ELSEVIER

Food and Chemical Toxicology 42 (2004) 1089–1125

Food and Chemical Toxicology

www.elsevier.com/locate/foodchemtox

Unintended effects and their detection in genetically modified crops☆

F. Cellini[a], A. Chesson[b], I. Colquhoun[c], A. Constable[d], H.V. Davies[e], K.H. Engel[f,*], A.M.R. Gatehouse[g], S. Kärenlampi[h], E.J. Kok[i], J.-J. Leguay[j], S. Lehesranta[h], H.P.J.M. Noteborn[i], J. Pedersen[k], M. Smith[l]

[a]*Metapontum Agrobios, SS Jonica Km 448.2, I-75010 Metaponto Matera, Italy*
[b]*Nutritional Chemistry, Rowett Research Institute, Greenburn Road, Aberdeen AB21 9SB, UK*
[c]*Food Quality and Materials Science Division, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK*
[d]*Quality and Safety Assurance, Food Safety Group, Nestlé Research Centre, Route de Jorat, CH-1000 Lausanne, Switzerland*
[e]*Cellular and Environmental Physiology Department, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK*
[f]*Department of General Food Technology, Technical University of Munich, Am Forum 2, D-85350 Freising-Weihenstephan, Germany*
[g]*Department of Agricultural and Environmental Sciences, University of Newcastle, Ridley Building, Newcastle upon Tyne NE1 7RU, UK*
[h]*Department of Biochemistry, University of Kuopio, Savilahdentie 6, FIN-70211 Kuopio, Finland*
[i]*RIKILT- Institute of Food Safety, Wageningen University & Research Centre, Bornsesteeg 45, PO Box 230, NL-6700 AE Wageningen, The Netherlands*
[j]*Department of Plant Ecophysiology and Microbiology, DEVM Building 177, CEA Cadarache, F-13108 Saint-Paul-lez-Durance Cédex, France*
[k]*Institute of Food Safety and Toxicology, Danish Veterinary and Food Administration, 19 Moerkhoej Bygade, DK-2860 Søborg, Denmark*
[l]*Unilever Health Institute, Unilever Research Vlaardingen, P.O. Box 114, NL-3130 AC Vlaardingen, The Netherlands*

Received 16 November 2003; accepted 4 February 2004

Abstract

The commercialisation of GM crops in Europe is practically non-existent at the present time. The European Commission has instigated changes to the regulatory process to address the concerns of consumers and member states and to pave the way for removing the current moratorium. With regard to the safety of GM crops and products, the current risk assessment process pays particular attention to potential adverse effects on human and animal health and the environment. This document deals with the concept of unintended effects in GM crops and products, i.e. effects that go beyond that of the original modification and that might impact primarily on health. The document first deals with the potential for unintended effects caused by the processes of transgene insertion (DNA rearrangements) and makes comparisons with genetic recombination events and DNA rearrangements in traditional breeding. The document then focuses on the potential value of evolving ''profiling'' or ''omics'' technologies as non-targeted, unbiased approaches, to detect unintended effects. These technologies include metabolomics (parallel analysis of a range of primary and secondary metabolites), proteomics (analysis of polypeptide complement) and transcriptomics (parallel analysis of gene

Abbreviations: 2-D, two-dimensional; 2-DE, two dimensional electrophoresis; *A. comosa*, *Aegilops comosa*; AMDIS, automated mass spectral deconvolution and identification system; AT, adenosine and thymidine nucleotides; Bt, *Bacillus thuringiensis*; CBB, Coomassie Brilliant Blue; cDNA, DNA complementary to an RNA strand; CVA, canonical variates analysis; DFA, discriminant function analysis; DIGE, difference gel electrophoresis; DNA, deoxyribonucleic acid; DSB, double-strand break; *E. carotovora*, *Erwina carotovora*; *E. coli*, *Escherichia coli*; ENTRANSFOOD, European network safety assessment of genetically modified food crops; EPSPS, 5-enolpyruvyl-shikimate-3-phosphate synthase; EST, expressed sequence tag; ES, electrospray; ESI, electrospray ionisation; FID, flame ionisation detection; FISH, fluorescence in situ hybridisation; FT-ICR, Fourier transform- ion cyclotron resonance; FTIR, Fourier transform infrared; FTMS, Fourier transform mass spectrometry; GC, gas chromatography; GISH, genomic in situ hybridisation; GM, genetically modified; GMOCARE, EU-project on new methodologies for assessing the potential of unintended effects in genetically modified food crops; GP, genetic programming; HPLC, high-performance liquid chromatography; HR, homologous recombination; ICAT, isotope-coded affinity tags; IgE, immunoglobulin E; IR, Infrared; LC, liquid chromatography; LDA, linear discriminant analysis; MALDI, matrix-assisted laser desorption ionisation; MAR, matrix-attachment region; mRNA, messenger RNA; MS, mass spectrometry; NHEJ, non-homologous end joining; NMR, nuclear magnetic resonance; ORF, open reading frame; PC, principal component; PCA, principal component analysis; PCR, polymerase chain reaction; Ph, locus that suppresses recombination between homoeologous chromosomes; pI, iso-electric point; PLS, partial least squares; *R. solanacearum*, *Ralstonia solanacearum*; RNA, ribonucleic acid; SAGE, serial analysis of gene expression; SDS-PAGE, sodium dodecyl sulphate- polyacrylamide gel electrophoresis; SIMCA, soft independent modelling of class analogy; S/MAR, scaffold/matrix-attachment region; T-DNA, transferred region of the tumour-inducing plasmid of *Agrobacterium*; TF, transcription factor; Ti, tumour-inducing; TOF, time of flight; totRNA, total RNA; UV, ultraviolet.

☆ All authors contributed equally to the paper

* Corresponding author. Tel.: +49-8161-714250; fax: +49-8161-714259.
E-mail address: k.h.engel@lrz.tu-muenchen.de (K.H. Engel).

0278-6915/$ - see front matter
doi:10.1016/j.fct.2004.02.003

expression). The technologies are described, together with their current limitations. Importantly, the significance of unintended effects on consumer health are discussed and conclusions and recommendations presented on the various approaches outlined.

Keywords: Genetic modification; Substantial equivalence; Comparative analysis; Targeted analysis; Non-targeted analysis; Unpredictable effects; Unexpected effects

1. Introduction

The approaches used in the safety assessment of crops and foods derived from genetically modified organisms have been developed in collaborative work with international agencies such as the Organisation for Economic Co-ordination and Development (OECD, 1993) and the United Nations World Health Organisation/Food and Agricultural Organisation (FAO/WHO, 1991, 2000). The approach involves the concept of substantial equivalence, whereby the characteristics of the modified crop/food are compared to an existing food/crop with a history of safe use. This is most usually the parent crop from which the modifications were made. The process involves a targeted compositional analysis (profile of major nutrients and toxicants), and the expected intake and role in the diet of the novel food. This comparison provides the basis on which to focus further toxicological requirements for a safety assessment. Three scenarios may be considered (European Commission, 1997). Firstly, the novel food is equivalent to an accepted traditional food or ingredient, in which case no further testing is needed. Secondly, the novel food is equivalent to the traditional counterpart except for some well defined differences; safety assessments will be targeted to these differences. Finally, the novel food differs from the traditional counterpart in multiple and complex respects, or there are no traditional counterparts; such a novel food would require an extensive safety assessment. A stepwise procedure to carry out the safety assessment has been discussed by König et al. (2004).

Concerns have been raised that the current approach of using targeted analyses to compare the composition of GM crops to their traditional counterparts is biased (Millstone et al., 1999) and does not take into account the possibility of unintended effects and unexpected effects that could result directly or indirectly from the genetic modification. The potential occurrence of such "unintended effects" is currently one of the concerns being raised regarding the application of recombinant DNA techniques in the production of foods. In this report aspects related to the detection of unintended effects are discussed. As a basis for the paper, the following definitions have been adopted:

- "Intended effects" of genetic engineering are those that are targeted to occur from the introduction of the gene(s) in question and that fulfil the original objectives of the genetic transformation process.
- "Unintended effects" represent a statistically significant difference in the phenotype, response, or composition of the GM plant compared with the parent from which it is derived, but taking the expected effect of the target gene into account. Such comparisons should be made when GM and non-GM counterparts are grown under the same regimes and environments.
- "Predictable unintended effects" are those unintended changes that go beyond the primary expected effect(s) of introducing the target gene(s), but that may be explicable in terms of our current knowledge of plant biology and metabolic pathway integration and interconnections.
- "Unpredictable unintended effects" are those changes falling outside our present level of understanding.

Predictable and unpredictable unintended effects may or may not prove to have relevance in terms of product safety, but must be taken into account when assessing risk. In addition, there is little guidance for crop producers on which parameters should be measured for the comparison, which analytical methods should be used, and which sampling procedures should be followed to provide statistically sound analyses. With the development of new molecular techniques such as profiling techniques, it has been thought possible to address these concerns.

Safety assessments follow the well accepted paradigm, which includes hazard identification and characterisation, exposure assessments, and subsequent risk characterisation. The aim of the safety assessment of novel foods, including those produced by GM technologies, is to demonstrate that the novel food is as safe as its traditional counterpart (where one exists) and as such does not introduce any additional or new risks to the health of the consumer. The relevance of this issue with regard to consumer acceptance of GM crop-derived foods has been discussed by Frewer et al. (2004). Predictable and unpredictable unintended effects may or may not prove to have relevance in terms of product safety, but must be taken into account when assessing risk.

The present paper aims to critically discuss the detection of unintended effects in GM crops. However, prior

to this, it is necessary to describe the mechanisms whereby unintended effects may arise during (GM) crop breeding and how this compares to natural DNA recombination in plants. The ways in which unintended effects are dealt with in conventional breeding programmes are outlined. General issues relating to analytical procedures to detect unintended effects are discussed and the present status of targeted approaches are reported. Current developments in non-targeted approaches, i.e., the profiling techniques of genomics, proteomics, and metabolomics are presented with a critical discussion of their potentials and limitations. This last point is closely linked to the relevance of searching for unintended effects with respect to safety assessments and raises the question as to whether the more information we have available would, in reality, reduce the uncertainties in the safety assessment.

2. Transgene insertion in plant DNA in the context of natural DNA recombination

2.1. Classical plant breeding

Plant breeding has always exploited genetic methods both by using natural genetic variation combined with artificial selection and by inducing new variability by artificial means. In this sense, plant breeding can also be defined as ''applied plant genetics''. For plant breeders, genetic variability has not only been a matter of chance, but has also been induced, controlled, and exploited by artificial techniques. Plant population genetic structure has been widely changed by breeding practices (Allard, 1999), a fact that indicates the deep influence of breeding practices on the genetic make-up of plant crops. Without doubt, chromosomal recombination mechanisms are central to plant breeding. Indeed, chromosomal recombination influences not only the genetic variability, but also the speed and the efficiency with which the desired combination of genes is achieved. Since plant breeders always select for recombinants, it is not surprising that they also select, inadvertently, for plant genotypes with a higher frequency of chiasma. It has long been known that the frequency of chiasma is under genetic control and there is good evidence that modern plant cultivars have a higher frequency of chiasma than the natural population from which they have been derived (Rees, 1993).

A striking example of how the control of chromosome recombination is important in plant breeding is represented by the development of the wheat variety 'Compair' (Riley et al., 1968). In this case, an *Aegilops comosa* chromosome M2 addition line was created by back-crossing the wheat variety 'Chinese Spring' (2n = 2x = 42), susceptible to the fungus *Puccinia striiformis*, with the wild relative *A. comosa* (2n = 2x = 14), containing a gene which conferred resistance to the pathogen. Whilst chromosome M2 carried the resistance to the *Puccinia* it also carried other genes that had many unintended effects. The chromosomal M2 trait that carried the genes of interest could not be introgressed into the homoeologous wheat chromosome group 2, because 'Chinese Spring' carried the *Ph* locus that suppresses recombination between homoeologous chromosomes in favour of recombination between homologous ones. The problem was brilliantly overcome by crossing the addition M2 line with the wild relative *Aegilops speltoides*, a diploid with a complement of 14 chromosomes, which was able to suppress *Ph* activity. The tetraploid addition hybrid offspring was repeatedly back-crossed to Chinese Spring. A plant with 21 chromosome pairs was then selected through resistance to *Puccinia* and then selfed to give rise to the commercial variety 'Compair'.''

2.1.1. Genetic recombination

Plant genetic recombination mechanisms can be grouped into the two major mechanisms present in eukaryotes: (i) homologous recombination (HR) and (ii) illegitimate recombination, a form of the more general non-homologous end joining (NHEJ) process (for more details, see also Haber, 2000; Schnable et al., 1998; Camerini-Otero and Hsieh, 1995). Both mechanisms are explained by the currently most widely accepted models, which are based on the double-strand break (DSB) repair model (Szostak et al., 1983; also reviewed in Britt, 1996). NHEJ is the predominant form of recombination in plants. A summary (Gorbunova and Levy, 1999) of recent studies on the NHEJ process in plants indicated that:

- End-joining by simple ligation, with no sequence alteration, is rare.
- End-joining is usually associated with deletions ranging from 1bp up to >1kb.
- Rejoining frequently occurs at short repeats.

In summary, available data indicate that DSB repair in plants is more error prone than in other organisms. Since in plants the prominent mechanism of repair is via end-joining, errors that change the original sequence occur at very high frequency.

2.1.2. Localisation of recombination breakpoints

Studies on the evolution of disease resistance genes (Richter et al., 1995) and genes influencing quality traits (Fridman et al., 2000) suggest that genes arose from the shuffling of sequence domains between members of the gene family. The pattern is similar to that observed for DSB repair in plants.

In both HR and NHEJ, recombination could theoretically occur randomly along the length of the chromosomes. However, in plants, as in other organisms, the

presence of hot spots for recombination breakpoints has long been known (reviewed in Schnable et al., 1998; Lichten and Goldman, 1995). In many plant species, it has now been well established that large differences in the recombination rate over a large portion of the genome are present. In wheat, for example, genes are grouped into close clusters, and gene-rich regions are hot spots for recombination (Gill et al., 1996). Generally, centromeres, which are heterochromatin rich, display a very low level of recombination rate compared to telomeres, in which genes are preferentially located (Barakat et al., 2000; Faris et al., 2000; Gill et al., 1996; Tanksley et al., 1992).

The fact that genes per se represent hot spots for recombination is supported by several pieces of evidence. Good examples are represented by the maize anthocyaninless1-shrunken2 (*a1-sh2*) interval (Xu et al., 1995), the rice *wx* locus (Inukai et al., 2000), and the *Lycopersicon pennellii* apoplastic invertase gene *Lin5* (Fridman et al., 2000), in which the recombination frequency is about 10 times greater than the average for the genome. This bias for gene-rich regions is an important feature of plant DNA recombination, which from an evolutionary perspective, has allowed the emergence of new alleles with novel characteristics within plant populations (reviewed in Schnable et al., 1998; Clegg et al., 1997). As examples, recombination at the *Rp1* locus in maize has generated novel resistance genes to the pathogen *Puccinia sorghi* (Richter et al., 1995) and the recombination hotspot in the tomato *Lin5* gene has created multiple isogenic chimeric alleles that assign the QTL for fruit sugar content to a defined sequence (Fridman et al., 2000).

The mechanisms and the features which explain why genes per se are hot spots for recombination are not well known as yet, but some studies indicate that chromosomal structural characteristics are important. In particular, DSB occurs in nucleosome-free regions, which are susceptible to the attack by nucleases (Faris et al., 2000). Chromatic regions that are gene-rich domains are structurally accessible to the transcription process, and thus likely prone to the action of the recombination machinery.

2.2. Molecular plant breeding

Exogenous DNA can be integrated into plant genomic DNA using a range of techniques defined as ''gene transfer technologies'' (reviewed in Hansen and Wright, 1999). With plants, the two most commonly used methods of DNA delivery are the biolistic or microprojectile bombardment system, and *Agrobacterium*-mediated transformation, the latter being the most widely used system (Birch, 1997).

The biolistic method is based on a physical delivery of DNA-coated gold or tungsten microprojectiles into plant target tissue by acceleration. The mechanism by which DNA-coated accelerated particles are able to deliver DNA into living plant cells without damage is still not clear.

Agrobacterium-mediated transformation exploits the biological ability of this soil-borne bacterium to copy and transfer a specific portion of DNA (termed T-DNA) present on a tumour inducing (Ti) plasmid into the plant cell nucleus, where it can be integrated into chromosomes. The most recent model for T-DNA transfer, supported by numerous experiments and data (see Gelvin, 1998, 2000; Ward and Zambryski, 2001), is based on three schematic steps: initiation, bacterium-to-plant transfer, and nucleus targeting; these processes involve the bacterial *vir* proteins and some plant factors. T-DNA integrates into the plant genome in the absence of any homology with plant DNA sequences through the process of illegitimate recombination or NHEJ (Gheysen et al., 1991; Ohba et al., 1995; Mayerhofer et al. 1991; reviewed in Gelvin 1998, 2000). As explained above, NHEJ is the predominant plant recombination mechanism for DNA integration, which requires a DSB mechanism (Britt, 1996). The importance of DSB repair for the insertion of exogenous DNA has been clearly demonstrated by Salomon and Puchta (1998).

2.2.1. Transgene integration in plants

Insertion of transgenes into chromosomal DNA can result in either single copy or repeated and multiple insertions (Pawlowski and Somers, 1996). Multiple insertions can occur in linked or unlinked sites (Laufs et al., 1999; Koncz et al., 1989).

In addition to ''perfect'' integrations, several rearrangements of the transgene construct and of the target site can occur. The presence of rearrangements is often associated with instability in transgene expression (Fladung, 1999; Kumar and Fladung, 2000). *Agrobacterium*-mediated integration can result in complex integration patterns, including directed and inverted repeats (Krizkova and Hrouda, 1998). Filler DNA of plant origin is often found between T-DNA repeats, but direct T-DNA fusion is also observed (Kumar and Fladung, 2000). At the site of insertion in the plant DNA chromosomal rearrangements linked to T-DNA insertion are also possible. These include inversion (Laufs et al., 1999) and translocations (Castle et al., 1993). The presence of 'backbone' sequences, i.e. sequences belonging to bacterial plasmid outside the left and right borders, is also often observed (De Buck et al., 2000). T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation (Kononov et al., 1997). The genomic context of inserted transgenes has been investigated in some plants. Sequencing of the flanking regions of T-DNA insertions in tobacco has

highlighted the recurring presence of motifs that flank transgenes (Matzke and Matzke, 1998). These include:

- AT rich sequences characteristic of matrix-attachment regions (MARs)
- Microsatellite sequences
- Retroelements
- Tandem repeats.

In contrast to other detected sequences, MARs have been associated with the stability of transgene expression (Iglesias et al., 1997). Furthermore, AT rich regions in *Arabidopsis* resembling motifs characteristic of a scaffold/matrix-attachment region (S/MAR) have been proposed as the preferred sites for transgenes introduced by particle bombardment (Sawasaki et al., 1998). These results suggest that the recombination process of T-DNA and "naked" DNA delivered by the biolistic method probably uses the same plant machinery. This hypothesis is strengthened by results from direct DNA transfer experiments in rice which revealed that in the absence of protein factors involved in exogenous DNA transfer through *Agrobacterium*, the qualitative and/or quantitative efficiency of transformation events is not compromised (Kohli et al., 1998)

2.2.2. Distribution of transgene integration sites in transgenic plants

The previous paragraph emphasised that transgene integration occurs in plants through illegitimate recombination mechanisms, in which no homology is required. Thus, it is not surprising that there is no preference for specific sequences in the genome for the integration process. Currently, it is inherently impossible to predict the fate and the site of the integration of a particular transgene construct into the plant genome, given its nucleotide sequence.

Several genetic mapping studies have revealed that transgene integration occurs throughout the entire genome and that there is no chromosomal preference for gene integration in several plant species, e.g. tomato (Thomas et al., 1994) and potato (Jacobs et al., 1995). However, transgenes and, in particular, T-DNA containing transgenes for which more data are available have preferences for gene rich-regions. The fact that T-DNA integration into genes can cause mutations due to the loss of gene function was soon recognised (Koncz et al., 1992). The discovery of phenotypic mutants led to the possibility of developing a new strategy of gene isolation, based on T-DNA tagging. This primary strategy, along with some of its variants, (reviewed in Walbot, 1992; Koncz et al., 1992; Weigel et al., 2000; Matsuhara et al., 2000) is today widely implemented in order to isolate genes based on the observation of phenotypic mutants and co-segregation of the mutations with the T-DNA.

