid and polar rice extracts, respectively, and were completely recovered in the respective fractions. Co-elution with naturally occurring rough rice constituents was not observed.

4.3.7 Repeatability

Sufficient repeatability is one of the conditions to be met in order to apply the developed methodology to the differentiation of rice lines and cultivars. Data obtained for selected compounds belonging to various classes demonstrated the good intralaboratory repeatability of the fractionation procedures as well as of the overall metabolite profiling methodology involving extraction, fractionation, derivatisation and gas chromatographic analysis with flame ionisation detection. As shown in Table 12 for selected compounds, relative standard deviations (RSD) of normalised peak heights were equal or lower than 10 % and lower than those calculated for absolute peak heights. This proves the suitability of the selected internal standards for peak height normalisation.

Table 12: Repeatability of the metabolite profiling methodology, calculated for selected compounds

	RSD [%][a]		
compound	fractionation (n=5) (absolute peak heights)	overall method (n=3) (absolute peak heights)	overall method (n=3) (normalised peak heights)
fatty acid methyl esters			
C16:1 (cis 9)	12	7	1
C17:0	11	9	2
C23:0	8	5	7
hydrocarbons			
squalene	16	7	4
C29:0	13	6	3
free fatty acids[b]			
C18:0	13	8	4
C24:0	11	8	10
alcohols[b]			
C28:0	7	9	1
C32:0	12	10	3
sterols and triterpenic alcohols[b]			
campesterol	13	10	3
β-sitosterol	11	10	3
24-methylenecycloartanol	12	12	4
phenolic compounds[b]			
methyl ferulate	9	8	2
sugars[b]			
glucose	3[c]	4	1
sucrose	5[c]	6	1
raffinose	9[c]	6	4
amino acids[b]			
glycine	10	7	6
serine	9	8	5
glutamic acid	9	9	9
organic acids[b]			
citric acid	9	9	7
malic acid	11	9	5

[a]: RSD: relative standard deviation; [b]: Precision data were calculated for trimethylsilylated derivatives. [c]: n = 4

4.3.8 Recovery

A sufficient recovery of metabolites from rice material is a further prerequisite to apply the profiling method to the detection of differences between rice lines and cultivars. Recovery experiments were performed for compounds identified in fractions I - IV which belong to various classes.

As a first approach increasing amounts of freeze-dried rough rice flour were extracted and fractionated by the described procedure and analysed by GC-FID. The resulting peak areas were plotted against the flour amounts extracted (Figure 25). Extraction of the metabolites from 200 mg of rice flour with 4 mL of dichloromethane (lipids) and 10 mL of methanol/water 80+20, v+v (polar compounds) assures analysis of the respective compounds within the linear range of the method.

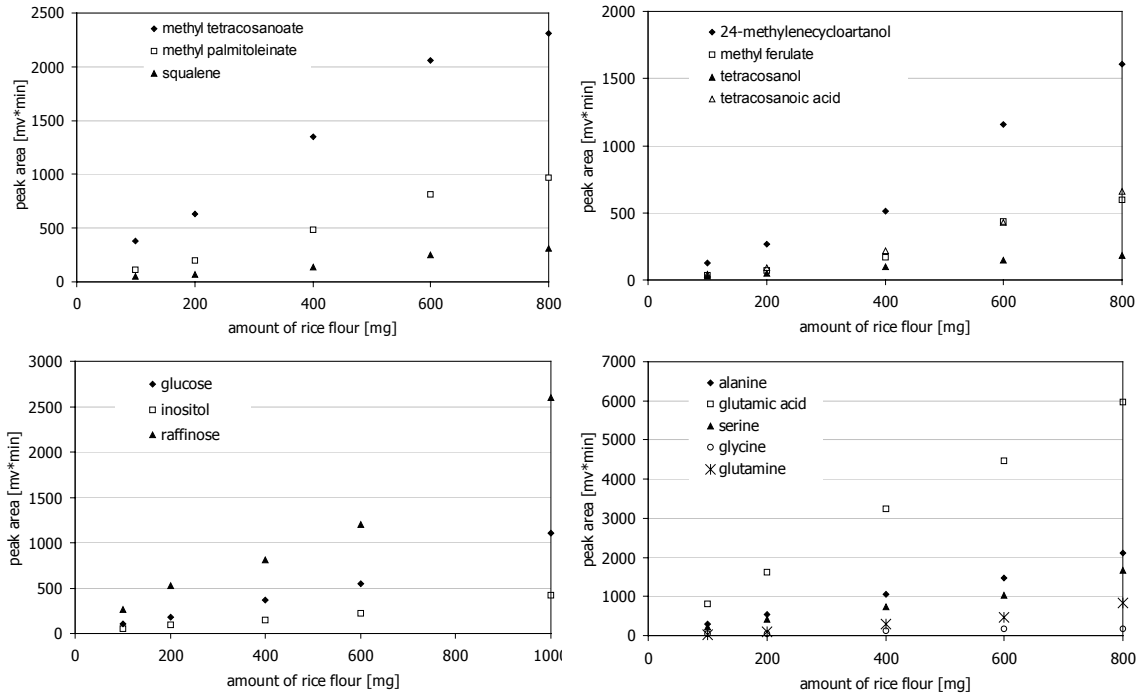

Figure 25: Peak areas of compounds identified in fractions I - IV vs. amounts of rice flour extracted

This recovery experiment proves that the developed metabolite profiling method can detect relative changes in concentrations of a broad spectrum of polar and unpolar rice constituents.

In order to determine the percentages of compounds recovered from rice material after extraction and fractionation, known amounts of standards were spiked at the start of the analytical procedure. Results gained by GC-FID analysis of fractions obtained from spiked samples were compared to those gained for fractions spiked

prior to GC analysis. Recoveries determined for selected compounds (sugars, organic acids, amino acids, triglycerides, hydrocarbons, fatty acids, sterols) ranged from 67 % (glutamine) to 99 % (glucose) and from 88 % (5α - cholestan-3β-ol) to 104 % (δ - tocopherol). This experiment demonstrates the suitability of the developed methodology to isolate compounds belonging to various classes from freeze-dried rice flour and to separate them quantitatively into different fractions. Recoveries determined for compounds used as internal standards were in the same range as those of rice constituents (Table 13).

Table 13: Mean recoveries of compounds belonging to various classes from rice determined for the metabolite profiling method

fraction/compound	mean recovery (n=2) [%]	fraction/compound	mean recovery (n=2) [%]
fraction I		**fraction IV**	
tripalmitin	93	glutamine	67
squalene	95	glutamic acid	84
fraction II		glycine	85
palmitic acid	98	lysine	91
δ-tocopherol	104	citric acid	76
stigmasterol	93		
fraction III		**internal standards**	
fructose	96	tetracosane	90
glucose	98	5α-cholestan-3β-ol	88
sucrose	82	phenyl-β-D-glucopyranosid	76
raffinose	70	p-chlorophenylalanine	85

4.3.9 Conclusions

The first prerequisite of an unbiased profiling approach is the use of a technique isolating a broad spectrum of compounds from the sample matrix. This in turn results in a complex extract exhibiting strong differences in concentrations of abundant constituents on one hand and minor but nutritionally or toxicologically relevant compounds on the other hand. The work-up of such a mixture must always be a compromise because clean-up procedures established for the targeted analysis of single compounds cannot be applied. The usefulness of a sub-fractionation of the

extract has to be evaluated in the light of the additional information gained and the input of work required.

The profiling method developed allows the separate isolation of lipids and polar compounds. For further fractionation two routinely applicable approaches were selected. For the lipid extract, column chromatography on commercially available cartridges proved suitable to separate major from minor constituents. For the polar compounds, a strategy based on the hydrolytic stability of silylated derivatives was modified and applied for the first time to an extract from plant material.

This sub-fractionation is more laborious than the direct analysis of total extracts as described previously [69, 77]. However, the procedure decreases the probability of co-elutions in the resulting gas chromatograms and thus founds the basis to use GC-FID as routine screening mode and GC-MS for final confirmation of differences detected. The method will allow valuable additional information especially on the minor constituents. It should be applicable to plants with similar distributions of constituents and the principles involved might serve as a basis for compositional analysis of other crops.

Trimethylsilylation turned out to be a universal method for the detection of a broad spectrum of compounds. However, the necessity to select the appropriate silylating reagent became obvious. Compounds exhibiting thermal instability and/or low volatility cannot be detected by this method. Profiling techniques based on liquid chromatography have the potential to close this gap [54].

The GC-MS identification of about 100 compounds is to be considered as a first step. The number of peaks detected in gas chromatograms obtained from rice by far exceeds that of peaks to which chemical structures could be assigned. However, following a non-targeted approach, the entirety of peaks detected has to be taken into consideration for data analysis. Coupling of the GC column to the mass spectrometer will allow to gain structural information on peaks which exhibited statistically significantly different abundances in comparative GC-FID analyses. The repeatability and recovery data determined indicate the potential of the metabolite profiling methodology to identify metabolic phenotypes, to differentiate rice cultivars, and to get an idea of the natural degree of variation in the metabolites identified. Owing to the complexity of the chromatograms obtained, automated data treatment and assessment will be required.

The screening of a broad spectrum of compounds combined with an effective separation of major and minor constituents may help to increase the chance to detect unintended effects due to the application of recombinant DNA techniques.

4.3.10 Summary

A metabolite profiling methodology was developed using rice as model crop. The approach is based on consecutive extraction of lipids and polar compounds and subsequent fractionations of both extracts. Transesterification/solid phase extraction (lipids) and selective hydrolysis of silylated derivatives (polar compounds) are applied to separate major from minor constituents. The method covers a broad spectrum of chemical classes. Preliminary investigation of the four fractions by GC-MS analysis of silylated derivatives resulted in the identification of more than 100 compounds. This unbiased non-targeted approach can serve as an additional analytical tool for the safety assessment of genetically modified crops and may help to increase the chance to detect unintended effects due to the application of recombinant DNA techniques.

4.4 Method for automated comparative analysis of metabolite profiling data

4.4.1 Introduction

Metabolite profiling aims to analyse the entirety of metabolites present in biological systems in a non-targeted way. In addition to the chemical - analytical techniques involved, which should be as universal as possible, methods for unbiased analysis of metabolite profiling data must be available. They have to offer the possibility to include the entirety of generated data for analysis. This means to extract relevant information from complex raw data files and to present results in a suitable form for further evaluation.

The evaluation of metabolite profiling data is based on a comparative approach taking into consideration a set of comparators and the natural variability. With increasing data complexity manual comparison of the profiles becomes time - consuming. The chromatograms obtained by the metabolite profiling method described in 4.3 are characterised by two-dimensional data structure, *i.e.* detector responses at corresponding time points. High data complexity is reflected by the detection of more than 100 peaks in a single run. The peak heights observed ranged over three orders of magnitude. The chromatograms were generally well resolved (see 4.3.4 and 4.3.5). However, patterns formed by groups of partially resolved peaks were still observed. Visual inspection of consecutively acquired chromatograms revealed retention time differences between corresponding peaks, which were frequently higher than a peak width (0.10 min). Software programmes for automated comparison of such chromatograms are available. They allow multivariate analysis of multiple chromatograms [129] and univariate peak-by-peak comparison of two similar profiles [130]. Because computer-assisted analysis of chromatographic data is mainly complicated by shifting retention times, the programmes involve methods for automated peak alignment.

The objective of this study was to develop a set of tailored tools which allows automated comparison of single as well as of multiple chromatograms using a univariate peak-by-peak approach. In order to meet the challenge of retention time instability, tools for automated and manual correction of retention time differences

should be included. This should lead to the generation of standardised chromatographic data suitable for establishment of databases and subsequent automated chromatogram comparison.

The performance of the software tools will be exemplarily demonstrated using model chromatograms obtained according to the procedure described in 3.2.1. Therefore, a minor lipid fraction was obtained from rough rice spiked with palmitic acid, δ-tocopherol, 5α-cholestan-3β-ol, and stigmasterol. A minor lipid fraction yielded from the unspiked rough rice served as the comparator.

4.4.2 Tools

An overview on the software tools developed is given in Figure 26.

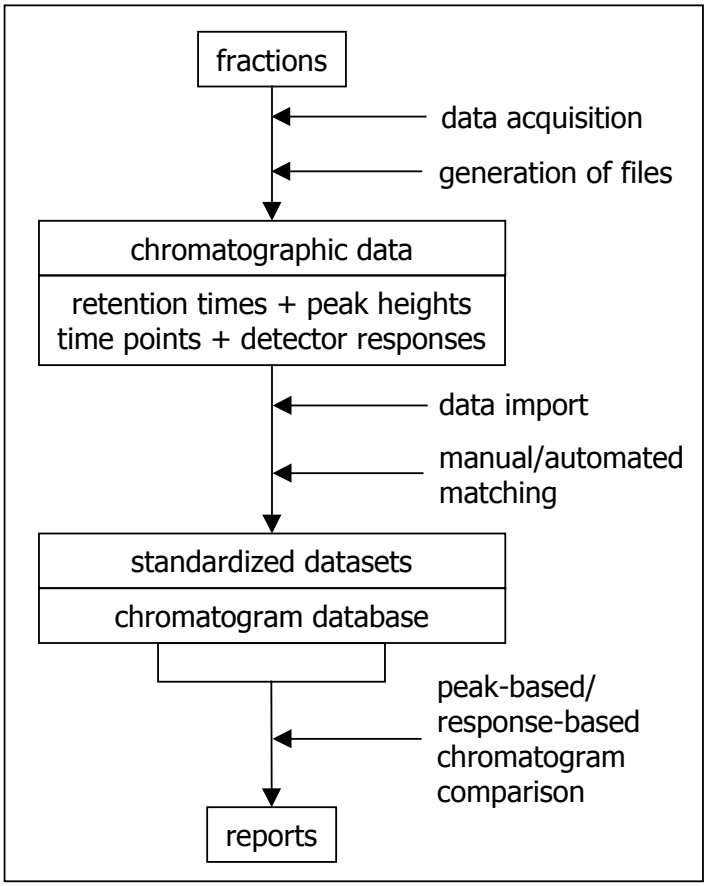

Figure 26: Approach for automated comparative analysis of metabolite profiling chromatograms.

The GC-FID data acquired are processed directly and after automated peak integration using commercially available software. This results in the generation of data files containing "detector responses + corresponding time points" and "peak heights + retention times", respectively. Tools were developed for automated and manual chromatogram matching. This allows transfer of standardised datasets into

databases and eventually automated chromatogram comparison. Tools were tailored for comparing the sample chromatogram to the reference. Chromatographic data calculated from databases can also be used as comparators. Chromatogram comparison is performed peak-based and response-based using "standardised peak heights + corrected retention times" and "normalised detector responses + corrected time points", respectively. Results of data analysis are summarised in graphical and tabular reports.

4.4.3 Standardisation of chromatograms

The tools developed for standardisation of chromatographic metabolite profiling data include automated and manual methods for correction of retention times, standardisation of peak heights/areas and normalisation of detector responses. The automated procedure is based on the use of standard compounds. A graphical user interface is available for manual standardisation of chromatograms.

4.4.3.1 Automated retention time matching and normalisation of peak heights

Standard compounds, *e.g.* hydrocarbons or carboxylic acids, are commonly applied for standardisation of retention times [131]. Considering the high complexity of the metabolite profiles obtained, a procedure based on the presence of five hydrocarbons (C_{11}, C_{16}, C_{24}, C_{30}, C_{38}) was implemented into the software. Thus, the methodology follows principles routinely used for electropherograms [132, 133]. Automated standardisation of retention times of all peaks detected in a sample chromatogram is achieved on the basis of two regression functions valid for early and late eluting peaks, respectively. They are calculated on the basis of the retention time differences observed for the standard compounds in the sample and the reference chromatogram [134].

Figure 27 illustrates the application of this automated approach to the model chromatograms. The hydrocarbons were added to the fractions prior to GC analysis. The chromatograms obtained from the spiked rough rice (sample chromatogram) and from the unspiked rough rice (reference chromatogram) are shown in Figure 28a/b. Automated integration using commercially available software provided lists of peak heights and corresponding retention times. Retention time differences observed for hydrocarbons in the sample and the reference chromatogram ranged from

-0.32 min (C_{11}) to +0.11 min (C_{38}). After using the self-tailored tool, the remaining retention time discrepancies to the reference ($\leq$ 0.03 min) were lower than one peak width (0.10 min).

For complex chromatograms determination of peak heights, which is less error-prone than that of peak areas, was preferred for quantification. In accordance with commonly applied standard procedures, normalisation of peak heights was achieved on the basis of the peak height ratio calculated for an internal standard peak in the sample and in the reference chromatogram. For the model chromatograms shown in Figure 28a/b, triacontane was used for peak height normalisation. Eventually, standardised datasets containing corrected retention times and standardised peak heights were obtained for the sample and the reference chromatogram.

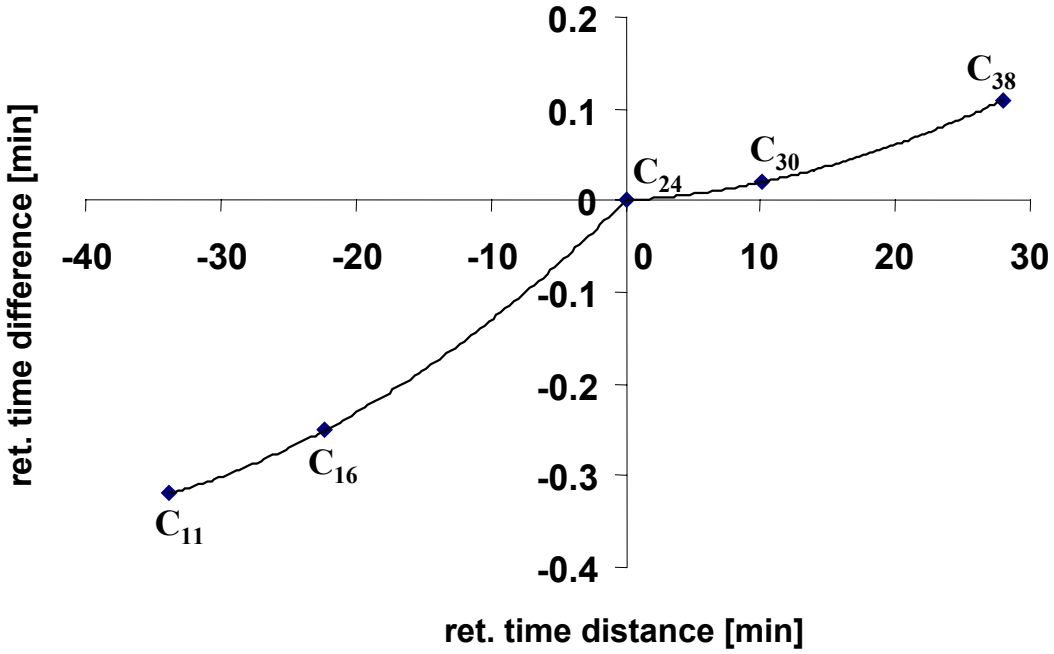

Figure 27: Graphs of quadratic regression functions calculated for automated retention time matching of a sample chromatogram to a reference chromatogram. Retention time distance: distances of the standard compound C_{24} to C_{11}, C_{16}, C_{30}, C_{38}, respectively, observed in the sample chromatogram; retention time difference: differences in retention times of corresponding standard compounds observed between the sample chromatogram and the reference chromatogram.

4.4.3.2 Manual chromatogram matching and normalisation of detector responses

In addition to the methodology for automated correction of retention times, a graphical user interface was developed which allows manual chromatogram matching. The tool provides a set of functions enabling manipulation of a sample chromatogram until it fits accurately to the reference, *i.e.* shifting, stretching,

compressing along the time/response axis, and baseline correction. Stretching and compressing along the time axis is automatically restricted to noisy chromatogram intervals. This ensures conservation of the peak shape in the course of chromatogram matching. In contrast to the automated methodology, matching accuracy can be controlled by eye.

This tool was applied to the model chromatograms obtained from the spiked (sample chromatogram) and the unspiked rough rice (reference chromatogram). It starts with lists of time points and corresponding detector responses which resulted from direct processing of the GC-FID data. An additional list containing retention times and corresponding peak heights obtained by peak integration was also imported. Starting from retention time differences ranging from –0.32 min to +0.11 min, manual matching led to a sample chromatogram fitting accurately to the reference (Figure 30). For response normalisation, the triacontane peak was used as the reference. Eventually, datasets containing normalised detector responses and corresponding corrected time points as well as corrected retention times and standardised peak heights were provided for the sample and the reference chromatogram.

4.4.4 Comparison of chromatograms

The objective of software development was to provide tools for comparison of the datasets generated by automated and manual standardisation of chromatograms. This faces the challenges of data complexity and of the non-targeted character of a metabolite profiling approach. Tools were developed for unbiased comparison of standardised datasets using a univariate peak-by-peak approach. Results are reported in graphical and tabular form.

4.4.4.1 Peak-based comparison of chromatograms

The tool tailored for peak-based comparison of chromatograms uses standardised peak heights and the corresponding corrected retention times for analysis. Comparison is performed peak-by-peak and results in a graphical and a tabular report. The graphical report indicates the magnitudes of percentage peak height differences and the height of additional peaks as dots along the retention time axis. The tabular report provides a rapid overview on the similarity of the comparators by

listing the number and magnitudes of differences and the heights of additional peaks.

Figure 28c/d shows the graphical report obtained by peak-based comparison of the standardised model chromatograms. It indicates two additional peaks and two peak height increases for the spiked sample compared to the unspiked reference chromatogram. The detection of two additional peaks in the sample chromatogram resulted from spiking with compounds which are absent (5α-cholestan-3β-ol) or present in very low concentrations (δ-tocopherol) in rough rice. The extra addition of naturally occurring rough rice constituents (palmitic acid and stigmasterol) led to the indication of percentage peak height differences.

These results are in accordance to those provided in the tabular report (Figure 29). Data complexity is reflected by the total number of peaks (307) detected in the two model chromatograms. In order to meet this challenge and to get meaningful results, the software tool allows to pass parameters influencing the outcome of data analysis. By setting a peak height threshold (2000 µV) and blank intervals (4 – 5 min, 30 – 37 min), very small peaks and overloaded intervals were excluded and 72 out of 307 peaks were taken into account for comparison. In combination with the adjustment of a detection threshold for percentage peak height differences (35 %), the four alterations due to sample spiking but not those owing to analytical variations were reported.

In the course of software development it became obvious, that appropriate setting of the parameter "width" was the most critical step. The parameter was introduced in order to manage the small discrepancies of corrected retention times between the sample and reference chromatograms. The parameter defines the maximum retention time difference acceptable for two peaks to be considered as comparators. Based on a high matching accuracy, a parameter value (0.03 min) lower than the width of the narrowest peak (0.10 min) ensured precise chromatogram comparison.

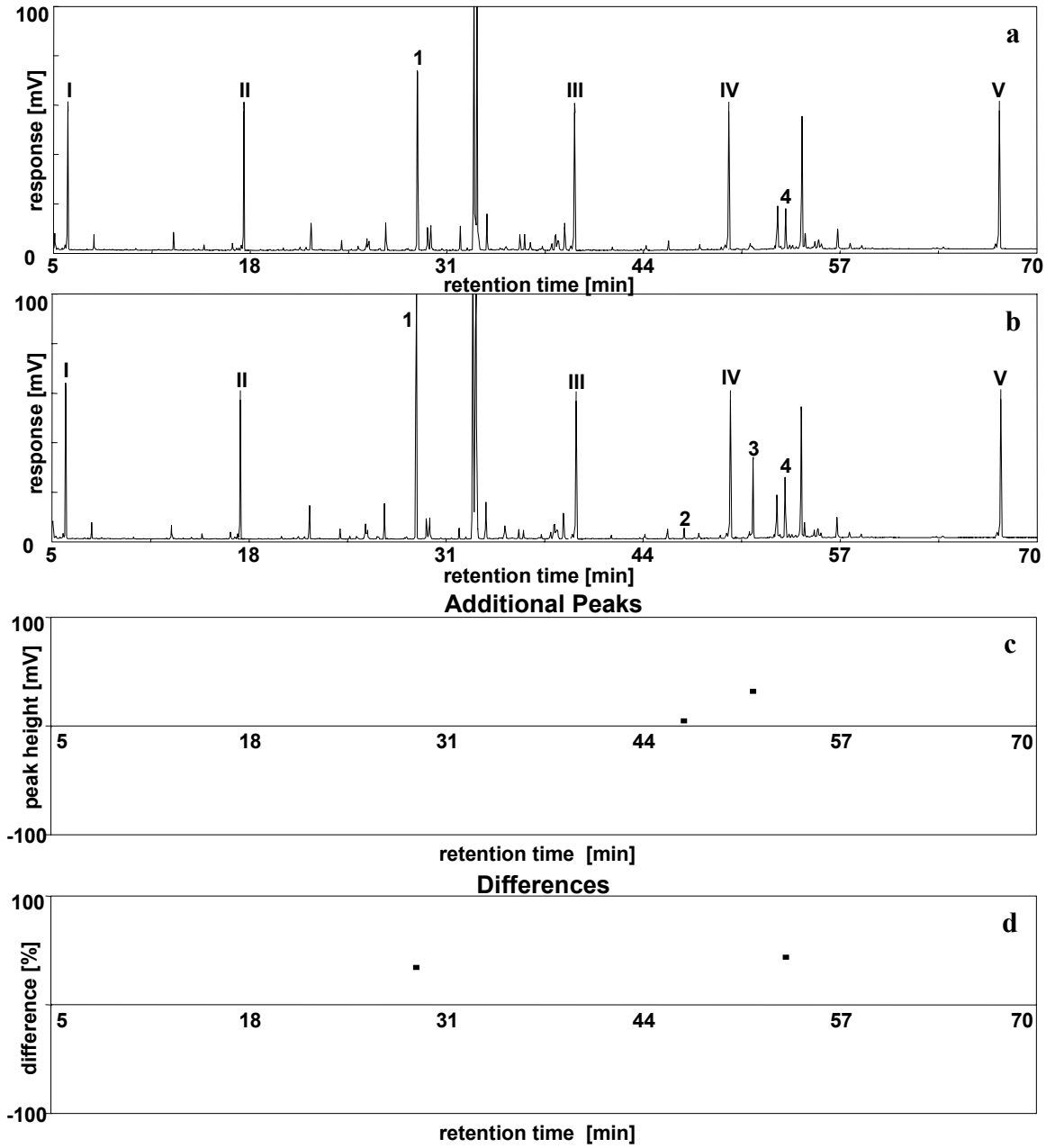

Figure 28: Automated peak-based comparison of model chromatograms. **a/b**: GC-FID chromatograms obtained from the minor lipid fractions (fraction II) of (**a**) an unspiked and (**b**) a spiked rough rice sample: **1**: palmitic acid-TMS, **2**: δ-tocopherol-TMS, **3**: 5α-cholestan-3β-ol-TMS, **4**: stigmasterol-TMS, **I**: undecane (C_{11}), **II**: hexadecane (C_{16}), **III**: tetracosane (C_{24}), **IV**: triacontane (C_{30}), **V**: octatriacontane (C_{38}). **c/d**: graphical report generated by peak-based comparison of the chromatograms; heights of additional peaks (**c**) and magnitudes of percentage peak height differences (**d**) are indicated as dots along the retention time axis.

statistics: comparison of	reference	and	sample
Parameters			
min. peak height:	2000 µV		
detect differences higher than:	35 %		
width:	0.03 min		
total peaks:	307		
peaks included:	72		
total alterations:	**4**		**8%**
Additional Peaks			
total:	**2**		
peak height: 100 - 500:	0		0%
501 - 1000:	0		0%
1001 - 2000:	0		0%
2001 - 5000:	1		50%
> 5000:	1		50%
Differences			
total:	**2**		
20 - 30 %:	0		0%
30 - 50 %:	2		100%
50 - 100 %:	0		0%
>100 %:	0		0%
Blanks			
Blank 1	4.00 min	to	5.00 min
Blank 2	30.00 min	to	37.00 min

Figure 29: Tabular report obtained by means of peak-based comparison of the model chromatograms shown in **Figure 28**a/b.

