KUOPION YLIOPISTON JULKAISUJA C. LUONNONTIETEET JA YMPÄRISTÖTIETEET 203
KUOPIO UNIVERSITY PUBLICATIONS C. NATURAL AND ENVIRONMENTAL SCIENCES 203

SATU LEHESRANTA

Proteomics in the Detection of Unintended Effects in Genetically Modified Crop Plants

Doctoral dissertation

To be presented by permission of the Faculty of Natural and Environmental Sciences of the University of Kuopio for public examination in Auditorium L22, Snellmania building, University of Kuopio, on Saturday 16th December 2006, at 12 noon

Institute of Applied Biotechnology
University of Kuopio

KUOPION YLIOPISTO

KUOPIO 2006

Distributor: Kuopio University Library
P.O. Box 1627
FI-70211 KUOPIO
FINLAND
Tel. +358 17 163 430
Fax +358 17 163 410
http://www.uku.fi/kirjasto/julkaisutoiminta/julkmyyn.html

Series Editors: Professor Pertti Pasanen, Ph.D.
Department of Environmental Sciences

Professor Jari Kaipio, Ph.D.
Department of Physics

Author's address: Institute of Applied Biotechnology
University of Kuopio
P.O. Box 1627
FI-70211 KUOPIO
FINLAND
Tel. +358 17 163 068
Fax +358 17 163 148

Supervisors: Professor Sirpa Kärenlampi, Ph.D.
Institute of Applied Biotechnology
University of Kuopio

Professor Howard Davies, Ph.D.
Quality, Health and Nutrition Programme
Scottish Crop Research Institute
Dundee, Scotland, UK

Reviewers: Professor Teemu Teeri, Ph.D.
Department of Applied Biology
University of Helsinki

Harry Kuiper, Dr.
DLO-RIKILT - Institute of Food Safety
Wageningen, The Netherlands

Opponent: Prof. Dr. Karl-Heinz Engel
Technische Universität München
Lehrstuhl für Allgemeine Lebensmitteltechnologie
München, Germany

ISBN 951-27-0681-4
ISBN 951-27-0456-0 (PDF)
ISSN 1235-0486

Kopijyvä
Kuopio 2006
Finland

Lehesranta, Satu. Proteomics in the detection of unintended effects in genetically modified crop plants. Kuopio University Publications C. Natural and Environmental Sciences 203. 2006. 71 p.
ISBN 951-27-0681-4
ISBN 951-27-0456-0 (PDF)
ISSN 1235-0486

Abstract

Within Europe at least genetic modification of crop plants still remains controversial. One of the major issues is the possibility of unintended effects caused, for example, by the site of transgene integration (e.g. interruption of important open reading frames or regulatory sequences). These could result in modified metabolism, novel fusion proteins or other pleiotropic effects which could compromise safety by e.g. producing new allergens or toxins. Comparative safety assessment includes targeted analysis of key nutrients and anti-nutritional factors. However, arguably, unintended effects are less likely to be detected with this conventional targeted analysis of a relatively limited number of molecules compared with non-targeted methods such as transcriptional, protein and metabolite profiling. Thus broader scale profiling or 'omics' methods could increase the chances of detecting possible unintended effects.

Using two-dimensional electrophoresis and mass spectrometry, it is often possible to visualize, quantify and identify hundreds or even thousands of proteins in a given tissue or cell sample. However, little information is still available on the extent of natural variation in the proteome caused by genetic background, environmental influences and other factors. Consideration of the extent of natural variation in the proteome is important in the comparative analysis of genetically modified (GM) crops, because observed differences in GM lines might be well within normal variation observed in non-modified material.

The aim of the present Thesis was to evaluate the applicability of proteomic techniques in characterising several GM potato lines for possible unintended effects. In addition, an insight into the extent of natural variation in potato tuber was provided by analysing a wide range of potato genotypes, the life cycle of the tuber, and the effect of alternative cultivation techniques. The studies conducted here suggest that compared to the extensive natural variation found between different non-GM varieties and landraces, the effects of genetic modification on the proteome are considerably less pronounced. In fact, no clear differences between the protein patterns of the GM lines and their controls were found. Furthermore, major changes in protein profiles due to developmental/physiological stages and, to a lesser extent, alternative agricultural systems were observed.

Universal Decimal Classification: 504.73, 575.113, 577.112, 582.951, 632.8, 633.491
CAB Thesaurus: transgenic plants; genetic engineering; crops; adverse effects; side effects; food safety; pleiotropy; allergens; potatoes; metabolism; proteins; tubers; development; protein modifications; genotypes; genetic variation; recombination; electrophoresis; mass spectrometry; organic farming

Acknowledgments

This work was carried out at the Institute of Applied Biotechnology (previously Department of Biochemistry) at the University of Kuopio during the years 2001–2006.

I am very grateful to all the people who have participated in the work towards this Thesis in some way during these years; they are too many to be all listed here. Especially I wish to thank:

My principal supervisors prof. Sirpa Kärenlampi and prof. Howard Davies for support, discussions and encouragement, especially during the final crunch, and for providing the facilities for the work.

Seppo Auriola for helping me in my analyses and for introducing me to the world of mass spectrometry. Kaisa Koistinen, for teaching me a great deal of practical MS skills and helping in many other ways.

The whole Plant Biotechnology group - it has been a great deal of fun to work with you all. Especially I wish to thank Harri Kokko, Arja Tervahauta, Marjo Tuomainen, Viivi Hassinen, Kati Hanhineva, Mikko Anttonen, Soile Suomalainen and Tuomas Salemaa for all the help they have given me during these years. Eija Sedergren-Varis and Raisa Malmivuori for excellent technical assistance.

Louise Shepherd, Naoise Nunan, Nathalie Massat, Jim McNicol and Carlo Leifert for great collaboration and co-authorship. Special thanks to Jim who has taught me a great deal about the value of statistics and good experimental design.

Professor Ykä Helariutta for kindly supporting the final preparation of my Thesis by providing flexible working conditions.

My parents, Markku and Terttu Lehesranta, for their support and encouragement during my undergraduate and PhD studies.

Finally, my deepest and warmest gratitude to Kaisa for all the support, love and everything else without which this work could not have been done.

This work was financially supported by the European Commission (GMO-CARE, Contract No QLK1-1999-00765; SAFE FOODS, Contract No Food-CT-2004-506446) and a grant from Suomen Kulttuurirahasto (the Finnish Cultural Foundation).

Kuopio, December 2006

Satu Lehesranta

Abbreviations

2-DE	Two-dimensional electrophoresis
ANOVA	Analysis of variance
CaMV	Cauliflower mosaic virus
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CV	Coefficient of variance
cv.	Cultivar
DIGE	Difference gel electrophoresis
DSB	Double-strand break
DTT	Dithiothreitol
EST	Expressed sequence tag
FDR	False discovery rate
FTICR	Fourier transform-ion cyclotron resonance
FT-IR	Fourier transform infrared spectroscopy
GC	Gas chromatography
GM	Genetically modified
HPLC	High-performance liquid chromatography
ICAT	Isotope-coded affinity tag
IPG	Immobilized pH gradient
LC	Liquid chromatography
MIAME	Minimum information about a microarray experiment
MPSS	Massively parallel signature sequencing
MS	Mass spectrometry
MudPIT	Multidimensional protein identification technology
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
pI	Isoelectric point
NMR	Nuclear magnetic resonance
RNAi	RNA interference
SD	Standard deviation
SAGE	Serial analysis of gene expression
SED	Standard error difference
SELDI	Surface-enhanced laser desorption/ionization
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
T-DNA	Transferred DNA
Ti	Tumour-inducing

List of original publications

This Thesis is based on the following publications referred to in the text by their Roman numerals:

I Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., Nunan, N., McNicol, J.W., Auriola, S., Koistinen, K.M., Suomalainen, S., Kokko, H.I., Kärenlampi, S.O. (2005) Comparison of tuber proteomes of potato (*Solanum* sp.) varieties, landraces and genetically modified lines. *Plant Physiology* 138, 1690–1699.

II Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., Koistinen, K.M., Massat, N., Nunan, N., McNicol J.W., Kärenlampi, S.O. (2006) Proteomic analysis of the potato tuber life cycle. *Proteomics* 6, 6042–6052.

III Lehesranta, S.J., Koistinen, K.M., Massat, N., Davies, H.V., Shepherd, L.V.T., McNicol, J.W., Cakmak, I., Cooper, J., Lück, L., Kärenlampi, S.O., Leifert, C. (2006) Effects of agricultural production systems and their components on protein profiles of potato tubers. (Submitted for publication)

This Thesis also contains unpublished material.

Contents

1 Introduction **13**

2 Review of the literature **15**
2.1 Transgene integration in the context of natural genetic variation in plants . . . 15
2.1.1 Genetic variation of crop plants . . . 15
2.1.2 The mechanisms of genetic recombination in plants . . . 16
2.1.3 Transgenic techniques introduce new variation . . . 17
2.2 Safety considerations in genetic modification . . . 18
2.2.1 Unintended effects in GM plants . . . 18
2.2.2 Assessment of the safety of GM food crops . . . 22
2.3 Profiling methods and detection of variation . . . 23
2.3.1 Protein profiling . . . 24
2.3.2 Other profiling methods . . . 28
2.3.3 Considerations for the interpretation of profiling data . . . 30
2.3.4 Natural variability of protein profiles in plants . . . 31
2.4 Conclusions . . . 32

3 Aims of the present study **33**

4 Materials and methods **34**
4.1 Plant material from GM and non-GM potato genotypes (**I**) . . . 34
4.2 Plant material for the tuber life cycle analysis (**II**) . . . 34
4.3 Plant material and field experimental design for the analysis of alternative cultivation practices (**III**) . . . 35
4.4 Extraction of total soluble protein (**I, II, III**) . . . 35
4.5 Two-dimensional electrophoresis (**I, II, III**) . . . 36
4.6 Image and data analysis (**I, II, III**) . . . 36
4.7 Protein identification by HPLC-electrospray tandem mass spectrometry (**I, II, III**) . . . 36

5 Results **38**
5.1 Validation of the proteomic procedure . . . 38
5.1.1 Effect of freeze-drying of potato tuber material on 2-DE . 38
5.1.2 Sources of variation in 2-DE of tuber proteins . . . 38
5.2 Protein profiles of genetically modified potato lines (**I**) . . . 40
5.3 Protein profiles of non-GM potato genotypes (**I**) . . . 40
5.4 Comparison of genetically modified and non-modified potato lines and varieties (**I**) . . . 43

5.5 Changes in protein patterns during tuber life cycle (**II**) 44
5.6 The effect of alternative cultivation practices (**III**) 45

6 Discussion **47**
6.1 Applicability and reproducibility of potato protein profiling by 2-DE 47
6.2 Natural variability of protein patterns in potato 48
6.2.1 Variability caused by different genetic backgrounds 48
6.2.2 Variability caused by developmental and physiological states 49
6.2.3 Variability caused by alternative cultivation techniques . . 50
6.3 The effect of genetic modification on the proteome is small 51
6.4 Natural variation appears much higher than variation between GM lines 52

7 Conclusions **53**

References **55**

APPENDIX: Original publications

1 Introduction

It is generally accepted that traditional food is safe for the majority of consumers. To introduce a new variant or cultivar developed from a traditional crop plant, few analyses are required to prove its safety. The situation is quite different if the crop is developed by using genetic modification.

In a majority of cases seen so far, a new gene has been introduced to a non-predetermined location in the genome. It is quite feasible to ask the question whether the new gene products (proteins) are safe or not. Therefore, for all genetically modified (GM) crop plants, the safety of the newly introduced proteins needs to be demonstrated before the plants can be released into the market. Another point of concern is the random integration of the new gene into the plant genome. Both the new gene itself and its site of integration may give rise to unintended adverse effects. For example, transgene integration might interrupt regulatory sequences or open reading frames, modifying the metabolism. These modifications could compromise the safety of the food crops by, for instance, leading to the production of new allergens or toxins. Having the gene and the integration site well characterised provides the basis for the safety assessment. However, it is a common practice today to perform a large number of targeted analyses of key macronutrients, micronutrients, antinutrients and toxins, to demonstrate the substantial equivalence of the novel crop with the conventional counterpart. Still, there is criticism that possible unintended effects are not covered by these targeted analyses.

Integration of exogenous DNA (transgene) occurs via the same mechanism as does the natural recombination. Several types of rearrangements are thus observed both in transgene integration sites and in natural recombination sites. While this mechanism provides natural variation for breeders to select from, it is also a source of unintended effects similar to that in genetically engineered crop plants. In the light of the variation generated by the natural recombination and by the repertoire of traditional breeding technologies exploited for decades, the question arises how much variation might the transfer and integration of a single gene generate in the overall genetic makeup of a crop plant compared to the variation already existing. A related question is how probable are unintended effects that extend beyond this variation.

Non-targeted methods such as transcriptional, protein and metabolite profiling offer potentially unbiased approaches to the detection of unintended effects. How feasible are profiling techniques in general as tools to provide additional data for the risk assessment of GM crops? Do they provide added value worth the investment? Do they give reassurance that unintended adverse effects have not occurred? While it is clear that a comprehensive coverage of all constituents present in a given tissue is difficult to obtain with the current technologies, pro-

teins are the key molecules of interest, as they are potential allergens and catalyse the synthesis of metabolites some of which are potential toxins.

To assess the observed differences in the context of natural variation in the composition, comparative data of 'normal' protein levels are needed to understand the effect of genetic background, developmental stages, physiological states, environmental conditions and cultivation techniques, and to be able to set the criteria against which to determine what is a significant difference worth considering a possible safety risk. Currently there is very little information publicly available on protein patterns in major food crops.

This Thesis evaluates the applicability of proteomic techniques in characterising genetically modified lines for possible unintended effects. Furthermore, it describes natural variation due to different genetic backgrounds, developmental and physiological states and cultivation practices to establish a baseline of protein expression using potato tuber as the model system.

2 Review of the literature

2.1 Transgene integration in the context of natural genetic variation in plants

Transgenic and other molecular techniques are the most recent additions in the plant breeder's toolbox and allow for the introduction of novel genetic material even from distant sources. As both natural recombination and the repertoire of traditional breeding technologies are ample sources of genetic variation, a major question is how much variation might the transfer and integration of a single gene generate in the overall genetic composition of a genetically modified (GM) crop plant compared to natural variation. Another key question is how likely are unintended effects that introduce novel variation beyond what is naturally observed.

2.1.1 Genetic variation of crop plants

Plant breeding has always employed natural or artificially induced genetic variation and selection for introducing new cultivars. As the gene pool used during early domestication and modern breeding has for many crop species been limited, cultivated varieties represent only a small fraction of the variability among their wild relatives [46]. Landraces are the earliest form of cultivars available and, in comparison to modern-day cultivars, are highly heterogeneous. Despite difficulties in interspecific breeding, introgression breeding with wild relatives has made a considerable contribution to the development of modern-day varieties of many important food crops.

In addition to utilising natural variation present in these populations by crosses and selection, modern plant breeding employs intensive methods for the modification of plant genetic composition, such as intervarietal hybrids, wide interspecific crosses, protoplast fusion, mutagenesis (by chemicals or irradiation) and ploidy modification. Each of these has the potential to produce abundant pleiotropic effects on gene structure and trait expression in plants. For instance, it is known that irradiation-induced mutagenesis can lead to chromosomal rearrangements [166] and possibly to increased recombination frequency [105]. It has been shown that interspecies crosses and ploidy modification lead to major genomic changes in wheat [74, 132]. Various species and genera, including many important crop plants, naturally display extensive chromosomal rearrangements, and widely different chromosome numbers have been detected within a species and/or a genus [114]. Genome doubling significantly affects gene expression, resulting in epigenetically induced gene silencing [1]. In general, study of natural and laboratory generated polyploids has revealed extensive and rapid genomic changes within a few generations in some groups, including sequence rearrangements, homoe-

ologous recombination, sequence elimination, changes in DNA methylation and gene silencing, often leading to novel phenotypes not present in the contributing species [112, 131]. The genomes of some species, such as maize, are characterised by high variability due to insertions of transposable elements and differential presence of genes and gene fragments [23]. A comparison of the genomes of maize inbred lines has shown that transposable elements continue to change the genome, profoundly affecting genetic diversity within the species [102, 120]. Apart from affecting genome structure by insertions and excisions, transposable elements have the potential to activate and alter the expression of adjacent genes in wheat [85]. Therefore, it appears that plant genomes are highly dynamic and may by natural or human-induced means undergo major rearrangements, leading to a staggering amount of variability in natural and breeding populations.