The preference of T-DNA for gene-rich regions is supported by several pieces of evidence and is not surprising, since T-DNA integration shares the common pathway of DSB repair with the plant natural recombination mechanism, in which genes constitute recombination hot spots. The use of a promoterless *nptII* gene led to the conclusion that the integration of T-DNA shows a bias (greater than 70%) for transcriptionally active chromatin domains in tobacco and *Arabidopsis* (Koncz et al., 1989; Herman et al., 1990). Further support on gene-rich region preference of T-DNA comes from the correlation between the pattern of integration of transgene in Arabidopsis and rice and the different gene distribution in the genome of the two (Barakat et al., 2000).

Transgene chromosomal location and structure can be detected by cytological methods such as GISH (genomic *in situ* hybridization) and FISH (fluorescence *in situ* hybridisation) (Iglesias et al, 1997; Pedersen et al., 1997) and by direct sequencing of flanking DNA after rescue of clones carrying transgene/genomic DNA junctions (Babiychuk et al., 1997; Fladung, 1999; Krizkova and Hrouda, 1998; Liu et al., 1995; Parinov et al., 1999; Sawasaki et al., 1998; Spertini et al., 1999; Thomas et al., 1994; Winkler et al., 1998).

Localising the exact integration site and its structure also has a strong relevance for the identification of stably expressed genes. In fact, transgenic loci with unstable expression or epistatic silencing actions are highly correlated with complex integration patterns (Fladung, 1999; Laufs et al., 1999; Matzke and Matzke, 1998). Furthermore, the transgene locus and its surrounding regions are important for stable expression of the new gene (Gelvin, 1998; Iglesias et al., 1997; Sawasaki et al., 1998). Stable inserts with no binary vector sequences have been preferentially located to the chromosome telomers in tobacco (Iglesias et al., 1997). Analogously, in monocot species (Pedersen et al., 1997) and petunia (ten Hoopen et al., 1999), transgenes are preferentially integrated into the distal part of the chromosome arms, a region particularly rich in genes (Barakat et al., 2000; Faris et al., 2000; Gill et al., 1996; Tanksley et al., 1992).

2.3. Relevance for safety assessment

A possible consequence of the random integration of transgenes in the plant DNA is the disruption of endogenous gene function due to the insertional mutagenesis process. Given this scenario, two pertinent questions can arise for safety evaluation:

1. Is transgene integration in plant chromosomes any more likely to result in DNA disruption than the natural recombination mechanisms?
2. Is there an increased chance of a transgene being integrated into active gene-rich regions compared to other chromosomal locations?

Evidence supports the random integration of transgenes into plant genomic DNA by the NHEJ process that repairs DNA by a DSB mechanism. This process is identical to the preferred recombination mechanism that occurs in plant cells, especially during mitosis. It is evident that transgene integration in plants uses the natural ability of plant cells to exchange genetic material during DNA duplication by NHEJ. Natural NHEJ in plants is an error-prone pathway that introduces new filler DNA into the recombination zone resulting in a DNA patchwork. This is similar to what has been observed for transgene integration sites. The presence of several types of rearrangements in transgene integration is also observed during DSB repair in natural plant recombination mechanisms.

There is good evidence that exogenous genes preferentially integrate into transcriptionally active regions of chromosomes. This feature is being widely used to develop specific T-DNA tagging technologies in plants. DNA analysis of transgene flanking regions in potato shows that 50% of all insertions are localised in low copy DNA (Conner and Jacobs, 1999). Given that DSB is the preferred mechanism for DNA integration in plants, the data reflect the fact that genes per se and low copy DNA are hot spots for natural recombination. In this respect, transgene DNA apparently behaves no differently from endogenous plant DNA.

While the recombination mechanism provides plants with a "natural" means to develop new genetic variability, such modifications could be a source of unintended effects both in classical and modern breeding approaches. Since genes are hot spots for recombination, the production of deletions and filler DNA could result in changes to gene sequences, leading to gene disruption and/or the production of new proteins in plants. One of the advantages of genetic engineering vs. classical plant breeding is the inherently higher accuracy of the technology and the fact that molecular approaches can be used to provide essential information on what genetic elements have been inserted into the transgenic plant and the regions into which they have been inserted. This is not possible with classical breeding approaches.

3. Unintended effects in conventional breeding

As discussed previously, the occurrence of unintended effects is not a phenomenon specific to genetic engineering. In classical breeding programmes, extensive backcrossing procedures are applied in order to remove unintended effects. Fig. 1 indicates the steps taken to select lines in a traditional plant breeding programme using barley, an in-breeder, as an example.

This process would apply to most cereals. It is possible that with some parents the resulting F1 generation produces no lines with obvious desirable traits. The

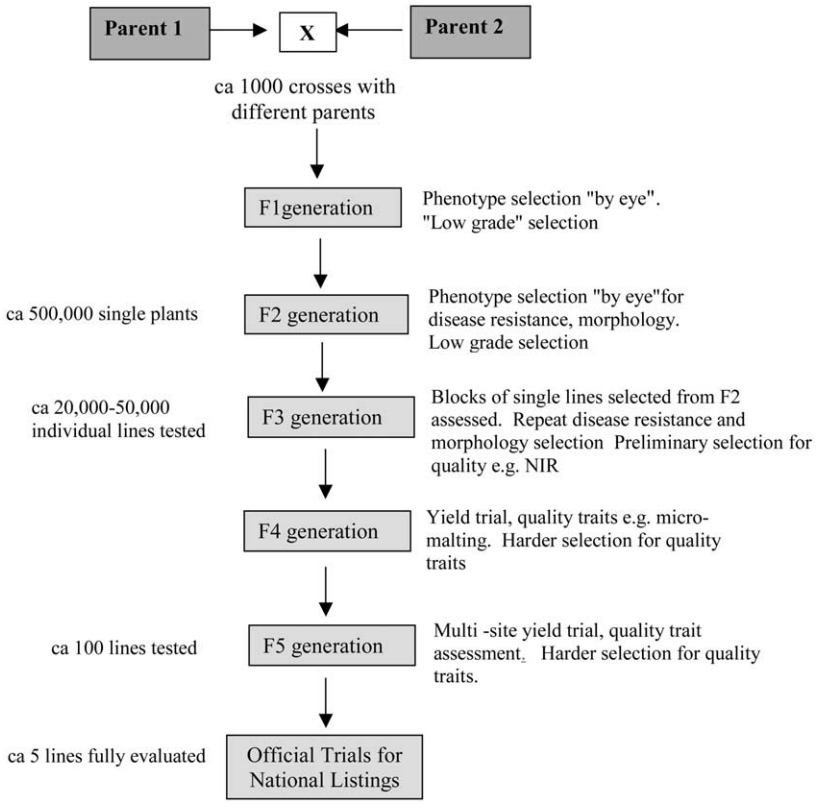

Fig. 1. Line selection in conventional breeding.

selection of starting parental material is therefore paramount and is an iterative process requiring years of experience dealing with diversity in available genetic resources. In the example given, each generation represents one year of selection. In the early stages, lines are selected or discarded using low grade selection procedures based on visible phenotype, extent of disease etc. With regard to the use of statistics in breeding, quantitative analysis only becomes important once lines have entered a field trial phase. The usual approach is to apply analysis, such as "least significant difference", to compare controls with new lines. Controls may be represented by several cultivars, with known characteristics and performance, against which the new lines can be evaluated.

Due to the common practice of selecting favourable lines and discarding those exhibiting unwanted properties in the course of the breeding programme, documented reports on unintended effects in conventional breeding are rare. Some of the few examples actually published in the literature are listed in Table 1.

4. General issues related to comparative analysis

4.1. Introduction

The principle of "substantial equivalence", which has been elaborated in several international consultations since the beginning of the 1990s (FAO/WHO, 1991; WHO, 1995; OECD, 1993), became a key element in the safety assessment of foods derived from genetically modified organisms. The concept is used to identify similarities and differences between the genetically modified (GM) food and a comparator with a history of safe use, which subsequently guide the safety assessment process (FAO/WHO, 2000). There are several critical issues to be considered in the practical application of this concept, which is eventually directed towards the detection of unintended effects.

4.2. Selection of compounds

The result of the described comparative approach heavily depends on the compounds selected for comparison. Compositional data are based on proximate analysis (e.g. protein, fat, moisture, ash) as well as on the analysis of key compounds, such as naturally occurring toxicants and essential nutrients.

Key nutrients or anti-nutrients are those components of a food that may have a substantial impact in the overall diet. They may be major (e.g. fats, proteins, carbohydrates) or minor constituents (e.g. minerals, vitamins). Key toxicants are naturally occurring toxicologically relevant compounds, e.g. glycoalkaloids in potato tubers (Codex Alimentarius Commission, 2003). There are compilations on natural toxicants in food (MAFF, 1996) and attempts have been made to set up networks for the evaluation and assessment of these data (NETTOX, 1998a–i). Other approaches focus on setting up minimum lists of compounds to be analyzed for specific crops in the course of the comparison of GM lines and conventional counterparts (Nordic Council of Ministers, 1998). Consensus documents on key compounds in soybean, canola, potato, sugar beet, maize, and bread wheat have been prepared at international level (OECD, 2001a,b, 2002a,b,c, 2003), while those for rice, cotton, sunflower, forage legumes, barley, and tomato are still in preparation.

4.3. Selection of appropriate comparators

With crops that propagate vegetatively, comparative analyses would include the parental variety used to produce the transgenic lines, whereas with crops that reproduce sexually, comparators are most likely to include appropriate isogenic lines. Since many crops are developed using back-crossing, it is important that the most appropriate controls are used and that analyses do not rely solely on comparisons with the original parental material. In commercial breeding, for example, specific male pollinator lines may be used to produce the final cultivar. Knowledge of the "natural" variation in metabolite composition for the species in question will be extremely important in the risk assessment process. A scenario can be envisaged whereby a GM line may differ significantly from its non-GM parent in the content of specific nutrients or antinutritional factors, yet compositionally the GM line may fall well within the published natural variation for the pertinent compounds in cultivars with a history of safe use. Ideally,

Table 1
Reports on unintended effects resulting from conventional breeding

Host plant/trait	Unintended effect	Reference
Barley Powdery mildew resistance	Low yield	Thomas et al. (1998)
Celery Pest resistance	High furanocoumarins content	Beier (1990)
Corn High lysine content	Low yield	Villegas et al. (1992)
Potato Pest resistance	Low yield High glycoalkaloid content	Harvey et al. (1985)
Squash/*Zucchini* Pest resistance	High cucurbitacin content	Coulston and Kolbye (1990)

the comparator with a known history of safe use should be the isogenic parental line. However, for many commercial food crops, F1 hybrid varieties are used. Compositional analysis must then be performed on the advanced (backcrossed) breeding material and the conventional counterpart.

4.4. Field trials

Protocols for field trials performed with genetically modified and control crops must be specified and documented with respect to:

1. number of locations, growing seasons, geographical spreading, and replicates,
2. statistical models for analysis, confidence intervals,
3. baseline used for consideration of natural variations.

Guidelines being elaborated internationally emphasise that the location of trial sites should be representative of the range of environmental conditions under which the plant varieties are expected to be grown. The number of trial sites should be sufficient to allow accurate assessment of compositional characteristics over this range and trials should be conducted over a sufficient number of generations to allow adequate exposure to the variety of conditions encountered in nature (Codex Alimentarius Commission, 2003).

Other regulatory bodies have been more specific in their requirements, e.g. minimum number of sites (6), minimum number of years (2), and minimum number of replicates (3) have been recommended (ACNFP, 1999).

4.5. Statistical significance

Experimental design should be consistent with the intended method of statistical analysis. Field trial data should be treated statistically by applying an appropriate analysis of variance. A completely randomised design, for example, may allow for analysis of variance of normally distributed data and thus show whether the experimental factors (location, year, climatic conditions, plant variety) interact with each other. The confidence intervals used for statistical analysis (normally 95%) have to be specified.

5. Targeted approaches

5.1. Demonstration of "substantial equivalence" by investigation of defined constituents

The comparison of the chemical composition of the genetically modified plant to that of a traditionally obtained counterpart has been a key element in the safety assessment of genetically modified crops. Such a comparative approach will reveal similarities as well as differences between the transgenic crop and the selected comparator and will thus give information on the "status of equivalence" (König et al., 2004).

This concept has been applied in the pre-market assessment of the first generation of genetically modified crops. Comprehensive analysis demonstrated that apart from the intended presence of the gene expression products (e.g. the Bt-protein in insect-resistant corn) or specific metabolites (e.g. laurate in canola), there were no compositional differences between GM and conventionally bred plants which would go beyond the natural variability. Examples are listed in Table 2.

One of the most critical points of this procedure is the selection of compounds to be analyzed. "Targeted approaches" are based on the qualitative/quantitative analysis of defined constituents. As shown in Table 2, the investigations included proximate analysis as well as analysis of major (e.g. amino acids, fatty acids) and minor constituents (e.g. minerals). Special emphasis has been placed on naturally occurring toxicants (e.g. glycoalkaloids) and essential nutrients (e.g. vitamins) or anti-nutrients (e.g. trypsin inhibitors). Analyses also include compounds that, from current biological and biochemical knowledge, might be predicted to change in response to the specific genetic modification. For example, the analysis of phenylalanine, tyrosine, and tryptophan levels is relevant to the determination of possible "predictable unintended effects" in Roundup-Ready soybeans, since the EPSPS gene confers Roundup resistance through modifications to the synthesis of aromatic amino acids.

5.2. Demonstration of unintended effects by phenotypic selection and by investigation of defined constituents

During the development of transgenic plant varieties and for any given trait(s), a large number of transformants/clones that do not perform up to the required expectations will be discarded through assessment in the laboratory, glasshouse, and small scale field trials. In all cases, new cultivars produced by genetic engineering are extensively tested and screened prior to commercial release. Evaluations of plant vigour, growth habit, yield, crop quality, and insect and disease susceptibility would be performed. This would result in the elimination of major unintended effects which are more easily screened for, leaving the more subtle differences to deal with. The biological relevance of unintended effects resulting from genetic modification refers to the implications of these effects for the agronomic performance of the plant. This does not necessarily mean that there will be any relevance to safety for human health and the environment.

Several examples exist that show that, following genetic modification events, unintended effects can have an impact on potential agronomic performance (Table 3). For example, the capacity for fructan biosynthesis, when introduced into the cytoplasm of potato tuber cells, results in transformants with impaired carbohydrate transport and perturbed tuber development (Dueck et al., 1998; Turk and Smeekens, 1999). Another example results from the overexpression of phosphatidyl serine-synthase in wheat with the aim of increasing resistance to aluminium toxicity; necrotic lesions developed in these plants (Delhaize et al., 1999). Such phenotypes are obviously detrimental to any further commercial development of the transgenic lines in question.

Other types of unintended effects with biological relevance have only been detected after growing plants under specific environmental conditions. For example,

Table 2
Demonstration of "substantial equivalence" by investigation of defined constituents

Plant	Genetic modification	Parameters	Reference
Canola	High laurate	• Amino acids fatty acids • Erucic acid glucosinolates	Redenbaugh et al. (1995)
Corn	Insect resistance (Cry I A(b))	• Proximate analysis • Amino acids fatty acids • Calcium, phosphorus	Sanders et al. (1998)
Cotton	Herbicide tolerance (glyphosate)	• Proximate analysis • Amino acids fatty acids • Gossypol • α-Tocopherol • Aflatoxins	Nida et al. (1996)
Cotton	Herbicide tolerance (bromoxynil)	• Amino acids fatty acids • Gossypol cyclopropenoid fatty acids	Redenbaugh et al. (1995)
Soybean	Herbicide tolerance (Roundup Ready glyphosate resistance)	• Proximate analysis • Fatty acids • Amino acids (phenylalanine, tyrosine, tryptophan) • Trypsin inhibitors lectins phytoestrogens urease stachyose, raffinose phytate	Fuchs et al. (1996)
Potato	Herbicide tolerance (chlorsulfuron)	• Proximate analysis • Amino acids valine, leucine, isoleucine	Monro et al. (1993)
Potato	Insect resistance (Cry III A)	• Proximate analysis • Glycoalkaloids • Vitamins B_6, C folic acid • Potassium	Lavrik et al. (1995)
Tomato	Insect resistance (Cry I A(b))	• Proximate analysis • α-Tomatine • Vitamin C • Minerals (Ca, P, Na, K, Mg, Fe, Cl)	Noteborn et al. (1995)
Tomato	Anti-sense polygalacturonase (Flavr Savr)	• Protein • Tomatine • Vitamins A, B_6, C thiamine, riboflavin • Minerals (Ca, Mg, P, Na)	Redenbaugh et al. (1995)

Table 3
Demonstration of unintended effects by phenotype selection or investigation of defined constituents

Host plant	Trait	Unintended effect	Reference
Potato	Expression of bacterial levansucrase	Adverse tuber tissue perturbations Impaired carbohydrate transport in the phloem	Turk and Smeekens (1999); Dueck et al. (1998)
Wheat	Expression of phosphatidyl serine synthase	Necrotic lesions	Delhaize et al. (1999)
Soybean	Expression of glyphosate (EPSPS) resistance	Splitting stems and yield reduction (up to 40%) at high soil temperatures (45 °C) Higher lignin content (20%) at normal soil temperatures (20 °C)	Gertz et al. (1999)
Wheat	Expression of glucose oxidase	Phytotoxicity	Murray et al. (1999)
Rice	Expression of soybean glycinin	Increased Vit. B6-content (+50%)	Momma et al. (1999)
Potato	Expression of soybean glycinin	Increased glycoalkaloid content (+16-88%)	Hashimoto et al. (1999a); Hashimoto et al. (1999b)
Potato	Expression of yeast invertase	Reduced glycoalkaloid content (−37-48%)	Engel et al. (1998)
Canola	Overexpression of phytoene-synthase	Multiple metabolic changes (tocopherol, chlorophyll, fatty acids, phytoene)	Shewmaker et al. (1999)
Rice	Expression of carotenoid biosynthetic pathway	Formation of unexpected carotenoid derivatives (beta-carotene, lutein, zeaxanthin)	Ye et al. (2000)

glyphosate resistant transgenic soybean has shown stem splitting and yield reduction (up to 40%) under high soil temperatures (45 °C), and a 20% higher lignin content at normal temperatures (25 °C) (Gertz et al, 1999). It has been hypothesised that the expression of the EPSPS gene, which confers glyphosate resistance, might alter the distribution of metabolites in the shikimate pathway, which is central to the synthesis of aromatic amino acids, lignin, some vitamins, and other secondary metabolites. However, expression of the EPSPS gene does not appear to modify amino acid biosynthesis. Some types of biologically relevant unintended effects may therefore only become manifest under specific environmental regimes.

The use of targeted approaches for comparative analysis of defined constituents is clearly extremely useful in identifying those differences between parental and GM lines that go beyond the intended effects of the genetic modification. For example, the expression of a nutritionally valuable protein (soybean glycinin) in rice and potatoes increased the content of a nutritionally important compound (vitamin B_6) as well as that of toxicologically relevant compounds (glycoalkaloids). These unintended effects were neither predicted nor explained by the authors (Hashimoto et al., 1999a,b). In another example, expression of a yeast invertase gene in potato tubers resulted not only in an altered carbohydrate metabolism (predictable) but also in a reduced glycoalkaloid content (Engel et al., 1998). Detailed investigations revealed that the lower glycoalkaloid content was associated with differences in plant maturity between the transgenic lines and controls. This confirms that links exist between metabolic and developmental processes.

Another example of how unintended effects have been identified by targeted approaches comes from work on GM canola (Shewmaker et al., 1999). Overexpression of an enzyme in isoprenoid metabolism (phytoene-synthase) resulted not only in an altered level of the metabolite downstream of the target of genetic modification (phytoene), but also in changes in the levels of compounds (tocopherols, chlorophyll) whose biosynthesis is linked to the modified pathway. Based on our knowledge of biochemical pathways, this is a good example of a predictable unintended effect. Similarly, with regard to the so-called "Golden Rice", introduction of phytoene synthase and phytoene desaturase into the endosperm of a single transformant was expected to induce the formation of lycopene. Surprisingly, α-carotene, β-carotene, lutein, and zeaxanthin were synthesized instead of lycopene. This was explained by a further conversion of the newly introduced substrate lycopene by lycopene cyclases inherently present in normal rice endosperm (Ye et al., 2000).

5.3. Limitations of targeted approaches

The targeted approach for analysis of single compounds with special focus on important nutrients and critical toxicants has been widely accepted by international bodies as part of the concept of "substantial equivalence". It has been successfully applied to the safety assessment of the first generation of GM crops. Indeed, genetically modified plants belong to the best

analysed foods we know. Whilst the selection of compounds for analysis is the first step in the targeted approach, there are no generally accepted and harmonised guidelines that define the full extent of analyses required to fulfil statuary risk assessment procedures. Furthermore, some critics consider the targeted approach to be biased and focused on known compounds and expected/predictable changes (Millstone et al., 1999). Whilst the newer profiling technologies (discussed later) will vastly increase the numbers of compounds analysed (thereby reducing the level of uncertainty), targeted approaches will still be required to address effects on the key nutritional and antinutritional components of crops.