4.4.4.2 Response-based comparison of chromatograms

The objective of software development was to provide a method which allows comparison of chromatograms without prior peak integration. This was achieved by means of a self-tailored tool which uses the manually standardised raw data, *i.e.* normalised detector responses and corresponding corrected time points, for analysis. In contrast to the peak-based approach, selection of time points for comparison is based on an adjustable threshold value which refers to the first derivative (slope) of the chromatogram. This allows to take peak apexes and valleys (slope = 0) as well as shoulders (slope $\geq$ 0) into account for comparison.

Figure 30 shows a section of the graphical report generated by response-based comparison of the model chromatograms using a threshold value of zero. It consists of an overlay of the standardised comparators. The detection of an additional peak and of a percentage response difference resulting from spiking with 5α-cholestan-3β-ol and stigmasterol is indicated by filled circles. Points where no difference has been detected, *e.g.* at the campesterol peak, are labelled by empty circles. The tool allows

the adjustment of two parameters defining thresholds analogous to those used for peak-based comparison. By setting the response threshold to 3500 µV and the detection threshold for percentage response differences to 35 %, the results obtained by response-based comparison of the model chromatograms were in accordance to those gained by peak-based comparison. In Figure 30, the potential usefulness of the response-based approach is illustrated at the point labelled by an arrow. The point is part of a partially unresolved pattern, which is difficult to be integrated. The use of the response-based tool (slope = 0) instead of the peak-based methodology ensures appropriate consideration of the shoulder for comparison.

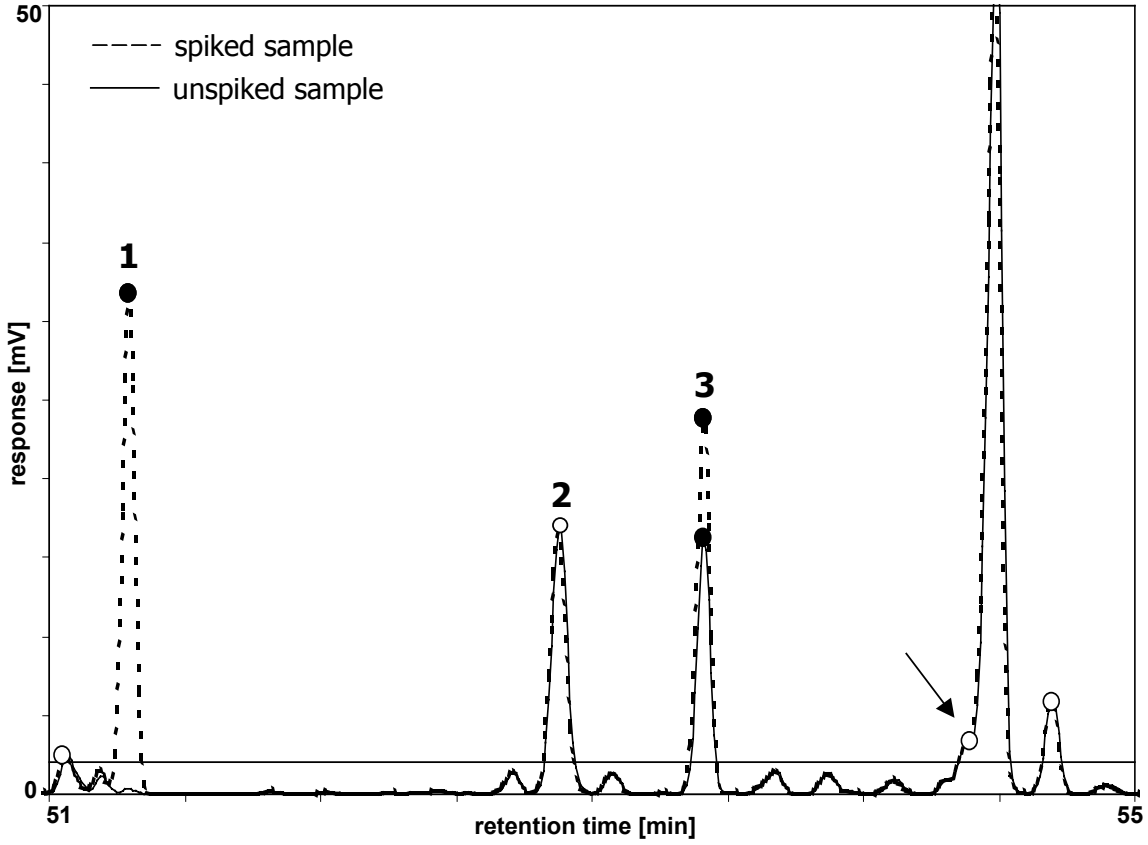

Figure 30: Graphical report generated by response-based comparison of model chromatograms obtained from rough rice and from a rough rice sample spiked with 5α-cholestan-3β–ol and stigmasterol. Points of the superimposed standardised comparators, *i.e.* peak apexes and a shoulder (see arrow), taken into account for comparison are labelled by circles. Responses lower than the response threshold (see horizontal line) were excluded. Detection of response differences is indicated by filled circles. Points where no difference has been detected, *e.g.* for campesterol-TMS, are labelled by empty circles: **1**: 5α-cholestan-3β–ol-TMS, **2**: campesterol-TMS, **3**: stigmasterol-TMS.

4.4.5 Concept for utilisation of chromatogram databases

The generation of a large number of standardised datasets requires the establishment of a chromatogram database. The database is an appropriate medium to store the datasets in a catalogued form for subsequent comparison.

In addition to the tools allowing the comparison of two chromatograms, procedures were developed for univariate comparative analysis of multiple chromatograms. This is achieved by the use of theoretical chromatographic data / theoretical chromatograms calculated from the database as comparators. Starting from a series of pre-selected peak-based datasets, theoretical chromatographic data, *i.e.* maximum, minimum, mean, standard deviation, are derived by combining all areas/heights of those peaks detected within a retention time window (parameter "width"). In order to obtain a theoretical chromatogram from a series of pre-selected response-based datasets, all normalised detector responses at a corresponding corrected time point are combined to calculate a mean, a minimum or a maximum.

Visualisation of the resulting theoretical rather than all single chromatograms proved to be advantageous for comparative analysis of large sample sets by eye.

Automated univariate comparison of the data obtained from two series of datasets can be performed by means of the same methodologies as described for single datasets. The use of multiple rather than single chromatograms as comparators offers the opportunity to assess the differences observed. Tools were developed which exclusively indicate statistically significant differences between two series of peak-based datasets. Furthermore, the software provides the flexibility to use theoretical data calculated by different methods as comparators (Figure 31). For example, comparison of a mean to a maximum chromatogram results in the detection of extreme differences between two series of datasets. This might be useful in order to assess differences between samples regarding their relevance in the light of the overall variability observed for a set of samples.

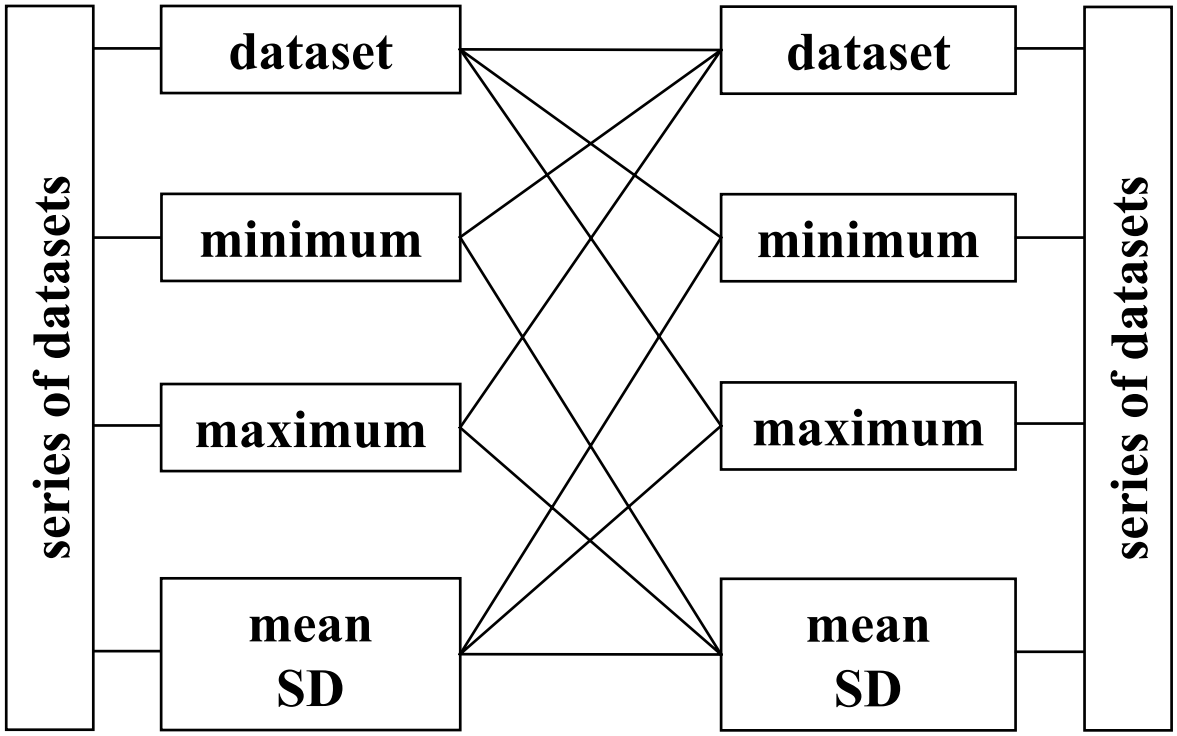

Figure 31: Concept for utilisation of chromatogram databases for comparative analysis of metabolite profiling data. Single datasets and theoretical chromatographic data / theoretical chromatograms (maximum, minimum, mean, standard deviation) calculated from a series of datasets can be used as comparators. SD = standard deviation

Another possibility to utilise a chromatogram database is to subject a series of pre-selected standardised datasets to multivariate comparison by means of principal component analysis, hierarchical cluster analysis, or discriminant analysis. Multivariate analysis of complex datasets is particularly attractive, because it reduces data complexity to information most useful to describe differences between samples. Graphical presentation of results gained by these techniques allows to obtain a rapid overview on the similarity of a huge number of datasets at a glance.

A tool was developed which transforms a series of pre-selected standardised peak-based datasets into a matrix suitable for multivariate analysis by means of commercially available statistics packages, *e.g.* [135]. The tool assigns a uniform retention time to all those peaks exhibiting corrected retention times within the same retention time window. The width of the retention time window can be adjusted by means of the parameter "width".

4.4.6 Conclusions

Metabolite profiling aims to provide a comprehensive picture of the low molecular weight constituents present in biological systems. The methods applied for analysis of the data obtained have to take into account the non-targeted character of such a screening approach. This leads to the necessity to start data analysis from minimally pre-processed data files.

The use of raw chromatograms [129] as well as of data obtained from a large number of pre-selected compounds [84] has been described for analysis of complex chromatograms. Multivariate and univariate statistics have been applied to evaluate metabolite profiling data [84].

The tools described here for comparison of metabolite profiling data are based on a non-targeted, univariate peak-by-peak approach. Considering the inherent unbiased character of the underlying metabolite profiling methodology and the large number of peaks to be expected, the tools were tailored to use the entire information of the chromatograms rather than a number of pre-selected peaks for comparative analysis. This assures equal treatment of data obtained from structurally identified, unknown or unexpected compounds. The application of univariate methods allows to directly assign differences between samples to compounds. The use of chromatogram databases and theoretical chromatographic data / theoretical chromatograms calculated thereof as comparators offers the opportunity to assess the differences observed.

The combination of tools developed for automated correction of retention times and automated peak-based comparison allows rapid assessment of the similarity of chromatograms. The tools tailored for manual matching and response-based data analysis provide the flexibility to compare chromatograms, if accurate peak integration is not possible. Manual matching controlled by eye is more time-consuming than previously described methodologies, which rely on algorithms for automated alignment of raw chromatograms [129, 130]. However, the manual procedure guarantees for accurate results, especially when highly dissimilar chromatograms exhibiting large retention time differences have to be matched.

The tools described here were developed for the non-targeted processing of GC-FID chromatograms. The underlying methodology should also be applicable to the analysis of chromatographic data with similar two-dimensional structure, *i.e.* detector

responses and corresponding time points. Consequently, the tools developed and the concepts applied should be useful in a wide area of applications, *e.g.* quality control and authenticity assessment. The importance of public-domain databases as a medium to store, catalogue and publish metabolite profiling data has been recognised [67, 136]. The tools described here might support efforts to establish such databases.

The non-targeted screening of a broad spectrum of compounds combined with the unbiased peak-by-peak analysis of chromatograms can serve as a powerful analytical tool for the safety assessment of genetically modified crops. This may improve the methodologies for detection and assessment of unintended effects due to the application of recombinant DNA techniques.

4.4.7 Summary

Software tools for comparative analysis of metabolite profiling chromatograms were developed. They include methods for manual and automated correction of retention times and responses by means of corresponding standards. This allows transfer of normalised datasets into databases and eventually automated comparison of chromatograms. Chromatogram comparison is based on a non-targeted, univariate peak-by-peak approach using peak heights / areas and corresponding retention times or detector responses and corresponding time points. Tools were developed for comparing single chromatograms. Comparison of multiple chromatograms is achieved by the use of theoretical chromatographic data or theoretical chromatograms calculated from databases as comparators. Results of data analysis may be summarised in graphic as well as in tabular reports listing the number of additional peaks and the magnitudes of (statistically significant) differences.

4.5 Profiling of systems with known effects on the metabolite level

4.5.1 Introduction

In chapter 4.3 a metabolite profiling method for simultaneous analysis of a broad spectrum of low molecular weight constituents in rice was described [137]. Tools for automated comparison of metabolite profiling data have been developed [138, 139]. In validation experiments it was demonstrated that the metabolite profiling method allows precise determination of compounds belonging to various classes in rice with high recovery. It was shown, that changes on the metabolite level due to spiking of low molecular weight compounds to rice can be detected by means of the self-tailored software.

Another approach to prove the reliability of the developed metabolite profiling methodology is to compare the outcomes obtained to those gained by validated standard protocols. Therefore, this chapter focuses on results yielded by comparative metabolite profiling of cereal samples exhibiting known differences on the metabolite level which had been detected by targeted investigations. Different compounds belonging to the four fractions which are isolated by means of the described fractionation methodology will be covered. Conventional rice samples with different contents of a nutritionally interesting minor lipid (γ-oryzanol) will serve as one sample set. In order to demonstrate the applicability of the developed methodology to other cereal crops, results obtained by the investigation of maize lines with differences in major lipids (fatty acid distribution) will be shown. Barley samples at different malting stages will be used to evaluate the suitability of the profiling approach for comprehensive high-throughput-analysis of polar and unpolar metabolites in metabolically active tissue. Results obtained by manual comparison of chromatograms as well as by means of the self-tailored software tools will be demonstrated.

4.5.2 Phytosteryl ferulates (γ-oryzanol) in rice

γ-Oryzanol comprises a mixture of phytosteryl ferulates located in rice bran. Cycloartenyl ferulate, 24-methylenecycloartanyl ferulate (Figure 32), campesteryl

ferulate, β-sitosteryl ferulate and campestanyl ferulate have been identified as the major components [125]. In addition to its technological usefulness, *e.g.* stabilisation of vegetable oils at frying temperature, physiological properties, such as antioxidative activity, and hypocholesterolemic effects have been reported for γ-oryzanol [140-142]. Therefore, γ-oryzanol represents a nutritionally interesting compound class which should be taken into consideration in the course of a safety assessment of GM rice. Specific methods for the analysis of γ-oryzanol in rice should allow the differentiation of the individual components, since studies indicate that the physiological effects depend on the composition of γ-oryzanol [141, 142].

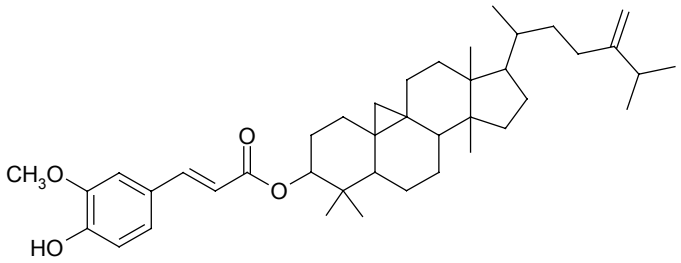

Figure 32: 24-methylenecycloartanyl ferulate – a major component of γ-oryzanol

Recent studies by means of a newly developed on-line LC-GC method revealed statistically significant differences in total γ-oryzanol contents as well as in γ-oryzanol compositions between conventional rice cultivars [143, 144]. Figure 33 exemplarily shows results obtained for rice cultivars *Cripto* and *Thaibonnet* by targeted on-line LC-GC analysis (A. Miller, personal communication). *Cripto* exhibited higher concentrations of total γ-oryzanol as well as of individual phytosteryl ferulates (campestanyl-, 24-methylenecycloartanyl-, and cycloartenyl ferulate).

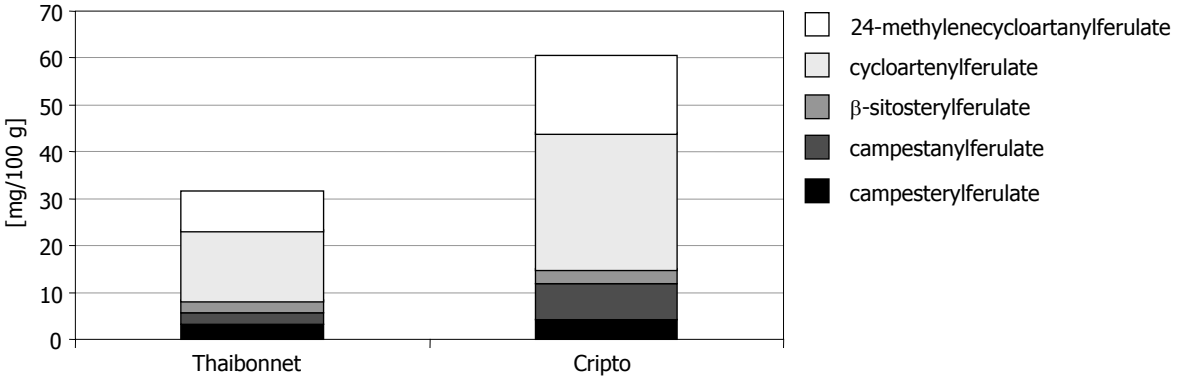

Figure 33: Mean total contents of γ-oryzanol and mean distribution of phytosteryl ferulates in rice cultivars *Thaibonnet* and *Cripto* (n = 6) determined by on-line LC-GC analysis.

In contrast to the specific on-line LC-GC method which allows GC analysis of intact phytosteryl ferulates, sample preparation according to the developed metabolite profiling method (see 4.3.4) involves conversion of γ-oryzanol into methyl ferulate and phytosterols by transesterification. Subsequent fractionation of transesterified lipids by silica gel column chromatography leads to the enrichment of cleaved γ-oryzanol components in fraction II (minor lipids). Figure 34 illustrates the usefulness of the fractionation approach for detection of methyl ferulate. Analysis of total lipids instead of individual fractions would result in co-elution of the minor compound methyl ferulate with methyl palmitate, a major lipid formed through transmethylation of triglycerides containing palmitic acid. This would render GC-FID analysis of methyl ferulate nearly impossible.

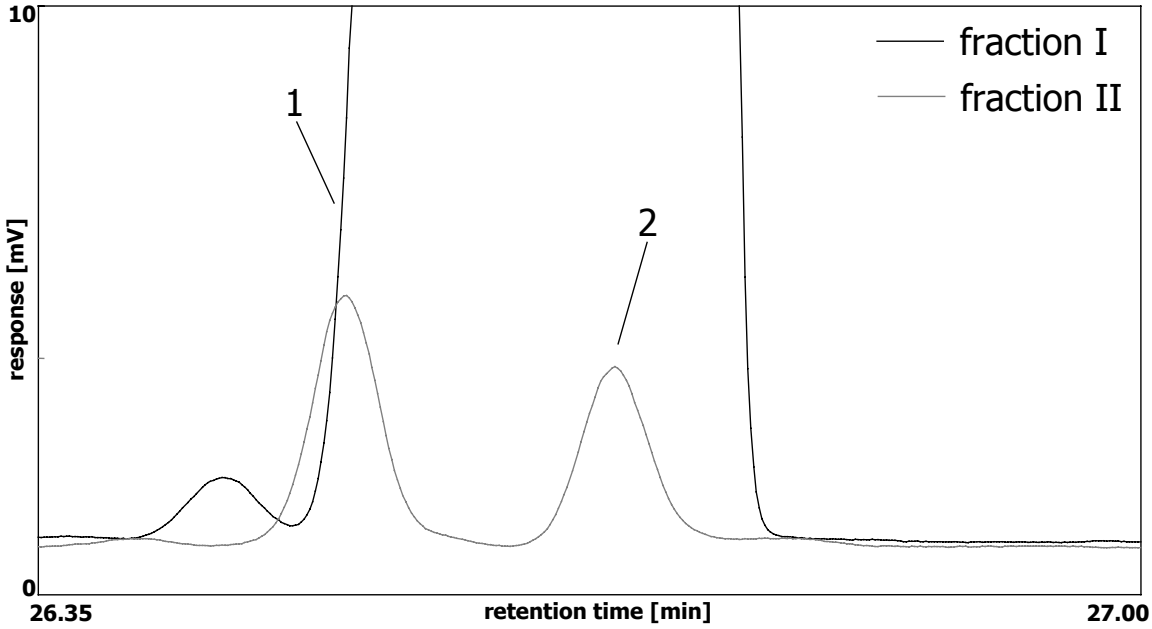

Figure 34: Overlay of GC-FID chromatograms obtained from lipid fractions I and II of the rice cultivar *Thaibonnet* by means of the metabolite profiling method. Sample preparation and GC analysis were performed as described in 3.2.1. **1**: methyl palmitate, **2**: methyl ferulate.

Figure 35A shows a chromatogram which resulted from the GC-FID investigation of a minor lipid fraction (fraction II) isolated from the cultivar *Thaibonnet*. The graphical report (Figure 35B) obtained by automated peak-based comparison of chromatograms yielded from minor lipid fractions of *Thaibonnet* (n = 3) and *Cripto* (n = 3) indicated a statistically significant difference in mean methyl ferulate abundances between the cultivars. Furthermore, the report shows a good correlation of the difference detected for the acid moiety (methyl ferulate) to those observed for two phytosteryl components derived from γ-oryzanol (campestanol, 24-methylene-

cycloartanol). This indicated the detected difference in methyl ferulate abundances to result from different γ-oryzanol concentrations in the cultivars.

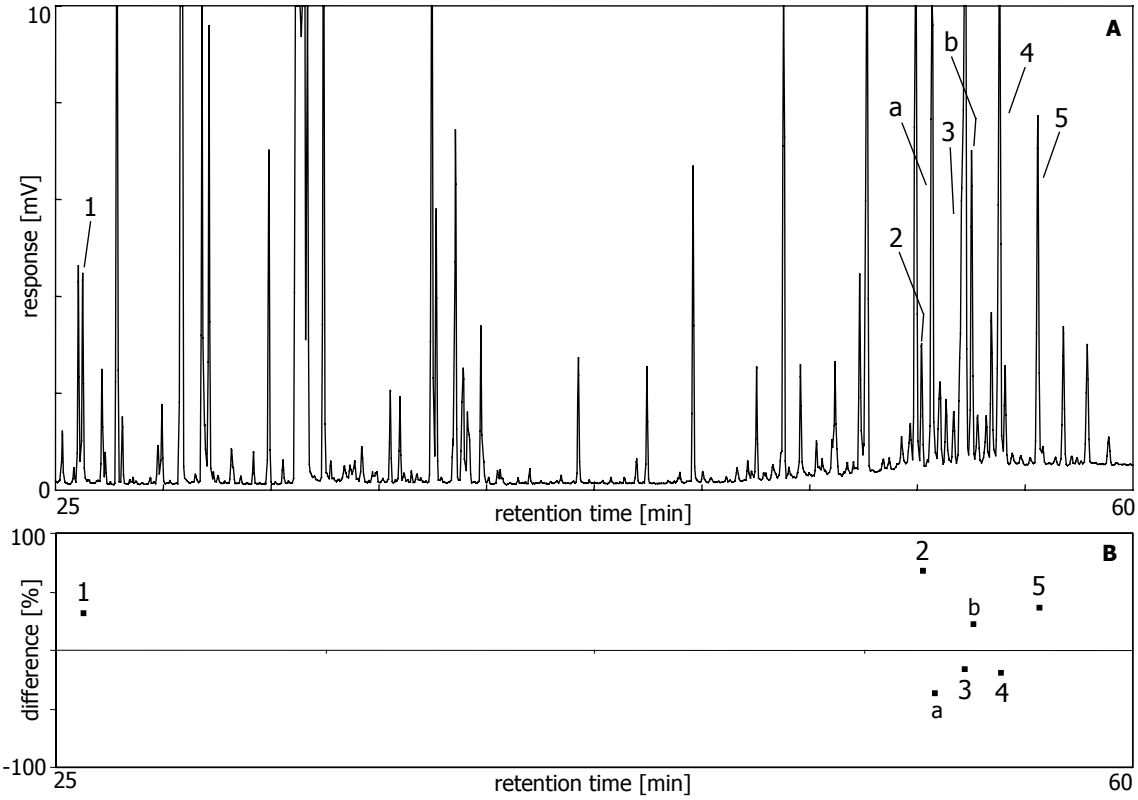

Figure 35: **A**: Metabolite profiling chromatogram obtained from the minor lipid fraction of the rice cultivar *Thaibonnet*. **B**: Graphical report obtained by automated peak-based comparison of chromatograms yielded from minor lipid fractions of cultivars *Thaibonnet* (n = 3) and *Cripto* (n = 3). Statistically significant differences between mean peak heights, indicated by dots along the retention time axis, were detected for methyl ferulate-TMS (**1**), campestanol-TMS (**2**), β-sitosterol-TMS (**3**), cycloartenol-TMS (**4**), and 24-methylenecycloartanol-TMS (**5**). Additional statistically significant differences were detected for stigmasterol-TMS (**a**) and Δ^5-avenasterol-TMS (**b**).

Results obtained by automated data analysis were confirmed by manual comparison of the metabolite profiling chromatograms (Figure 36). The profiling data obtained for methyl ferulate-TMS, campestanol-TMS, and 24-methylenecycloartanol-TMS were in good accordance to those gained by means of the specific on-line LC-GC method (Figure 33). This indicated the reliability of the developed metabolite profiling approach.

Quantification of methyl ferulate by means of the metabolite profiling method allows estimation of total γ-oryzanol content in rice. However, accurate determination of total γ-oryzanol contents is not possible since presence of methyl ferulate from sources different from γ-oryzanol [145, 146] cannot be excluded. The distribution of individual phytosteryl ferulates in rice cannot be deduced from metabolite profiling data since differentiation of free phytosteryls from those released by cleavage

through transesterification is not possible. This is illustrated in Figure 36 showing phytosteryl compositions which are different from those determined for the phytosteryl ferulate components of γ-oryzanol by means of the targeted on-line LC-GC method (Figure 33). Hence, evidence for differences in γ-oryzanol contents between rice cultivars obtained by means of the developed metabolite profiling approach should trigger detailed investigations by a specific method, *e.g.* the on-line LC-GC technique.