2.1.2 The mechanisms of genetic recombination in plants

Recombination is without doubt crucial to plant breeding. There are differences in the recombination rate over portions of genomes [158], and this phenomenon might be of general importance for higher plants and their genome evolution. Gene-rich regions are hot spots for recombination e.g. in wheat [61]. Because of this bias for gene-rich regions, new variation in the form of new alleles with novel characteristics has emerged within plant populations [31] and in crop plants [158]. Studies on the evolution of disease resistance genes [143] and genes influencing quality traits [54] suggest that novel alleles have arisen from the shuffling of sequence domains between members of the gene family in a pattern that is similar to that observed for double-strand break (DSB) repair in plants [27].

Two major genetic recombination mechanisms have been identified in plants, and both mechanisms are currently explained by the DSB repair model [63]. Generally, DSBs can be repaired via two different pathways, either via homologous recombination or via non-homologous end-joining, also known as illegitimate recombination. Non-homologous end joining is the predominant form of recombination in somatic cells of plants [140] and also plays a role in meiotic recombination [81]. Natural DSBs are induced by factors such as transposons, radiation, chemicals and endonucleases [140]. Since the DSB repair system involved in the recombination is more error-prone in plants than in other organisms, errors that change the original sequence occur at a high frequency. DSB repair rarely occurs without any sequence alterations, and usually gives rise to deletions ranging up to more than 1 kb and introduction of new filler DNA [63, 64]. Using DSB repair, various kinds of genomic sequences that are available for a copying process can be inserted into new genomic positions [63, 150]. DSB repair is also a prominent source of deletions.

Integration of exogenous DNA (transgene) appears to utilize the same mech-

anisms as does the natural recombination [179, 167]. Thus it is expected that several types of rearrangements would also be observed in transgene integration sites.

2.1.3 Transgenic techniques introduce new variation

In genetic transformation mediated by *Agrobacterium*, a portion of DNA (T-DNA) present on a Ti plasmid is transferred to the plant cell nucleus [179]. The T-DNA is defined by its left and right borders, 25-base pair direct repeats that, in theory, direct the processing of T-DNA and the genetic material to be integrated [197]. Alternatively, the biolistic method can be used to introduce 'naked' pieces of DNA in species that are recalcitrant to *Agrobacterium* transformation [30]. Currently, it is not possible to introduce a defined number of transgenes into the genome of a higher plant nor target efficiently the foreign DNA to specific positions in the genome, although transgene integration and gene replacement employing homologous recombination may be feasible in the future [180]. While it cannot be guaranteed that the transferred DNA is integrated intact, its structure and integration site can be determined by molecular analyses. Typically, transgenes show variable expression patterns in independent transgenic plants, and often there is little correlation between transgene copy number and expression level [60].

As the integration of exogenous DNA transferred by either transformation technique appears to utilize the DSB repair mechanism, several types of rearrangements are observed both in transgene integration sites and in natural recombination sites. Integration of transgenes into chromosomal DNA can occur either as single copies or repeated and multiple insertions [60], and filler sequences occur between T-DNA repeats and also at the junctions of T-DNA and plant DNA [192]. T-DNA vector 'backbone' sequences are also frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation [91]. At the site of insertion in the plant DNA, chromosomal rearrangements, such as inversions and translocations, have been observed in several species [52, 104, 122, 170, 174]. *Agrobacterium*-mediated transformation results in a higher proportion of transgene loci with a single integrated complete T-DNA and little damage to the genomic DNA, as compared with direct DNA delivery [36]. However, the characterization of transgene integration sites suggests that the resulting variation in transgene locus structures is more likely to be determined by plant genomic factors and the mechanisms of illegitimate recombination than how or in what form the DNA is delivered to the nucleus [167].

An important question for the variation generated by transgene introduction is how frequently transgenes integrate into active genes. Large-scale gene tagging studies in several plant species have shown that gene integration occurs throughout the genome along the length of all chromosomes [52, 111, 177]. However, Alonso

et al. [6] observed a highly nonuniform distribution of T-DNA integration events using genome-wide insertional mutagenesis in *Arabidopsis*, with the density of integration correlating with gene density along each chromosome. Results from e.g. rice corroborate that integration preferably occurs in gene-rich locations along the genome [29]. The proportion of T-DNA insertions disrupting known coding gene sequences in tagging experiments has varied in the range of 24-28% in *Arabidopsis* [52] to 58% in rice [29]. At the gene level, T-DNA integration events in *Arabidopsis* have been suggested to show a preference for insertion in sequence compositions that occur frequently in the 5' and 3' flanking regions compared to protein coding sequences [6, 159].

The stability of expression and the trait is a major concern in the genetic modification of plants [121]. Instability of transgene expression is often associated with complex multicopy patterns of transgene integration at the same locus, as well as position effects due to random integration [92]. However, transcript level-mediated posttranslational silencing is likely to be the main cause for the large variability in transgene expression seen among transformants, rather than position effects [160]. Overexpression of transgenes can result in the silencing of the introduced transgenes and, strikingly, often involves the suppression of homologous endogenous gene copies. It was first reported as variation in petunia flower coloration after the introduction of a chalcone synthase (*chs*) transgene under the control of the 35S promoter [123, 185]. Small interfering RNAs (siRNAs) specific for the transcribed transgene sequence were later shown to be a hallmark of posttranscriptional silencing of transgenes [73]. Recently, a great deal of research has been devoted to explaining the mechanisms of gene silencing. Silencing can result either from abolished transcription of the introduced gene (transcriptional gene silencing) or the degradation of transgene RNA (posttranscriptional gene silencing) [121]. Although it is a problem in some GM applications, it does also provide another tool to add specific traits to plants. As silenced genes can under certain circumstances silence the expression of homologous genes located elsewhere in the genome, the expression of endogenous plant genes can be downregulated [121]. However, research on RNA interference (RNAi) based methods for the suppression of plant genes is very recent and heavily ongoing, and the true potential of these methods for genetic modification of crop plants remains to be seen [98].

2.2 Safety considerations in genetic modification

2.2.1 Unintended effects in GM plants

Currently, there is a substantial public concern about the food safety of GM crops. One issue is the random integration of the new gene into the plant genome and

the possibility that both the new gene itself and its site of integration may give rise to unintended adverse effects. As T-DNA insertion frequently occurs within known coding genes, an insertion may disrupt existing genes, resulting in inactive mutant forms of the gene that may become apparent in later generations when the transgene is homozygous [6, 44, 45]. It is also possible that an insertion results in the production of novel fusion proteins, if the flanking DNA sequence is read through into the inserted DNA or vice versa. Such fusion proteins have been detected at low frequencies using a reporter gene [86]. These modifications could compromise the safety of the food crops by, for instance, leading to the production of new allergens or toxins, at least in theory [95].

Most crops naturally produce allergens, toxins, or other antinutritional substances [7, 125]. A single transgene could therefore in some cases lead to unexpected pleiotropic effects that may influence the expression of inherent plant toxins or allergens, having a direct impact on health or allergenic potential. These effects could be e.g. increased activity of the naturally occurring metabolic pathways, increased synthesis caused by increased gene activation, or reduced decomposition; the possibility of gene activation, especially if it leads to the production of harmful compounds, raises the most concern for food safety [125].

Although not directly due to the integration of a transgene, somaclonal variation may be a problem to the efficiency of transgenic plant production [17]. It was originally defined as genetic and phenotypic variation among clonally propagated plants [103] and is often not meiotically inherited but is only observed in primary regenerants [84]. Meiotically heritable variation is important in situations where the end product is propagated. It is manifested as changes in chromosome structure, polyploidy, phenotypic mutation, sequence change, and gene activation or silencing [106]. The occurrence of somaclonal variation is increased by either prolonged or harsh tissue culture [14, 101].

Novel gene combinations arising from the genetic modification or of existing genes through conventional breeding techniques may, therefore, introduce unintended and unexpected effects. An intended effect may be defined as a result of genetic transformation that is targeted to occur from the introduction of the gene(s) in question and that fulfils the original objective of the transformation process; an unintended effect may then be defined as a 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 [27]. These effects may or may not be explicable in terms of known biology or prove to have relevance to safety, but must be considered in a risk assessment.

Unintended effects are not confined to the use of GM technology, but are observed also as a result of conventional plant breeding and, although they are rare, there are various examples in the literature. However, plant breeders routinely assess phenotypic appearance and agronomic performance of new candidate va-

rieties in the laboratory, glasshouse, and small scale field trials, including GM crops, and discard lines that express undesirable characteristics. Therefore new cultivars produced by genetic modification or by conventional breeding are extensively tested and screened prior to commercial release. This would result in the elimination of major unintended effects which are more easily screened for, possibly leaving the more subtle differences to deal with [27].

Examples of unintended effects both in transgenic and traditionally bred crops are shown in Table 1. In these cases, the effects were demonstrated by phenotype selection or by analysis of defined constituents.

Several examples show that unintended effects can have a negative impact on potential agronomic performance. Such phenotypes are obviously detrimental to any further commercial development of the transgenic lines in question [27]. However, genotype $\times$ environment interactions may be significant, leading to inconsistent differences in performance across different locations. For instance, transgenic *Bt* maize was reported to contain more lignin in stems, possibly reducing its digestibility and feeding value [155], while another experiment did not identify biologically significant differences in lignin content due to the transgene [83].

An example of an unintended effect detected by targeted analysis of several plant compounds comes from GM canola, in which transgenic expression of phytoene synthase not only resulted in an altered level of a metabolite downstream of the target of modification but also in changes in the levels of compounds the biosynthesis of which is linked to the modified pathway [164]. Based on what is known of the biosynthesis of isoprenoids, this is an example of a predicted unintended effect.

Expression of genes in different organisms can potentially result in differences in folding or posttranslational modification of proteins and may also be a concern for safety [80]. In a recently published case, Prescott *et al.* [139] showed that an α-amylase inhibitor from bean, introduced to convey resistance to pea weevils, was differently glycosylated when the gene was introduced to pea, leading to an altered immunogenic reaction in mice.

In the future, GM approaches are likely to produce plants with improved nutritional properties and increased complexity of the genetic modification, with the potential to more far-reaching effects on metabolic processes that cannot easily be predicted by current knowledge of plant biology [27]. For example, overproduction of fruit phytoene synthase to increase carotenoid production led into dwarfism in transgenic tomato plants probably by redirecting metabolites from gibberellin synthesis [53]. Kristensen *et al.* [93] have shown that it is possible to engineer transgenic plants with the insertion of multiple genes expressing a whole biosynthetic pathway with marginal inadvertent effects, while the insertion of an incomplete pathway can lead to significant alterations in plant morphology,

Table 1: Examples of unintended effects both in transgenic and traditionally bred crops.

Crop	Ref.	Unintended effect
Transgenic crops		
Pea	[139]	Expression of a bean α-amylase inhibitor in peas resulted in altered structure and immunogenicity of the protein
Potato	[16]	Potato plants transformed with lectin genes to enhance insect resistance had lower levels of glycoalkaloids in leaves
Rice	[116]	Transgenic rice with reduced glutelin showed increase in the levels of prolamine, compensating reduced total protein
Maize	[155]	Transgenic *Bt* maize contained more lignin in stems, possibly reducing its digestibility
Barley	[77]	Barley transformed with the *bar* gene showed reduced field performance compared to a conventional variety
Rape	[5]	Oilseed rape transformed with a herbicide resistance gene controlled by the cauliflower mosaic virus (CaMV) 35S promoter was rendered herbicide sensitive by CaMV infection
Rice	[119]	Rice expressing transgenic glycinin contained 20% more protein, possibly due to higher glycinin, but also 50% more vitamin B6
Canola	[164]	Transgenic expression of phytoene synthase led to an increase in carotenoids, but also in altered fatty acid composition of canola seeds
Tomato	[53]	Constitutive expression of a fruit phytoene synthase in transgenic tomatoes caused dwarfism by redirecting metabolites from the gibberellin pathway
Potato	[32]	Transgenic potato lines showed unexpected changes in phenotype and field performance, either due to transgene or tissue culture
Traditionally bred crops		
Potato	[196]	Potato variety Lenape contained very high levels of toxic solanine
Barley	[12]	Breakdown of resistance of barley to powdery mildew by a relief of soil water stress was attributed to the genetic background rather than the resistance allele *mlo*
Potato	[186]	Potato breeding lines with wild species in their pedigrees produced novel toxic glycoalkaloids
Celery	[8]	Pest-resistant celery variety contained high levels of psoralen, which produced rashes in agricultural workers
Squash	[87]	High levels of cucurbitacin in traditionally bred squash led to food poisonings

transcriptome and metabolome.

2.2.2 Assessment of the safety of GM food crops

It is generally assumed that traditional food is safe for the majority of consumers. Few analyses are required to prove the safety of a newly introduced cultivar developed from a traditional crop plant. In some cases, e.g. for potato and oilseed rape, maximum limits have been set to the content of known toxins [41]. The situation is quite different if the crop is developed by using genetic engineering, as within Europe at least genetic modification still remains highly controversial.

Since the 1990's, FAO, OECD and WHO have raised the question of how to deal with novel foods, in particular those derived from genetic modification. The principle of substantial equivalence was proposed as a rational and practical approach [126, 43]. The concept of substantial equivalence includes the idea that existing food or plant sources known to be safe can be used as a basis for comparison when assessing the safety of a new or modified food. If it is known that a new food is derived from an organism whose new traits have been well characterised and there is reasonable certainty of no harm as compared with a conventional counterpart, a new food can be considered substantially equivalent and may be treated in the same way regarding safety. When there is extensive knowledge of the range of relevant characteristics of the modified traditional food, the comparison of a new product is straightforward. The application of substantial equivalence becomes more complicated if less background information is available or if there is no directly comparable conventional counterpart [126].

The most direct way to predict unintended effects is to analyse the transgene flanking regions to establish whether the insertion has taken place within or near an endogenous gene. Possible alterations in the phenotype may be identified through a comparative analysis of growth performance, yield, disease resistance, chemical composition etc. [94]. It is a common practice today to perform a large number of analyses, so-called targeted analyses, to demonstrate the substantial equivalence of the novel crop with the conventional counterpart [100]. Targeted analyses include key macronutrients, micronutrients, antinutrients and toxins. For example, for potato, and other Solanaceae in general, a main safety issue is the glycoalkaloid content of the tubers, a concern earlier identified in classical breeding, as glycoalkaloid levels have been known to be affected in crosses of *S. tuberosum* and wild *Solanum* species [186]. In certain cases, toxicity studies on experimental animals are advised [100]; however, the usefulness of animal studies in general food safety testing is limited [43].

The targeted approach has its limitations, as the selection of compounds that can be analysed is restricted and may be biased. Furthermore, it is not possible to detect unknown toxicants or anti-nutrients using targeted analysis [94]. Thus, ar-

guably, unintended effects are less likely to be detected with conventional targeted analysis of a relatively limited number of molecules.

2.3 Profiling methods and detection of variation

To increase the chances of detecting unintended effects, profiling methods have been proposed as tools for characterising changes in the composition of GM plants [43, 96]. In theory, an unbiased, untargeted comparison using molecular profiling techniques offers almost unlimited possibilities for monitoring differences in cells and tissues [94]. These technologies include transcriptomics (profiling of mRNA levels), proteomics (proteins) and metabolomics (metabolites).

Data sets produced by these profiling methods could be used, either alone or in combination, in comparative studies to assess the composition and quality of food as affected by the environment, genetic background, naturally occurring or induced mutations, and genetic modification. Detected differences in the levels of transcripts, proteins or metabolites may be an indication of unintended side effects of the genetic modification and provide information for further investigations of possible toxicological relevance [27]. The new technologies also have potential for various other food safety applications. An important benefit is also the generation of comprehensive data that may contribute to basic research and a better understanding of crop biochemistry and food composition.

However, profiling methods have been criticised for not being fully developed and validated and having certain limitations [43, 94]. The genome of *Arabidopsis*, for instance, has been estimated to contain ca. 25 000 genes [9], with the potential to produce a corresponding number of mRNAs and proteins. The protein complement of the genome is further complicated by posttranslational modifications. It has been suggested that the number of metabolites is an order of magnitude lower; on the other hand, it has been estimated that up to 200 000 different metabolites occur across the plant world [49]. Profiling methods employ extraction procedures and separation and detection methods that are not selective and allow a wide range of compounds to be isolated; yet it is very difficult to monitor all potentially interesting molecules, and the methods are often a compromise between the range of molecules to be detected and the potential for technical variation [62]. For the majority of profiling applications, the absolute value is, however, unimportant, and the relative value is sufficient [50].

Transcriptomics provides, at least for specific plant species, the most complete coverage of gene products and thus also potential unintended effects; however, whole genome arrays are currently not available even for many important food crops [94, 118]. A comprehensive coverage is harder to obtain in the case of proteomes and metabolomes, which are complex mixtures of molecules with widely different chemical properties. However, proteins and metabolites as potential al-

lergens, toxicants and antinutrients are the actual functional molecules of interest [27]. The benefits and problems of each of these methods are discussed in the following chapters.

