TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Allgemeine Lebensmitteltechnologie

Applicability of proteomics to assess effects due to genetic engineering in the context of natural variability using *Arabidopsis thaliana* as a model organism

Martin Christian Rübelt

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

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Gerhard Wenzel
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Karl-Heinz Engel
	2. apl. Prof. Dr. Angelika Görg
	3. apl. Prof. Dr. Klaus-Dieter Jany, Universität Stuttgart

Die Dissertation wurde am 17.12.2004 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 21.01.2005 angenommen.

Acknowledgements

I am most grateful to the members of my dissertation committee: Professor Dr. Klaus-Dieter Jany for inspiring my interest in the topic, helping to make possible a scholarship at Monsanto Company, ensuring that I stay on the right track with personal interest and encouragement at every step of the work; Professor Dr. Karl-Heinz Engel for valuable guidance, strategic direction, and critical review of my thesis; Professor Dr. Angela Görg for kindly agreeing to serve as a committee member and for evaluating my work in light of her extensive expertise in proteomics; and Professor Dr. Gerhard Wenzel for chairmaning the dissertation committee.

I am also deeply indebted to my advisor team at the Monsanto Company, St. Louis: my supervisor James Astwood, PhD, for consistent confidence in my work, his enthusiasm in my project and regular interest in its progress, as well as useful advice on thesis related issues; Markus Lipp, PhD, for his professional input on method validation and on other crucial aspects of the work, and stimulating my creativity and motivation with his friendship and support; Tracey Reynolds, PhD, for dedicating her time and effort to review countless versions of my thesis, providing insightful comments on content, style, and presentation (in other words, for teaching me how to write), and for the overall unfaltering interest in the progress of my work; and Cherian George, PhD, for assistance in the early stages of this work.

The completion of this thesis would also not be possible without the following individuals: Margaret Nemeth, PhD, who provided advice on statistics and experimental design, and always patiently explaining complex concepts to an incredibly thickheaded individual like me; Nancy Leimgruber, who introduced me to the dark art of 2DE and always welcomed my presence in her laboratory; Richard Thoma, who deepened my skills in mass spectroscopy; Bala Karunanandaa, PhD, who opened my eyes to the model plant *Arabidopsis thaliana*, which I mistakenly considered "an uninteresting weed"; Kanthasamy (Karu) Karunanandaa, PhD, who kindly took me under his wing and brought me in touch with various teams and people at Monsanto, which proved crucial for the successful completion of my work; Daniel Free, who assisted me with the cultivation of the *Arabidopsis* plants; Professor Dean DellaPenna, PhD, of Michigan State University and Jon Schmuke, who provided me with the transgenic *Arabidopsis* lines and characterization data; and all the employees of Regulatory Science, Crop Analytics, Biotechnology, and Library Services at Monsanto Company, St. Louis, who contributed in numerous ways to this work and made me feel comfortable during its various stages.

Finally, I am grateful to Monsanto Company for providing generous financial assistance and access to top-notch facilities for my research in the form of a three-year scholarship. This amazing opportunity has given me incredible inspiration for my future endeavors as a scientist and development as a person.

This thesis is dedicated to my parents

Dr. Christian Rübelt and Doris Rübelt.

Abbreviations

2DE	two-dimensional gel electrophoresis
ADI	acceptable daily intake
BSA	Bovine Serum Albumin
CBB	Coomassie Brilliant Blue
cDNA	complementary DNA
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CV	coefficient of variation
CYMAL	cyclohexyl-b-D-maltoside
DNA	deoxyribonucleic acid
DNase I	deoxyribonuclease I
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid
FAO	Food and Agriculture Organization of the United Nations
FDA	Food and Drug Administration
FOS-Choline-10	n-dodecylphosphosphocholine
fw	fresh weight
g	gravitational acceleration
ID	identification
IEF	isoelectric focusing
IOD	integrated optical density
IPG	immobilized pH gradient
kD	kilo Dalton
MALDI	matrix-assisted-desorption-ionization
MS	mass spectrometry
Mw	molecular weight
N	sample size
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
pI	isoelectric point
PMF	peptide mass fingerprinting
PMSF	phenylmethanesulfonyl fluoride
R2	coefficient of determination for line regression
RNA	ribonucleic acid
RNase A	ribonuclease A.
rpm	revolution per minute
RT	room temperature
SD	standard deviation
SDS	sodium dodecyl sulfate
TBP	tributylphosphine
TCA	trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
T-DNA	transfer DNA
TOF	time-of-flight
Vh	volt-hours
WHO	World Health Organization

1 Introduction and aim of study

Crops developed through modern biotechnology, i.e. genetic engineering, have been marketed and used by farmers since the mid-1990s. In the eight-year period from 1996 to 2003, there has been a 40-fold (1.7 to 67.7 million hectares) increase in transgenic crop cultivation. Seven million farmers in 18 countries grow crops produced using recombinant DNA techniques. Herbicide tolerance continues to be the major trait with 73 percent of the global area of transgenic crops, followed by insect resistance and the stacked genes of herbicide tolerance and insect resistance with 18 and 8%, respectively [1].

Although different approaches are used to regulate products of biotechnology worldwide, all regulatory systems have in common the ultimate goal that the genetically modified organism (GMO) must be safe. In order to identify any potential hazard to the health of humans, animals, or the environment, GMOs have to undergo a risk and safety assessment. The core of the risk and safety assessment is the comparison of the GMO with a traditional counterpart, which has an established history of safe use. This approach, known as the concept of substantial equivalence, was developed by the OECD Working Group in 1993 [2]. Any identified difference between the GMO and its counterpart is then the subject of the safety assessment, which can include nutritional, toxicological, and immunological testing. The comparative analysis takes into account agronomic, phenotypic, genetic, and compositional aspects. The substantial equivalence approach is a very powerful and flexible tool for safety assessment and has worked well in the safety assessment of genetically modified crops [3]. Over 50 biotech crops have been considered to be as safe and nutritious as their conventional counterparts [4]. Worldwide, there has been no verifiable adverse effect resulting from the cultivation and consumption of products from GM crops [5, 6] after eight years of production.

However, concerns exist that the spectrum of compounds used for the comparison is not comprehensive enough and is based solely on experience gained from analysis of food crops obtained via conventional breeding. Thus, unintended effects altering the levels of toxicants or anti-nutrients might remain undetected [7, 8]. This could be of particular importance for GM crop plants of the second generation, where the genetic modifications become more complex, such as the introduction of a new metabolic pathway in the so-called Golden Rice [9]. The insertion of multiple genes involved in different pathways will possibly increase the likelihood for unintended and unexpected effects, either resulting from insertion or pleiotropy [10, 11].

In order to improve the probability of detecting unintended effects, profiling techniques are considered as complementary analytical tools [11-13]. Profiling technologies such as transcriptomics, proteomics, and metabolomics allow the simultaneous measurement and comparison of thousands of plant components without prior knowledge of their identity. These 'unbiased' approaches are considered to facilitate a more holistic analysis. The development and evaluation of "unbiased" non-targeted techniques providing additional information on unintended changes that might remain undetected using targeted approaches has been discussed by international advisory bodies since 1997 [11, 14].

Aim of study

The aim of this study was to explore the applicability of proteomics to investigate changes in the plant proteome due to natural variability and genetic engineering. *Arabidopsis (A.) thaliana* was used as a model organism. The three phases of the study are summarized below.

In the first phase, a reproducible proteomics method based on two-dimensional gel electrophoresis (2DE) will be established for the qualitative and quantitative analysis of the proteome of *A. thaliana* seeds and validated for repeatability, sensitivity, and linearity. In the second phase, the validated method will be used to determine the qualitative and quantitative changes in protein profiles within a given *A. thaliana* ecotype and among a set of *A. thaliana* ecotypes. These data will be used to ascertain the biological and genetic variation in protein profiles within and between *A. thaliana* ecotypes. In the final phase, insertional and pleiotropic effects on the proteome due to genetic engineering will be investigated using transgenic *A. thaliana* lines representing two different strategies: (i) transgene has no impact on an endogenous metabolic pathway and (ii) transgene has impact on an endogenous metabolic pathway. The data generated from the natural variability study (phase two) will be used to interpret potential differences between these transgenic lines and their non-transgenic parental lines.

These studies will help to understand the advantages and limitations of 2DE based proteomics as an analytical tool for the investigation of changes due to genetic engineering of plants in the context of natural variability. This work will also serve as a model for the evaluation of differences in the proteome of transgenic and conventional plants and the potential use of proteomics as a complementary tool to the current safety assessment program for genetically engineered plants.

2 Background

2.1 Biotechnology

The term "biotechnology" was first coined in 1919 by the Hungarian engineer Karl Ereky to describe the derivation of products from raw materials with the help of living organisms [15]. The United Nations Convention on Biological Diversity [16] further defined 'biotechnology' as the use of "biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use." As such, biotechnology has been practiced by human society since the beginning of recorded history in such activities as making bread (6000 BC), cheese (4000 BC), beer (6000 BC), or breeding domestic animals or food crops.

Traditional biotechnology

Using traditional selective plant breeding techniques, scientists have been working to improve plants for human benefit for hundreds of years. Selective breeding includes techniques such as hybridization (to generate 'hybrid vigor'), alteration of ploidy (haploids, tetraploids, etc.), induction of mutations through irradiation or chemical mutation agents, somaclonal selection, embryo rescue (enabling 'wide crosses'), anther culture, protoplast fusion, and marker-assisted breeding. All of our agricultural crop varieties have been originally derived from wild plants by domestication and selective breeding using traditional methods. Corn hybrids are a good example for improvement in grain yield by traditional breeding techniques. After the introduction of the single cross F_1 hybrid seed, corn yields have increased rapidly in the United States [17]. Also, crosses between members of two different genera have been performed and commercialized by traditional breeding. *Triticale*, for example, is a crop species resulting from a plant breeder's cross between wheat (*Triticum*) and rye (*Secale*). Fertile lines are achieved by doubling the chromosomes of the sterile hybrid resulting from the cross [18]. However, because of the lack of control over the gene mixing that occurs during sexual reproduction, conventional plant breeding remains time-intensive and, to a certain extent, a trial-and-error approach.

Modern biotechnology

Modern biotechnology was born with the development of recombinant DNA (rDNA) technology by Cohen and Boyer in 1973 [19, 20]. Recombinant DNA technology has

enabled scientists to add new beneficial traits to plants, animals, and microorganisms for food production. It involves the isolation, manipulation, and introduction of specific gene(s) encoding desired trait(s) into target organisms. The most frequently used methods to introduce DNA into a plant are *Agrobacterium*–mediated transformation [21, 22] and particle bombardment method ('gene-gun') [23].

The *Agrobacterium*-based method uses the ability of the soil bacterium *Agrobacterium tumefaciens* to transfer part of its genetic information (T-DNA) on the Ti (tumor inducing) plasmid to plant cells. This results in the integration of the T-DNA into the plant genome. For genetic engineering of plants, the T-DNA sequence of the Ti-plasmid is replaced by the gene of interest. *Agrobacterium*-mediated transformation is the method of choice for dicots, such as soybean, tomato, potato, canola, and pea [24-26]. Recent development has made it possible to apply this technology also to monocots, such as maize [27], rice [28], and wheat [29]. For the particle bombardment method, plant embryos are bombarded with small, highly accelerated gold or tungsten particles that are coated with DNA containing the gene of interest. In this manner, transgenes can be delivered into the cell's genome, and transgenic plants can then be regenerated from these cells. This method has a low transformation efficiency. The particle bombardment method has been successfully used for monocots, e.g. maize, rice, barley, and wheat [30-33].

At this time, either methods results in random integration of one or more copies of the DNA sequence into the genome of the plant. Although both methods have been successfully employed for the transformation of plants, *Agrobacterium*-mediated transformation has now become the method of choice. Its success is due to the versatility of crop species that can be utilized and to higher transformation efficiency. Plants and their products developed through modern biotechnology are referred to as genetically modified (GM), genetically engineered (GE), transgenic, biotech, or recombinant.

Genetically modified food crops

The first commercialized genetically modified food crop was the Flavr Savr™ tomato developed by Calgene, and it was deemed as safe by the FDA in 1994 [34]. *Agrobacterium*-mediated transformation was used to introduce a polygalacturonase (PG) gene in the antisense direction. The transgene product suppresses the production of the endogenous PG enzyme, which is associated with the breakdown of pectin in tomato fruit. The transgenic

variety ripens normally but experiences less pectin degradation and, therefore, has increased thickness and consistency, which benefits all stages of harvesting and processing.

Since the introduction of the Flavr Savr™ tomato in 1994, there have been over 100 transgenic approved globally for food consumption [24, 35]. The major transgenic crops are herbicide-tolerant soybean (61 per cent of total acreage in 2003), insect- and/or herbicide-tolerant maize (23 per cent), insect- and/or herbicide-tolerant cotton (11 per cent), and herbicide-tolerant canola (5 per cent) [1]. In addition, crops like chicory, flax (linseed), papaya, potato, rice, squash, sugar beet, and tomato have been genetically modified to contain traits such as resistance to herbicides, tolerance to insects and viruses, delayed ripening, and male sterility [35]. These crops, with enhanced agronomic traits (input traits), are considered to be the first generation GM crops. They help farmers by reducing production costs or increasing yield. The GM crops also provide indirect benefits for consumers, such as stable or declining food prices, by increasing agricultural productivity and safer food and feed supply through reduced mycotoxin levels [1, 36].

Second generation biotech products are currently being developed that will provide direct benefits for consumers (output traits) in the form of enhanced nutritional or health-promoting characteristics, such as Golden Rice. Rice is a major staple food for hundreds of millions of people but it lacks several essential nutrients such as beta-carotine, the precursor of vitamin A [37]. Vitamin A deficiency is one of the leading causes of night blindness in humans and has also been correlated with increased occurrence of several diseases such as diarrhea, respiratory ailments, tuberculosis, malaria, and ear infections [38]. According to the World Health Organization (WHO), around 2.8 million children under five years of age currently exhibit a severe clinical manifestation of vitamin A deficiency known as xerophthalmia [39]. The idea of Golden Rice is the fortification of rice with provitamin A through plant breeding. Therefore, Potrykus et al. [37, 40] established a biosynthetic pathway *de novo* in rice endosperm, enabling the accumulation of provitamin A. This was achieved using rDNA technology. It is thought that after further improvements, Golden Rice may alleviate vitamin A deficiency. Further examples of second-generation crops in the development stage or commercialized are listed in Table 1.

Table 1 Nutritionally enhanced GM products [41]

Product	Benefit	References
Enhanced vitamins/minerals		
Golden Rice	Enhanced level of beta-carotene	[9]
Golden Mustard	Enhanced level of beta-carotene	[42, 43]
High iron rice	Enhanced level of iron	[44]
Enhanced vegetable oils		
High laurate canola*	Enhanced levels of lauric acid	[45]
High oleic soybean oil*	Improved fatty acid profile	[46]
High stearate soybean oil	Reduced need for hydrogenation and generation of trans fatty acids	[47]
High omega-3 vegetable oil	Enhanced level of stearidonic acid	[48]
Enhanced protein		
Protein-enhanced potato	Enhanced nutritional quality of potato protein	[49]
Enhanced protein rice	Enhanced nutritional quality of potato protein	[50-52]
Reduced undesirable components		
Reduced allergen soy	Reduced allergenic P34 protein	[53]
Reduced polyphenol oxidase in potatoes	Reduced bruising upon exposure to air	[54, 55]
Reduced raffinose soy	Improved digestibility	[56]
Feed grain improvements		
Maize and soy with improved essential amino acid levels	Reduced nutrients needed to be added to feed	[38, 57]
Grains with fungal phytase	Increased bioavailability of phosphorous, reduced bio-waste	[38]

* Products are commercialized.

GM Crops of the third generation will be used as 'bio-reactors' to produce enzymes, antibodies, recombinant proteins, or pharmaceutically active ingredients [58, 59]. The potential of molecular farming in plants is demonstrated by the commercial success of plant-derived avidin and β-glucuronidase at a lower cost than native proteins [60].

2.2 Regulation of genetically modified organisms (GMOs)

Worldwide, various regulatory approaches based on varying philosophical perceptions are used to regulate GMOs and food derived from GMOs. They differ significantly in intent (product vs. process), legislative basis (existing vs. technology-specific laws), transparency, and location of decision-making authority (institutional vs. political) [61].

The regulatory approach of the USA as described under the Coordinated Framework for Regulation of Biotechnology (1986) [62], for example, is based on the principle that techniques of biotechnology are not inherently risky and that biotechnology should not be regulated as a process, but rather that the product of biotechnology should be regulated in the same way as products of other technologies. As such, the U.S. government has implemented no new legislation or risk assessment procedures to regulate them. The existing laws are considered adequate for the oversight of biotechnology products. For example, food or food additives derived from rDNA technology are subject to existing requirements of the Federal Food, Drug, and Cosmetic Act (FFDCA) governing food additives and substances that are 'generally recognized as safe' (GRAS) and, therefore, need no FDA pre-market approval if they are substantially equivalent to existing food or food additives already on the market. The FDA has recently proposed to move to a mandatory pre-market notification policy for which companies must submit a scientific and regulatory assessment of the GM food 120 days before the GM food is marketed [63]. However, this has yet to be enacted.

In Europe, the use of recombinant DNA technology as such requires regulation [64] and pre-market approval of genetically engineered food. In addition, labeling of GM foodstuff is mandatory in the European Union (EU), not in the USA. Notwithstanding the ideological and political differences, all regulation systems require that GM foods not present any risk to human or animal health, or the environment. Several international agreements like the revised General Agreement on Tariffs and Trade (GATT) [65] from 1994, the Sanitary and Phytosanitary Measures Agreement (SPS) [66] from 1994, and the Cartagena Protocol [67] in force since September 2003, emphasize this principle and demand that any regulation must be based on science. They also encourage the use of international standards, guidelines and recommendations for the safety assessment of GMOs and GM food. Because of its complexity and rigor, the European regulation framework was chosen as an example to explain the regulation of GMOs in the following section.

2.2.1 EU regulation of genetically modified organism/food

The European Union has regulated genetically modified organisms (GMOs) since the early 1990s, and these regulations have been constantly extended and refined. Genetically modified organisms have to be approved before they can be released into the environment or placed on the market.

EU regulation of the deliberate release of GMOs to the environment

Directive 2001/18/EC [68], applicable since October 17, 2002, amended the earlier Directive 90/220/EEC from April 23, 1990 [69]. The Directive regulates the authorization procedure for the deliberate release of genetically modified organisms into the environment. It also outlines principles for the environmental risk assessment. A GMO or a product containing GMOs has to be evaluated regarding its risk to human health and the environment before it can be released into the environment or placed on the market. That also includes GMOs that are not intended to be cultivated in the EU but imported to the EU for uses in food or industrial processing. The Directive sets the principles for environmental risk assessment and requires mandatory post-market monitoring. Approval is limited to a maximum of ten years. Currently, eight GMOs used for food or feed, have been authorized for the commercial release in the EU according to the former Directive 90/220/EEC. They are listed in Table 2.

Table 2 Food and feed-relevant GMOs approved under Directive 90/220/EEC [70]

Crop/Event	Company	Description	Uses
Canola			
MS1, RF1 => PGS1	Bayer CropScience	Male-sterility, herbicide-tolerant	Import, processing
MS1, RF2 => PGS2	Bayer CropScience	Male-sterility, herbicide-tolerant	
HCN92 (Topas 19/2)	Bayer CropScience	Herbicide-tolerant	Import, processing
Maize			
Bt-176	Syngenta Seeds	Insect-resistant, herbicide-tolerant	Cultivation, import, processing
Bt-11	Syngenta Seeds	Insect-resistant, herbicide-tolerant	Import, processing
MON 810	Monsanto Co.	Insect-resistant	Cultivation, import, processing
T25	Bayer CropScience	Herbicide-tolerant	Cultivation, import, processing
Soybean			
Roundup Ready (GTS 40-3-2)	Monsanto	Herbicide-tolerant	Import, processing

Applications for the authorization procedure under Directive 2001/18/EC have been submitted for 19 food and feed relevant GMOs (Table 3). Because of a 'de facto' moratorium on the use of genetically modified organisms in the EU, no new GMOs have been authorized for planting or use in the EU since October 1998.

Table 3 Food and feed relevant GMOs for which notifications are submitted to the EU Commission under Directive 2001/18/EC as of February 2004 [according to: 71]

Crop/Event	Company	Description	Uses
Beet (fodder and sugar)			
A5/15 Roundup Ready® Fodder beet	DLF-Trifolium, Monsanto, Danisco Seed	Herbicide-tolerant	Cultivation, feed
H7-1 Roundup Ready® Sugar Beet	KWS SAAT AG and Monsanto	Herbicide-tolerant	Cultivation, processing
T9100152 Roundup Ready® Sugar Beet	Monsanto and Syngenta	Herbicide-tolerant	Cultivation, processing
Canola			
GT 73	Monsanto Company	Herbicide-tolerant	Import, feed, processing
Ms8, Rf3	Bayer CropScience	Herbicide-tolerant	Cultivation, import, feed, processing
FALCON, GS40/90pHoe6Ac	Bayer CropScience	Herbicide-tolerant	Cultivation, import
Liberator pHoe6/Ac	Bayer CropScience	Herbicide-tolerant	Cultivation, import
T45	Bayer CropScience	Herbicide-tolerant	Import, feed, processing
Cotton			
281-24-236 3006-210-23	Dow AgroSciences and Agrigenetics	Insect-resistant	Import, processing
531	Monsanto Company	Insect-protected	Cultivation, import, processing
1445 Roundup Ready® Cotton	Monsanto Company	Insect-protected, herbicide-tolerant	Cultivation, import, processing
BXN, 10215 and 10222	Monsanto (Calgene)	Herbicide-tolerant	Import, processing
Maize			
NK603 x MON810	Monsanto Company	Insect-resistant, herbicide-tolerant	Cultivation, import, feed, processing
NK603 Roundup Ready® maize	Monsanto Company	Herbicide-tolerant	Cultivation, import, feed, processing
MON863 x MON810	Monsanto Company	Resistant against certain insect pests	Import, processing
1507	Pioneer/Mycogen Seeds	Insect-resistant, herbicide-tolerant	Cultivation, import, feed, processing
Bt-11	Syngenta Seeds	Insect-resistant and herbicide-tolerant	Cultivation, feed, processing
Rice			
LL 62	Bayer CropScience	Herbicide-tolerant	Import, feed, processing
Soybean			
A 2704-12 / A 5547-127	Bayer CropScience	Herbicide-tolerant	Import

EU regulation of GM foods

Although the Directive 2001/18/EC contains provisions for a risk assessment for human health, its main objective is environmental sustainability (risk assessment) of GMOs. Food and feed derived from these GMOs have to undergo a separate authorization procedure before they can be placed on the market as food or feed product. On April 18, 2004, the Regulation (EC) 1829/2003 on GM Food and Feed [72] was instated and replaced the GM part of Novel Foods Regulation (EC) 258/97 [73] of January 27, 1997. The GM Food and Feed Regulation (Regulation 1829/2003) sets out rules for the labeling and authorization of food and feed products containing, consisting or produced from GMOs. Unlike the Novel Foods Regulation, labeling requirements are now extended to all food products produced from GMOs, including food ingredients, regardless of the detectability of DNA or protein of transgenic origin in the final product. To facilitate accurate labeling, the GM Food and Feed Regulation is supported by Regulation (EC) 1830/2003 [74], which requires that relevant information concerning genetic modification along the food chain be available, from production to delivery of GM products to the market (traceability).

According to Regulation 1829/2003, all genetically modified food/feed or products thereof have to be authorized prior to placement on the market regardless of their substantial equivalence to existing food/feed or food/feed ingredients. Similar to non-GM food, GM food must not have adverse effects on human health, animal health, or the environment; mislead the consumer; or differ from food, which it is intended to replace to the extent that its normal consumption would be nutritionally disadvantageous for the consumer. However, unlike 'normal' food, GM food has to undergo a pre-market safety assessment before being authorized to be placed on the European market. The European Food Safety Authority (EFSA) performs the evaluation of new GMOs and GMO-derived products for food and feed purposes. According to Article 5 (8) of Regulation (EC) 1829/2003, EFSA is supposed to publish detailed guidance concerning the scientific aspects and the presentation of information necessary to demonstrate that a product complies with the criteria referred to in Article 4 of the Regulation. Since the guidance has not been published yet by EFSA, risk assessors and appliers use the 'Guidance Document for the Risk Assessment of Genetically Modified Plants and Derived Food and Feed' prepared by the "Joint Working Group on Novel Foods and GMOs" and published March 6/7, 2003 [13]. Up to November 7, 2003, products of 14 genetically modified plant crops were approved in accordance with the former Regulation (EC) 258/97 (Table 4).

Table 4 Food or food products derived from genetically modified plants that can legally be marketed in the EU - notification pursuant Regulation 258/97 [70].