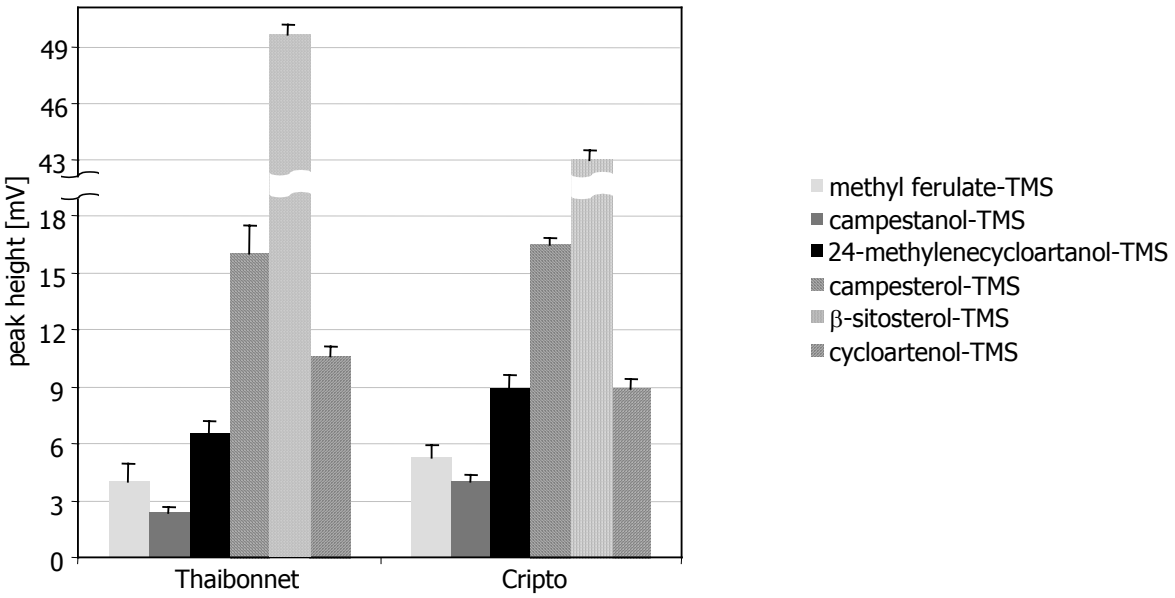

Figure 36: Mean abundances (n = 3) of methyl ferulate-TMS as indicator for total γ-oryzanol contents and of campestanol-TMS, 24-methylenecycloartanol-TMS, campesterol-TMS, β-sitosterol-TMS and cycloartenol-TMS formed from the phytosteryl ferulate components of γ-oryzanol. Data were obtained by triplicate analysis of rice cultivars *Thaibonnet* and *Cripto* by means of the metabolite profiling method. Error bars represent confidence intervals (error probability: 5 %).

4.5.3 Fatty acids in maize

One of the objectives of metabolite profiling method development was to involve sample preparation techniques which are adaptable to the work-up of other cereals. Therefore, the elaborated metabolite profiling methodology was tested for maize grains. Results will be exemplarily demonstrated for fatty acid methyl esters found in the major lipid fraction (fraction I). Maize grains are a valuable source of nutritionally essential fatty acids. Therefore, fatty acid distribution should be taken into consideration for safety assessment of transgenic maize.

Insect-resistance had been conferred to maize plants by introduction of a synthetic *cry1Ab* gene from *Bacillus thuringiensis*. Roß *et al.* [92] performed comprehensive targeted investigations in order to compare the fatty acid compositions of Bt-maize

grains (event 176) to conventionally bred material. Despite some statistically significant differences observed at single locations, evaluation of data collected from field trials in three seasons (1998 – 2000) at six locations in Bavaria led to the conclusion, that transgenic Bt-maize is substantially equivalent to conventional maize in terms of fatty acid distribution [92]. In order to prove the applicability of the developed metabolite profiling approach for analysis of maize grains, GM and parental lines found to exhibit differences in fatty acid contents were chosen for analysis.

Figure 37 exemplarily shows results obtained by means of targeted investigation of fatty acids in GM breeding line Navares and in its conventional counterpart Antares which had been grown in Neuhof in 2000. The application of a validated standard method for analysis [147] resulted in the detection of statistically significant differences in methyl myristate, methyl palmitoleinate, and methyl palmitate contents between transgenic and parental line.

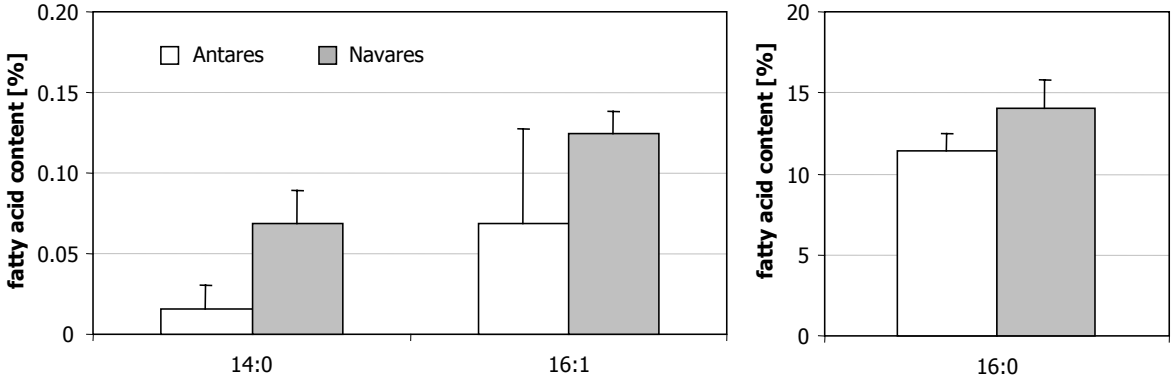

Figure 37: Contents of methyl myristate (14:0), methyl palmitoleinate (16:1), and methyl palmitate (16:0) expressed as percentages of total fatty acid contents in transgenic Bt – maize (breeding line Navares, n = 3) and in its conventional counterpart (Antares, n = 3) determined by means of targeted investigation of fatty acids using a validated standard protocol [147]. Error bars represent confidence intervals (error probability: 5 %).

These lines exhibiting equal contents of total fatty acids (data not shown) were analysed by the developed metabolite profiling method. Sample preparation according to the protocol described in 4.3 led to separate extracts containing lipids and polar compounds. Subsequent transmethylation of lipids under alkaline conditions resulted in the formation of fatty acid methyl esters from ester-bound fatty acids. In contrast to the method for targeted investigation of fatty acid distribution, the metabolite profiling protocol involves an additional fractionation step which led to the enrichment of fatty acid methyl esters in fraction I (major lipid

fraction). Triplicate isolation and subsequent GC-FID investigation of fractions I from Antares and Navares resulted in a total of six chromatograms which were compared by means of the self-tailored software tools.

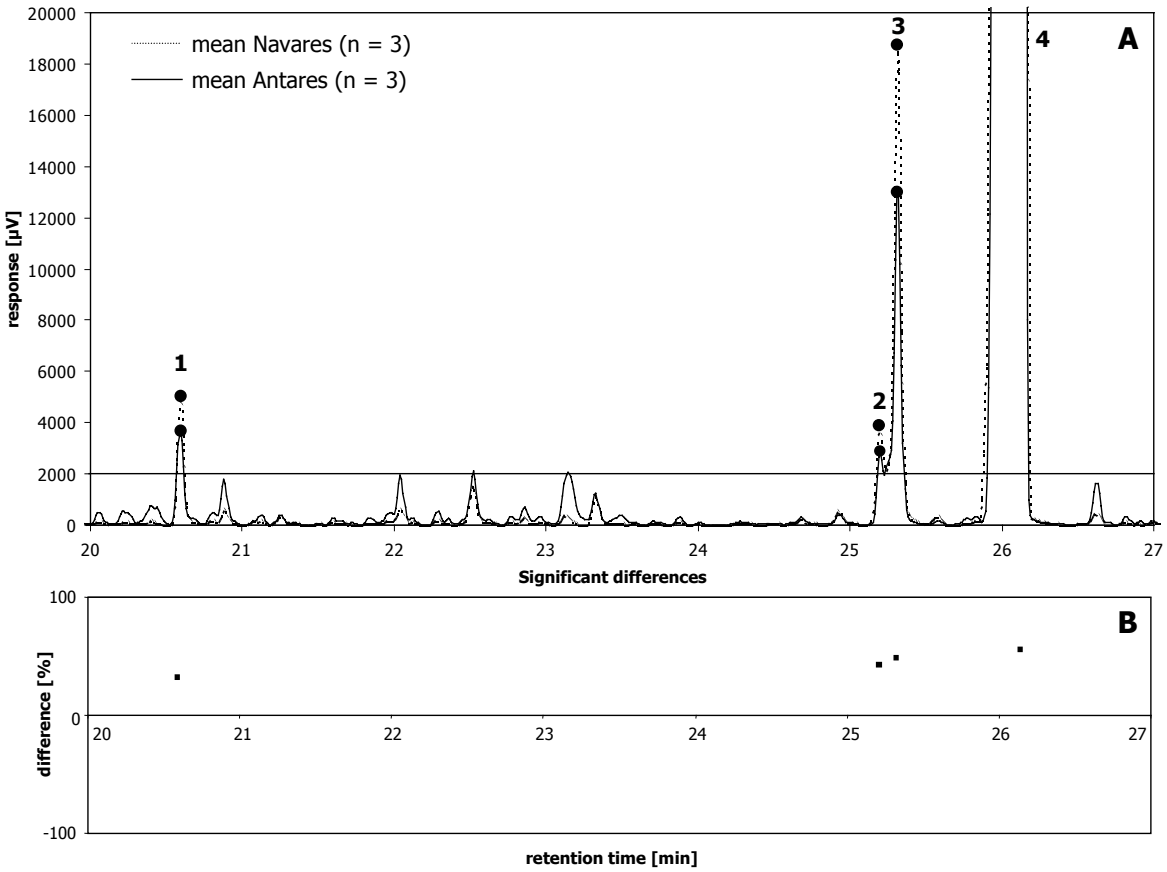

Figure 38: Reports generated by automated comparison of metabolite profiling chromatograms obtained by GC-FID investigation of fraction I isolated from maize lines Antares (conventional maize breeding line, n = 3) and Navares (genetically modified breeding line, n = 3). **A**: Graphical report generated by response-based comparison of means calculated from chromatograms obtained from Antares and Navares, respectively, after manual standardisation. Differences are indicated by filled circles for (**1**) methyl myristate, (**2**) methyl palmitolinoleate, (**3**) methyl palmitoleinate, and (**4**) methyl palmitate. **B**: Graphical report generated by peak-based comparison of mean peak heights calculated from chromatographic data obtained from Antares and Navares, respectively, after manual standardisation. Statistically significant differences are indicated by dots.

Figure 38A shows the graphical report generated by response-based data analysis. It contains an overlay of mean chromatograms calculated from manually standardised chromatographic data obtained from Antares and Navares, respectively. The availability of two matched mean chromatograms instead of six single chromatograms strongly facilitated analysis of the profiles by eye. Automated comparison of means led to the detection of differences in methyl myristate, methyl palmitoleinate, and methyl palmitate contents between transgenic and parental line which was in agreement with results gained by targeted investigation. Metabolite profiling of the maize lines revealed an additional difference for methyl

palmitolinoleate whose content was below the quantification limit of the standard method.

Results obtained by response-based data analysis were in full accordance to those gained by automated peak-based comparison of the six GC-FID chromatograms. The graphical report (Figure 38B) indicated statistically significant differences between mean heights for the same peaks which had been labelled in the response-based report (Figure 38A).

In summary, the developed metabolite profiling methodology was adaptable to maize grains and allowed comprehensive analysis of fatty acid distribution without any modification of the protocol developed for rice. The method involved for synthesis of fatty acid methyl esters is very similar to that routinely applied for targeted investigation of fatty acid content. Owing to an additional fractionation step, the metabolite profiling method has a lower quantification limit than the standard protocol.

4.5.4 Polar and unpolar metabolites in germinating barley

Germination results in profound changes in the levels of low molecular weight barley constituents. Therefore, analysis of barley grains at different malting stages represents a suitable positive control to test the developed metabolite profiling methodology. Because of its significant role in beer brewing, the malting process has been extensively investigated by means of classical analytical methodologies. Data on the time-dependent changes in the levels of sugars, amino acids, organic acids, and lipids are available from the literature [94].

In order to investigate the applicability of the developed metabolite profiling methodology to detect the expected effects on the metabolite level in the course of the malting process, barley grains at different malting stages were analysed by means of the protocol described in 4.3. Malting barley (cultivar Pasadena) was used as material. It was subjected to a standardised micromalting procedure [94]. Samples were taken after 5, 24, 48, 72, 96, 120, 144, and 168 hours (1^{st} – 7^{th} day). An additional sample was obtained after kilning (192 hours). Grains including the shoot were immediately frozen in liquid nitrogen and subsequently freeze-dried. Barley flour was analysed in accordance to the metabolite profiling protocol developed for rice. Owing to the expected high contents of polar metabolites in

malted barley, the amount of the polar extract subjected to fractionation was reduced compared to rice (see 3.2.3.3). Samples were investigated in triplicate and the resulting GC-FID chromatograms were analysed manually as well as by means of the developed software tools. Univariate and multivariate statistics were used for data evaluation. Barley constituents were identified by comparing retention times with those of silylated and methylated reference compounds or by comparing mass spectra obtained by GC-MS investigations with the entries of mass spectra libraries [80, 81]. Results on the time courses of barley metabolite levels during germination gained by means of metabolite profiling will be shown exemplarily for selected compounds in the following chapters. They will be assessed in the light of data reported in literature.

4.5.4.1 Amino acids / organic acids

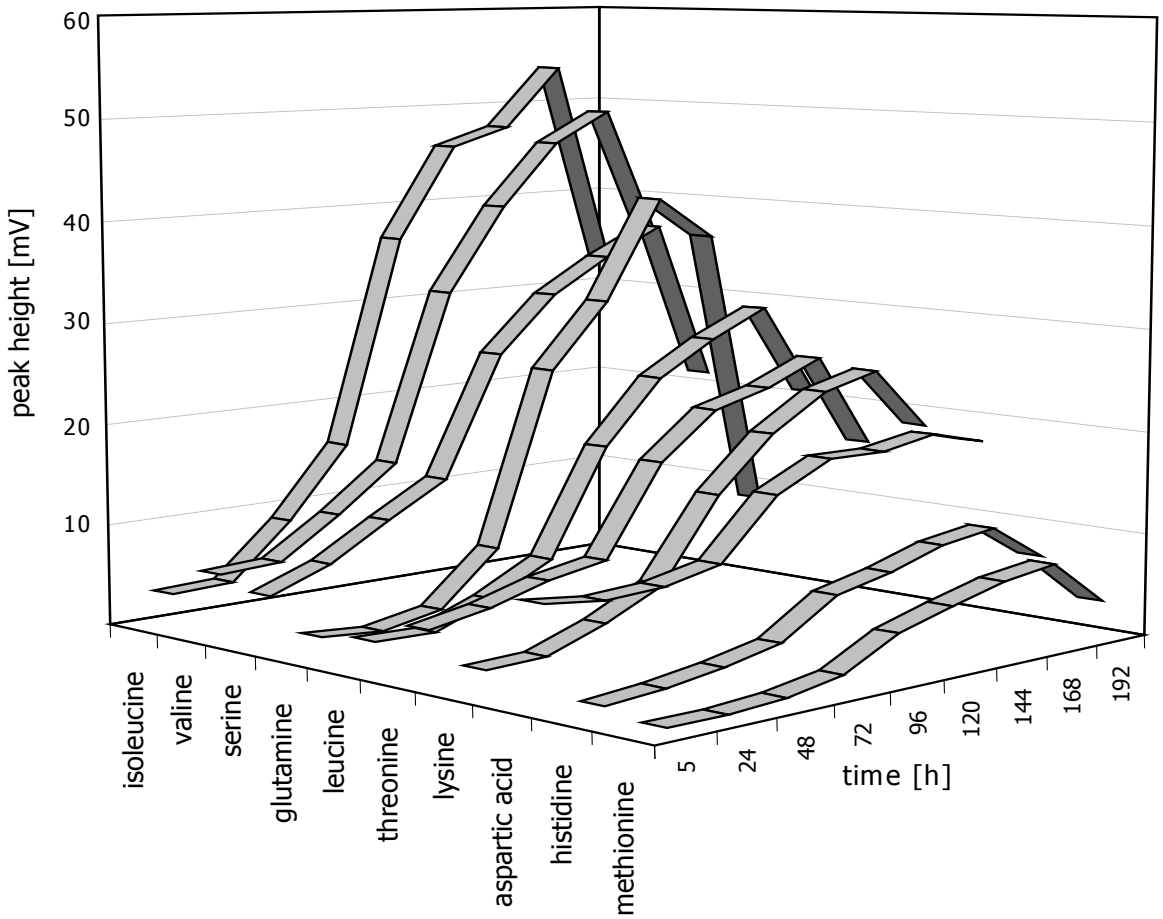

Figure 39: Time courses of the levels of selected amino acids during barley malting determined by means of metabolite profiling.

As shown in Figure 39, investigation of the amino acids-containing fractions of malted barley grains revealed statistically significant increases in the levels of lysine, valine, leucine, serine, threonine, aspartic acid, methionine, and glutamine in the course of the malting process.

These data obtained by means of metabolite profiling are in agreement with literature data determined with "classical" methods [94, 148-152].

At the beginning of micromalting, levels of free amino acids were shown to be relatively low. Biosynthesis of proteases during steeping and subsequent enzyme-catalysed cleavage of proteins at the start of germination resulted in an up to 18-fold rise (isoleucine) of free amino acid levels. The strongest increase was observed 2 – 3 days after start of the malting process. Formation of *Maillard* products led to the expected extreme decrease of free amino acid contents (up to 75 % for glutamine) during kilning [153].

Proline exhibited a time course which was qualitatively similar to those shown in Figure 39. However, in accordance to literature data [94, 154], proline content of germinated barley was found to be approx. three times higher than those of other free amino acids (Figure 40).

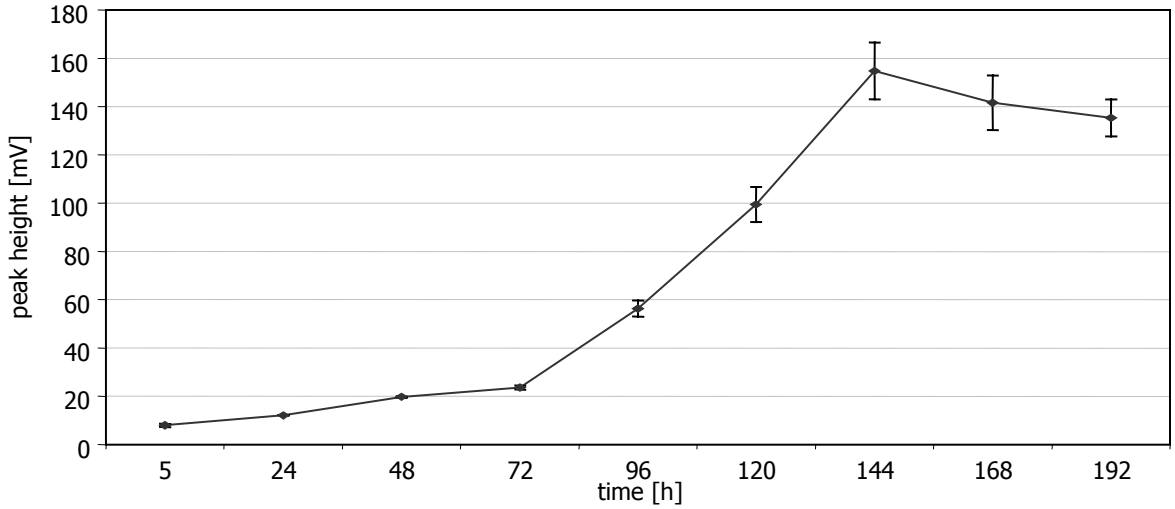

Figure 40: Time course of proline abundances during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

Free proline is liberated from proline-rich hordeins during germination and is essential for biosynthesis of root proteins. In order to meet the high proline requirement, additional quantities of this amino acid are produced by *de-novo* synthesis from glutamic acid [94, 154]. The observed slight decrease of proline content (5 %) during kilning was also reported in literature. It was explained by the

absence of a feed-back inhibition of proline formation. This leads to a steady rise of proline levels at the start of kilning, which counteracts the expected decrease during curing at high temperature [154].

According to literature data, malate contents in malted barley increase slightly during steeping and drop down steadily in the course of germination [155]. This was also observed when using the developed metabolite profiling method for analysis (Figure 41). The extraordinary time course of the malate level was explained by its role in the *Krebs*-cycle and in amino acid metabolism. During germination, malate serves as educt for biosynthesis of oxalacetic acid which is further transformed to aspartic acid. This results in a continuous decrease of malate contents in barley grains in the course of the malting process [94, 155-157].

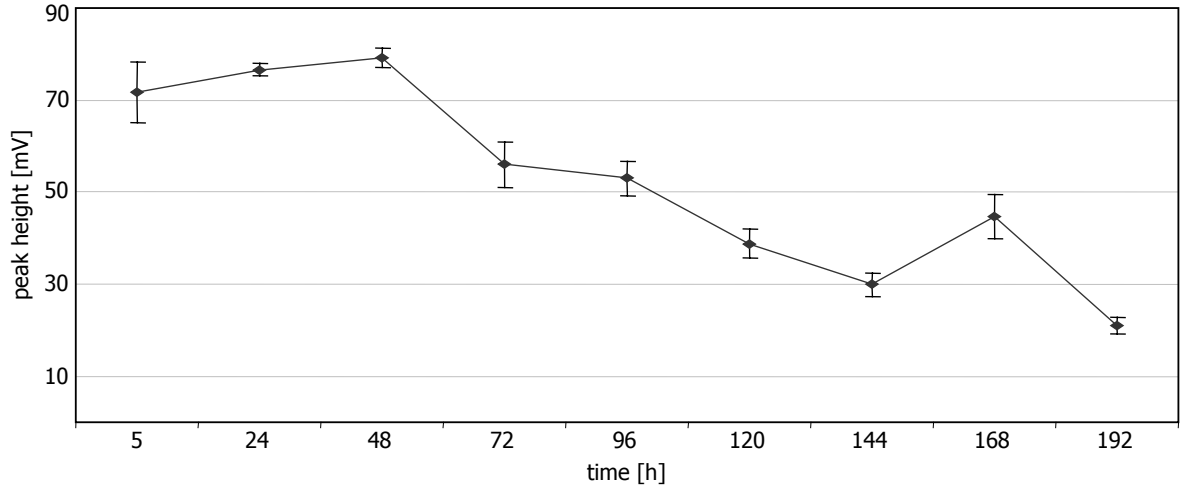

Figure 41: Time course of malate abundances during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

The homoserine level exhibited a time course during barley malting which was different to those discussed above. As shown in Figure 42, homoserine abundances were close to the detection limit of the metabolite profiling method from day 1 to day 7. Kilning resulted in a statistically significant increase of the homoserine content. These results are in agreement with the reported formation of homoserine by cleavage of S-methylmethionine during curing at high temperature [94] (Figure 43).

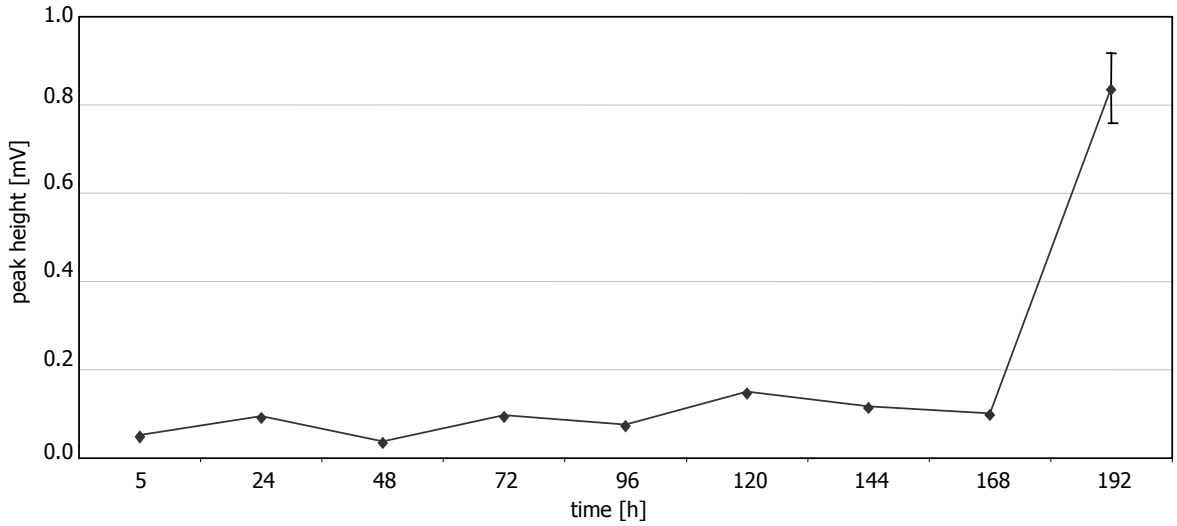

Figure 42: Time course of homoserine abundances during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

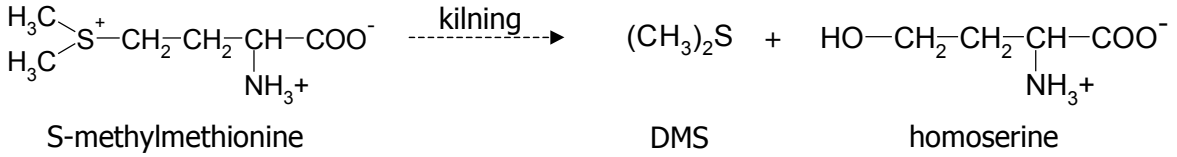

Figure 43: Cleavage of S-methylmethionine into dimethylsulfide (DMS) and homoserine at high temperature during kilning [94].

The additional homoserine peak in the chromatograms obtained from cured malt (Figure 44A) was also detected by automated data analysis using the developed self-tailored software tools (Figure 44B). This was achieved by peak-based comparison of mean chromatographic data calculated from GC-FID chromatograms obtained from cured malt (n = 3) to the maximum of a database containing standardised data of the 24 GC-FID chromatograms which resulted from the investigation of the green malt samples at different malting stages (day 1 to day 7, n = 3). Considering the fact that each of the 27 chromatograms contained approximately 100 peaks, ca. 2700 peaks were taken into consideration for automated comparison. The agreement of results gained by automated comparison with those obtained manually represents a successful positive control for the developed self-tailored software tools.