2.3.1 Protein profiling

Proteins have a strong influence on the phenotype and safety of a crop plant, either directly through protein content or function or indirectly through the relationship of a protein with metabolites [42]. The term proteomics is more widely used and can be considered to encompass the analysis of various aspects of protein expression, structure and function. Here it mostly refers to quantitative profiling of a large number of proteins in a given sample. It is possibly the most challenging of the 'omics' or profiling methods, as proteins vary enormously in their structure and chemical properties, which makes the monitoring of all proteins in a given plant tissue complicated. The number of chemically distinct proteins far exceeds the number of genes due to, for instance, alternative splicing of transcripts prior to translation [134]. Posttranslational modifications, such as phosphorylation, acylation, or proteolytic processing, can alter protein activity, location and stability; also, protein concentrations extend over a far greater dynamic range than do those of nucleic acids [134]. Yet changes in phenotype, e.g. proteins and other functional molecules, and not changes in genes, are of interest to food safety and nutrition. Most allergens are proteins [99] and many food allergens are present as major components (1–18% of total protein) [80]. The case of the α-amylase protein that showed altered glycosylation when expressed in transgenic pea [139] highlights the importance of the characterisation of posttranslational modifications of proteins. Expression of genes in different organisms can potentially result in differences in folding or posttranslational modification of proteins, and these should to be taken into account in the assessment [80].

The workflow of proteomic analysis includes several common steps regardless of the analytical methods employed (Figure 1). The key elements include a proper experimental design taking into account the biological system, extraction of as many proteins as possible, protein or peptide separation by gel-based methods or liquid chromatography (LC), quantification of spots (image analysis), peaks etc., statistical analysis of data, identification of proteins/peptides of interest by comparison of mass spectra against sequence databases and integration of the data with other data sets [145]. A standard representation of proteomic experiments similar to the MIAME guidelines for transcriptomics [21] has been proposed [172].

For profiling purposes, a quantitative method for monitoring protein levels is an absolute requirement, combined with a high throughput. The most widely used technique for protein separation to date, two-dimensional electrophoresis (2-

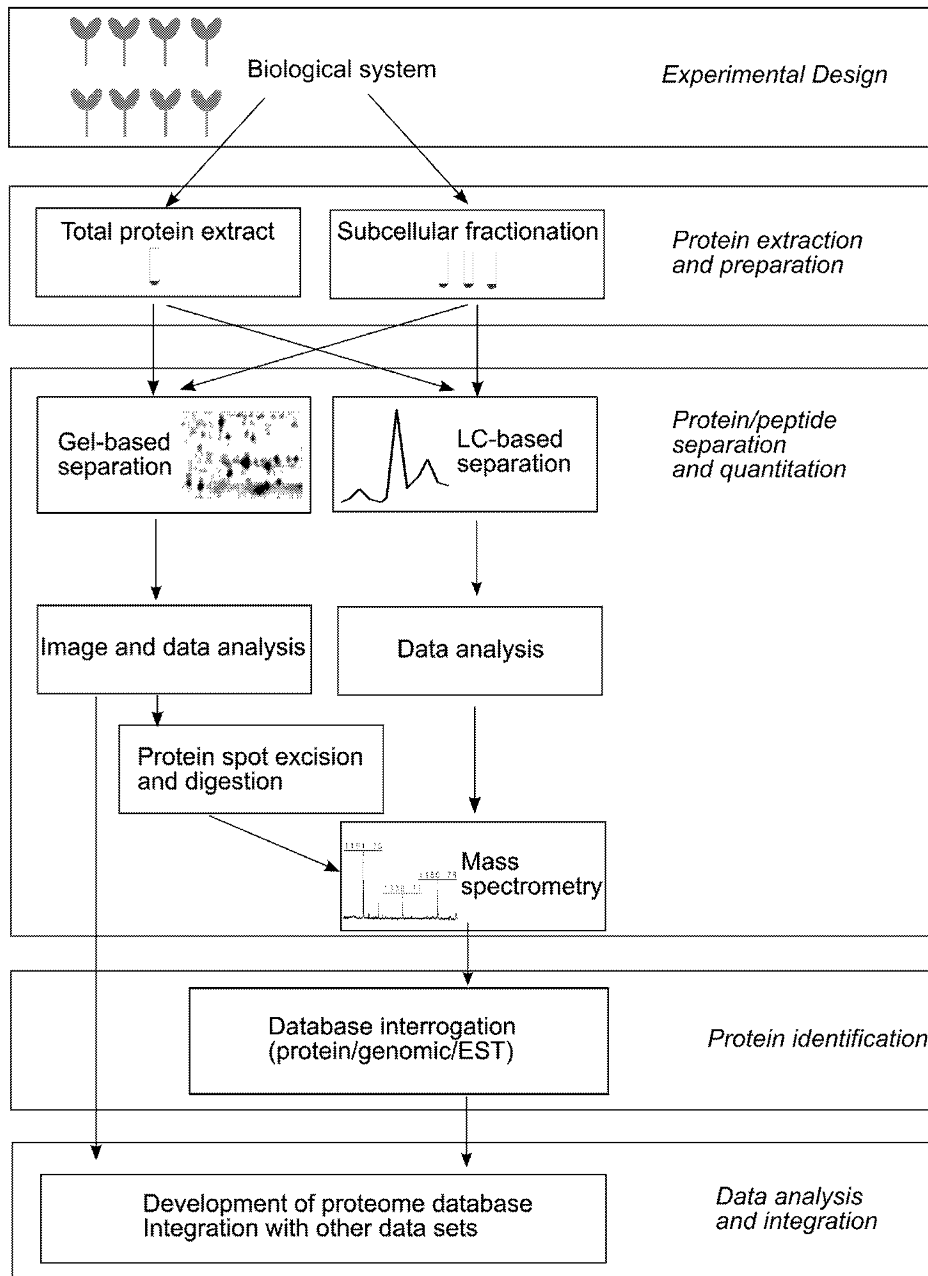

Figure 1: Overview of a typical workflow of proteomic analysis. Modified from Rose *et al*. [145].

DE) [129], is still popular, powerful, mature, and relatively sensitive [141]. Using 2-DE and mass spectrometry in combination, it is often possible to visualize, quantify and identify hundreds or even thousands of proteins in a given tissue or cell sample, and proteome analysis is increasingly used in functional plant studies [25]. Current fluorescent-based stains offer a broader dynamic range and good sensitivity for protein detection in 2-DE and, due to different electrophoretic mobilities and detection methods, various important posttranslational modifications can also be visualised, such as phosphorylation, glycosylation and peptide signal cleavages [133]. Newly developed techniques such as difference gel electrophoresis (DIGE) [181] allow for a direct comparison of two or more samples in the same 2-DE gel and the inclusion of an internal standard in a 2-DE run [4].

There are, however, still some limitations. The large range of protein expression levels limits the ability of the 2-DE approach to analyse proteins of low abundance [67, 141]. Current proteomic studies have revealed that the majority of identified proteins are abundant housekeeping proteins that are present at high levels per cell, whereas proteins such as receptor molecules present at much lower concentrations are usually not detected. Prefractionation of the protein sample may help in this, as may concentrating on subproteomes or narrow-range pH gradients in 2-DE [70]. The outcome of a proteomic analysis is strongly dependent on the extraction method [62]. Compounds such as phenols and polysaccharides interfere with protein extraction and separation [70], making plant tissues that commonly contain relatively low amounts of protein but a significant amount of various interfering substances a difficult starting material. In the steps involved in 2-DE and also probably in protein extraction, there is extensive and variable loss of proteins, which means that spot intensities in 2-DE do not always represent the absolute amounts of proteins present in original sample [194]. Usually the most abundant proteins dominate the separation, and some types of proteins are underrepresented, such as large or highly hydrophobic proteins and membrane-associated proteins in 2-DE [135, 153]. 2-DE is somewhat expensive and labour intensive, and is difficult to automate in a high throughput setting [113].

Recent years have brought about improved mass spectrometric technology and instrumentation [2] that holds much promise for proteomic analyses. A range of mass spectrometry and non-gel-based methods for protein separation and identification have recently been described, such as multidimensional protein identification technology (MudPIT). It couples multidimensional chromatography and tandem mass spectrometry of peptide mixtures for an automated protein identification [110, 191]. While MudPIT has less inherent biases than 2-DE and can therefore identify more membrane proteins and proteins of low abundance, in particular when combined with highly sensitive and accurate mass spectrometry systems such as FTICR-MS (Fourier transform-ion cyclotron resonance mass spectrometry) [195], it is more of an exploring technique and not readily quantitative

as such. The more common and versatile use of large-scale MS-based proteomics has been to document the expression of proteins present in a cell or tissue [2].

Quantitative applications employing 'heavy' and 'light' labelling have also been developed. A heavy isotope may be added to the growth medium of an organism [190]; however, this is only applicable in situations where metabolic labelling is feasible. Various tags such as isotope coded affinity tags (ICAT) [68] have been developed, and commercial reagent kits for several modifications of the ICAT strategy with varying chemistries such as the iTRAQ [146] reagents are now available. Regardless of the improvements, targeting low-abundance and membrane proteins remains a challenge [67], and the performance of the LC-MS strategy for large sample sets is not clear at the moment, although it shows promise for large-scale quantitative analysis [58]. A highly comprehensive proteome exploration of rice using both 2-DE and MudPIT [90] was able to detect a total of over 2500 proteins in different parts of rice. While MudPIT provided a larger coverage of proteins than 2-DE (2363 and 556 proteins, respectively), there were 165 proteins that were only detected in 2-DE and 1972 only in MudPIT. This suggests that neither of the methods is fully comprehensive but rather complementary [134].

Overall, MS analysis and identification of proteins is based on the assumption that an adequate database of amino acid and nucleotide sequences exists for the organism being analyzed or that sufficient homology exists with more fully characterized species [2]. This is a major limitation in the analysis of many plant proteins at the present time, and proteomics is indeed most effective with fully sequenced genomes [134].

Other approaches have also surfaced in the recent years but have not yet gained popularity compared to existing methods. Analytical protein arrays for the purpose of protein profiling typically comprise a library of peptides or antibodies arrayed on a surface. By far, the greatest obstacle in developing functional protein microarrays is the construction of a comprehensive expression clone library from which a large number of distinct protein samples can be produced [15]. Protein array approaches are being developed, particularly for functional studies, but will not be available for many plant species for some time. Other approaches include the SELDI (surface-enhanced laser desorption/ionization) platform that utilizes the binding of whole proteins on affinity arrays [78].

Corpillo *et al.* [34] have described the use of proteomics as a tool to improve investigation of substantial equivalence in GM organisms. They used 2-DE to determine whether there are significant differences between parent and GM tomato seedlings modified for resistance to tomato spotted wilt virus. However, only one GM line and 40 major proteins were included in the study. Recently, Ruebelt *et al.* [149] compared the seed proteomes of transgenic *Arabidopsis* lines to nontransgenic lines. Differences in spot quantity were found to be in the range

observed for *Arabidopsis* accessions [148] or were related to the introduced gene.

2.3.2 Other profiling methods

Analysis of differential gene expression has been proposed as a method to determine the substantial equivalence of GM organisms [94, 187]. Several methods for large-scale gene expression analysis have been developed, including differential display [109], serial analysis of gene expression (SAGE) [188], massively parallel signature sequencing (MPSS) [22] and microarrays [156], of which microarrays appear the most promising for GM applications. Microarrays may be constructed of short oligonucleotides or complete cDNA clones and provide a rapid way to monitor in parallel the expression of thousands of transcripts; even whole-genome arrays are available for some model plants [118]. Apparent quantitative differences in mRNA levels may be confirmed by analysis of individual transcripts by quantitative real-time PCR or Northern analysis. cDNA or oligonucleotide microarrays are, however, not completely unbiased, as they require a priori sequence information for each gene that is to be monitored, and the most crucial part in the detection of altered gene expression using microarray technology is the construction of the array [187]. Microarrays, in general, do not account for differences resulting from naturally occurring differences in a gene sequence between organisms. Overall, published cross-platform analyses suggest that the conclusions derived from a microarray analysis may be largely dependent upon the type of platform used in the experiment [118] - a problem also inherent to other current profiling methods [55]. Standardization of experimental design and methods would facilitate comparison of array data produced in different experiments or laboratories. A standard set of technical details, MIAME [21] requires the reporting of enough details to ensure that the results of a microarray experiment could be interpreted or repeated. A particular issue regarding the use of microarrays for detecting variation due to different genetic backgrounds is that mRNA expression tends to change and fluctuate considerably [27]. Analysis of mRNA levels may thus not always measure expression relevant to the actual end product, as the levels of transcripts and proteins do not always correlate very well [69, 175].

The third ‘omics’ technology, metabolomics, forms a direct link between gene expression and the function of the metabolic network, reflecting the phenotype most closely in many cases [49]. It can also be applied without genome information [55, 72]. However, no single technique meets all the requirements for an ideal metabolomics method. Developments involving gas chromatography (GC) have been responsible for the recent upsurge of interest in plant metabolomics, as GC-MS is highly sensitive and can detect a wide variety of organic compounds [49, 72]. Liquid chromatography is often used to detect compounds not well cov-

ered by other methods, as it suits for involatile as well as volatile compounds. Methods such as nuclear magnetic resonance (NMR), Fourier transform infrared spectroscopy (FT-IR) or FTICR-MS, when applied to crude extracts without a separation step, are considered mainly as fingerprinting methods for rapid screening and sample classification [47]. Several analytical techniques may be needed and are often used in combination for a full description of the metabolome [144], and many of them are fast enough for high-throughput measurements. The most difficult problem with metabolomic approaches is probably the determination of the exact chemical structure of the individual compounds seen, as the majority of compounds observed remain unknown [144]. Unknown peaks can however be used as target analytes without prior knowledge of their exact chemical structure or identified by comparing mass spectra to available libraries [50]. Extraction procedure is a key step, as it should lead to efficient extraction of a wide range of compounds across the different chemical classes but in practise is always a compromise [49]. Standards for reporting metabolomics analyses are also emerging [79].

The application of several techniques of metabolic profiling for the characterization of GM plant lines has been demonstrated in literature. Roessner *et al.* [144] have evaluated the applicability of metabolic profiling (GC-MS) for the characterisation of transgenic potato lines modified in sucrose metabolism. The lines showed various metabolic perturbations, supported by evidence from earlier characterisation by classical biochemical approaches, and were classified by differences in metabolic profiles that could be assigned by the genetic modifications. Le Gall *et al.* [107] analysed the ^{1}H NMR spectra of transgenic tomato with altered flavonoid metabolism and concluded that the levels of 15 other metabolites were significantly different in the transgenic fruits compared to the controls; however, the changes in mean values were relatively minor and within the natural variation that would be observed in a field-grown crop. Defernez *et al.* [39] studied potatoes of various GM lines modified in primary carbon metabolism, starch synthesis, glycoprotein processing, or polyamine/ethylene metabolism using NMR and HPLC-UV profiling. The most obvious differences were seen between the two varieties used. There were significant differences in the amounts of several compounds in GM lines with altered polyamine metabolism; however these were lines that gave rise to a very abnormal phenotype. Differences in the mean values amounted to no more than 2–3-fold. Kristensen *et al.* [93] used both metabolic and transcript profiling to characterize *Arabidopsis* plants and demonstrated that the effects of transformation with partial biosynthetic pathways of dhurrin led to changes at the metabolite and transcript levels, while the transformation of the whole biosynthetic pathway did not induce major differences. Gregersen *et al.* [66] have described the use of microarrays for the comparison of gene expression profiles of wheat transformed for phytase expression against wild type. They

found no significant effects apart from slight differences that were attributed to minor differences in the developmental status of the seeds. Baudo *et al.* [13] analysed the transcriptome and Baker *et al.* [11] the metabolome of transgenic wheat expressing additional subunits of glutenin.

2.3.3 Considerations for the interpretation of profiling data

Although profiling methods are not fully comprehensive, it is clear that they all are capable of producing vast amounts of data, and the analysis and interpretation of omics data requires advanced bioinformatics [42]. Traditionally data analysis has been lagging somewhat behind the analytical methods, at least in proteomics. Interpretation of mRNA, protein or metabolite levels is difficult because biochemical pathways are linked and highly regulated. Thus it is to be expected that correlations of individual compounds and gene functions will be more clearly distinguished by multivariate data mining techniques [50, 198]. These advanced statistical methods, including principal component analysis (PCA), partial least squares regression, clustering techniques etc. take into account the fact that hundreds or even thousands of compounds may be measured, often with a limited number of samples, and provide visualization tools for highly complex multidimensional data [50, 65, 72].

The ability to assign plant samples to groups on the basis of their profiles using techniques such as PCA offers many possibilities for functional plant profiling. While such groups are likely to be dominated by differences between genotypes, the use of PCA allows for the separation of profiles according to several factors to be distinguished [50]. PCA can furthermore be used to analyze which compounds exert the largest influence on the components and may also highlight individual compounds that are not significant by classical statistical tests such as ANOVA [50, 65].