The next generation of GM crops is likely to include those with improved nutritional properties and more far reaching effects on metabolic processes, i.e. an increased complexity of the genetic modification. The introduction of new biosynthetic pathways in plants and genetic modifications targeting key enzymes in primary and secondary metabolism could result in metabolic perturbations not explicable in terms of our current knowledge of plant biology and metabolic pathway integration and interconnections. This could lead to the occurrence of unpredictable unintended effects and in altered levels of metabolites not revealed by a targeted approach.

6. Non-targeted approaches

6.1. Introduction

The major plant genome sequencing projects (*Arabidopsis*, rice) provide information that does not change with time or circumstance and thus can be considered as a huge reference book. Microarrays and proteomics, however, measure components that are subject to constant change. These changes may be cyclical, developmental, or responses to changes in the environment. The two approaches provide comprehensive 'snapshots' of the cell, tissue, or organ at the levels of messenger RNA (the expressed genes or *transcriptome*) and protein (*proteome*), respectively, and will eventually reveal how the components function and interact as working parts of the cell machinery. The enzymes, which themselves may be up- or down-regulated by genetic modification, control the formation and transformation of cellular metabolites such as sugars, acids, fats, and many minor compounds that, in food, may be essential, beneficial, or even harmful to health. Proteomics is becoming an important tool in biological sciences and will contribute greatly to the understanding of gene function, particularly as it is known that mRNA levels do not correlate well with the protein levels (Gygi et al., 1999). The entire collection of metabolites in the cell is called the *metabolome* and the science of measuring it (metabolomics) forms the third of the new '-omics' technologies. It uses established methods of analytical chemistry, but in a novel way by seeking to measure 'everything at once' rather than isolating single compounds of interest. The results again reflect a changeable composition measured at a specific moment in time.

The development of these three technologies allows us to think in terms of whole biological systems, leading to understanding and prediction of how a change in one part of the system will affect other parts and ultimately the whole. It is worth considering the scale of the undertaking, however, on each of the three levels. The genomes of yeast and *Arabidopsis* contain about 6000 and 25,000 genes, respectively, with potential to produce the corresponding numbers of proteins. The effective amount of variation among the proteins is even greater because of post-translational modifications. Proteomics techniques are required to detect these modifications. It has been stated that the number of metabolites is an order of magnitude lower than the number of genes, i.e. the total for yeast would be about 600. On the other hand, it has been estimated that about 80,000 different metabolites have been identified across the plant world. Current technology allows all the genes of yeast to be represented on a single chip. For plants with their much larger genomes, a range of microarrays has to be prepared, tailored to specific purposes. For example, arrays with up to 10,000 elements are being constructed for different potato tissues to focus on the patterns of gene expression connected with disease resistance. Truly comprehensive coverage is harder to obtain in the cases of the proteome and metabolome. If the figures above are accepted, the task would appear to be less daunting for the metabolites than for the proteins. There is much on-going research to improve these techniques since proteins (potential allergens) and metabolites (potential toxicants) are the actual functional molecules of interest. The problems are mostly attributable to questions of separation techniques, resolution, sensitivity, and concentration range, as discussed in the following chapters.

Despite the present lack of true comprehensiveness, it is not in doubt that any one of the techniques is capable of generating volumes of data that are greater than anything that biologists have faced before. In applying these techniques to food crops, it is essential that multiple control samples are studied in order to assess inherent environmental variability, thus enabling the experimental samples to be evaluated in the context of the conventional crop as a whole. The need for multivariate analysis (principal component analysis, clustering techniques etc.) of the data has already been recognised and demonstrated with model organisms, where the external sources of variation can be much better controlled, and the same methods will clearly be mandatory for field crops. There are some early examples

of this approach being taken with microarrays and metabolomics; however, less appears to have been carried out in proteomics, probably because of experimental problems involved in matching the data between different samples. This is certainly an area that is ripe for development.

6.2. Functional genomics

6.2.1. Introduction

The term genomics is generally used for the combined activities to analyse the nucleotide sequence of entire genomes, or larger parts thereof. The term functional genomics, on the other hand, relates to the functionality of these sequences, i.e. of the transcribed genes (transcriptome) and the related regulatory elements, and thus provides the opportunity for gaining further insight into the (complex) metabolic relationships within an organism. In practice, functional genomics refers to the study of direct expression products of genes, the mRNAs converted into more stable cDNAs, using microarray technology. This technology miniaturises the scale of well-established molecular biological methodologies, such as hybridisation of a probe to a target sample. The advantage of the technology is the use of a large variety of individual identified probes in a single experiment by arraying the probes to a solid surface. Subsequently, all probes are hybridised simultaneously to the labelled sample under investigation. In this way, gene expression profiles can be established of individual or mixed tissue samples of transgenic plant varieties and compared to the same type of samples from control plants that are either unmodified or have integrated the 'empty' vector sequence, i.e. the vector sequence without the target sequence. Detected differences in gene expression profiles may be an indication for unintended side effects of the genetic modification and provide information for further investigations on toxicological relevance .

6.2.2. Principles

6.2.2.1. mRNA fingerprinting. Conventional methods for analysis of differential gene expression include Northern blotting (Alwine et al., 1977), S1 nuclease protection (Berk and Sharp, 1977), differential display (Liang and Pardee, 1992), arbitrarily primed PCR (Welsh et al., 1992), and serial analysis of gene expression (SAGE) (Velculescu et al., 1995). Northern blotting is the method most often used to confirm the expression of the introduced gene sequence. Differential display has been applied in order to detect altered gene expression in genetically modified plant material compared to the non-modified parent line (Kok et al., 1998, 2001). Based on the original concept of Liang and Pardee (1992), specific subsets of the mRNA population of both a genetically modified tomato plant and the unmodified parent line were amplified. The results demonstrated that a targeted primer design will result in an amplified mRNA subset that includes the transcription products of the introduced gene sequence. It was, however, also shown that this method is very labour intensive and not well suited for routine detection of altered gene expression as part of a risk assessment strategy.

6.2.2.2. DNA microarray technology. As stated, microarrays have the advantage of the parallel screening of a large number of (identified) gene sequences for differences in gene expression in tissues of variable origin (Van Hal et al., 2000). The immobilised probes on microarrays for gene expression will either be cDNA sequences or oligonucleotides. The probes can either be synthesised on a solid support (oligonucleotides, especially genechips from Affymetrix) or synthesised prior to spotting in array format (both oligonucleotides and cDNA microarrays, spotting equipment available from various companies [among others Cartesian Technologies, Biorobotics, GeneMachines etc.]). The approach of in-situ synthesis has resulted in very high density arrays of oligonucleotides (up to 250.000 spots/cm^2) (Lockhart and Winzeler, 2000). The oligonucleotides are usually short (20–25 bp) and this requires very strict hybridisation conditions and the confirmation of the resulting signal by additional probes for the same cDNA (leading to a significant reduction of sequences that can be identified, but still amounting to approximately 10.000 genes/cm^2). The approach whereby available sequences are spotted in array format can also result in several thousand spots per cm^2 and has advantages with respect to costs and flexibility of the approach. This approach has therefore found numerous applications in fundamental research and allows cDNAs or sequences that are well-characterised and functionally identified to be used. Furthermore, ESTs that are known to be associated to a specific physiological, developmental, or environmental condition can also be utilised in order to investigate metabolic pathways related to this condition. Preferably, the spotted sequences have comparable lengths that allow for specific hybridisation under the selected hybridisation conditions. The next step is to isolate all mRNAs from the samples of interest and label them directly by incorporation of a labelled nucleotide during a reverse transcription reaction or indirectly by attachment of reactive molecules to the (backbone of the) RNA molecules. This latter step is then followed by a second step after hybridisation, when a fluorescently labelled second substance is linked to the reactive side of the bound molecule. In both array systems, it is possible to quantify (relative) gene expression on the basis of the fluorescent signal of individual spots and compare gene expression profiles in different samples under investigation.

Other types of arrays have been developed in the last few years. Examples include the electroarray system,

where the labelled negative DNA solution can be directed to individual, positively charged spots (Nanogen, 2003), or suspension arrays, where the probes are bound to uniquely coded polystrene beads, as opposed to gene-specific PCR products dotted or synthesised onto the surface of a solid support (Nolan and Sklar, 2002).

6.2.3. Applications

Application of informative array systems can very rapidly generate large amounts of data on gene expression under different developmental, physiological, and environmental conditions. Genes that are differentially expressed under a specific set of conditions can be studied further by either additional microarray experiments or other targeted gene expression studies to increase our knowledge of metabolic pathways and physiological interactions. In relation to food safety strategies of plant products, gene expression studies should focus on metabolic routes leading to the formation of anti-nutrients, including natural toxins, as well as on the metabolism of positive nutritional factors (micro- and macronutrients) to monitor for possible unintended changes. In addition, other cDNAs can be spotted on the array to screen for alterations in gene expression in other metabolic systems of the plant that may be of relevance to the safety or nutritional value of the plant.

There are no published examples available to date on the application of genomics (DNA microarray technology) to the detection of unintended effects in GM products. Within the EU Fifth Framework project GMOCARE (GMOCARE, 2003), the potential for analysing differential gene expression using DNA microarrays as a means of contributing to future improved food safety evaluation strategies is currently being assessed.

6.2.4. Possibilities and limitations

Microarray technology may prove to be a very straightforward technique to detect possible changes in the metabolism of a genetically modified organism when compared to its nearest comparator and more informative than a classical component analysis to reveal changes in the 'black box' of the physiology of the plant. The most important advantages are:

- potential for detection of changes in many different routes in a single hybridisation experiment, including any metabolic pathway that may have been activated by the genetic modification, but is not usually switched on in the tissue under investigation
- use of information available in relation to spotted sequences (the annotation of the sequences with their functional identification or the type of library they originate from that may give further clues as to their expression under a specific physiological or developmental condition) to directly investigate the relevance of the detected changes for the safety of the derived food products (Brown and Botstein, 1999).

Functional genomics is thus likely to provide insight into pathways that are relevant for the safety of food crops and also into the natural variation in the expression of genes, e.g. under different environmental conditions (Sommerville and Sommerville, 1999). Furthermore, this approach enables expression profiles of individual tissues to be carried out, e.g. in specific tissues that are consumed.

DNA microarrays can also be used in alternative strategies for the analysis of effects of environmental stress and different climatological and soil conditions. In principle, these conditions can be mimicked under strict controlled conditions in growth chambers. Exposure studies can be performed under extreme, but realistic, conditions and the effects measured using functional genomics approaches. The scale and resolution of DNA microarrays will facilitate the detection of alterations in gene expression and, if the relevant pathways are known, in possible consequences for food safety. The use of growth chambers, facilitated by the use of large- scale analysis technology, rather than field conditions, enables analyses to be done in months rather than years.

Before applying the microarray approach for the safety evaluation of GMO-derived products, it is important to decide upon the type of information that is required. The most decisive factor is the number and type of sequences to be spotted, since these will determine the maximum amount of information that can be obtained after analysis of the hybridisation results. This means that as many sequences as possible should be spotted from all relevant tissues of the organism under investigation. The spotted sequences should preferably be non-redundant, with individual sequences represented only once and (parts of) individual genes represented only in very limited numbers, but whereby they may still be informative (Franssen-van Hal et al., 2002). Preferably, the array should also contain sequences from metabolic routes that are not usually expressed in the tissue under investigation, but that may have been activated due to the genetic modification. In practice, this may not yet be feasible and the best arrays will consist of an informative selection of the different sets of sequences mentioned. It should also be kept in mind that the annotations of ESTs in genomic databases are not always reliable, and thus the functional identity may have to be confirmed before the toxicological relevance of detected differences can be discussed (Somerville and Sommerville, 1999). Similarly, hybridisation experiments should be performed with RNA populations

(either totRNAs or more purified mRNAs) that are isolated from the most relevant individual tissues or a mixture of these tissues. It is clear that sampling is therefore a crucial step in the microarray analysis as differences in the RNA of sampled tissues that are hybridised to the array are likely to be a much larger source of variation compared to possible differences in gene expression due to a genetic modification. With the increase of knowledge on the natural variation in the expression levels in individual tissues, it will become easier to use reference sequences that will also control for correct sampling procedures. However, for the time being, detected differences will have to be confirmed by other methods to exclude this obvious source of difference in gene expression profiles. The availability of sufficient RNA from relevant, individual tissues may also form a problem. To date, different methods have been developed to amplify mRNA from a specific tissue or even a single cell, but it is likely that differences will occur in the relative abundance of the individual mRNAs before and after the amplification step (L. Pellis, personal communications).

Data analysis of the hybridisation results is the major challenge related to the microarray approach. Informative results can only be obtained by rigorous experimental set-up, which has been designed to answer specific questions, in combination with appropriate software to carry out data analysis. Depending on the question being asked, the number of necessary hybridisations may vary in a microarray experiment. For the detection of differences in gene expression between a GMO-derived plant and its traditional counterpart, it will be necessary to decide:

- how many tissues are relevant,
- whether the relevant tissues need to be analysed individually or can be analysed simultaneously in a mixture,
- how many controls should be included to correct for differences in natural variation in the sampled tissues that are not related to the genetic modification(s). Both traditional lines (to control for the sampling procedure), control lines related to the genetic modification (to distinguish, for example, effects related to the place of insertion from [secondary] effects of the inserted gene sequences), etc. should be used.

Since hybridisation experiments result in thousands of data points, software packages to handle this large amount of data are thus essential. Fortunately, the availability of such software is rapidly improving, making it easier to compare large sets of data in order to find similar changes in gene expression profiles that may be related to a genetic modification. Pertinent in this respect are developments in data handling and storage, which should, in the end, make it possible not only to use any results from earlier experiments, but also related observations or hypotheses, to help answer new questions. There are already examples of databases that combine sequence information with information on genetics, gene expression, homology, regulation, function, interactions and phenotype information (Lockhart and Winzeler, 2000; Gerstein, 2000; Bassett et al., 1999). It is feasible that this type of database will in the future be set up for all important organisms in relation to food supply. Such databases will enable us to better understand observed differences in gene expression and related phenotypic alterations, and hence the subsequent consequences for food safety.

6.3. Proteomics

6.3.1. Principles

In essence, proteomics is an amalgam of three technologies, high-resolution 2-dimensional electrophoresis (2-DE) to separate the proteins present in a tissue, image analysis to aid comparisons of separations, and mass spectrometry (MS) to determine the identity of the proteins of interest.

6.3.1.1. Two-dimensional gel electrophoresis. Currently, separation is largely based on two-dimensional protein gel electrophoresis, in which separation in the first dimension is by charge and in the second dimension by molecular weight. Wide-range pH gradients (3-12) allow the separation of highly acidic or basic proteins and an overview of total cellular extracts (Görg et al., 1999). These overviews are, however, dominated by the most abundant protein species. Narrow pH gradients of 1-1.5 pH units stretch protein patterns, allowing a more detailed investigation by providing enhanced resolution and aiding in the detection of minor components (Görg et al., 2000; Wildgruber et al., 2000). These ultrazoom gels have been reported to allow the detection of proteins with an abundance as low as 300 copies per cell (Hoving et al., 2000).

Despite the high capacity of 2-DE in protein separation, it may be necessary to divide the proteome of a given sample into subproteomes. Proteins may be extracted from specific sub-cellular compartments, or according to their relative solubilities (Cordwell et al., 2000). Subcellular fractions not only display qualitative and quantitative differences in their proteins, but also in the ease of protein isolation. Different methods may have to be developed for cytosolic proteins and soluble cytoskeletal elements, membrane (Santoni et al., 1998, 1999) and organellar (Rotig and Chauveau, 1987; Peltier et al., 2000) proteins, nuclear membrane and soluble nuclear proteins, and detergent-resistant cytoskeletal filaments with nuclear matrix proteins. Examples of 'functional proteomics' approaches include affinity

purification to obtain binding partners of a protein of interest (calcium binding proteins, glycoproteins, hydrophobic or membrane proteins) and purification of a multi-protein complex with a defined function (Andersen and Mann, 2000; Lopez et al., 2000; Santoni et al., 2000).

6.3.1.2. Detection and quantification of proteins. Quantitative comparison of two 2-D gels requires linear, uniform, and reproducible detection methods. Coomassie Brilliant Blue (CBB) dyes G and R stain fairly uniformly, but are limited by sensitivity (~100 ng) (Rabilloud, 2000). Silver staining provides low-nanogram range sensitivity and a good contrast. However, the most sensitive silver staining methods may not be compatible with mass spectrometric analysis and only 77% of the silver-stained spots have been shown to have a linear relationship with the total amount of protein present (Costa and Plomion, 1999). The more recently introduced SYPRO fluorescence dyes (Patton, 2000) allow the detection of 1–10 nanograms of protein and responses are linear over three orders of magnitude. This compares favourably with CBB and silver staining where the linear range is only about 40-fold and which may vary from protein to protein (Merril, 1990; Steinberg et al., 1996). A disadvantage of fluorescent dyes is the need for specific detection and image analysis systems. In large-scale proteomic projects, quantification and post-translational modification need to be analysed by 'proteomatic' (i.e. proteome automatic bioinformatic analysis directly from the gel) techniques.

Replicate 2-D gels are never identical and, despite the availability of purpose-designed image analysis programmes, exact matching of spots may be difficult. Difference gel electrophoresis (DIGE) circumvents some of the problems by enabling two samples covalently labeled with different fluorescent dyes to be run on the same gel (Ünlü et al., 1997; Ünlü, 1999). Cyanine-based dyes maintain the pI and mobility of labeled proteins, provide a sensitivity equal to silver staining, and improve comparative accuracy. As only 1–2% of all protein molecules are labeled by the dye, the method does not prevent mass spectrometric analysis. This may be the preferred approach in the future since it eliminates one source of variation but is presently limited by cost.

6.3.1.3. Image analysis. The 2-DE gel image analysis typically consists of several steps. First, the gel image is scanned prior to computer-based processing, in which the gel image is processed to remove background noise and artifacts. The exact positions of individual spots are determined and a quantitative value for each spot is calculated from its shape and intensity. The gels are matched to each other to allow the comparison of identical spots. Finally, the data are analysed to find qualitative and quantitative changes in the protein profiles and to evaluate the quality of the data. Several computer software systems for gel matching, image detection, and analysis are available (Fig. 2). Most software systems can handle large numbers of gels, while some allow combining of replicate gels into so-called "synthetic" or "average" gels that may then be compared to each other. The results of the analysis can be saved in 2-DE databases that link descriptive information (name, molecular weight and pI, expressional data, post-translational modifications etc.) to spots in a representative gel image.

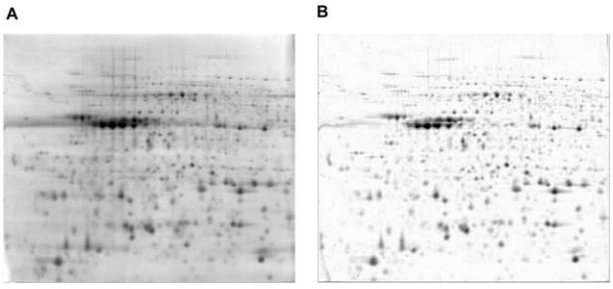

Fig. 2. Protein profile of potato variety Cara (A) immediately after staining and scanning of the gel, and (B) after processing of image (A) with PDQuest image analysis program.

6.3.1.4. Identification of proteins and post-translational modifications. Proteins may be identified by matching the images to existing 2-DE databases, provided that methods and protocols (e.g. protein isolation and electrophoretic run) are standardised. However, only a very limited number of plant protein 2-DE databases are presently publicly available (Table 4).

The basic method used to identify proteins separated by 2-DE is matrix assisted laser desorption ionisation time-of-flight (MALDI-TOF) mass spectrometric analysis (Andersen and Mann, 2000). The detection of post-translational modifications has also become much easier and more sensitive with this method. With MALDI-TOF, the masses of a set of peptides, usually derived from a tryptic digest of a single protein isolated from the gel, may be measured. The resultant peptide mass map may be identified against a protein database generated 1) from experimental data on proteins, or 2) theoretically from genomic sequence data.

For the identification of proteins, peptide mass mapping is not always sufficient. High-resolution (electrospray ionisation) tandem mass spectrometry, however, allows for further fragmentation of selected ion species, thus increasing the chances to identify the protein. Sequencing of peptides from fragment spectra is limited. Methods for de novo sequencing apply derivatization and fragmentation of peptides, facilitating far more selective sequence database searches and definitive protein identification (Bauer et al., 2000; Keough et al., 2000).