Product	Crop	Event(s)	Trait(s)	Company
Refined oil	Canola	TOPAS 19/2	Herbicide-tolerant	Bayer CropScience
Refined oil	Canola	MS1Bn, RF1Bn, RF2Bn, MS1xRF1, MS1xRF2	Male-sterility, fertility restore	Bayer CropScience
Refined oil	Canola	GT73	Herbicide-tolerant	Monsanto
Refined oil	Canola	FALCON GS 40/90	Herbicide-tolerant	Bayer CropScience
Refined oil	Canola	Liberator L62	Herbicide-tolerant	Bayer CropScience
Refined oil	Canola	MS8, RF, MS8xRF3	Herbicide-tolerant	Bayer CropScience
Oil	Cotton	1445	Herbicide-tolerant	Monsanto
Oil	Cotton	531	Insect-resistant	Monsanto
Various	Maize	MON 810	Insect-resistant	Monsanto
Various	Maize	T25	Herbicide-tolerant	Bayer CropScience
Various	Maize	Bt-11x2022	Insect-resistant	Syngenta Seeds
Various	Maize	MON 809	Insect-resistant	Pioneer
All*	Maize	Bt-176	Insect-resistant	Syngenta Seeds
All*	Soybean	GTS 40-3-2	Herbicide-tolerant	Monsanto

* approved under Directive 90/220/EEC

All products, which have been lawfully placed on the market before the Regulation (EC) 1829/2003 entered into force, have to undergo a further safety assessment regarding the requirements of Regulation (EC) 1829/2003. Authorizations according to Regulation (EC) 1829/2003 are valid for a maximum of 10 years and can be renewed.

2.3 Safety assessment of genetically modified food

Since the end of the 1980s, scientific expert-groups and international organizations have taken a lead in establishing international guidelines for the risk and safety assessment of GMOs and food derived from GMOs. They include the Food and Agriculture Organization of the United Nations (FAO), the World Health Organization (WHO) [11, 75-77], the International Food Biotechnology Council (IFBiC) [78], the International Life Science Institute (ILSI) [79], and the Organization for Economic Co-operation and Development (OECD) [2, 80],

The aim of the risk and safety assessment is to identify, to characterize, and to assess any potential hazard to the health of humans, animals, or the environment due to the GMOs or products derived thereof.

Rationale of the concept of substantial equivalence

Safety assessment of whole foods is difficult and only a few studies had been conducted before the introduction of GMOs [11]. Traditional toxicological studies, mainly based on bioassays such as animal tests, have been used to ensure the safety of single, well-characterized compounds like food additives, pesticides, or other chemicals, with little nutritional value and low anticipated level of exposure. Long-term animal feeding studies, for example, are usually used to determine that there are no observed adverse effect levels (NOAELs) and are then used to establish acceptable daily intakes (ADI = NOAEL x safety factor) for these substances. The safety factor, usually 100, provides a safety margin taking into account the difference between the animal model and humans. Contrary to single compounds, whole foods are complex mixtures of tens of thousands of compounds with a wide range of compositions and nutritional values and, potentially, high levels of exposure. It is generally recognized [2, 11, 12, 75, 81] that animal feeding studies have limitations, as the test compound will comprise a fraction of the experimental diet and reflect low quantities of the anticipated human intake. High intake levels are impossible since the diet must be formulated to avoid nutritional imbalance (e.g., lack of specific vitamins), which itself can lead to adverse effects unrelated to the toxicity of the food. In addition, many toxic or anti-nutritional substances are known to be naturally present in conventional crops. Some examples of such are solanine in potatoes; proteases inhibitor in soybeans, maize, barley, and potatoes; lectins in soybeans, wheat, and potatoes; and isoflavones in soybeans [82, 83]. Thus, many food crops cannot be consumed until they have been properly processed, which alters or eliminates their toxicity.

Due to the difficulties in applying traditional toxicological testing to whole foods, alternative approaches were required for the safety assessment of GM foods. Based on the assumption that food prepared and used in conventional ways has an established history of safe use [2], the OECD Working Group developed the principle that GM foods or plants can be compared with their traditional counterparts in order to show that they are as safe as their traditional counterparts [2]. Any identified differences between the GM food and its counterparts are then subjects of a safety assessment, which can include appropriate nutritional, toxicological, and immunological testing. This approach, known as the concept of substantial equivalence, has been further elaborated by WHO/FAO [11, 77, 84] and is now the basis of all GM food safety and risk assessment procedures used by regulatory authorities worldwide, e.g., US FDA 1992, Health Canada 1994, Japanese Ministry of Health and Welfare, 1996, and EU

GM Food and Feed Regulation 2003. The extent of the safety assessment of GM products is decided on a case-by-case basis, depending on the type of genetic modification and the outcome of this comparative assessment.

EU safety assessment approach

The European Union's safety assessment approach is based on the principle of substantial equivalence and international guidelines published by FAO/WHO. The Joint Working Group on Novel Foods and GMOs of the Scientific Committee on Food compiled a guidance document that describes the different elements considered in the risk assessment process of genetically modified plants and derived food and feed [13] regarding the frameworks of Directive 2001/18/EC and Regulation (EC) 1829/2003.

According to the EC Guidance Document the safety assessment of GM plants and products has to include the following elements [13]:

- the molecular characterization of the introduced gene(s) and gene product(s),
- the potential environmental impact following a deliberate release,
- the safety and risk evaluation of the GM plant or products derived thereof, including agronomic and compositional characteristics of the GM crop, nutritional assessment, as well as potential toxicity and allergenicity aspects of the inserted gene(s), the newly expressed protein(s) (if applicable) and relevant metabolites (if applicable)

The molecular characterization of the GMO takes into account the characteristics of the donor and recipient organisms and provides information on methods used for the genetic modification, DNA sequence used in transformation, inserted/deleted sequence, insertion site, copy number, expression of the insert, protein identity (primary sequence, molecular weight, post-translational modification, function), inheritance, and stability.

If the GMO is intended to be released into the environment, the characteristics of the GMO and its potential impact on the ecosystem, i.e. wild plants, non-modified crops, target and non-target organisms, have to be assessed.

Comparative analysis takes into account agronomic, phenotypic, genetic, and compositional aspects. Such comparisons are made with GM and non-GM counterparts grown side-by-side at multiple geographic locations during more than one growing season. The analysis of the

composition includes the assessment of key macro- and micro-nutrients, toxicants, and anti-nutritional compounds. In order to harmonize and standardize data requirements, the OECD has compiled science-based consensus documents on key nutrients, anti-nutrients, and toxicants characteristic for wheat, maize, potatoes, sugar beet, soybeans, and canola [85-90]. The nutritional evaluation of GM foods includes an analysis of the composition of nutrients, an evaluation of the biological efficacy and availability of nutrient components, and an assessment of dietary intake and nutritional impact.

The toxicological assessment should include studies on newly expressed protein(s) and any changes beyond normal variation due to the genetic modification. Such changes could be the presence of novel constituents and alterations in the levels of naturally occurring constituents. Toxicological studies include a sequence homology search to proteins known to cause adverse effects, protein stability, and a 28-day oral toxicity study according to OECD guideline 407 [91]. For processed foods derived from GMOs, any effect of the production process should be assessed with respect to concentration, elimination, denaturation, or degradation of the recombinant-DNA and the novel protein in the final product. The strategy of an allergenicity assessment focuses on IgE-mediated responses and follows an integrated, stepwise, case-by-case approach in accordance to the recommendations of ILSI/IFBiC and the Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology [92, 93] and the Codex *ad hoc* Intergovernmental Task Force on Foods Derived from Biotechnology [77]. It includes identification of potential linear IgE binding epitopes, pepsin-resistance tests, and specific serum screening. The allergenicity assessment focuses mainly on the newly expressed protein but also considers changes in the pattern of allergenic proteins naturally present in the conventional plant or product.

2.4 Unintended effect(s)

The assessment of the safety of genetically modified organism aims to address both intended and unintended effects as a result of the genetic modification. The European Joint Working Group on Novel Foods and GMOs describes intended and unintended effects as follows [13]:

"**Intended effects** are those that are targeted to occur from the introduction of the gene(s) in question and which fulfill the original objectives of the genetic transformation process.
Unintended effects are considered to be consistent differences between the GM plant and its

appropriate control lines, which go beyond the primary expected effect(s) of introducing the target gene(s)."

The Codex Alimentarius [84] further divides unintended effects into two groups: those which are "**predictable**", based on the knowledge of insertion site and the metabolic connection to the inserted trait, and those that are "**unexpected**", mainly due to unknown methods of gene regulation and gene-gene interactions.

Origin of unintended effects

Unintended effects can be the result of secondary or pleiotropic effects of the transgene expression and insertional effects resulting from transgene integration into plant genomes [94-97]. Cell metabolism is based on complex and poorly understood interactions (networks) between genes, proteins, metabolites, and the environment. The introduced gene or the changed expression level of an existing gene may interact in an unpredictable way with other biochemical pathways, resulting in changes of the crop composition.

The introduction of exogenous DNA sequences into the plant genome by *Agrobacterium*-mediated transformation or particle bombardment method, as described under 2.1, is a random process leading to physical disruption. This could inactivate endogenous genes or modify gene expression. Activation of silent genes and formation of fusion proteins by transcriptional read-through processes are also theoretically possible [95].

It should be noted that the occurrence of unintended effects is not unique to the application of rDNA techniques. It is an inherent and general phenomenon that occurs frequently in traditional breeding, e.g. due to hybridizations (potato breeding lines with novel, toxic glycoalkaloids [98]), natural genetic recombination, natural chromosomal rearrangements (translocations and inversions), activity of transposable elements in plant genomes, cell fusion, or chemical and radiation induced mutations [12, 77, 95]. Unintended effects due to traditional or recombinant DNA approaches may lead to potential food hazards. These comprise an accumulation of toxins or anti-nutrients, a decrease of important nutrients, changes in bioavailability of nutrients (micro- and macro-), and a formation of novel toxins including allergens [99].

Table 5 lists examples of unintended effects due to traditional breeding and genetic engineering to demonstrate that unintended effects are not specific to rDNA techniques.

Unintended effects due to traditional or recombinant DNA approaches may lead to potential food hazards. These comprise an accumulation of toxins or anti-nutrients, a decrease of important nutrients, changes in bioavailability of nutrients (micro- and macro-), and a formation of novel toxins including allergens [99].

Table 5 Unintended effects due to traditional breeding and recombinant DNA techniques [100]

Plant	Trait	Unintended effect	Reference
Traditional breeding			
Barley	Powdery mildew resistance	Yield ↓	[101]
Celery	Pest resistance	Furanocoumarins ↑	[102]
Corn	High lysine content	Yield ↓	[103]
Potato	Pest resistance	Glycoalkaloids ↑ and yield ↓	[104]
Zucchini	Pest resistance	Cucurbitacin ↑	[105]
Genetic engineering			
Canola	Overexpressing phytoene-synthase	Multiple metabolic changes (tocopherol, chlorophyll, fatty acid, phytoene)	[42]
Potato	Expressing yeast-invertase	Glycoalkaloids ↓	[106]
	Expressing soybean glycinin	Glycoalkaloids ↑	[107]
	Expression of bacterial levansucrase	Adverse tuber tissue perturbations	[108]
Rice	Expressing soybean glycinin	Vitamin B6 ↑	[51]
	Expressing provitamin A biosynthetic pathway	Formation of unexpected carotenoid derivatives	[109]
Wheat	Expression of phosphatidyl serine synthase	Necrotic lesions	[110]

Detection of unintended effects

In order to identify unintended effects, a comparative approach is used that focuses on the determination of similarities and differences between the GM plant and its conventional counterpart. Comparative analyses can be applied at several points along the production chain. If possible, they should be carried out with the GM plant and the raw agricultural commodity rather than processed food as indications of an unintended effect may get lost during food processing steps. The traditional counterpart is the nearest line of comparable genetic background, the parental variety or the equivalent non-GM hybrid. These lines are not always readily available because the commercial GM crops may have undergone different traditional breeding techniques, such as inbreeding, back-crossing, and outcrossing. This makes a direct comparison to a line of a comparable genetic background difficult. Therefore, in order to take such genetic differences into account, closely related varieties should be

included as a baseline for the evaluation of substantial equivalence. Figure 1 shows the scheme for a comparative and sequential approach to assess the safety of GM crops. According to this approach, parental and transgenic lines are compared to regard both intended and unintended effects. The differences found are considered in the light of the natural baseline generated by conventional varieties due to genetic variation and environmental conditions. Differences outside of this natural baseline are assessed for their safety relevance. Additional nutritional and toxicological testing may be warranted.

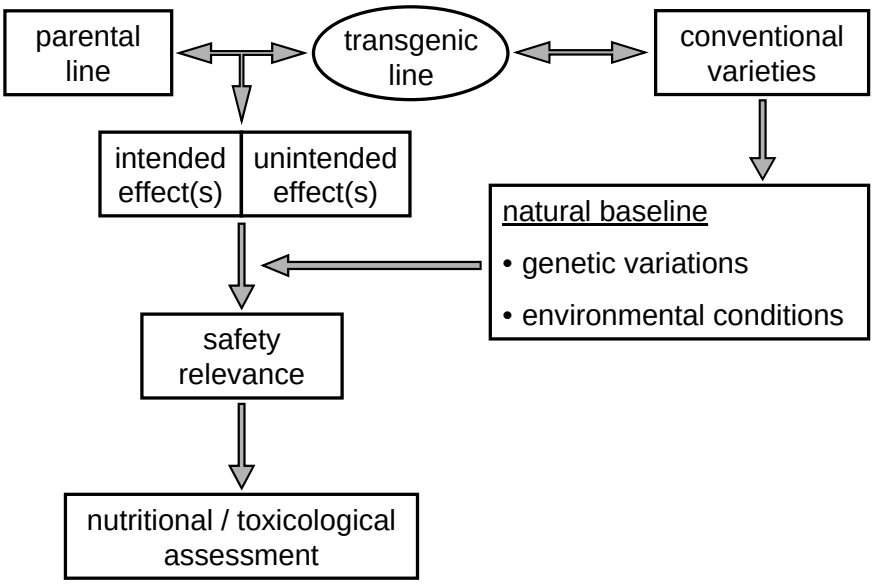


Figure 1 Sequential comparative approach for the safety assessment of genetically modified crops

For comparative analysis, the GM crop and the control lines need to be grown under identical environmental conditions to obtain informative data. To account for environmental, climatological, and geographical influences, GM and non-GM varieties should preferably be grown at multiple geographical locations and seasons [13]. The comparison to data from literature and databases [111] is also a valuable tool to assess the biological and safety relevance/significance of any detected difference.

The comparison between GM and non-GM crop usually comprises agronomic/phenotypic characteristics and the composition of the crop. Agronomic and phenotypic characteristics are very sensitive indices of alterations in metabolism and potential genetic pleiotropy, and, hence, robust indicators of equivalence [112]. Table 6 exemplarily shows the phenotypic and agronomic parameters assessed for corn.

Table 6 Phenotype and agronomic characteristics used to evaluate equivalence of corn [112]

Morphological and agronomic characteristics	
Stand establishment	Early plant vigor
Leaf orientation	Leaf color
Plant height	Root strength (lodgin)
Silk date	Silk color
Ear height	Ear shape
Ear tipfill	Tassel color
Tassel size	Reaction to fungicides/herbicides
Dropped ears	Late season staygreen/appearance
Stalk rating	Susceptibility to pathogens/pests
Above ear intactness	Yield

Compositional studies are based on an analysis of key macro- and micronutrients, anti-nutrients, and known toxins and allergens. The OECD Task Force on the Safety of Novel Foods and Feed has compiled consensus documents for a number of economically important food crops. These describe recommended constituents for a comparative analysis and include their background levels. As of February 2004, consensus documents on wheat, maize, potatoes, sugar beets, soybeans, and canola have been published, while additional documents on cotton, rice, sunflowers, barley, forage legumes, and tomatoes are in preparation [113].

Table 7 shows the key nutritional and anti-nutritional parameters suggested to be analyzed for soybean seeds and maize kernels.

Table 7 Suggested nutritional and compositional parameters to be analyzed in soybean and maize matrices for human food use [86, 89].

Parameter	Soybean seeds	Maize kernels
Proximates[1)]	X	X
Minerals[2)]		X
Vitamins[3)]		X
Amino acids	X	X
Fatty acids	X	X
Toxicants/antinutrients		
Phytic acid	X	X
Raffinose		X
Furfural		X
Ferulic acid		X
p-coumaric acid		X
Trypsin inhibitors	X	
Lectins	X	
Isoflavones	X	

[1)] Proximates includes protein, fat, total dietary fibre, ash and carbohydrates
[2)] Minerals for maize include Na, K, Ca, P, Mg, Fe, Cu, Se, Zn
[3)] Vitamins for maize include the Vitamines A, B_1, B_2, B_6, C, E, Folate, and Niacin

The final number of analytes depends on the type of GM crop, the intended effect, and any "predictable" effect and will be determined on a case-by-case basis.

If none of the compared parameters exceed natural variation, the plant crop is considered to be substantially equivalent. It is then estimated that the likelihood of any relevant toxicologically effects has been reduced [97].

Concerns

The substantial equivalence approach is a powerful and flexible tool for the safety assessment and has worked well in the safety assessment of genetically modified crops [3]. The substantial equivalence evaluation of over 50 biotech crops worldwide has lead to the conclusion that foods and feeds derived from these biotech crops are as safe and nutritious as foods and feeds derived from traditional crops [4]. Worldwide, there has been no verifiable unintended toxic or nutritionally harmful effects resulting from the cultivation and consumption of products from GM crops [5, 6] after eight years of production.

However, concerns exist that the spectrum of compounds used for the comparison is not comprehensive enough and that it is based solely on experience gained from analysis of food crops obtained via conventional breeding. Unintended effects altering the level of new toxicants or anti-nutrients might remain undetected [7, 8]. This could be of particular importance for GM crop plants of the second generation, where the genetic modifications become more complex. One example is the introduction of a new metabolic pathway in rice (Golden Rice: [9]). The insertion of multiple genes involved in different pathways will possibly increase the likelihood for unintended and unexpected effects [10, 11], either resulting from insertion or pleiotropy. In order to improve the detection of unintended effects, profiling techniques are considered an alternative for detecting such changes [11-13].

2.5 Profiling techniques

The development and evaluation of "unbiased" non-targeted techniques providing additional information on unintended changes that might remain undetected using targeted approaches has been discussed by international advisory bodies since 1997 [11, 14]. Profiling technologies such as transcriptomics, proteomics, and metabolomics allow the simultaneous measurement and comparison of thousands of plant components without prior knowledge of their identity. These 'unbiased' approaches are considered to facilitate a more holistic analysis. Three main cell constituent groups are targeted by these technologies: the transcriptome, i.e. the total of initial transcription products (mRNA) per cell at a given time; the proteome, the total of translation products (protein) per cell at a given time; and the metabolome, the total of metabolites per cell at a given time. These technologies will be discussed in detail below.

2.5.1 Transcriptomics

The total set of initial transcription products (mRNA) in a cell at a given time is called the transcriptome; it provides information about gene expression at a specific time in a specific tissue [63]. Insertion mutations or pleiotropy can potentially alter the expression of endogenous genes. Thus, an analysis of the transcriptome might be a powerful tool to detect such effects on the transcript level.

Lockhart and Winzeler [114] presented a very informative overview of the technology of gene expression analyses. Many methods are available to analyze differences in gene expression (mRNA abundance), including northern blotting [115], differential display [116, 117], arbitrarily primed PCR [118], serial analysis of gene expression (SAGE) [119], MPSS (massively parallel signature sequencing) [120], and DNA microarrays [121].

Kok et al. [117] investigated the usefulness of the differential display method to detect altered gene expression in genetically modified plant varieties compared to their traditional counterparts. The concept of differential display is to use short arbitrary primers in combination with the anchored oligo-dT primers to systematically amplify and visualize most of the mRNA in the samples. The polymerase chain reaction (PCR) products of the samples are analyzed by polyacrylamide gel electrophoresis. Kok et al. were able to demonstrate that the method was reproducible and able to discriminate between two developmental stages

(green and red) of the same tomato line. However, the authors concluded that the differential display method is rather laborious, dependent on individual skills, and, thus, is not suited for routine application to detect altered gene expression as part of a risk assessment strategy [117, 122, 123].

DNA microarray technology is another method for gene-expression analysis and is considered to be the most promising among the transcriptomics techniques to support the safety assessment of GMOs [122]. The principle underlying this method is base-pairing (i.e., A-T and G-C for DNA; A-U and G-C for RNA) or hybridization of complementary nucleotide sequences of mRNA or cDNA to a high-density array of immobilized target sequences, each corresponding to a specific gene. The target sequences, also called probes, can be amplified cDNA (500 to 5,000 bases long) from genomic DNA or cDNA libraries, or synthetic oligonucleotides (25 to 60 bases long) that are immobilized to a solid surface. RNA samples (e.g., total mRNA isolated from tissue of two different plant lines) are labeled by reverse transcription in the presence of fluorescently tagged nucleotide precursors. The mRNA of each plant line is labeled with different fluorescent dyes to allow quantitative comparison. The two fluorescently labeled cDNA samples are co-hybridized to a DNA microarray. Each spot of the array represents one gene of interest. The array is scanned at two different wavelengths to detect the relative transcript abundance for each plant line [114].

The advantages of microarray methods are their impressive sensitivity – one fragment in a pool of up to 500,000 [124] – and their potential to measure mRNA levels 'quantitatively' (relative to control) for thousands of genes, or even entire genomes [114, 125] simultaneously with a single DNA microarray. However, at this time, there are no arrays available that cover entire transcriptomes of individual (crop) plant species. The commercially available arrays contain ESTs or cDNAs of specific tissues of the plant in specific developmental stages [122]. Therefore, DNA microarray analysis is not truly unbiased as it only provides expression information for the subset of gene sequences on the array. Expression changes of other genes will remain undetected. Another challenge is the distinction of similar mRNAs because of cross-hybridization of other mRNAs with related sequences with the probe, e.g. different members of the same gene family [122]. Current estimates are that sequences with >70% nucleotide identity over >200 bp are likely to exhibit some degree of cross-hybridization [126]. This may hinder the detection of mRNAs of fusion proteins.

Furthermore, as mRNA is not a health issue, it would be necessary to confirm a detected difference in mRNA abundance at the protein or metabolite level in order to assess the biological and toxicological significance of mRNA changes [122]. This translation may not always be simple because the correlation between mRNA and protein level or its biological activity is poor (coefficient <0.5) as a result of different turnover rates of mRNA and protein, alternative splicing and post-translational modification [127-131]. The correlation between transcript and metabolite levels is also small; only 2% of possible pairs (26,616) showed significant (P<0.01) correlation in potato tuber [132].

Urbanczyk-Wochniak et al. [132] compared the discrimination power of RNA and metabolic profiling to distinguish between different potato tuber systems (transgenic vs. wild type). A principle-component analysis of the transcriptional profile of two transgenic potato tuber samples (expressing a bacterial sucrose phosphorylase [133] or a yeast invertase [134]) and their wild-type tuber sample was not able to discriminate either from each other. Although, the PCA of their metabolic profile clustered completely independently all three lines [132]. The authors concluded that the establishment of substantial equivalence between transgenic and conventional crops should be performed at more than one level.

A more comprehensive overview of applications of microarray technology in various areas is provided in current reviews, e.g. by Fadiel and Naftolin [135] (biomedical applications) and by Aharoni and Vorst [136] (plant biological applications).

2.5.2 Proteomics

Proteomics can describe protein expression and its changes under the influence of biological perturbations (e.g., disease or mutation) in a comprehensive and quantitative way [137]. The proteome, coined by Wilkins et al. [138], is the protein compliment of the genome and the result of genetic expression, ribosomal synthesis, and proteolytic degradation. Contrary to the genome, which is constant for an organism, the proteome of an organism is in flux; it depends on cell cycle, environmental influences, and tissue/cell type. Although there is no single, fixed proteome, the proteome nevertheless remains a direct product of the genome. Proteins play a vitally important role in metabolic pathways and other cell processes. Therefore, if a genetic modification affects the genome (e.g. insertional mutation) or gene regulation (pleiotropic effect) of a plant in a way that it changes metabolic pathways or produces a new protein (e.g., inserted gene product, fusion protein, or activation of a silent

gene), the proteome will also be altered. Protein function is highly regulated by post-translational modifications such as phosphorylation and glycosylation. As a consequence of such post-translational modifications as well as the discrepancies between the levels of mRNA and their proteins, only the study of the proteins themselves can provide information on their actual amounts and activities at a given time in a given tissue [139].

Expression proteomics is achieved by two-dimensional gel electrophoresis (2DE). 2DE combined with mass spectrometry (MS) is the most widely used approach to analyze proteomes [140]. Proteins are separated in the first dimension as polypeptides according to their isoelectric point (pI) and in the second dimension according to their molecular weight. Hundreds to a few thousand proteins can be separated and quantified on a single gel. Proteins of interest can be subsequently sequenced and identified by MS or MS/MS.

Improvements in 2D gel electrophoresis techniques now offer highly reproducible resolution for protein separation (more details under 2.6.1). Moreover, recent advances in mass spectrometry have allowed an analysis of low levels of proteins separated by 2D gels.

The power of 2DE-based proteomics is in the comparison of proteomes and in identifying differentially expressed proteins [141]. Such an approach has already proven successful for the investigation of differential protein expression in *Arabidopsis* due to developmental or environmental changes or mutations [142-150]. Proteomics, using 2D gel electrophoresis technique as the core method, was also successfully employed to detect pleiotropic effects of single mutations [144, 145, 151, 152). The use of the 2DE approach to assess genetic variability at the level of expressed proteins is reviewed by Thiellement et al. [153].

A 2DE based proteomics approach was also utilized by Corpillo et al. [154] to compare the protein profiles of a tomato hybrid genetically modified for resistance to tomato spotted wilt virus (TSWV) and the same non-transgenic hybrid. Total soluble protein was extracted from tomato seedlings and separated by 2DE. The transgenic TSWV nucleoprotein was undetectable by immunoblotting. There was no actual data regarding the number of compared spots. However, quantitation or statistical evaluation was presented in the publication. The authors state that “no reproducible difference was found between the wild-type and the GM tomato in terms of appearance, and disappearance or shift of any spot” [154].