The various profiling methods will obviously work best if information obtained through different technologies can be combined. Based on comparisons of metabolic and transcript profiling of potato tubers it has been suggested that comparisons of transgenic and conventional crops should be performed at more than one level [182]. These systems biology approaches aim at a holistic understanding of interactions across several molecules and phenotypes by integrating data from various levels [42, 62].

Despite the great promise of these techniques, there are significant technical and biological issues associated with achieving validated quantitive data sets from profiling experiments. Great care has to be taken in experimental design, and the variability of the chemical and technical methods needs to be characterised [47]. Furthermore, oscillations in plant homeostasis related to e.g. local environment, development, lighting conditions and diurnal rhythms are common and need to be

considered in sampling [72].

In applying these techniques to food crops, it is essential that multiple control samples are studied in order to assess inherent environmental variability, so that the experimental samples may be evaluated in the context of the conventional crop as a whole [27]. A successful application of profiling techniques to the safety evaluation of GM foods will therefore require databases that contain information on variations in profiles associated with differences in developmental stages and environmental conditions [94]. Furthermore, the safety implications of the majority of proteins and metabolites are still largely unknown, and most are likely not to have significance for food safety. While limits for maximum content of some known plant toxins and antinutrients such as glucosinolates in oilseed rape have been set by the European Union [41] and consensus documents for recommended levels of certain constituents have been compiled [127, 128], no internationally agreed ranges for acceptable concentrations and variations are given for many such compounds [125].

It has been argued that substantial equivalence is difficult to define, especially in studies involving the profiling of hundreds or thousands of compounds [26, 34]. When statistically significant differences are found, their biological significance should be assessed. It is easier to find at least some differences than no differences at all; however, the differences may well be within natural variation [27]. Catchpole *et al.* [26] have therefore suggested that if samples or GM lines cannot be separated in PCA (except for the molecules that have been modified by the transgene), then they may be considered substantially equivalent.

2.3.4 Natural variability of protein profiles in plants

Currently little comparative profiling data are available for individual crop plants, although proteome analysis (and other profiling methods) is increasingly used in functional plant studies [25]. Plant proteomics has been clearly shown to be capable of characterizing genetic variations, including mutant lines, genetic distances and phylogenetic distances [176]. These approaches have demonstrated the capacity of 2-DE to distinguish genotypes of model plants such as the species in the Brassicaceae family [115] and accessions of *Arabidopsis* [148] and *Thlaspi caerulescens* [178]. The extent of variation in cultivars of important crop species varies, for instance for the protein profiles of grape berry mesocarp between six cultivars [154], wheat populations [38] and maize inbred lines [24, 33]. It has been suggested that at the proteome level differences may be more pronounced than at the genome level [176]. Even multiple effects of single mutations on protein expression have been described, e.g. the *o2* mutation affecting zein proteins in maize endosperm [33, 37] and *Arabidopsis* [151, 152] and tomato [76] mutations with alterations in morphology and physiology.

An issue to be considered in proteomic studies is that the proteome of a plant is not constant. Instead, protein expression of a plant clearly changes during plant growth and development and between tissues and cell types [3]. Variation in different developmental or physiological stages or organs has been well demonstrated at the proteome level. Storage tissues such as seeds deposit large amounts of storage protein and undergo many physiological changes during their development and germination that have been observed by proteomic studies in model plants such as *Arabidopsis* [56] and *Medicago* [57] and in important crop species such as oilseed rape [71] and barley [51]. However, for many commercially and nutritionally important plant species and developmental processes comparative proteomic and other profiling data are still lacking.

2.4 Conclusions

Crop breeding by both conventional means and by genetic modification have at least a theoretical potential to modify the plant composition beyond the particular trait intended, thus leading to unintended effects. Analytical detection of unintended effects is a huge undertaking with many technical challenges. A further challenge is to determine the significance of any unintended effect on consumer health, as unintended effects do not automatically imply a health hazard. Risks may be considered if the nutritional profile of the plant has been altered, if proteins have been altered in a way that affects their allergenic potential, or if new or increased levels of potentially toxic secondary metabolites are produced. However, unintended effects may also have absolutely no impact on health, or may even be beneficial.

In risk assessment, the aim is to evaluate the safety of a crop by comparing it to a conventional counterpart, where available, that has a safe history of use. Therefore, knowledge of natural variation in conventional crops and a thorough insight into the biology of the crop is essential for the interpretation of a possible unintended effect.

3 Aims of the present study

The aims of the thesis were to:

1. set up an experimental method for monitoring protein expression of potato tuber material and assess its reproducibility
2. study several genetically modified lines of potato for possible unintended effects (**I**)
3. provide an insight into the natural variation of protein expression due to tuber-to-tuber variation, genetic background (**I**), physiological state (**II**) or alternative agricultural practices (**III**)

4 Materials and methods

4.1 Plant material from GM and non-GM potato genotypes (I)

Non-GM potato (*Solanum tuberosum* L.) material included 32 genotypes, more precisely 21 named cultivars of tetraploid potato, eight landraces and three diploid genotypes including accessions and named cultivars of *S. phureja* adapted for long daylength conditions.

Genetically modified lines selected for the analyses have been developed at the Scottish Crop Research Institute over several years [97, 171, 173]. These included wild-type tubers of cv. Desirée, tubers generated from non-modified plants produced via tissue culture, transgenic tubers transformed with an 'empty vector' construct (W2GBSS VO4 and VO9; Mal1 2V1) and tubers transformed to express various genes (Mal1 2A, 5A and T41A; Sam35S 1 and 3). *Agrobacterium*-mediated transformation was used to generate the GM lines. It is noted that the GM lines used are experimental ones that were selected to aid development of the profiling methodologies and are not intended for commercial purposes.

Tubers of each independent line were grown in a randomized field plot according to standard agricultural practices at the Scottish Crop Research Institute. For each replicate, a single average-sized tuber (usually between 80 to 100 g fresh weight per tuber, depending on the line and construct) was selected for analysis, resulting in four individual tuber samples for each line. Tubers were stored at 15 °C for one week after harvest and at 5 °C prior to analysis.

4.2 Plant material for the tuber life cycle analysis (II)

To provide material at a range of physiological stages (**II**, Table 1), 50 plants (*S. tuberosum* cv. Desirée) were grown from tubers under containment. Plants were harvested sequentially over ca. 5 months to obtain stolons/tubers at predefined developmental stages (**II**, Table 1). Material was frozen in liquid nitrogen without cutting and the frozen tissues were freeze-dried and milled. For stages 4 to 8, three plants were harvested at each stage. Two opposite eighths were removed from each tuber within a replicate, bulked together, frozen in liquid nitrogen, freeze dried and milled.

At stage 4 tubers were removed from three plants and then reburied within the compost. These detached tubers were harvested after three days, in parallel with tubers from three intact plants at the same developmental stage. Mature tubers harvested from senesced plants (stage 6) were stored at either 5 °C or 10 °C in the dark for four weeks.

Tubers harvested at stage 6 were stored at 5 °C in the dark until sprouts were just visible. Tubers were then transferred to 10°C in the dark for two weeks to

stimulate further sprout development, and tubers were sampled when sprouts were ca. 1 cm long (stage 7). Remaining sprouting tubers were planted and allowed to form shoots ca. 20 cm in length in the dark before tubers were sampled (stage 8).

4.3 Plant material and field experimental design for the analysis of alternative cultivation practices (III)

Tubers (cv. Santé) samples for the comparison of conventional and organic management practices were produced in the Nafferton factorial systems study, a group of long-term, replicated factorial field experiments designed to identify the effect of (a) fertility management methods, (b) crop protection practices and (c) rotational designs (i.e. pre-crops) used in organic, low input and conventional production systems in the UK. The field experiment was performed in a 3-factorial randomised split-plot design with pre-crop (winter wheat or grass/clover), crop protection (organic or conventional) and fertilisation (organic and conventional) as factors (**III**, Table 1).

The crop protection and fertilization protocols used were the same as used by the commercial conventional and organic farming businesses. All fertilisation treatments (**III**, Supplementary Table 2) were applied 4 weeks prior to planting of tubers and no irrigation was used. Crop protection and defoliation treatments were applied as described in **III**, Supplementary Table 1. After defoliation, tubers were left in the ground for 4 weeks to allow skin maturation.

Tubers were sampled after 8 weeks of storage at ambient temperature (8-12 °C). About 800 g fresh weight from 4-5 tubers were combined. Two opposite eighths (to minimise gradient effects within the tuber) were removed from each tuber within a replicate and bulked together. The resulting ca. 200 g of fresh weight were chopped, frozen in liquid nitrogen, freeze-dried and milled, resulting in a total of 64 samples.

4.4 Extraction of total soluble protein (I, II, III)

Total soluble protein was extracted from ca. 1 g of freeze-dried powder or from ca. 3 g of powder from fresh tuber material homogenized under liquid nitrogen as previously described by Koistinen *et al.* [89]. The protein pellet was dissolved in 2-DE sample buffer containing 9.5 M urea, 2% CHAPS, 1% DTT, 0.8% Bio-Lyte 3/10 ampholyte (Bio-Rad, Hercules, CA, USA). Protein was quantified using the Bio-Rad Protein Assay Dye reagent. Ca. 300 μg or 150 μg of total protein, depending on the experiment, was loaded on each 2-DE gel.

4.5 Two-dimensional electrophoresis (I, II, III)

The first dimension isoelectric focusing was performed using 24 cm IPG strips (Amersham Biosciences, Uppsala, Sweden) with linear pH range 4–7 in Ettan IPGPhor isoelectric focusing system according to the manufacturer's instructions. The second dimension was run in the Hoefer DALT system (Amersham Biosciences) in 19×23 cm homogeneous 12% SDS-PAGE gels according to the manufacturer's instructions.

The gels were stained with the SYPRO Ruby fluorescent stain (Bio-Rad). Gel images were acquired with the FLA-3000 fluorescent image analyzer (Fuji Photo Film Co. Ltd., Tokyo, Japan) using excitation wavelength filter of 470 nm and emission wavelength filter of 580 nm.

4.6 Image and data analysis (I, II, III)

Gel image analysis was performed with PDQuest software v7.1 (Bio-Rad). Protein spot intensities were normalised to the total intensity of valid spots to minimize possible errors due to differences in the amount of protein and staining intensity, and an extensive manual validation of spot detection and matching was performed.

The majority of protein spots with low normalized quantity values (depending on the experiment) and quality values of 0 as given by PDQuest were considered background noise and not indicative of real protein quantity. The spot quantities were transformed using a square-root or log(z+1) transformation to normalise the data, and all subsequent statistical analyses were performed using the transformed data.

The statistical methods used were ANOVA and the non-parametric Kruskal-Wallis test to identify individual protein spots with significantly different expression levels, and principal component analysis (PCA) to explore whether one or more lines, varieties or other experimental treatments would separate from others and to identify groups of proteins which, in combination, had different expression levels among the gels. Statistical analyses were performed using the Genstat software package (VSN International, Hemel Hempstead, UK).

4.7 Protein identification by HPLC-electrospray tandem mass spectrometry (I, II, III)

In-gel digestion for protein identification was performed according to Koistinen *et al.* [89] and the tryptic peptides were analysed by two different mass spectrometry systems. Tryptic peptides were separated using Ultimate/Famos capillary LC system (LC Packings, Amsterdam, the Netherlands). The LC was connected to

a mass spectrometer with a Protana platform (Protana, Odense, Denmark). Mass spectra were recorded with a LCQ quadrupole ion trap mass spectrometer (Thermoquest, San Jose, CA, USA). The peptides were identified with Xcalibur software (Thermoquest) and the Sequest algorithm. Alternatively, the LC was connected to mass spectrometer with a nanoES ion source (Protana). The positive TOF mass spectra were recorded on a QSTAR XL hybrid quadrupole TOF instrument (Applied Biosystems, Foster City, CA, USA). The peptides were identified using ProID software (Applied Biosystems) and the Mascot interface [136].

Peptides were identified against the NCBI nonredundant protein database (Viridiplantae), the TIGR potato EST database (`http://www.tigr.org/`) and a potato EST library kindly supplied by Dr. J.P. van Dijk, RIKILT Institute of Food Safety, The Netherlands. Matches of MS/MS spectra against sequences in the databases were also verified manually.

5 Results

5.1 Validation of the proteomic procedure

5.1.1 Effect of freeze-drying of potato tuber material on 2-DE

Freeze-drying is a convenient method that allows for easier handling, storage and shipping of potato tuber material, e.g. from laboratory to laboratory or country to country (reduces transportation costs, minimises potential metabolic changes and allows long term storage of samples). To find out whether this process had any effect on protein extraction and 2-DE data, two parallel transverse slices of eight tubers were prepared: one cross-section was freeze-dried and extracted, while the other was frozen under liquid nitrogen and extracted. These subsamples were then analysed by 2-DE in the same gel batch. Both freeze-dried and fresh material produced 2-DE gels with good protein separation. Fewer spots were detected in gels run from freeze-dried material (1397±200 [SD]), compared with fresh material (1606±132, P=0.009). Spots that were matched to at least 14 of the 16 gels were analysed further, i.e. altogether 838 spots. There was a significant difference between fresh and freeze-dried samples in 140 spots (P<0.01) and within this group of 140, the P value was <0.001 for 44 spots that were randomly scattered across molecular weights and isoelectric points. Furthermore, principal component analysis (PCA) was performed for the 838 spots. The freeze-dried and fresh subsamples separated in the first principal component (Figure 2). None of the other PCs showed any differences, indicating that the overall spot intensity was the major factor differentiating freeze-dried from fresh material. The average CVs for these spots were similar for the freeze-dried and fresh samples, 31% and 27%, respectively. Thus, although less spots were detected in gels run from freeze-dried material, the quality of this material was concluded to be satisfactory.

5.1.2 Sources of variation in 2-DE of tuber proteins

The variability of the 2-DE system as a whole was assessed in an experiment that included protein extraction and biological variation between individual tubers grown in the same conditions (3 tubers × 3 extractions per tuber × 2 gels per extract). A total of 862 spots were deemed valid. Some reports have determined analytical variance by selecting a dataset of, for example, 50 representative protein spots of various molecular weights and isoelectric points [10, 82]. Therefore, in our study, a set of 50 representative spots of different levels of expression and covering a wide range of pI and M_r values, present in all analysed gels, were selected at random. The average coefficients of variance (CVs) of these 50 spots was determined to be 19.5% across the 18 gels. This suggests that the variation observed by us is similar to the values reported by others for analytical variation due

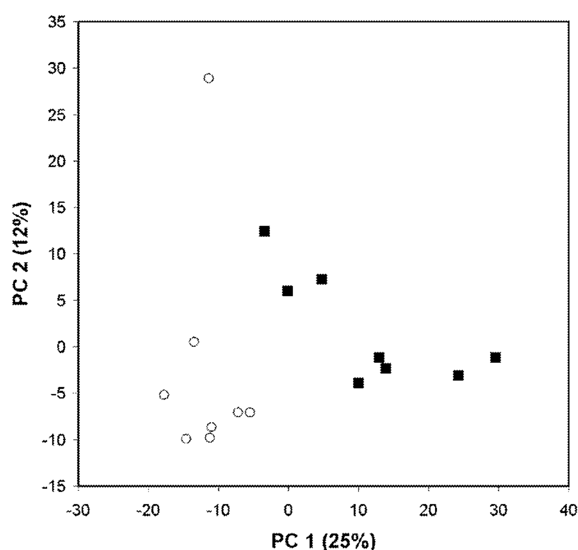

Figure 2: Principal component scores for 2-DE from fresh (○) and freeze-dried (■) potato tuber material. Values in brackets indicate the percentage of total variation accounted for by each principal component.

to the 2-DE step alone [10, 28, 82, 124, 147]. Furthermore, a CV was calculated for each of the 862 spots across all 18 gels. While the CVs varied considerably between different spots, the average of all CVs was 39.0% for all spots and 34.8% for the 813 spots matched across at least 14 gels. The CVs correlated negatively with spot quantity, as spots with low quantity tended to have a higher CV. Spots that were detected in 13 or fewer gels tended to have very high coefficients of variance.

The variance components of each contributing factor, i.e. tuber to tuber variation and variation due to extraction and 2-DE, were determined by ANOVA. Distributions of variance for each spot indicated that variances due to 2-DE were clearly higher in general (Figure 3). Variance associated with extraction was also greater than tuber to tuber variance. Variance associated with extraction and tuber to tuber differences were more important than that associated with 2-DE in only 62 and 17 of the 813 spots, respectively. In many cases, the variance due to protein extraction and tuber was 0; this is probably because the variation due to these factors was so small compared to that due to 2-DE that it was not possible to estimate it exactly. PCA was used to further explore patterns in variation. The first six PCs accounted for 11.1%, 9.5%, 7.4%, 6.8%, 6.3% and 5.8% of total variation. The fact that none of the PCs accounted for a large part of the total variation suggests that there was no overriding effect of a single source of variation in the

experimental setup. Major differences highlighted by PCA were the significant difference between the tubers in PC 2 (P=0.035) and between the gel runs in PC 4 (P=0.018). Although the differences were significant according to ANOVA, there was considerable overlap of the groups (Figure 4).