Table 4
Plant protein 2-DE databases available via internet

WORLD-2DPAGE: Index to 2-D PAGE databases and services (plants) http://www.expasy.ch/ch2d/2d-index.html
Arabidopsis protein 2D map http://www.expasy.ch/cgi-bin/map2/def?ARABIDOPSIS
The Arabidopsis mitochondrial proteome project (Abteilung Angewandte Genetik, Universität Hannover, Germany) http://www.gartenbau.uni-hannover.de/genetik/AMPP
Plant plasma membrane database (PPMdb) (Gent University, Belgium) Database contains several 2D gels from Arabidopsis (callus, leaves, roots) and identification of several proteins in them and some gels from plasma membrane fraction; Santoni et al 1998, 1999 http://sphinx.rug.ac.be:8080/ppmdb/index.html
The potato mitochondrial proteome (Abteilung Angewandte Genetik, Universität Hannover, Germany) http://www.gartenbau.uni-hannover.de/genetik/Page4.html
Maize Genome Database (INRA, Gif-sur-Yvette, France) http://db_Pub_home/moulon.inra.fr/imgd/
Proteomic database of maritime pine (INRA, Bordeaux, France) Needles, xylem, germinated megagametophytes; Costa et al. 1999, 2000 http://www.pierroton.inra.fr/genetics/2D/
Meta-Database Catalog of 2D gel images found in Web databases-2DWG (all organisms, no plants) Collected by Laboratory of Experimental and Computational Biology (LECB) http://www-lmmb.ncifcrf.gov/2dwgDB/2DWG.html
Arabidopsis protein 2D map http://www.expasy.ch/cgi-bin/map2/def?ARABIDOPSIS The Arabidopsis mitochondrial proteome project (Abteilung Angewandte Genetik, Universität Hannover, Germany) http://www.gartenbau.uni-hannover.de/genetik/AMPP Plant plasma membrane database (PPMdb) (Gent University, Belgium) Database contains several 2D gels from Arabidopsis (callus, leaves, roots) and identification of several proteins in them and some gels from plasma membrane fraction; Santoni et al 1998, 1999? http://sphinx.rug.ac.be:8080/ppmdb/index.html The potato mitochondrial proteome (Abteilung Angewandte Genetik, Universität Hannover, Germany) http://www.gartenbau.uni-hannover.de/genetik/Page4.html Maize Genome Database (INRA, Gif-sur-Yvette, France) http://orsay1.moulon.inra.fr/imgd/ Proteomic database of maritime pine (INRA, Bordeaux, France) Needles, xylem, germinated megagametophytes; Costa et al. 1999, 2000 http://www.pierroton.inra.fr/genetics/2D/ Meta-Database Catalog of 2D gel images found in Web databases-2DWG (all organisms, no plants) Collected by Laboratory of Experimental and Computational Biology (LECB) http://www-lmmb.ncifcrf.gov/2dwgDB/2DWG.html

Some mass spectrometric applications try to avoid 2-DE separation of proteins. Haynes and Yates (2000) reviewed the current state of analytical methods in proteomics. Two recently published methods that offer an alternative approach were presented and discussed, with emphasis on how they can provide information not available via 2-DE. These two methods were the isotope-coded affinity tags (ICAT) approach (Gygi et al. 1999), of particular value for studies of differential expression (Mosely, 2001), and the multidimensional liquid chromatography and tandem mass spectrometry approach as presented by Link et al. (1999). Both techniques represent significant advances in analytical methodology for proteome analysis. Combined with affinity purification, these could be very useful for the identification and quantification of some post-translational modifications.

6.3.2. Applications

The applications of proteomics can be divided into three main areas: (1) identification of proteins, their precursors, and post-translational modifications; (2) 'differential display' proteomics for quantification of the variation in the amounts of proteins; (3) studies of protein-protein interactions. Determination of components of a protein complex or a cellular structure is central in functional analysis and may be the area of greatest promise in proteomics.

Proteomics is potentially useful for the detection and assessment of unintended or unexpected effects in GM crops or crops developed by any other means of plant breeding, although it has yet to be used for this purpose. The potential is illustrated in the following example, closest to, but not actually deploying, proteomics. In low-glutelin rice developed by antisense technology, unintended increases in the levels of other proteins such as prolamins were discovered using SDS-PAGE (Kubo, 2000). This might not be detected by standard nutritional analysis (total protein/amino acid profiles), but could affect nutritional quality and allergenic potential if used as food. In the EU Fifth Framework project GMOCARE, several transgenic potato and tomato lines are being studied for any unintended effects using proteomics.

Proteomics has many applications in crop science and is an important tool in plant breeding. Proteomic tools have been successfully applied for the characterisation of individuals or lines, estimation of genetic variability within and between plant populations, establishment of genetic distances that can be used in phylogenetic studies, and characterization of mutants (Thiellement et al., 1999). 2-DE is an abundant and cheap source of good-quality genetic markers. In the context of genome mapping projects, such markers are physiologically relevant, since they reveal loci whose transcripts are translated in the organ analysed. Improvements in software have

permitted quantification of the proteins, leading to the concept of ''protein quantity loci'', which, combined with the ''quantitative trait loci'' approach, may reveal candidate proteins with a possible role in the metabolism or phenotype under study. Also the effects of, for example, plant protection products, fertilizers and elicitors can be studied at the biochemical level.

Antibodies increase the possibilities for detection and identification of proteins from the proteome immunoblot. This may be particularly valuable for the detection of food/pollen allergens by using patients' immune sera. 2-DE followed by immunoblotting was used by Posch et al. (1997) to characterise and identify the IgE-reactive proteins of *Hevea* latex, which is the main cause of the latex type I allergy. The major latex allergens have been localized on 2-D maps. Microheterogeneity and post-translational modifications can also be detected by 2-DE and protein analysis. By 2-DE blotting using patients' IgE and monoclonal antibodies, Petersen et al. (1997) detected IgE-reactive isoforms and found single amino acid substitutions in different-sized group I grass pollen allergens. When a recombinant group I grass allergen was expressed in *E. coli* and in *Pichia pastoris*, 2-DE immunoblotting demonstrated microheterogeneity in molecular mass and pI. Expressed in *Pichia,* the allergen was hyperglycosylated compared to the natural form. This is an example of the application of proteomics to examine the equivalence of natural and recombinant proteins.

6.3.3. Possibilities and limitations

The advantage of 2-DE over other separation techniques is fractionation with very high resolution. The separated proteins are embedded in a matrix where they can be detected with very high sensitivity. The isolated proteins can be readily extracted for further characterisation, including full sequence analysis. Up to 10,000 individual protein species have been resolved in a single gel, similar in magnitude to the estimated number of expressed proteins in a eukaryotic cell.

The ease of protein extraction varies depending on the source (plant, tissue, cellular compartment, protein structure). Hydrophobic protein species and very large/small and basic/acidic proteins are easily lost. Low-abundant proteins may remain undetected. Reproducibility problems can arise from sample treatment and electrophoresis, although some differential methods allow the test and control samples to be electrophoresed on the same gel. The sensitivity of staining procedures sets some limits to the amount of protein needed for loading, identification, and quantification, while the statistical treatment of 2-DE pattern has not yet been fully resolved.

The comparison of datasets between different laboratories requires a standardised method to be developed for sample isolation and electrophoresis. This is a major challenge. Small variations in this multistep procedure may have major influences on the resulting protein pattern. Therefore, mechanisation and standardisation of gel casting, loading, and running are essential for reproducible results. Staining time and linearity (silver staining *vs.* other methods) need standardisation; any standardisation procedure should include within-laboratory and inter-laboratory validation. Further improvements are also desirable in the following areas: sensitivity, tools for comparative purposes and reduction in sample size required for analysis. In parallel, more background data are needed on natural variability. For studies of minor changes in protein expression or modification in a tissue, it is desirable that the entire array of proteins be displayed in one gel.

Several factors need to be considered when samples are taken to detect unintended effects using proteomics. A potato tuber or a tomato fruit is not homogeneous and different tissues and cell types will produce different polypeptide profiles. Developmental processes, storage regimes, genetic, agronomic, and environmental factors also influence the proteome. Less well recognised is the fact that the sample may contain ''non-host proteins'' due to, for example, endophytic microorganisms (fungi, bacteria, viruses). Infected plant material has a different protein pattern from that of a healthy one. For example, Lehesranta et al. (unpublished) discovered tomato mosaic virus protein in a 2-D gel of tomato fruit proteins.

6.3.4. Conclusions

1. Proteomics approaches are crucial to our understanding of development, structure, and metabolism and may be a promising method for the detection and understanding of unintended effects in GM food crops in the future.
2. Proteomics complements other ''omics'' based approaches.
3. The ability to identify proteins from 2-D gels will improve as databases are enlarged and refined.
4. Proteomics combined with immunoblotting with appropriate human sera should help to identify potential allergens in the crop. In some cases, proteomics may be useful not only for the crop, but also for food ingredients; for example, in case of protein isolates and concentrates (e.g. new emulsifiers, foam stabilisers) that may have increased contents of allergenic proteins.
5. Proteomics can be a useful tool for the identification of precursors, post-translational modifications and degradation products of proteins.
6. Proteomics, combined with immunological detection, may be useful for studying whether new fusion proteins are produced from transgene integration sites.

6.4. Metabolomics

6.4.1. Principles and general approaches

The terms *metabolome* and *metabolomics* were defined in section 6.1. To place metabolomics in context, Fiehn has distinguished four types of metabolic investigation (Fiehn, 2001):

1. *Target compound analysis* (analysis of specific compounds most directly affected by a modification or experiment).
2. *Metabolic profiling* (analysis of selected compounds from the same chemical group or compounds linked by known metabolic relationships).
3. *Fingerprinting* (rapid screening for sample classification, e.g. by global analysis of spectroscopic data, no identification and quantification of individual compounds).
4. *Metabolomics* (identification and quantification of as many individual compounds as possible across all compound classes).

The main techniques used to date for metabolomics have been gas chromatography (GC), high performance liquid chromatography (HPLC), and nuclear magnetic resonance (NMR). Analyses of types *1* and *2* have traditionally relied upon chromatographic separations coupled with well-developed calibrations for the specific analytes. Methods like NMR, Fourier transform infrared (FTIR) spectroscopy or mass spectrometry, when applied to crude extracts without any prior separation step, are seen mainly as fingerprinting tools. Metabolomics combines elements of *1, 2,* and *3*. It aims to ensure that valid *quantitative* comparisons can be made with adequate precision, recognising that extraction and measurement procedures cannot be optimal for every single compound that is measured.

6.4.1.1. Experimental design, sampling and extraction procedures. Experimental Design. The initial objective is to be able to detect metabolic changes between GM and appropriate control populations when the two groups of plants are grown under 'identical' conditions. For this, it is necessary to separate the effects of biological variability from metabolic effects arising from the modification, i.e. the number of replicate plants needs to be established. There are potentially hundreds of metabolites to consider, each with its own mean and distribution of values within the positive and control groups. The number of samples required to determine if there is a significant difference will depend on both the compound and the growing conditions. Some metabolites are known to show much greater variability than others; samples grown in the greenhouse will show less variation than those grown in the field. The best available guidelines for experimental design probably come from plant breeding trials in which analyses of specific compounds (e.g. sugars, glycoalkaloids) have traditionally been carried out.

Sampling. Decisions must be taken on whether to measure metabolites in fresh tissue, frozen stored tissue, or freeze dried materials, on which part(s) of the plant to analyse, and on whether individual fruits, tubers etc from one plant should be mixed together or analysed separately. It may be necessary to standardise sizes of fruits or tubers to be analysed (matching of experimental and control samples) since different proportions of flesh to peel, for example, could affect final metabolite levels. Measuring many compounds simultaneously may give added confidence in the sampling procedure if a few compounds differ greatly between comparison groups, whilst the majority do not.

Extraction procedures. The sample preparation step should lead to efficient extraction of a wide range of compounds across the different chemical classes, taking care to suppress enzyme action that would change the metabolite composition of the extracted solution. Material for extraction can be prepared from some fresh tissues by freezing and grinding (in liquid nitrogen) or by freeze-drying. Methanol or methanol/water mixtures are typically used for extraction followed by addition of chloroform/water if separation of polar and lipophilic metabolites is required. Fiehn reviewed the different extraction methods (Fiehn, 2002) but pointed out that there have been no direct comparisons of performance as regards recovery and reproducibility.

6.4.1.2. Methods of measurement. No single technique meets all the requirements for an ideal metabolomics measurement method. There is little doubt, however, that developments involving gas chromatography have been responsible for the recent upsurge of interest in plant metabolomics. GC provides high resolution compound separations and can be used in conjunction with a flame ionisation detector (GC/FID) or a mass spectrometer (GC/MS). In the GC/MS mode, the mass spectrum is scanned (e.g. from m/z 50 to 600) once or twice a second throughout the duration of the GC run, typically 50 minutes. Both detection methods are highly sensitive and universal, i.e. capable of detecting almost any organic compound, regardless of its class or structure. However, most of the metabolites found in plant extracts are too involatile to be analysed directly by GC methods. The compounds have to be converted to less polar, more volatile derivatives before they are applied to the GC column. Efficient derivatisation methods are available (see below), but this represents one complication of the GC method as compared with HPLC or NMR, when there are many samples to be examined.

The earliest application of GC to plant metabolic profiling employed GC/FID (Sauter et al., 1991). Compound identification was by comparison of retention times and mass spectra with those of authentic standards. GC/MS was used only for structure identification in a few 'typical' extracts, not as a routine measurement method. Compensations for variations in instrument performance were made by adding to each sample (i) an internal standard for intensity correction, and (ii) a homologous series of retention time standards, permitting conversion of retention times to retention coefficients. Most compounds give rise to only one peak in the chromatogram, but potentially hundreds of major, minor and trace components are detectable. Since the interest is in the magnitude of any *relative changes* between samples the measurement of trace components is as important as the measurement of major ones. The number of compounds is so large that there will always be some that share the same retention time. In more recent studies, the GC/MS mode has been used for routine detection (rather than GC/FID) and it has been shown that it allows the quantification of minor peaks, even when they are badly overlapped by major ones (Fiehn et al., 2000a). Further details of the GC/MS and updated GC/FID methods of operation are given below (Section 6.4.2).

The second major separation technique is *high performance liquid chromatography* (HPLC). HPLC, with UV detection, is probably the most common method used for targeted analysis of plant materials and for metabolic profiling of individual classes. A derivatisation step is not essential (unless needed for detection), since involatile and volatile substances may be measured equally well. Selection of compounds arises initially from the type of solvent used for extraction (as with all methods that use an extraction step) and then from the type of column and detector. For example, HPLC/UV will only detect compounds with a suitable chromophore; a column selected for its ability to separate one class of compounds will not generally be useful for other types. HPLC profiling methods all rely to a great extent on comparisons with reference compounds. The full UV spectrum (measured for each peak when UV-diode array detectors are used) gives some useful information on compound identity in complex profiles, but often indicates the class of the compound rather than its exact identity.

LC/MS (and in future LC/NMR) techniques offer powerful solutions to the problems of detector generality and structure determination (Mellon et al., 2000). LC/MS can be used to detect compounds that are not well covered by other methods, because they are not well derivatised or are outside the scanned mass range in GC/MS. The electrospray ionisation (ESI) technique has made polar molecules accessible to direct analysis by MS, as well as being compatible with HPLC separations. Quantification of multiple compounds in crude extracts can, in principle, be carried out in the same way as described for GC/MS, although automation of the procedure presents greater practical difficulties. The low sensitivity of LC/NMR relative to LC/MS means that, at present, it is more frequently used for the structural characterisation of unknowns (Wolfender et al., 2001) than for comparative analysis of numerous samples. However, NMR is a very general detection method and can provide unique structural information, so with improvements in sensitivity, its use will grow.

In addition to the methods described above, there are a number of fingerprinting techniques (NMR, direct injection MS, FTIR) that can be used for rapid profiling of large numbers of samples and, to different extents, are capable of providing specific chemical information. Samples are examined after solvent extraction (as in HPLC, no derivatisation is required), or as intact tissues (magic angle spinning NMR), juices or purees (NMR and FTIR), or dried materials (FTIR).

In principle, ^{1}H NMR can detect any metabolites containing hydrogen. Signals can be assigned by comparison with libraries of reference compounds or by two-dimensional NMR (Fan, 1996; Gil et al., 2000). ^{1}H NMR spectra of plant extracts are inevitably crowded, not only because there is a large number of contributing compounds, but also because of the low overall chemical shift dispersion. ^{1}H spectra are also complicated by spin-spin couplings that add to signal multiplicity, although they are an important source of structural information. In ^{13}C NMR, the chemical shift dispersion is twenty times greater and spin-spin interactions are removed by decoupling. Despite these advantages, the low sensitivity of ^{13}C NMR prevents its routine use with complex extracts. Sensitivity can be enhanced when seedlings are grown in the presence of ^{13}C enriched carbon dioxide, but this is obviously only an option for laboratory based studies.

The important role of LC/MS was mentioned above, but it is also possible to obtain metabolite 'mass profiles' without any chromatographic separation. Such profiles are obtained by injecting crude extracts into the electrospray ionisation source of the mass spectrometer (Smedsgaard and Frisvad, 1996). ESI results largely in the production of quasimolecular ions such as $[M+H]^{+}$, $[M+\text{cation}]^{+}$, or $[M\text{-}H]^{-}$ for each species present in the mixture, with a very restricted amount of fragmentation. Thus a fingerprint spectrum is obtained with a single dominant peak for each metabolite and peaks separated according to the molecular mass. The fingerprint can be used as a classification tool, for example in taxonomy. Some mass analysers are capable of very high resolution and permit the mass to be determined to four or five decimal places. This allows unique formulae to be assigned to peaks with masses of a few hundred or so. The coupling of high sensitivity with high resolution provides a method of determining how many

metabolites are present and a valuable first indication, from the formulae, of their identities.

The advantage of FTIR spectroscopy as a fingerprinting method is the ease of sample preparation, the speed with which data can be acquired, and the high degree of reproducibility attainable. Samples that can be poured or spread to make good contact with a flat surface can be measured by the attenuated total reflectance method whereas powdered or dried samples are measured by diffuse reflectance. The spectra are less easily interpreted than with the other methods, but extremely subtle differences may be picked out using chemometrics, providing a powerful classification tool.

6.4.1.3. Data analysis. The objective of the data analysis is to determine whether samples from GM and conventional groups can be differentiated and, if so, to identify the factors underlying the discrimination. The use of computer programmes in data analysis is almost essential but, whatever the measurement method, checking or standardisation of the input data is required to ensure that comparisons are valid, especially between different batches or runs.

Plant extracts are very complex in composition and, when many samples are examined, it is difficult to make any meaningful comparison of their individual spectra or chromatograms 'by eye'. Nevertheless a useful first step is to compare average spectra of GM and conventional samples and attempt to identify the chemical origins of the most obvious differences. Graphical displays (box plots) and standard *univariate* statistical methods (*t*-test, ANOVA) may be applied to individual compounds (or peaks) to identify significant differences in metabolite levels between GM and control samples. If the compounds are explicitly identified and quantified (as with GC/MS) a direct compositional comparison is possible and may be sufficient for an extended substantial equivalence study, using data that has been collected in a non-targeted way.

Multivariate statistical methods give a much better overall picture of how a given sample relates to other samples, a central issue in metabolomics. This applies both to cases where the metabolites are explicitly quantified (GC/MS) and to the fingerprinting techniques, where multivariate analysis is essential. All compounds or peaks are treated simultaneously, but this presents some mathematical problems when the number of variates (compounds measured or data points in a spectrum) exceeds the number of samples. Some procedures for group discrimination and sample classification can be applied to the original data but more often a data compression step is required.

Principal component analysis (PCA) is the main method of data compression (Kemsley, 1998). It is most useful when the original data is highly correlated. PCA transforms the original data (e.g. intensities for hundreds of data points in a spectrum) into a set of 'scores' for each sample on the principal component (PC) axes. The PC axes replace the original variates but are (a) ordered, with successive PCs accounting for decreasing amounts of variance and (b) orthogonal, with no correlation between the scores on different axes. Thanks to these properties, a small number of PCs can replace the many original variates without any loss of information. Scatter plots of the scores on the first few PC axes provide an excellent means of visualizing and summarising the data and often reveal patterns that cannot be discerned in the original data. The scores plots may show clustering of similar samples, separation of different types (GM and control) or the presence of outliers. Plots of the loadings (the second output from a PCA) may also be used to indicate which compounds are most responsible for group separation, as compounds (peaks) with high loading values are most significant. Examples are given in the next section.

6.4.2. Applications

6.4.2.1. GC/FID. A novel metabolite profiling methodology has been developed using rice as a model crop (Frenzel et al., 2002). The role of rice grains as storage organs results in a complex composition characterised by extreme differences in concentrations of compounds. In order to meet this challenge, a method to fractionate the total rice extracts was developed enabling the GC analysis of a broad spectrum of major and minor constituents. The approach is based on consecutive extraction of lipids and polar compounds. Transesterification and solid phase extraction are applied to separate major (fatty acid methylesters) from minor lipids (e.g. phytosterols). Selective hydrolysis of silylated derivatives results in separate fractions of major (sugars) and minor (organic acids/amino acids) polar constituents. Profiles of silylated/methylated compounds are obtained by means of GC/FID; subsequent identification can be achieved by GC/MS.