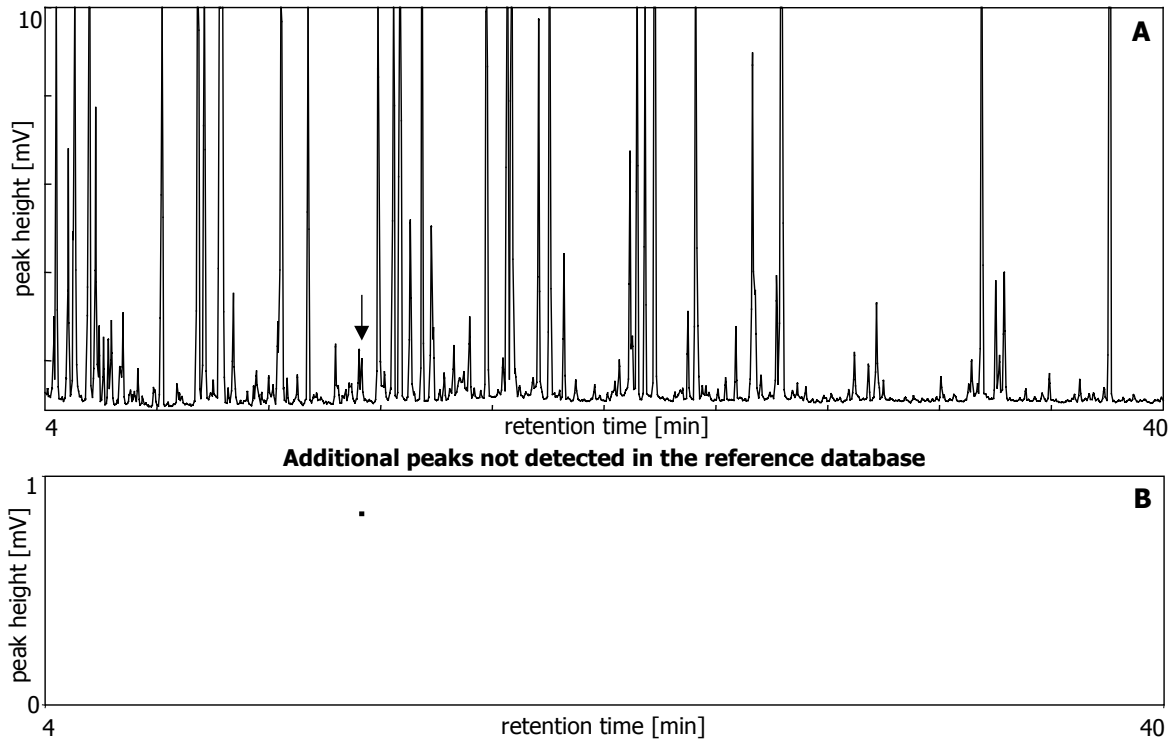

Figure 44: Data analysis of GC-FID chromatograms obtained from cured malt. **A**: GC-FID chromatogram obtained from fraction IV of a cured malt sample by means of metabolite profiling. For conditions see 3.2.3. The homoserine peak is labelled by an arrow. **B**: Graphical report generated by automated peak-based comparison of mean chromatographic data calculated from GC-FID chromatograms obtained from cured malt to the maximum of a reference database containing standardised data of 24 GC-FID chromatograms which resulted from the investigation of the green malt samples at different malting stages (day 1 to day 7, n = 3). The dot indicates the detection of a peak (homoserine) in chromatograms obtained from cured malt samples which was not found in any chromatogram stored in the reference database.

In addition to automated univariate peak-by-peak comparison, the 27 chromatograms obtained from the amino acid / organic acid fractions of barley at different malting stages were subjected to multivariate analysis. Using the self-tailored software tools, standardised peak-based chromatographic data were automatically transformed to a matrix which was subsequently imported into a commercial statistics package (XLSTAT, Addinsoft, France) for Principal Component Analysis (PCA). For graphical presentation of results, scores calculated for the first principal component were plotted against those calculated for the second one (Figure 45). Each chromatogram is represented by a dot.

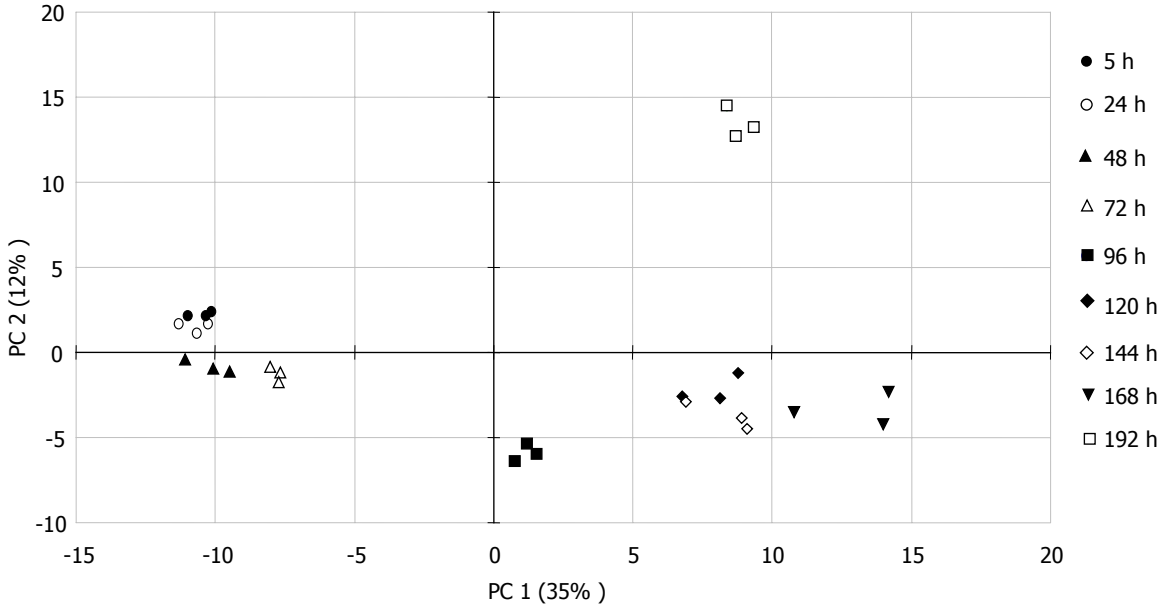

Figure 45: Principal component analysis of standardised peak-based chromatographic data obtained from organic acid / amino acid fractions of barley at different malting stages.

Chromatograms of organic acid / amino acid fractions from barley at different malting stages were grouped into four visually distinguishable clusters. Samples which had been taken on the first three days (5 h – 72 h) of the malting process are part of the first group. During that time, steeping leads to biosynthesis of proteolytic enzymes but not to extreme differences in metabolite levels (see Figure 39). The strong increase of amino acid contents on days 4 and 5 (72 h - 120 h) resulted in the formation of separate clusters comprising samples taken on these two days. The weaker rise of peak abundances after day 5 (120 h – 168 h) is reflected by the similarity of chromatograms obtained from the respective samples which group together with those taken on day 5. Groups comprising samples which had been taken before kilning (5 h – 168 h) were mainly separated through differences in scores of the first principal component. Kilning resulted in an extreme change in the amino acid / organic acid distribution. Consequently, chromatograms obtained from cured malt form a separate group which differs from green malt samples in the scores of the second principal component.

4.5.4.2 Sugars

Time courses of sugar metabolites in the malting process were determined by GC-FID investigation of fractions III obtained from barley at different malting stages. As regards the most abundant monosaccharides (arabinose, xylose, glucose, fructose), time courses determined by metabolite profiling were very similar to those observed

for free amino acids. Amylolytic and cytolytic enzymes biosynthesised during steeping catalyse the cleavage of polysaccharides which results in an extreme increase of monosaccharide contents (up to 20-fold for glucose) at the start of germination (48 h – 120 h). Kilning leads to a drop of monosaccharide abundances through formation of *Maillard* products at high temperature (Figure 46). Results obtained by means of the developed metabolite profiling methodology are in full accordance with data reported in literature. Fructose and glucose are mainly biosynthesised by cleavage of starch and sucrose. Arabinose and xylose are liberated from pentosanes located in barley hulls. The hexose - pentose ratio determined (ca. 100) is in agreement with the fact, that starch content in barley exceeds that of pentosanes by a factor of 100 [158, 159]. Glucose concentrations in germinated barley are higher than those of fructose owing to a slower utilisation of glucose for catabolic processes and the additional formation of glucose through maltose degradation [94].

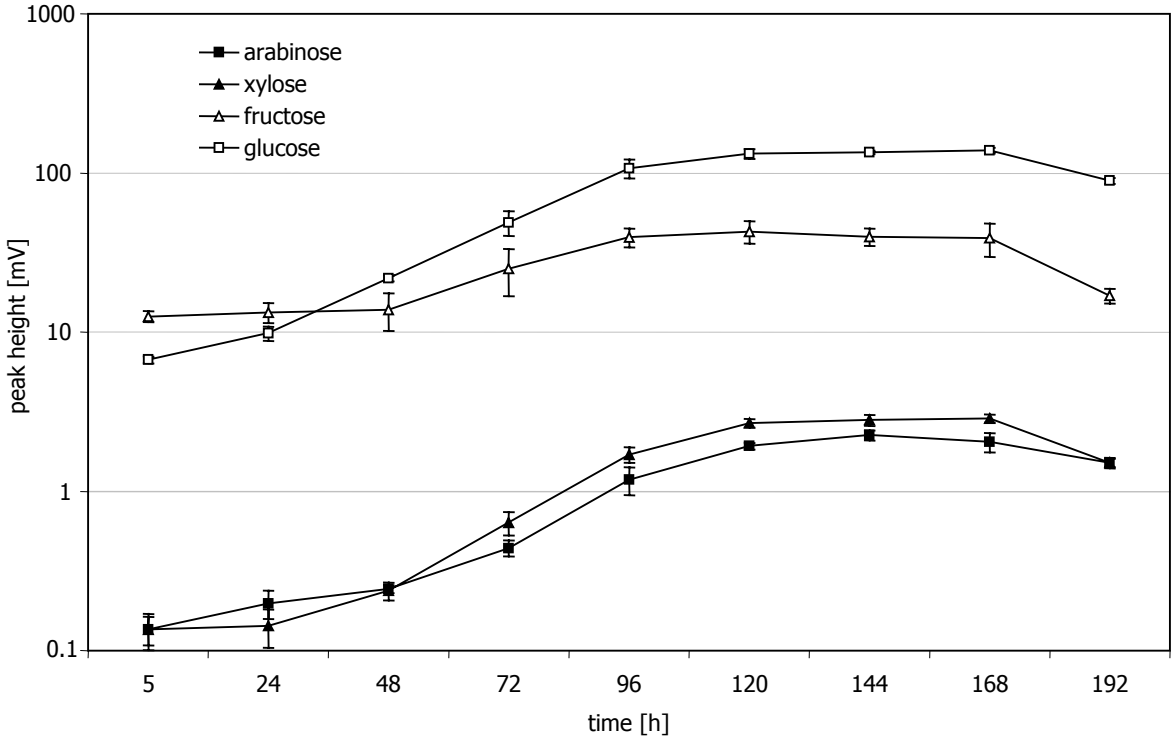

Figure 46: Time courses of monosaccharide abundances during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

Sucrose exhibited a time course which was different from those observed for the monosaccharides. Owing to the relatively low contents of glucose and fructose in barley grains, sucrose is utilised for catabolic purposes at the start of the malting process which results in a decrease of its concentration during steeping. With increasing availability of glucose biosynthesised by cleavage of starch via maltose,

sucrose contents rise steadily through *de-novo* synthesis from glucose [94]. This is even accelerated during kilning. Due to the fact that sucrose is not involved in the *Maillard* reaction its content is higher in cured than in green malt. According to [94], the maltose content of germinating barley increases significantly but slowly in the course of the malting process. This was explained by the cleavage of the disaccharide to two glucose molecules catalysed by maltase. As shown in Figure 47, results obtained by means of the developed metabolite profiling methodology were in agreement with the reported metabolic connections between glucose, maltose and sucrose.

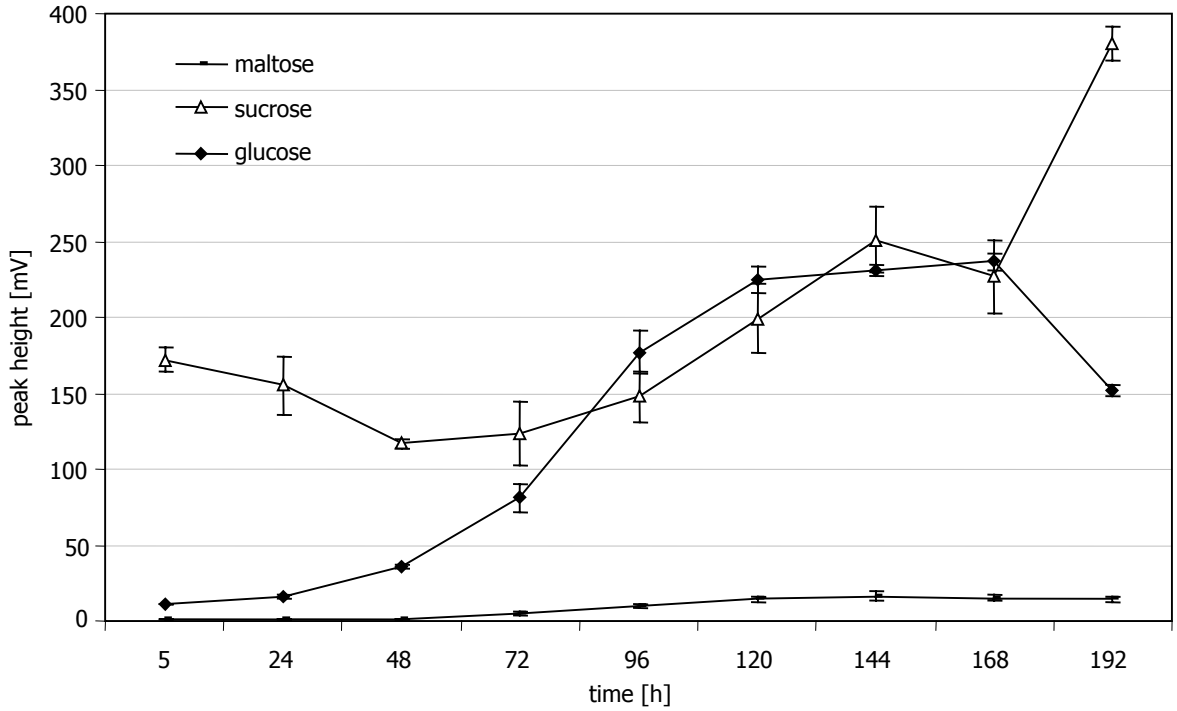

Figure 47: Time courses of maltose, glucose, and sucrose abundances during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

According to Figure 48, levels of raffinose drop rapidly during the first 24 hours of the malting process. Forty eight hours later the abundance of the raffinose peak was below the detection limit of the metabolite profiling method. This time course has been explained by the liberation of sucrose from the trisaccharide which is essential for energy supply at the start of the malting process [94, 158].

In accordance with the procedure described for organic acid / amino acid fractions (see 4.5.4.1), the 27 chromatograms obtained by the GC-FID investigation of sugars were subjected to multivariate analysis by means of PCA (Figure 49). In comparison to Figure 45, this resulted in an even better separation of samples at different

malting stages. The non-parallel time courses of mono-, di-, and trisaccharides led to visually distinguishable groups comprising samples taken after 48, 72, 96, and 120 hours. In agreement with the organic acid / amino acid fractions, chromatograms obtained from cured malt formed a cluster which differed in scores of the second principal component from those of green malt samples.

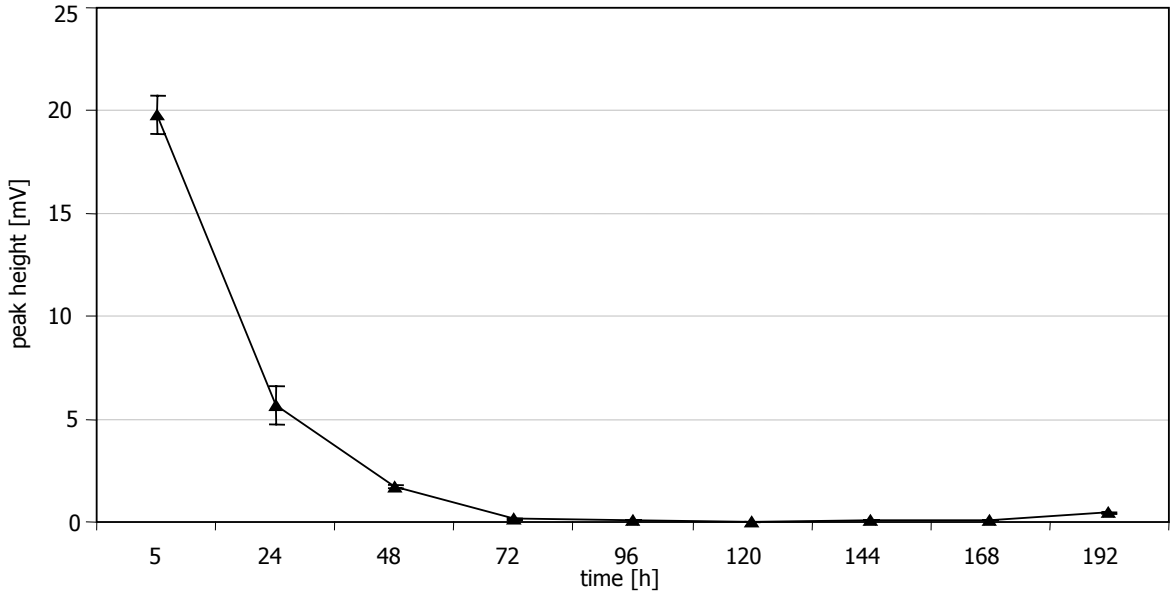

Figure 48: Time course of raffinose abundances during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

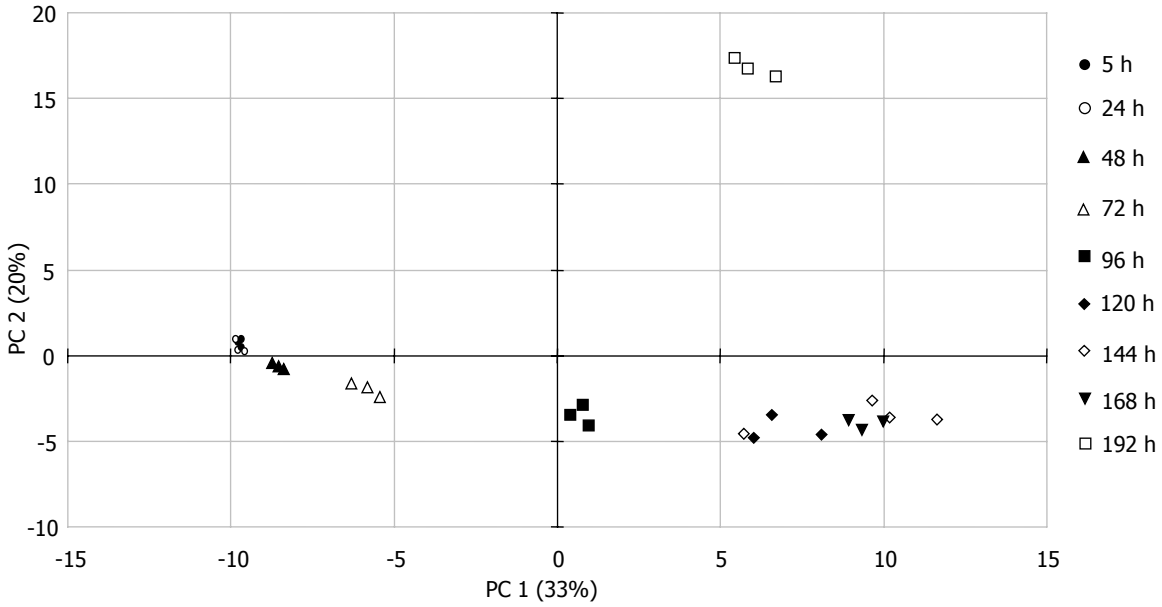

Figure 49: Principal component analysis of standardised peak-based chromatographic data obtained from sugar fractions of barley at different malting stages.

4.5.4.3 Lipids

Time courses of lipids during the malting process were determined by GC-FID investigation of fractions I and II obtained from barley grains at different malting stages. The lipid content of barley is relatively low (2 % - 3 %). Triglycerides, free fatty acids, mono- and diglycerides, phytosterols, phospholipids, and waxes account for approximately 90 % of total lipids [94]. Transesterification of the lipid extract which had been obtained in the same way as described for rice (see 3.2.1.3) and subsequent silica gel chromatography resulted in fraction I containing mainly fatty acid methyl esters derived from ester bound fatty acids and hydrocarbons. Free fatty acids and phytosterols are the major components of lipid fraction II obtained from barley grains.

Compared to the polar low molecular weight constituents (see 4.5.4.1 and 4.5.4.2) changes of barley lipids in the course of the malting process were relatively small. Significant changes of peak abundances similar to those observed for free amino acids and free sugars were hardly detected. One exception was the hydrocarbon pentacosane. As shown in Figure 50, peak abundances of this compound identified in fraction I steadily increased from the first to the sixth day of malting.

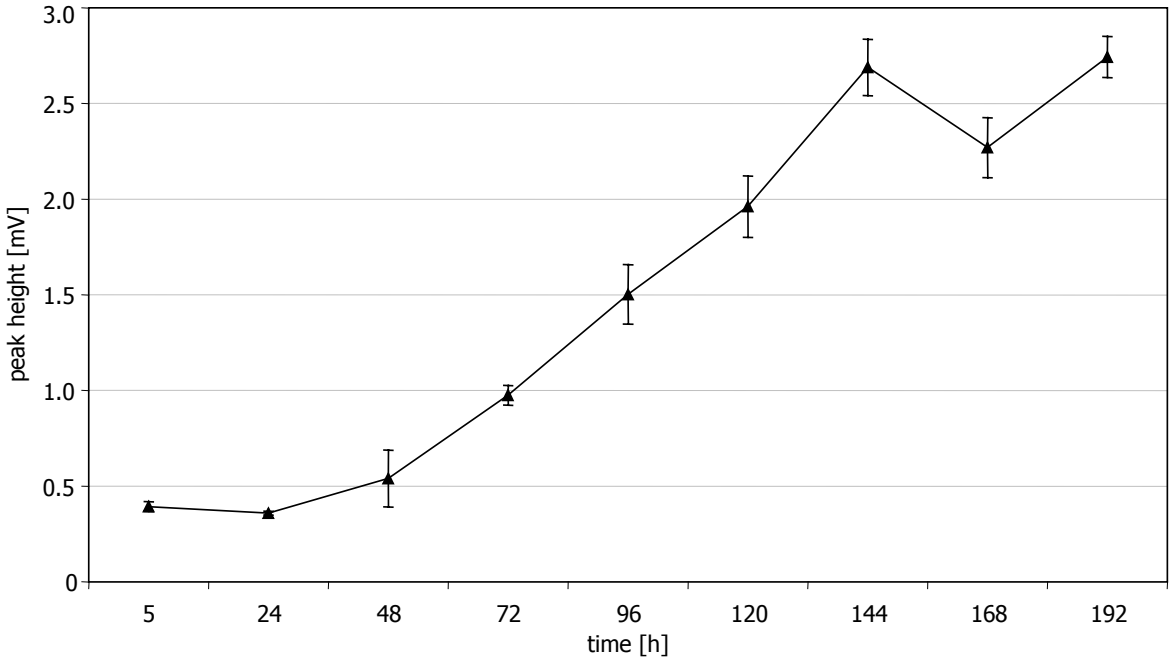

Figure 50: Time course of the abundances of the pentacosane peak during barley malting, determined by metabolite profiling (n = 3). Error bars represent confidence intervals (error probability = 5 %).

So far, *de-novo* biosynthesis of hydrocarbons in the course of barley malting has not been reported in literature. However, hydrocarbons are well-known constituents of starch lipids which are difficult to extract by means of dichloromethane [160]. Germination of barley involves cytolytic, amylolytic and proteolytic degradation of the grain matrix [161] which might result in an improved extractability of grain lipid constituents associated with polysaccharide molecules. This might be a possible explanation for the observed increase of the pentacosane abundance.

According to literature, germination of barley grains leads to a decrease of total lipid content by less than 25 % [94]. In addition to carbohydrates, free fatty acids cleaved by lipase-catalysis from triglycerides are used for catabolic processes. Moreover, fatty acids represent essential building blocks of cell membranes in the newly formed shoots and roots [94, 162]. Saturated fatty acids are preferentially liberated from triglycerides in the course of germination. In order to meet the need for unsaturated fatty acids, additional double bonds are included into saturated ones before build-up of new lipid structures.

Results obtained by means of metabolite profiling were in full agreement with the metabolic connections described above. Figure 51 exemplarily illustrates the detected decrease (-15 %) of the contents of esterified saturated fatty acids and the rise of free unsaturated fatty acids (+50 %) in the primary stages of the malting process, using stearic (Figure 51A) and linoleic acid (Figure 51B) as representatives for these compound classes.

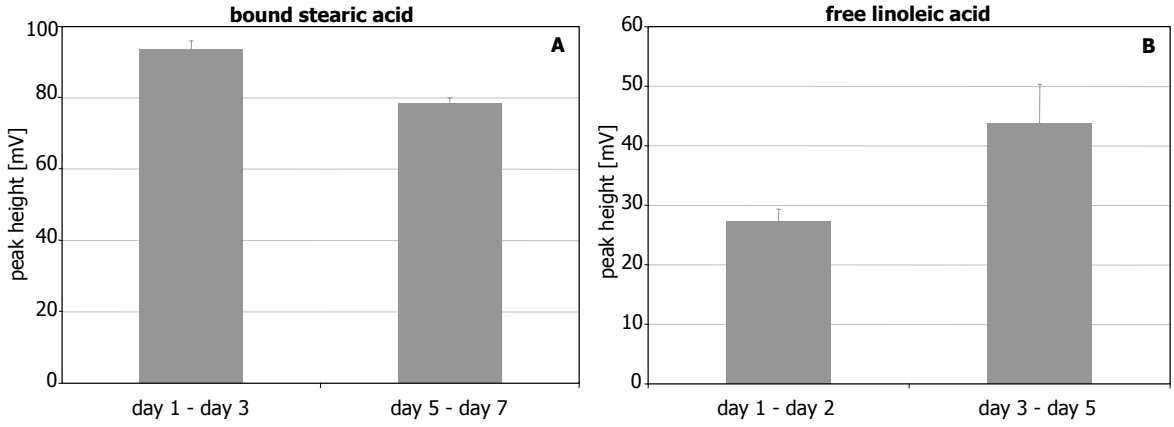

Figure 51: Mean contents of bound stearic acid (**A**) and free linoleic acid (**B**) in barley grains at different malting stages. Mean peak heights (n = 3) of methyl stearate (**A**) and linoleic acid-TMS (**B**) determined in GC-FID chromatograms obtained from barley grains at the different malting stages were averaged over the time periods indicated below the columns. Error bars indicate confidence intervals (error probability: 5 %).

In agreement with the minor changes in barley lipids detected by univariate comparison, multivariate analysis of the GC-FID chromatograms (fractions I and II) resulted in a less clear classification of samples according to different malting stages than those observed for sugars (fraction III) and organic acids / amino acids (fraction IV). However, as exemplarily shown in Figure 52 for fraction II, visually distinguishable groups comprising chromatograms obtained from early (5 h – 24 h, 48 h - 72 h) and late (96 h – 168 h) malting stages were formed. In contrast to the fractions containing polar metabolites, chromatograms obtained from malted barley after kilning were not clearly distinguishable from those which yielded from green malt.

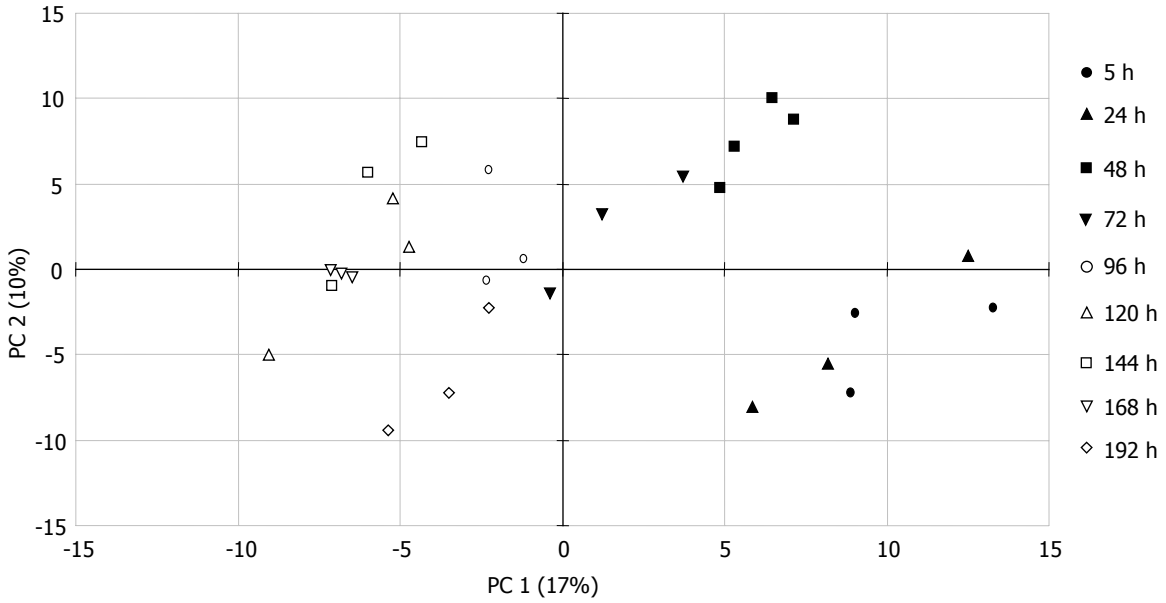

Figure 52: Principal component analysis of standardised peak-based chromatographic data obtained from lipid fraction II of barley at different malting stages.