5.2 Protein profiles of genetically modified potato lines (I)

A range of genetically modified potato lines was selected for proteomic analysis along with their appropriate controls. Qualitatively the protein patterns on the 2-DE gels were similar across the lines. ANOVA revealed significant differences (P<0.01) in the quantities of seven protein spots out of 240 analysed (**I**, Table III). Analysis with the Kruskal-Wallis test also revealed that spot intensity was significantly different among the lines in two cases out of 490. Seven of these proteins were identified by HPLC electrospray tandem mass spectrometry (**I**, Table I, Figure 3). These proteins included some defence-related proteins as well as some proteins involved in protein destination and storage and some proteins with unknown function. One of these spots was matched to a patatin protein; however, other spots that were previously identified as patatins did not appear to differ between the lines. Also other spots previously matched to Kunitz-type enzyme inhibitors and aspartic proteinases did not show significant differences among the lines. Line-specific proteins, i.e. proteins expressed in all replicates of only one or few GM lines, were not found.

Principal component analysis (PCA) was assessed for its capacity to differentiate between genotypes based on protein quantities, and to identify groups of proteins responsible for differentiating the genotypes. PCA was carried out on the 240 protein spots (**I**, Table III). No pronounced separation between the lines was observed in the plots of PCA scores (data not shown). PCA analysis was also carried out on protein sets where significant difference among potato lines was found by ANOVA. Again, there was no clear separation between the lines in the PCA plots (**I**, Figure 4). The lines Mal1 2A and Mal1 5A were slightly separated from all the other lines in the first and second components, suggesting that these particular lines were primarily responsible for the seven protein spots identified as significantly different by ANOVA (**I**, Table III).

5.3 Protein profiles of non-GM potato genotypes (I)

A total of 32 non-GM potato genotypes were selected to represent a range of genetic variation: 21 named cultivars of tetraploid potato, eight landraces and three diploid genotypes including accessions and named cultivars of *S. phureja* adapted for long daylength conditions. Between 800-1200 polypeptides were detected depending on genotype. A total of 1932 polypeptides were detected, when all spots

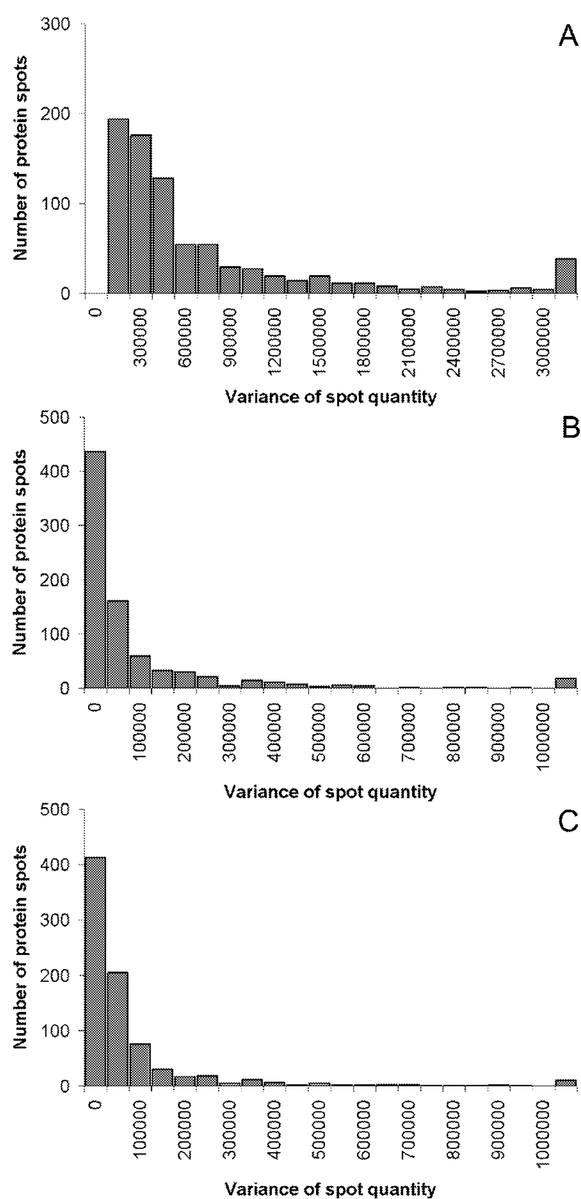

Figure 3: Distributions of variance of spot quantity for each source of variation: A, two-dimensional electrophoresis; B, protein extraction; and C, tuber to tuber. Analysis of variance was used to calculate the three variance components for the quantity of each of the 813 spots detected in at least 14 of the 18 gels.

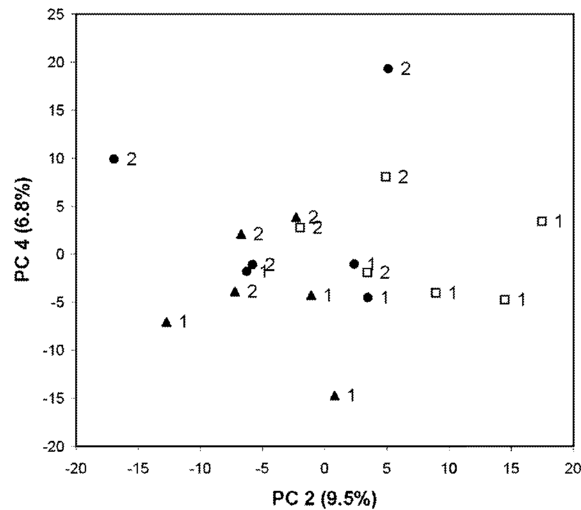

Figure 4: Principal component scores for 2-DE samples derived from three individual tubers (▲, □ or •) and run in gel batch coded 1 or 2. The percentage of the total variation accounted for by each principal component is shown in brackets.

in all genotypes were combined. There were clear qualitative and quantitative differences in polypeptide profiles between the genotypes (**I**, Figure 2). One of the most obvious differences occurred in proteins with the molecular weight of ca. 40 000 to 45 000 and pI of 4.5 to 5.5, which corresponds to various isoforms of patatin, the major storage protein of potato tuber [137, 157]; many genotypes showed distinctive patterns of putative patatin isoforms.

Analysis of individual proteins revealed that for 1077 spots out of 1111, the expression was significantly different among the genotypes. Interestingly, only 34 of the protein spots detected did not appear to differ significantly between genotypes. Among these proteins that did not differ between genotypes were proteins matched to two triosephosphate isomerases, a putative malate dehydrogenase, GRP1 and annexin p34 (**I**, Table I). Proteins that were expressed in less than 27 of the 127 samples were not analysed statistically, because these proteins were rarely detected in all four replicates of the different potato genotypes. Genotype-specific proteins, i.e. spots expressed in only one or a few genotypes, were not examined any further; there were ca. 600 of these proteins. Qualitative differences observed in 2-DE are likely to be due to allelic variations and possibly posttranslational modifications.

PCA was carried out on the 393 protein spots also analysed by ANOVA (**I**, Table II) and was found to differentiate several genotypes from each other (**I**, Figure 3). For example, in the second and third components the line TBR3302 (2)

and the three *S. phureja* genotypes (accession PHU.4637 and named cultivars Inca Sun and Mayan Gold) separated from the tetraploid *S. tuberosum* named varieties and landraces. In the fourth component, varieties Glenna, Morag, Maris Piper and Pentland Javelin were slightly separated from the other varieties (not shown), but the differences between named varieties were not as obvious. Nor were most landraces separated very clearly from other genotypes.

Several of the proteins with very large positive or large negative loadings in the PCA components two, three and four were proteins which also differed among genotypes. These proteins were targeted for identification (**I**, Table I, Figure 1). Among those contributing to the second component were several which were related to disease and defence responses. Most of the proteins with high loadings in the third component appeared to be involved in the glycolytic pathway, such as several enolases, triosephosphate isomerase, and fructose-bisphosphate aldolase. This component also included two polypeptides with homologies to ascorbate peroxidases and UTP-glucose-1-phosphate uridylyltransferases. Protein spots with low or high loadings in the fourth component included some proteinases, a Kunitz-type enzyme inhibitor and some proteins or polypeptides matching to EST sequences of currently unknown function.

A total of 77 proteins were tentatively identified by HPLC electrospray tandem mass spectrometry using cv. Desirée as the model genotype and those were included in a 2-DE reference map (**I**, Table I, Figure 1).

5.4 Comparison of genetically modified and non-modified potato lines and varieties (I)

As the results indicated more differences between non-GM genotypes than between GM lines and their controls, some varieties and lines were selected for a second set of experiments. The GM line Sam35S 3, vector-only control W2GBSS VO4, two wild-type control Desirée lines, the variety Maris Piper and an accession of *S. phureja* (PHU.4637) were analysed together, with three tubers of each genotype.

In PCA (**I**, Fig. 5), the genotypes Maris Piper and *S. phureja* were clearly separated from each other and from all Desirée samples, whether GM or non-GM. No separation was observed between wild-type Desirée and transformed lines in the components, again indicating that there was much less variation between Desirée and the GM lines than between the different non-transgenic varieties.

5.5 Changes in protein patterns during tuber life cycle (II)

To obtain a comprehensive set of data of tuber protein profiles, proteins were extracted from 13 independent groups of samples covering a range of developmental stages and predicted metabolic activities (**II**, Table 1). There were clear qualitative and quantitative differences in protein profiles resolved by 2-DE between the developmental stages (**II**, Figure 1).

Principal component analysis was used to investigate whether or not tuber developmental stages could be separated on the basis of their protein profiles. The first two components accounted for 35% and 19.7% of the total variance, respectively. Stages 1 and 2 (non-swelling and swelling stolons) clearly separated from other stages along PC2, whereas small developing tubers (stage 3), larger developing tubers (stage 4) and tubers from senesced plants (stages 5 to 8) separated from each other along PC1 (**II**, Figure 2A). Stages 7 and 8 (sprouting tubers) also separated from the other stages on PC1 versus PC4 (5.2% of the total variance [**II**, Figure 2B]).

Previous work has shown that tuber excision can stimulate substantial metabolic changes in a relatively short time scale which mimic the sink-source transition found during normal maturation and senescence [130]. Excising developing tubers from the mother plant did not seem to have a major influence on protein levels compared with leaving the tubers attached, as stage 4a separated only slightly from 4b and, overall, resembled 4b more closely than stages 5 to 8. In contrast, storing the tubers from senesced potato plants at low temperature (5 °C, stage 6a) had a small but recognisable effect on proteins: spot 0314 (not shown), a spot matched against a putative mitochondrial processing peptidase, an isoform of ascorbate peroxidase and a putative pyridoxine biosynthesis protein, as well as a spot identified as a glycine-rich RNA-binding protein were more abundant compared to stages 5 or 6b (storage at room temperature) (**II**, Table 2 and Supplementary table 1). Conversely, one protein spot was less abundant at stage 6a compared to stages 5 or 6 (data not shown). Although storage at low temperatures often leads to changes in sugar levels [189], none of the proteins identified as involved in carbohydrate metabolism appeared to change significantly in stage 6a.

Cluster analysis of the profiles (defined as the set of estimated mean intensities at each of the tuber stages) for 150 spots selected based on their low false discovery rate (FDR<0.02%) revealed five major types of profiles designated A to E (**II**, Table 2 and Figure 3), the largest of which included proteins that were present at high levels in developing tubers but decreased during tuber maturation. 59 of the proteins were identified by HPLC-ESI-MS/MS (**II**, Table 2); in addition, 50 proteins that were not included in the cluster analysis were identified, including several novel proteins. Most noticeably, the development process was

characterized by the accumulation of the major storage protein patatin isoforms and enzymes involved in disease and defence reactions. Furthermore, enzymes involved in carbohydrate and energy metabolism and protein processing were associated with early developmental stages but decreased during tuber maturation.

5.6 The effect of alternative cultivation practices (III)

To assess the effect of different production systems and cultivation practices, potato plants were grown in a carefully designed field trial using three management regimes with the primary variants being fertilizer (mineral or compost), crop protection (conventional or organic), and different pre-crops. Other parameters, including soil type and variety, were kept consistent across the trial.

Protein profiles of potato tuber samples were separated by 2-DE. The protein profiles resolved were very similar among all the 64 samples (**III**, Figure 1). A total of 1097 spots were matched across the 2-DE gels and considered in the statistical analyses. Principal component analysis (PCA) was used to investigate whether the agricultural treatments could be separated on the basis of the tuber protein profiles. Tubers grown with either conventional or organic fertilization separated along the 1st and 2nd principal components, which accounted for 11.5% and 8%, respectively, of all the variability in the data (**III**, Figure 2). The fertilizer effect was independent of the other factors. When ANOVA was performed on individual spots, it highlighted significant differences between the two fertilization regimes for at least 160 proteins selected on the basis of a false discovery rate of 5%.

However, there was no effect of different crop protection protocols and pre-crop (grass and clover vs. wheat) on the level of expression of different proteins (**III**, Supplementary Figures 1 and 2, respectively). There were also no significant interactions between the three factors.

The 160 protein spots that showed significant differences between the two fertilization regimes showed two types of profiles: 17 were present at higher levels in tubers grown with conventional fertilization regime, while 143 were more abundant in tubers grown with the organic fertilization regime (**III**, Figure 3). The standard error differences (SEDs) between the mean protein content in tubers grown under these two fertiliser types were all in the range of 2.5 to 5. A total of 46 of these protein spots were identified (**III**, Supplementary Table 5). Proteins that were more abundant in tubers grown with compost included several proteins involved in protein synthesis, folding and degradation (several heat shock proteins, chaperonins, proteases and subunits of the 20S proteasome) and in other hydrolytic reactions (β-xylosidase). Furthermore, several identified spots were matched against enzymes involved in glycolysis and energy metabolism. Various proteins that are commonly upregulated in stress responses (Kunitz-type enzyme

inhibitor, superoxide dismutase, ascorbate proxidase, glyoxalases I and II) were also increased in compost-grown tubers. Proteins that were more abundant in tubers grown with mineral fertilizer included a small heat shock protein, a proteinase inhibitor protein and some polypeptides identified as patatins. These spots were of smaller molecular weight than previously identified isoforms (**I**, **II**) and may be degradation products. Overall, the major isoforms of patatin did not show any major differences in this experiment.

6 Discussion

6.1 Applicability and reproducibility of potato protein profiling by 2-DE

For vegetatively propagated crops with complex genetics such as the potato, genetic modification for crop improvement offers significant potential. As a member of the Solanaceae, the potato was among the first crop plants to be accessible for transgenic approaches [48]. As the fourth most important food crop consumed worldwide, it is a realistic model for a field grown food crop and was thus selected as the model plant for this work. On the other hand it is a difficult one for proteomic studies, as the edible part, the tuber, contains little protein (ca. 2% of fresh weight) and large amounts of starch (ca. 15–25% [48]).

2-DE was deemed the most suitable approach for a complex plant for which there is little sequence information available, in particular as it was desirable to profile the whole complex mixture and as many proteins as possible. Data from a total of 316 2-DE gels were included in this Thesis; thus there was a need for a sufficiently high throughput for the large number of samples from field experiments. When this work was started, quantitative MS-based methods and DIGE were not available to the extent that they are now. In this Thesis, the 2-DE approach was applied to several sample sets and was shown capable of detecting changes in the levels of potato proteins, such as changes in protein expression during tuber development. The methods used allowed for the separation of 800–1100 protein spots in the tuber. While this is likely to represent only a subset of all potato proteins, the tuber is a relatively specialised storage organ, and its protein and transcript complements tend to become simpler as tubers develop and mature [48, 88].

The reproducibility of the 2-DE method was similar to those reported in the literature [10, 28, 82, 124, 147]. In previously published work, analysis has, however, been limited to a much smaller number of proteins. The results also indicate that the reproducibility of spot quantity depends on the protein and its quantity and that including a large number of low-intensity spots is likely to increase the variation observed. The 2-DE gels themselves were the main source of variation observed compared to protein extraction and tuber-to-tuber variation, which is not unexpected, as there are many steps involved in running a 2-DE gel, many of which contribute to differences in spot and background intensity. One should be aware of this while designing experiments, and if the number of samples requires several gel batches, samples should be appropriately randomised across batches. The results also imply that the reliability of 2-DE results could be improved by increasing the number of replicate gels and not necessarily by increasing the number of biological replicates. Under the sampling regimes used in this experiment, variation between mature tubers of the same variety did not appear very high.