A set of tools was developed and tailored in order to achieve automated analysis of standardised GC-FID profiling data (Frenzel et al., 2003). As shown in Fig. 3, they include methods for automatic and manual correction of retention times and responses by means of corresponding standards. This allows transfer of normalised datasets into databases and, eventually, automatic comparison of chromatograms. The GC data obtained are further processed directly (FID responses vs. time) or after peak integration (peak areas/heights and retention times). Results of data analysis may be summarised in graphic-, as well as in tabular- reports listing the numbers of new peaks and (significant) differences

6.4.2.2. GC/MS. A complete procedure for metabolomics based on GC/MS measurement has been described

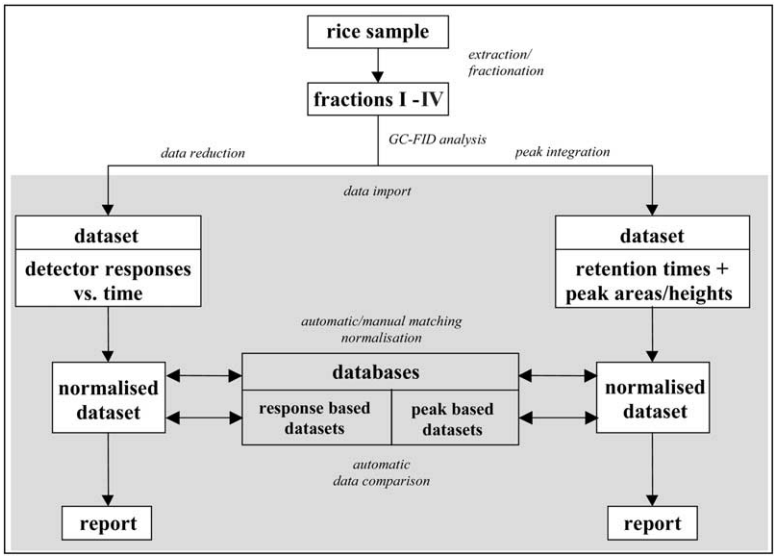

Fig. 3. Approach for automated analysis of standardised metabolite profiling data.

and validated using potato tubers (Roessner et al., 2000) and *Arabidopsis* leaves (Fiehn et al., 2000a). Extraction/fractionation was by addition of methanol to frozen crushed plant tissue, followed by extraction at 70 °C and subsequent separation of polar and lipophilic fractions after additions of water and chloroform. Optimised derivatisation procedures have been described: oximation—trimethylsilylation for the polar fraction and transmethylation—trimethylsilylation for the non-polar fraction.

There were two stages to the measurement procedure. In the first stage, a reference table was constructed containing for each component the retention time, mass spectrum, and characteristic ion used for quantification. The sources of data for this table were (a) mass chromatograms of known metabolites, run as reference compounds and (b) components whose mass spectra were derived from chromatograms of the extracts. These might be components of known structure (identified by searching standard libraries) or unknowns, not found in the library, but with distinct mass spectra. Both the known and unknown components were added to the reference table. An advanced GC/MS method for structure identification of certain unknowns has been developed (Fiehn et al., 2000b). Reference tables containing several hundred 'target' compounds have been placed on the web (Max Planck Institute, 2003).

The extract chromatograms were tested for the presence of each target compound by examining mass spectral scans in a narrow window about the tabulated retention time of the target compound (Fiehn et al., 2000a). After a target compound had been found in the extract (by matching its mass spectrum with the one in the reference table), it was quantified by integrating a characteristic peak from the mass spectrum. The characteristic ion was extracted from the full scan chromatogram, i.e. selected ion monitoring was not used. The absolute quantity of the compound could have been determined using a calibration procedure, but to compare different extracts, the peak area of the compound was divided by the peak area of an internal standard, added to all the extracts. Quantification was routinely automated, but manual integration was required for some components.

Using this approach, over 300 compounds have been quantified in *Arabidopsis* extracts (polar and non-polar fractions). About half of them corresponded to known structures, but the remainder are currently unknown. PCA of GC/MS metabolic data has been used to demonstrate clustering of *Arabidopsis* genotypes (Fiehn et al., 2000a). The samples were two ecotypes with one wild-type and one single-locus mutant in each case, giving four distinct groups. Many plants (28–45 from each group) were grown under 'identical' conditions. After PCA the metabolic composition of each plant could be represented as a point in a 3-dimensional principal component space. The points fell into clusters corresponding to the four groups of plants. Complete elimination of environmental and sample processing effects is never possible, so each cluster had a spread of points determined by the 'environmental' influence, whilst the inter-cluster distance indicated the 'systematic' effects of the genetic differences on composition. The two ecotypes were far apart with a much smaller separation between each wild-type and its associated mutant: in one case, there was a clear internal separation between wild-type and mutant, whereas in the other case, the two clusters partly overlapped. Thus, not only was information on the overall effect of the various genetic differences on metabolic composition obtained, but also

information as to how metabolomics and multivariate analysis could contribute to a new definition of substantial equivalence.

GC/MS measurement of metabolites in soil- and 'in vitro'-grown potato tubers, two systems that had previously been considered highly comparable, showed tenfold and greater differences across a range of compounds (Roessner et al., 2000). In the same study, a comparison was made of tuber composition in three types of transgenic potato (with modified sucrose metabolism or inhibited starch synthesis) and a common wild-type line. It revealed an unexpected appearance in one transgenic line of disaccharides (trehalose, maltose, isomaltose) not detected in the wild-type and not found in previous analyses of the transgenic, simply because these disaccharides had not been among the targets. Further studies (Roessner et al., 2001a,b) on related transgenic lines have identified more unexpected changes and have explored the extent to which the profiles of polar metabolites can be modified by environmental treatments (incubation of tuber discs in sugar solutions). Some initial approaches to the investigation of correlated metabolic changes were also described (Roessner et al., 2001a), which should lead to a deeper understanding of how metabolic networks operate and ultimately provide even greater control over the outcome of genetic modifications.

6.4.2.3. HPLC. Extraction of plant tissues by aqueous alcohol has been used in combination with HPLC/UV to detect phenylpropanoid, tryptophan, and indole derivatives plus some miscellaneous aromatic compounds (Graham, 1991, 1998). Many studies have used HPLC to determine polyphenols in food: for example the flavonoids in fruits, vegetables, and tea were quantified by HPLC/UV after HPLC/MS had confirmed the origin of some closely eluting peaks (Justesen et al., 1998). A comprehensive HPLC profiling procedure for plant isoprenoids (chloroform extraction, monitoring different wavelengths with a diode array detector) has been described with applications to tomato and *Arabidopsis* (Fraser et al., 2000).

LC/MS is now becoming the favoured method for structure determination and quantitative analysis of individual toxicants and bioactive compounds from plants, such as glycoalkaloids, glucosinolates, and phenolics (Mellon et al., 2000). These applications all employed reverse-phase columns, which are appropriate for the separation of relatively hydrophobic compounds. Another type of column must be used for oligosaccharides and other highly polar metabolites that are not detected by the GC/MS method described above. It has been shown that hydrophilic interaction columns may be coupled to electrospray mass spectrometers, for structure determination (by MS/MS) and quantification of polar metabolites (Tolstikov and Fiehn, 2002).

HPLC and other types of chromatography have been used in metabolomics studies outside the plant sciences. Microbial extracts have been analysed for glycolytic intermediates and nucleotides by LC/ESI-MS (Buchholz et al., 2001) and for 27 anionic metabolites by capillary electrophoresis coupled with ESI-MS (Soga et al., 2002). Simultaneous analysis of many redox active metabolites in mitochondria (Kristal et al., 1998) and serum (Vigneau-Callahan et al., 2001) has been carried out by HPLC coupled with Coulometric array detection, a method that offers very high sensitivity. The influence of growth conditions on the bacterial metabolome has been studied by two-dimensional thin-layer chromatography of extracts from *E. coli* grown in media containing ^{14}C-glucose (Tweeddale et al., 1998).

6.4.2.4. NMR. ^{1}H NMR profiling was carried out to compare two types of transgenic tomato (BT and antisense-exogalactanase) with their appropriate controls (Noteborn et al., 2000). The problem of NMR sensitivity was overcome by starting the aqueous extraction with a large amount of material (equivalent to about 2.5g freeze-dried powder). The major components, mainly sugars and other polar compounds, were subsequently removed in the first fraction to reduce the overlap problem. The crude mixture was separated into four fractions by solid phase extraction, which greatly increased the number of compounds that could be resolved by ^{1}H NMR at 400 MHz during examination of these fractions. For example, the amount of starting material was sufficient to permit detection of tomatine in one fraction. A separate chloroform/methanol treatment was used to extract non-polar components. A univariate statistical analysis of modified and control samples was applied to all consistently occurring peaks in the spectra after applying customised software for peak picking and registration. The comparison was carried out for each of the 3000 NMR intensities obtained from fractionated tomato extracts.

A sizeable number of significant differences was detected between Bt-tomatoes and controls grown adjacent to one another on three different occasions. These differences, most of which amounted to no more than two-fold changes of the mean, were revealed to be statistical 'false positives' when results from the three separate crops were combined. For the second transgenic line, the number of differences similarly diminished when further controls were added to the original transgenic/control pair. The one difference that remained when all controls were included concerned α-lycopene (2–4 fold higher in the transgenic), not itself a direct target of the genetic modification, but perhaps changed as a consequence of delayed fruit ripening (Lommen et al., 1998; Noteborn et al., 2000).

Genetic modification has been used to up-regulate flavonoid biosynthesis in order to enhance the anti-oxidant

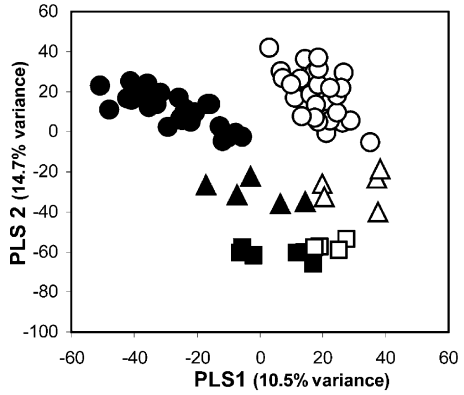

Fig. 4. First two PLS scores for 80 transgenic and control tomatoes at three stages of maturity ○ red transgenic; ● red control; △ turning transgenic; ▲ turning control; □ green transgenic; ■ green control. Reprinted with permission from Journal of Agricultural and Food Chemistry, 2003, 51, 2447–2456. Copyright 2003 American Chemical Society.

capacity of tomatoes. The transgenic tomatoes were generated by simultaneous over-expression of two maize regulatory genes, Leaf colour (*LC*) and Colourless-1 (*C1*), resulting in greatly enhanced levels of flavonoid glycosides in the flesh of *LC*/*C1* tomatoes (Bovy et al., 2002). The enhancement in flavonoid glycosides was expected, but because the modification resulted in increased expression of several genes in the pathway, it was difficult to predict exactly which compounds would be produced. Therefore, the flavonoids in the modified tomatoes were separated, identified by LC/NMR and LC/MS, and quantified by LC/UV to determine changes relative to azygous controls grown at the same time (Le Gall et al., 2003a). This is an example of the metabolite profiling approach.

Extracts of the whole tomatoes (in 70% methanol) were then analysed by ^{1}H NMR spectroscopy in order to detect any unintended effects of the genetic manipulation. The spectra of modified and control tomato fruits at the green, turning, and red stages were reduced to sample scores by applying partial least squares (PLS) data compression to the full spectra. Fig. 4 shows these scores for the first two PLS factors. PLS is a variant of principal component analysis, in which the input data includes the category to which the samples belong as well as the spectra of the samples. The NMR spectra clearly discriminated between GM and control samples on the first PLS axis, and between green, turning, and red tomatoes on the second axis (Le Gall et al., 2003b). Furthermore Fig. 4 shows that the differences between GM and controls become more pronounced as the fruits ripen.

Examination of mean spectra of the different types of sample or the PLS loadings shows which compounds are mainly responsible for the discrimination seen in Fig. 4. For example, Fig. 5 shows segments of the mean spectra of the red transgenic and control fruits highlighting a few selected compounds. The increased level of flavonoid glycosides in the *LC*/*C1* tomatoes is largely responsible for the discrimination between transgenics and controls. Several non-flavonoid compounds (e.g. sucrose, phenylalanine) also showed statistically significant differences between transgenic and control groups, but, in contrast to the flavonoid glycosides, two- to three-fold changes in mean level were the largest observed. Such changes are considered very small against the background of natural variability, confirming that the effects of modification have essentially been confined to the targeted pathway. These results suggest that changes of background composition will be of little importance in this case (although additional sets of samples should be studied and more comprehensive profiling procedures used). The flavonoids themselves included a number of compounds that were new in tomato, mostly glycosides of the commonly occurring aglycone, kaempferol. The compounds had all been identified previously in other plants, but little was known about the absorption and biological activity of these specific compounds in man. Further studies on the activity of the 'intended' products of the modification would therefore be desirable.

One proposed strategy for metabolic phenotyping of silent mutations has used NMR profiling in conjunction with multivariate classification methods (Raamsdonk et al., 2001). No attempt was made to interpret the NMR spectra or loadings in the way described above. Discriminant function analysis (DFA, also known as canonical variates analysis) was used to distinguish six

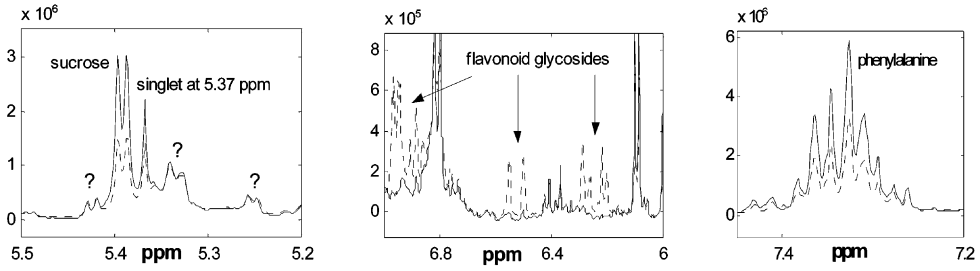

Fig. 5. Details of the mean ^{1}H NMR spectra of red tomatoes showing signals of a selection of metabolites: LC/C1 transgenics (dashed line) and controls (solid line) Adapted from Journal of Agricultural and Food Chemistry, 2003, 51, 2447–2456. Copyright 2003 American Chemical Society.

clusters representing yeast strains with different single gene deletions. The DFA effected a further reduction in dimensionality (following a PCA on the NMR spectra of yeast extracts) and provided the best possible two-dimensional display in terms of group separations. Strains in which the deleted genes had similar functions formed clusters that were close to each other in the two-dimensional scores plot. This suggested a method by which the function of an unknown gene could be discovered if its deletion was found to have the same metabolic effect as deletion of a fully characterised gene.

6.4.2.5. Direct injection MS and FTIR. Mass fingerprints of the secondary metabolites in crude extracts of fungal cultures have been obtained by direct injection of the extracts into an electrospray mass spectrometer. Run time was about one minute. The fingerprints from ten known *Penicillium* species were placed in a database and used for identification of new samples by means of a standard mass spectrometer library search routine (Smedsgaard and Frisvad, 1996). Metabolite profiling of five different types of wine without any need for prefractionation has been described. It employed positive- and negative-ion electrospray ionisation with a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Cooper and Marshall, 2001). The high resolution of the instrument allowed peaks from many compounds to be resolved and the mass accuracy permitted unique elemental compositions to be calculated for most of them. The molecular weight range of the detected compounds was from approximately 300 to 900 and the identified compounds were mostly phenolics plus some carbohydrates. The FTMS method has now been applied to plant metabolomics (Aharoni et al., 2002). The mass profiles of strawberry fruits measured at successive stages of development showed many changes, whereas few differences were found between GM tobacco flowers overexpressing a strawberry MYB transcription factor and their controls. The transgenic tobacco had an altered petal colour and one of the compounds affected was found to be a cyanidin glycoside, an important flower pigment.

The diffuse reflectance method has been used to measure FTIR spectra of homogenised, dried tomato tissues. The tomatoes had been grown in a hydroponic system under normal and 'high saline' conditions. The aim was to classify the tomatoes according to treatment and to identify the compositional differences responsible for the classification (Johnson et al., 2000). The standard multivariate methods of data analysis like PCA could be used for the classification, but the derivation of useful chemical information from the loadings is usually very difficult for FTIR spectra. In this case, the authors used genetic programming (GP) to build classification models that were based on the absorbance at a very small number of frequencies (as few as five out of over 800 measured). The advantage of GP is that it automatically selects those features of the spectrum that are responsible for the discrimination. In this case the frequencies selected were highly characteristic of cyanide group vibrations, and an increased level of cyanide containing compounds is consistent with present understanding of the effects of a high salt treatment. FTIR spectroscopy clearly provides a powerful and convenient method of classifying samples at the whole tissue level. The example discussed highlighted one particular consequence, but the contrast in treatments would be expected to result in much wider metabolic changes. The extent to which IR spectroscopy will be able to specify such changes across a broad range of metabolites remains to be seen.

6.4.3. Possibilities and limitations

6.4.3.1. Summary of requirements. The different metabolomics methods can be judged against the following criteria:

- Sample preparation—extractions should be simple, convenient, repeatable, and non-selective.
- Measurement method—needs to be rapid, reproducible, and capable of automation.
- Good sensitivity, linear response, and high resolving power.
- Simultaneous detection of major and minor components covering extremely wide concentration ranges.
- Possibility of assigning unknown peaks in chromatograms or spectra.
- Consistency of peak positions or ability to 'register' these between different samples.
- Automated extraction of information (compound identity, amount) from complex data sets.
- Data presented in suitable form for statistical analysis.

6.4.3.2. Review of methods discussed. At present, GC/MS is the method that most closely meets the criteria listed above. One advantage of the GC/MS method is its ability to quantify metabolites whose structures are not yet known; another advantage is that use of characteristic ions for quantification means that components with very similar retention times (but different characteristic ions) can be quantified successfully. However, the degree to which the analysis is genuinely 'open' clearly depends on building the widest possible range of target compounds into the reference tables. The number of components that can be distinguished in GC/MS chromatograms is increased with the help of deconvolution programmes, of which the best known is the Automated Mass Spectral Deconvolution and Identification System, AMDIS (Stein, 1999; Halket et al., 1999; programme may be downloaded from http://chemdata.

nist.gov/mass-spc/amdis/). Rapid scanning of the mass range is possible with time of flight (TOF) mass spectrometers and provides the highly digitised chromatograms required by the deconvolution software. Indeed, the use of deconvolution software with GC/TOF-MS has permitted a shortening of GC run times and increased sample throughput. Despite these advances, compounds with molecular weights outside the mass range scanned or compounds that are not derivatized successfully will not be detected. For these reasons, it is likely that other metabolomics methods will continue to be developed as well as GC/MS itself. The method picks up many new compounds that have not traditionally been analysed, but the identification of these compounds is not straightforward. Complete automation of the data extraction procedure is desirable for high throughput studies, but may still present problems. Data standardisation procedures are still evolving. Standardisation is required to correct for variations in instrument performance before the data are statistically analysed. The original procedure used normalisation to a single internal standard (Roessner et al., 2000; Fiehn et al., 2000a), but in a more recent procedure a set of wild type samples was run with every batch of transgenics; subsequently the data were normalised to the average wild type value *for every compound measured* (Roessner et al., 2001a).

A review (Molnár-Perl, 1999) has compared the relative merits of GC and HPLC for simultaneous analysis of sugars and acids in foods and opted in favour of GC on most counts. In the wider context of metabolomics, the two methods are complementary. GC/MS has proved successful in the analysis of small polar primary metabolites; HPLC has been more generally used for larger non-polar secondary metabolites, including most of the natural toxicants in plants. HPLC generally shows good sensitivity, but somewhat poorer resolution compared to GC. Two related problems affect its suitability for high-throughput metabolomics. First is the lack of stability of retention times, which hinders automated input of data to the statistical treatments. Some correction methods have been proposed (Nielsen et al., 1998) and used to classify *Penicillium* species from HPLC profiles (Nielsen et al., 1999). Second is the limited possibility to identify the origin of unknown peaks from HPLC/UV data. LC/MS (as well as LC/NMR and in future LC/NMR/MS) techniques offer some solutions to these problems. Retention index/mass spectral libraries will be constructed with programmes such as AMDIS and then used in conjunction with the instrument manufacturers' own quantification software to reduce raw LC/MS data sets to sample versus compound spreadsheets, with individual cells holding concentration-related intensities. However, unlike GC/MS, no single LC/MS method has yet been demonstrated that allows detection and quantification of all the different classes of compound present in crude plant extracts.