In summary, results obtained by investigation of polar and unpolar metabolites in germinating barley were in full accordance with literature data. Based on a low experimental variability (high intra-laboratory repeatability), the metabolite profiling methodology developed originally for rice also allowed to follow the metabolic changes in germinating barley and to detect the expected effects on the metabolite level. It was demonstrated, that the combination of metabolite profiling and automated data analysis is suitable to detect extreme concentration differences (*e.g.* proline) and additional peaks (*e.g.* homoserine) in broad samples sets. Univariate data analysis was suitable to trace the time courses of individual metabolites. Multivariate data evaluation by means of PCA gave a rapid overview on the similarity

of the samples, taking into account the entirety of peaks detected in the chromatograms.

4.5.5 Conclusions

In order to evaluate the reliability of the developed metabolite profiling approach, the combination of GC-FID profiling and automated data analysis was applied to comparative analysis of samples with known differences on the metabolite level. Assessment of results has to consider the degree of accordance to reported data and has to take into account that a profiling approach must always be a compromise between comprehensiveness and specificity.

Results obtained by means of the new methodology were in all cases in agreement with the expected differences. Investigations involved low molecular weight compounds belonging to various classes (triglycerides, free fatty acids, phenolic compounds, organic acids, amino acids, sugars) in three different cereals (rice, maize, and barley). It was shown that the fractionation approach developed for rice is also applicable to crops with similar distribution of major and minor polar and unpolar constituents. GC-FID/MS was suitable for comprehensive qualitative and quantitative analysis of silylated / methylated metabolites. In contrast to specific methods for analysis of single compounds which usually involve a calibration procedure, differences were detected by comparison of peak heights. In all cases, analytical variability was lower than the magnitude of differences in metabolite abundances. In combination with its applicability to constituents in a broad concentration range (three orders of magnitude), the metabolite profiling method allowed the detection of small (-15 %) as well as of extreme changes (20-fold increase) in metabolite levels. However, in order to further confirm such differences, the amounts of metabolites in a sample should be derived from detector responses on the basis of a calibration equation.

γ-Oryzanol in rice served as an example for a nutritionally interesting compound class which is very difficult to investigate even by specific analytical methodologies. It was demonstrated that metabolite profiling can give an indication for changes in total γ-oryzanol contents. However, owing to limited specificity of the profiling approach, hyphenated techniques (on-line LC-GC) must be applied for confirmation and for detailed investigation of the distribution of individual γ-oryzanol components.

As regards analysis of esterified fatty acids in maize grains, principles involved in the metabolite profiling method are similar to those applied for targeted determination of fatty acid distribution. In contrast to the standard protocol, the profiling approach includes an additional purification step and uses an unpolar GC column for chromatographic separation of fatty acid methyl esters. This results in a low detection limit and allows additional analysis of long chain fatty acids (> 22 carbon atoms) as well as of hydrocarbons. On the other hand, the targeted approach was superior to the profiling method in terms of fatty acid differentiation according to the configuration of double bonds (*cis* / *trans*) and regarding separation of abundant components (18 carbon atoms) according to the degree of unsaturation. This illustrates the above-mentioned compromise between comprehensiveness and specificity.

Metabolite profiling of low molecular weight constituents in germinating barley served as example for non-targeted investigation of a broad spectrum of compounds in metabolically active tissue. Compared to classical methods for malt analysis, the newly developed methodology was advantageous in terms of rapidness and comprehensiveness. Considering the fact that expected changes on the metabolite level could be detected, the experiment represents a convincing positive control as regards the application of the profiling approach to the detection of unintended effects. As regards pentacosane, the non-targeted approach revealed changes which had not been reported before. However, further specific investigation of hydrocarbons would be required in order to confirm that differences in peak abundances are really due to different contents in grains and not to differences in extractability from grains at different malting stages.

Application of the developed metabolite profiling method resulted in huge amounts of data. Based on robust standardisation procedures, the self-tailored software tools for comparison of chromatograms were successfully applied for automated data analysis. As demonstrated for investigation of fatty acid distribution in maize grains, (direct) response-based treatment of chromatographic data proved to be useful for comparison of GC chromatogram series by-eye. However, based on the applied fractionation procedure and the resulting sufficient resolution in the GC-FID chromatograms obtained, the more rapid peak-based approach was preferred for data analysis. The ability of the software tools to detect statistically significant

differences between two series of chromatographic data sets turned out to be extremely useful for data evaluation. The possibility to establish chromatogram databases and to calculate theoretical chromatographic data thereof provided the basis for assessment of differences.

The detection of homoserine in chromatograms obtained from cured malt served as example for identification of a new peak not present in the huge number of chromatograms yielded from green malt at different malting stages. Comparison of the maximum chromatogram calculated from the database for green malt samples to the chromatograms obtained from cured malt directly led to the detection of the unique peak.

Automated univariate comparison of the chromatograms obtained from rice samples different in γ-oryzanol content revealed the usefulness of the generated graphical report for elucidating the role of the compounds underlying the difference. The good correlation between the changes in methyl ferulate and phytosterol abundances may facilitate the identification of phytosteryl ferulates as possible candidates causing unspecific effects in the profiles.

Multivariate analysis of standardised peak-based chromatographic data by means of PCA allowed a rapid overview on the degree of similarity within huge amounts of data. In contrast to the univariate peak-by-peak comparison, PCA focuses on the major differences for sample classification. As regards chromatograms obtained from barley grains at different malting stages, PCA plots perfectly illustrated the metabolic changes during germination.

The combination of metabolite profiling and automated data analysis proved to be able to detect expected differences in abundances of a broad spectrum of low molecular weight constituents of rice, maize, and barley. This represents an important prerequisite for application of the developed methodology for non-targeted unbiased investigation of metabolites in such cereals in order to detect unanticipated changes on the metabolite level.

4.5.6 Summary

The developed combination of metabolite profiling and automated data analysis was applied to comparative analysis of cereal grains with known differences on the metabolite level. This included the investigation of rice grains with different γ-

oryzanol contents and maize grains different in distributions of fatty acids. Germinating barley at different malting stages served as material for analysis of metabolically active tissues with expected changes in metabolite abundances.

It was demonstrated that the metabolite profiling approach developed for rice involving fractionation of total lipid and polar extracts is also applicable to maize and barley grains. Results obtained by metabolite profiling were in accordance to those gained by specific targeted analyses (rice and maize) and in agreement with literature data (barley). Automated univariate peak-by-peak comparison of metabolite profiles proved to be useful to detect statistically significant differences in peak heights between series of standardised chromatographic data sets and allowed to trace differences directly to individual peaks. Multivariate analysis of chromatogram series by means of principal component analysis on the basis of the entirety of detectable peaks allowed to classify samples according to similarity.

The provided evidence that the developed metabolite profiling approach is able to detect expected changes in metabolite abundances forms the basis for its application to discover unexpected effects on the metabolite level.

4.6 Metabolite profiling of GM and conventional rice

4.6.1 Introduction

The comparison of the chemical composition of a GM plant to that of a traditional counterpart by means of targeted investigation of defined constituents has been successfully used in the course of the safety assessment of GM foods [15]. The major limitation of this approach is its biased character focusing on known compounds and expected / predictable changes.

The objectives of this study were to exemplarily demonstrate the suitability of the developed metabolite profiling methodology to the non-targeted comparative analysis of genetically modified food, and to apply the software tools developed for automated comparative analysis of metabolite profiling data and for establishment of metabolite profiling databases.

Genetically modified Bt-rice grains expressing a synthetic *cry1Ab* gene from *Bacillus thuringiensis* (line KMD1) [89] and a lectin gene from *Galanthus nivalis* (line GNA) [91], respectively, were used as model crops. They were compared to the corresponding parental lines (Xiushui 11 and ASD16) which had been grown together with their GM counterparts in two field trials (Xiushui 11 – KMD1) and a glass house trial (ASD16 – GNA). This represents a suitable model since expression of insecticidal proteins (Cry1Ab and *Galanthus nivalis* agglutinin, GNA, respectively) does not aim at intended changes on the metabolite level. For assessment of potential differences between GM and parental lines, a model database was established containing metabolite profiles obtained from conventional rice cultivars as well as from a broad spectrum of commercial rice samples. Univariate and multivariate methods were evaluated regarding their suitability to discover statistically significant differences in abundances of individual metabolites within the broad series of complex data sets.

4.6.2 Metabolite profiling model database for conventional rice samples

In order to get an idea on the extent of natural variability, a set of 23 conventional brown rice samples was analysed by means of the developed metabolite profiling method. One sub-set comprised a total of ten samples obtained from the cultivars *Balilla*, *Cripto*, and *Thaibonnet* (sample codes BA1-4; CR1-4; Thai1-2 in Table 3).

They were different in terms of growing season and origin. *Balilla* samples BA2 - BA4, *Cripto* samples CR3 and CR4 as well as the *Thaibonnet* samples were obtained as different lots from the same supplier. Another sub-set consisted of a total of ten commercial brown rice samples for which affiliation of the cultivar was unknown (sample codes E1; E8; E10; E15; H5-7; H9; H16; J5 in Table 4). They differed in grain shape (short, medium and long grain rice), origin (Italy, US, Spain, India, Guyana, France, Japan), and other characteristics (Basmati rice, Red Camargue rice) and had been selected from a set of more than 50 commercial lots (Table 4) in order to obtain a collection of rice samples which show significant differences. The two parental Xiushui 11 lines and line ASD16 served as the third sub-set of conventional rice material.

In order to establish a metabolite profiling model database for conventional rice material, the 23 brown rice samples were extracted, fractionated and analysed by GC-FID as described in chapter 4.3. Samples obtained from cultivars *Balilla, Cripto and Thaibonnet* (n = 10) and parental lines (n = 3) were subjected in triplicate to the complete metabolite profiling procedure. The commercial lots (n = 10) were analysed once only. The resulting GC-FID chromatograms were standardised automatically and manually by means of the developed software tools. Eventually, a database was built consisting of 49 standardised peak-based data sets and a total of approximately 32,000 data points. Every data set contained the informations on ca. 650 peaks.

For characterisation of the database, the degree of overall variability in corresponding peak abundances was compared to that observed within the sub-sets of conventional cultivars (intra-cultivar variability) and to that determined by triplicate analysis of the same sample (analytical variation). Therefore standardised heights of corresponding peaks were related to their median according to equation 6 (see chapter 3.2.1.11). This resulted in equal weighting of peak abundances which were twice (+100 %) and half the median (-100 %). This mode of calculation allowed the comparison of degrees in variability irrespective of the magnitudes of absolute standardised peak heights. Results obtained for selected compounds identified in rice fractions I – IV are shown in Figure 53. The data are representative for rice constituents exhibiting high and low variability within the 23 samples included in the model database.

The highest degrees of variability within the 23 conventional samples in the model database were observed for free sugars and polyols identified in rice fraction III (see Figure 53A). Phytosterols (fraction II) and fatty acid methyl esters (fraction I) derived from ester-bound fatty acids exhibited the lowest variabilities (Table 14). In general, variability in peak abundances was lower for lipids (Figure 53C,D) than for polar compounds (Figure 53A,B). However, as exemplarily shown for β-sitosterol and free stearic acid, variabilities can differ considerably between individual compounds identified in the same fraction.

As illustrated exemplarily for eight rice constituents in Figure 53, overall variabilities within the 23 rice samples in the model database (columns 1) were higher than those observed within *Balilla* (columns 2) and *Cripto* samples (columns 3). Analytical variation determined by triplicate analysis of the same *Thaibonnet* sample (columns 4) was low enough to differentiate individual samples obtained from the same cultivar. This was even true for β-sitosterol (Figure 53D) exhibiting a remarkably low intra-cultivar variability.

Ratios of overall and intra-cultivar variabilities to analytical variation calculated for the eight compounds are summarised in Table 14. Except for methyl docosanoate and β-sitosterol, ratios were equal to or higher than four.

However, because intra-cultivar variability was as high as inter-cultivar variability, no differentiation between the cultivars *Balilla* and *Cripto* was possible on the basis of the eight rice constituents. In other words, magnitudes of differences in metabolite abundances due to changes in environmental conditions (origin, growing season) were equal to or higher than those due to different characteristics of cultivars.

This leads to the following conclusions regarding the comparison of cultivars on the basis of metabolite profiling data: In order to reveal the influence of cultivar characteristics on metabolite concentrations, metabolite profiling data obtained from samples grown under identical environmental conditions must be used for comparison. Differences in metabolite profiling data between cultivars should be evaluated in the light of natural variability determined for a set of appropriate comparators. Magnitudes of differences should be assessed on a compound-by-compound basis taking into account the individual degrees of variability.

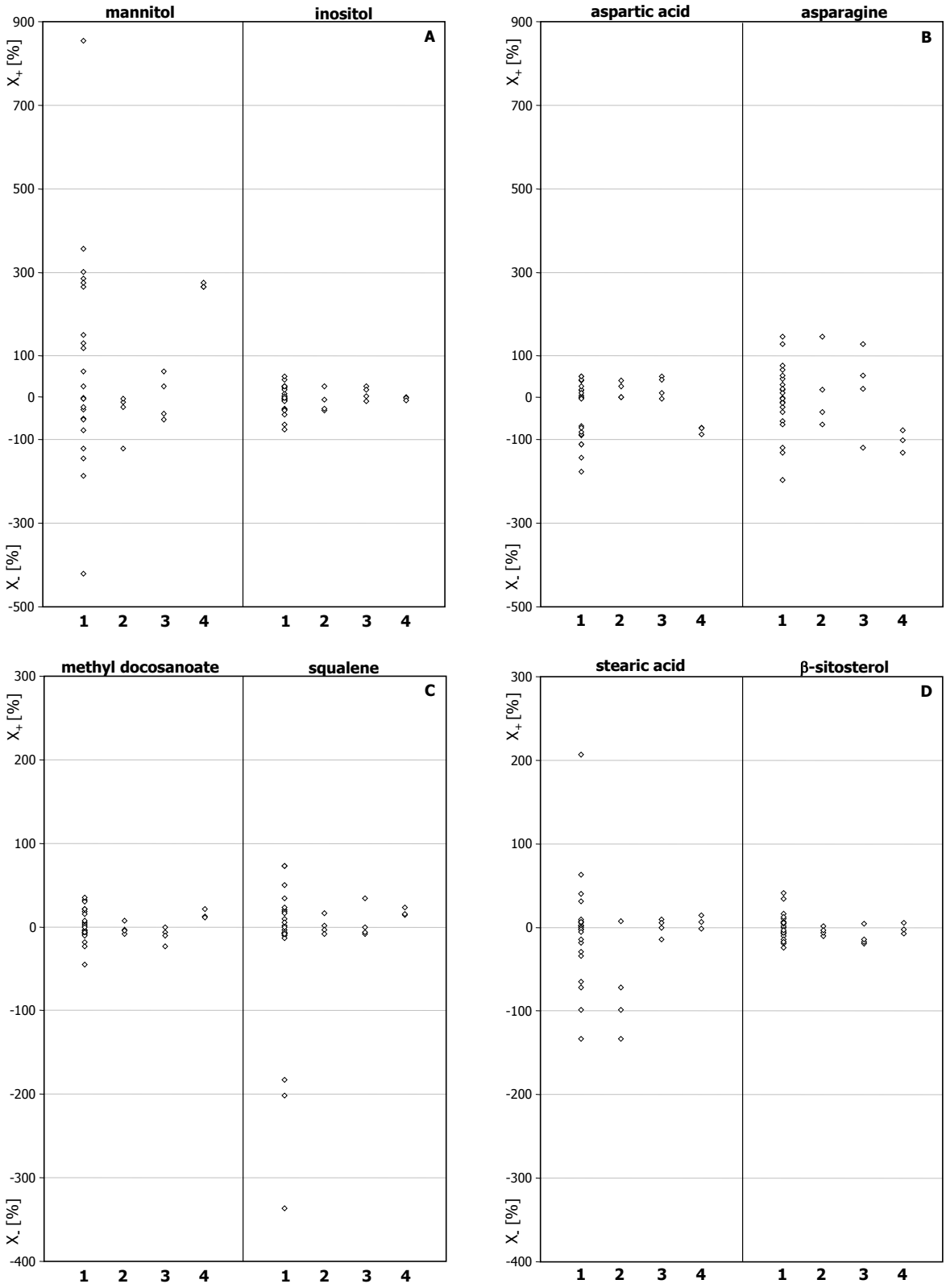

Figure 53: Characterisation of a model database containing standardised metabolite profiling data of 23 conventional rice samples. X_+ / X_- values expressing the magnitudes of divergence of standardised peak heights to the median were calculated according to **equation 6** for selected compounds identified in fraction III (**A**), fraction IV (**B**), fraction I (**C**), and fraction II (**D**). Positions of dots in columns illustrate the variability of peak heights observed within the 23 samples of the model database (**1**), within *Balilla* (**2**), and *Cripto* (**3**) samples. Column **4** illustrates the analytical variation determined by triplicate analysis of the same *Thaibonnet* sample.

Table 14: Natural variability and ratios of natural and intra-cultivar variability, respectively, to analytical variation calculated for abundances determined by metabolite profiling analysis of eight compounds in 23 conventional rice samples.

	natural variability S [%]*	natural variability / analytical variation*	*Balilla* variability* / analytical variation
mannitol	1275	425	36
squalene	409	58	4
asparagine	344	12	7
stearic acid	340	24	9
aspartic acid	229	29	4
inositol	127	21	10
methyl docosanoate	80	9	2
β-sitosterol	66	6	1.7

*: S, *Balilla* variability and analytical variation were calculated according to **equation 6**.

4.6.3 Comparative metabolite profiling of genetically modified rice

4.6.3.1 Multivariate approach

As shown by the investigation of germinating barley (see chapter 4.5.4), multivariate analysis of metabolite profiling data by means of Principal Component Analysis (PCA) can give a rapid overview on the similarity of samples regarding the entirety of constituents on the metabolite level. In order to gain a first impression on the similarity of the genetically modified rice lines to conventional counterparts, metabolite profiling data obtained from Bt- and GNA rice grains as well as those yielded from corresponding parental lines and from conventional rice cultivars (sample codes BA1-4; CR1-4; Thai1-2 in Table 3) were subjected to PCA. This involved the standardisation of a total of 192 chromatograms. Subsequently, matrices (one per rice fraction) were generated from standardised peak-based data sets by means of the developed software tools. The unbiased consideration of the entirety of peaks detectable in the 192 chromatograms resulted in a total of 32,155 data points which were subjected to PCA analysis. For graphical presentation of results, scores of the first principal component were plotted against those calculated for the second one. The plots obtained for each rice fraction are shown in Figures 54 and 55.

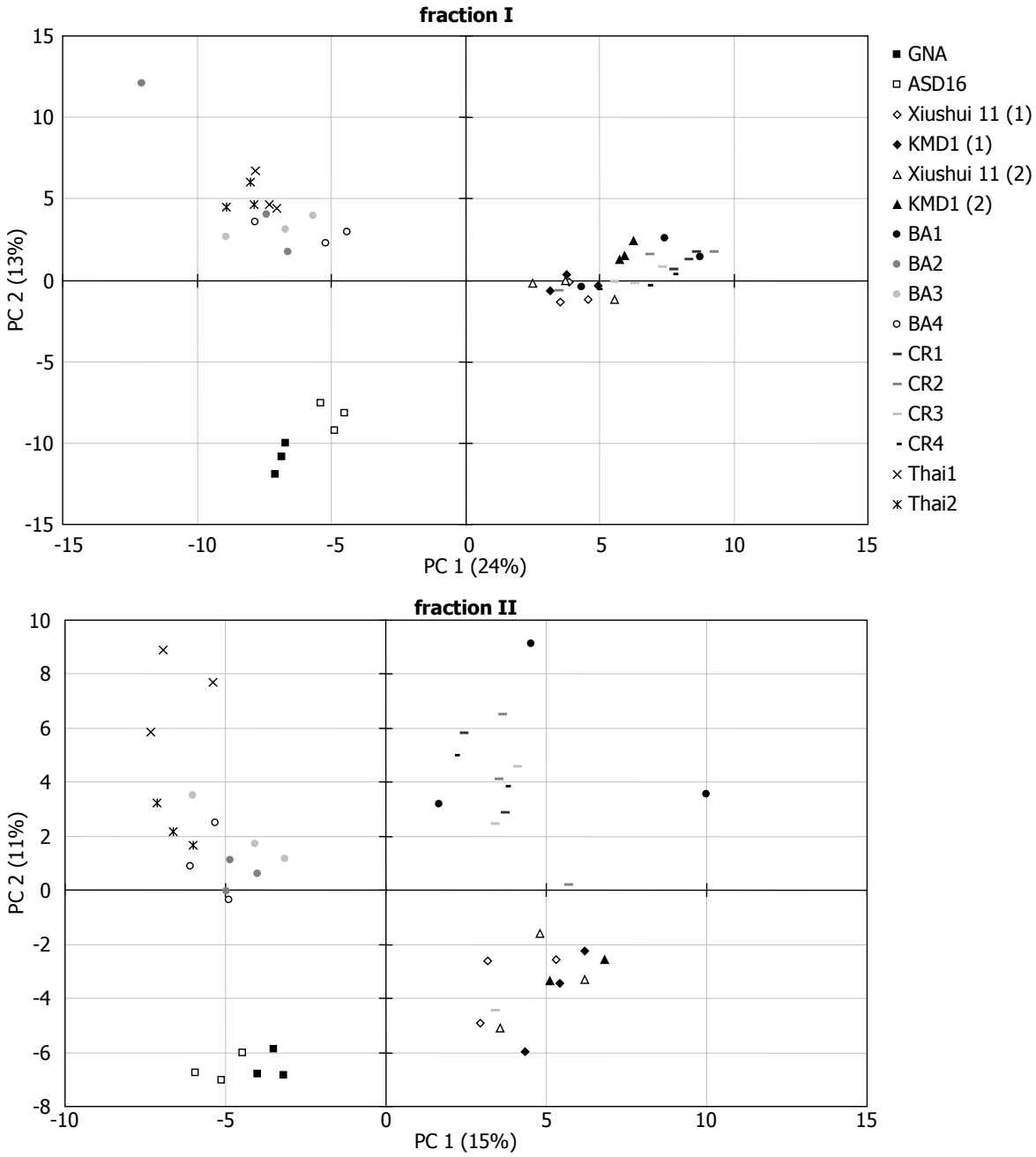

Figure 54: Principal Component Analysis of GC-FID metabolite profiling data obtained from fractions I and II (major and minor lipids) of genetically modified Bt- and GNA rice (lines KMD1, GNA), corresponding parental lines (Xiushui 11, ASD16), and conventional cultivars *Balilla* (BA1 – BA4), *Cripto* (CR1 – CR4), *Thaibonnet* (Thai1, Thai2). For sample codes see **Table 3**.

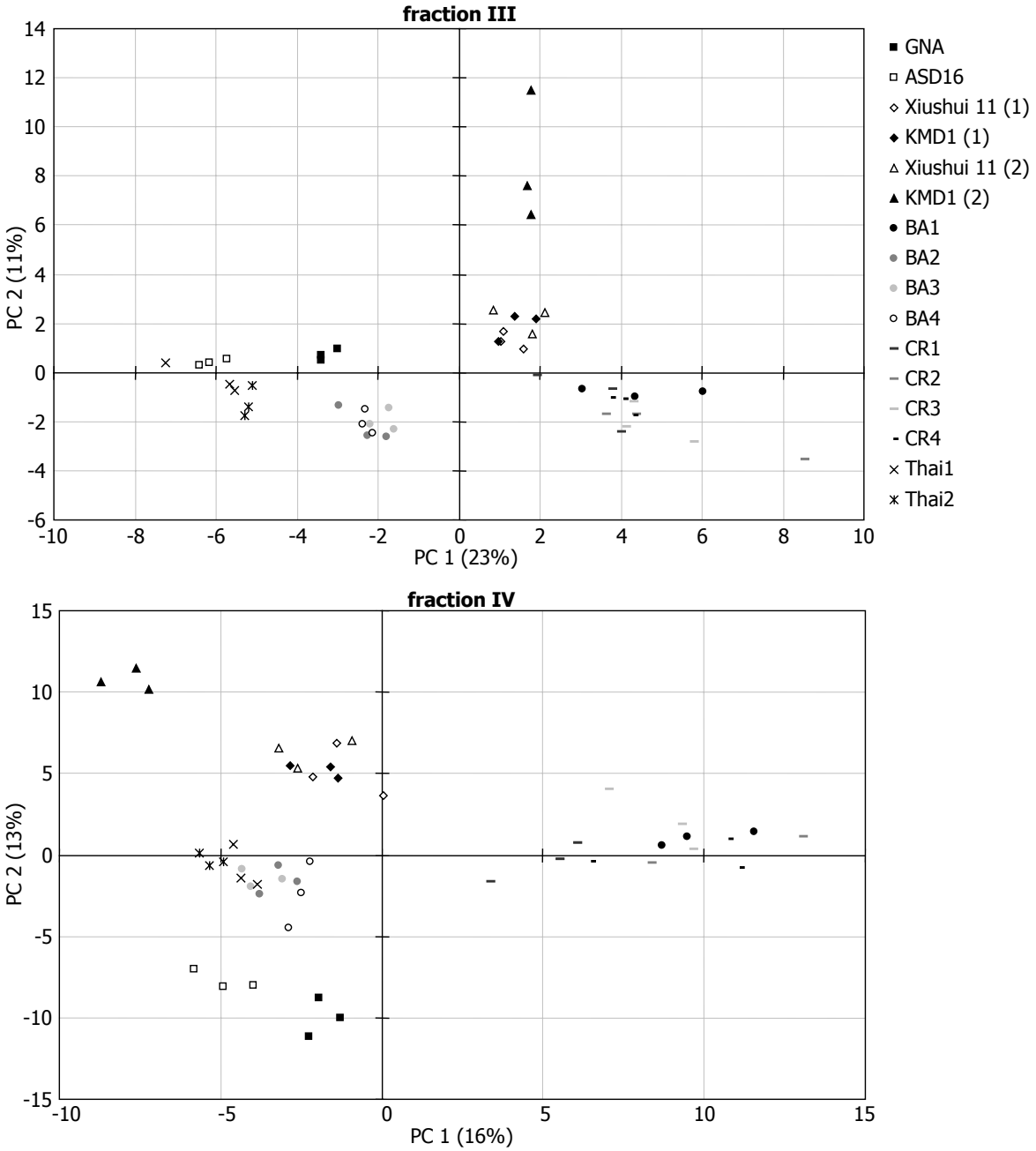

Figure 55: Principal Component Analysis of GC-FID metabolite profiling data obtained from fractions III and IV (sugars and organic acids / amino acids) of genetically modified Bt- and GNA rice (lines KMD1, GNA), corresponding parental lines (Xiushui 11, ASD16), and conventional cultivars *Balilla* (BA1 – BA4), *Cripto* (CR1 – CR4), *Thaibonnet* (Thai1, Thai2). For sample codes see **Table 3**.