6.2 Natural variability of protein patterns in potato

6.2.1 Variability caused by different genetic backgrounds

For potato, almost no comparative data for protein profiling has so far been available. Proteomics of maize [24, 33], wheat [38] and other plant species have revealed a large extent of genetic variability in protein expression, and it seems that the large genetic variability of protein expression is commonplace [40]. Studies on barley cultivars have shown that single amino acid substitutions are sufficient to explain pI differences of β-amylase spots [51]. Thus qualitative differences observed in 2-DE are likely to be due to allelic variations and possibly posttranslational modifications.

Potato belongs to the Solanaceae family, which encompasses about 90 genera and 2800 species such as tomato (*Lycopersicon esculentum*) and tobacco (*Nicotiana tabacum*). The genus, *Solanum*, includes all tuber-bearing wild species, as well as *S. tuberosum*, the species to which all common potato cultivars belong [75]. The potato has the richest genetic resources of any cultivated plant, and these genetic resources are generally easily incorporated into cultivars. Yet only a small number of related species have actually been used for introgression of resistance traits into cultivars, because of the introduction of undesirable 'wild' traits together with the desired trait [59]. *S. tuberosum* can be diploid (2n=24) or tetraploid (4n=48); however, all common cultivars over the world are tetraploid [75]. The genome of the potato has not been sequenced yet (`http://www.potatogenome.org/`). Thus, sequence information that may be used is somewhat limited, making MS-based proteomics a challenge for potato. Many proteins identified during this work were matched against proteins from other species and/or potato EST sequences.

Although the potato gene pool used by European breeders has for historical reasons been limited, a large number of very different cultivars has been developed [19]. Variation in protein patterns in this wide selection of (non-GM) genetic backgrounds turned out to be very extensive. Many of the proteins that contributed to the separation of the non-GM genotypes appeared to be involved in disease and defense responses, sugar and energy metabolism or protein targeting and storage, presently considered to convey no safety risk.

The qualitative and quantitative differences in patatin isoforms across the non-GM genotypes were striking. Allergenic reactions caused by potato are not considered particularly common, but of potato proteins, patatin (Sol t 1) plays a role in IgE-mediated food allergy [161]. Patatin is a major protein, comprising as much as 40% of total protein [142]. While patatin has been considered storage protein, lipase and β-1,3-glucanase activities have been described for some isoforms, indicating that patatins may play a role in defence reactions [165]. In-

dividual members of the patatin gene family exhibit differential expression patterns e.g. depending on the tuber development stage [169], and the patatin protein isoforms also showed differential accumulation during the tuber life cycle. Also several potato proteins similar to soybean trypsin inhibitors are associated with allergic responses to raw potato [162]. These proteins were observed at high levels in potato tubers and they showed a great deal of variation between the different genotypes. It therefore appears that the allergen content in potato tubers shows high variability due to several natural factors.

6.2.2 Variability caused by developmental and physiological states

The tubers are derived from underground stems (stolons) that in favourable conditions enlarge to form tubers. The active growth of tubers is accompanied by major changes in the physiology and metabolism that lead to large depositions of starch and storage proteins [138]. Tubers also decrease their general metabolic activity and behave as typical storage sinks [48]. Since, unlike in the seeds of many important species, dehydration does not occur during tuber maturation, the tubers remain metabolically active (although metabolic rates do change [48]) through maturation and eventually into the dormant phase. Dormancy break and sprouting subsequently lead to plant establishment in the field, completing the life cycle. Potatoes can be eaten at several stages of development and storage. An understanding of the extent of variation consumers might be naturally exposed to forms an important backdrop to data from other experiments.

Proteomic profiling by 2-DE combined with multivariate analysis was able to separate stages corresponding to major phases of the potato tuber life cycle, i.e. stolon to tuber and sink to source transitions. Furthermore, some individual stages could be separated, such as sprouting tubers. Altogether, over 100 proteins could be identified, including several novel proteins.

Studies at the transcript level using cDNA microarrays [88] have recently provided an overview of gene expression during tuber development and storage. Although the experimental conditions described by Kloosterman *et al.* [88] were slightly different from the ones used in our experiment, at the stages of early tuber growth designated 3a, 3b, 3c and 3d, several genes with an upregulated transcript profile such as proteinase inhibitors, lipoxygenase and patatins also showed an increasing protein profile corresponding to profile types A and B. The largest transcript profile type corresponded to a decrease in transcription over the same developmental stages and was similar to our profile type E (**II**, Figure 3). The small number of distinct major profiles of protein abundances identified, with the largest group being proteins that decrease during development, is in accordance with the results obtained from metabolomics (N. Massat, personal communication) and transcriptomics and is likely to reflect the greater metabolic activity of

the developing tuber in comparison with the mature tuber. Characteristic profiles were in many cases clearly different between protein family members, being thus consistent with the results obtained by transcript profiling [88].

The results thus suggest that, while mature tubers are the most relevant from a food safety perspective, protein expression appears highest during early tuber development, and a number of proteins of interest are likely to be missed, if only mature tubers are analysed. Protein profiles of mature tubers, however, remain relatively stable during storage and sprouting. Thus the sampling of mature tubers is practical and likely to lead to relatively uniform material, although one should still keep in mind that tubers of similar size may not be physiologically equivalent.

6.2.3 Variability caused by alternative cultivation techniques

Plants, especially field grown crops, respond to environmental conditions and agronomic practices, and differences in protein profiles due to these factors are likely to occur. Furthermore, there is currently some debate on whether practices such as organic cultivation lead to differences in crop composition. There are significant differences in the agronomic practices used in intensive conventional and organic food production systems, in particular to fertility management, crop protection protocols and rotational designs [108]. Agronomic practices used in organic and 'low input' production systems have repeatedly been claimed to deliver environmental or biodiversity, agronomic and food quality benefits compared to intensive conventional systems. However, evidence for differences in composition of nutritionally relevant compounds is currently insufficient and often contradictory [18, 193]. Also there is to our knowledge little information on the effects of agricultural production systems on gene expression and protein profiles in crops that could elucidate the underlying mechanisms.

The main aim of the cultivation practice experiment was to assess the effect of three main factors or components of agricultural production systems (rotational design, fertilization and crop protection) on protein expression patterns in potato tubers. The fertilisation regime was shown to be the main source of differences in the protein profiles, while the crop protection regime and the previous crop did not appear to lead to changes at the protein level. Interestingly, tubers treated with conventional crop protection could not be separated from those treated with organic crop protection, nor was there any obvious effect of pre-sowing plots with either grass and clover or wheat. According to existing literature, if an effect of crop protection is present, it is most likely small, and the effects occurring in different cultivation systems may to some extent balance each other out by imposing different types of stress [20]. Both crop protection regimes were relatively effective measures against various diseases in this experiment, and disease incident was low.

The accumulation of chaperones and proteins involved in the degradation of proteins and other large molecules indicates that protein synthesis and turnover and other hydrolytic reactions were more active in the tubers grown using the organic fertilisation regime. Furthermore, the levels of several enzymes involved in glycolysis and energy metabolism were higher in tubers grown using the organic fertilisation regime, suggesting a higher rate of cell respiration. Many of the identified proteins are also highly expressed in defence reactions. Many of these mechanisms are also known to increase as responses to various stresses, suggesting that the compost fertilisation induced a stress response in the tubers. In agreement with this, it has been previously shown that compost as soil amendment can result in the activation of systemic resistance and increase in the activities of defence-related proteins [184]. On the other hand, nitrogen metabolism and responses to nitrate fertilisation are complex and frequently linked to carbon metabolism and glycolysis [35, 168]. Thus the differences seen between the two fertilization regimens could also, at least in part, be due to different levels of nitrogen available, although a similar amount of nitrogen was applied both in the form of compost and mineral fertilization. With the mineral fertilization, all nitrogen applied will have been immediately available for the plant after application and available nitrogen levels will have subsequently declined due to uptake by plants and to nitrogen losses, while the nitrogen supply for compost is expected to have been very low immediately after planting and have subsequently increased throughout the growing season.

6.3 The effect of genetic modification on the proteome is small

Compared with the natural variation observed in the non-GM samples, the effects of genetic modification on the proteome were considerably less pronounced. Indeed, statistical analysis showed no clear differences between the protein patterns of the GM lines and their controls. No new proteins unique to individual GM lines were observed. Therefore, on the basis of this analysis, there was no evidence for any major changes in protein pattern in the GM lines tested. This in itself is interesting as some of the lines, e.g. Mal1 and Sam35S, produced extremely stunted plants with low tuber yield [97, 171, 173]. The chemical composition of the tubers did not appear to be clearly affected either [163]. This suggests that a strong morphological phenotype does not necessarily mean that there are changes in composition, and similarly, a change in composition does not necessarily mean changes in agronomic properties. However, nine proteins showing statistically significant differences in expression were detected, and seven could be tentatively identified. It would require observations over several years and climatic conditions to confirm that these are truly unintended effects.

6.4 Natural variation appears much higher than variation between GM lines

In these studies, the main sources of variation in the protein profiles of potato tubers were genetic backround and developmental/physiological state. Genetic modification, along with factors such as cultivation technique and tuber-to-tuber variability appeared to cause much less variation.

These results have been corroborated recently by Catchpole *et al.* [26] who compared several GM potato lines and cultivars by metabolic profiling. The authors found differences between the GM lines only in those metabolites that were targets of the genetic modification; apart from those compounds, the GM lines could not be distinguished from their controls in PCA. On the other hand, all cultivars could be clearly distinguished from one another. This led the authors to suggest that the GM lines could be considered substantially equivalent to the non-GM parent line.

Several other studies demonstrating the substantial equivalence of GM plants have recently been published using different model plants and different profiling methods: potato metabolomics [39], tomato proteomics [34] and metabolomics [107], wheat transcriptomics [13, 66] and metabolomics [11], and *Arabidopsis* proteomics [148, 149]. Thus there is increasing evidence that only few statistically significant differences, apart from the intended effect of the modification, are usually found between GM lines and/or controls and that these differences are generally within the range of natural variation due to physiological, genetic or environmental factors. In targeted analyses of a range of transgenic potato tubers, statistically significant differences were observed between controls and GM lines but they appeared to be random and were also observed between wild-type controls and tubers derived from tissue culture only or tubers transformed with an empty vector [163]. The authors also raised the possibility that somaclonal variation may be responsible for some differences observed between specific GM lines or tissue culture/empty vector controls.

Apart from the factors considered here, at the transcriptome and metabolite level, differences due to developmental stages [88, 182] and diurnal changes [183] can be extensive, and changes in the composition due to factors such as climatic conditions and plant responses to herbivore and pathogen attack can be expected. For instance, toxic secondary metabolites in potato are increased in response to stresses [117]. Non-GM genotypes of wheat grown on different sites and in different years also showed differences in their metabolite profiles [11]. These effects clearly should be taken into account in profiling studies. Furthermore, the fact that differences due to an alternative cultivation technique could be detected using protein profiling suggests that the methods should also be useful in other applications, such as the food safety assessment of different agricultural practices.

7 Conclusions

2-DE was used as a profiling tool to monitor the protein contents of potato tubers of GM and non-GM lines. The results from these protein profiling studies imply that transgene insertion *per se* has a much smaller effect on protein expression in the tuber than genetic recombination and other processes involved in conventional breeding. It is not surprising considering what is now known about the nature of plant genome and its dynamics. Even though genetic modification does not appear to generate major changes apart from the ones targeted to occur, proteins with increased level in the GM line compared to the conventional counterpart might be worth further attention if the level clearly falls outside the normal variation. This is to exclude any risks from e.g. increased expression of potent allergens. As the current profiling methods produce huge amounts of data, it is almost inevitable that some statistically significant differences will be found. Therefore the focus should be in truly consistent differences.

Furthermore, extensive differences in the protein profiles of the tubers were found to be caused by development and storage and alternative cultivation techniques. Such factors should be taken into account when characterising GM lines and determining whether possibly observed differences are outside of variation caused by diverse physiological and environmental conditions.

Because there are many factors that may affect the composition of crops, proteomic profiling used in a random comparison of products available on the market is likely to be difficult, as the interpretation of the results is not straightforward. Instead, proteomics should be much more useful in the early stages of plant breeding as a tool to screen for possible alterations in protein composition that might trigger a more thorough analysis and assessment of safety. Proteomics could be worthwhile particularly with future GM crops having increasingly complex modifications and potentially more extensive alterations in their composition.

Like other profiling methods, proteomic screening is not yet in routine use when assessing the safety of GM products. While many allergens are frequently present at high levels in plant material, many proteins with possible safety implications may be of very low abundance. A comprehensive and sensitive coverage of all proteins is therefore important but difficult to achieve with current analytical and sample preparation methods. For practical safety assessment purposes, proteomic and other profiling methods of the present day still require expensive equipment and well trained people, but current active development of e.g. more robust MS-based methods may lead to automated, fully validated analytical tools in the coming years.

As a whole, profiling methods have the potential to reduce uncertainty by providing much more information on crop composition than targeted analysis alone. The combined development and application of validated metabolomic, proteomic

and transcriptomic approaches in plant biology will contribute to our knowledge of biological systems but there may also be clear benefits in the area of food safety in the future.

References

[1] **Adams KL, Wendel JF** (2005) Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* 8, 135–141.

[2] **Aebersold R, Mann M** (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207.

[3] **Agrawal GK, Yonekura M, Iwahashi Y, Iwahashi H, Rakwal R** (2005) System, trends and perspectives of proteomics in dicot plants Part II: proteomes of the complex developmental stages. *J Chromatogr B* 815, 125–136.

[4] **Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I** (2003) A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 3, 36–44.

[5] **Al-Kaff NS, Kreike MM, Covey SN, Pitcher R, Page AM, Dale PJ** (2000) Plants rendered herbicide-susceptible by cauliflower mosaic virus-elicited suppression of a 35S promoter-regulated transgene. *Nat Biotechnol* 18, 995–999.

[6] **Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR** (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.

[7] **Ames BN, Profet M, Gold LS** (1990) Dietary pesticides (99.99% all natural). *Proc Natl Acad Sci USA* 87, 7777–7781.

[8] **Ames BN, Profet M, Gold LS** (1990) Nature's chemicals and synthetic chemicals: comparative toxicology. *Proc Natl Acad Sci USA* 87, 7782–7786.

[9] **The Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

[10] **Asirvatham VS, Watson BS, Sumner LW** (2002) Analytical and biological variances associated with proteomic studies of *Medicago truncatula* by two-dimensional polyacrylamide gel electrophoresis. *Proteomics* 2, 960–968.

[11] **Baker JM, Hawkins ND, Ward JL, Lovegrove A, Napier JA, Shewry PR, Beale MH** (2006) A metabolomic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotechnol J* 4, 381–392.

[12] **Baker SJ, Newton AC, Crabb D, Guy, DC, Jefferies RA, MacKerron DKL, Thomas WTB, Gurr SJ** (1998) Temporary partial breakdown of *mlo*-resistance in spring barley by sudden relief of soil water-stress under field conditions: the effects of genetic background and *mlo* allele. *Plant Pathol* 47, 401–410.

[13] **Baudo MM, Lyons R, Powers S, Pastori GM, Edwards JH, Holdsworth MJ, Shewry PR** (2006) Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. *Plant Biotechnol J* 4, 369–380.

[14] **Beaujean A, Sangwan RS, Lecardonnel A, Sangwan-Norreel BS** (1998) *Agrobacterium*-mediated transformation of three economically important potato cultivars using sliced internodal explants: an efficient protocol of transformation. *J Exp Bot* 49, 1589–1595.

[15] **Bertone P, Snyder M** (2005) Advances in functional protein microarray technology. *FEBS J* 272, 5400–5411.

[16] **Birch ANE, Geoghegan IE, Griffiths DW, McNicol JW** (2002) The effect of genetic transformations for pest resistance on foliar solanidine-based glycoalkaloids of potato (*Solanum tuberosum*). *Ann Appl Biol* 140, 143–149.

[17] **Birch RG** (1997) Plant transformation: problems and strategies for practical application. *Annu Rev Plant Physiol Plant Mol Biol* 48, 297–326.

[18] **Bourn D, Prescott J** (2002) A comparison of the nutritional value, sensory qualities, and food safety of organically and conventionally produced foods. *Crit Rev Food Sci Nutr* 42, 1–34.

[19] **Bradshaw JE, Mackay GR** (1994) Breeding strategies for clonally propagated potatoes. In Bradshaw JE, Mackay GR, eds, Potato genetics. CABI, Wallingford, UK, pp 467–497.

[20] **Brandt K, Mølgaard JP** (2001) Organic agriculture: does it enhance or reduce the nutritional value of plant foods? *J Sci Food Agric* 81, 924–931.

[21] **Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M** (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat Genet* 29, 365–371.

[22] **Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K** (2000) Gene expression analysis by massively

parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18, 630–634.

[23] **Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A** (2005) Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* 17, 343–360.