A more in-depth discussion of the technical aspects of 2DE will be provided in section 2.6.

Multidimensional protein identification technology (MudPIT), introduced by Yates' group [155] is an emerging alternative to 2DE. For the MudPIT approach, extracted proteins are digested to peptides with a protease prior to separation by two-dimensional liquid chromatography (2D-LC) facilitating a direct interface to the ion source of a MS. So far, primarily studies on yeast have been published using the MudPIT technology [128, 156-158]. Washburn et al. [157, 159] were able to resolve and identify 1484 proteins from yeast (*Saccharomyces cerevisiae*), representing 24% coverage of the predicted ORFs of this organism. The authors demonstrated that even low-abundance proteins can be clearly identified. In a recent study, Washburn et al. [128] studied the mRNA and protein expression changes of yeast when cultured on either rich media or minimal media. They used metabolic labeling by growthing yeast cells in media with either ^{14}N or ^{15}N as the sole nitrogen source for the quantitative proteomics analysis. The overall correlation between mRNA and protein expression was weakly positive with a correlation coefficient of 0.45 for 678 loci. The shortcoming of metabolic labeling is that it is limited to species that can be grown under such conditions. Koller et al. [160] recently applied MudPIT in combination with 2DE to analyze the plant proteome of rice. Using these two approaches, they were able to detect and to identify 2528 unique proteins. One of the major drawbacks of this technology for the analysis of GM crops is that quantitative proteome analysis is not readily possible.

Another quantitative LC-MS approach is based on isotope-coded affinity tag labeling (ICAT). The proteins from two different samples are labeled by ICAT with different molecular masses [161]. The labeled proteins of the two samples are combined and digested with a protease such as trypsin and analyzed by LC-MS/MS. The difference in protein quantity is determined by measuring the peptides labeled with heavy and light ICAT reagent. This technique was successfully used for the first time in plants by Islam et al. [162] to analyze differences in protein compositions between wheat cv. Chinese Spring and its chromosome deletion lines. Smolka et al. [163] used the ICAT reagent to label yeast proteins prior to 2DE separation. The separation by 2DE and mass spectrometric analysis of differential isotopically labeled proteins provided an accurate quantification of each the protein contained in protein spots of comigrating proteins.

Typically, these LC-MS approaches have been used for the proteome analysis of unicellular prokaryotes and eukaryotes (yeast) [128, 161, 164, 165], because protein identification is based on the comprehensive knowledge of the investigated organism genome sequence

information of peptides. Recent studies have shown that different proteomics approaches, such as 2DE-MS and ICAT-MS/MS or MudPIT, applied to the same sample provide complementary information and therefore contribute to a more complete understanding of the proteome [160, 166]. One of the great challenges of LC-MS/MS approaches is managing the huge data sets generated by high-throughput 2-D LC-MS/MS. A typical 2-D LC-MS/MS run (MudPIT) generates between 20,000 and 30,000 individual MS/MS spectra that have to be further analyzed using software routines such as SEQUEST, MASCOT, or ProFound for peptide identification [167].

Proteomics has already been successfully employed to study changes in the protein composition of genetically modified food crops. The decrease in glutelin levels of genetically engineered low-glutelin rice, based on anti-sense technology, has been associated with an unintended increase in prolamin levels [11]. The change in prolamin levels would not have been detected by standard nutritional analyses such as total protein and amino acid profiles. Instead it was observed following sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

2.5.3 Metabolomics

Metabolites are often seen as the end products of the cascade of gene – transcript – protein (enzyme) – metabolite. Unexpected changes, like mutations at the gene level or pleiotropic effects, may alter the metabolome of a plant tissue. The metabolome has been defined, in a microbial context, as the total complement of metabolites in a cell [168]. It is estimated that the number of metabolites present in the plant kingdom exceeds 200,000 [169], yet only approximate 10,000 of them have been described so far [170]. Unlike targeted compositional analyses, metabolomics aims to detect as many metabolites as possible for an unbiased and comprehensive assessment of all metabolites present in a biological sample [171]. In contrast to the transcriptome that consists of linear polymers of four nucleotides with highly similar chemical properties, the metabolome is a complex mixture of small molecules with different carbon skeletons, functional groups and physicochemical properties including volatility, polarity, hydrophobicity, solubility, and charge status. The chemical diversity and complexity of the metabolome make it extremely challenging to cover all metabolites simultaneously and no single analytical technique exists that would cover all metabolites [172]. The highest number of metabolites that has been detected and quantified with a single

method (GC-TOF MS after chloroform/methanol/water extraction) is 652 metabolites for *Arabidopsis* leaves [173]. In order to gain a more comprehensive view of the metabolome of a plant tissue, multiple methods need to be deployed in combination. History, background, advantages, and limitations of techniques used for metabolomics are discussed in current reviews [169, 172, 174-177]. Numerous techniques exist but the most promising and commonly used ones are combinations of chromatographic separation with spectrometric detection methods, e.g. gas chromatography coupled to mass spectrometry (GC-MS), liquid chromatography coupled to mass spectrometry (LC-MS), and LC coupled to nuclear magnetic resonance (off-line LC-NMR).

Methods based on gas chromatography coupled to mass spectrometry have been successfully applied to plant metabolite profiling [175] and have allowed the detection of 150 compounds in potato (*Solanum tuberosum*) tuber tissue [178] and of 326 metabolites in *Arabidopsis thaliana* leaves [179]. In a recent publication, Weckwerth et al. [173] detected 652 metabolites in a single *Arabidopsis* leaf sample by GC coupled with time-of-flight (TOF) MS. Frenzel et al. [180] sub-fractionated the crude extract of rice into four fractions in order to increase the resolution power for minor constituents. More than 100 compounds were identified in this manner by GC/MS.

Another powerful tool for plant metabolite analysis is high-resolution nuclear magnetic resonance spectroscopy (NMR) with or without prior LC separation. In order to obtain a metabolite fingerprint of *Arabidopsis thaliana,* Ward et al. [181] applied ^{1}H-NMR to crude plant extracts of nine ecotypes and differences were observed in both the carbohydrate region and the aliphatic region. Noteborn et al. [182] were able to detect glutamic acid and citric acid as the two metabolites that were altered in the GM long-ripening tomato compared to the parent line by using off-line LC-NMR. Another study on genetically modified tomatoes was conducted by Le Gall et al. [183], who analyzed non-transgenic and transgenic tomatoes overexpressing two maize transcription factors by ^{1}H-NMR. The comparison of the ^{1}H-NMR spectra revealed that apart from the intended increase of flavonoids, there were statistically significant changes in levels of other compounds, including citric acid, sucrose, and phenylalanine.

These profiling techniques provide a broader view of plant constituents than the targeted methods and, thus, additional opportunities to identify unintended effects. However, the targeted methods are standardized and validated, and they provide robust, quantitative, and

comparable data. Furthermore, reference data exists in published literature and public databases such as the ILSI Crop Composition database [ILSI/IFBiC: 111]. The untargeted, profiling techniques provide semi-quantitative information. Thus, responses/signals can be quantified, and relative values between GM sample and control can be calculated. However, the enormous complexity of the data sets generated through profiling analyses and the lack of baseline data of natural variations will provide a major challenge for a meaningful interpretation of such comparative analyses. The key question will be: Are the detected differences of biological relevance?

The following requirements are necessary before these methods can be deployed within the safety assessment of genetically modified crops:

- Standardization of sample collection, sample preparation, and sample extraction
- Standardization and validation of methods
- Establishment of verified baseline range of natural variations
- Improvement of bioinformatic systems

The utility of profiling methods for the safety assessment of genetically modified food plants is currently under scrutiny. Two research projects in Europe, GMOCARE (QLK1-1999-00765 [184] and UK Foods Standards Agency G02 [185] are assessing the applicability and practicability of such profiling techniques to detect unintended effects due to genetic modification on various crops. No research data have been published from these groups yet. So far, no profiling method or protocol has been developed that would be suitable to complement the tools for the safety assessment of genetically food. These methods may be useful for identifying changes

2.6 2DE-based proteomics

The goal of differential expression proteomics is to find the protein expression that is induced (seen as new protein), inhibited (seen as missing protein), or changed (seen as differences in protein abundance) between biological samples, e.g. genetically modified versus non-transgenic or parental individuals. It requires a qualitative (presence of protein) and quantitative (abundance of protein) comparison of protein profiles and can be achieved by 2DE [186]. Two-dimensional gel electrophoresis (2DE) was first described by O'Farrell [187] and by Klose [188] in 1975. It demonstrated for the first time the separation and

visualization of several hundred proteins of a complex protein mixture on a single gel. 2DE utilizes two independent protein characteristics for the separation of protein mixtures: the isoelectric point (pI; pH at which the net charge of a protein is zero) and the molecular weight (M_w) of a protein. In the first dimension, proteins are separated according to their pI by isoelectric focusing (IEF) and in the second dimension according to their M_w by SDS-PAGE. The stained 2DE gel displays a two-dimensional protein spot pattern or protein profile. Each visualized spot on the resulting 2DE gel corresponds to at least one protein in the sample. Thousand of different proteins can thus be separated, and information such as the protein pI, the apparent molecular weight, and the relative amount of each protein can be obtained. A further strength of 2DE is that it can reveal post-translational modifications (PTMs) of proteins, such as phosphorylation, geranylation, glycosylation, proteolytic cleavage, translational induction, alternative splicing, or any other modification that affects molecular weight or isoelectric point.

Improvements in 2DE techniques, like the introduction of immobilized pH gradient gels supported by a plastic backing (IPG strips) [189, 190], now offer highly reproducible resolution for protein separation. Moreover, recent advances in mass spectrometry (MS) [191-193] and genome analysis have made it possible to identify small quantities of protein separated by 2DE in a high throughput manner. The combination of 2DE with mass spectrometry and bioinformatics is, therefore, still one of the most powerful and widely used tools for proteome analysis [194, 195]. Applications of 2DE in various areas of proteomics were recently reviewed [137, 195-197]. They include expression and functional proteomics as well as identification (e.g., taxonomy, forensic work), the study of genetic variation and relationships, the detection of stages in cellular differentiation, studies of growth cycles, the examination of pathological states and diagnosis of disease (disease marker), cancer research, monitoring of drug action, purity checks, and micro-scale protein purification.

2.6.1 2DE: State of the art

Proteome analysis based on 2DE consists basically of six steps (Figure 2): sample preparation, protein separation (IEF coupled with SDS-PAGE), visualizing of proteins (staining), detection of protein spots of interest (image analysis), identification of proteins, and function analysis (bioinformatics). The six steps are discussed in more depth below.

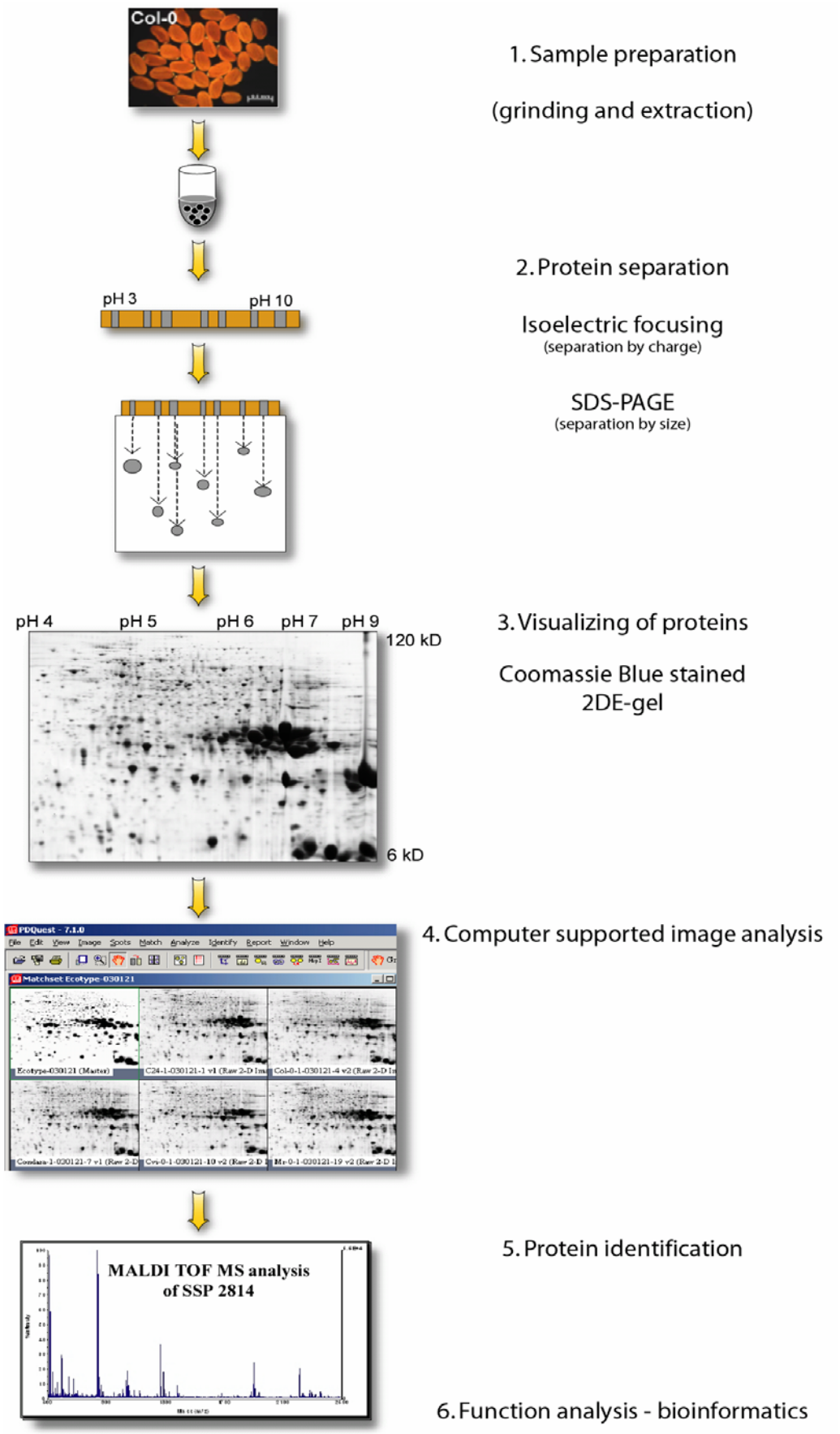


Figure 2 Workflow of 2DE

2.6.1.1 Sample preparation

Sample preparation is the most crucial and critical step of 2DE [149, 198]. It directly affects 2DE results with respect to the protein classes (membrane proteins, cytosolic proteins), the number and stability of proteins, and the abundance of proteins displayed in the 2DE gel, as well as the resolution quality of the 2DE pattern [195]. An ideal sample preparation would (i) allow for solubilization of all protein components, including hydrophobic membrane proteins; (ii) prevent protein aggregation and degradation; (iii) be compatible with isoelectric focusing (IEF); and (iv) remove all contaminants (e.g., DNA, RNA, metal ions, and salt), which interfere with the IEF separation [199].

Plant samples, such as leaves, roots, seedlings, or seeds, are usually homogenized in a liquid nitrogen cooled mortar to maintain protein stability [200]. Proteins are then solubilized from the sample powder by extraction (lysis) buffers. In order to obtain and maintain protein solubilization, it is necessary that the extraction buffer completely disrupt all inter- and intra-molecular interactions and minimizes protein modifications. Chemicals used to achieve these goals can be classified as chaotropes, detergents, reducing agents, carrier ampholytes, protease inhibitors, and nucleases. Examples are given for each group in Table 8. Chaotropes disturb non-covalent interactions, such as hydrogen bonds and hydrophobic interactions, which lead to protein unfolding and denaturation. Detergents are surface-active molecules and support the denaturation of proteins by preventing hydrophobic interactions [201]. Reducing agents are used to disrupt intra- and intermolecular disulfide bonds. Carrier ampholytes aid protein solubilization, inhibit interaction between hydrophobic proteins and immobilized ampholytes of the IEF gel strip, which leads to streaking effect at the basic end [202], scavenge cyanate ions [203], and help in the precipitation of nucleic acids during centrifugation [204]. The denaturing property of the sample buffer is not always sufficient to inhibit the action of proteases [149, 201, 205]. Protease inhibitors are added to the extraction buffer to prevent protein degradation by proteases (e.g., carboxypeptidases and endoproteinases) and nucleases are added to degrade nucleic acids, which can bind proteins and ampholytes and lead to a horizontal streak pattern [199].

Table 8 Commonly used chemicals for protein sample preparation [199]

Group	Chemical
Chaotropes	Urea, Thiourea
Detergents	
Ionic	SDS, deoxycholic acid
Nonionic	Triton X-100, Tween 20, Nonidet NP-40, CYMAL-5, CYMAL-6
Zwitterionic	Zwittergent 3-10, CHAPS, FOS-Choline-10
Reducing agents	
Thiols	DTT, Mercaptoethanol
Phosphines	TBC, TCEP
Protease inhibitor	PMSF, Cocktails
Chelating agents	EDTA, EGTA
Nuclease	DNase I, RNase A

Many advances have been made in recent years to improve sample preparation protocols, in particular to improve the solubilization of hydrophobic membrane proteins. Among the improvements are the introduction of the chaotrope thiourea [206], of new detergents (compared by Henningsen et al. [207]), such as caprylyl sulphobetaine (SB 3-10) [208], amidosulphobetaine 14 (ASB 14) [209, 210], and C8∅ [209], and of the reducing agent tributyl phosphine (TPB) [211]. Further improvements, as well as details about function, advantages, and disadvantages of buffer components, have been thoroughly discussed by Righetti et al. [195, 212], Molloy [194], Shaw et al. [204], and Herbert et al. [211, 213].

Different extraction procedures have been used to extract plant proteins to enhance the protein extraction efficiency and the 2DE pattern resolution. The most frequently used methods are (i) TCA/acetone precipitation of proteins and resolubilization in an IEF compatible extraction buffer [214] and (ii) direct extraction with extraction buffer [187]. Precipitation of proteins is advantageous because it: (i) concentrates the proteins from low protein sources (leaf tissues), (ii) inactivates proteases and phenoloxidases [205], and (iii) removes interfering compounds, such as salts, polysaccharides, organic acids, lipids, polyphenols, pigments, and terpenes [149, 201, 205]. The main drawback of methods based on precipitation/resolubilization steps is the risk of losing proteins due to incomplete precipitation or resolubilization [215, 216]. Direct solubilization in an IEF buffer does not have this risk, but its utilization is limited by the amount of interfering compounds in the plant tissue.

Because of differences in physicochemical properties and the large dynamic range of proteins in the sample tissue, there is no single method that is able to display all proteins on a single 2DE gel. Prefractionation of subcellular organelles [217] or complex protein samples prior to 2DE [reviewed by 218-221]; or sequential extraction procedures [208, 215] aim to solve these problems. These methods allow the detection of low-abundance proteins [219, 222], membrane proteins [219], and significantly increased the total number of detected proteins [215, 223]. However, quantification of proteins is more complicated or not possible [222] because one protein is often distributed into many fractions, and protein recovery rates are difficult to determine [224].

The optimal protocol depends greatly on the type of sample, e.g., leaf, root, seedling, seed, and the experimental objectives. These include investigation of expression changes and analysis of the protein composition of a given cell type (display of whole or organelle proteome), of classes of proteins (cytosolic, membrane, nuclear), or of protein-protein interaction (protein aggregates). It has to be determined empirically.

2.6.1.2 First dimension: IEF with immobilized pH gradients

Isoelectric focusing is used in the first dimension to separate proteins according to their isoelectric point (pI). The pI is the pH at which a protein has no net charge and hence does not migrate in an electric field. The net charge of a protein is the sum of all negative and positive charges of the amino acid side chains of the protein. Side chains of lysine, arginine, and histidine are positively charged (protonated) at low pH values and neutral at high pH values, while side chains of tyrosine, cysteine, glutamate, and aspartate are negatively charged (deprotonated) at high pH values and neutral at low pH values. Glycosylation and phosphorylation of the protein also influence its net charge. During the IEF separation, proteins migrate in a pH gradient towards the anode or the cathode to the pH values where their net charges are zero: their isoelectric points (pI). Should a protein diffuse away from its pI, it would gain a charge and the electric field would move it back toward its pI. The proteins become highly concentrated at their pIs, which results in a high sensitivity for detection. Charge differences as small as 0.01 pH can be differentiated [225].

In the traditional method, described by O'Farrell [187] and Klose [188], the IEF step was carried out with carrier ampholytes (CAs) generated pH gradients in tube gels. Carrier ampholytes are synthetically derived low molecular mass compounds with ionisable amino

and carboxyl groups. During CA-IEF, a pH gradient is formed in an electric field by the migration of CAs to their respective pIs. Drawbacks of this method include the susceptibility to cathodic drift during long focusing times [194, 226], distortion of the pH gradients due to high salt concentration of samples [189] and high protein load [187], and the fragility of the tube gels. The introduction of plastic film-supported 'immobilized pH gradients' (IPGs) in the 1980s has helped to overcome many of the limitations of CA-IEFs [189, 190, 226]. For IPGs, the pH gradient is formed by copolymerization of buffering compounds (Immobiline™) with acrylamide and cross-linker into the polyacrylamide matrix [227]. IPGs permit higher resolution [190] and higher protein loading capacity [227]. They also eliminate the problem of cathodic drift [228] and allow separation of basic and acidic proteins under equilibrium conditions [200, 228]. In addition, IPGs can be precisely tailored (linear, stepwise, sigmoidal) to maximize separation of specific proteins, and the reproducibility of the 2DE pattern is thus improved [226, 229]. Interlaboratory comparisons with 2DE gels using IPG gel strips showed that spot position reproducibility is high, with a mean standard deviation along the x-axis (pH) of approximately 1.5 mm for a 180-mm IPG gel [229]. Although IPGs provide superior loading capacity to CAs, investigations have shown that the increase in protein loading did not result in the visualization of more protein *per se*. Protein spots start to merge with other spots at high protein loads while others remain at the limit of visualization [225]. In addition, proteins in complex samples can co-migrate to the same position in a gel [230]. Higher resolution can be obtained using narrow pH range IPGs (down to 1 pH unit) [200, 231]. The enormous resolution power of narrow IPG technology was demonstrated for two different organisms, *Eschericia coli* and *Saccharomyces cerevisiae.* Within 1 pH unit, up to 1000 *E. coli* proteins which corresponded to 85% of the E. coli proteome of this pI range (pH 5.1-5.6), and 2700 *S. cerevisiae* protein spots (pH 4.8-5.5) were resolved [225]. IPG gels strips are commercially available as a broad range (pH 3-11 [232]) or as a narrow range (down to 1 pH units).

Samples can be applied to the IPG strips by applying them directly to the rehydrated strip via sample cups (cup-loading) during IEF [190] or by in-gel rehydration [233, 234] prior to IEF. For the in-gel sample rehydration, the protein sample is included in the rehydration (reswelling) solution of the IPG strip. Although cup-loading is the classic method, it often resulted in protein precipitation at the application point and streaking of unsolubilized proteins. The use of in-gel rehydration has eliminated sample streaking and precipitation at the application point and has also allowed larger quantities of proteins to be loaded and

separated [233, 234]. Cup-loading is still preferable for the separation of basic proteins (e.g., pH 7-11) as it provides higher resolution and better gel-to-gel reproducibility for those basic proteins than the in-gel rehydration method [235].

2.6.1.3 Second dimension: SDS-PAGE

In the second dimension, proteins are separated according to their molecular weight. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in a Tris-chloride / Tris-glycine buffer [236] is commonly used as the second dimension for 2DE. The mobility of proteins is inversely proportional to their masses and is also affected by the sieving properties (pore size) of the gel. Improved resolution of low molecular weight peptides (< 14 kDa) can be achieved with Tris-tricine gels, according to Schaegger and von Jagow [237]. Although similar results are achieved with vertical and horizontal flatbed systems [238], the vertical system is more practical because multiple gels can be run in one unit at the same time. Zhan and Desiderio (2003) [239] found higher repeatability of spot position and spot quantity when a vertical system was used. The availability of homogeneous (single-percentage acrylamide) and gradient precast gels in multiple sizes contributes to the reproducibility of 2DE pattern within and among laboratories [138, 229].

Before application of the IPG strips to the SDS-PAGE gels, the separated proteins in the IPG strips have to be saturated (coated) with SDS. SDS is a denaturing anionic detergent and forms negatively charged SDS-protein complexes with constant negative net charges per unit mass (1.4 g SDS/g protein). Hence, the intrinsic net charge of the proteins is masked and the separation is exclusively based on the molecular mass of the proteins. In addition to protein coating by SDS, all cysteine (Cys) residues are reduced and alkylated in a 2-step equilibration process according to Görg et al. [190]. The alkylation of thiol groups on proteins prevents their reoxidation and minimizes unwanted reactions of cysteine residues during electrophoresis. This process may result in streaking and other artifacts on the 2DE gel as well as hampering the protein identification by mass spectrometry.