As regards genetically modified GNA rice, PCA resulted in visually distinguishable groups comprising data obtained from the GM line and from the corresponding parental line (ASD16). Assessment of PCA plots for fractions I, III, and IV also revealed visually distinct differences between the Bt-rice line (KMD1 (2)) and the parental counterpart (Xiushui 11 (2)) which had been grown in the second field trial.

However, when taking into account data from the first field trial (KMD1 (1), Xiushui 11 (1)) no differences between the GM and the parental lines were visually detectable.

In general, distances between dots representing data for GM and corresponding parental lines were small in comparison to those between dots for different conventional rice cultivars. As regards *Balilla* and *Cripto* samples, intra-cultivar variability due to different origins was as high as inter-cultivar variability. PCA plots for all rice fractions indicated separate groups comprising data for the *Balilla* sample grown in Spain (BA1) and for the samples from Italy (BA2 – BA4). The groups containing data of the Spanish *Balilla* sample were always superimposed by those comprising *Cripto* data. This is in full agreement with the conclusions drawn in the course of the characterisation of the metabolite profiling database for conventional cultivars (see Figure 53).

4.6.3.2 Univariate comparative analysis of Bt-rice

Detailed comparative investigation of metabolites in Bt-rice grains was performed following the univariate approach. The procedure shown in Figure 56 was followed peak-by-peak, and comparative data analysis was carried out automatically by means of the developed software tools.

1. Comparison of mean metabolite profiling data obtained from GM and corresponding parental lines grown under identical environmental conditions:

   → *consistent* statistically significant differences between the means of GM and parental lines?

2. Assessment of *consistent* statistically significant differences by comparison of the means of GM lines to the mean/maximum/minimum of a metabolite profiling database for conventional comparators:

   → statistically significant differences between the means of GM lines and conventional comparators?

3. Further assessment of *consistent* statistically significant differences between the means of GM and parental lines by comparison of the magnitudes of differences to intra-cultivar variabilities and to natural ranges observed for conventional comparators.

Figure 56: Procedure for detection and evaluation of differences in metabolite profiling data between genetically modified lines and conventional comparators.

As a first step, standardised metabolite profiling data obtained from GM lines and from the corresponding parental lines grown under identical environmental conditions were compared. Differences between mean peak abundances (n = 3) were assessed for statistical significance by means of Student's *t* test (error probability: 5 %) on a peak-by-peak basis. Peaks exhibiting heights lower than 1 mV were not taken into consideration. Statistically significant differences between GM and parental lines which were detectable in both field trials (*consistent* statistically significant differences) triggered the additional comparison of GM lines to a database containing metabolite profiling data sets obtained from conventional comparators. For that purpose the model database described in chapter 4.6.2 was used which comprises data sets from conventional rice cultivars and from a broad spectrum of commercially obtained rice samples. Means of GM lines were compared to the mean and to the maximum/minimum calculated using the data sets stored in the database. The statistically significant differences observed were further assessed regarding their magnitudes compared to the intra-cultivar variability determined for the conventional cultivars *Balilla*, *Cripto*, and *Thaibonnet*, and to the overall natural range observed for commercial samples.

The performance of this procedure will be demonstrated by using metabolite profiling data sets obtained from fractions I – IV of genetically modified Bt-rice and conventional samples. Results of comparison will be illustrated using the graphical reports generated by the developed software tools for automated comparative analysis of metabolite profiling data.

Fraction I

The graphical reports generated by following the described procedure for comparative analysis of metabolite profiling data sets obtained from Bt-rice fractions I are presented in Figure 57. The report shown in Figure 57B indicates one statistically significant difference (-29 %) between the Bt-rice line and the corresponding parental line which had been grown in the first field trial. As indicated in Figure 57C this difference (-53 %) and four additional ones were observed in the second field trial. Hence, one *consistent* statistically significant difference was detected in fraction I. The compound "behind" the difference (peak number 1 in

Figure 57A) was identified as methyl stearate formed by transmethylation of ester-bound stearic acid. The difference was confirmed by targeted analysis of fatty acid distribution [163] which required further comparative investigations according to the procedure described in Figure 56.

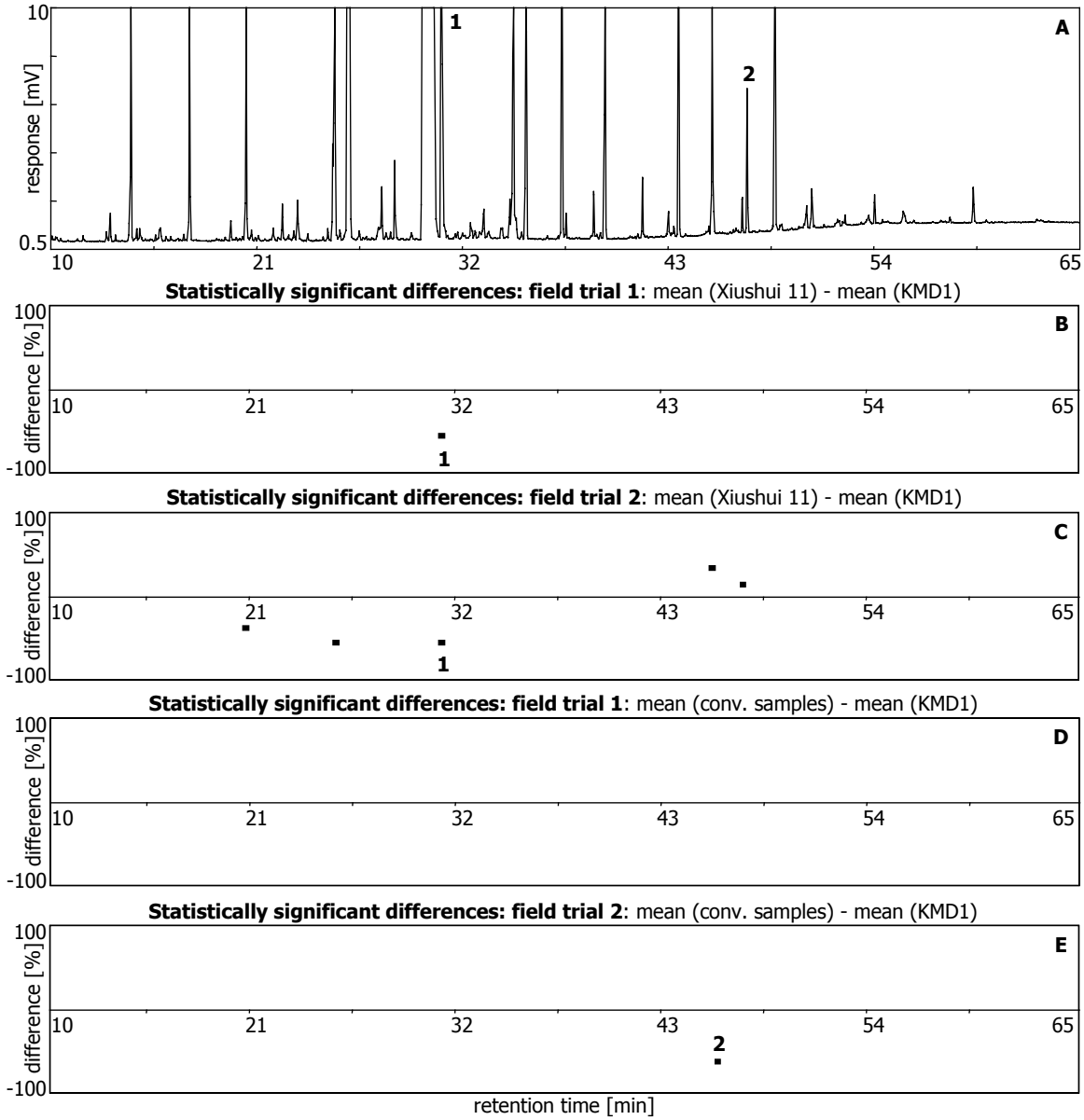

Figure 57: Comparative analysis of metabolite profiling data obtained from rice fractions I of genetically modified Bt-rice lines (KMD1, n = 3). **A**: Standardised GC-FID chromatogram obtained from the corresponding parental line (Xiushui 11) grown in the first field trial. **1**: methyl stearate, **2**: methyl hexacosanoate. **B** – **E**: Graphical reports generated by automated comparison of mean metabolite profiling data obtained from GM lines (n = 3) to those from parental lines (**B**, **C**, n = 3) and to those calculated from a database for conventional rice samples (**D**, **E**, n = 23).

As shown in Figure 57D,E, graphical reports did not indicate a statistically significant difference between the methyl stearate abundances in GM rice lines and the mean calculated from the database of conventional samples. The difference detected for

methyl hexacosanoate in Figure 57E (peak and dot number 2) had not been detected in the course of the comparison of the GM line grown in the second field trial to the corresponding parental line (Figure 57C). This indicated that methyl hexacosanoate abundances in the GM line as well as in the parental line are different from the mean of the database. According to the procedure shown in Figure 56, results obtained for fraction I did not require further assessment.

Fraction II

Figure 58 exemplarily shows a tabular report generated by automated comparison of mean metabolite profiling data obtained from rice fractions II of Bt and parental rice grains grown in the first field trial. It indicates that calculation of a total of 365 mean peak heights was performed. Ninety seven means were higher than 1 mV and were taken into account for comparison. Eventually, this led to the detection of 2 statistically significant differences (+20 %; -39 %) which were indicated in the corresponding graphical report shown in Figure 59B.

statistics: comparison of	**M_Xiushui 11**	**and**	**M_KMD1**
n =	**3**	**n =**	**3**
Parameters			
min. peak height:		1000 µV	
error probability:		5 %	
width:		0.1 min	
total peaks:		365	
peaks included:		97	
total alterations:		**2**	**4%**
Additional peaks			
total:		**0**	
peak height:	100 - 500:	0	0%
	501 - 1000:	0	0%
	1001 - 2000:	0	0%
	2000 - 5000:	0	0%
	>5000:	0	0%
Statistically significant differences			
total:		**2**	
10 - 30 %:		1	50%
30 - 50 %:		1	50%
50 - 100 %:		0	0%
>100 %:		0	0%
Blanks			
Blank 1	0.00 min	to	0.00 min
Blank 2	32.80 min	to	33.50 min

Figure 58: Tabular report generated by automated comparison of mean metabolite profiling data obtained from rice fractions II of GM (KMD1, n = 3) and parental rice grains (Xiushui 11, n = 3) grown in field trial 1.

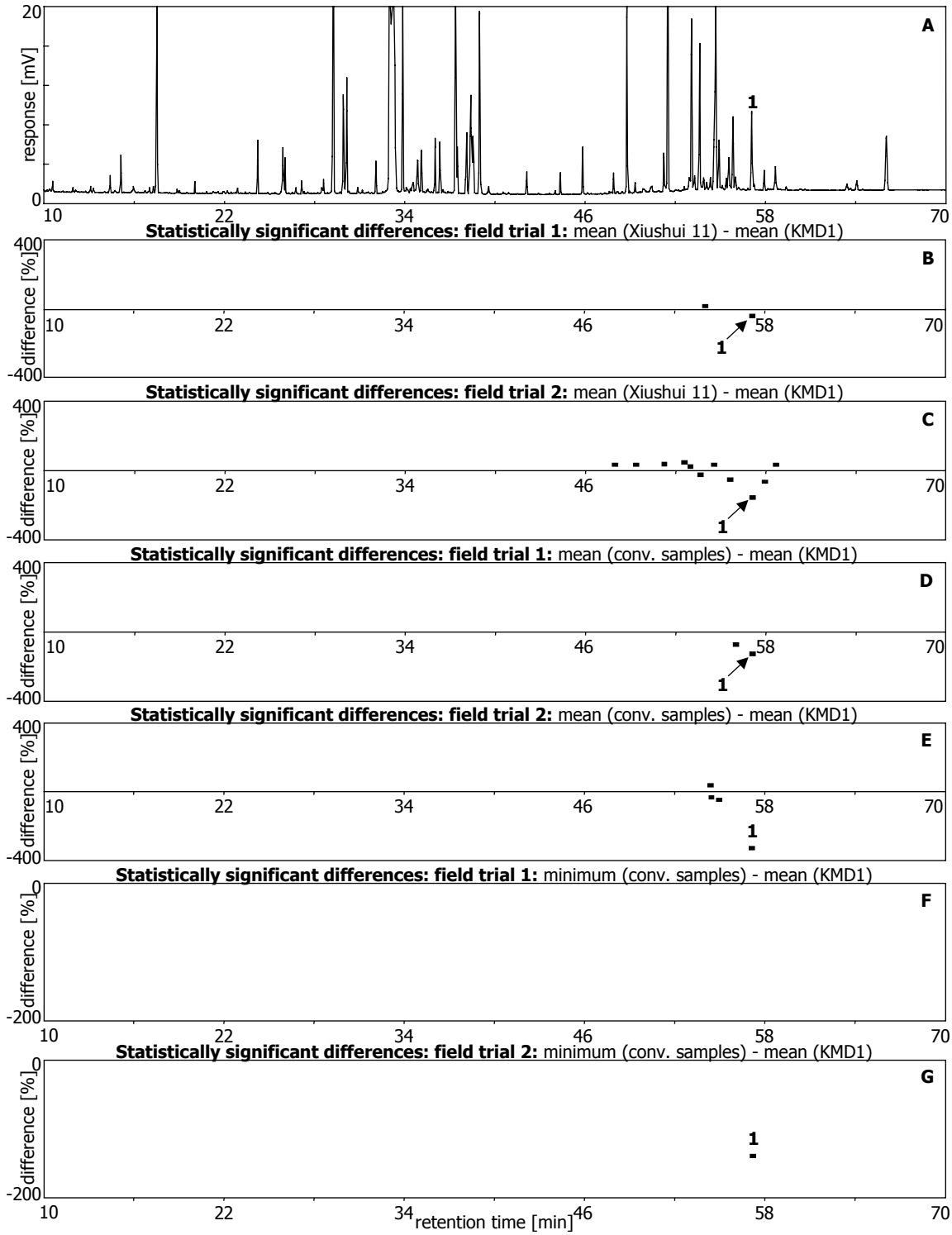

Figure 59: Comparative analysis of metabolite profiling data obtained from rice fractions II of genetically modified Bt-rice lines (KMD1, n = 3). **A**: Standardised GC-FID chromatogram obtained from the corresponding parental line (Xiushui 11) grown in the first field trial. **1**: 24-methylenecycloartanol-TMS. **B** – **G**: Graphical reports generated by automated comparison of mean metabolite profiling data obtained from GM lines (n = 3) to those from parental lines (**B**, **C**, n = 3) and to means (**D**, **E**) and minima (**F**, **G**) calculated from a database for conventional rice samples (n = 23). *Consistent* statistically significant differences were indicated for 24-methylenecycloartanol-TMS (**1**).

The additional graphical reports presented in Figure 59 revealed that one of the two differences was also detected by comparative analysis of metabolite profiling data obtained from the second field trial (Figure 59C). The compound "behind" this *consistent* statistically significant difference was identified as 24-methylenecycloartanol-TMS (24-MC, peak number 1 in Figure 59A). In contrast to the situation described for fraction I, 24-MC abundances in GM lines were statistically significantly lower than the mean and minimum 24-MC abundances calculated from the database of conventional rice samples (Figure 59D - G). According to the procedure described in Figure 56, this triggered further assessment of the differences.

As shown in Figure 60, the magnitude of the difference observed in the second field trial was higher than that observed in the first one. However, the magnitudes of the differences in 24-MC contents observed in both field trials were lower than the intra-cultivar range observed for *Cripto* samples from different origins grown in different seasons.

The 24-MC abundance determined in the GM lines is outside of the natural range determined for conventional rice samples. From a biochemical point of view this should be confirmed in additional field trials.

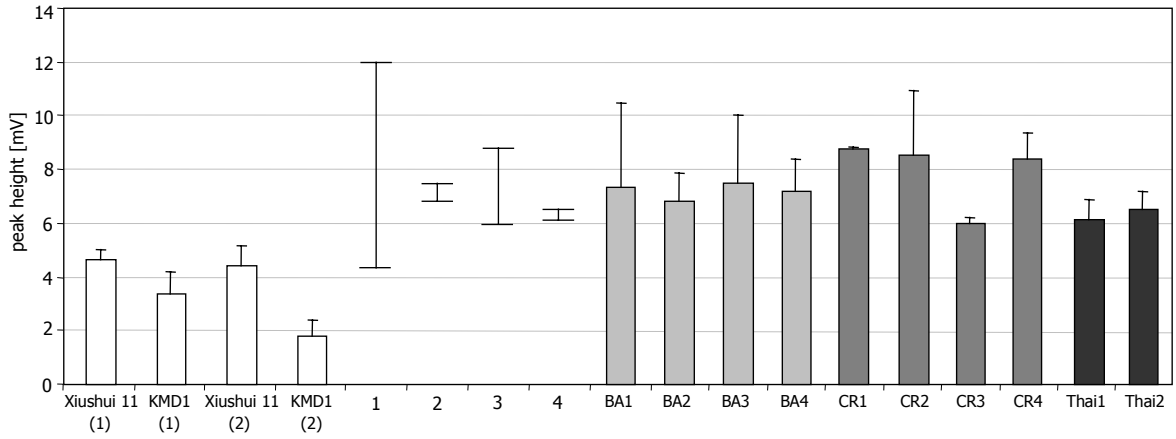

Figure 60: Abundances of 24-methylenecycloartanol-TMS in GM and conventional rice material. Bars display mean abundances (n = 3) of 24-methylenecycloartanol-TMS calculated from GC-FID chromatograms obtained from rice fractions II of genetically modified Bt- (KMD1) and corresponding parental lines (Xiushui 11) and of the conventional rice cultivars *Balilla*, *Cripto*, and *Thaibonnet*. For sample codes see 3.1.1. Error bars display confidence intervals (error probability: 5 %). Vertical lines display the ranges (from maximum to minimum) of mean 24-methylenecycloartanol-TMS abundances determined for (**1**) conventional rice samples (natural range, N = 23), (**2**) *Balilla* samples (N = 4), (**3**) *Cripto* samples (N = 4), and (**4**) *Thaibonnet* samples (N = 2).

Fraction III

Principal Component Analysis had revealed a clear separation between metabolite profiling data obtained from the GM line grown in field trial 2 to those of the corresponding parental line (see Figure 55). This was substantiated by means of the univariate approach. As shown in Figure 61D, three additional peaks were detected in Bt-rice grains grown in the second field trail which were identified as xylitol, mannitol, and trehalose. However, this was not confirmed by comparative analysis of metabolite profiling data obtained from the first field trial (Figure 61B) and must be interpreted in consideration of the high degree of natural variability found for sugars and polyols (see chapter 4.6.2). The only *consistent* statistically significant difference in fractions III obtained from GM and corresponding parental lines was detected for raffinose-TMS (Figure 61A,C,E). Its levels were also statistically significantly lower than the corresponding mean (data not shown) and minimum (field trial 2, Figure 61G) calculated from the database for conventional rice samples. As shown in Figure 62, the magnitudes of the differences between GM and parental lines were comparable to the intra-cultivar range in *Balilla* samples.

Fraction IV

The sequence of comparisons between metabolite profiling data obtained from GM lines and those from conventional cultivars was identical to that described for fraction II (see Figure 59). The appearance of a separate group in the PCA plot comprising data yielded from GM line grown in field trial 2 (see Figure 55) was substantiated by the detection of a relatively high number of 20 statistically significant differences to the corresponding parental line. However, only two of them were confirmed through evaluation of data gained from the first field trial. Subsequent comparison of GM lines to the database for conventional rice samples resulted in the indication of one peak whose abundances were statistically significantly lower than the mean and the minimum calculated from the database. Identification of the compound "behind" the difference through matching of its mass spectrum to entries in mass spectra libraries [80, 81] was not achieved.

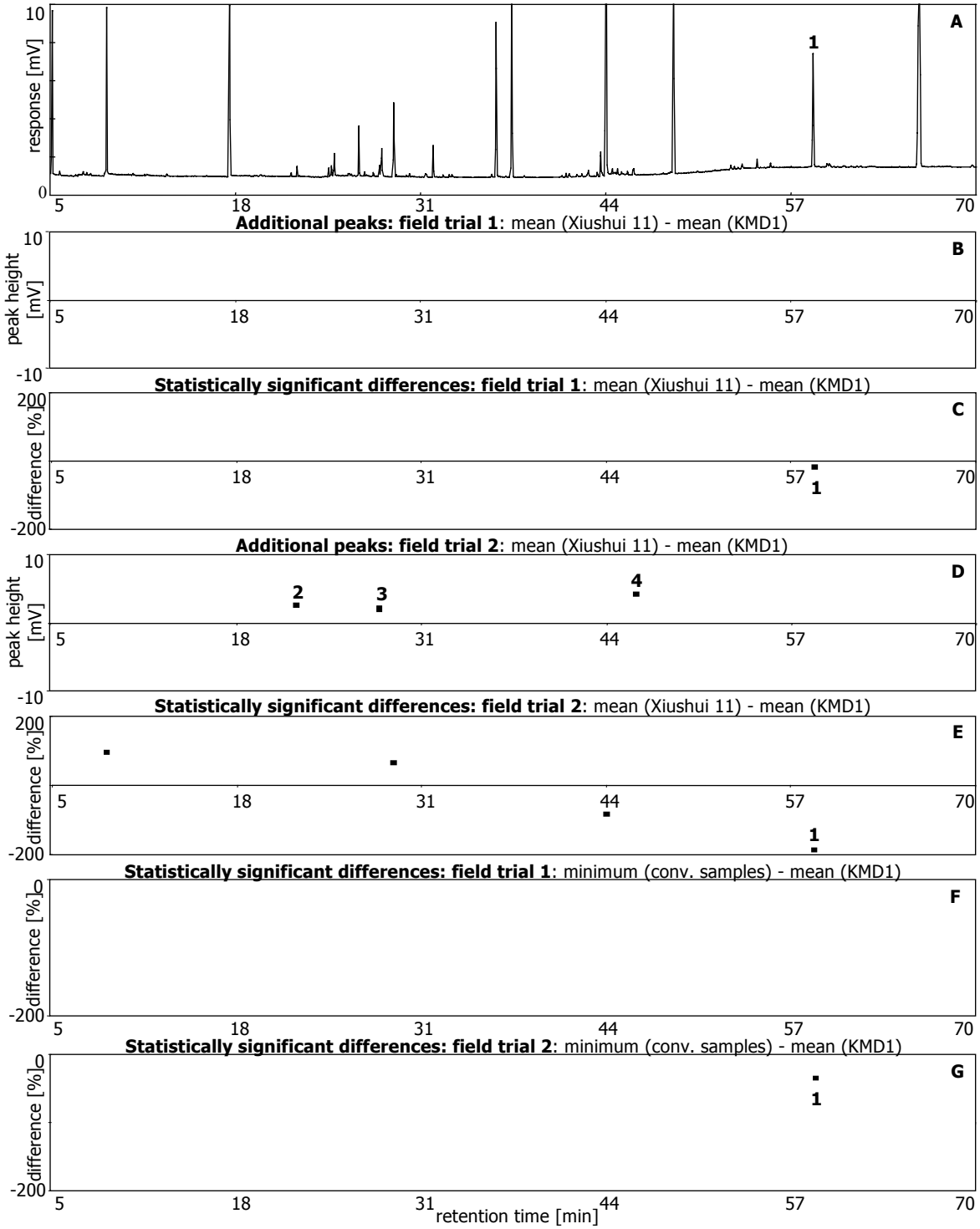

Figure 61: Comparative analysis of metabolite profiling data obtained from rice fractions III of genetically modified Bt-rice lines (KMD1, n = 3). **A**: Standardised GC-FID chromatogram obtained from the corresponding parental line (Xiushui 11) grown in the first field trial. **1**: raffinose-TMS. **B** – **G**: Graphical reports generated by automated comparison of mean metabolite profiling data obtained from GM lines (n = 3) to those from parental lines (**B** – **E**, n = 3) and to minima (**F**, **G**) calculated from a database for conventional rice samples (n = 23). Three additional peaks were detected in chromatograms obtained from the GM line grown in the second field trial and were identified as xylitol (**2**), mannitol (**3**), and trehalose (**4**). *Consistent* statistically significant differences between GM and corresponding parental lines were indicated for raffinose-TMS (**1**).

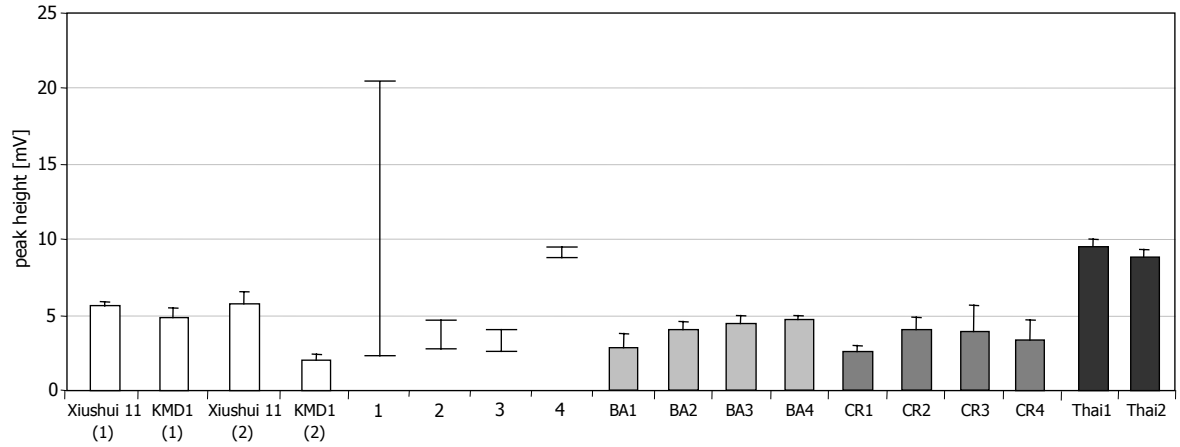

Figure 62: Abundances of raffinose-TMS in GM and conventional rice material. Bars display mean abundances (n = 3) of raffinose-TMS calculated from GC-FID chromatograms obtained from rice fractions III of genetically modified Bt- (KMD1) and corresponding parental lines (Xiushui 11) and of the conventional rice cultivars *Balilla*, *Cripto*, and *Thaibonnet*. For sample codes see 3.1.1. Error bars display confidence intervals (error probability: 5 %). Vertical lines display the ranges (from maximum to minimum) of mean raffinose-TMS abundances determined for (**1**) conventional rice samples (natural range, N = 23), (**2**) *Balilla* samples (N = 4), (**3**) *Cripto* samples (N = 4), and (**4**) *Thaibonnet* samples (N = 2).

Table 15 summarises the results gained by non-targeted comparison of metabolite profiling data obtained from genetically modified Bt-rice grains to those of conventional rice samples. Out of 163 peaks taken into account for analysis, five were found to be *consistently* statistically significantly different to those in the corresponding parental lines grown under identical environmental conditions. Three of the five were outside the natural range determined for 23 conventional rice samples.