NMR of crude extracts is one of the fastest means of obtaining a metabolite profile when sample preparation and run times are considered together. Whether a specific compound is detected or not depends upon both the amount of compound present (NMR is the least sensitive of the methods discussed) and on where the signals appear in the spectrum. Minor components (e.g. sugar alcohols) that happen to have all their signals in the same region as major components (e.g. glucose, fructose and sucrose) are difficult to resolve. Because of these limitations more elaborate fractionation methods were devised (Lommen et al., 1998; Noteborn et al., 2000), but these procedures greatly lengthen the sample preparation time and do not solve the problem of resolution for compounds of similar polarity. Use of very high field NMR spectrometers with crude extracts improves sensitivity, signal dispersion, and the ability to identify novel metabolites, but at considerable financial cost.

Signal strength is not a reliable guide to the relative abundance of different molecular species in electrospray-MS of crude extracts because ionisation efficiency depends on the acidity (negative-ion ES) or basicity (positive-ion ES) of the neutral molecules (Cooper and Marshall, 2001). The presence of many different species in the crude extracts may lead to suppression of some important signals (for example no peaks with m/z below 300 were detected for the wine samples discussed above). It remains to be seen how reliable the method will prove for quantitative comparisons in real metabolomics applications. Although it may not be absolutely comprehensive, the number of compounds resolved is potentially greater than for any other method. The combination of accurate mass measurements (not necessarily requiring FTMS equipment) with MS/MS techniques would also allow the identity of many of these compounds to be established.

6.4.3.3. The role of chemometrics. Metabolite profiles may eventually be incorporated into a safety screening strategy in different ways. The simplest of these would be to identify peaks from known toxicants and compare relative amounts in test and control samples. However, a more demanding search for 'unintended effects' would make full use (in the multivariate sense) of all the data available. To date, most applications of chemometrics in plant metabolomics have used visual assessments of the dendrograms, scores plots, and loadings provided by the 'unsupervised' methods of cluster analysis and PCA: typically, clusters of the different groups of samples have been clearly revealed on plots of PC1 vs PC2. Such clear separations have been observed since the transgenic lines selected for analysis have had fundamental modifications made to their central metabolism

(e.g. changes in the pathway of sucrose catabolism in potato tubers), resulting in widespread changes in metabolite levels of both the predicted and unpredicted kind. In future 'real life' applications, data analysis tools will be required to identify less obvious changes against backgrounds of greater variability: initial studies have only used plants grown in the laboratory under closely controlled conditions.

Novel foods might be tested for 'substantial equivalence' with respect to an accumulated database of 'acceptable' samples in the same way as GC/MS (Halket et al., 1999) and NMR data (Lindon et al., 2001) have been used for diagnosis of metabolic disorders. Pattern recognition methods can highlight abnormal samples against a background of 'normal' ones, even though 'normal' covers an enormous range of compositions. Chemometric methods such as SIMCA (soft independent modelling of class analogy) (Kemsley, 1998) can deal with this type of problem but will require a large number of samples (covering different genotypes, growing seasons and locations, as well as stress conditions) to model the 'acceptable' class. There will certainly be a need to develop methods of data standardisation to make inter-laboratory use of such databases possible.

Diagnosis of metabolic errors is often dependent on accumulation of high levels of a few individual metabolites that are absent in normal samples. Such changes could correspond to the 'intended effects' of genetic modification in plants. Other changes are likely to be more subtle in the case of GM/conventional crop comparisons, especially in those crops developed to a commercial stage where there are no differences in phenotype or yield, but a rigorous test of their equivalence (or non-equivalence) will still be required. Chemometric tools are available for situations where successful discrimination depends upon, including differences between many, possibly minor, components. Multivariate statistical models, involving more than the two or three dimensions examined in a scatter plot, can be constructed from the PC scores using classification methods such as linear discriminant analysis (LDA) and canonical variates analysis (CVA) (Kemsley, 1998). A decision must be taken on how many PCs to retain in the model; however, this requires information on sample origin (GM or control), so that the classification success rate of the model can be determined. These methods are run in two stages (training and validation), where the training set samples are used to construct the model and an independent validation set is used to test its assignment power. If the number of samples is rather small because of analytical limitations, then cross-validation can be used to determine how many PCs to include. It ensures that any discrimination observed is genuine and not a result of overfitting. The degree of difference between the groups and the variability within each one determine whether a simple PCA is sufficient or whether modelling is needed. The training/validation approach described has not yet been used in metabolomics, but would be required in any regulatory context.

With metabolomics it may always be easier to find *some* differences between two sets of samples than to prove conclusively that there is *no* difference. However, as long as the differences are well defined, their importance can be assessed and any uncertainty reduced.

7. Significance of unintended effects on consumer health

7.1. Safety assessment of unintended effects

Traditional plant sources of food with a long history of use have not been evaluated for safety in a systematic way. Typically, it was by trial and error where the plant was incorporated into the diet, often after some form of processing, e.g. cooking, to make it acceptable from both a taste and safety point of view. Traditional varieties of food crops are known to contain both beneficial components (nutrients and other compounds), as well as compounds with a toxic potential (natural plant toxins, allergens, and anti-nutritional factors, which reduce the availability of nutrients). However, the toxic potential (hazard) will only be expressed (as risk) if the person consuming the foodstuff is exposed to amounts that are sufficient to cause toxic effects.

An additional consideration is the balancing of risk of possible harm against the known nutritional benefits of consuming the plant food (as part of a balanced and diversified diet). Such balancing for traditional foods has taken place over the years in a subliminal way (again by trial and error) and is more recently becoming enshrined in nutritional and dietary guidelines. Therefore, with the introduction of a novel or modified plant into the food supply, it is essential to view its safety in the context of what is already safely used in food.

As explained in previous chapters, crop breeding by both conventional means and by genetic modification has theoretically the potential to modify the plant composition beyond that particular trait that was intended, thus resulting in 'unintended effects'. To analytically determine all possibilities of unintended effects is a huge undertaking with many technical challenges. A further challenge is to determine the real significance of any unintended effect on consumer health. Unintended effects do not automatically imply a health hazard. Hazards may be considered if the nutritional profile of the plant has been altered, if proteins have been altered in such a way so as to affect their allergenic potential, or if new or increased levels of potentially toxic secondary metabolites are produced. However, unintended effects may have absolutely no impact on health, or may even be beneficial by reducing potentially toxic substances.

The aim of the risk assessment procedure for a GM crop is to evaluate its safety when comparing it with the safety of its conventional counterpart (where it exists) that has a safe history of food use. An initial step of this assessment procedure is the determination of compositional equivalence. Based on these data, potential hazards can be identified and subsequently characterised. Exposure assessments are essential, since the intake of only toxicologically significant amounts can lead to health effects, even for relatively toxic substances. The consequences of food processing and consumption patterns must be taken into account. Risk characterisation integrates and interprets the information collected in the previous steps to estimate the probability of occurrence and severity of adverse events in a given population under defined exposure conditions.

The compositional comparative approach leads to three situations:

- No significant differences are identified: no further testing required.
- Only a few well-defined or specific components are identified as different. Further safety assessment would focus on these components. Evaluation is simplified if the identity and function are known and if nutritional/toxicological information is already available. If the identity or function is not known, approaches seeking structural alerts (protein similarities to known allergens, structural similarities of metabolites to known toxic compounds), knowledge of metabolic pathways in which the compound participates, and assessment of likely intakes could be used to determine which, if any, specific components should be evaluated further.
- Numerous differences to the conventional counterpart are detected. From a practical and ethical point of view, it would not be possible to carry out a huge programme of safety testing to evaluate the toxic potential of each of the new or altered components detected. Even if this were possible, the effects of combinations of different compounds may not be reliably predicted from the data of individual compounds. In this case, safety (wholesomeness) testing on the food as a whole must be considered.

The technical approaches described in this document generate the compositional data on which to determine the degree of compositional equivalence. The two approaches of targeted analyses and profiling techniques provide different types, quality, and quantity of information. In both approaches, only significant differences which would lie outside the range of natural variation would be investigated further.

The advantage of targeted analyses is that the availability of both nutritional and toxicological information on components of known concern for a given crop facilitates the choice of compounds to target, as well the interpretation of results. The extent of our knowledge on the natural variation of these individual compounds may, however, be limited. The targeted approach may also include an analysis of environmental contaminants to determine altered susceptibility to the accumulation of heavy metals, pesticides and their metabolites, and fungal toxins. For example, Bt maize expressing insecticidal Cry proteins have been shown to have decreased *Fusarium* infections, resulting in lower levels of the mycotoxin Fumonisin (Munkvold et al., 1997, 1999). A second level of targeted analysis would depend on knowledge of (a) metabolic pathways in which the product of the transgene could participate and (b) the genomic region into which the transgene construct has been inserted and the function of any subsequently disrupted genes. Toxicologically relevant effects may, however, be difficult to predict due to the complexity of biochemical pathways.

Profiling (non-targeted) approaches measure a wide range of parameters, which are not defined prior to analysis. This is advantageous in that the food as a whole is being analysed. Genomics, proteomics, and metabolomics have the potential for generating massive amounts of data, simultaneously characterising the presence and amounts of potentially thousands of molecules (of known and unknown function). This would certainly lead to a more complete compositional profile of the crop, and increase the possibility of understanding complex metabolic pathways. The safety significance of data generated from profiling techniques should, however, be interpreted with great caution. The correct interpretation of the data generated by such techniques relies heavily on the availability of adequate comparative data on compositional profiles and crop variation and the establishment of criteria to determine what constitutes a significant difference. This reference data should be generated from crops with an accepted safety and used as crop specific 'benchmarks' to which to compare the profile of the GM crop. At present, few such data are available. A lack of comparative compositional profiles may lead to mis-interpretation of data, and so to unneccessary, or lack of, concern. The development of such reference data is therefore of fundamental importance to the usefulness of profiling techniques in providing a basis for compositional equivalence and subsequent safety assessments.

7.2. Conclusions

- Unintended effects do not automatically infer health hazards. The significance of unintended effects on consumer health must be evaluated in

the risk assessment process, and must take into account the intended use of the food.

- Compositional data on GM crops compared to non-GM counterparts is neccessary to determine the degree of compositional equivalence on which to focus any further safety testing in the process of risk assessment.
- Only crops which show significant differences to the conventional counterpart should be evaluated further.
- The compilation of consensus documents identifying parameters important for safety/nutritional evaluations and providing acceptable 'normal' ranges would be valuable in standardising current targeted analyses.
- The development of reference databases is essential to the interpretation of data generated from profiling techniques and should be required before the acceptance of such techniques as part of the routine strategy for safety assessment
- A combination of targeted and non-targeted methods of analysis (to be decided on a case by case basis) is likely to be the best way forward to evaluate the safety of GM and conventionally bred crops.

8. Overall conclusions

The aim of crop breeding is to apply selection aimed at specific characteristics, such as improving nutritional quality and yield. The major source of natural variation and of breeding programmes is the natural molecular mechanisms of DNA exchange and repair. These mechanisms are the same for all crops, irrespective of whether the DNA has been specifically modified by genetic engineering techniques or has been altered via conventional crossing of different varieties. In addition to the introduction of selected characteristics (intended effects), unintended effects may also occur. There is no indication that unintended effects are more likely to occur in GM crops than in conventionally bred crops. Unintended effects may have positive, negative, or indeed no consequences on the agronomical vigour or safety profile of the crop. The same field selection processes apply to both conventional and GM breeding. This selection process takes many years and removes major unintended effects.

The introduction of crops produced by novel technologies, such as genetic engineering methods, onto the market place is regulated under the Novel Foods Regulation (EC 258/97). A thorough pre-market safety assessment is required. This is not a requirement for the introduction of new seed varieties bred by conventional breeding, although unintended effects may also be present in these crops. The safety of conventionally bred crops is taken for granted based on a history of safe use. However, some cases (extremely rare) have been reported where unintended effects have given rise to safety concerns. These were identified after the crop had already entered the market. Characterisation of GM crops is a legal requirement, and is part of the safety assessment. GM crops are therefore better characterised than conventionally bred crops, including knowledge on the site and nature of the genetic modification. GM crop characterisation currently includes compositional analysis of pre-selected nutrients and toxins, while a comparison is made relative to the composition of conventional crops. Criticisms of this current strategy are that it is open to bias and will never pick up unexpected unintended effects. Profiling techniques may help address these potential limitations.

Profiling techniques such as genomics, proteomics, and metabolomics provide a 'global' overview of gene expression and chemical composition within the crop, be it GM or non-GM. These techniques aim to be unbiased with regard to the choice of genes, proteins, and metabolites profiled. Methodologies are still rapidly developing; they are not yet (and may never be) comprehensive. All current profiling approaches are based on comparison of GM materials with selected controls in self-contained experiments. The data generated have a great potential to increase our knowledge of plant physiology and metabolic networks, and will also improve targeted analyses. This provides advantages for all types of breeding programmes. Vast quantities of data can be generated from these methods; however, subsequent interpretation of the data is at present limiting. To date, there is a lack of data on which to determine the useful contribution of these techniques to GM crop safety assessments.

Unintended effects do not automatically infer health hazards. Ideally, only those parameters that fall outside the range of natural variation would be considered further in safety assessment. However, there is a lack of information on the natural variation within and between given plant cultivars for all the parameters that may now be measured. Safety assessment is simplified if the identification and safety significance of differences is known. A major drawback would be the lack of adequate toxicity databases to interpret the safety significance of compounds with unknown identity and/or function. Numerous differences and/or differences of unknown function may lead to the consideration of more extensive safety testing.

Further research is required to validate profiling methodologies. Informatic tools need to be developed to extract relevant information from raw data sets. The comprehensiveness of coverage given by all profiling techniques needs to be improved. The development of publicly available databases of crop composition and profiles is an absolute requirement to determine natural variation of compounds within and between given plant

species. As information is gathered, a benchmark (expanding) to which to compare new crops could be envisaged. These databases would also greatly aid the robustness of targeted analyses.

The safety assessment of GM crops should focus primarily on the intended novel traits (target gene(s) and product(s)). Unintended effects occur in both GM and non-GM crops; however, GM crops are better characterised. It may be suggested that the two should be treated the same in safety assessments, bearing in mind that safety assessments are not required for non-GM crops. Profiling techniques should not at present be an official requirement. However, their further development should be actively encouraged. Government funding will be required to develop crop composition databases.

Appendix

To avoid duplication and to learn from results obtained in earlier EU framework programmes, a comprehensive review of the ECLAIR, FLAIR, AIR, FAIR and the Fifth framework programme was carried out. The objectives were to identify either previously known unintended effects or techniques that might be relevant to the identification of unintended effects in genetically modified food crops. However, the nature of the available reports from these programmes only allows for a brief review of potentially interesting projects since very few details are available. For further evaluation of their usefulness, direct contact with the participating partners in these projects will be necessary.

Only in one of the reviewed projects were results on observed unintended effects described. In ECLAIR AGRE.001 (The research development and production of low temperature storage tolerant chipping (crisping) potato cultivates), the overall objectives were to prevent the accumulation of reducing sugars during storage of potato tubers at low temperatures. After cloning of genes encoding enzymes that have a regulatory role in the pathway of starch respiration and sugar metabolism, transgenic plants with a high constitutive levels of phosphofructokinase were generated. However, although cold storage of these tubers demonstrated the desired inhibition of cold-induced sucrose and reducing sugar synthesis, in some cases such plants also displayed an undesirable phenotypic effect on shoot and tuber growth.

In project AIR3-CT94-2311 (Development of new methods for safety evaluation of transgenic food crops), interesting methods were developed for carbohydrate profiling and chemical fingerprinting. Both these methods are very promising and are currently being used with the purpose of identifying unintended effects in the ongoing GMOCARE project.

In several other framework programmes, techniques for identification of mostly specific compounds are evaluated and might be of interest for this project. A list of examples of possible interesting projects is given below, together with a short review of project AIR3-CT94-2311.

ECLAIR AGRE.0021: Optimisation of lignin in crop and industrial plants through genetic engineering

The objective was to alter lignification in plants through anti-sense RNA techniques. Changes in transformed plants were analysed by chemical and cytochemical analysis.

ECLAIR AGRE.0048: To increase the use of Pisum and other grain legumes by improving the composition and nutritional value of the seed by genetic and by processing techniques

The objectives of the project were to increase knowledge of the composition and nutritional value in monogastic animals of pea and other grain legume spices. A large database on composition and nutritional value of different lines of peas was created. An infrared spectroscopy calibration method was set-up for this information. The generated database might be an extended set of data which could be very useful for comparison when generating transgenic peas (“normal level”)

ECLAIR AGRE.0064: Isolation of new starter cultures from cheese and fermented milk

The objectives were to isolate and characterize new starter cultures biochemically, biologically, as well as genetically. Gas chromatographic techniques (GC/MS) were developed to profile flavour production by different strains.

ECLAIR AGRE.0001: The research development and production of low temperatures storage tolerant chipping (crisping) potato cultivates

The overall objective is to prevent the accumulation of reducing sugars during storage of potato tubers at low temperatures. The specific objectives were to clone genes which code for enzymes that have a regulatory role in the pathways of starch respiration and sugar metabolism. Several transgenic potato lines were obtained, where some that expressed high constitutive levels of phosphofructokinase either in the cytosol or in the plastids, yielded plants displaying undesirable phenotypic effects on shoot and tuber growth. However, cold-storage of these tubers demonstrated the desired inhibition of cold-induced sucrose and synthesis of reducing sugars.

FAIR-CT97-3676: Development of new strategies for resistance engineering in transgenic potato toward Ralstonia (pseudomonas) solanacearum *and* Erwina carotovora *(Bacterial resistance)*

The objective of this project is to develop new, specific and effective resistance strategies toward *R. solanacearum* and *E. carotovora* by means of genetic engineering. http://www.uni-hohenheim.de/fair/project.htm

FAIR-CT97-3761: Engineering rice for resistance to insects (ERRI)

The objectives of this project are to develop striped stemborer (*Chilo suppressalis*) resistant rice through transfer of *Bacillus thuringiensis* (B.t.) genes encoding insecticidal toxins and plant proteinase inhibitor (p.i.) genes into the most widely used rice varieties in Europe. Several transgenic plants have been generated and evaluated.

FAIR-CT98-3844: Public perceptions of agricultural biotechnologies in Europe (PABE)

The objectives are to identify the content and pattern of background factors that might affect public perceptions (such as views of scope and responsiveness of risk-management regulatory systems, of the ethical acceptability of gene transfer from animals to plants, or of the industrialisation of food generally) and to assist in more informed and constructive public debate in this area. http://domino.lancs.ac.uk/ieppp/Home.nsf

FAIR-CT98-4239: Lower application of insecticides by the production of insect-resistant crops using novel protease inhibitor genes (NOVEL PI CROPS)

The objectives are to generate transgenic crops expressing broader insecticidal genes like protease inhibitor (PI) genes. The project concentrates on potato, tomato, maize and melon.

FAIR5-3493: TOMSTRESS - Engineering tomato against environmental stress

The main objective is to increase knowledge on the impact of environmental stress on yield and quality of tomato. Specifically to understand the role of ethylene and active oxygen species in oxidative stress and to develop genetic engineering strategies to reduce the oxidative stress. Several partners in the project have ongoing methods to study and evaluate stress in plants. This includes gene expression, changes in secondary metabolism, measurements of antioxidant enzyme activities, and evaluation of flavour/aroma volatiles by HPLC. http://www.gsf.de/Forschung/institute/biop.phtml

QLK3-1999-00531: Development of high-throughput PNA-based molecular diagnostic systems (EUPNA)

The main objectives are to develop new powerful parallel genetic analysis procedures based on the use of the DNA mimic, peptide nucleic acid (PNA), as probe. The project aims at producing a library of PNA probes designed to detect specific gene mutations in microarray- and capillary electrophoresis-based diagnostic systems as well as designing and improving methods to separate PNA-oligonucleotides in capillary electrophoresis.

QLG2-1999-00876: Regulatory gene initiative in Arabidopsis *(REGIA)*

The main objectives of this study are to generate transcription factors (TF) gene arrays and a normalised TF ORF library in a 2-Hybrid vector (to study interactions between TFs).

QLK3-2000-40698: Characterisation of glycosylation structures of pharmaceutical proteins produced in plants with DNA-micro arrays and glycosylation fingerprinting

The objectives are to produce recombinant proteins in plants, analyse glycosylation patterns by fingerprinting techniques and RNA profiling by DNA microarrays.

QLK1-1999-00156: Food safety in Europe: Risk assessment of chemicals in food and diet

The objectives of the project include a detailed state-of-the-art evaluation of all stages involved in risk assessment and seek to integrate these in the most relevant manner for assessing risk using a matrix approach. Specifically to explore means to improve the principles applied to risk assessment, identify gaps in knowledge that might lead to differences in interpretation of toxicological and exposure data, as well as to determinate the nature and level of testing needed for safety evaluation relevant to the nature of chemicals.