[24] **Burstin J, de Vienne D, Dubreuil P, Damerval C** (1994) Molecular markers and protein quantities as genetic descriptors in maize. I. Diversity among 21 inbred lines. *Theor Appl Genet* 89, 943–950.

[25] **Canovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock HP, Rossignol M** (2004) Plant proteome analysis. *Proteomics* 4, 285–298.

[26] **Catchpole GS, Beckmann M, Enot DP, Mondhe M, Zywicki B, Taylor J, Hardy N, Smith A, King RD, Kell DB, Fiehn O, Draper J** (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc Natl Acad Sci USA* 102, 14458–14462.

[27] **Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Gatehouse AMR, Kärenlampi S, Kok EJ, Leguay J-J, Lehesranta S, Noteborn HP, Pedersen J, Smith M** (2004) Unintended effects and their detection in genetically modified crops. *Food Chem Toxicol* 42, 1089–1125.

[28] **Chang J, Van Remmen H, Cornell J, Richardson A, Ward WF** (2003) Comparative proteomics: characterization of a two-dimensional gel electrophoresis system to study the effect of aging on mitochondrial proteins. *Mech Ageing Dev* 124, 33–41.

[29] **Chen S, Jin W, Wang M, Zhang F, Zhou J, Jia Q, Wu Y, Liu F, Wu P** (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. *Plant J* 36, 105–113.

[30] **Christou P** (1992) Genetic transformation of crop plants using microprojectile bombardment. *Plant J* 2, 275–281.

[31] **Clegg MT, Cummings MP, Durbin ML** (1997) The evolution of plant nuclear genes. *Proc Natl Acad Sci USA* 94, 7791–7798.

[32] **Conner AJ, Williams MK, Abernethy DJ, Fletcher PJ, Genet RA** (1994) Field performance of transgenic potatoes. *New Zeal J Crop Hort* 22, 361–371.

[33] **Consoli L, Damerval C** (2001) Quantification of individual zein isoforms resolved by two-dimensional electrophoresis: genetic variability in 45 maize inbred lines. *Electrophoresis* 22, 2983–2989.

[34] **Corpillo D, Gardini G, Vaira AM, Basso M, Aime S, Accotto GP, Fasano M** (2004) Proteomics as a tool to improve investigation of substantial equivalence in genetically modified organisms: The case of a virus-resistant tomato. *Proteomics* 4, 193–200.

[35] **Coruzzi GM, Zhou L** (2001) Carbon and nitrogen sensing and signaling in plants: emerging 'matrix effects'. *Curr Opin Plant Biol* 4, 247–253.

[36] **Dai S, Zheng P, Marmey P, Zhang S, Tian W, Chen S, Beachy RN, Fauquet C** (2001) Comparative analysis of transgenic rice plants obtained by Agrobacterium-mediated transformation and particle bombardment. *Mol Breeding* 7, 25–33.

[37] **Damerval C, Le Guilloux M** (1998) Characterization of novel proteins affected by the *o2* mutation and expressed during maize endosperm development. *Mol Gen Genet* 257, 354–361.

[38] **David JL, Zivy M, Cardin ML, Brabant P** (1997) Protein evolution in dynamically managed populations of wheat: adaptive responses to macro-environmental conditions. *Theor Appl Genet* 95, 932–941.

[39] **Defernez M, Gunning YM, Parr AJ, Shepherd LVT, Davies HV, Colquhoun IJ** (2004) NMR and HPLC-UV profiling of potatoes with genetic modifications to metabolic pathways. *J Agric Food Chem* 52, 6075–6085.

[40] **De Vienne D, Bost B, Fiévet J, Zivy M, Dillmann C** (2001) Genetic variability of proteome expression and metabolic control. *Plant Physiol Biochem* 39, 271–283.

[41] **EC** 1999. Commission Regulation (EC) No 2316/1999 of 22 October 1999 laying down detailed rules for the application of Council Regulation (EC) No 1251/1999 establishing a support system for producers of certain arable crops. Official Journal of the European Communities, L280, 43–65.

[42] **Edwards D, Batley J** (2004) Plant bioinformatics: from genome to phenome. *Trends Biotechnol* 22, 232–237.

[43] **FAO/WHO** (2000) Safety aspects of genetically modified foods of plant origin. Report of a joint FAO/WHO consultation. World Health Organization, Geneva.

[44] **Feldmann KA** (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J* 1, 71–82.

[45] **Feldmann KA, Marks MD, Christianson ML, Quatrano RS** (1989) A dwarf mutant of Arabidopsis generated by T-DNA insertion mutagenesis. *Science* 243, 1351–1354.

[46] **Fernie AR, Tadmor Y, Zamir D** (2006) Natural genetic variation for improving crop quality. *Curr Opin Plant Biol* 9, 196–202.

[47] **Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L** (2004) Metabolite profiling: from diagnostics to system biology. *Nat Rev Mol Cell Biol* 5, 763–769.

[48] **Fernie AR, Willmitzer L** (2001) Molecular and biochemical triggers of potato tuber development. *Plant Physiol* 127, 1459–1465.

[49] **Fiehn O** (2002) Metabolomics - the link between genotypes and phenotypes. *Plant Mol Biol* 48, 155–171.

[50] **Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L** (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18, 1157–1161.

[51] **Finnie C, Melchior S, Roepstorff P, Svensson B** (2002) Proteome analysis of grain filling and seed maturation in barley. *Plant Physiol* 129, 1308–1319.

[52] **Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R** (2003) A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. *Plant Mol Biol* 52, 161–176.

[53] **Fray RG, Wallace A, Fraser PD, Valero D, Hedden P, Bramley PM, Grierson D** (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. *Plant J* 8, 693–701.

[54] **Fridman E, Pleban T, Zamir D** (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc Natl Acad Sci USA* 97, 4718–4723.

[55] **Fukusaki E, Kobayashi A** (2005) Plant metabolomics: potential for practical operation. *J Biosci Bioeng* 100, 347–354.

[56] **Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D** (2001) Proteomic analysis of Arabidopsis seed germination and priming. *Plant Physiol* 126, 835–848.

[57] **Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J** (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol* 133, 664–682.

[58] **Gaspari M, Verhoeckx KCM, Verheij ER, van der Greef J** (2006) Integration of two-dimensional LC-MS with multivariate statistics for comparative analysis of proteomic samples. *Anal Chem* 78, 2286–2296.

[59] **Gebhardt C, Valkonen JP** (2001) Organization of genes controlling disease resistance in the potato genome. *Annu Rev Phytopathol* 39, 79–102.

[60] **Gelvin SB** (1998) The introduction and expression of transgenes in plants. *Curr Opin Biotechnol* 9, 227–232.

[61] **Gill KS, Gill BS, Endo TR, Taylor T** (1996) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* 144, 1883–1891.

[62] **Glinski M, Weckwerth W** (2005) The role of mass spectrometry in plant systems biology. *Mass Spectrom Rev* 25, 173–214.

[63] **Gorbunova V, Levy AA** (1999) How plants make ends meet: DNA double-strand break repair. *Trends Plant Sci* 4, 263–269.

[64] **Gorbunova V, Levy AA** (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucl Acid Res* 25, 4650–4657.

[65] **Gottlieb DM, Schultz J, Bruun SW, Jacobsen S, Sondergaard I** (2004) Multivariate approaches in plant science. *Phytochemistry* 65, 1531–1548.

[66] **Gregersen PL, Brinch-Pedersen H, Holm PB** (2005) A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic Res* 14, 887–905.

[67] **Gygi SP, Rist B, Aebersold R** (2000) Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol* 11, 396–401.

[68] **Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R** (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994–955.

[69] **Gygi SP, Rochon Y, Franza BR, Aebersold R** (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19, 1720–1730.

[70] **Görg A, Weiss W, Dunn MJ** (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4, 3665–3685.

[71] **Hajduch M, Casteel JE, Hurrelmeyer KE, Song Z, Agrawal GK, Thelen JJ** (2006) Proteomic analysis of seed filling in *Brassica napus*. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. *Plant Physiol* 141, 32–46.

[72] **Hall RD** (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol* 169, 453–468.

[73] **Hamilton AJ, Baulcombe DC** (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952.

[74] **Han FP, Fedak G, Ouellet T, Liu B** (2003) Rapid genomic changes in interspecific and intergeneric hybrids and allopolyploids of Triticeae. *Genome* 46, 716–723.

[75] **Hawkes JG** (1990) The potato: Evolution, biodiversity and genetic resources. Belhaven Press, London.

[76] **Herbik A, Giritch A, Horstmann C, Becker R, Balzer HJ, Bäumlein H, Stephan UW** (1996) Iron and copper nutrition-dependent changes in the protein expression in a tomato wild type and the nicotianamine-free mutant *chloronerva*. *Plant Physiol* 111, 533–540.

[77] **Horvath H, Jensen LG, Wong OT, Kohl E, Ullrich SE, Cochran J, Kannangara CG, von Wettstein D** (2001) Stability of transgene expression, field performance and recombination breeding of transformed barley lines. *Theor Appl Genet* 102, 1–11.

[78] **Hutchens TW, Yip TT** (1993) New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 7, 576–580.

[79] **Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall R, Kopka J, Lane GA, Lange BM, Liu JR, Mendes P, Nikolau BJ, Oliver SG, Paton NW, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner LW, Wang T, Walsh S, Syrkin Wurtele E, Kell DB** (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* 22, 1601–1606.

[80] **Jonas DA, Antignac E, Antoine JM, Classen HG, Huggett A, Knudsen I, Mahler J, Ockhuizen T, Smith M, Teuber M, Walker R, de Vogel P** (1996) The safety assessment of novel foods. *Food Chem Toxicol* 34, 931–940.

[81] **Jones GH, Armstrong SJ, Caryl AP, Franklin FCH** (2003) Meiotic chromosome synapsis and recombination in *Arabidopsis thaliana*; an integration of cytological and molecular approaches. *Chromosome Res* 11, 205–215.

[82] **Jorge I, Navarro RM, Lenz C, Ariza D, Porras C, Jorrin J** (2005) The Holm Oak leaf proteome: Analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry de novo sequencing and sequence similarity searching. *Proteomics* 5, 222–234.

[83] **Jung HG, Sheaffer CC** (2004) Influence of Bt transgenes on cell wall lignification and digestibility of maize stover for silage. *Crop Sci* 44, 1781–1789.

[84] **Kaeppler SM, Kaeppler HF, Rhee Y** (2000) Epigenetic aspects of somaclonal variation in plants. *Plant Mol Biol* 43, 179–188.

[85] **Kashkush K, Feldman M, Levy AM** (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genetics* 33, 102–106.

[86] **Kertbundit S, De Greve H, Deboeck F, Van Montagu M, Hernalsteens JP** (1991) *In vivo* random β-glucuronidase gene fusions in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 88, 5212–5216.

[87] **Kirschman JC, Suber RL** (1989) Recent food poisonings from cucurbitacin in traditionally bred squash. *Food Chem Toxicol* 27, 555–556.

[88] **Kloosterman B, Vorst O, Hall RD, Visser RGF, Bachem CW** (2005) Tuber on a chip: differential gene expression during potato tuber development. *Plant Biotechnol J* 3, 505–519.

[89] **Koistinen KM, Hassinen VH, Gynther PAM, Lehesranta SJ, Keinänen SI, Kokko HI, Oksanen EJ, Tervahauta AI, Auriola S and Kärenlampi SO** (2002) Birch PR-10c is induced by factors causing oxidative stress but appears not to confer tolerance to these agents. *New Phytol* 155, 381–391.

[90] **Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, Hays L, Schieltz D, Ulaszek R, Wei J, Wolters D, Yates JR III** (2002) Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci USA* 99, 11969–11974.

[91] **Kononov ME, Bassuner B, Gelvin SB** (1997) Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. *Plant J* 11, 945–957.

[92] **Koprek T, Rangel S, McElroy D, Louwerse JD, Williams-Carrier RE, Lemaux PG** (2001) Transposon-mediated single-copy gene delivery leads to increased transgene expression stability in barley. *Plant Physiol* 125, 1354–1362.

[93] **Kristensen C, Morant M, Olsen CE, Ekstrøm CT, Galbraith DW, Møller BL, Bak S** (2005) Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. *Proc Natl Acad Sci USA* 102, 1779–1784.

[94] **Kuiper HA, Kok EJ, Engel KH** (2003) Exploitation of molecular profiling techniques for GM food safety assessment. *Curr Opin Biotechnol* 14, 238–243.

[95] **Kuiper HA, Kleter GA, Noteborn HPJM, Kok EJ** (2001) Assessment of food safety issues related to genetically modified foods. *Plant J* 27, 503–528.

[96] **Kuiper HA, Noteborn HPJM, Peijnenburg AACM** (1999) Adequacy of methods for testing the safety of genetically modified foods. *Lancet* 354, 1315–1316.

[97] **Kumar A, Taylor MA, Arif SAM, Davies HV** (1996) Potato plants expressing antisense and sense S-adenosylmethionine decarboxylase (*SAMDC*) transgenes show altered levels of polyamines and ethylene: antisense plants display abnormal phenotypes. *Plant J* 9, 147–158.

[98] **Kusaba M** (2004) RNA interference in crop plants. *Curr Opin Biotechnol* 15, 139–143.

[99] **Kvasnička F** (2003) Proteomics: general strategies and application to nutritionally relevant proteins. *J Chromatogr B* 787, 77–89.

[100] **König A, Cockburn A, Crevel RWR, Debruyne E, Grafstroem R, Hammerling U, Kimber I, Knudsen I, Kuiper HA, Peijnenburg AACM, Penninks AH, Poulsen M, Schauzu M, Wal JM** (2004) Assessment of the safety of foods derived from genetically modified (GM) crops. *Food Chem Toxicol* 42, 1047–1088.

[101] **Labra M, Vannini C, Grassi F, Bracale M, Balsemin M, Basso B, Sala F** (2004) Genomic stability in *Arabidopsis thaliana* transgenic plants obtained by floral dip. *Theor Appl Genet* 109, 1512–1518.

[102] **Lai J, Li Y, Messing J, Dooner HK** (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability in maize. *Proc Natl Acad Sci USA* 102, 9068–9073.

[103] **Larkin PJ, Scowcroft WR** (1981) Somaclonal variation - a novel source of variability from cell cultures for plant improvement. *Theor Appl Genet* 60, 197–214.

[104] **Laufs P, Autran D, Traas J** (1999) A chromosomal paracentric inversion associated with T-DNA integration in Arabidopsis. *Plant J* 18, 131–139.

[105] **Lebel EG, Masson J, Bogucki A, Paszkowski J** (1993) Stress-induced intrachromosomal recombination in plant somatic cells. *Proc Natl Acad Sci USA* 90, 422–426.

[106] **Lee M** (1988) The chromosomal basis of somaclonal variation. *Ann Rev Plant Physiol Plant Mol Biol* 39, 413–437.

[107] **Le Gall G, Colquhoun IJ, Davis AL, Collins GJ, Verhoeyen ME** (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J Agric Food Chem* 51, 2447–2456.

[108] **Lewis WJ, van Lenteren JC, Phatak SC, Tumlinson, JH III** (1997) A total system approach to sustainable pest management. *Proc Natl Acad Sci USA* 94, 12243–12248.

[109] **Liang P, Pardee AB** (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967–971.

[110] **Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III** (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17, 676–682.

[111] **Liu Y, Mitsukawa N, Oosumi T, Whittier RF** (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* 8, 457–463.

[112] **Liu B, Wendel JF** (2003) Epigenetic phenomena and the evolution of plant allopolyploids. *Mol Phylogen Evol* 29, 365–379.

[113] **Lopez MF** (2000) Better approaches to finding the needle in a haystack: Optimizing proteome analysis through automation. *Electrophoresis* 21, 1082–1093.

[114] **Lönnig WE, Saedler H** (2002) Chromosome rearrangements and transposable elements. *Annu Rev Genet* 36, 389–410.

[115] **Marqués K, Sarazin B, Chané-Favre L, Zivy M, Thiellement H** (2001) Comparative proteomics to establish genetic relationships in the Brassicaceae family. *Proteomics* 1, 1457–1462.

[116] **Maruta Y, Ueki J, Saito H, Nitta N, Imaseki H** (2001) Transgenic rice with reduced glutelin content by transformation with glutelin A antisense gene. *Mol Breeding* 8, 273–284.

[117] **Matthews D, Jones H, Gans P, Coates S, Smith LMJ** (2005) Toxic secondary metabolite production in genetically modified potatoes in response to stress. *J Agric Food Chem* 53, 7766–7776.

[118] **Meyers BC, Galbraith GW, Nelson T, Agrawal V** (2004) Methods for transcriptional profiling in plants. Be fruitful and replicate. *Plant Physiol* 135, 637–652.

[119] **Momma K, Hashimoto W, Ozawa S, Kawai S, Katsube T, Takaiwa F, Kito M, Utsumi S, Murata K** (1999) 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* 63, 314–318.