Table 15: Comparison of metabolite profiling data obtained from genetically modified Bt-rice grains to those of conventional rice samples.

		statistically significant differences			
		to parental lines		**to database**	
fraction	peaks included for comparison	compound	mean ratio GM/parental	mean	maximum/ minimum
I	34	methyl stearate	0.71	**-**	**-**
II	57	24-methylene-cyloartanol-TMS	0.55	**+**	**+**
III	18	raffinose-TMS	0.61	**+**	**+**
IV	54	aspartic acid-TMS	1.45	**-**	**-**
		unknown	0.69	**+**	**+**
total numbers	**163**	**5**		**3**	**3**

4.6.3.3 Univariate comparative analysis of GNA rice

Univariate comparative analysis of genetically modified GNA rice by means of metabolite profiling was performed according to the same procedure as described for Bt-rice. The non-targeted consideration of 169 mean peak heights calculated automatically from GC-FID chromatograms obtained from the GM line for comparison resulted in the detection of 50 statistically significant differences to the corresponding parental line (ASD16). This was in full agreement with the plots generated by multivariate data analysis (Figures 54, 55) which had revealed a slight but clearly visible separation between the two lines. Magnitudes of the differences ranged from -104 % (citrostadienol-TMS) to +265 % (unknown constituent). In contrast to Bt-rice, material for comparative analysis was limited to brown rice grains grown in a single glass house trial.

Statistically significant differences detected between GM and parental line were assessed in the light of the natural range represented in the established model database for conventional rice samples. Automated data analysis by means of the developed software tools revealed a total of 8 constituents of GNA rice exhibiting mean abundances which were statistically significantly different to the means calculated from the model database. Two of them (valine-2TMS and cycloartenol-TMS) were found to be outside the natural range determined for 23 conventional rice samples. The data are summarised in Table 16.

In order to evaluate the total number as well as the magnitudes of differences observed between the GM and the corresponding parental line, they were compared to those determined by comparative analysis of conventional cultivars. Results are summarised in Figure 63. With rising diversity in terms of origin, growing season, and cultivar, the total number as well as the number of wide differences (>100 %) observed between conventional rice samples increased. Comparison of two *Balilla* samples grown in Italy in 2000 resulted in the detection of 11 statistically significant differences. Only two of them were higher than 50 %. The use of a *Thaibonnet* sample as comparator revealed a total of 51 statistically significant differences and 19 magnitudes higher than 50 %. As regards the total number of statistically significant differences, results obtained for ASD16 and GNA (50 differences) were comparable to inter-cultivar variability determined for cultivars *Balilla* and *Thaibonnet*. However, as indicated by the dotted line in Figure 63, this was mainly

due to the high proportion of small discrepancies (<50 %) between GM and parental line. On the basis of the number of wide differences (>100 %), diversity of ASD16 and GNA was in the same range as observed for intra-cultivar variability in different *Cripto* samples.

Table 16: Comparative analysis of genetically modified GNA rice (n = 3) and conventional rice samples by means of metabolite profiling

		statistically significant differences [%][a]		
fraction	compound	to parental line[b]	to the mean of conventional samples[c]	to the maximum/minimum of conventional samples[d]
I[e]	C14:0	+51	-	-
	C15:1	-42	-	-
	C16:2	+32	-	-
	C16:1	+32	+50	-
	C16:0	+16	-	-
	C17:1	+38	-	-
	C17:0	+42	-	-
	C18:0	+55	+42	-
	C20:2	-21	-	-
	C20:1	-15	-	-
	C20:0	+27	+42	-
	C22:1	+35	-	-
	C23:0	+16	-	-
	C24:1	+28	-	-
	squalene	-54	-336	-
	nonacosane	+41	-	-
	C28:0	+30	-	-
II[f]	C18:0	+27	-	-
	C20:1	-25	-	-
	n. i.[g]	-38	-	-
	n. i.	-12	-	-
	Δ^7-campestenol	-37	-	-
	n. i.	-60	-	-
	β-sitosterol	-45	-	-
	Δ^5-avenasterol	-74	-	-
	gramisterol	-41	-	-
	cycloartenol	-48	-99	-21
	Δ^7-avenasterol	-50	-33	-
	24-methylenecycloartanol	51	-	-
	citrostadienol	-104	-	-
III[f]	glycerol	+15	-	-
	sorbitol	additional peak	+502	-
	glucose	+138	-	-
	inositol	+43	-	-
	sucrose	+47	-	-
	trehalose	missing peak	-	-
	raffinose	-56	-	-
IV[f]	alanine	+33	-	-
	valine	+49	+35	+7
	phosphate	-27	-	-
	proline	+152	-	-
	leucine	+74	-	-
	serine	+77	-	-
	threonine	+53	-	-
	n. i.	-146	-	-
	malate	-71	-	-
	aspartic acid	+81	-	-
	phenylalanine	+61	-	-
	glutamic acid	+55	-	-
	n. i.	+265	-	-

[a]: determined by Student's *t*-test, error probability: 5 %; [b]: n = 3; [c, d]: n = 23; [e]: compounds (except for squalene and nonacosane) were identified as fatty acid methyl esters; [f]: compounds were identified as trimethylsilylated derivatives; [g]: not identified

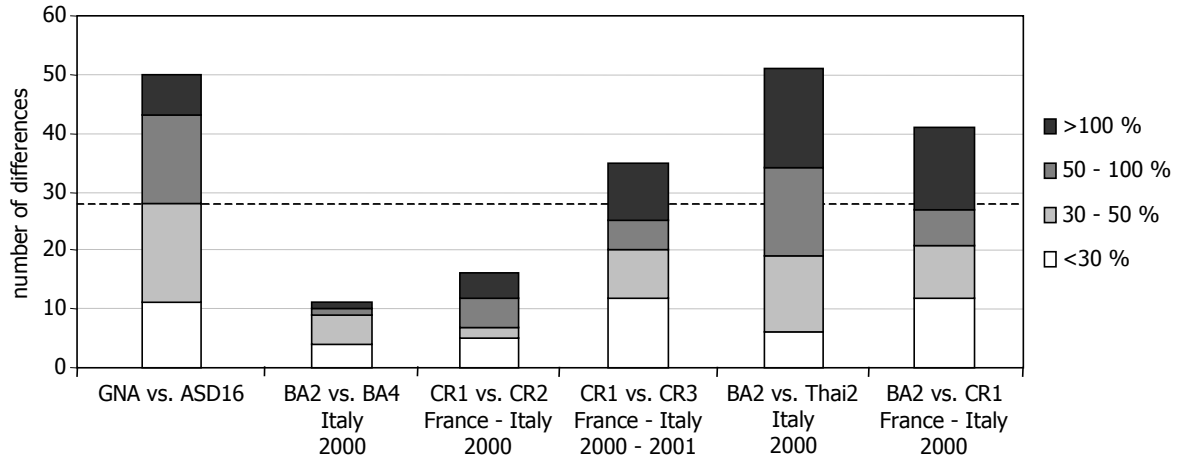

Figure 63: Number and magnitudes of statistically significant differences determined by comparative metabolite profiling analysis of genetically modified GNA rice and its corresponding parental line (ASD16) and of conventional rice cultivars from different origins grown in different seasons. For sample codes see **Table 3**.

4.6.4 Conclusions

The developed combination of metabolite profiling and automated data analysis was applied to comparative compositional analysis of genetically modified rice grains. This was performed in order to investigate the applicability of a non-targeted profiling approach for chemical characterisation within the safety assessment framework of GM foods. The evaluation of the methodology focused on the question whether additional exploitation of such an approach may yield in additional information which is useful to guide the subsequent safety assessment process.

The unbiased non-targeted character of profiling techniques is the most striking argument for their additional use in the course of a safety assessment. The methodologies applied must be as comprehensive as possible and have to offer the opportunity to identify and quantify the metabolites which are "behind" the differences observed. Analytical data have to be provided in a suitable form for further evaluation.

Genetically modified rice grains expressing genes for biosynthesis of insecticidal proteins were used as model material. As the type of genetic modification did not aim at a change on the metabolite level, the study design allowed to focus evaluation of the chemical composition of the GM lines on the potential occurrence of unexpected effects on metabolite levels.

In accordance to the recommended procedure for targeted analysis of single compounds, metabolite profiling of GM lines was performed following a comparative approach. Corresponding parental lines grown under identical environmental conditions were taken into account as the primary comparators. In contrast to specific analyses, a method was used isolating a broad spectrum of compounds from the sample matrix. GC-FID investigation of pre-fractionated extracts resulted in highly resolved chromatograms exhibiting hundreds of peaks. Comparison of their mass spectra to entries of mass spectra libraries led to the identification of more than one hundred compounds. The necessity to derivatise compounds to make them volatile and limited possibilities to collect them from the eluate of the GC column for further investigations are the main drawbacks of a GC based approach. Compounds exhibiting thermal instability and low volatility cannot be detected by the developed profiling methodology. Presently, it does not cover metabolites "typical" for food quality (*e.g.* vitamins) and food safety (*e.g.* mycotoxins). Although mass spectrometry provided valuable information on structural identity of compounds, *de-novo* identification of unknown metabolites was difficult to achieve. Metabolite profiling methods applying liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance may help to overcome these limitations.

Considering the specific requirements of a safety assessment, data analysis of a metabolite profile obtained from a GM food has to reveal the entirety of differences to that obtained from the conventional counterpart. The software tools developed allow to meet these requirements and are suitable to take into consideration for analysis the entirety of peaks detectable in metabolite profiles. Principal Component Analysis resulted in the formation of visually distinguishable groups comprising metabolite profiling data obtained from GM and parental lines used in the model study. The use of multivariate techniques is particularly attractive because they allow to reduce data complexity to information most useful to describe differences between samples. Analysis of PC loadings has been suggested in order to find the differences in peak abundances which led to a differentiation of sample groups in PCA plots. However, the use of a univariate peak-by-peak approach for comparison of metabolite profiles allows to direct assignment of differences in profiles to the underlying peaks and assures detection of even minor qualitative and quantitative changes.

The software tools developed were tailored in order to achieve non-targeted univariate comparison of metabolite profiling data. Their application resulted in the unbiased consideration of all peaks higher than a fixed response threshold for data analysis.

The results obtained indicate the potential of the developed metabolite profiling methodology to detect statistically significant differences between metabolite profiling data obtained from genetically modified lines and conventional counterparts. This meets the objective of this study to demonstrate the applicability of a non-targeted approach for comprehensive comparative compositional analysis of GM foods.

Investigation of material obtained from additional field trials could be required in order to prove the correlation between differences in composition and the application of recombinant DNA techniques.

Differences in compositions between GM and conventional crops do not mean a safety issue *per se*. Evaluation of unintended changes regarding their biochemical, nutritional, toxicological, and safety relevance is one of the most challenging tasks arising from the exploitation of metabolite profiling techniques but goes beyond the scope of the present study. The scheme elaborated (see Figure 56) may help to standardise the procedure for comparison of a GM to a broader spectrum of conventional foods as a first step to assess unintended differences to the corresponding parental line. Considering the huge amounts of data which must be generated in the course of such an approach, databases are required to store the data. The software tools developed may help to establish and automatically utilise metabolite profiling databases.

From a safety point of view, a database describing the natural variability of metabolite abundances for the parental line should be the primary basis for evaluation of *consistent* statistically significant differences in metabolite profiling data between GM and parental lines grown under identical environmental conditions. Databases reflecting different developmental stages and environmental conditions are essential. Inter-laboratory co-operation might help to accelerate the build-up of such databases. However, this requires stringent standardisation of the protocols applied for generation of metabolite profiling data. Different approaches have been described for establishment of metabolite profiling databases [67]. So far, no

metabolite profiling database exists which can be utilised to assess metabolite profiles in the course of a safety assessment.

Application of metabolite profiling techniques for safety assessment of GM foods may in principle provide relevant information regarding metabolic consequences as a result of genetic modification. In theory, an unbiased non-targeted comprehensive comparison between GM organisms and their traditional counterparts offers almost unlimited possibilities for detection of unintended effects. However, due to the structural diversity of metabolites, analytical coverage of the "metabolome" by means of a single method will not be possible. The potential of metabolite profiling techniques for characterisation of the chemical composition of GM foods is obvious, but further exploration with respect to specificity and sensitivity of the methods and validation is needed in order to make it really useful for safety evaluation of GM foods.

4.6.5 Summary

In a model study, the developed combination of metabolite profiling and automated data analysis was applied to the characterisation of the chemical compositions of genetically modified rice grains expressing genes for biosynthesis of the insecticidal proteins *Cry1Ab* and *GNA*, respectively. Data obtained from GM lines were compared to those of corresponding parental lines grown under identical environmental conditions and statistically significant differences were evaluated in the light of the variability in metabolite profiling data stored in a model database for conventional rice samples. The study revealed the potential of the methodology for unbiased screening of a broad spectrum of compounds in order to reveal statistically significant differences in the compositions of genetically modified and conventional foods. The suitability of the developed software tools for establishment of databases and for assessment of the detected differences was demonstrated. This paves the way towards application of a non-targeted metabolite profiling approach in order to screen for unintended effects due to the application of recombinant DNA techniques.

5 Outlook

The global area of transgenic crops is steadily increasing [3]. In addition to crops with altered agronomic traits, a number of genetically modified plants obtained through extensive genetic modifications with the purpose of improving food quality traits will soon enter the commercial market [6, 164]. Food safety assessment of the resulting products will become increasingly challenging. New innovative techniques such as DNA microarray technology, proteomics, and metabolomics may assist in characterising the complex interactions of food components at the molecular and cellular levels.

In addition to the assessment of unintended effects regarding their potential influence on food safety, characterisation of unexpected changes on the metabolite level may be extremely useful to understand metabolic networks. Recently, large-scale approaches have been initiated involving targeted over-expression and inhibition of key enzymes in plant metabolic pathways and subsequent determination of the metabolic consequences by means of non-targeted metabolite profiling techniques [165].

Metabolic fingerprinting as well as metabolite profiling may be applied for high-throughput analysis of GM plants for unexpected shifts in metabolite distributions. The developed NIRS method offers the possibility to screen large sets of rice genotypes for unintended effects regarding phytic acid contents in a non-destructive way. It may also be used to identify low-phytate-mutants thus exhibiting improved nutritional properties. The performance of NIRS to quantitatively analyse several compounds on the basis of a single spectrum is particularly attractive.

In addition to the frequently cited "Golden Rice" [36], GM rice plants with altered polyamine [166] and trehalose levels [27] are in the pipeline. In order to assess the food safety of products derived from such plants with intended changes on the metabolite level, unbiased investigation of metabolites will be especially important. The metabolite profiling method described allows simultaneous non-targeted analysis of a broad spectrum of cereal constituents including sugars, amino acids, and isoprenoids which are involved in key metabolic pathways. However, combination of approaches rather than a single technique will be required to comprehensively cover the metabolome.

The potential of metabolite profiling techniques for compositional analysis is obvious but their suitability to meet regulatory requirements and their usefulness for food safety assessment still needs to be demonstrated. This will certainly involve considerable efforts regarding standardisation and validation. Furthermore, metabolite profiling databases will be an important prerequisite, in order to be able to store, publish, and evaluate the large amounts of data generated. Establishment of databases storing metabolite profiles together with raw data and detailed metadata (*i.e.* information about the context of the data, such as the experimental protocol) has been proposed [67]. This type of databases may be extremely useful if comparison of metabolite profiling data generated in different laboratories is required. The software tools described offer the possibility to provide raw data in a suitable form for establishment of such databases.

Exploitation of a combination of specific methods for analysis of single compounds and profiling techniques will increase the chances to detect unintended effects resulting from the application of recombinant DNA techniques.

6 Summary

Analysis of the chemical composition represents an important step in the safety assessment of genetically modified foods. In addition to the targeted investigation of single compounds, non-destructive techniques and metabolite profiling methodologies can be applied for analysis of low-molecular weight food constituents. In the present study, the suitability of these approaches for the analytical characterisation of rice grains within a food safety assessment framework was investigated. Potential and limitations of the techniques were exemplarily demonstrated by analysis of low-molecular weight compounds in conventional and genetically modified (GM) rice grains.

The suitability of Near Infrared Spectroscopy (NIRS) for non-destructive determination of minor compounds was investigated by using phytic acid, an anti-nutritional rice grain constituent, as example. Applicability of the methodology for rapid screening of rice samples from different origins regarding low, medium, and high phytic acid contents was demonstrated. Results obtained by NIRS for GM and parental rice lines were in agreement with those determined by the traditional method. Sample preparation is reduced to a minimum by the non-destructive approach. However, extensive calibration and validation is required in order to gain reliable results by means of NIRS. Assessment of NIRS data obtained from GM lines has to take into account spectral features of corresponding parental lines.

A gas chromatographic metabolite profiling method for unbiased screening of a broad spectrum of compounds was elaborated using rice as model crop. It is based on a fractionation approach which eventually allows non-targeted investigation of major and minor rice grain constituents. General applicability of the technique for analysis of cereals with metabolite compositions similar to rice was demonstrated. Reliability of the profiling approach was proven by investigation of rice, maize and barley samples with known and expected differences on the metabolite level. Software tools for comparative analysis of metabolite profiling data were developed. They include methods for manual and automated correction of retention times and responses by means of corresponding standards. This allows transfer of normalised datasets into databases and eventually automated unbiased chromatogram comparison. Application of the metabolite profiling methodology to the analysis of

GM and conventional rice revealed the suitability of a non-targeted approach for detection of statistically significant differences between metabolite levels in different genotypes. The usefulness of databases established through investigation of a broad spectrum of conventional rice cultivars for evaluation of differences in metabolite profiling data was demonstrated.

The unbiased screening of a broad spectrum of compounds combined with powerful methods for data analysis may help to increase the chances to detect unintended effects due to the application of recombinant DNA techniques.

7 Zusammenfassung

Die Bestimmung der chemischen Zusammensetzung stellt einen wichtigen Bestandteil der Sicherheitsbewertung gentechnisch veränderter Lebensmittel dar. Neben der spezifischen Analyse von Einzelverbindungen können nicht-destruktive Verfahren und Screening–Techniken zur Untersuchung niedermolekularer Lebensmittelinhaltsstoffe eingesetzt werden. In der vorliegenden Arbeit wurden diese Methodenansätze hinsichtlich ihrer Anwendbarkeit zur chemisch-analytischen Charakterisierung von Reis im Rahmen einer Sicherheitsbewertung beurteilt. Möglichkeiten und Grenzen dieser Techniken wurden beispielhaft anhand der Untersuchung konventioneller und gentechnisch veränderter Reissorten aufgezeigt.

Die Eignung der Nahinfrarotspektroskopie (NIRS) zur nicht-destruktiven Analyse von Minorbestandteilen wurde am Beispiel der Phytinsäure, einem Reisinhaltsstoff mit antinutritiven Eigenschaften, untersucht. Es wurde gezeigt, dass die NIRS bei minimalem Probenvorbereitungsaufwand zur schnellen Einteilung von Reissorten nach hohem, mittlerem und niedrigem Phytinsäuregehalt geeignet ist. Die mit der NIRS erhaltenen Ergebnisse für konventionelle und gentechnisch veränderte Reissorten stimmten gut mit denen der klassischen nasschemischen Referenzmethode überein. Dies setzte jedoch einen erheblichen Aufwand für die Kalibration und Validation voraus. Weiterhin stellte sich heraus, dass die Beurteilung quantitativer NIRS Daten für die gentechnisch veränderten Linien nur unter Einbeziehung der Daten für die Elternlinien möglich ist.

Im Rahmen dieser Arbeit wurde ein gaschromatographisches Analysenverfahren entwickelt, welches das Screening eines breiten Spektrums niedermolekularer Stoffe in Reis als Modell-Lebensmittel erlaubt. Das Prinzip der Methode besteht in der Anwendung von Fraktionierungsschritten, die schließlich eine separate Untersuchung der Haupt- und Minorbestandteile von Reis ermöglichen. Die allgemeine Anwendbarkeit der zugrundegelegten Analysenprinzipien für Getreidesorten mit einer dem Reis ähnlichen Zusammensetzung wurde demonstriert. Die Verlässlichkeit der mit dem Screeningverfahren erhaltenen Ergebnisse wurde durch die vergleichende Untersuchung von Reis-, Mais- und Gerstenproben mit bekannten bzw. vorhersehbaren Unterschieden in der Metabolitenzusammensetzung bestätigt. Zur vergleichenden Auswertung der komplexen Daten wurde eine Software erarbeitet.

Sie ermöglicht eine Normalisierung von Chromatogrammen mit Hilfe von Werkzeugen zur automatischen und manuellen Korrektur von Rententionszeit- und Responseunterschieden. Dies erlaubt die Aufnahme standardisierter Datensätze in Datenbanken und schließlich einen automatisierten Chromatogrammvergleich. Bei der vergleichenden Untersuchung von gentechnisch verändertem und konventionellem Reis wurde gezeigt, dass mit Hilfe des Screeningverfahrens statistisch signifikante Unterschiede in der Metabolitenzusammensetzung zwischen verschiedenen Genotypen detektiert werden können. Zur Beurteilung des Ausmaßes von Unterschieden können Datenbanken herangezogen werden, die durch Untersuchung einer großen Anzahl von Reissorten erstellt wurden.

Das Screening eines breiten Spektrums niedermolekularer Verbindungen in Kombination mit der Anwendung leistungsfähiger Methoden zur Datenauswertung erhöht die Wahrscheinlichkeit, unerwartete Effekte in gentechnisch veränderten Lebensmitteln zu detektieren.

8 References

[1] Heß D. Biotechnologie der Pflanzen – Eine Einführung. Eugen Ulmer, Stuttgart, Germany, pp. 180-200, **1992**

[2] Redenbaugh K., Hiatt W., Martineau B., Emlay D. Regulatory assessment of the FLAVR SAVR tomato. Trends Food Sci. Technol. 1994, 5:105-110.

[3] James C. Preview: Global status of commercialized transgenic crops: 2002. ISAAA Briefs No. 27, International Service for the Acquisition of Agri-Biotech Applications, Ithaca, NY, **2002**

[4] Gianessi L.P. and Carpenter J.E. Agricultural biotechnology: Updated benefit estimates. January 2001, National Centre for Food and Agricultural Policy, Washington, DC, **2001**

[5] Engel K.-H., Frenzel Th., Miller A. Current and future benefits from the use of GM technology in food production. Toxicology Letters 2002, 127:329-336.

[6] Kuiper H.A., Kleter G.A., Noteborn H.P.J.M., Kok E.J. Assessment of the food safety issues related to genetically modified foods. Plant J. 2001, 27:503-528.

[7] Kuiper H.A., Noteborn H.P.J.M., Pejinenburg A.A.C.M. Adequacy of methods for testing the safety of genetically modified foods. Lancet 1999, 354:1315-1316.

[8] Safety evaluation of foods derived by modern biotechnology: concepts and principles. Organization for Economic Co-Operation and Development, Paris, **1993**

[9] Workshop – Food safety evaluation. Organization for Economic Co-Operation and Development, Paris, **1996**

[10] Report of a joint consultation: Biotechnology and food safety. World Health Organization, Geneva, **1997**

[11] Safety aspects of genetically modified foods of plant origin. Report of a joint FAO/WHO expert consultation on foods derived from biotechnology. World Health Organization, Geneva, **2000**

[12] Guidance document on the information needed for the risk assessment of genetically modified plants and derived food and feed. European Commission, Health & Consumer Protection Directorate-General, Scientific Steering Committee, **2003**

[13] Lavrik P.B., Bartnicki D.E., Feldman J., Hammond B.G., Keck P.J., Love S.L., Naylor M.W., Rogan G.J., Sims S.R., Fuchs R.L. Safety assessment of potatoes resistant to Colorado potato beetle. In: Engel K.-H., Takeoka G.R., Teranishi R. (eds.) Genetically Modified Foods: Safety Issues, The American Chemical Society, Washington, DC, pp. 148-158, **1995**

[14] Hamilton K.A., Goodman R.E., Fuchs R.L. Safety assessment of insect-protected cotton. In: Thomas J.A. and Fuchs R.L. (eds.) Biotechnology and safety assessment, 3rd edition, Academic Press, London, UK, pp. 436-465, **2002**

[15] Glennon J.R., Bookout J.T., Duncan D.R., Fuchs R.L. Lavrik P.B., Love S.L., Mueth M., Olson T., Owens E.D., Raymond P.J., Zalewski J. Compositional analysis of tubers from insect and virus resistant potato plants. J. Agric. Food Chem. 2000, 48:5936-5945.

[16] Millstone E., Brunner E., Mayer S. Beyond substantial equivalence. Nature 1999, 401:525-526.

[17] European network safety assessment of genetically modified food crops. http://www.entransfood.com

[18] Kochian L.V., Garvin D.F. Agricultural approaches to improve phytonutrient content in transgenic plants: an overview. Nutr. Rev. 1999, 57:13-18.

[19] Down R.E., Gatehouse A.M.R., Hamilton W.D.O., Gatehouse J.A. Snowdrop lectin inhibits development and decreases fecundity of the glasshouse potato aphid (*Aulacorthum solani*) when administered *in-vitro* and via transgenic plants both in laboratory and in glass house trials. J. Insect Physiol. 1996, 42:1035-1045.

[20] Leple J.C., Bonade-Bottino M., Augustin S., Pilate G., Le Tan V.D., Delplanque A., Cornu D., Jouanin L.,. Toxicity *to Chrysomela tremulae* (Coleoptera: Chrysomelidae) of transgenic poplars expressing a cysteine proteinase inhibitor. Mol. Breeding 1995, 1:319-328.

[21] Takahashi M., Nakanishi H., Kawasaki S., Nishizawa N.K., Mori S. Enhanced tolerance of rice to low iron availability in alkaline soils using barley nicotianamine aminotransferase genes. Nat. Biotechnol. 2001, 19:466-469.

[22] De la Fuente J.M., Ramirez-Rodriguez V., Cabera-Ponce J.L., Herrera-Estrella L. Aluminum tolerance in transgenic plants by alteration of citrate synthesis. Science 1997, 276:1566-1568.

[23] Karakas B. Salinity and drought tolerance of mannitol accumulating transgenic tobacco. Plant Cell Environ. 1997, 20:609-616.

[24] Van Breusegem F., Slooten L., Stassart J.M., Moens T., Botterman J., Van Mantagou M., Inze D. Overproduction of *Arabidopsis thaliana* FeSOD confers oxidative stress tolerance to transgenic maize. Plant Cell Physiol. 1999, 40:512-523.

[25] Apse M.P., Aharon G.S., Snedden W.A., Blumwald E. Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in *Arabidopsis*. Science 1999, 285:1256-1258.

[26] Mitra J. Genetics and genetic improvement of drought resistance in crop plants. Curr. Sci. 2001, 80:758-763.

[27] Garg A.K., Kim J.K., Owens T.G., Ranwala A.P., Choi Y.D., Kochian L.V., Wu R.J. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc. Natl. Acad. Sci. 2002, 99:15889-15903.

[28] Kossmann J., Abel G., Buttcher V., Duwening E., Emmermann M., Loberth R., Springer F., Virgin I., Welsh T., Willmitzer L. Transgenic plants as a tool to understand starch biosynthesis. In: Petersen S.B., Svensson B., Pedersen S. (eds.) Carbohydrate Bioengineering, Elsevier Science, Amsterdam, The Netherlands, pp. 271-278, **1995**

[29] Hellwege E.M., Czapla S., Jahnke A., Willmitzer L., Heyer A.G. Transgenic potato (*Solanum tuberosum*) tubers synthesize the full spectrum of inulin molecules naturally occurring in globe artichoke (*Cynara scolymus*) roots. Proc. Natl. Acad. Sci. 2000, 97:8699-8704.

[30] Friedt W., Lühs W. Recent developments and perspectives of industrial rapeseed breeding. Lipid 1998, 100:219-226.

[31] Hongtrakul V., Slabaugh M.B., Knapp S.J. A seed specific Δ-12 oleate desaturase gene is duplicated, rearranged, and weakly expressed in high oleic acid sunflower lines. Crop Sci. 1998, 38:1245-1249.

[32] De Lumen B.O., Krenz D.C., Jamela M. Molecular strategies to improve the protein quality of legumes. Food Technol. 1997, 51:67-70.

[33] Vail I.K., Anderson O.D. Genetic engineering of wheat gluten. Trends Plant Sci. 1997, 2:292-297.