Recently, Herbert et al. [240, 241] and Galvani et al. [242, 243] have suggested that the reduction and alkylation of sample proteins should be carried out prior to IEF in order to improve 2DE resolution in the alkaline region, above pH 8, and to abolish acrylamide adduct formation during the SDS-PAGE run. They also point out that SDS can strongly quench the alkylation reaction [243] and that thiourea in the rehydration buffer depletes the

iodoacetamide used for alkylation [242]. However, the equilibration method according to Görg et al. [190] has continued to be the most common method utilized for 2DE.

2.6.1.4 Detection of proteins

In order to analyze the presence and abundance of separated proteins in 2DE, the proteins must be stained. The major challenge is the large dynamic range of proteins in biological samples; it can span to 7 or 8 orders of magnitude [137, 225, 244] and no staining method can accurately quantitate proteins over this range [245]. Even the most sensitive gel staining methods are only quantitative up to five orders of magnitude [244]. Another challenge is the wide range of amino acid compositions, isoelectric points, molecular weights, and post-translational modifications, all of which can affect the staining ability of those proteins [246, 247]. Subjecting identical gels to different stains will result in the identification and quantitation of different protein pattern [247]. It is critical to utilize a stain for proteome analysis that allows for accurate quantitation of proteins and for maximizing protein visualization. The technique used depends on the desired sensitivity, dynamic range, linearity, and ease of use. The most common detection methods are radioactive labeling, Coomassie staining, silver staining, and the fluorescent staining or labeling methods. Brief descriptions of each method are provided below.

Radioactive labeling with fluorography or phosphor imaging is the most sensitive method available for protein detection (~ 200 fg/protein spot [248]). It provides a dynamic range of more than five orders of magnitude when storage phosphor imagine technology is utilized as the detection method [245]. This method is limited, as the sample must be radiolabeled *in vivo* or *in vitro* during the cellular growth phase using ^{35}S-methionine or ^{14}C-amino acids to employ this technique. The utilization of radioactive labeling for the investigation of proteomes is not practical. However, it is a sensitive method for measuring changes in protein expression due to environmental stress or chemical treatment.

Silver staining offers the greatest sensitivity (~ 0.5 ng protein [249]) of the non-radioactive protein detection methods [245] and has become the standard method for detecting trace amounts of proteins [249]. Many protocols have been published for silver staining of 2DE gels based on two kinds of silver staining: acidic silver nitrate and alkaline silver diamine [250]. Both approaches utilize the reduction of silver ions that have been attached to the proteins to metallic silver. The physico-chemical principles underlying these two methods

have been reviewed by Rabilloud et al. [251]. Although silver staining is a standard method in many proteomics laboratories, it has many disadvantages, such as (i) the time-consuming and laborious multi-step staining process, requiring numerous solution changes and precise timing, (ii) the extreme protein-to-protein variability in staining [252] (based on amino acid composition and post-translational modifications) [203, 245, 251, 253], some proteins are un-stained [245] and the linear dynamic range is narrow (4-60 ng protein [252]) [249, 254, 255], and (iii) the lack of specificity. DNA and lipopolysaccharides are also stained by this method [249] and can contribute to interfering background staining in the gels. Standard silver staining protocols use glutaraldehyde and formaldehyde, which are known to alkylate α- and ε-amino groups of proteins [250]. Richert et al. [256] suggest that there is a formation of protein reticulation by methylene bridges due to the formaldehyde reaction with amino acid reactive side chains. This interferes with subsequent protein analysis by N-terminal sequencing or mass spectrometry. New silver staining protocols omit glutaraldehyde and formaldehyde and thus overcome the compatibility problems with mass spectroscopy. The staining sensitivity of such modified silver staining methods is decreased [218, 252, 256].

Coomassie Brilliant Blue (CBB) staining is the most widely used staining method. While not as sensitive as silver staining methods, it has a broader linear dynamic range (50-1000 ng [249]) [203]. CBB dye stains proteins in an acidic environment due to ionic interactions between the dye and the basic amino acid residues of the protein, as well as due to hydrogen bonding, van der Waals attraction, and hydrophobic bonding [245]. Coomassie protocols based on Coomassie Brilliant Blue R-250 dye, methanol, and acetic acid or trichloroacetic acid (TCA) require destaining steps and provide only a moderate sensitivity of 40 ng protein per spot [138]. The introduction of colloidal CBB staining using CBB G-250 (dimethylated CBB R-250) by Neuhoff et al. [257] provided a fast, sensitive, and background-free detection method of proteins in polyacrylamide gels. The addition of methanol and the increased ammonium sulfate concentration in the dye solution further improved the method [258, 259] and allowed complete staining of protein bands/spots throughout the entire cross section of the gel layer. Neuhoff et al. [258] were able to clearly detect 0.5 ng of BSA, while other authors indicate that 8 to 10 ng of protein is the limit of detection for colloidal CBB stain [250; 186]. The absolute sensitivity and staining linearity, both depend on the proteins being stained. However, Berggren et al. [252] showed that for four proteins the protein-to-protein variation in staining intensity was much smaller with CBB than with silver staining. Mahon and Dupree [254] demonstrated that protein quantities up to 8 μg per 2DE spot could be

quantified. CBB staining methods are compatible with mass spectrometry or protein sequencing methods [260] as long as TCA / alcohol mixtures are omitted from the staining solution to minimize modifications of glutamic and aspartic acid side chains [252].

Recently introduced fluorescent dyes, sold under the trade name SYPRO (Molecular Probes, Inc., Eugene, OR), promise to improve the detection and comparison of proteins in 2DE gels. They can be divided into two categories according to their staining mechanism. Both groups interact noncovalently with protein but SYPRO Orange and SYPRO Red interact with the SDS-protein micelle [261, 262], whereas SYPRO Ruby interacts with the basic amino acids in proteins at an acidic pH by a CBB-type mechanism [250]. SYPRO Orange and SYPRO Red can detect protein amounts as little as 4-10 ng and demonstrate a broad linear dynamic range of over 2.5 to 3 orders of magnitude [262]. SYPRO Ruby is even more sensitive, detecting 1-2 ng protein with a linear dynamic range of 1 to 1000 ng [252]. Another advantage of these stains is their compatibility with downstream analysis, such as MALDI-TOF-MS, sequencing by LC-tandem MS [263], and Edman sequencing [250]. Mackintosh et al. [264] recently introduced a new fluorescent dye based on a polyketide natural product. Epicocconone, the active ingredient of the dye, binds to the SDS of the protein-SDS micelle. Lightning Fast (commercial name: Deep Purple™ Total Protein stain) was found to be an order of magnitude more sensitive than SYPRO Ruby and exhibited quantitative linearity over more than four orders of magnitude. Also, Lightning Fast is compatible with subsequent analyses such as N-terminal sequencing and MALDI-MS [264]. The authors estimated that Lightning Fast can detect protein (50 kD) as low as 12 copies per cell [264].

Another very promising proteomics approach using fluorescent dyes is two-dimensional differential gel electrophoresis (2D DIGE) [265]. Three succinimidyl esters of the cyanine dyes are commercially available: Cy2, Cy3, and Cy5. They are structurally similar but have spectrally different fluorescence [265]. The dyes react with the epsilon amino group of lysine residues in proteins. Thus, sample proteins can be covalently labeled prior to 2DE with any of those three dyes. This enables the running of three different samples simultaneously on the same 2DE gel, which reduces method related variation and eases the qualitative and quantitative comparison of protein profiles [265-267]. Because the protein labeling with the cyanine dyes is covalent, the pI of proteins are decreased and the molecular weight increases. This effect is more noticeable for very basic and lower molecular weight proteins [247, 265].

Although these shifts do not affect the 2D DIGE analysis, it complicates the identification of proteins by MS and the comparison with 2DE profiles in public databases [247, 265].

2.6.1.5 Image acquisition and analysis

The complexity of the data in 2DE gels requires the use of computer-aided analysis. Therefore, images of stained gels have to be digitized for computer analysis by either an image scanner, a laser densitometer, a phosphor imager (autoradiography), a fluorescence scanner, or a charge-coupled device (CCD) camera. Most systems digitize gels with a resolution (pixel size) of 100 to 200 μm, yet some laser densitometers have resolution up to 35 μm. These systems can detect a wide range of densities or grayscales depending on their bit depth (e.g., 8 bit system = 256 grayscales, and 16 bit system = 65,536 grayscales). CDD cameras and image scanners (reviewed in [268, 269]) are applicable for a variety of stains but also more susceptible to grayscale saturation effects than other devices, such as laser densitometers [141].

Following image acquisition, gel images are subjected to analysis software packages for background correction, removal of vertical and horizontal streaking, spot detection, spot quantitation, and spot matching (spot assignment across different gels). A standard spot (SSP) number is assigned to each spot that contains vertical and horizontal positional information. Spot quantity (volume) is calculated by spot area and spot density. A number of image analysis software packages have been developed to facilitate rapid and accurate differential comparisons. Raman et al. [270] provide a good overview of such commercial packages. Among the most frequently used programs are ImageMaster 2D Elite (Amersham Pharmacia Biotech, Uppsala, Sweden), PDQuest (Bio-Rad Laboratories, Hercules, CA, USA), Melanie 3 (Geneva Bioinformatics (GeneBio) S.A., Geneva, Switzerland), Z3 (Compugen, Tel Avivi, Israel), and Progenesis (Nonlinear USA Inc, NC, USA).

Various authors [249, 254, 270, 271] have tested the precision and reliability of several analysis programs and have found that all the systems suffer from errors in spot recognition, quantification, and matching. Mahon and Dupree [254] found that the relative quantitation error in a repeated scan experiment fell between 1-10%. The error was found to be five to ten-fold larger when different gels with the same sample were analyzed. Roger et al. [271] compared the performance characteristics of the widely used gel image analysis packages. They showed that ImageMaster is the most accurate package for spot detection with up to

85% true positive hits, whereas PDQuest is the most robust for spot overlap. These data were obtained with synthetic gels. Real gels provide more challenges with vertical and horizontal streaking, pattern distortion, and slight differences in gel size. An automatic gel processing system not requiring user intervention is not available.

2.6.1.6 Protein identification

The goal of differential expression proteomics is to find the protein that is induced (new spots), inhibited (missing spots), or changed (differences in spot intensity) between biological samples that have been subjected to environmental stress, chemical treatment, genetic modification, or between genotypes. Numerous gels are required for the statistical separation of protein differences from background noise. Proteins that show differences in expression can then be identified. The position of the protein in the 2DE gel already provides information about the protein of interest. It indicates the isoelectric point (pI) and molecular weight (Mw) of the protein. Accurate estimations of protein pI and Mw can be obtained by using known proteins on a reference map. Gel analysis software programs use these data to construct standard curves of pI and molecular weight, and they calculate an estimate of the pI and Mw of unknown proteins in the 2DE gel. It must be noted, however, that proteins carrying post-translational modifications may migrate to unexpected pI or Mw positions during electrophoresis. As a consequence, the pI and Mw does not give enough information for the identification of a protein.

Traditionally, proteins from 2D gels have been identified by techniques such as immunoblotting, N-terminal micro-sequencing, internal peptide sequencing, co-migration of unknown proteins with known proteins, or by over-expression of homologous genes of interest in the organism under study. While these techniques are powerful identification tools, they are expensive and time- and labor-intensive as well as limited to already known and/or highly abundant proteins [138]. N-terminal sequencing is slow (10 hours for 20 amino acids), expensive, does not provide information about post-translational modifications, and is not easily applicable because more than 50% of the plant proteins are N-terminally blocked [272]. The development of highly sensitive mass-spectrometry techniques, like matrix-assisted-desorption-ionization-time-of-flight-mass-spectrometry (MALDI-TOF-MS) [273, 274] and the availability of EST (expressed sequence tag), genome and protein sequence databases provide very powerful tools for the rapid identification of proteins.

The process of protein identification typically involves several steps. First, an in-gel digest of proteins is performed with a protease such as trypsin. The resulting proteolytic peptides of the protein spot are subjected to MALDI-TOF-MS. MALDI allows soft (nondestructive) ionization of large biomolecules, such as peptides and proteins. Peptides are embedded in light-absorbing matrix and the mixture is desorbed and ionized by a pulsed UV laser. After that, the single-charged peptide ions [275] are accelerated in an electric field and induced into the flight-tube of the mass spectrometer. The time needed to reach the detector is proportional to the square root of their mass-to-charge ratios and provides a highly precise measure of their molecular weights. The peptide masses are subsequently matched against theoretical peptide libraries generated from protein sequences. The matching is usually undertaken in an interactive manner, whereby peaks of mass 500 - 3000 D are selected and compared under various search parameters including Mw of protein, mass accuracy of peptides, artifactual modifications introduced by electrophoresis (acrylamide adduct to cysteine and the oxidation of methionine), and the number of missed enzyme cleavages allowed. The correct protein identity is the protein that has the most peptide masses in common with the unknown sample. The combination of different fragment masses from the same protein, the 'peptide mass fingerprint' (PMF), reveals its identity [276]. The technique will identify proteins from a fully sequenced organism [277], but if a protein is degraded, modified or simply unknown, then the composition and partial sequence information of its proteolytic fragments can be obtained by using tandem mass spectrometry (MS/MS) [278, 279]. Tandem mass spectrometry initially determines the mass of peptides, then subjects the peptides to fragmentation by collision with a gas, and finally determines the mass of the fragments. The resulting spectra give information about a peptide's amino acid sequence (*de novo* sequencing) and post-translational modifications. This sequence tag and the original peptide mass can then be used for highly specific searches of protein and EST databases [277]. These techniques are also capable of dealing with peptide fragment mixtures from two or more proteins (proteins with identical mobility on 2DE gels) [280].

3 Materials and methods

3.1 Materials

3.1.1 Plant lines

3.1.1.1 *A. thaliana* ecotypes

Seeds for the *A. thaliana* ecotypes were obtained from the Arabidopsis Biological Resource Center in Ohio (http://arabidopsis.org/abrc/) with the exception of seeds for Col-0, which were obtained from Monsanto Co., MO, USA. The stock numbers are: C24 (CS906), Condara (CS6175), Cvi-0 (CS6675), Ll-0 (CS6781), Ma-0 (CS6789), Mr-0 (CS6795), Mt-0 (CS6799), Nd-0 (CS6803), Oy-0 (CS6824), Tsu-0 (CS6874), and Ws (CS6891).

3.1.1.2 Transgenic *A. thaliana* lines

GUS-lines

Monsanto Co., MO, USA, provided seeds for six transgenic *Arabidopsis* lines (T3 generation) containing an inserted *β*-glucuronidase (*gus*) gene. The transgenic *Arabidopsis* events were obtained by *Agrobacterium*-mediated transformation with a T-DNA containing the *β*-glucuronidase (*gus*) gene as well as the neomycin phosphotransferase II (*nptII*) gene for antibiotic selection.

Enhanced tocopherol lines

Professor Dr. Dean DellaPenna from the Michigan State University, MI, USA, provided seeds for six tocopherol-enhanced transgenic *Arabidopsis* lines and their parental line (WT-P). The transgenic *A. thaliana* lines contain an inserted p-hydroxyphenylpyruvate dioxygenase (*hppd*) gene, or an inserted γ-tocopherol-methyltransferase (*gtmt*) gene [281, 282].

3.1.2 Chemicals

All chemicals were reagent-grade or better quality and supplied by the firms Sigma-Aldrich (St. Louis, MO, USA), Bio-Rad (Hercules, CA, USA), and Amersham Biosciences (Piscataway, NJ, USA). Chemicals obtained from other suppliers are as follows:

- α-cyano-4-hydroxycinnamic acid (Ciphergen, Palo Alto, CA, USA)
- Atropine Standard (CE Elantech, Inc., Lakewood, NJ, USA)
- Chromic Oxide Granules (CE Elantech, Inc., Lakewood, NJ, USA)

- Indicating Drierite Anhydrous Calcium Sulfate® (W.A. Hammond Drierite Company Ltd., Xenia, Ohio, USA)
- Lysing Matrix E (Qbiogene, Inc., Irvine, CA, USA)
- Metro-Mix 200 soil (Hummert™ International, Earth City, MO, USA)
- Phosphate Buffered Saline (PBS), pH 7.4 (Roche Diagnositics GmbH, Mannheim, Germany)
- Proteaese inhibitor cocktail Complete™ (Roche Diagnositics, Indianapolis, IN, USA)
- Quartz Wool (CE Elantech, Inc., Lakewood, NJ, USA)
- Reduced Copper Wires (CE Elantech, Inc., Lakewood, NJ, USA)
- Silvered Cobaltous – Cobaltic Oxide (CE Elantech, Inc., Lakewood, NJ, USA)
- Trypsin, sequencing grade modified (Promega Co., Madison, WI, USA)

3.1.3 Equipment

- Analytical balance AT262 Delta Range® (Mettler Toledo, Columbus, OH, USA)
- Beckman Φ 360 pH/Temp/mV Meter (Beckman Coulter Inc., Fullerton, CA, USA)
- Bench-top centrifuge Eppendorf 5415 C (Brinkmann Instruments Inc., Westbury, NY, USA)
- Bench-top centrifuge Eppendorf 5804 R (Brinkmann Instruments Inc., Westbury, NY, USA)
- Clay Adams Nutator orbital mixer (Becton-Dickinson, Sparks, MD, USA)
- Costar 96-well microtiter plate (polystyrene) (Corning Incorporated, Corning, NY, USA)
- Costar siliconized microcentrifuge tube (0.65 and 1.7 mL) (Corning Inc., Corning, NY, USA)
- Criterion cell (Bio-Rad Laboratories, Hercules, CA, USA)
- Criterion Dodeca cell (Bio-Rad Laboratories, Hercules, CA, USA)
- FlashEA™ 1112 Protein Analyzer (Thermo Finnigan distributed by CE Elantech, Inc., Lakewood, NJ, USA)
- Gladware® Container Entrée (700 gram polypropylene container with lid) (local grocery store)
- GS-800™ Calibrated Densitometer (Bio-Rad Laboratories, Hercules, CA, USA)
- L-8800 Amino Acid Analyzer (Hitachi High Technologies America, Inc., San Jose, CA, USA)
- Mega Grinder (Monsanto Company, St. Louis, MO, USA)
- Microplate spectrophotometer PowerWave™ Xi (Bio-Tek Instruments Inc., Winooski, VT, USA)
- Molecular Imager® FX System with external Laser (Bio-Rad Laboratories, Hercules, CA, USA)
- Pipet, repeating 100 mL (VWR International, Batavia, IL, USA)
- PowerPac 1000 (Bio-Rad Laboratories, Hercules, CA, USA)
- Precision balance PC4400 Delta Range® (Mettler Toledo, Columbus, OH, USA)
- PROTEAN® IEF cell (Bio-Rad Laboratories, Hercules, CA, USA) with Epson TM-T88III Thermal Printer (Epson, Long Beach, CA, USA)
- ProteomeWorks Spot cutter (Bio-Rad Laboratories, Hercules, CA, USA)
- Quartz Reactor Tube (CE Elantech, Inc., Lakewood, NJ, USA)

- Refrigerated circulator PolyScience, Model 1160-A (VWR Scientific, West Chester, PA, USA)
- Sarstedt screw cap micro tubes 2 mL (VWR Scientific, West Chester, PA, USA)
- Savant Speed-Vac® Concentrator Systems (Savant Instruments, Holbrook, NY, USA)
- Tin Capsules (5 x 9 mm) (CE Elantech, Inc., Lakewood, NJ, USA)
- U.S.A. Standard Testing Sieve, E-11, No. 45 (VWR Scientific, West Chester, PA, USA)
- Vortex Mixer (VWR Scientific, West Chester, PA, USA)
- Voyager-DE™ Pro MALDI-TOF MS (Applied Biosystems, Foster City, CA, USA)
- VWR Heated water bath (VWR Scientific, West Chester, PA, USA)
- Water purification system Milli-Q® plus (Millipore, Billerica, MA, USA)

3.1.4 Software

- Eager 300 for EA 1112 software (CE Elantech, Inc., Lakewood, NJ, USA)
- JMP The Statistical Discovery Software® version 5 (SAS Institute Inc., Cary, NC, USA)
- KC4™ software for PowerWave™ X (Bio-Tek Instruments Inc., Winooski, VT, USA)
- Microsoft® Excel 2000 (Microsoft Corporation, CA, USA)
- MINITAB® version 13 Statistical Software (Minitab Inc., State College, PA, USA)
- PDQuest™ 2-D Analysis Software versions 6 and 7 (Bio-Rad Laboratories, Hercules, CA, USA)
- Phylip 3.6 (available on http://evolution.genetics.washington.edu/phylip.html)
- SAS® version 9 (SAS Institute Inc., Cary, NC, USA)

3.2 Methods

3.2.1 Growth of plants

Seeds were sown in Metro-Mix 200 soil saturated with tap water in individual 2.5-inch pots. The pots were arrayed in a 4 x 8 grid in a standard greenhouse tray and covered with germination domes. After sowing, the flats were cold-treated (vernalized) at 4°C for four days, then placed in a growth chamber at 20°C, 70% relative humidity, 16h light (150-200 $\mu E/s/m^2$). The plants were watered and fertilized (100 ppm Peter's 20:20:20) twice weekly via sub-irrigation. Germination domes were removed after one week from plant date. The plants were thinned to one plant per pot and the pots were placed in a checkerboard pattern within the tray to ensure adequate lighting and ventilation for all the plants. After first siliques were formed, the upper part of the plants (inflorescence) was wrapped into transparent plastic tubes. The tubes were fastened just above the rosette leaves and kept widely open to reduce the collection of moisture from transpiration.

3.2.2 Harvesting

Seeds were harvested after complete maturity of all the seeds on a plant by cutting the branches above the rosette leaves and threshing the seeds in a plastic tube. The seeds were then separated from chaff and plant debris by sieving and aspirating before being placed in weighed paper envelopes.

3.2.3 Storage

In order to avoid seed deterioration, the paper envelopes were placed in freezer bags containing desiccant (Drierite Anhydrous Calcium Sulfate®) and stored at 4°C.

3.2.4 Phenotypic analysis and methodology

Four phenotypic traits, i.e. first flowering date (FFD), rosette diameter (RD), seed yield, and seed protein content, were assayed. The FFD is the number of days from the date of planting until the opening of the first flower and was assayed by daily inspection of the plants. The RD is the diameter (in cm) of the leaf rosette at the time of first flowering. The seed yield is the amount of harvested seeds for one plant. It was measured by weighing the paper envelopes before and after adding the seeds and subtracting the envelope weight from the total weight. Leaf and stem morphology was visually assessed by the overall shape, length, thickness, and pubescence.

The seed protein content was determined using the FlashEA™ 1112 Protein Analyzer. The protein analyzer operates according to the Dynamic Flash Combustion technique, a modified Dumas method. The seed samples were weighted in a tin capsule and introduced into the combustion reactor. After combustion (1000°C; oxygen flow: 280 mL/min), the produced gases are carried by a helium flow (145 mL/min) to a second reactor (reduction column; 780°C) filled with copper, then swept through CO_2 and H_2O traps, a GC column and finally detected by a thermal conductivity detector. The protein content was calculated with the Eager 300 software using the protein factor 6.25. The analysis was performed in replicates of the pooled seed samples, also used for the 2DE analysis.

3.2.5 Protein extraction

The carrier ampholytes stock used in preparation of all solutions was a 2:1:3 mixture of Ampholine 3.5-9.5, Pharmalyte 5-8, and Bio-Lyte 3-10.

3.2.5.1 Extraction buffers

Low detergent extraction buffer (EB 1)

7 M urea, 2 M thiourea, 0.75% (w/v) CHAPS, 0.75% (v/v) Triton X-100, 100 mM DTT, 1% (v/v) carrier ampholytes, 20% (v/v) isopropanol, protease inhibitor cocktail (1 tablet "complete" for 50 mL extraction buffer).

High CHAPS extraction buffer according [142] (EB 2)

7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.2% (v/v) Triton X-100, 14 mM DTT, 1% (v/v) carrier ampholytes, 18 mM Tris-HCl, 50 U/mL DNase, 5 Kunitz U/mL RNase A, protease inhibitor cocktail (1 tablet "complete" for 50 mL extraction buffer).

CHAPS/SB3-10 extraction buffer according [208] (EB 3)

5 M urea, 2 M thiourea, 2% (w/v) CHAPS, 2% (w/v) SB 3-10, 100 mM DTT, 0.5% (v/v) carrier ampholytes, 40 mM Trizma base, protease inhibitor cocktail (1 tablet "complete" for 50 mL extraction buffer).

Non-denaturing extraction buffer: PBST (EB 4) and PBST+DTT (EB 4+DTT)

1 mM KH_2PO_4, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4, 0.1% Tween, 100 mM dithiothreitol (only EB4+DTT), and protease inhibitor cocktail (1 tablet "complete" for 50 mL extraction buffer).

Extraction buffers were prepared without DTT and stored in aliquots at -20 °C. DTT was added before use.

3.2.5.2 Mortar and pestle grinding method

Seeds (~150 mg) were ground into a fine powder in a mortar and pestle in the presence of liquid nitrogen. The powder was resuspended directly in the mortar with an extraction buffer with continued homogenization until the homogenate reached room temperature (RT). The homogenate was transferred to a screw cap tube and mixed on a Nutator mixer for one hour at RT. The homogenate was centrifuged 16,000 × *g* for five minutes at RT. The supernatant was removed with a syringe and recentrifuged. The resulting supernatant was stored in aliquots at –80°C.