[120] **Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A** (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet* 37, 997–1002.

[121] **Muskens MWM, Vissers APA, Mol JNM, Kooter JM** (2000) Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. *Plant Mol Biol* 43, 243–260.

[122] **Nacry P, Camilleri C, Courtial B, Caboche M, Bouchez D** (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* 149, 641–650.

[123] **Napoli C, Lemieux C, Jorgensen R** (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. *Plant Cell* 2, 279–289.

[124] **Nishihara JC, Champion KM** (2002) Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* 23, 2203–2215.

[125] **Novak WK, Haslberger AG** (2000) Substantial equivalence of antinutrients and inherent plant toxins in genetically modified novel foods. *Food Chem Toxicol* 38, 473–483.

[126] **OECD** (1993) Safety evaluation of foods derived by modern biotechnology: concepts and principles. Organisation for Economic Cooperation and Development, Paris.

[127] **OECD** (2002a) Consensus documents on compositional considerations for new varieties of potatoes: key food and feed nutrients, anti-nutrients and toxicants. Series on the Safety of Novel Foods and Feeds No. 4, Consensus Documents. Organisation for Economic Cooperation and Development, Paris.

[128] **OECD** (2002b) Consensus documents on compositional considerations for new varieties of maize (*Zea mays*): key food and feed nutrients, anti-nutrients and secondary plant metabolites. Series on the Safety of Novel Foods and Feeds No. 6, Consensus Documents. Organisation for Economic Cooperation and Development, Paris.

[129] **O'Farrell PH** (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007–4021.

[130] **Oparka KJ, Davies HV, Wright KM, Viola R, Prior DAM** (1990) Effect of sink isolation on sugar uptake and starch synthesis by potato-tuber storage parenchyma. *Planta* 182, 113–117.

[131] **Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee H, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA** (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends Genet* 19, 141–147.

[132] **Ozkan H, Levy AA, Feldman M** (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13, 1735–1747.

[133] **Patton WF** (2002) Detection technologies in proteome analysis. *J Chromatograph B* 771, 3–31.

[134] **Peck SC** (2005) Update on proteomics in Arabidopsis. Where do we go from here? *Plant Physiol* 138, 591–599.

[135] **Pedersen SK, Harry JL, Sebastian L, Baker J, Traini MD, McCarthy JT, Manoharan A, Wilkins MR, Gooley AA, Righetti PG, Packer NH, Williams KL, Herbert BR** (2003) Unseen proteome: Mining below the tip of the iceberg to find low abundance and membrane proteins. *J Proteome Res* 2, 303–311.

[136] **Perkins DN, Pappin DJ, Creasy DM, Cottrell JS** (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.

[137] **Pots AM, Gruppen H, Hessing M, van Boekel MA, Voragen AG** (1999) Isolation and characterization of patatin isoforms. *J Agric Food Chem* 47, 4587–4592.

[138] **Prat S, Frommer WB, Höfgen R, Keil M, Kossmann J, Koster-Topfer M, Liu XJ, Muller B, Pena-Cortes H, Rocha-Sosa M, Sánchez-Serrano JJ, Sonnewald U, Willmitzer L** (1990) Gene expression during tuber development in potato plants. *FEBS Lett* 268, 334-338.

[139] **Prescott VE, Campbell PM, Moore A, Mattes J, Rothenberg ME, Foster PS, Higgins TJV, Hogan SP** (2005) Transgenic expression of bean α-amylase inhibitor in peas results in altered structure and immunogenicity. *J Agric Food Chem* 53, 9023–9030.

[140] **Puchta H** (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J Exp Bot* 56, 1–14.

[141] **Rabilloud T** (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* 2, 3–10.

[142] **Racusen D, Foote M** (1980) A major soluble glycoprotein of potato tubers. *J Food Biochem* 4, 43–52.

[143] **Richter TE, Pryor TJ, Bennetzen JL, Hulbert SH** (1995) New rust resistance specificities associated with recombination in the *Rp1 complex* in maize. *Genetics* 141, 373–381.

[144] **Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR** (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13, 11–29.

[145] **Rose JKC, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS** (2004) Tackling the plant proteome: practical approaches, hurdles and experimental tools. *Plant J* 39, 715–733.

[146] **Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ** (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154–1169.

[147] **Ruebelt MC, Leimgruber NK, Lipp M, Reynolds TL, Nemeth MA, Astwood JD, Engel KH, Jany KD** (2006) Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 1. Assessing analytical validation. *J Agric Food Chem* 54, 2154–2161.

[148] **Ruebelt MC, Lipp M, Reynolds TL, Astwood JD, Engel KH, Jany KD** (2006) Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 2. Assessing natural variability. *J Agric Food Chem* 54, 2162–2168.

[149] **Ruebelt MC, Lipp M, Reynolds TL, Schmuke JJ, Astwood JD, DellaPenna D, Engel KH, Jany KD** (2006) Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 3. Assessing unintended effects. *J Agric Food Chem* 54, 2169–2177.

[150] **Salomon S, Puchta H** (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* 17, 6086–6302.

[151] **Santoni V, Bellini C, Caboche M** (1994) Use of two-dimensional protein-pattern analysis for the characterization of *Arabidopsis thaliana* mutants. *Planta* 192, 557–566.

[152] **Santoni V, Delarue M, Caboche M, Bellini C** (1997) A comparison of two-dimensional electrophoresis data with phenotypical traits in *Arabidopsis* leads to the identification of a mutant (*cri1*) that accumulates cytokinins. *Planta* 202, 62–69.

[153] **Santoni V, Molloy M, Rabilloud T** (2000) Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* 21, 1054–1070.

[154] **Sarry JE, Sommerer N, Sauvage FX, Bergoin A, Rossignol M, Albagnac G, Romieu C** (2004) Grape berry biochemistry revisited upon proteomic analysis of the mesocarp. *Proteomics* 4, 201–215.

[155] **Saxena D, Stotzky G** (2001) *Bt* corn has a higher lignin content than non-*Bt* corn. *Am J Bot* 88, 1704–1706.

[156] **Schena M, Shalon D, Davis RW, Brown PO** (1995) Quantitative monitoring of gene expression patterns with a complementary cDNA microarray. *Science* 270, 467–470.

[157] **Schmidt MH, Raulf-Heimsoth M, Posch A** (2002) Evaluation of patatin as a major cross-reactive allergen in latex-induced potato allergy. *Ann Allergy Asthma Immunol* 89, 613–618.

[158] **Schnable PS, Hsia A, Nikolau BJ** (1998) Genetic recombination in plants. *Curr Opin Plant Biol* 1, 123–129.

[159] **Schneeberger RG, Zhang K, Tatarinova T, Troukhan M, Kwok SF, Drais J, Klinger K, Orejudos F, Macy K, Bhakta A, Burns J, Subramanian G, Donson J, Flavell R, Feldmann KA** (2005) *Agrobacterium* T-DNA integration in *Arabidopsis* is correlated with DNA sequence compositions that occur frequently in gene promoter regions. *Funct Integr Genomics* 5, 240–253.

[160] **Schubert D, Lechtenberg B, Forsbach A, Gils M, Bahadur S, Schmidt R** (2004) Silencing in Arabidopsis T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell* 16, 2561–2572.

[161] **Seppälä U, Alenius H, Turjanmaa K, Reunala T, Palosuo T, Kalkkinen N** (1999) Identification of patatin as a novel allergen for children with positive skin prick test responses to raw potato. *J Allergy Clin Immunol* 103, 165–171.

[162] **Seppälä U, Majamaa H, Turjanmaa K, Helin J, Reunala T, Kalkkinen N, Palosuo T** (2001) Identification of four novel potato (*Solanum tuberosum*) allergens belonging to the family of soybean trypsin inhibitors. *Allergy* 56, 619–626.

[163] **Shepherd LVT, McNicol JW, Razzo R, Taylor MA, Davies HV** (2006) Assessing the potential for unintended effects in genetically modified potatoes perturbed in metabolic and developmental processes. Targeted analysis of key nutrients and anti-nutrients. *Transgenic Research* 15, 409–407.

[164] **Shewmaker CK, Sheehy JA, Daley M, Colburn S, Ke DY** (1999) Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *Plant J* 20, 401–412.

[165] **Shewry PR** (2003) Tuber storage proteins. *Ann Bot* 91, 755–769.

[166] **Shirley BW, Hanley S, Goodman HM** (1992) Effects of ionizing radiation on a plant genome: Analysis of two Arabidopsis *transparent testa* mutations. *Plant Cell* 4, 333-347.

[167] **Somers DA, Makarevitch I** (2004) Transgene integration in plants: poking or patching holes in promiscuous genomes? *Curr Opin Biotechnol* 15, 126–131.

[168] **Stitt M** (1999) Nitrate regulation of metabolism and growth. *Curr Opin Plant Biol* 2, 178–186.

[169] **Stupar RM, Beaubien KA, Jin W, Song J, Lee MK, Wu C, Zhang HB, Han B, Jiang J** (2006) Structural diversity and differential transcription of the patatin multicopy gene family during potato tuber development. *Genetics* 172, 1263–1275.

[170] **Takano M, Egawa H, Ikeda JE, Wakasa K** (1997) The structures of integration sites in transgenic rice. *Plant J* 11, 353–361.

[171] **Taylor MA, George LA, Ross HA, Davies HV** (1998) cDNA cloning and characterisation of an α-glucosidase gene from potato (*Solanum tuberosum* L.). *Plant J* 13, 419–425.

[172] **Taylor CF, Paton NW, Garwood KL, Kirby PD, Stead DA, Yin Z, Deutsch EW, Selway L, Walker J, Riba-Garcia I, Mohammed S, Deery MJ, Howard JA, Dunkley T, Aebersold R, Kell DB, Lilley KS, Roepstorff P, Yates JR 3rd, Brass A, Brown AJ, Cash P, Gaskell SJ, Hubbard SJ, Oliver SG** (2003) A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat Biotechnol* 21, 247–254.

[173] **Taylor MA, Ross HA, McRae D, Stewart D, Roberts I, Duncan G, Wright F, Millam S, Davies HV** (2000) A potato α-glucosidase gene encodes a glycoprotein-processing α-glucosidase II-like activity. Demonstration of enzyme activity and effects of down-regulation in transgenic plants. *Plant J* 24, 305–316.

[174] **Tax FE, Vernon DM** (2001) T-DNA-associated duplication/translocations in Arabidopsis. Implications for mutant analysis and functional genomics. *Plant Physiol* 126, 1527–1538.

[175] **Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE** (2004) Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics* 3, 960–969.

[176] **Thiellement H, Zivy M, Plomion C** (2002) Combining proteomic and genetic studies in plants. *J Chromatography B* 782, 137–149.

[177] **Thomas CM, Jones DA, English JJ, Carroll BJ, Bennetzen JL, Harrison K, Burbidge A, Bishop GJ, Jones JDG** (1994) Analysis of the chromosomal distribution of transposon-carrying T-DNAs in tomato using the inverse polymerase chain reaction. *Mol Gen Genet* 242, 573–585.

[178] **Tuomainen MH, Nunan N, Lehesranta SJ, Tervahauta AI, Hassinen VH, Schat H, Koistinen KM, Auriola S, McNicol J, Kärenlampi SO** (2006) Multivariate analysis of protein profiles of metal hyperaccumulator *Thlaspi caerulescens* accessions. *Proteomics* 6, 3696–3706.

[179] **Tzfira T, Li J, Lacroix B, Citovsky V** (2004) *Agrobacterium* T-DNA integration: molecules and models. *Trends Genet* 20, 375–383.

[180] **Tzfira T, White C** (2005) Towards targeted mutagenesis and gene replacement in plants. *Trends Biotechnol* 23, 567–569.

[181] **Ünlü M, Morgan ME, Minden JS** (1997) Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* 18, 2071–2077.

[182] **Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR** (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep* 4, 989–993.

[183] **Urbanczyk-Wochniak E, Baxter C, Kolbe A, Kopka J, Sweetlove LJ, Fernie AR** (2005) Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. *Planta* 221, 891–903.

[184] **Vallad GE, Goodman RM** (2004) Systemic acquired resistance and induced systemic resistance in conventional agriculture. *Crop Sci* 44, 1920–1934.

[185] **van der Krol AR, Mur LA, Beld M, Mol JNM, Stuitje AR** (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* 2, 291–299.

[186] **van Gelder WMJ, Scheffer JJC** (1991) Transmission of steroidal glycoalkaloids from *Solanum vernei* to the cultivated potato. *Phytochemistry* 30, 165–168.

[187] **van Hal NLW, Vorst O, van Houwelingen AMML, Kok EJ, Peijnenburg A, Aharoni A, van Tunen AJ, Keijer J** (2000) The application of DNA mciroarrays in gene expression analysis. *J Biotechnol* 78, 271–280.

[188] **Velculescu VE, Zhang L, Vogelstein B, Kinzler KW** (1995) Serial analysis of gene expression. *Science* 270, 484–487.

[189] **Viola R, Davies HV** (1994) Effect of temperature on pathways of carbohydrate metabolism in tubers of potato (*Solanum tuberosum* L.). *Plant Sci* 103, 135–143.

[190] **Washburn MP, Ulaszek R, Deciu C, Schieltz DM, Yates JR 3rd** (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* 74, 1650–1657.

[191] **Washburn MP, Wolters D, Yates JR III** (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19, 242–247.

[192] **Windels P, De Buck S, Van Bockstaele E, De Loose M, Depicker A** (2003) T-DNA integration in Arabidopsis chromosomes. Presence and origin of filler DNA sequences. *Plant Physiol* 133, 2061–2068.

[193] **Woese K, Lange D, Boess C, Bögl KW** (1997) A comparison of organically and conventionally grown foods - results of a review of the relevant literature. *J Sci Food Agric* 74, 281–293.

[194] **Zhou S, Bailey MJ, Dunn MJ, Preedy VR, Emery PW** (2005) A quantitative investigation into the losses of proteins at different stages of a two-dimensional gel electrophoresis procedure. *Proteomics* 5, 2739–2747.

[195] **Zimmer JSD, Monroe ME, Qian WJ, Smith RD** (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev* 25, 450-482.

[196] **Zitnak A, Johnston GR** (1970) Glycoalkaloid content of B5141-6 potatoes. *Am Potato J* 47, 256–260.

[197] **Zupan JR, Zambryski P** (1995) Transfer of T-DNA from *Agrobacterium* to the plant cell. *Plant Physiol* 107, 1041–1047.

[198] **Zweiger G** (1999) Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends Biotechnol* 17, 429–436.

Kuopio University Publications C. Natural and Environmental Sciences

C 185. Luomala, Eeva-Maria. Photosynthesis, chemical composition and anatomy of Scots pine and Norway spruce needles under elevated atmospheric CO2 concentration and temperature.
2005. 137 p. Acad. Diss.

C 186. Heikkinen, Lasse M. Statistical estimation methods for electrical process tomography.
2005. 147 p. Acad. Diss.

C 187. Riihinen, Kaisu. Phenolic compounds in berries.
2005. 97 p. Acad. Diss.

C 188. Virkutyte, Jurate. Heavy metal bonding and remediation conditions in electrokinetically treated waste medias.
2005. 133 p. Acad. Diss.

C 189. Koistinen, Kaisa. Birch PR-10c: multifunctional binding protein.
2006. 79 p. Acad. Diss.

C 190. Airaksinen, Sanna. Bedding and manure management in horse stables: its effect on stable air quality, paddock hygiene and the compostability and utilization of manure.
2006. 91 p. Acad. Diss.

C 191. Asikainen, Arja. Use of computational tools for rapid sorting and prioritising of organic compounds causing environmental risk with estrogenic and cytochrome P450 activity.
2006. 51 p. Acad. Diss.

C 192. Ålander, Timo. Carbon composition and volatility characteristics of the aerosol particles formed in internal combustion engines.
2006. 54 p. Acad. Diss.

C 193. Molnár, Ferdinand. Structural analysis of the ligand-binding domains of human and mouse CAR, human VDR and human PPARs.
2006. 115 p. Acad. Diss.

C 194. Kasurinen, Anne. Soil-related processes of young silver birch trees grown under elevated CO_2 and O_3 .
2006. 64 p. Acad. Diss.

C 195. Metsärinne, Sirpa. Degradation of Novel and Conventional Complexing Agents.
2006. 138 p. Acad. Diss.

C 196. Heijari, Juha. Seed origin, forest fertilization and chemical elicitor influencing wood characteristics and biotic resistance of Scots pine.
2006. 39 p. Acad. Diss.

C 197. Hakulinen, Mikko. Prediction of density, structure and mechanical properties of trabecular bone using ultrasound and X-ray techniques.
2006. 84 p. Acad. Diss.

C 198. Al Natsheh, Anas. Quantum Mechanics Study of Molecular Clusters Composed of Atmospheric Nucleation Precursors.
2006. 55 p. Acad. Diss.