[34] Zeh M., Casazza A.P., Kreft O., Rössner U., Bieberich K., Willmitzer L., Höfgen R., Hesse H. Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. Plant Physiol. 2001, 127:792-802.

[35] Gawienowski A.M. Integration of the metabolic pathways of steroids, carotenoids, and retinoids. Crit. Rev. Biochem. Mol. Biol. 1999, 34:405-410.

[36] Ye X., Al-Babili S., Klöti A., Zhang J., Lucca P., Beyer P., Potrykus I. Engineering the provitamin A (betacarotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 2000, 287:303-305.

[37] Lucca P., Wünn J., Hurrell R.F., Potrykus I. Development of iron-rich rice and improvement of its absorption in humans by genetic engineering. J. Plant Nutr. 2000, 23:1983-1988.

[38] Dixon R.A., Steele C.L. Flavanoids and isoflavanoids - a gold mine for metabolic engineering. Trends Plant Sci. 1999, 4:394-400.

[39] Regulation (EC) No 258/97 of the European Parliament and of the Council of 27 January 1997 concerning novel foods and novel food ingredients. Official Journal of the European Commission 1997, L43:1-7.

[40] Consensus document on compositional considerations for new varieties of soybean: Key food and feed nutrients and antinutrients. Environment, Health and Safety Publications, Series on the Safety of Novel Foods and Feeds No 2. Environment Directorate, OECD, Paris, **2001**
http://www.olis.oecd.org/olis/2001doc.nsf/LinkTo/env-jm-mono(2001)15

[41] Consensus document on key nutrients and key toxicants in low erucic acid rapeseed (canola). Environment, Health and Safety Publications, Series on the Safety of Novel Foods and Feeds No 1. Environment Directorate, OECD, Paris, **2001**
http://www.olis.oecd.org/olis/2001doc.nsf/LinkTo/env-jm-mono(2001)13

[42] Schilter B. and Constable A. Regulatory control of genetically modified (GM) foods: likely developments. Toxicology Letters 2002, 127:341-349.

[43] Fuchs R.L., Re D.B., Rogers S.G., Hammond B.G. Safety evaluation of glyphosate tolerant soybeans. OECD documents - Food safety evaluation, OECD, Paris, pp. 61-70, **1996**

[44] Hashimoto W., Momma K., Katsube T., Ohkawa Y., Ishige T., Kito M., Utsumi S., Murata K. Safety assessment of genetically engineered potatoes with designed soybean glycinin: compositional analyses of the potato tubers and digestibility of the newly expressed protein in transgenic potatoes. J. Sci. Food Agric. 1999, 79:1607-1612.

[45] Momma K., Hashimoto W., Ozawa S., Kawai S., Katsube T., Takaiwa F., Kito M., Utsumi S., Murata K. Quality and safety evaluation of genetically engineered rice with soybean glycinin: analyses of the grain composition and digestibility of glycinin in transgenic rice. Biosci. Biotechnol. Biochem. 1999, 63:314-318.

[46] Engel K.-H., Gerstner G., Roß A. Investigation of glycoalkaloids in potatoes as example for the principal of substantial equivalence. In: Novel Foods regulation in the EU - Integrity of the process of safety evaluation. Federal Institute of Consumer Health Protection and Veterinary Medicine, Berlin, pp. 197-209, **1998**

[47] Shewmaker C.K., Sheely J.A., Daley M., Colburn S., Ke D.Y. Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. Plant J. 1999, 22:401-412.

[48] Fiehn O., Kloska S., Altmann T. Integrated studies on plant biology using multiparallel techniques. Curr. Opin. Biotechnol. 2001, 12:82-86.

[49] Van Hal N.L.W., Vorst O., Van Houwelingen A.M.M.L., Kok E.J., Peijnenburg A., Aharoni A., Van Tunen A.J., Keijer J. The application of DNA microarrays in gene expression analysis. J. Biotechnol. 2000, 78:271-280.

[50] Anderson J.S., Mann M. Functional genomics by mass spectrometry. FEBS Lett. 2000, 480:25-31.

[51] Raamsdonk L.M., Teusink B., Broadhurst D., Zhang N., Hayes A., Walsh M.C., Berden J.A., Brindle K.M., Kell D.B., Rowland J.J., Westerhoff H.V., van Dam K., Oliver S.G. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat. Biotechnol. 2001, 19:45-50.

[52] Taylor J., Goodacre R., Wade W.G., Rowland J.J., Kell D.B. The deconvolution of pyrolysis mass spectra using genetic programming: application to the identification of some *Eubacterium* species. FEMS Microbiol. Lett. 1998, 160:237-246.

[53] Johnson H.E., Gilbert R.J., Winson M.K., Goodacre R., Smith A.R., Rowland J.J., Hall M.A., Kell D.B. Explanatory analysis of the metabolome using genetic programming of single interpretable rules. Genet. Program Evolv. Mach. 2000, 1:243-258.

[54] Fraser P.D., Pinto M.E.S., Holloway D.E., Bramley P.M. Application of high performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. Plant J. 2000, 24:551-558.

[55] Kawano S. Application to agricultural products and foodstuffs. In: Siesler H.W., Ozaki Y., Kawata S., Heise H.M. (eds.) Near infrared spectroscopy, Wiley-VCH, Weinheim, Germany, pp. 269-287, **2002**

[56] Osborne B.G., Fearn T., Hindle P.H. Practical NIR spectroscopy with applications in food and beverage analysis, 2nd edition, Longman, Harlow, UK, **1993**

[57] Kawata S. Instrumentation for Near-Infrared Spectroscopy. In: Siesler H.W., Ozaki Y., Kawata S., Heise H.M. (eds.) Near infrared spectroscopy, Wiley-VCH, Weinheim, Germany, pp. 43-73, **2002**

[58] Anonymus. User manual for Spectrum Quant software, Perkin Elmer, Rodgau Jügesheim, Germany, **2000**

[59] Heise H.M., Winzen R. Chemometrics in Near-Infrared Spectroscopy. In: Siesler H.W., Ozaki Y., Kawata S., Heise H.M. (eds.) Near infrared spectroscopy, Wiley-VCH, Weinheim, Germany, pp. 43-73, **2002**

[60] De la Roza B., Martinez A., Santos B. The estimation of crude protein and dry matter degradability of maize and grass silages by near infrared spectroscopy. J. Near Infrared Spectrosc. 1998, 6:145-151.

[61] Fontaine J., Hörr J., Schirmer B. Near-Infrared Reflectance Spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products and poultry meal. J. Agric. Food Chem. 2001, 42:2726-2731.

[62] Fontaine J., Hörr J., Schirmer B. Near-Infrared Reflectance Spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran, and sorghum. J. Agric. Food Chem. 2002, 50:3902-3911.

[63] Tillmann P., Reinhardt T.C. Nahinfrarotspektroskopie (NIRS) zur schnellen Qualitätsbestimmung von Raps am Beispiel der Glucosinolate, Institut für Pflanzenbau und Pflanzenzüchtung, Georg-August-Universität, Göttingen, Germany, **1993**

[64] Büning-Pfaue H., Hartmann R., Harder J., Kehraus S., Urban C. NIR-spectrometric analysis of food. Methodical development and achievable performance values. Fresenius J. Anal. Chem. 1998, 360:832-835.

[65] Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol. Biol. 2002, 48:155-171.

[66] Oliver S.G., Winson M.K., Kell D.B., Baganz R. Systematic functional analysis of the yeast genome. Trends Biotechnol. 1998, 16:373-378.

[67] Mendes P. Emerging bioinformatics for the metabolome. Brief. Bioinform. 2002, 3:134-145.

[68] Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp. Funct. Genom. 2001, 2:155-168.

[69] Katona Z.F., Sass P., Molnár-Perl I. Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography – mass spectrometry. J. Chromatogr. A 1999, 847:91-102.

[70] Sauter H., Lauer M., Fritsch H. Metabolic profiling of plants: A new diagnostic technique. In: Baker D.R., Fenyes J.G., Moberg W.K. (eds.) Synthesis and chemistry of agrochemicals II, ACS Symposium Series 443, American Chemical Society, Washington, DC, pp. 288-299, **1991**

[71] Folch J., Lees M., Stanley G.H.S. A simple method for the isolation and purification of total lipids from animal tissue. J. Biol. Chem. 1957, 226:497-509.

[72] Noteborn H.P.J.M., Lommen A., Weseman J.M., van der Jagt R.C.M., Groenendijk F.P.F. Chemical fingerprinting and *in vitro* toxicological profiling for the safety evaluation of transgenic food crops. In: Horning M. (ed.) Report of the demonstration programme on food safety evaluation of genetically modified foods as a basis for market introduction. Ministry of Economic Affairs, The Hague, The Netherlands, pp. 51-79, **1998**

[73] Rössner U., Wagner C., Kopka J., Trethewey R.N., Willmitzer L. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J. 2000, 23:131-142.

[74] Tretheway R.N. and Willmitzer L. Metabolic profiling: wet chemistry and data processing. Abstract of Lectures, First International Congress on Plant Metabolomics, Wageningen, The Netherlands, **2002**

[75] Tolstikov V.V. and Fiehn O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. Anal. Biochem. 2002, 301:298-307.

[76] Kimura M., Yamamoto T., Yamaguchi, S. Automated metabolic profiling and interpretation of GC/MS data for organic acidemia screening: a personal computer-based system. Tohoku J. Exp. Med. 1999, 188:317-334.

[77] Fiehn O., Kopka J., Dörmann P., Altmann T., Trethewey R.N., Willmitzer L. Metabolite profiling for plant functional genomics. Nat. Biotechnol. 2000, 18:1157-1161.

[78] Fiehn O., Tolstikov V., Costisella B., Einert T., Kose F., Gu Y., Weckwerth W. Metabolomic correlation networks of *Arabidopsis* and potato leaves generated by combining GC-TOF and HILIC-ion trap MS data. Abstract of Lectures, First International Congress on Plant Metabolomics, Wageningen, The Netherlands, **2002**

[79] Levsen K., Preiss A., Godejohann M. Application of high-performance liquid chromatography coupled to nuclear magnetic resonance and high performance liquid chromatography coupled to mass spectrometry to complex environmental samples. Trends Anal. Chem. 2000, 19:27-47.

[80] Mass spectra library of The National Institute of Standards and Technology, NIST, Gaithersburg, MD, **1995**

[81] Metabolite mass spectra library, Max-Planck-Institute of Molecular Plant Physiology, Golm, Germany, **2000**
http://www.mpimp-golm.mpg.de/mms-library/tms-e.html

[82] Hughey C.A., Rodgers R.P., Marshall A.G. Resolution of 11,000 compositionally distinct components in a single electrospray ionization fourier transform ion cyclotron resonance mass spectrum of crude oil. Anal. Chem. 2002, 74:4145-4149.

[83] Stein S.E. An integrated method for spectrum extraction and compound identification from gas chromatography / mass spectrometry data. J. Am. Soc. Mass Spectrom. 1999, 10:770-781.

[84] Rössner U., Lüdemann A., Brust D., Fiehn O., Linke T., Willmitzer L., Fernie A.R. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. The Plant Cell 2001, 13:11-29.

[85] Rössner U., Willmitzer L., Fernie A.R. High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. Plant Physiol. 2001, 127:749-764.

[86] Horning M.C. and Horning M.G. Metabolic profiles: gas-phase methods for analysis of metabolites. Clin. Chem. 1971, 17:802-809.

[87] Tretheway R.N., Krotzky A.J., Willmitzer L. Metabolic profiling: a rosetta stone for genomics? Curr. Opin. Plant Biol. 1999, 2:83-85.

[88] Noteborn H.P.J.M., Lommen A., van der Jagt R.C., Weseman J.M. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. J. Biotechnol. 2000, 77:103-114.

[89] Shu Q., Ye G., Cui H., Cheng X., Xiang Y., Wu D., Gao M., Xia Y., Hu C., Sardana R., Altosaar I. Transgenic rice plants with a synthetic *cry1Ab* gene from *Bacillus thuringiensis* were highly resistant to eight lepidopteran rice pest species. Mol. Breed. 2000, 6:433-439.

[90] Wu G., Cui H., Ye G., Xia J., Sardana R., Cheng X., Li Y., Altosaar I., Shu Q. Inheritance and expression of the *cry1Ab* gene in Bt (*Bacillus thuringiensis*) transgenic rice. Theor. Appl. Genet. 2002, 104:727-734.

[91] Rao K.V., Rathore K.S., Hodges T.K., Fu X., Stoger E., Sudhakar D., Williams S., Christou P., Bharathi M., Brown D.P., Powell K.S., Spence J., Gatehouse A.M.R., Gatehouse J.A. Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper. Plant J. 1998, 15:469-477.

[92] Roß A., Eder J., Zellner A., Engel K.-H. Sicherheitsbewertung gentechnisch veränderter Lebensmittel - Einfluß von Sorte, Standort und Reifegrad auf Inhaltsstoffe von Bt-Mais. Schriftenreihe der Bayerischen Landesanstalt für Bodenkultur und Pflanzenbau 1999, 3:35-46.

[93] Amtliche Sammlung von Untersuchungsverfahren nach § 35 LMBG. L 16.01-1. Beuth Verlag GmbH, Berlin, Germany, **2000**

[94] Narziss L. Die Technologie der Malzbereitung, 7th edition, Enke, Stuttgart, Germany, pp. 439-440, **1999**

[95] Zimmermann M.B., Hurrell R.F. Improving iron, zinc and vitamin A nutrition through plant biotechnology. Curr. Opin. Biotechnol. 2002, 13:142-145.

[96] Mendoza C., Viteri F.E., Lönnerdal B., Young K.A., Raboy V., Brown K.H. Effect of genetically modified, low-phytic acid maize on absorption of iron from tortillas. Am. J. Clin. Nutr. 1998, 68:1123-1127.

[97] Adams C.L., Hambidge M., Raboy V., Dosch J.A., Siam L., Westcott J.L., Krebs N.F. Zinc absorption from a low-phytic acid maize. Am. J. Clin. Nutr. 2002, 76:556-559.

[98] Tuntawiroon M., Sritongkul N., Rossander-Hultén L., Pleehachinda R., Suwanik R., Brune M., Hallberg L. Rice and iron absorption in man. Eur. J. Clin. Nutr. 1990, 44:489-497.

[99] Hurrell R.F. Influence of vegetable protein sources on trace element and mineral bioavailability. J. Nutr. 2002, in press.

[100] Raboy V. Low-phytic-acid grains. Food Nutr. Bull. 2000, 21:423-427.

[101] Lucca P., Hurrell R., Potrykus I. Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. Theor. Appl. Genet. 2001, 102:392-397.

[102] Matthäus B., Lösing R., Fiebig H.-J. Determination of inositol phosphates IP3 – IP6 in rapeseed and rapeseed meal by an HPLC method, Part 1: Method. Fat Sci. Technol. 1995, 97:289-291.

[103] Hunt J.R., Johnson L.K., Juliano B.O. Bioavailability of zinc from cooked Philippine milled, undermilled, and brown rice, as assessed in rats by using growth, bone zinc, and zinc-65 retention. J. Agric. Food Chem. 2002, 50:5229-5235.

[104] Oberleas D. and Harland B.F. Analytical methods for phytate. In: Graf E. (ed.) Phytic acid: Chemistry & Applications. Pilatus Pr., Minneapolis, MN, pp. 77-100, **1986**

[105] Parrish F.W., Madacsi J.P., Phillippy B.Q., Wilfred A.G., Buco S.M. Determination of phytic acid in cottonseed by Near-Infrared Reflectance Spectroscopy. J. Agric. Food Chem. 1990, 38:407-409.

[106] De Boever J.L., Eeckhout W., Boucque C.V. NIRS determination of phytate phosphorous in vegetable feedstuffs. Neth. J. Agric. Sci. 1994, 4:357-369.

[107] Latta M., Eskin M. A simple and rapid colorimetric method for phytate determination. J. Agric. Food Chem. 1980, 28:1313-1315.

[108] Juliano B.O., Bechtel D.B. The rice grain and its gross composition. In: Juliano B.O. (ed.) Rice – chemistry and technology, 2nd edition, The American Association of Cereal Chemists, St. Paul, MN, pp. 17-58, **1985**

[109] Kikunaga S., Katoh Y., Takahashi M. Biochemical changes in phosphorus compounds and in the activity of phytase and α-amylase in the rice (*Oryza sativa*) grain during germination. J. Sci. Food and Agric. 1991, 56:335-343.

[110] Yoshida T., Tanaka K., Kasai Z. Phytase activity associated with isolated aleurone particles of rice grains. Agric. Biol. Chem. 1975, 39:289-290.

[111] Perez D.P., Sanchez M.T., Cano G., Carrido A. Authentication of green *Asparagus* varieties by Near-Infrared Reflectance Spectroscopy. J. Food Sci. 2001, 66:323-327.

[112] Smith T.N., Pesti G.M., Bakalli R.I., Kilburn J., Edwards H.M. The use of Near Infrared Reflectance Spectroscopy to predict the moisture, nitrogen, calcium, total phosphorus, gross energy, and phytate phosphorus contents of broiler excreta. Poultry Sci. 2001, 80:314-319.

[113] Kwon Y.K. and Cho R.K. Identification of rice varieties using NIR spectroscopy. J. Near Infrared Spectrosc. 1998, 6:67-73.

[114] Turza S., Toth A.I., Varadi M. Multivariate classification of different soyabean varieties. J. Near Infrared Spectrosc. 1998, 6:183-187.

[115] Hurburgh C.R., Rippke G.R., Heithoff C., Roussel S.A., Hardy C.L. Detection of genetically modified grains by near infrared spectroscopy, Proceedings PITTCON 2000 – Science for the 21st Century, #1431, New Orleans, FA, pp. 12-17, **2000**

[116] Biedermann M., Grob K., Mariani C. Transesterification and on-line LC-GC for determining the sum of free and esterified sterols in edible oils and fats. Fat Sci. Technol. 1993, 95:127-133.

[117] Schmarr H.G., Gross H.B., Shibamoto T. Analysis of polar cholesterol oxidation products: evaluation of a new method involving transesterification, solid phase extraction, and gas chromatography. J. Agric. Food Chem. 1996, 44:512-517.

[118] Amtliche Sammlung von Untersuchungsverfahren nach § 35 LMBG. L 16.01-3. Beuth Verlag GmbH, Berlin, Germany, **2000**

[119] Seetharamaiah G.S. and Prabhakar J.V. Oryzanol content of Indian rice bran oil and its extraction from soap stock. J. Food Sci. Technol. 1986, 23:270-273.

[120] Heberle J. and Simchen G. Silylating agents, 2nd edition, Fluka Chemie AG, Buchs, Switzerland, **1995**

[121] Clay K.L. and Murphy R.C. New procedure for isolation of amino acids based on selective hydrolysis of trimethylsilyl derivatives. J. Chromatogr. 1979, 164:417-426.

[122] West I.R. and Moskowitz G.J. Improved gas chromatography method for the quantitation of saccharides in enzyme-converted corn syrups. J. Agric. Food Chem. 1977, 25:830-832.

[123] Nikolov Z.L. and Reilly P.J. Isothermal capillary column gas chromatography of trimethylsilyl disaccharides. J. Chromatogr. 1983, 254:157-162.

[124] Gehrke C.W. and Leimer K. Trimethylsilylation of amino acids – derivatization and chromatography. J. Chromatogr. 1971, 57:219-238.

[125] Xu Z. and Godber J.S. Purification and identification of components of γ-oryzanol in rice bran oil. J. Agric. Food Chem. 1999, 47:2724-2728.

[126] Kamal-Eldin A. Seed lipids of *Sesamum indicum* and related wild species in Sudan. The Sterols. J. Sci Food Agric. 1992, 59:327-334.

[127] Toivo J.T., Philips K., Lampi A.-M., Piironen V. Determination of sterols in foods: recovery of free, esterified, and glycosidic sterols. J. Food Comp. Anal. 2001, 14:631-643.

[128] Starke I., Holzberger A., Kamm B., Kleinpeter E. Qualitative and quantitative analysis of carbohydrates in green juices (wild mix grass and alfalfa) from a green biorefinery by gas chromatography / mass spectrometry. Fresenius J. Anal. Chem. 2000, 367:65-72.

[129] Chemometrics package "Pirouette", Infometrix Inc., Woodinville, WA, **2002**

[130] Gorenstein M.V., Cohen S.M., Henriksen J.R. Chromatographic pattern matching – a new capability for Millenium^{32}, Waters Corporation, Milford, MA, **2000**

[131] Kováts E. Gas chromatographic characterization of organic substances in the retention index system. Advan. Chromatog. 1965, 1:229-47.

[132] Sapirstein H.D. and Bushuk W. Computer-aided analysis of gliadin electropherograms. I. Improvement of precision of relative mobility determination by using a three reference band standardization. Cereal Chem. 1985, 62:372-377.

[133] Wrigley C.W., Batey I.L., Bekes F., Gore P.J., Margolis J. Rapid and automated characterisation of seed genotype using Microgrand electrophoresis and pattern-matching software. Appl. Theor. Electrophor. 1992, 3:69-72.

[134] Nakatsuka K. and Shinei I. Data processing system for chromatography. United States Patent. US 4.740.930., **1985**

[135] XLSTAT Add-in for Microsoft Excel, Addinsoft, Paris, France.

[136] Desire F., German B., Watzke H., Pfeifer A., Saguy S. Bioinformatics and data knowledge: the new frontiers for nutrition and foods. Trends Food Sci. Technol. 2002, 12:215-229.

[137] Frenzel Th., Miller A., Engel K.-H. Metabolite profiling – a fractionation method for analysis of major and minor compounds in rice grains. Cereal Chem. 2002, 79:215-221.

[138] Frenzel Th., Miller A., Engel K.-H. A methodology for automated comparative analysis of metabolite profiling data. Eur. Food Res. Technol., in press.

[139] Frenzel Th., Miller A., Engel K.-H. Metabolite profiling: analysis of major and minor rice grain constituents by means of GC-FID/MS and automatic data treatment. In: Pfannhauser W., Fenwick G.R., Khokhar S. (eds.), Biologically-active phytochemicals in food. The Royal Society of Chemistry, Cambridge, UK, pp. 271-273, **2001**

[140] Gertz C., Klostermann S., Kochar S.P. Testing and comparing oxidative stability of vegetable oils and fats at frying temperature. Eur. J. Lipid Sci. Technol. 2000, 102:543-551.

[141] Nakayama S., Manabe A., Suzuki J., Sakamoto K. Comparative effects of two forms of gamma-oryzanol in different sterol compositions on hyperlipidemia induced by cholesterol diet in rats. Japan J. Pharmacol. 1987, 44:135-143.

[142] Xu Z., Hua N., Godber J.S. Antioxidant activity of tocopherols, tocotrienols, and γ-oryzanol components from rice bran against cholesterol oxidation accelerated by 2,2'-Azobis(2-methylpropionamidine). J. Agric. Food Chem. 2001, 49:2077-2081.

[143] Miller A., Frenzel Th., Schmarr H.-G., Engel K.-H. Development of an on-line LC-GC method for the rapid analysis of γ-oryzanol in rice bran oil. In: Pfannhauser W., Fenwick G.R., Khokhar S. (eds.). Biologically-active phytochemicals in food. The Royal Society of Chemistry, Cambridge, UK, pp. 274-276, **2001**

[144] Miller A., Frenzel Th., Schmarr H.-G., Engel K.-H. Coupled liquid chromatography-gas chromatography for the rapid analysis of γ-oryzanol in rice lipids. J. Chromatogr. A 2003, 985:403-410.

[145] Tanaka A., Kato A., Tsuchiya T. Isolation of methyl ferulate from rice bran oil. J. AOCS 1971, 48:95-97.

[146] Krishna A.G.G, Khatoon S., Shiela P.M., Sarmandal C.V., Indira T.N., Mishra A. Effect of refining of crude rice bran oil on the retention of oryzanol in the refined oil. J. AOCS 2001, 78:127-131.

[147] Schulte E., Weber K. Rapid preparation of fatty acid methyl esters from fats with trimethylsulfonium hydroxide or sodium methylate. Fett Wissenschaft Technologie 1989, 91:181-183.

[148] Jones M. and Pierce J.S. Biochemistry of amino acid production in germinating barley. A. S. B. C. Proceedings 1966, 10-17.

[149] Mändl B., Franzky G., Wullinger F., Wagner D., Piendl A. Zusammensetzung verschiedener Malzfraktionen an Aminosäuren. Brauwelt 1973, 113:770-776.

[150] Jones M. and Pierce J.S. Amino acids in brewing. Development of individual amino acids during malting and mashing. Eur. Brew. Conv., Proc. Congr. 1963, 11:101-134.

[151] Jones M. and Pierce J.S. Some factors influencing the individual amino acid composition of wort. A. S. B. C. Proceedings 1964, 130-136.

[152] Piendl A. Enzyme activities in germinating barley. Proc. Biochem. 1971, 6:19-26.

[153] Mändl B., Wullinger F., Wagner D., Piendl A. Einfluss der Abdarrtemperatur auf den Aminosäuregehalt der Malze und Würzen von verschiedenen Gerstensorten. Brauwelt 1970, 110:437-443.

[154] Jones M. and Pierce J.S. The role of proline in the amino acid metabolism of germinating barley. J. Inst. Brew. 1967, 73:577-583.

[155] Schormüller J., Brandenburg W., Langer H. Organische Säuren in Kaffee und Kaffee-Ersatzstoffen sowie in Trockenpulver aus Kaffee-Ersatzstoffen und Kaffee. Z. Lebensm. Untersuch. Forsch. 1961, 115:226-235.

[156] Stryer L. Biochemie, 4th edition, Spektrum, Heidelberg, Germany, p. 540, **1996**

[157] Gehlhoff R. and Piendl A. Einfluss der Abdarrtemperatur auf die Zusammensetzung des Malzes an Glycerin und organischen Säuren von verschiedenen Gerstensorten. Brauwissenschaft 1971, 24:109-110.

[158] Latzko E. and Kotze J.P. Veränderungen in Gehalt von D-Glucose, D-Fructose, Saccharose, Raffinose und anderen β-D-Fructosiden sowie ihre Verteilung in keimender Sommergerste. Brauwissenschaft 1965, 18:93-95.

[159] Piendl A. Carbohydrate composition of malt. Proc. Biochem. 1971, 6:32-33.

[160] Juliano B.O. Polysaccharides, proteins, and lipids of rice. In: Juliano B.O. (ed.) Rice – chemistry and technology. 2nd edition. The American Association of Cereal Chemists, St. Paul Minnesota, p. 148, **1985**

[161] Ketterer M., Miedaner H., Narziss L. Lipidstoffwechsel von Braugerste beim Mälzungsprozess. BioEngineering 1992, 8:48-52.

[162] Jensen S.A., Heltved F. Visualization of enzyme activity in germinating cereal seeds using a lipase sensitive fluorochrome. Carlsberg Res. Commun. 1982, 47:177-181.

[163] Danier J. Bayerische Hauptversuchsanstalt für Landwirtschaft, personal communication, **2002**

[164] Kleter G.A., Noordam M.Y., Kok E.J., Kuiper H.A. New developments in crop plant biotechnology and their possible implications for food products safety. RIKILT Report 2000.004. National Institute for Quality Control of Agricultural Products (RIKILT), Wageningen, The Netherlands, **2000**
www.rikilt.wageningen-ur.nl/Publications/Publications/Tekstrapport2000%20004.htm

[165] Trethewey R.N. Gene discovery via metabolic profiling. Curr. Opin. Biotechnol. 2001, 12:135-138.

[166] Lepri O., Bassie L., Safwat G., Thu-Hang P., Trung-Nghia P., Hölttä E., Christou P., Capell T. Over-expression of a cDNA for human ornithine decarboxylase in transgenic rice plants alters the polyamine pool in a tissue specific manner. Mol. Genet. Genomics 2001, 266:303-312.