3.2.5.3 Mega Grinder method

The Mega Grinder utilized in these studies was designed and built by Monsanto Company and Automated Technology, Inc. (Fenton, MO, USA) to allow quick, efficient grinding of seed samples ranging from a single seed to bulk samples. It utilizes extremely rapid shaking (up to 3,000 strokes/minute) to impact grinding beads into intact seeds causing the seeds to become pulverized. The Mega Grinder consists of a two horsepower electric motor that drives a crankshaft via a belt. The crankshaft drives a piston that holds the sample trays in an up-and-down motion. The speed of the up-and-down motion can be adjusted from 0 to 1600 rpm. Seed samples (30-40 mg) were weight into a screw cap tube and the grinding beads were placed on top of the samples. The tubes were put into a 24-format aluminum shaker tray and pre-cooled at –80°C for at least two hours. The filled shaker trays were then placed in the Mega Grinder and shaken for 45 seconds at 1500 rpm. The sample tubes were placed on dry ice immediately after the grinding process. Extraction buffer (0.7 mL) was added to each tube and after brief vortexing, the samples were mixed on a Nutator for one hour at RT and subjected to centrifugation at 16,000 × *g* for five minutes at RT to remove insoluble material. Finally, the supernatants were removed with a syringe, and after an additional centrifugation step, they were stored in aliquots at –80°C.

3.2.5.4 Defatting of seed

Hexane (1 mL) was added to the seed powder (0.7 mg), vortexed, and incubated in a water bath for 15 minutes at 50°C. After centrifugation at 16,000 × *g* for 10 minutes, the supernatant was removed with a glass pipette. This process was repeated three more times with fresh hexane. After vacuum desiccation, the defatted samples were stored at -80°C.

3.2.5.5 Protein quantification

Protein concentration was estimated using the Bio-Rad Protein Assay (Catalog No. 500-0002) based on the dye-binding method of Bradford [283]. Protein quantitation was performed according to the instruction manual (Bulletin No. 9004) using the microtiter plate format. Bovine serum albumin (BSA) was used to generate the standard curve. The samples and BSA were diluted in a solution containing 3 M urea and 3 M guanidine-HCl to avoid protein precipitation.

3.2.6 IEF-PAGE

Protein extracts were diluted in sample buffer containing 6 M urea, 2% (w/v) CHAPS, 2% (v/v) carrier ampholytes, and 15% glycerol. A 4% self-cast 12-well Criterion PAGE gel containing 6 M urea, 2% (w/v) CHAPS, and 2% (v/v) carrier ampholytes was prefocused at 50 V for 20 minutes before the diluted samples (10 μg) were loaded on the cathodic side. The IEF running program was 30 min at 50 V, 1 h at 100 V, and 2.5 h at 150 V. The gel was fixed in 3.5% (w/v) 5-sulfosalicylic acid, 12% TCA for 30 min and stained with Bio-Rad IEF gel stain.

3.2.7 Two-dimensional gel electrophoresis (2DE)

3.2.7.1 First dimension

Protein extracts were diluted in rehydration buffer containing 7 M urea, 2 M thiourea, 0.75% (w/v) CHAPS, 0.75% (v/v) Triton X-100, 100 mM DTT, 0.3% (v/v) carrier ampholytes, 10% (v/v) isopropanol, 12.5% (v/v) water saturated isobutanol, 1 tablet/50ml protease inhibitor and a trace bromophenol blue. The diluted samples were vortexed and centrifuged at 16,000 × *g* for 3 min prior to strip rehydration. Thirteen centimeter long immobilized pH gradient gel strips (IPG) with nonlinear pH 3-10 gradients (Amersham Biosciences) were cut 0.5 cm on both sides to fit in the IEF unit. The gel strips were rehydrated using 230 μl of diluted sample (150 μg total protein). Each strip was overlaid with approximately 1 mL of mineral oil and left undisturbed for 20-23 h at room temperature. The rehydrated strips were transferred to the focusing tray and covered again with mineral oil. The IEF was carried out using a Bio-Rad PROTEAN® IEF cell with a controlled cell temperature of 20°C and with a maximum current of 50 μA/strip. The running conditions were: from 0 to 500 V in 3 h, from 500 to 4,000 V in 6 h, and a final phase of 4,000 V to a total of 35,000 Vh. After IEF was completed, the strips were stored at - 80°C until required for the second dimension.

3.2.7.2 Second dimension

Prior to applying the focused IPG strips to the second dimension, the IPG strips were equilibrated first for 10 min in 50 mM Tris-HCl (pH 8.8), 6 M urea, 30% (w/v) glycerol, 2.3% (w/v) SDS, 1% (w/v) DTT, bromophenol blue and then another 10 min in the same solution except DTT was replaced with 4% (w/v) iodoacetamide. The equilibrated IPG strips were then applied to 8-16% Tris-HCl linear gradient Criterion™ gels and sealed with 1x

Tris/Glycine/SDS running buffer (Bio-Rad) with 0.9% (w/v) low-melt agarose and a trace of bromophenol blue. The gels were run for the first 15 min at 130 V and then at 180 V until the tracking dye reached the bottom of the gel. Twelve Criterion gels were run in one Bio-Rad Criterion Dodeca cell at the same time for increased reproducibility. The refrigerated circulator was set to 9.5°C with a high pump speed. The buffer in the lower tank was stirred with a stir bar to maintain a constant buffer temperature throughout the entire cell. When individual Criterion cells were used for the second dimension, the 1x Tris/Glycine/SDS running buffer was cooled down to 4°C prior to use.

3.2.7.3 Molecular weight standards

Bio-Rad prestained broad range SDS-PAGE standards were diluted 1:3 in hot 1x Laemmli sample buffer (Bio-Rad) with 100 mM DTT and 1% (w/v) agarose. The mixture was drawn into 1.5 mm diameter glass tubes, solidified, and then cut into 1-2 mm long pieces. A molecular weight standard in agarose was positioned near the basic end of the IPG strip and sealed into place with a drop of hot agarose solution used to overlay the IPG strips. For the Fluorescent stain, unstained Bio-Rad broad range molecular weight standards were used and prepared the same way as described above except the standards were diluted 1:80.

3.2.7.4 Staining methods

3.2.7.4.1 Silver stain

The commercially available silver staining kits, ProteoSilver™ Plus Silver Stain Kit (Sigma-Aldrich) and Silver Quest™ Silver Staining Kit (Invitrogen Corporation, Carlsbad, CA, USA), were used according to their instruction manuals (Technical Bulletin PROT-SIL2 and SilverQuest™ Version C IM-6070). Imaging of the stained proteins was performed using the GS-800™ Calibrated Densitometer (Silver stain gel application, white filter, transmissive mode, 36.3x36.3 microns scan resolution, 0-3.0 OD absorbance range).

3.2.7.4.2 Fluorescent stain

SYPRO® Orange

SYPRO® Orange (Bio-Rad Laboratories) protein staining was performed according to a protocol optimized for 2DE gels by Malone et al. [284]. The gel was incubated in 100 mL fixative (40% ethanol, 2% acetic acid, 0.0005% SDS) for 2 h. The fixative solution was

discarded and replaced with 100 mL washing solution containing 2% acetic acid, 0.0005% SDS. The gel was incubated in this solution for 30 minutes. This wash step was repeated two additional times for a total of three washes. Finally, the gel was incubated in 100 mL staining solution (1:5000 dilution of SYPRO® Orange protein stain in 2% acetic acid, 0.0005% SDS) for 2 h. After briefly rinsing the gel in Milli-Q deionized water, the Bio-Rad Molecular Imager™ FX was used to image the gel using the pre-programmed application 'SYPRO® Orange Protein Stain Gel'.

SYPRO® Ruby

Protein staining using SYPRO® Ruby (Bio-Rad) was completed by following the manufacturer's instructions. In short, the 2DE gel was washed for 30 minutes in 150 mL of 10% methanol, 7% acetic acid. After replacing the washing solution with 100 mL SYPRO® Ruby gel stain, the gel was stained for 3 h with continuous gentle agitation. To reduce background fluorescence and increase sensitivity, the gel was washed in 10% methanol, 7% acetic acid for 30 minutes. The Bio-Rad Molecular Imager™ FX was used to image the gel using the pre-programmed application function 'Protein Stain Gel/SYPRO® Ruby'.

All incubation steps were performed at room temperature on an orbital rotator. Because SYPRO® Orange and SYPRO® Ruby dyes are light-sensitive, the gel staining was carried out in aluminum foil-wrapped staining dishes (Gladware® Container Entrée).

3.2.7.4.3 Colloidal Coomassie Brilliant Blue

Colloidal Coomassie Brilliant Blue (CBB) protein staining was performed according to Neuhoff et al. [258]. Upon completion of electrophoresis, the gel was incubated in 100 mL fixative (30% methanol, 7% acetic acid) for 1 h. The fixative solution was discarded and replaced with 100 mL CBB staining solution. The CBB staining solution was prepared fresh by mixing 80 mL of 0.1% (w/v) CBB G-250 in 2% (w/v) phosphoric acid, 11% w/v ammonium sulfate with 20 mL of methanol. The gel was incubated in this solution for three days. After carefully washing the gel with Milli-Q deionized water to remove any excess of CBB dye, the gel was detained for 2 h in 100 mL Milli-Q deionized water. A paper towel was added to bind excess CBB. For stable fixation and storage, the gel was transferred into 100 mL 20% (w/v) ammonium sulfate in water and incubated for a minimum of 1 h before scanning the gel. All incubation steps were performed at room temperature on an orbital rotator. Imaging of the stained proteins was performed using the GS-800™ Calibrated

Densitometer (Coomassie Blue gel application, red filter, transmissive mode, 36.3x36.3 microns scan resolution, 0-3.0 OD absorbance range).

3.2.7.5 Image analysis

3.2.7.5.1 Spot detection

The scanned images of the 2DE gels were processed and analyzed with PDQuest 2-DE Gel Analysis software Version 6 or 7.1 (Bio-Rad). The images were cropped and oriented using the image editing controls of the program. All images were processed with the following software settings for spot detection and background subtraction: Sensitivity: 40; Size scale: 3; Min peak: 400; PowerMean 3x3, Floater 97, Speckles filter. This spot detection parameters were chosen as they allowed the detection of the majority of protein spots above limit of detection ($LOD = OD_{background} + 3 \times SD_{background}$ [285, 286]) without detecting image noise as spots. Spots detected by the software program were manually verified. False positive spots (e.g., artifacts and multiple spots in a cluster) were manually removed; false negative spots (obviously missed spots with OD > LOD) were added to the images. A spot was considered to be reproducibly present/absent when it is present/absent in all three replicate gels of one extraction.

3.2.7.5.2 Protein expression comparison

In order to compare spots across gels, a match set was created from the images of the gels in an experiment. A standard gel (Master) was generated out of the image with the greatest number of spots. Spots, reproducible present in a match set member but not present in the image with the most spots, were manually added to the standard gel. The automated matching tool of the PDQuest software package was used to match spots across the gels. A few landmarks were manually defined to improve the automated matching results. All spots matched by the software program were manually verified. The spots were quantified by 2-D Gaussian modeling. Spot quantities of all gels were normalized to remove non-expression-related variations in spot intensity, so the raw quantity of each spot in a gel was divided by the total quantity of all the spots in that gel that have been included in the standard. Data were exported to Excel and from there to JMP and/or SAS for statistical analysis. The repeatability of spot position and quantity was calculated with JMP. The coefficient of determination (R^2) and lack of fit for the linearity study were calculated with the statistical

software package JMP. One-way analysis of variance (ANOVA) was performed with JMP or SAS for each spot to identify significant differences in spot quantities of 2-fold and greater with a significance level of 0.05. To avoid overestimation of quantitative differences due to inaccurate quantification of poorly resolved protein spots, all spots with the maximum value and an average spot quantity less than or equal to 40 were excluded from the data set. In addition, the detected statistically significant difference (P<0.05) must have a power above 80%. Protein patterns resulting from triplicate 2DE gels were compared.

3.2.7.5.3 Mw and pI calibration

The Mw and the pI were calibrated by using a commercial mixture of 2DE standard proteins (2-D SDS-PAGE standards, Bio-Rad) covering a mass range of 17.5-76 kD and a pI range of 4.5-8.5 (Table 9). The 2DE standard (13 µL) was added to an *A. thaliana* (Col-0) seed extract (34 mg total protein) and 2DE was performed. Figure 3 shows the spot positions of the 2DE standard proteins within the *A. thaliana* (Col-0) seed proteome. The molecular weight and isoelectric point values for the known protein spots as well as the molecular weight values for the Bio-Rad pre-stained broad range SDS-PAGE standards (not shown in Figure 3) were entered and PDQuest calculated Mw and pI values for all the spots in the match set. Fifteen spots of the *A. thaliana* (Col-0) seed proteome were chosen as Mw/pI reference spots and their Mw and pI values were used to determine the Mw and pI of the proteins in all studies.

Table 9 Constituent proteins of Bio-Rad 2-D SDS-PAGE Standards

Protein	pI	Mw (D)
1. Hen egg white conalbumin	6.0, 6.3, 6.6	76,000
2. BSA	5.4, 5.5, 5.6	66,200
3. Bovine muscle actin	5.0, 5.1	43,000
4. Rabbit muscle GAPDH	8.3, 8.5	36,000
5. Bovine carbonic anhydrase	5.9, 6.0	31,000
6. Soybean trypsin inhibitor	4.5	21,500
7. Equine myoglobin	7.0	17,500

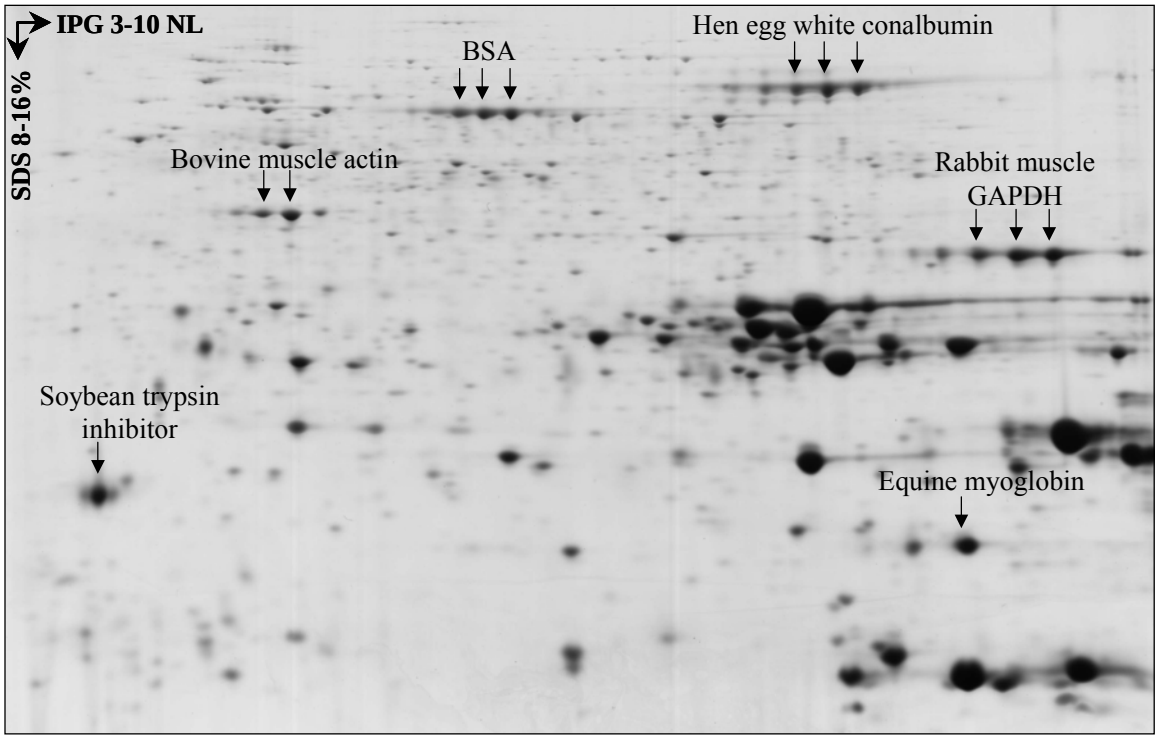


Figure 3 The 2DE pattern of Bio-Rad 2-D SDS-PAGE standards within the *A. thaliana* seed proteome. Standards (13 µL) and *A. thaliana* seed extract (34 µg total protein) were loaded onto a pH 3-10 NL IPG and separated on an 8-16% SDS-PAGE Criterion gel, then stained with colloidal CBB.

3.2.8 Protein identification by MALDI-TOF MS

3.2.8.1 In-gel tryptic digestion

Protein spots were excised from the stained 2DE gels using Bio-Rad's Spot cutter and placed into siliconized microcentrifuge tubes. Proteins were in gel digested with trypsin according to the published procedure [287], except that the alkylation step was omitted, having been included in the 2DE gel procedure. Briefly, the gel pieces were washed with 100 µL of 50 mM ammonium bicarbonate for 15 min at room temperature, dehydrated by addition of 100 µL acetonitrile (50% (v/v) final concentration), and incubated for additional 15 min at room temperature. These three steps were repeated two additional times to give a total of three washes prior to digestion. The gel pieces were dried to completion for 1 h in a Speed-Vac. Digestion was performed by incubating each gel piece in 30 µL trypsin solution for 16 h while shaking at 37°C. The trypsin solution was prepared by diluting sequencing grade modified trypsin in 25 mM ammonium bicarbonate to a final concentration of 33 µg/mL. Peptides were extracted by one change of 5% (v/v) trifluoroacetic acid and three changes of

5% (v/v) trifluoroacetic acid in 50% (v/v) acetonitrile. Each of the four extraction steps was performed for 30 min at room temperature with shaking. The pooled peptide solutions for each protein were concentrated using a Speed-Vac.

3.2.8.2 MALDI-TOF analysis

The mass spectra of the tryptic digests were acquired with a Voyager-DE™ Pro MALDI-TOF mass spectrometer equipped with a nitrogen laser (λ = 337 nm). The instrument was operated in delayed extraction reflector mode. The samples were reconstituted in 5 µL 0.1% (v/v) trifluoroacetic acid. In some cases, the samples were desalted with ZipTip® C_{18} (Millipore) according to the instruction manual before the MALDI-TOF MS analysis. Peptides were co-crystallized 1:2 (v/v) with matrices consisting of saturated α-cyano-4-hydroxycinnamic acid prepared in 60% (v/v) acetonitrile / 36% (v/v) methanol / 4% (v/v) water. The spectra were either internally calibrated using known trypsin autocatalytic fragment peaks or externally calibrated using a standard peptide mixture. Monoisotopic peptide masses were assigned and searched against the NCBInr database using MS-Fit [288, 289] and against the Swiss-Prot/TrEMBL databases using PeptIdent [290, 291] in order to identify the proteins. The searchin parameters were set up as follows: cysteine as carbamidomethyl-cysteine; maximum allowed peptide mass error of 0.5 Dalton; consideration of one incomplete cleavage per peptide; minimum number of matched-peptides was 4; and the searching range was within the experimental pI value ± 1 pH unit and experimental Mw ± 20%.

3.2.9 Phenetic tree

A phenetic tree was constructed according to Marques et al. [292] based on the pairwise comparison of the qualitative (presence / absence of spot) protein profiles of the *A. thaliana* ecotypes. The pairwise comparisons were done by counting the number of spots presents in both ecotypes (N_{AB}) and specifically present in one (N_{A0}) or the other (N_{0B}) of the two considered ecotypes. The Jaccard or Dissimilarity index (Equation 4.3) was used to compute a dissimilarity matrix based on these data. The matrix was run with the Phylip 3.6 software package [293] using the Neighbor Joining algorithm to calculate an unrooted phenetic tree.

4 Result and Discussion

4.1 2DE method development and optimization

The first task of this study was to develop and optimize a gel-based proteomics method for the qualitative and quantitative comparison of transgenic and non-transgenic *Arabidopsis thaliana* seed proteomes. It was aimed at developing a robust method that covers as many proteins as possible and provides clear, reproducible 2DE patterns by minimizing the degradation or alteration of proteins. An appropriate sample preparation and 2-D electrophoresis conditions are absolutely essential to achieve this goal. Therefore, every important step of the method was optimized for the proteome analysis of *Arabidopsis* seed.

4.1.1 Grinding

Tissue disruption is an important initial step in every 2DE protocol so that proteins are accessible to the extraction buffer. There are several grinding methods available for the very small (~500 µm long [294]) seeds of *Arabidopsis.* Accurate and reproducible protein extraction requires sample grinding of a fine, uniform consistency. Using a mortar and pestle is the most commonly used grinding method for *Arabidopsis* seeds [142, 143, 295]. Although highly efficient, this method is not appropriate for large sample numbers, and the difficulties involved in cleaning the tools increase the risk of cross-contamination. Because of its high efficiency in grinding seeds, the mortar-and-pestle method was used as a standard for other grinding techniques to compare to.

Grinding attempts with disposable pellet pestles in centrifuge tubes according to [296] failed because the applicable force was not sufficient to break the hard seed coat of *Arabidopsis* seeds. Therefore, other grinding techniques were investigated that were better suited to handle the large number of samples for this study.

In a preliminary experiment, seeds were ground with a Mega Grinder (described under Material) under two different conditions (Table 10). The two Mega Grinder methods differed in terms of bead type, bead number, time of grinding, seed weight, buffer volume, and tube volume. The most efficient grinding method, measured by the highest level of total extracted protein, was the Mega Grinder method ID 3 (1 metal bead, 30 sec., 74 mg seed, 0.7 mL buffer, 1 mL tube), which extracted 14.6%. In comparison, the Mega Grinder method ID 2 (3

glass beads, 60 sec., 152 mg seeds, 1.2 mL buffer, 2 mL) extracted 7.9%, and the mortar-and-pestle method extracted 10.6% protein.

Table 10 Comparison of grinding efficiency based on total extracted protein quantity

ID	Method	Seed[1)] (mg)	Tube (mL)	Buffer[2)] (mL)	Yield: Protein (% fw)
1	Mortar and pestle	150	n/a	1.2	10.6
2	Mega Grinder, 3 glass beads, 60 sec.	152	2	1.2	7.9
3	Mega Grinder, 1 metal bead, 30 sec.	74	1	0.7	14.6

[1)] *Arabidopsis thaliana* Columbia seeds, harvested April 2001
[2)] High CHAPS extraction buffer (see 4.1.2.2, page 63)

Due to the large differences in grinding efficiency observed between these two Mega Grinder conditions, it seems reasonable to further optimize the grinding procedure. A factorial designed experiment was chosen to investigate the influence of bead material, tube volume, seed weight, and extraction time on grinding efficiency (Table 11).

Table 11 Fractional factorial design to determine seed grinding conditions; pooled *A. thaliana* seeds (harvest April 2001), 0.7 mL High CHAPS extraction buffer, 45 seconds at 1500 rpm, metal bead (4 mm), ceramic beads (2.4 to 3.8 mm)

ID	Beads	Tube	Seed[1)]	Extraction time	Yield: Protein (% fw)
1	1 Metal	1 mL	30 mg	60 min	13.7
2	1 Metal	2 mL	30 mg	30 min	12.9
3	3 Ceramic	1 mL	30 mg	30 min	15.5
4	3 Ceramic	2 mL	30 mg	60 min	13.9
5	1 Metal	1 mL	70 mg	30 min	11.8
6	1 Metal	2 mL	70 mg	60 min	13.0
7	3 Ceramic	1 mL	70 mg	60 min	11.9
8	3 Ceramic	2 mL	70 mg	30 min	12.6

[1)] *Arabidopsis thaliana* Columbia seeds, harvested April 2001

The data from Table 11 were statistically analyzed using Minitab 13 and JMP 5. Since none of the two-factor interactions were significant ($P > 0.15$), only the main effects were included in the model (Figure 4).

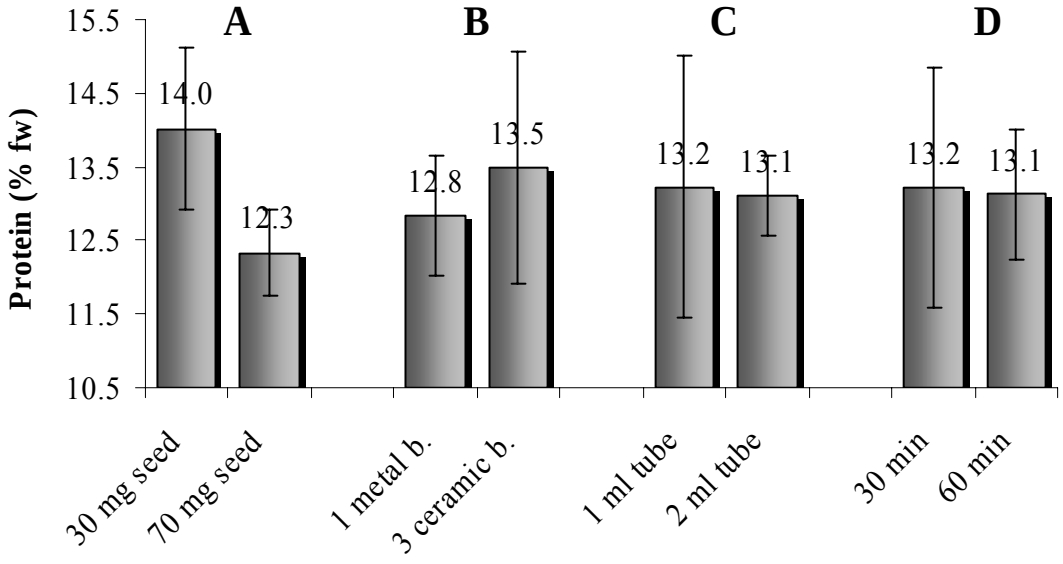


Figure 4 **Bar chart of the results of the fractional factorial design to determine seed grinding conditions. Grinding efficiency is based on the total extractable protein amount, expressed as% protein fw. The protein concentration was measured according to the Bio-Rad Protein assay using BSA as standard.**

A statistically significant ($P< 0.01$) difference in extraction efficiency was demonstrated for the seed-buffer ratio (Figure 4 A). The smaller seed-buffer ratio (30 mg to 0.7 mL) proved to be more efficient with 14.0% total extracted protein versus the larger seed-buffer ratio (70 mg to 0.7 mL) with 12.3% protein of fw. This may be due to an extraction saturation effect with increasing seed amount, as the extraction buffer volume stays constant. No statistically significant differences in grinding efficiency were found between the other parameter pairs (Table 11 and Figure 4). Due to the presence of intact seeds in the extract suspensions of the parameter combinations from Table 11, improvements to the grinding methods were completed further (Figure 5).