AIR3-CT94-2311: Development of new methods for safety evaluation of transgenic food crops

Participants:

DLO-RIKILT, Department of Risk Assessment and Toxicology, The Netherlands (RIKILT-DLO)

AgrEvo, Plant Genetic Systems NV, Belgium (PGS)

Vrije Universiteit Brussel, Department of Protein Chemistry, Belgium (VUB)

BIBRA International, United Kingdom (BIBRA)

Institute for Agrobiotechnology (IFA), Center of Analytical Chemistry, Austria (UBW.IC.IAB)

Scientific objectives

The objectives of project AIR3-CT94-2311 were to set up generic approaches to study and evaluate the safety of novel food crops. The scientific work consists of four main parts: Development of biochemical methods to isolate novel food proteins and to detect post-translational modification differences of newly introduced proteins (i.e. glycosylation). Development of a set of analytical test methods in order to identify unintentional changes in the composition of nutrients and naturally occurring toxicants, which may occur as a result of genetic modification (i.e. chemical fingerprinting). Characterization of the allergenic potency of the newly introduced proteins and modified food. Development of *in vitro* test methods to screen the toxicological properties of newly expressed proteins and modified food (i.e. toxicological profiling).

In project AIR3-CT94-2311, several conclusions on the topic of substantial equivalence were highlighted regarding the importance of selecting the appropriate counterparts. Thus, in order to differentiate between possible compositional changes due to genetic changes, to genetic variability or to environmental variations, several comparisons have to be considered: (i) isogenic lines grown under identical growth conditions; (ii) isogenic lines bred at multiple locations; (iii) the influence of downstream processing; (iv) an extended range of controls by entering current commercial varieties; and (v) the transformant and the range of that crop. Thus, the crop mean values and differences were of final interest, while these values were compared to those of an extended control population of commercial varieties of that crop, including external factors such as timing, processing, and environmental variations.

For carbohydrate profiling, two different purification methods were developed. From tomato-fruit sixteen different N-glycans of the oligomannosidic and complex-type or truncated oligosaccharides could be identified. Investigation of both non-transgenic and transgenic Bt-tomatoes (green and red-ripe fruit) revealed no difference in the relative levels either across seasons or to the non-modified extended control. It was concluded that carbohydrate profiling shows that as far as N-linked oligosaccharides were involved, it was not possible to detect significant alterations in the alpha-1,3-fucose content of tomatoes due to this genetic modification. However this technique appears to be a very powerful screening tool for the detection of compositional changes in N-glycans. Similarly, Off-line Liquid Chromatography combined with high resolution proton (^{1}H)-nuclear magnetic resonance (LC-NMR) was also shown to be a potential method for the screening for changes in complex matrices due to environmental effects, as well as for unintended effects in genetically modified crops. Some of the important features of these fingerprinting methods include: (i) a very broad detection method for low molecular compounds; (ii) a large dynamic range in concentrations to be detected in one spectrum (up to a factor of 10,000); (iii) a high degree of resolution enabling detection of many compounds simultaneously and (iv) a means of identifying compounds by delivering structural information.

To evaluate the ability of *in vitro* methods to screen for differences in toxicity between genetically modified and conventional foods, and thus to be used as a tool for establishing substantial equivalence, several cytotoxicity test were evaluated. These tests included a gastro-intestinal cell line from rat, a human embryonic intestinal cell line, and a human colon tumor cell line, as well as several commercially available assays. Using extracts from tomatoes, only green tomatoes exhibited cytotoxic effects with no difference between modified and non-modified crops. It was concluded that more experimental work was required in order to evaluate the potential use of these systems as screening tools for food products.

In conclusion, it was suggested that carbohydrate profiling and chemical fingerprinting in some incidences, together with appropriate extended controls, might at least limit animal feeding studies when testing for unintended effects.

References

ACNFP, 1999. Annual Report 1999. Advisory Committee on Novel Foods and Processes, Ministry of Agriculture, Fisheries and Food, Department of Health, London. http://www.foodstandards.gov.uk/multimedia/pdfs/acnfp1999.

Aharoni, A., de Vos, C.H.R., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., Goodenowe, D.B., 2002. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. OMICS 6, 217–234.

Allard, R.W., 1999. History of plant population genetics. Annual Review of Genetics 33, 1–27.

Alwine, J.C., Kemp, D.J., Stark, G.R., 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences U.S.A. 74, 5350–5354.

Andersen, J.S., Mann, M., 2000. Functional genomics by mass spectrometry. FEBS Letters 480, 25–31.

Babiychuk, E., Fuangthong, M., van Montagu, M., Inzes, D., Kushnir, S., 1997. Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach. Proceedings of the National Academy of Sciences U.S.A. 94, 12722–12727.

Barakat, A., Gallois, P., Raynal, M., Mestre-Ortega, D., Sallaud, C., Guiderdoni, E., Delseny, M., Bernardi, G., 2000. The distribution of T-DNA in the genomes of transgenic Arabidopsis and rice. FEBS Letters 471, 161–164.

Bassett, D.E., Eisen, M.B., Boguski, M.S., 1999. Gene expression informatics—it's all in your mine. Nature Genetics 21 (Suppl.), 51–55.

Bauer, M.D., Sun, Y., Keough, T., Lacey, M.P., 2000. Sequencing of sulfonic acid derivatized peptides by electrospray mass spectrometry. Rapid Communications in Mass Spectrometry 14, 924–929.

Beier, R.C., 1990. Natural pesticides and bioactive components in food. Reviews of Environmental Contamination and Toxicology 113, 47–137.

Berk, A.J., Sharp, P.A., 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of SI endonuclease-digested hybrids. Cell 12, 721–732.

Birch, R.G., 1997. Plant transformation: Problems and strategies for practical application. Annual Review of Plant Physiology and Plant Molecular Biology 48, 297–326.

Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Almenar Pertejo, M., Muir, S., Collins, G., Robinson, S., Verhoeyen, M., Hughes, S., Santos-Buelga, C., van Tunen, A., 2002. High-flavonol tomatoes resulting from heterologous expression of the maize transcription factor genes LC and C1. Plant Cell 14, 2509–2526.

Britt, A.B., 1996. DNA damage and repair in plants. Annual Review of Plant Physiology and Plant Molecular Biology 47, 75–100.

Brown, P.O., Botstein, D., 1999. Exploring the new world of the genome with DNA microarrays. Nature Genetics 21 (Suppl.), 33–37.

Buchholz, A., Takors, R., Wandrey, C., 2001. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionisation tandem mass spectrometry. Analytical Biochemistry 295, 129–137.

Camerini-Otero, R., Hsieh, P., 1995. Homologous recombination proteins in prokaryotes and eukaryotes. Annual Review of Genetics 29, 509–552.

Castle, L.A., Errampalli, D., Atherton, T.L., Franzmann, L.H., Yoon, E.S., Meinke, D.W., 1993. Genetic and molecular characterization of embryonic mutants identified following seed transformation in Arabidopsis. Molecular and General Genetics 241, 504–514.

Clegg, M.T., Cummings, M.P., Durbin, M.L., 1997. The evolution of plant nuclear genes. Proceedings of the National Academy of Sciences U.S.A. 94, 7791–7798.

Codex Alimentarius Commission, 2003. Draft Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants, at Step 8 of the Elaboration Procedure. ALINORM 03/34, Appendix III. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organisation, Rome. ftp://ftp.fao.org/codex/alinorm01/al0134ae.pdf.

Conner, A.J., Jacobs, J.M., 1999. Genetic engineering of crops as potential source of genetic hazard in the human diet. Mutation Research 443, 223–234.

Cooper, H.J., Marshall, A.G., 2001. Electrospray ionization Fourier transform mass spectrometric analysis of wine. Journal of Agricultural and Food Chemistry 49, 5710–5718.

Cordwell, S.J., Nouwens, A.S., Verrills, N.M., Basseal, D.J., Walsh, B.J., 2000. Subproteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional electrophoresis gels. Electrophoresis 21, 1094–1103.

Costa, P., Plomion, C., 1999. Genetic analysis of needle proteins in maritime pine: 2. Variation of protein accumulation. Silvae Genetica 48, 146–150.

Costa, P., Pionneau, C., Bauw, G., Dubos, C., Bahrmann, N., Kremer, A., Frigerio, J.M., Plomion, C., 1999. Separation and characterization of needle and xylem maritime pine proteins. Electrophoresis 20, 1098–1108.

Costa, P., Pot, D., Dubos, C., Frigerio, J.M., Pionneau, C., Bodenes, C., Bertocchi, E., Cervera, M.-T., Remington, D.L., Plomion, C., 2000. A genetic map of Maritime pine based on AFLP, RAPD and protein markers. Theoretical and Applied Genetics 100, 39–48.

Coulston, F., Kolbye, A.C., 1990. Biotechnologies and food: assuring the safety of foods produced by genetic modification. Regulatory Toxicology and Pharmacology 12, S1–S196.

De Buck, S., De Wilde, C., van Montagu, M., Depicker, A., 2000. Molecular Breeding 6, 459–468.

Delhaize, E., Hebb, D.M., Richards, K.D., Lin, J.M., Ryan, P.R., Gardner, R.C., 1999. Cloning and expression of a wheat (*Triticum aestivum*) phosphatidylserine synthase cDNA. Overexpression in plants alters the composition of phospholipids. Journal of Biological Chemistry 274, 7082–7088.

Dueck, Th.A., van der Werf, A., Lotz, L.A.P., Jordi, W., 1998. Methodological Approach to a Risk Analysis for Polygene-Genetically Modified Plants (GMPs): a Mechanistic Study. AB Nota Vol. 50. Research Institute for Agrobiology and Soil Fertility (AB-DLO), Wageningen.

Engel, K.H., Gerstner, G., Ross, A., 1998. Investigation of glycoalkaloids in potatoes as example for the principle of substantial equivalence. In: Novel Food Regulation in the EU- Integrity of the Process of Safety Evaluation. Federal Institute of Consumer Health Protection and Veterinary Medicine, Berlin, pp. 197–209.

European Commission, 1997. 97/618/EC: Commission Recommendation of 29 July 1997 Concerning the Scientific Aspects and the Presentation of Information Necessary to Support Applications for the Placing on the Market of Novel Foods and novel Food Ingredients and the Preparation of Initial Assessment Reports under Regulation (EC) No 258/97 of the European Parliament and of the Council. Official Journal of the European Communities L253, 1–36.

Fan, T.W.M., 1996. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. Progress in Nuclear Magnetic Resonance Spectroscopy 28, 161–219.

FAO/WHO, 1991. Strategies for Assessing the Safety of Foods Produced by Biotechnology, Report of a Joint FAO/WHO Consultation. World Health Organization, Geneva.

FAO/WHO, 2000. Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, Geneva, Switzerland, 29 May-2 June 2000. Food and Agriculture Organisation of the United Nations, Rome. ftp://ftp.fao.org/es/esn/food/gmreport.pdf.

Faris, J.D., Haen, K.M., Gill, S., 2000. Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 154, 823–835.

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

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

Fiehn, O., 2001. Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks. Comparative and Functional Genomics 2, 155–168.

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

Fladung, M., 1999. Gene stability in transgenic aspen (Populus) I. Flanking DNA sequences and T-DNA structure. Molecular and General Genetics 260, 574–581.

Franssen-van Hal, N.L.W., Vorst, O., Kramer, E., Hall, R.D., Keijer, J., 2002. Factors influencing cDNA microarray hybridisation on silylated glass slides. Analytical Biochemistry 308, 5–17.

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

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

Frenzel, Th., Miller, A., Engel, K.-H., 2003. A methodology for automated comparative analysis of metabolite profiling data. European Food Research and Technology 216, 335–342.

Frewer, L., Lassen, J., Kettlitz, B., Scholderer, J., Beekman, V., Berdal, K.G., 2004. Societal aspects of genetically modified foods. Food and Chemical Toxicology 42. DOI:10.1016/j.fct.2004.02.002.

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. Proceedings of the National Academy of Sciences U.S.A. 97, 4718–4723.

Fuchs, R.L., Re, D.B., Rogers, S.G., Hammond, B.G., 1996. Safety

evaluation of glyphosate-tolerant soybeans. In: Food Safety Evaluation. Organisation for Economic Cooperation and Development, Paris, pp. 61-70.

Gelvin, S.B., 1998. The introduction and expression of transgenes in plants. Current Opinion in Biotechnology 9, 227–232.

Gelvin, S.B., 2000. Agrobacterium and plant genes involved in T-DNA transfer and integration. Annual Review of Plant Physiology and Plant Molecular Biology 52, 223–256.

Gerstein, M., 2000. Integrative database analysis in structural genomics. Nature Structural Biology 7 (Suppl.), 960–963.

Gertz, J.M., Vencill, W.K., & Hill, N.S., 1999. Tolerance of transgenic soybean (Glycine max) to heat stress. In: 1999 Brighton Crop Protection Conference: Weeds. Proceedings of an International Conference, Brighton, UK, 15-18 November 1999, Volume 3. British Crop Protection Council, Farnham, pp. 835–840.

Gheysen, G., Villarroel, R., van Montagu, M., 1991. Illegitimate recombination in plants: a model for T-DNA integration. Genes and Development 5, 287–297.

Gill, K.S., Gill, B.S., Endo, T.R., Taylor, T., 1996. Identification and high-density map of gene-rich regions in chromosome group 1 of wheat. Genetics 144, 1883–1891.

Gil, A.M., Duarte, I.F., Delgadillo, I., Colquhoun, I.J., Casuscelli, F., Humpfer, E., Spraul, M., 2000. Study of compositional changes of mango during ripening by use of nuclear magnetic resonance. Journal of Agricultural and Food Chemistry 48, 1524–1536.

GMOCARE, 2003. GMOCARE, New Methodologies for Assessing the Potential of Unintended Effects in Genetically Modified Food Crops. GMOCARE Secretariat. RIKILT Institute of Food Safety, Wageningen. http://www.entransfood.com/RTDprojects/GMOCARE/default.htm.

Gorbunova, V., Levy, A.A., 1999. How plants make ends meet: DNA double-strand break repair. Trends in Plant Science 4, 263–269.

Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W., 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21, 1037–1053.

Görg, A., Obermaier, C., Boguth, G., Weiss, W., 1999. Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. Electrophoresis 20, 712–717.

Graham, T.L., 1991. A rapid, high resolution, high performance liquid chromatography profiling procedure for plant and microbial aromatic secondary metabolites. Plant Physiology 95, 584–593.

Graham, T.L., 1998. Flavonoid and flavonol glycoside metabolism in Arabidopsis. Plant Physiology and Biochemistry 36, 135–144.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 17, 994–999.

Haber, J.E., 2000. Recombination: a frank view of exchanges and vice versa. Current Opinion in Cell Biology 12, 286–292.

Halket, J.M., Przyborowska, A., Stein, S.E., Mallard, W.G., Down, S., Chalmers, R.A., 1999. Deconvolution gas chromatography/mass spectrometry of urinary organic acids—potential for pattern recognition and automated identification of metabolic disorders. Rapid Communications in Mass Spectrometry 13, 279–284.

Hansen, G., Wright, M.S., 1999. Recent advances in the transformation of plants. Trends in Plant Science 4, 226–231.

Harvey, M.H., McMillan, M., Morgan, M.R., Chan, H.W., 1985. Solanidine is present in sera of healthy individuals and in amounts dependent on their dietary potato consumption. Human Toxicology 4, 187–194.

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

Hashimoto, W., Momma, K., Yoon, H.-J., Ozawa, S., Ohkawa, Y., Ishige, T., Kito, M., Utsumi, S., Murata, K., 1999b. Safety assessment of transgenic potatoes with soybean glycinin by feeding studies in rats. Bioscience Biotechnology and Biochemistry 63, 1942–1946.

Haynes, P.A., Yates, J.R. 3rd, 2000. Proteome profiling-pitfalls and progress. Yeast 17, 81–87.

Herman, L., Jacobs, A., van Montagu, M., Depicker, A., 1990. Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. Molecular and General Genetics 224, 248–256.

Hoving, S., Voshol, H., van Oostrum, J., 2000. Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. Electrophoresis 21, 2617–2621.

Iglesias, V.A., Moscone, E.A., Papp, I., Neuhuber, F., Michalowski, S., Phelan, T., Spiker, S., Matzke, M., Matzke, A.J., 1997. Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. Plant Cell 9, 1251–1264.

Inukai, T., Sako, A., Hirano, H.Y., Sano, Y., 2000. Analysis of intragenic recombination at wx in rice: correlation between the molecular and genetic maps within the locus. Genome 43, 589–596.

Jacobs, J.M.E., van Eck, H.J., Arens, P., Verkerk-Bakker, B., te Lintel Hekkert, B., Bastiaanssen, H.J.M., El-Kharbotly, A., Pereira, A., Jacobsen, E., Stiekema, W.J., 1995. A genetic map of potato (*S. tuberosum*) integrating molecular markers, including transposons, and classical markers. Theoretical and Applied Genetics 91, 289–300.

Johnson, H.E., Gilbert, R.J., Winson, M.K., Goodacre, R., Smith, A.R., Rowland, J.J., Hall, M.A., Kell, D.B., 2000. Explanatory analysis of the metabolome using genetic programming of simple interpretable rules. Genetic Programming and Evolvable Machines 1, 243–258.

Justesen, U., Knuthsen, P., Leth, T., 1998. Quantitative analysis of flavonols, flavones and flavanones in fruits, vegetables and beverages by high-performance liquid chromatography with photodiode array and mass spectrometric detection. Journal of Chromatography A 799, 101–110.

Kemsley, E.K., 1998. Discriminant Analysis and Class Modeling of Spectroscopic Data. Wiley, Chichester.

Keough, T., Lacey, M.P., Fieno, A.M., Grant, R.A., Sun, Y., Bauer, M.D., Begley, K.B., 2000. Tandem mass spectrometry methods for definitive protein identification in proteomics research. Electrophoresis 21, 2252–2265.

Kohli, A., Leech, M., Vain, P., Laurie, D.A., Christou, P., 1998. Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. Proceedings of the National Academy of Sciences U.S.A. 95, 7203–7208.

Kok, E.J., van der Wal-Winnubst, E.N.W., Van Hoef, A.M.A., Keijer, J., 2001. Differential display of mRNA. In: Akkermans, A.D.L., van Elsas, J.D., de Bruijn, F.J. (Eds.), Molecular Microbial Ecology Manual, Supplement 5, 8.1.1. Kluwer, Dordrecht, pp. 1–10.

Kok, E. J., Keijer, J., Van Hoef, A. M. A., Van der Wal-Winnubst, E. N. W., Henkens, M. H. C. and Kuiper, H. A., 1998. mRNA fingerprinting of transgenic food crops. In: Report of the Demonstration Programme on Food Safety Evaluation of Genetically Modified Foods as a Basis for Market Introduction. Ministry of Economic Affairs, The Hague, 37-49.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Korber, H., Redei, G.P., Schell, J., 1989. High-frequency T-DNA mediated gene tagging in plants. Proceedings of the National Academy of Sciences U.S.A 86, 8467–8471.

Koncz, C., Nemeth, K., Redei, G.P., Schell, J., 1992. T-DNA insertional mutagenesis in Arabidopsis. Plant Molecular Biology 20, 963–976.

König, A., Cockburn, A., Crevel, R.W.R., Debruyne, E., Grafstroem, R., Hammerling, U., Kimber, I., Knudsen, I., Kuiper, H.A., Peijnenburg, A.A.C.M., Penninks, A.H., Poulsen, M., Schauzu, M.,

Wal, J.M., 2004. Assessment of the safety of food derived from genetically modified (GM) crops. Food and Chemical Toxicology 42. DOI:10.1016/j.fct.2004.02.019.

Kononov, M.E., Bassuner, B., Gelvin, S.B., 1997. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. Plant Journal 11, 945–957.

Kristal, B.S., Vigneau-Callahan, K.E., Matson, W.R., 1998. Simultaneous analysis of the majority of low-molecular-weight, redox active compounds from mitochondria. Analytical Biochemistry 263, 18–25.

Krizkova, L., Hrouda, M., 1998. Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. Plant Journal 16, 673–680.

Kubo, T., 2000. Development of low-glutelin rice by *Agrobacterium*-mediated genetic transformation with an antisense gene construct. Biotech 00/12, Topic 10: Potential of foods from which unfavourable components have been removed. In: Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, Geneva, Switzerland, 29 May- 2 June 2000. Food and Agriculture Organisation of the United Nations, Rome. ftp://ftp.fao.org/es/esn/food/Bio-12.pdf.