Methods 1, 2, 4, 6, 7, and 8 are statistically significantly ($P< 0.05$) different from methods 3 and 5. No intact seeds were detected when methods 1, 6, 7, and 8 were applied. Therefore, one can conclude that larger tubes were a key contributor to the increase in extraction efficiency.

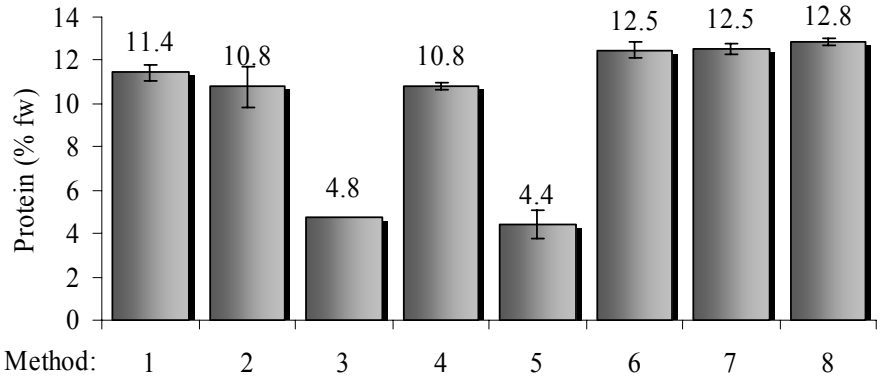


Figure 5 Protein extraction efficiency based on grinding method; pooled *A. thaliana* seeds (harvest February 2002), 0.7 mL low detergent extraction buffer, 45 seconds at 1500 rpm, 60 minutes extraction time, *N* = 2.

1 Mortar and pestle
2 Mega grinder, 1 mL centrifuge tube, 1 metal bead (4 mm)
3 Mega grinder, 1 mL centrifuge tube, 2 metal beads (3 mm), *N* = 1
4 Mega grinder, 1 mL centrifuge tube, 2 ceramic beads (2.4 – 3.8 mm)
5 Mega grinder, 1 mL centrifuge tube, BioMatrix E
6 Mega grinder, 2 mL centrifuge tube, 7 ceramic beads (2.4 – 3.8 mm)
7 Mega grinder, 2 mL centrifuge tube, 2 metal beads (3 mm)
8 Mega grinder, 2 mL centrifuge tube, 1 metal bead (4.75 mm)

Because the grinding efficiency of the mortar-and-pestle method and of the Mega Grinder methods 6, 7, and 8 were not significantly different ($P > 0.05$), the Mega Grinder was chosen as grinding method because it allows the grinding of multiple samples at the same time and excludes cross-contamination. Method 8 has additional handling advantages compared to the methods 6 and 7, as only one bead has to be added to the grinding tubes, and the single bead leaves enough space over the pellet to accurately remove the supernatant. Method 8 is also highly reproducible with a standard deviation of 0.29 ($N = 4$). Hence, the Mega Grinder method 8 with 1 metal bead in 2 mL centrifuge tubes was used as the grinding method for all the studies.

4.1.2 Protein extraction

Protein extraction is one of the most important steps, since it directly affects the outcome of 2DE. For example, proteins that are not solubilized during the extraction procedure will not appear on the 2DE gel. The pattern on the 2DE gel must reflect the protein composition of the sample without any losses or modifications. On the other hand, the extraction protocol should be as simple as possible to allow for high reproducibility and high comparative power. The goal was to develop a protein extraction protocol able to solubilize as many proteins as

possible as well as obtain 2DE gels of high resolution and quality while keeping sample manipulation to a minimum.

4.1.2.1 Defatting of seeds prior protein extraction

Lipids can interact with proteins, particularly membrane proteins. This can affect solubility, pI, and molecular weight of the proteins. Furthermore, lipids can also reduce the effectiveness of detergents to solubilize proteins by forming complexes with the detergents [201].

Because the *Arabidopsis* seeds used in this study contained on average 35% oil (fw basis), it was determined whether a defatting step prior to protein extraction had a beneficial effect on the 2DE gel resolution. Sample powder was defatted by four hexane extractions at 50 °C. Protein profiles of protein extracts from defatted and untreated ground seeds were visually compared in terms of resolution quality and spot pattern (Figure 6).

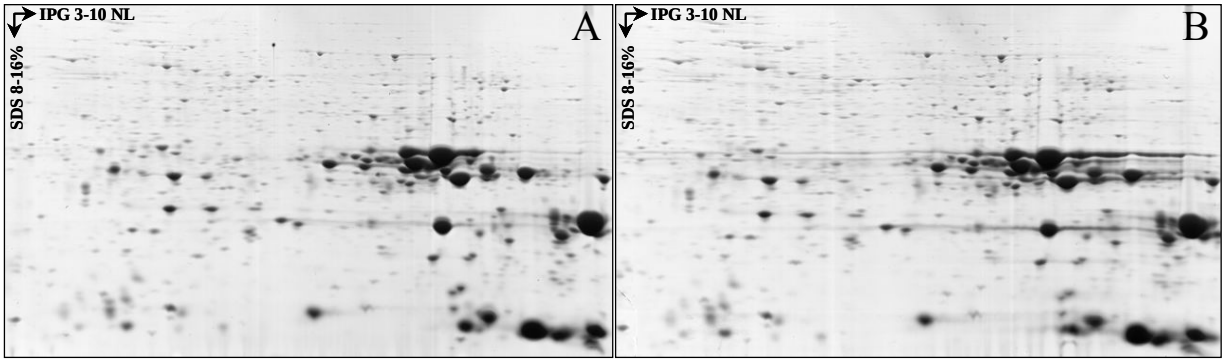


Figure 6 Influence of a defatting step prior to protein extraction on the 2DE pattern (Section pH 4-9, Mw 6-120 kD)
A: 2DE pattern of extract from original seeds
B: 2DE pattern of extract from defatted seeds

Two-dimensional gel electrophoresis patterns from the defatted seeds were visually similar to the untreated seeds. Neither treatment showed an obvious increase of resolution quality nor was an increase in spot number visually observed. Therefore, a defatting step was not included in the extraction protocol.

4.1.2.2 Extraction temperature

Seeds contain a lot of proteases [297], and the denaturing property of the extraction buffer is not always sufficient to inhibit the action of proteases [149, 201, 205]. Protease inhibitors added to the extraction buffer are supposed to prevent proteins from degradation by proteases. Seed proteins were extracted at RT and at 4 °C in order to examine the efficiency of the protease inhibitor cocktail used for the extraction buffer. The 2DE patterns from these extracts were then compared (Figure 7).

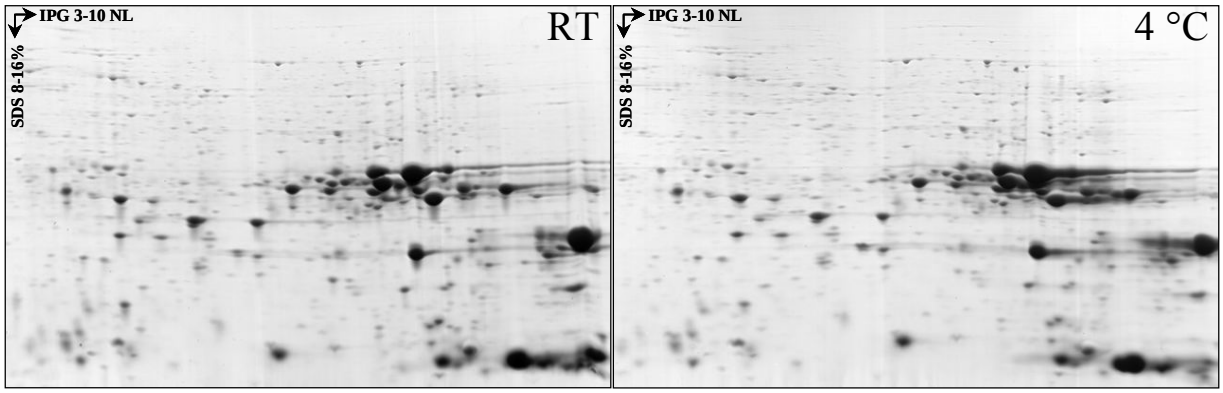


Figure 7 Influence of the extraction temperature on the extracted proteins
Protein extraction was conducted at room temperature (RT) and under chilled condition (4 °C). The showed 2DE gel sections encompass pH 4-9 and Mw 6-120 kD.

Special attention was paid to the loss of high molecular weight proteins and increase of low molecular weight proteins in the 2DE pattern of the sample extracted at room temperature compared to the 2DE pattern of the sample extracted at 4 °C. No qualitative differences could be visually detected between the 2DE patterns. The combination of high denaturing conditions, high DTT and protease inhibitor cocktail seems to be sufficient to prevent protein degradation due to protease activities. An extraction at 4 °C is not necessary.

4.1.2.3 Extraction buffer

Total proteins have been solubilized from mature *Arabidopsis* seeds with various extraction buffers either directly [142, 143, 296, 298] or after TCA/acetone precipitation [295, 299, 300]. Although TCA/acetone precipitation has been found useful for the inactivation of proteases and removal of interfering compounds [200], in particular when leaf and seedling tissues were used, the method bears the risk of protein loss due to incomplete precipitation or resolubilization [200, 215]. Direct solubilization in an IEF-compatible buffer avoids this risk,

reduces error, and is, therefore, preferred. Gallardo et al. [142, 143, 301] and Gruis et al. [298] obtained high quality 2DE gels from mature *Arabidopsis* seeds by direct solubilization of protein in an IEF-compatible extraction buffer.

In order to optimize the extraction protocol, three different extraction buffers were tested (see Table 12). In addition, a non-denaturing extraction method was included to compare the extraction of soluble versus total extractable proteins.

Table 12 Compositions of the three extraction buffers tested

	Extraction buffers		
Composition	Low detergent[a] (EB 1)	High CHAPS[b] (EB 2)	CHAPS/SB3-10[c] (EB 3)
Urea (M)	7	7	5
Thiourea (M)	2	2	2
CHAPS (% w/v)	0.75	4	2
SB 3-10 (% w/v)	-[d]	-[d]	2
Triton X-100 (% v/v)	0.75	0.2	-[d]
Ampholytes (% v/v)	1	1	0.5
Dithiothreitol (mM)	100	14	100
Protease inhibitor cocktail	Yes	Yes	Yes
Isopropanol (% v/v)	20	-[d]	-[d]
Tris-HCl (mM)	-[d]	18	-[d]
Trizma base (mM)	-[d]	14	40
DNase (U/mL)	-[d]	50	-[d]
RNase A (Kunitz U/mL)	-[d]	5	-[d]

[a]Monsanto Company; [b]Gallardo et al. [142]; [c]Molloy et al. [208]; [d]not contained

Low detergent extraction buffer (EB 1)

Rabilloud et al. [206] reported enhanced extraction efficiency of membrane-associated proteins by adding thiourea to a standard extraction buffer (modified O'Farrell's lysis buffer: 9 M urea, 2-4% (w/v) CHAPS, 1% (w/v) DTT, 2% (v/v) carrier ampholytes [200]). Giavalisco et al. [215] recently confirmed the beneficial effect of thiourea on protein solubilization resulting in increases of total protein yield and the number of detected proteins (mid-pI range), and more sharply focused spots [215]. It has long been known that the zwitterionic detergent CHAPS is far superior to Triton X-100 when urea alone is used as chaotrope [297, 302]. But when urea is used in combination with thiourea, Triton X-100

becomes more efficient than CHAPS for the solubilization of membrane proteins [303]. The detergent CHAPS was, therefore, partly substituted with Triton X-100. Isopropanol was added to the extraction buffer because of its beneficial effect of precipitating nucleic acid [247] and to control water flow at the basic end of the pH gradient during IEF [304, 305].

High CHAPS extraction buffer (EB 2)

The extraction buffer EB2 was used by Gallardo et al. to investigate processes during germination [142, 143, 301] and development [306] of *Arabidopsis* seeds.

CHAPS/SB3-10 extraction buffer (EB 3)

This buffer was first introduced by Rabilloud et al. [206] and then used by Molloy et al. [208 as 'enhanced solubilization' buffer in the last step of their 3-step differential extraction protocol. Méchin et al. [307] substituted DTT partly with TCEP and left out Trizma base to obtain their R2D2 buffer (5 M Urea, 2 M Thiourea, 2% CHAPS, 2% SB3-10, 20 mM DTT, 5 mM TCEP, 0.75% carrier ampholytes). Méchin et al. reported that the R2D2 buffer was very efficient for a large range of different samples and resulted in 2DE gels of high resolution and quality. They also demonstrated that 2% total detergent (CHAPS) was more efficient than 4% total detergent (CHAPS) to extract maize endosperm proteins [307].

Non-denaturing extraction buffer: PBST (EB 4) and PBST+DTT (EB 4+DTT)

The PBS buffer is a standard buffer used to extract soluble proteins under non-denaturing condition. The addition of the mild nonionic detergent Tween at 0.1% increases the solubilization efficiency but retains enzyme activities [204].

Comparison of protein extraction efficiencies

Arabidopsis seeds were ground and extracted using different extraction buffers. The protein solubilization efficiencies were compared based on the extracted protein yields, the IEF patterns, and the 2DE patterns.

Comparison on the basis of protein yields. Figure 8 shows a bar diagram of the protein yields as a percentage of fresh weight for each extraction buffer based on two replicates. No statistically significant differences in extracted protein yield were detected between EB1, EB2, and EB3. EB4 and EB4+DTT are less efficient than the low detergent buffer EB1.

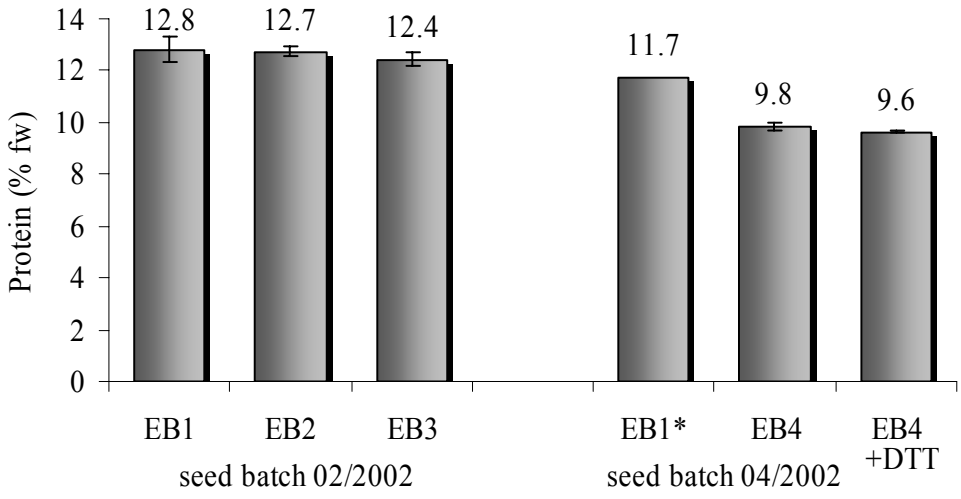


Figure 8 Comparison of extraction buffer (EB) efficiencies based on protein yields obtained from the two batches of *Arabidopsis* seeds.
* Standard deviation is not available because extracts were pooled prior protein assay.

Comparison on the basis of IEF. In a preliminary study (pH distribution of *Arabidopsis* seed proteins), the extraction efficiencies of EB 1 to 3 were compared in respect to pH protein groups. Although EB3 solubilized the same amount of proteins, the IEF-PAGE gel pattern of the EB3 extract revealed some weaker protein bands (see Figure 9, arrow at magnified gel area) compared to EB1 and EB2. It was concluded that EB3 is not as efficient as EB1 or EB2 in extracting as many proteins as possible. In addition, precipitation of an unknown with EB3 was frequently observed after buffer preparation. EB3 was discontinued from inclusion in the extraction buffer comparison.

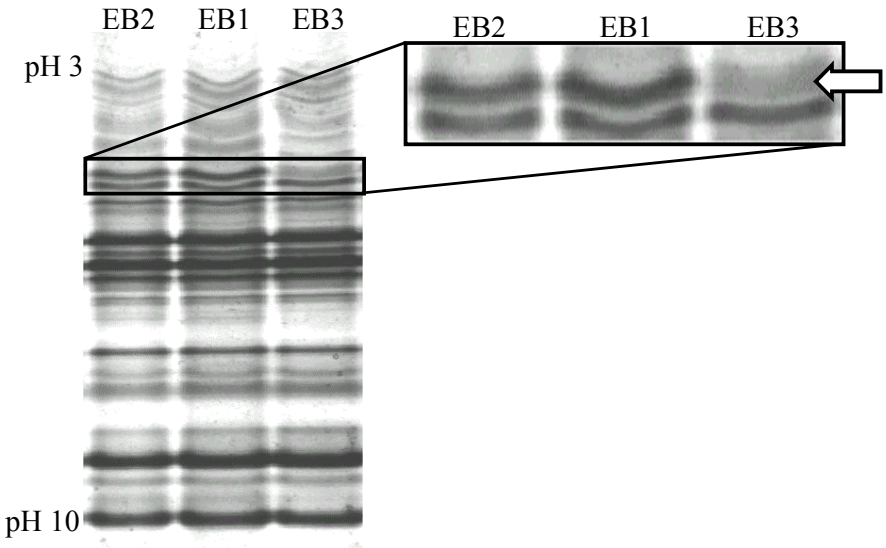


Figure 9 Comparison of extraction buffer efficiency based on their one dimensional isoelectrofocusing pattern pH 3-10

Comparison on the basis of 2DE patterns. The 2DE patterns from EB1- and EB2-extracts did not show obvious differences in the number of spots or resolution quality (Figure 10). The lower detergent concentration (1.5%) of EB1 did not result in a lower solubilization efficiency compared to the high CHAPS extraction buffer EB2 (4.2% total detergent). Méchin et al. [307] recently even showed that a lower detergent concentration (2% CHAPS) is more efficient in solubilizing proteins than a higher detergent concentration (4% CHAPS). 2DE patterns obtained with EB1 extracts were more consistent than with EB2 extracts (data not shown).

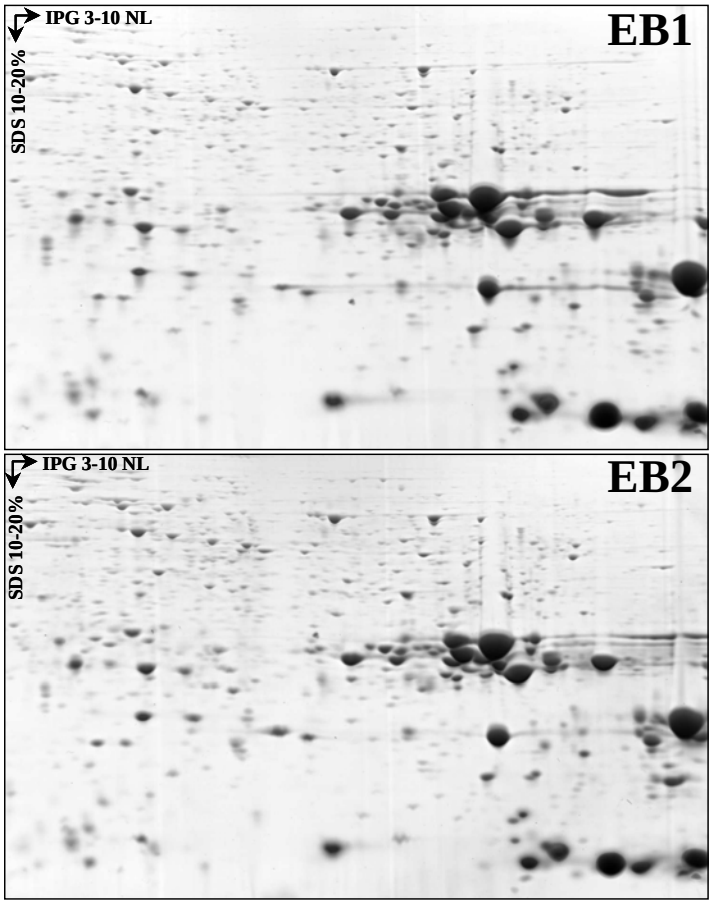


Figure 10 Comparison of extraction buffers based on the 2DE pattern (pH 4-9, Mw 6-120 kD)
EB1: Low detergent extraction buffer
EB2: High CHAPS extraction buffer

In order to get a more comprehensive understanding of the proteins extracted with the low detergent extraction buffer, the 2DE pattern of proteins extracted with the low detergent extraction buffer was also compared to 2DE patterns of proteins extracted with two non-denaturing buffers (EB4 and EB4+DTT). EB4+DTT was obtained by adding the reducing agent DTT to EB4. Non-denaturing buffers are used to solubilize cytosolic proteins (water-

soluble) proteins. As expected, the low detergent buffer (EB1) is with 11.7% extracted protein and approximately 690 resolved protein spots (Figure 8 and Figure 11) more efficient in the solubilization of proteins than the two non-denaturing buffers (EB4: 9.8% extracted protein and ~ 560 protein spots; EB4+DTT: 9.6% extracted protein and ~ 675 protein spots).

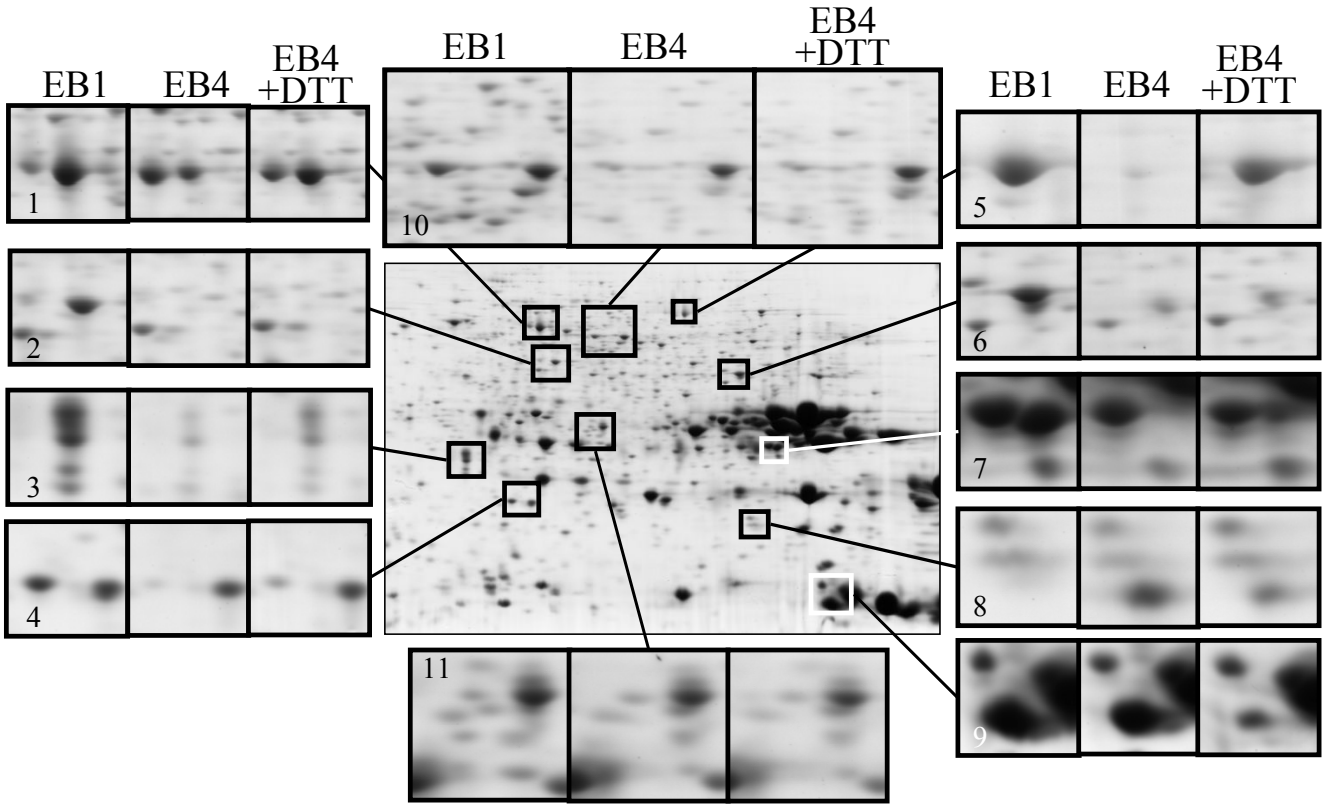


Figure 11 Comparison of extraction buffers based on 2DE pattern (pH 4-9, Mw 6-120 kD)
EB1: Low detergent extraction buffer (~ 690 protein spots)
EB4: PBST extraction buffer (~ 560 protein spots)
EB4+DTT: PBST extraction buffer + DTT (~ 675 protein spots)
The same rehydration buffer was used for all extracts to dilute and prepare the extracts for the IEF step.

However, the number of proteins extracted with the low detergent buffer (EB1: ~ 690 spots) is only marginally larger than the number of proteins extracted with the reducing non-denaturing buffer (EB4+DTT: ~ 675 spots). This suggests that mainly hydrophilic proteins are displayed on the 2DE gels. Giavalisco et al. [215] describe a similar situation and concluded that the urea/thiourea mixture increases the yield of water-soluble proteins. This can be seen in Figure 11-1, -3, -4, -6, -7, -9, and –10. However, some proteins were clearly only solubilized in EB1 (Figure 11-2, -10, -11) and a better resolution quality of the 2DE pattern was obtained when EB1 extracts were used. This is due to higher concentrations of salt in the PBST buffers, which interfere with the IEF step of the first dimension. Figure 11-8

shows a protein spot that is not present when proteins are solubilized in the low detergent buffer. It is possible that in spite of the added protease inhibitor cocktail and the chilled temperature during the extraction with the non-denaturing PBST buffer, the activity of proteases were not completely inhibited and that the protein is a degradation product (artifact).

4.1.3 Isoelectric focusing (IEF)

During IEF, proteins are separated according to their isoelectric point. Immobilized pH gradient gels bound on a plastic layer are used as a matrix. These so-called IPG strips are commercially available in sizes from 7 to 24 cm and with various pH ranges (e.g., broad range: pH 3-10 or narrow range: pH 5.3-6.5). In order to standardize and to improve the reproducibility [138, 229] of the 2DE method, commercially available gels were used for the first and second dimension. When this study was initiated, only one large-format gel system, the 18.3 x 19.3 cm SDS-PAGE PROTEAN II Ready Gels (Bio-Rad), was commercially available for the second dimension. In spite of many attempts and adjustments, it was not possible to achieve satisfying results with those large-format gels. The gel vendor indicated that other users observed similar findings and suspected the change to a new gel manufacturer and/or transportation problems as a possible reason. After a new gel shipment, these problems persisted, so it was decided that the commercially available medium-size format gel system (13.3 x 8.7 cm), the Criterion 2DE system (Bio-Rad), would be used for the second dimension. Criterion SDS-PAGE gels are used with 11 cm IPG strips.

4.1.3.1 Choice of the pH gradient

IPG gel strips with many different pH ranges are commercially available. The aim of the study was to cover a broad range of seed proteins and to achieve an equal spatial distribution of proteins in the gel in order to gain high-resolution 2DE and to avoid spot overlapping, which may hamper spot analysis. As illustrated in Figure 9, pIs of *Arabidopsis* seed proteins span the entire investigated range from pH 3 to 10. IPG gel strips with a wide pH interval of 3-10 are available as linear and non-linear pH gradients. The non-linear pH gradient increases resolution between pH 5-7 [248]. This may be advantageous since it was estimated that more than 40% of proteins in living systems have a pI between 4 and 6 [308].

To determine the appropriate pH range for these studies, the protein distribution obtained from linear and non-linear pH 3-10 IPG strips were compared (Figure 12). The 2DE pattern obtained with the linear IPG gel strip shows large areas (pH 3 to 5 and pH 7 to 10) without many protein spots, while in other areas (pH 5 to 7), the protein spots are compressed (Figure 12 A). Better separation and spatial distribution of protein spots was achieved by using the non-linear IPG gel strip (Figure 12 B). Hence, the non-linear pH 3-10 IPG strips were used for all studies.

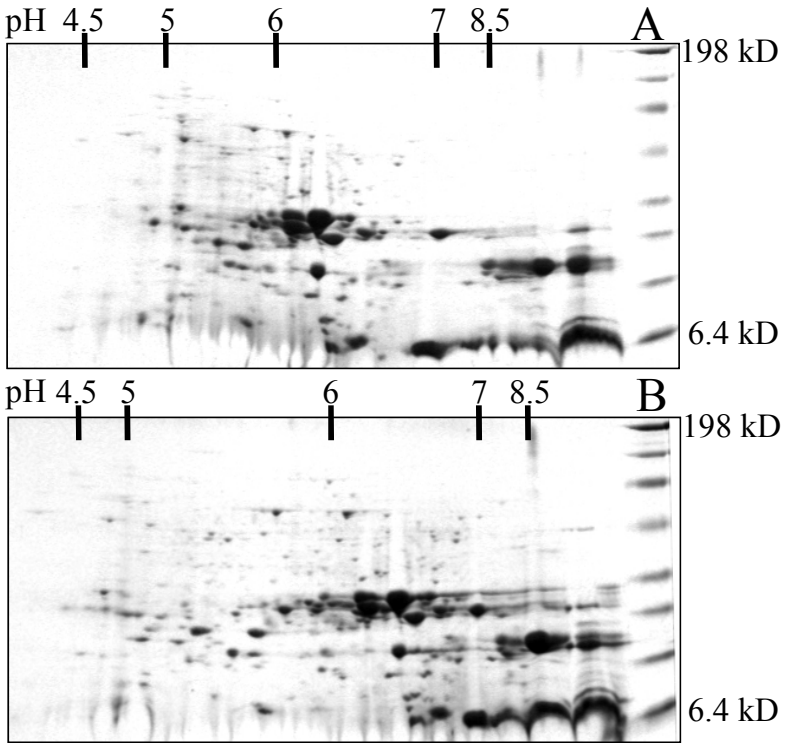


Figure 12 Comparison of pH gradient strips (non-linear vs. linear) for the first dimension
A: Bio Rad, immobilized linear pH gradient (3-10) strip
B: Amersham Biosciences, immobilized non-linear pH gradient (3-10) strip
total protein load: 200 µg; 10-20% Tris-HCl linear gradient SDS-PAGE gel

At the time of the method development, non-linear pH 3-10 IPG strips were not available as 11 cm strips. Therefore, 13 cm strips had to be cut on both ends to fit in the focusing trays of the equipment used. The manufacturer of the equipment used in this study recently introduced 11 cm non-linear pH 3-10 IPG strips and adjustable focusing trays on the market. Both the 11 cm non-linear pH 3-10 IPG strips and the adjustable focusing trays were employed, and the resulting 2DE patterns were compared to the 2DE pattern obtained with the developed method. The comparison showed that cutting the 13 cm IPG strip does not result in the loss of proteins, neither at the acidic nor at the basic end of the strip (data not shown).

4.1.3.2 Sample application

Protein samples were diluted in rehydration buffer and this sample-rehydration solution was used to rehydrate (reswell) the dry IPG strips. Sample proteins enter the gel during the rehydration; this in-gel rehydration has the advantage of reducing horizontal streaking, elimination of precipitation at the application point, and higher sample loading [233, 234]. Serious problems regarding reproducibility due to unequally rehydrated IPG strips were overcome after building rehydration trays optimized for the size of the IPG strips. The extraction buffer (EB1) was used as the basis for the rehydration buffer. Water movement during the IEF, caused by negatively charged groups fixed to the gel matrix, may lead to blurred zones. In order to control the water flow during IEF, isopropanol was partly substituted with water-saturated isobutanol. Leimgruber et al. [247] have reported that the combination of water-saturated isobutanol (12.5%) with isopropanol (10%) resulted in better control of water movement at the basic end of the IPG gel strip than by using isopropanol alone. The addition of 5% glycerol, though frequently suggested by different authors [247, 305], did not result in an improvement of the 2DE pattern in these studies. The ampholyte concentration was reduced from 1% in EB1 to 0.3% in the rehydration buffer in order to increase the voltage during IEF if necessary.

4.1.3.3 Isoelectrofocusing program

Isoelectrofocusing is an endpoint method as proteins reach the pH value of their pI in the pH gradient gel (equilibrium position) and thus stop migrating. However, focusing for too long can lead to horizontal streaking, distorted protein patterns, and protein loss due to active water transport [200]. If the focusing time is too short, proteins will not be completely focused and appear as horizontal streaks on the gel. Determing the optimum focusing time is crucial for 2DE pattern quality and reproducibility [200]. Factors such as the pH range and length of IPG strip as well as sample and rehydration buffer characteristics affect the amount of time required for complete focusing. Focusing time is expressed as an integral of voltage (volt-hours; Vh) and has to be determined empirically for optimal results. Amersham Biosciences running protocol [248] was used as a basis for the IEF program optimization. Various focusing times, maximal voltages, and additional focusing steps were tested in order to optimize the protocol. The combination of high voltage (7000 V) and prolonged focusing time (Vh) resulted in considerable water movement toward the anode, which led to distorted protein pattern at the acidic gel end (Figure 13).

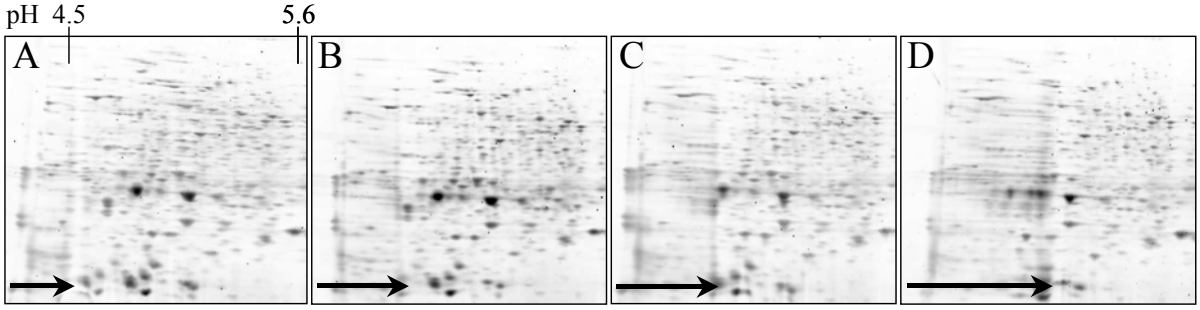


Figure 13 Comparison of different focusing times, maximum voltage: 7000 V
A: 25,000 Vh; B: 35,000 Vh; C: 40,000 Vh; D: 50,000 Vh
The enlarged 2DE gel regions encompass pH 4-5.6 and Mw 6-120 kD.
Arrow marks water movement.

Focusing for 25 kVh resulted in the least distortion of the 2DE pattern but also in incompletely focused protein spots. The best resolution was achieved by focusing for 35 kVh.

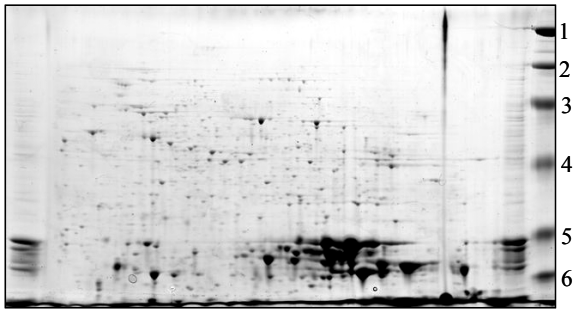
8000 V as suggested by the manufacturer could not be reached during any run. No difference in focusing quality was observed when either 4000 or 7000 V was used as the maximum voltage. Because of practical reasons (possibility to focus overnight), 4000 V was finally chosen as the maximal voltage.

4.1.4 Second dimension (SDS-PAGE)

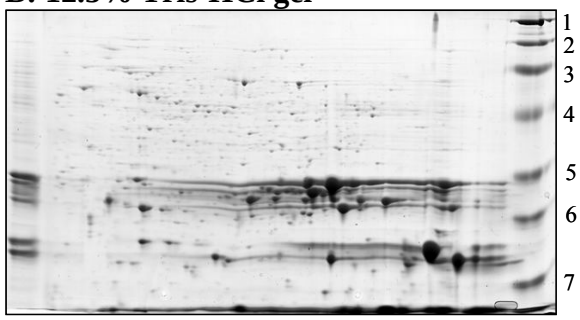
4.1.4.1 Choice of the SDS-PAGE gel gradient

Many different homogeneous (single-percentage acrylamide) and gradient precast gels are commercially available for the Criterion 2DE system. The percentage of acrylamide determines the pore size of a gel. A higher amount of acrylamide leads to smaller pore sizes and higher sieve effects. Single-percentage gels give excellent resolution of sample proteins that fall within a narrow molecular weight (Mw) range. Gradient gels allow proteins with a wide range of molecular weights to be analyzed simultaneously, and the decreasing pore size along the gradient has a focusing effect on the protein spots. Single-percentage gels are commonly used for the second dimension of 2DE as they are easy to cast without the requirement of sophisticated equipment. In order to select the optimal gel type for the second dimension, the *Arabidopsis* seed 2DE patterns obtained with two homogeneous and three gradient precast gels were compared (Figure 14).

A: 10% Tris-HCl gel

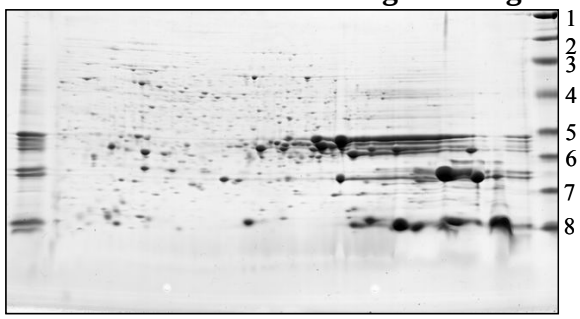


B: 12.5% Tris-HCl gel

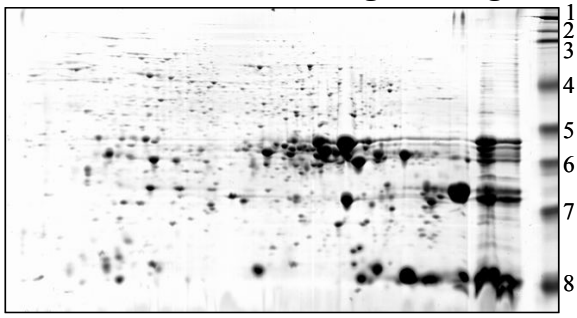


C: 10.5-14% Tris-HCl linear gradient gel

D: 10-20% Tris-HCl linear gradient gel



E: 8-16% Tris-HCl linear gradient gel



Molecular weight markers:

1:	198.0 kD
2:	115.0 kD
3:	93.0 kD
4:	49.8 kD
5:	35.8 kD
6:	29.2 kD
7:	21.3 kD
8:	6.4 kD

Figure 14 Comparison of different commercially available precast gels to determine the optimum acrylamide concentration for the second dimension.

Few protein spots appear above 115 kD as IPGs are limited in their ability to resolve high molecular weight proteins [225, 255, 309]. This may be due to size exclusion when the proteins are loaded onto the gel. The best resolution of proteins with molecular weights between 35.8 and 115 kD was achieved with the 10% Tris-HCl single-percentage gel. Proteins with Mw below ca 28 kD ran off the gel and were lost with the 10% Tris-HCl gel. Losses of low molecular weight proteins were also observed with the 12.5% and the 10.5-14% Tris-HCl gels. The gradient gels 10-20% and 8-16% displayed a larger range of extracted proteins, with the 8-16% gel showing a better spot distribution. The gradient gel 8-16% was used for all further studies.

4.1.5 Staining

After the proteins are separated by 2DE, they have to be visualized for the comparative proteome analysis. Coomassie Brilliant Blue (CBB) and silver staining are commonly used in proteomics studies. The recently introduced fluorescent stains SYPRO Orange and SYPRO Ruby increasingly gain popularity because of their high sensitivity and broad linear dynamic range [203, 250, 310]. The objective was to identify a staining method that stains a broad spectrum of proteins; allows the qualitative and quantitative comparison of seed protein profiles and protein identification by mass spectrometry or N-terminal sequencing; and is sensitive, reproducible, and easy to use. In order to determine the best staining method, a comparison was made between the 2DE gels of *Arabidopsis* seed proteins stained with two commercially available MS-compatible silver stains, a colloidal CBB stain, and two commercial fluorescent stains, SYPRO Orange and SYPRO Ruby.

4.1.5.1 Silver staining

Among the non-radioactive protein detection methods silver staining offers the greatest sensitivity [245] and allows the detection of protein quantities as little as 0.5 ng [249]. Although staining with silver has inherent disadvantages (see chapter 2.6.1.4), it is still considered the standard method for detecting minor proteins [249]. Two commercially available silver staining kits, ProteoSilver™ Plus Silver Stain Kit (Sigma-Aldrich, St. Louis, MO) and Silver Quest™ Silver Staining Kit (Invitrogen Corporation, Carlsbad, CA), were chosen because of their compatibility with mass spectrometry and their high protein sensitivity of 0.3 ng of bovine serum albumin (BSA) [manufactures instructions].

The staining procedures require multiple and carefully timed steps, including fixing, sensitization, staining, developing, stopping, and multiple washing steps in between. Although both staining methods utilize silver nitrate and have a very similar staining protocol, very different results were obtained (see Figure 15 A and D). Strong background staining and negatively stained spots were obtained with both staining methods (Figure 15 and Figure 16) but the appearance of negative-stained spots was much more pronounced with the SilverQuest staining kit (Figure 15 D). Figure 16 A and D show magnified areas of 2DE gels of Figure 15 as examples of negatively stained protein spots. Gels stained with ProteoSilver™ Plus also showed a stronger contrast between protein spots and gel background – spots appear to be darker. The shortcoming of silver staining to stain certain proteins is known and has been reported by Wirth and Romano [245]. Spot-saturation as a

reason for negative-staining is unlikely because the protein spots were also negatively stained when the protein loading was decreased from 200 µg (Figure 15 A and Figure 16 A-1) to 100 µg (Figure 15 C and Figure 16 C-1) total protein load.

Significant background staining is a serious challenge for analytical spot detection software programs, especially when the 2DE pattern is to be evaluated quantitatively. Therefore, the aim is to obtain gels with as little background staining as possible, or at least have a uniform background staining that can be subtracted by mathematic algorithms used by analytic software programs. Strong background staining and streaking can be caused by non-alkylated dithiothreitol (DTT) from the extraction, rehydration, and/or equilibration buffer as well as DNA and RNA in the sample [249].

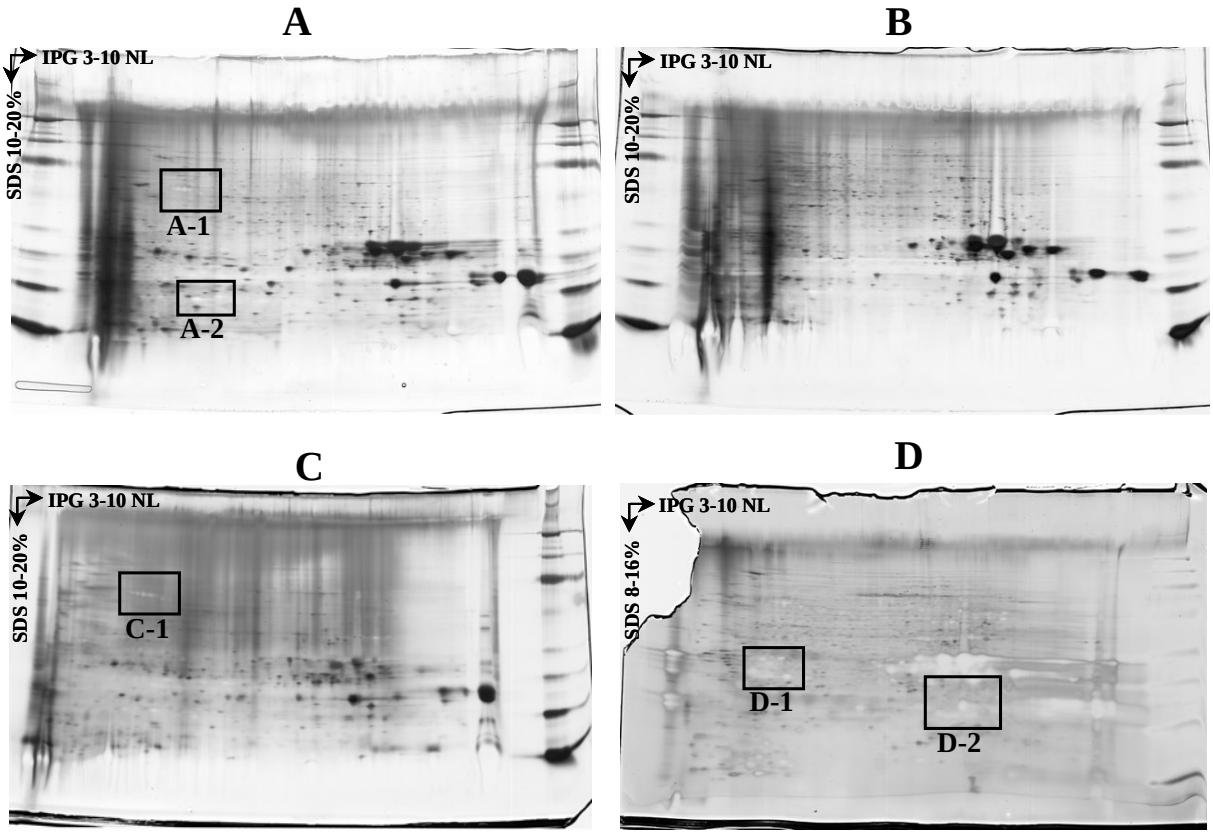


Figure 15 Comparison of 2DE gels of *Arabidopsis* seed proteins stained with two mass spectrometry-compatible silver staining kits:
A: ProteoSilver™ Plus Silver Stain Kit (Sigma-Aldrich)
200 µg total *Arabidopsis* seed protein, DTT as reducing agent
B: same as A but sample treated with Dnase and Rnase
C: same as A but 100 µg total protein and TBP as reducing agent
D: SilverQuest™ Silver Staining Kit (Invitrogen™)
100 µg total protein, DTT as reducing agent for extraction buffer, TBP as reducing agent for rehydration and equilibration buffer
Boxes correspond to gel regions enlarged in Figure 16.

Three different approaches were investigated to soften the background staining: (a) the protein sample was treated with DNase I and RNase A during protein extraction; (b) the reducing agent DTT was replaced in all steps (extraction, rehydration, and equilibration) by tributyl phosphine (TBP); and (c) the protein load was reduced from 200 to 100 µg total protein. As seen in Figure 15, neither the DNase/RNase treatment (Figure 15 B) nor a switch to TBP and reduction of the protein loading (Figure 15 C) decreased the background staining. Similar results using TBP and lower protein loading were obtained when the 2DE gels were stained with SilverQuest™ (Figure 15 D).

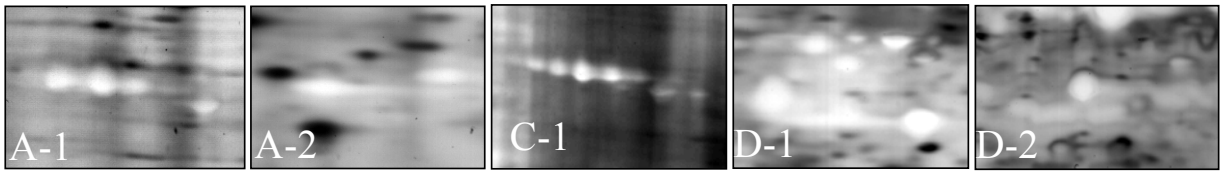


Figure 16 Examples for negative staining of proteins with silver stain. Enlarged gel regions 1 and 2 of 2DE gels of Figure 15 A and D.

Further attempts to use silver as a staining method were dismissed due to its inability to stain many proteins and the unspecific background staining. Other considerations were the narrow linear range of silver stain and its susceptibility to staining variability due to the inherent complexity of silver staining procedures [250].

4.1.5.2 Coomassie Brilliant Blue staining (CBB)

Coomassie Brilliant Blue is the most commonly used gel staining method [311]. It is characterized by its broad dynamic range, good quantitative linearity, compatiblty with MS, and comprehensive protein staining [263, 272]. The colloidal form of Coomassie Brilliant Blue G-250 does not utilize a destaining step [257]. Dye molecules are extracted from the colloids by the proteins whilst the colloidal particles do not penetrate the gel. The result is a stained gel with a clear background. By adding methanol and increasing the content of ammonium sulfate of the staining solution, Neuhoff et al. [258, 259] were able to shorten the staining time and to achieve the complete staining of protein bands throughout the entire gel layer.

The staining method according to Neuhoff et al. [258] was used to stain a 2DE gel of an *Arabidopsis* seed protein extract. Figure 17 shows the scanned colloidal CBB stained 2DE gel. The background was homogenous and almost clear with ODs below 0.02.