Kumar, S., Fladung, M., 2000. Transgene repeats in aspen: molecular characterisation suggests simultaneous integration of independent T-DNAs into receptive hotspots in the host genome. Molecular and General Genetics 264, 20–28.

Laufs, P., Autran, D., Traas, J., 1999. A chromosomal paracentric inversion associated with T-DNA integration in Arabidospis. Plant Journal 18, 131–139.

Lavrik, P.B., Bartnicki, D.E., Feldman, J., Hammond, B.G., Keck, P.J., Love, S.L., Naylor, M.W., Rogan, G.J., Sims, S.R., Fuchs, R.L., 1995. Safety assessment of potatoes resistant to colorado potato beetle. In: Engel, K.-H., Takeoka, G.R., Teranishi, R. (Eds.), Genetically Modified Foods- Safety Aspects, ACS Symposium Series 605. American Chemical Society, Washington DC, pp. 148–159.

Le Gall, G., DuPont, M.S., Mellon, F.A., Davis, A.L., Collins, G.J., Verhoeyen, M.E., Colquhoun, I.J., 2003a. Characterisation and content of flavonoid glycosides in genetically modified tomato (Lycopersicon esculentum) fruits. Journal of Agricultural and Food Chemistry 51, 2438–2446.

Le Gall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J., Verhoeyen, M.E., 2003b. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. Journal of Agricultural and Food Chemistry 51, 2447–2456.

Liang, P., Pardee, A.B., 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257, 967–971.

Lichten, M., Goldman, A.S.H., 1995. Meiotic recombination hot spots. Annual Review of Genetics 29, 423–444.

Lindon, J.C., Holmes, E., Nicholson, J.K., 2001. Pattern recognition methods and applications in biomedical magnetic resonance. Progress in NMR Spectroscopy 39, 1–40.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates, J.R., 1999. Direct analysis of protein complexes using mass spectrometry. Nature Biotechnology 17, 676–682.

Liu, Y.G., Mitsukawa, N., Oosumi, T., Whittier, R.F., 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant Journal 8, 457–463.

Lockhart, D.J., Winzeler, E.A., 2000. Genomics, gene expression and DNA arrays. Nature 405, 827–836.

Lommen, A., Weseman, J.M., Smith, G.O., Noteborn, H.P.J.M., 1998. On the detection of environmental effects on complex matrices combining off-line liquid chromatography and H-1-NMR. Biodegradation 9, 513–525.

Lopez, M.F., Kristal, B.S., Chernokalskaya, E., Lazarev, A., Shestopalov, A.I., Bogdanova, A., Robinson, M., 2000. High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation. Electrophoresis 21, 3427–3440.

MAFF, 1996. Inherent Natural Toxicants in Food. The Fifty-first Report of the Steering Group on Chemical Aspects of Food Surveillance, Paper No. 51. Ministry of Agriculture Fisheries and Food, London.

Matsuhara, S., Jingu, F., Takahashi, T., Komeda, Y., 2000. Heat-shock tagging: a simple method for expression and isolation of plant genome DNA flanked by T-DNA insertions. Plant Journal 22, 79–86.

Matzke, A.J.M., Matzke, M.A., 1998. Position effects and epigenetic silencing of plant transgenes. Current Opinion in Plant Biology 1, 142–148.

Mayerhofer, R.Z., Koncz-Kalman, C., Nawrath, G., Bakkeren, A., Cramer, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B., Koncz, C., 1991. T-DNA integration: a mode of illegitimate recombination in plants. EMBO Journal 10, 697–704.

Max Planck Institute, 2003. Metabolite Mass Spectra Library. Max Planck Institute of Molecular Plant Physiology, Golm. www.mpimp-golm.mpg.de/mms-library/index-e.html.

Mellon, F., Self, R., Startin, J.R., 2000. Mass Spectrometry of Natural Substances in Food. Royal Society of Chemistry, Cambridge.

Merril, C.R., 1990. Silver staining of proteins and DNA. Nature 343, 779–780.

Millstone, E., Brunner, E., Mayer, S., 1999. Beyond substantial equivalence. Nature 401, 525–526.

Molnár-Perl, I., 1999. Simultaneous quantitation of acids and sugars by chromatography: gas or high-performance liquid chromatography? Journal of Chromatography A 845, 181–195.

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. Bioscience Biotechnology and Biochemistry 63, 314–318.

Monro, J.A., James, K.A.C., Conner, A.J., 1993. Comparative Nutritional Evaluation of a Transgenic Herbicide-Resistant Potato and the Parent Cultivar. FoodInfo Report No.6. New Zealand Institute for Crop & Food Research, Christchurch.

Mosely, M.A., 2001. Current trends in differential expression proteomics: isotopically coded tags. Trends in Biotechnology 19, S10–S16.

Munkvold, G.P., Hellmich, R.L., Showers, W.B., 1997. Reduced fusarium ear rot and symptomless infection in kernels of maize genetically engineered for European Corn borer resistance. Phytopathology 87, 1071–1077.

Munkvold, G.P., Hellmich, R.L., Rice, L.G., 1999. Comparison of fumonisin concentration in kernels of transgenic Bt maize hybrids and nontransgenic hybrids. Plant Disease 83, 130. –138.

Murray, F., Llewellyn, D., McFadden, H., Last, D., Dennis, E.S., Peacock, W.J., 1999. Expression of the Talaromyces flavus glucose oxidase gene in cotton and tobacco reduces fungal infection, but is also phytotoxic. Mol. Breed 5, 219–232.

Nanogen, 2003. Core Technology. Nanogen, San Diego. http://www.nanogen.com/technology/core_technology.htm.

NETTOX, 1998a. List of Food Plant Toxicants, Information on Inherent Food Plant Toxicants, Report No. 1. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998b. List of Food Plants, Information on Inherent Food Plant Toxicants, Report No. 2. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998c. Proposal for the Structure and Organisation of a Future European Database, Information on Inherent Food Plant

Toxicants, Report No. 3. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998d. Compilation of Consumption Data, Information on Inherent Food Plant Toxicants, Report No. 4. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998e. Toxicological Guidelines for Safety Assessment of Plant Toxicants, Information on Inherent Food Plant Toxicants, Report No. 5. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998f. List of Food Plants Prioritised for Inclusion in a Future European Database, Information on Inherent Food Plant Toxicants, Report No. 6. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998g. Report of the NETTOX Graz Seminar on Inherent Food Plant Toxicants, Information on Inherent Food Plant Toxicants, Report No. 7. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998h. Report from the NETTOX Exchange of Scientists, Information on Inherent Food Plant Toxicants, Report No. 8. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

NETTOX, 1998i. Directory of Scientists, Information on Inherent Food Plant Toxicants, Report No. 9. NETTOX Plant/Food Composition/Human Exposure Research Group. Danish Veterinary and Food Administration, Søborg.

Nida, D.L., Patzer, S., Harvey, P., Stipanovic, R., Wood, R., Fuchs, R.L., 1996. Glyphosate-tolerant cotton: the composition of the cottonseed is equivalent to that of conventional cottonseed. Journal of Agricultural and Food Chemistry 44, 1967–1974.

Nielsen, N.-P.V., Carstensen, J.M., Smedsgaard, J., 1998. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. Journal of Chromatography A 805, 17–35.

Nielsen, N.-P.V., Smedsgaard, J., Frisvad, J.C., 1999. Full second-order chromatographic/spectrometric data matrices for automated sample identification and component analysis by non-data-reducing image analysis. Analytical Chemistry 71, 727–735.

Nolan, J.P., Sklar, L.A., 2002. Suspension array technology: evolution of the flat-array paradigm. Trends in Biotechnology 201, 9–12.

Nordic Council of Ministers, 1998. Safety Assessment of Novel Food Plants. Chemical Analytical Approaches to the Establishment of Substantial Equivalence. TemaNord 1998:591. Nordic Council of Ministers, Copenhagen.

Noteborn, H.P.J.M., Bienenmann-Ploum, M.E., van den Berg, J.H.J., Alink, G.M., Zolla, L., Reynaerts, A., Pensa, M., Kuiper, H., 1995. Safety assessment of the *Bacillus thuringiensis* insecticidal crystal protein CRYIA(b) expressed in transgenic tomatoes. In: Engel, K.-H., Takeoka, G.R., Teranishi, R. (Eds.), Genetically Modified Foods—Safety Aspects, ACS Symposium Series 605. American Chemical Society, Washington DC, pp. 134–147.

Noteborn, H.P.J.M., Lommen, A., van der Jagt, R.C., Weseman, J.M., 2000. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. Journal of Biotechnology 77, 103–114.

OECD, 1993. Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles. Organisation for Economic Cooperation and Development, Paris. http://www.oecd.org/pdf/M00033000/M00033002.pdf.

OECD, 2001a. Consensus Document on Key Nutrient and Key Toxicants in Low Erucic Acid Rapeseed (Canola), Series on the Safety of Novel Foods and Feeds No. 1., ENV/JM/MONO(2001)13. Organisation for Economic Cooperation and Development, Paris. http://www.olis.oecd.org/olis/2001doc.nsf/LinkTo/env-jm-mono(2001)13.

OECD, 2001b. Consensus Document on Compositional Considerations for New Varieties of Soybean: Key Food and Feed Nutrients and Anti-Nutrients, Series on the Safety of Novel Foods and Feeds No. 2, ENV/JM/MONO(2001)15. Organisation for Economic Cooperation and Development, Paris. http://www.olis.oecd.org/olis/2001doc.nsf/LinkTo/env-jm-mono(2001)15.

OECD, 2002a. Consensus Document on Compositional Considerations for New Varieties of Sugar Beet: Key Food and Feed Nutrients and Antinutrients. Series on the Safety of Novel Foods and Feeds No. 3, ENV/JM/MONO(2002)4. Organisation for Economic Cooperation and Development, Paris. http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/env-jm-mono(2002)4.

OECD, 2002b. Consensus Document 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, ENV/JM/MONO(2002)5. Organisation for Economic Cooperation and Development, Paris. http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/env-jm-mono(2002)5.

OECD, 2002c. Consensus Document 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, ENV/JM/MONO(2002)25. Organisation for Economic Cooperation and Development, Paris. http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/env-jm-mono(2002)25.

OECD, 2003. Consensus Document on Compositional Considerations for New Varieties of Bread Wheat (*Triticum aestivum*): Key Food and Feed Nutrients, Anti-Nutrients and Toxicants, Series on the Safety of Novel Foods and Feeds No. 7, ENV/JM/MONO(2003) http://www.olis.oecd.org/olis/2003doc.nsf/LinkTo/env-jm-mono(2003)7.

Ohba, T., Yioshioka, Y., Machida, C., Machida, Y., 1995. DNA rearrangements associated with the integration of T-DNA in tobacco: an example of multiple duplications of DNA around the integration target. Plant Journal 7, 157–164.

Parinov, S., Sevugan, M., Ye, D., Yang, W.C., Kumaran, M., Sundaresan, V., 1999. Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in Arabidopsis. Plant Cell 11, 2263–2270.

Patton, W.F., 2000. A thousand points of light: The application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. Electrophoresis 21, 1123–1144.

Pawlowski, W.P., Somers, D.A., 1996. Transgene inheritance in plants genetically engineered by microprojectile bombardment. Molecular Biotechnology 6, 17–30.

Pedersen, C., Zimmy, J., Becker, D., Jahne-Gartner, A., Lorz, H., 1997. Localization of introduced genes on the chromosomes of transgenic barley, wheat, triticale by fluorescence in situ hybridization. Theoretical and Applied Genetics 94, 749–757.

Peltier, J.B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., van Wijk, K.J., 2000. Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12, 319–341.

Petersen, A., Grobe, K., Lindner, B., Schlaak, M., Becker, W.M., 1997. Comparison of natural and recombinant isoforms of grass pollen allergens. Electrophoresis 18, 819–825.

Posch, A., Chen, Z., Dunn, M.J., Wheeler, C.H., Petersen, A., Leubner-Metzger, G., Baur, X., 1997. Latex allergen database. Electrophoresis 18, 2803–2810.

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., Van Dam, K., Oliver, S.G., 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nature Biotechnology 19, 45–50.

Rabilloud, T., 2000. Detecting proteins separated by 2-D gel electrophoresis. Analytical Chemistry 72, 48. A-55A.

Redenbaugh, K., Lindemann, J., Malyi, L., 1995. Application of the principles of substantial equivalence in the safety evaluation of

FLAVR SAVR tomato, BXN cotton and oil-modified rapeseed. In: Application of the Principles of Substantial Equivalence to the Safety Evaluation of Foods or Food Components from Plants Derived by Modern Biotechnology, Report of A WHO Workshop, WHO/FNU/FOS/95.1. World Health Organisation, Geneva, pp. 37-50.

Rees, H., 1993. Genetic systems, recombination and variability. In: Hayward, M.D., Bosemark, R.O., Romagosa, I. (Eds.), Plant Breeding: Principles and Prospects. Chapman and Hall, London.

Richter, T.E., Pryor, T.J., Bennetzen, J.L., Hulbert, S.H., 1995. New rust resistance specificities associated with recombination in the Rp1 complex in maize. Genetics 141, 373–381.

Riley, R., Chapman, V., Johnson, R., 1968. The incorporation of alien disease resistance in wheat by genetic interference with the regulation of meiotic chromosome synapsis. Genetical Research 12, 199–219.

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

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

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

Rotig, A., Chauveau, M., 1987. Polypeptide composition of potato tuber mitochondria and submitochondrial fractions analyzed by one- and two-dimensional gel electrophoresis. Plant Physiology and Biochemistry 25, 11–20.

Salomon, S., Puchta, H., 1998. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO Journal 17, 6086–6095.

Sanders, P.R., Lee, T.C., Groth, M.E., Astwood, J.D., Fuchs, R.L., 1998. Safety assessment of insect-protected corn. In: Thomas, J.A. (Ed.), Biotechnology and Safety Assessment 2nd ed. Taylor and Francis, Philadelphia, pp. 241–256.

Santoni, V., Doumas, P., Rouquie, D., Mansion, M., Rabilloud, T., Rossignol, M., 1999. Large scale characterization of plant plasma membrane proteins. Biochimie 81, 655–661.

Santoni, V., Molloy, M., Rabilloud, T., 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis 2, 1054–1070.

Santoni, V., Rouquie, D., Doumas, P., Mansion, M., Boutry, M., Degand, H., Dupree, P., Packman, L., Sherrier, J., Prime, T., Bauw, G., Posada, E., Rouze, P., Dehais, P., Sahnoun, I., Barlier, I., Rossignol, M., 1998. Use of a proteome strategy for tagging proteins present at the plasma membrane. Plant Journal 16, 633–641.

Sauter, H., Lauer, M., Fritsch, H., 1991. Metabolic profiling of plants—a new diagnostic technique. In: Baker, D.R., Fenyes, J.G. (Eds.), Synthesis and Chemistry of Agrochemicals II, ACS Symposium Series 443. American Chemical Society, Washington D.C, pp. 288–299.

Sawasaki, T., Takahashi, M., Goshima, N., Morikawa, H., 1998. Structures of transgene loci in transgenic Arabidopsis plants obtained by particle bombardment: junction regions can bind to nuclear matrices. Gene 218, 27–35.

Schnable, P.S., Hsia, A.-P., Nikolau, B.J., 1998. Genetic recombination in plants. Current Opinion in Plant Biology 1, 123–129.

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

Smedsgaard, J., Frisvad, J.C., 1996. Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts. Journal of Microbiological Methods 25, 5–17.

Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M., Nishioka, T., 2002. Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionisation mass spectrometry. Analytical Chemistry 74, 2233–2239.

Sommerville, C., Sommerville, S., 1999. Plant functional genomics. Science 285, 380–383.

Spertini, D., Beliveau, C., Bellemare, G., 1999. Screening of transgenic plants by amplification of unknown genomic DNA flanking T-DNA. BioTechniques 27, 308–314.

Stein, S.E., 1999. An integrated method for spectrum extraction and compound identification from gas chromatography/ mass spectrometry data. Journal of the American Society for Mass Spectrometry 10, 770–781.

Steinberg, T.H., Jones, L.J., Haugland, R.P., Singer, V.L., 1996. SYPRO Orange and SYPRO Red protein gel stains: One-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. Analytical Biochemistry 239, 223–237.

Szostak, J.W., Orr-Weaver, T.L., Rothestein, R.J., Stahl, F.W., 1983. The double strand-break repair model for recombination. Cell 33, 25–35.

Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Roder, M.S., Wing, R.A., Wu, W., Young, N.D., 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics 132, 1141–1160.

ten Hoopen, R., Montijn, B.M., Veuskens, J.T., Oud, O.J., Nanninga, N., 1999. The spatial localization of T-DNA insertions in petunia interphase nuclei: consequences for chromosome organization and transgene insertion sites. Chromosome Research 7, 611–623.

Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossignol, M., Santoni, V., de Vienne, D., Zivy, M., 1999. Proteomics for genetic and physiological studies in plants. Electrophoresis 20, 2013–2026.

Thomas, C.M., Jones, D.A., English, J.J., Carroll, B.J., Bennetzen, J.L., Harrison, K., Burbidge, A., Bishop, G.J., Jones, J.D., 1994. Analysis of the chromosomal distribution of transposon-carrying T-DNAs in tomato using the inverse polymerase chain reaction. Molecular and General Genetics 242, 573–585.

Thomas, W.T.B., Baird, E., Fuller, J.D., Lawrence, P., Young, G.R., Russell, J., Ramsay, L., Waugh, R., Powell, W., 1998. Identification of a QTL decreasing yield in barley linked to Mlo powdery mildew resistance. Molecular Breeding 4, 381–393.

Tolstikov, V.V., Fiehn, O., 2002. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. Analytical Biochemistry 301, 298–307.

Turk, S.C.H.J., Smeekens, S.C.M., 1999. Genetic modification of plant carbohydrate metabolism. In: Chopra, V.L., Malik, V.S., Bhat, S.R. (Eds.), Applied Plant Biotechnology. Science Publishers, Enfield, pp. 71–100.

Tweeddale, H., Notley-McRobb, L., Ferenci, T., 1998. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ('metabolome') analysis. Journal of Bacteriology 180, 5109–5116.

Ünlü, M., 1999. Difference gel electrophoresis. Biochemical Society Transactions 27, 547–549.

Ünlu, M., Morgan, M.E., Minden, J.S., 1997. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071–2077.

van Hal, N.L.W., Vorst, O., van Houwelingen, A.M.M.L., Kok, E.J., Peijnenburg, A., Aharoni, A., van Tunen, A.J., Keijer, J., 2000. The application of DNA microarrays in gene expression analysis. Journal of Biotechnology 78, 271–280.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science 270, 484–487.

Vigneau-Callahan, K.E., Shestopalov, A.I., Milbury, P.E., Matson, W.R., Kristal, B.S., 2001. Characterization of diet dependent meta-

bolic serotypes: analytical and biological variability issues in rats. Journal of Nutrition 131, 924. S-932S.

Villegas, E., Vasal, S.K., Bjarnason, M., 1992. Quality Protein Maize-what it is and how it was developed. In: Mertz, E.T. (Ed.), Quality Protein Maize. American Association of Cereal Chemists, St Paul, pp. 27–48.

Walbot, V., 1992. Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. Annual Review of Plant Physiology and Plant Molecular Biology 43, 49–82.

Ward, D.V., Zambryski, P.C., 2001. The six functions of Agrobacterium virE2. Proceedings of the National Academy of Sciences U.S.A 98, 385–386.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Frankhauser, C., Ferrandiz, C., kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.-Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., Chory, J., 2000. Activation tagging in Arabidopsis. Plant Physiology 122, 1003–1013.

Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D., McClelland, M., 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Research 2019, 4970–4976.

WHO, 1995. Application of the Principles of Substantial Equivalence to the Safety Evaluation of Foods or Food Components from Plants Derived by Modern Biotechnology. Report of a WHO Workshop. World Health Organisation, Geneva.

Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S.J., Larsen, P.M., Görg, A., 2000. Towards higher resolution: two-dimensional electrophoresis of Saccharomyces cerevisiae proteins using overlapping narrow immobilized pH gradients. Electrophoresis 21, 2610–2616.

Winkler, R.G., Frank, M.R., Galbraith, D.W., Feyereisen, R., Feldmann, K.A., 1998. Systematic reverse genetics of transfer-DNA-tagged lines of Arabidopsis. Plant Physiology 118, 743–750.

Wolfender, J.-L., Ndjoko, K., Hostettman, K., 2001. The potential of LC-NMR in phytochemical analysis. Phytochemical Analysis 12, 2–22.

Xu, X., Hsia, A.P., Zhang, L., Nikolau, B.J., Schnable, P.S., 1995. Meiotic recombination break points resolve at high rates at the 5′ end of a maize coding sequence. Plant Cell 12, 2151–2161.

Ye, X., Al Babili, S., Kloeti, A., Zhang, J., Lucca, P., Beyer, P., Potrykus, I., 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287, 303–305.