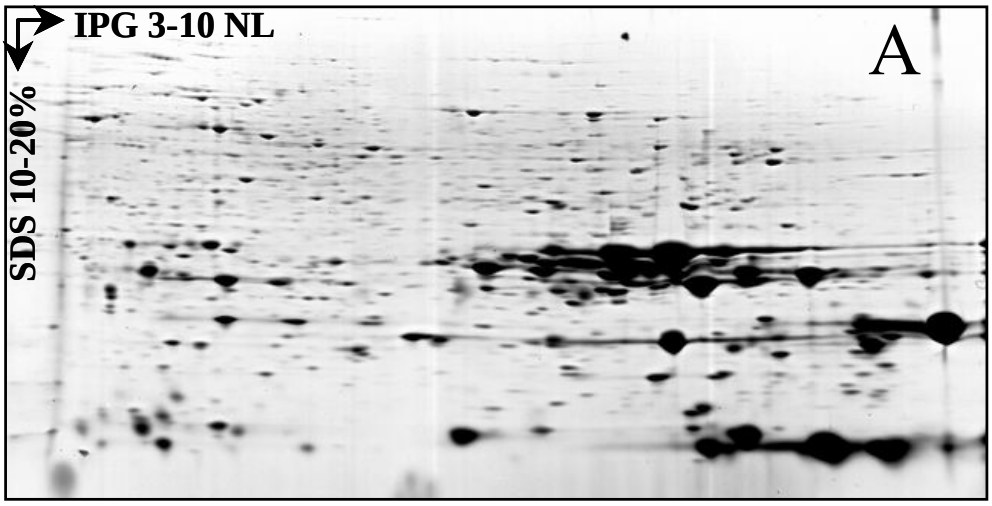


Figure 17 2DE gel of *A. thaliana* seed proteins stained with Colloidal Coomassie Brilliant Blue G-250.
(150 µg total protein load; 42,000 Vh; 7000 V maximum voltage; staining for 1d; ~ 676 spots detected by PDQuest; detection parameter are the same as for SYPRO Orange and SYPRO Ruby)

PDQuest detected around 676 protein spots in the colloidal CBB stained gel. Gallardo et al. [142] and Kamo et al. [295] used two different silver staining methods (not-MS compatible) to analyze the protein pattern of mature *Arabidopsis* seeds and were able to detect 1,272 and 984 protein spots, respectively, in large format 2DE gels (>180 x >200 mm). Taking into account the smaller dimension of the Criterion gels (110 x 80 mm), the detection of 676 proteins suggests that colloidal CBB is a fairly sensitive staining method.

4.1.5.3 Fluorescent staining

Fluorescent stains promise to provide silver-stain-comparable sensitivity in detecting proteins combined with a greater linear dynamic range than achieved with CBB [250, 284, 310]. They also allow the detection of glycoproteins, lipoproteins, low Mw proteins, and metalloproteins that are not stained well by other stains [312].

SYPRO Orange and SYPRO Ruby were used to stain *Arabidopsis* seed protein 2DE gels. SYPRO Orange staining was performed according to a protocol optimized for 2DE gels by Malone et al. [284], as the detection sensitivity of SYPRO Orange in 2DE gels is known to be less than in 1D SDS-polyacrylamide gels, due to interactions of the stain with nonionic detergents and carrier ampholytes present in the 2DE gel [252]. SYPRO Ruby staining was performed according to the manufacturers' instructions. Background-free staining was

achieved with both fluorescent stains (Figure 18). The computer program detected around 912 protein spots in the 2DE gel stained with SYPRO Ruby and around 490 spots in the SYPRO Orange-stained 2DE gel. The software program also detected spot saturation for some of the highly abundant proteins in the SYPRO Orange stained gel (Figure 18 A, arrows), but none of the proteins stained with SYPRO Ruby were saturated. It is apparent that SYPRO Ruby provides higher detection sensitivity and a broader dynamic range than SYPRO Orange.

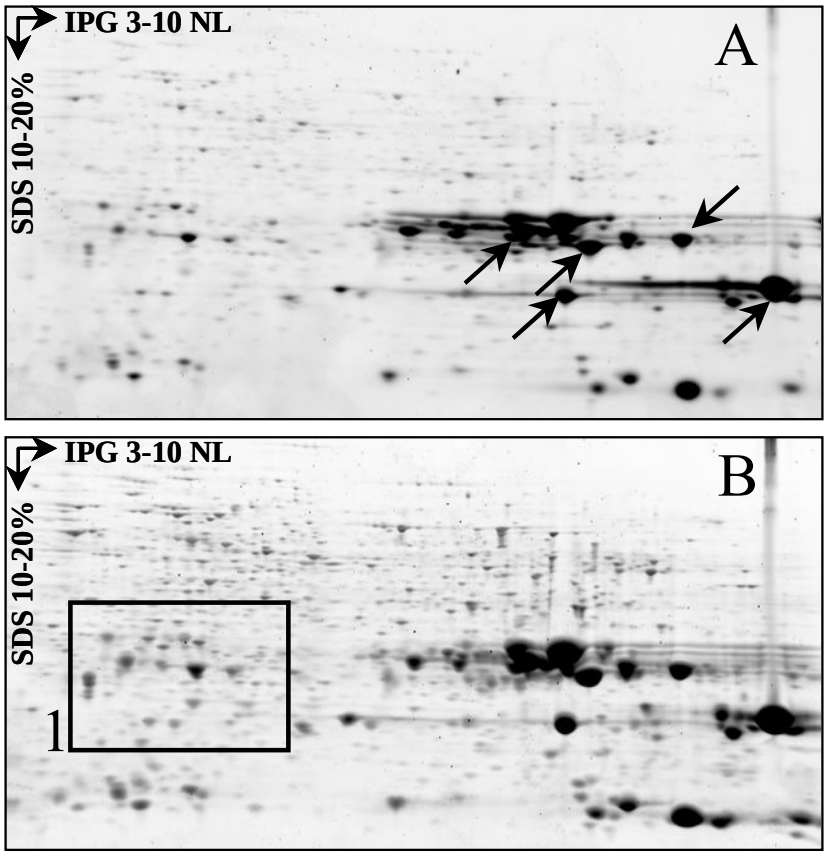


Figure 18 Comparison of two fluorescent staining methods.
(150 µg total protein load, 42,000 Vh, 7000 V)
A: SYPRO Orange, staining for 2 h, ~ 490 spots detected, arrows mark saturated spots
B: SYPRO Ruby, staining for 3 h, ~ 912 spots detected

Upon closer examination of the 2DE gels, randomly occurring flecks (speckles) were observed on the SYPRO Ruby-stained gels (Figure 19-1). According to Molecular Probes [313], the manufacturer of SYPRO dyes, these flecks are precipitated SYPRO Ruby dye. Yet a reason for the precipitation or information of what influences their generation are not given. The tendency of SYPRO Ruby to cause speckling on gels was also described by Mackintosh et al. [264]. After setting the minimum spot size of the analysis software high enough, most

of the speckles were ignored and only a few speckles had to be excluded manually from the data set.

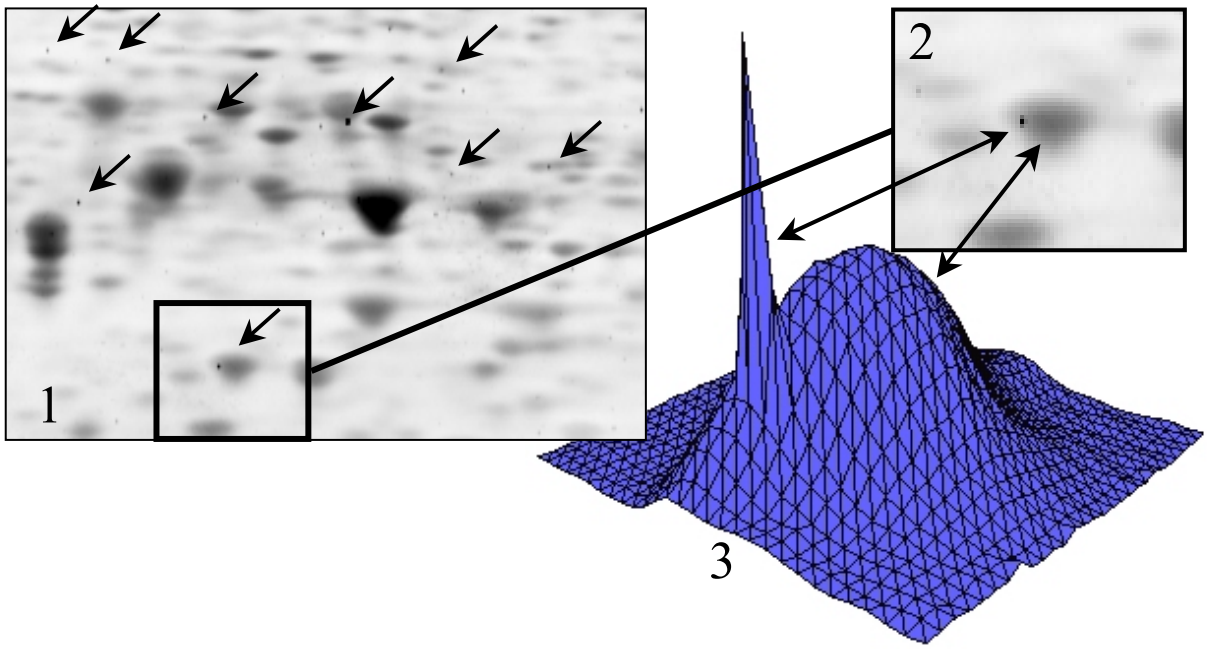


Figure 19 Effect of SYPRO Ruby speckles, which randomly occur in the gel, on protein spot quantitation.
1: Gel section from image Figure 18 B, arrows mark some of the speckles
2: Magnified area from gel section 1 with protein spot and SYPRO Ruby speckle
3: 3D rendering of a protein spot affected by a SYPRO Ruby speckle

The effect of SYPRO Ruby speckles, which randomly occur in the gel, on protein spot quantitation is demonstrated in Figure 19. Some of the speckles occur within the boundaries of a protein spot (Figure 19-2). In order to inspect the proper quantitation of these spots, they were viewed in a 3D mode (Figure 19-3). These flecks appear as large splitter on the shoulder of the 3D rendering for that protein and affect the protein spot quantitation.

Although SYPRO Ruby is the staining method of choice regarding sensitivity (~ 912 spots compared to CBB with ~ 676 detected spots) and the large dynamic range (at least 3 order of magnitude [249]), it was not employed for the comparison studies. The randomly occurring speckles hamper spot detection, and accurate spot quantitation cannot be guaranteed.

4.1.6 Summary

For the qualitative and quantitative analysis of the seed proteome of *Arabidopsis thaliana*, a proteomics method based on 2DE was developed. Every important step, from seed grinding to protein staining, was optimized.

The Mega Grinder method, using 1 metal bead in 2 mL centrifuge tubes and a seed-buffer ratio of 30 mg to 0.7 mL ratio, was chosen to be the grinding method because of its high grinding efficacy and reproducibility. The method also allows the grinding of multiple samples at the same time by excluding cross-contamination.

Different extraction conditions and buffers were compared. It was demonstrated that defatting of the seed prior to protein extraction or the extraction at 4 °C did not improve the quality of the 2DE protein pattern. The low detergent buffer EB1 with a total detergent concentration of 1.5% (0.75% CHAPS and 0.75% Triton X-100) provided the best combination of number of protein spots, quality of 2DE resolution, and consistency of 2DE pattern. The majority of the extracted proteins (98%) were hydrophilic proteins.

In order to standardize the 2DE procedure and to minimize the impact on the result from laboratory-specific variation in parameters, commercially available gels were used for the first and second dimension. The best separation and spatial distribution of *Arabidopsis* seed protein spots were achieved by using non-linear pH 3-10 IPG gel strips in the first dimension and 8-16% Tris-HCL linear gradient SDS-PAGE gels in the second dimension. Serious problems regarding reproducibility due to unequally rehydrated IPG strips were overcome after building customized rehydration trays. The best resolution was achieved by focusing for 35 kVh and a maximum voltage of 4000 V.

In the last optimization step, silver stain, colloidal CBB stain, and two fluorescent stains were compared in order to determine the best suitable staining method. SYPRO Ruby was the most sensitive staining method (~ 912 spots compared to CBB and SYPRO Orange with ~ 676 and ~ 490 detected spots, respectively). However, randomly occurring speckles of SYPRO Ruby dye makes spot detection more difficult, so accurate spot quantitation cannot be guaranteed. The colloidal CBB staining method was chosen to be the staining method because it is simple to use and is compatible with subsequent protein identification methods; also it provided background-free staining results and was more sensitive than SYPRO Orange. Silver staining was not considered to be the staining method because of strong background staining and negatively stained seed proteins.

4.2 Validation of the 2DE method

Any analytical method is subject to a certain degree of technical (methodological) variation. Technical variation of 2DE may be divided into qualitative (spot number and position) and quantitative (spot quantity) variation [239]. It arises from a combination of variations from each stage of the analysis. In order to design a meaningful experiment, it is essential to gain a comprehensive understanding of the scope of these variations inherent to the developed and optimized proteomics approach. Important methodological parameters like linearity and sensitivity will also be assessed to evaluate the power and quantitation limits of the developed method.

4.2.1 Repeatability

The goal of this study was to investigate the inherent variation in 2DE pattern data, i.e. the repeatability of the method. Repeatability refers to the precision of the method under conditions where data are obtained in the same laboratory by the same operator using the same equipment within short time intervals [314]. In many publications the term reproducibility is used instead of repeatability. Reproducibility is defined as the precision of a method under conditions where data are obtained in different laboratories with different operators using different equipment [314].

The repeatability of the method may be influenced by the extraction process and/or by the various steps of 2DE. The impact of the extraction procedure and 2DE on the repeatability of the spot pattern was investigated by comparing the qualitative and quantitative repeatability of the spot patterns among extracts and gels. This data will be used in Chapter 4.2.2 to define replicates (multiple gels from one extract or multiple extracts and one gel per extract) and to determine the number of replicates needed in the comparison studies.

The experimental design of the repeatability study is shown in Figure 20. Seeds harvested from six individual *Arabidopsis* plants (WT Col-0) were pooled to one seed pool. All *Arabidopsis* plants were grown under identical environmental conditions. To assess the extract-to-extract and the gel-to-gel variability, three protein extracts were prepared from this seed pool, and for each extract, 2DE was performed in triplicate.

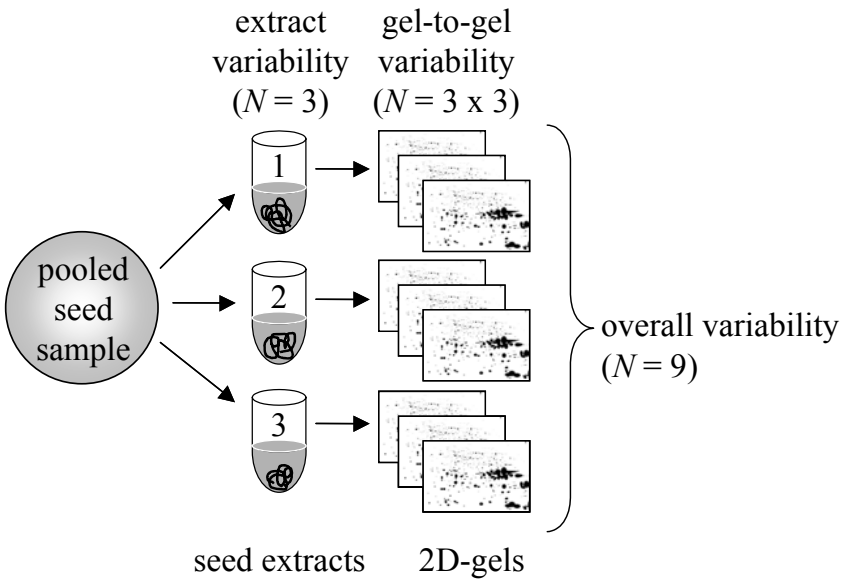


Figure 20 Experimental design for repeatability study.
Equal amounts of seeds obtained from six individual *Arabidopsis thaliana* plants were combined for the pooled seed sample.

The gels were digitized with a laser densitometer. The 2DE patterns of the nine gels were very similar (repeatable) as demonstrated in the displayed enlargements of two randomly picked regions of the *Arabidopsis* seed 2DE-pattern (Figure 21 and Figure 22).

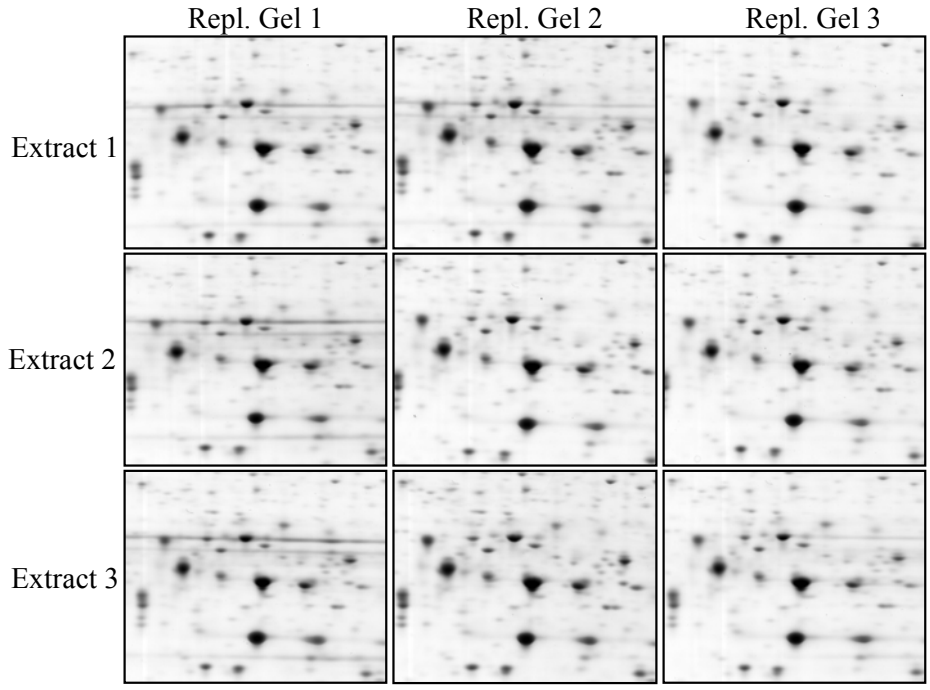


Figure 21 Enlargements of a region of the *A. thaliana* seed 2DE pattern of all nine 2DE gels of the repeatability study. The displayed portions encompass pH 4.7-5.5 and Mw 21-42 kD

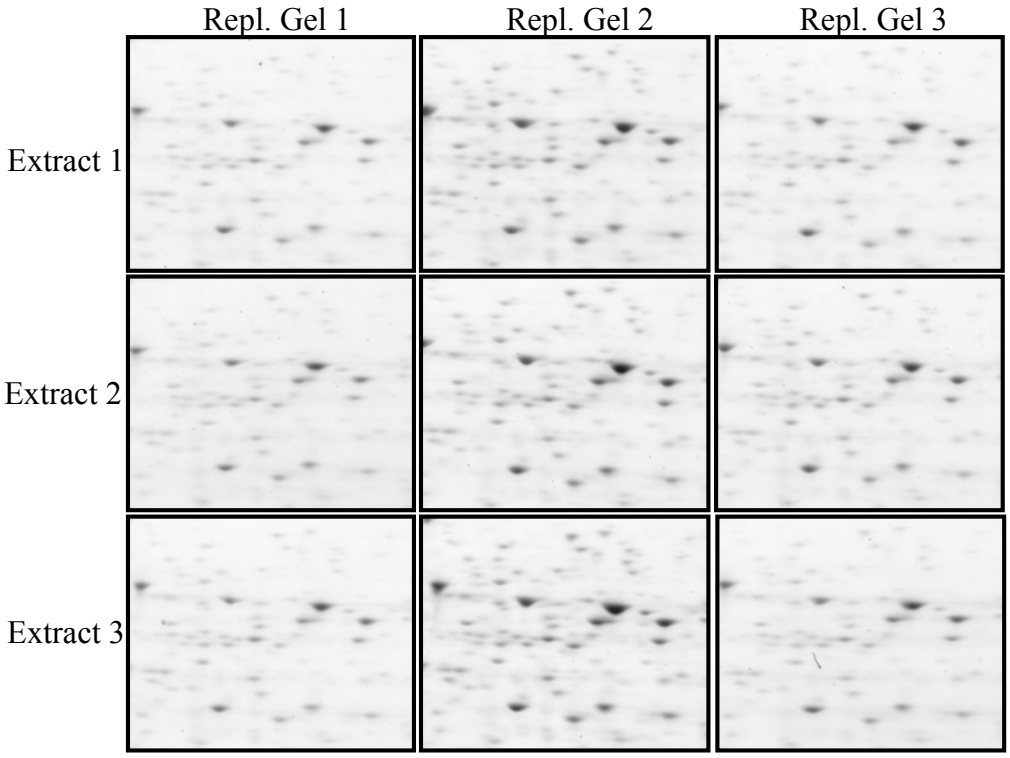


Figure 22 Enlargements of a region of the *A. thaliana* seed 2DE pattern of all nine 2DE gels of the repeatability study. The displayed portions encompass pH 5.2-5.6 and Mw 37-70 kD

The 2DE gel analysis software package PDQuest was used to process the digitized images. The statistical module of PDQuest was not used to analyze the data because it only calculates the population standard deviation (see equation 4.1) and not the sample standard deviation (see equation 4.2).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2} \tag{4.1}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{4.2}$$

Equation 4.1 is applicable if the whole population is measured. Since only small numbers ($N = 3$ and 9) of the set of all possible (infinite) measurements were used, the obtained standard deviation of the equation 4.1 may underestimate the population standard deviation, while equation 4.2 gives an unbiased estimate of the population standard deviation for a small number of measurements [285]. Therefore, the statistical software program JMP (SAS Institute Inc., NC, USA) was used to analyze the data.

Spots detected by the software program were manually verified. False positive spots (e.g., artifacts and multiple spots in a cluster) were manually removed; false negative spots (obviously missed spots with OD > LOD; LOD = $OD_{background}$ + 3 x $SD_{background}$ [285, 286]) were added to the images. A spot was considered to be reproducibly present/absent when it is present/absent in all three replicate gels of one extraction. Table 13 summaries the results from the spot detection using PDQuest software and manual verification.

Table 13 Results of spot detection and matching in the repeatability study

Extracts	Number of spots					Reproducible spots in	
	Gel 1	Gel 2	Gel 3	average	SD*	replicate gels	all gels
Extract 1	434	650	466	517	117	379	350
Extract 2	493	578	559	543	45	419	
Extract 3	483	706	482	557	129	403	
Extract-to-extract				539	20	400 ± 20	

* SD = standard deviation

An average of 539 distinct spots were discerned in each of the nine gels. The gel-to-gel standard deviations (45 to 129) were much larger than the extract-to-extract standard deviation of 20 (Table 13). Upon visual examination of the 2DE images, it became apparent that the replicate gel 2 for all extracts (1 to 3) had clearer and better resolved gels with less streaking and more faint spots above LOD than the other replicate gels (Figure 22). Thus, the numbers of detected spots in these gels were higher, and consequently, the SDs for these data sets were higher as well (Table 13). Although the same amount of total protein was loaded on each gel, it seemed that the spot quantity and the protein amount was higher for the replicate gel 2 (Figure 22). It is assumed that slight differences in the size of the rehydration lanes of the first version of customized rehydration trays might affect the rehydration efficiency. After using new customized rehydration trays, such obvious differences between replicate gels were no longer observed (data not shown).

In order to compare the position and the quantitative variation of spots, individual protein spots were matched between all replicate gels. The number of reproducible spots (present in all three replicate gels) for the extracts was on average 400 ± 20 spots. Spots with an average spot optical density (OD) below the limit of quantitation (LOQ = $OD_{background}$ + 10 x $SD_{background}$ [285]) were considered as faint spots and excluded from the quantitative comparison. No reproducible spot present in one extract was reproducibly missing (absent in all three replicates) in one of the other extracts.

Three hundred fifty reproducible spots were detected in all nine gels of the three extracts. These spots were used to calculate the degree of position variation in the 2DE method. The average gel-to-gel position standard deviation ($N = 9$) was found to be 1.9 ± 0.4 mm for the x-position (isoelectric point) and 0.9 ± 0.2 mm for the y-position (molecular weight). By taking into account the dimension of the mid-size 2DE gels (110 mm in x-position and 80 mm in y-position), the relative positional variation is 1.7 and 1.1% for the pI- and Mw-direction, respectively. This demonstrates the highly repeatable nature of the 2DE spot position and is consistent with published data (Table 14).

Table 14 Published research on position repeatability of 2DE.

Reference	*N*	x-Position (mm)	y-Position (mm)	Gel dimension (mm)
Present study data	9	1.9 ± 0.4	0.9 ± 0.2	110 x 90
Corbett et al. [229]	4	1.19 ± 0.41	2.88 ± 0.55	180 x 190
Blomberg et al. [226]	best 3/5	1.87 ± 1.06	1.20 ± 0.76	180 x 185
Norbeck et al. [315]	5	1.04 ± 0.28	0.81 ± 0.27	160 x 200
Li et al. [316]	3	0.87 ± 0.13	1.03 ± 0.21	180 x n/a*
Zhan et al. [239]	3	2.39 ± 0.81	0.81 ± 0.40	180 x 205
Zhan et al. [317]	4	1.95 ± 0.45	1.70 ± 0.53	180 x 205

n/a = not available

Spot variation may influence accurate matching of protein spots between gels and may hamper the detection of differences in protein patterns, such as protein shifts caused by phosphorylation, glycosylation. Variation in spot position did not occur independently from surrounding spots but rather in conjunction with the variation in neighboring spots. Figure 23 shows a 2DE image with the vector-offsets, which compares the corresponding spot positions between two 2DE gels (e.g. replicate gels 1 and 2 of extract 2). The vector-offsets demonstrate the distance (length of line) and the direction (angle) between the spot positions on the compared gels. They are good indicators to demonstrate the degree of gel distortions.

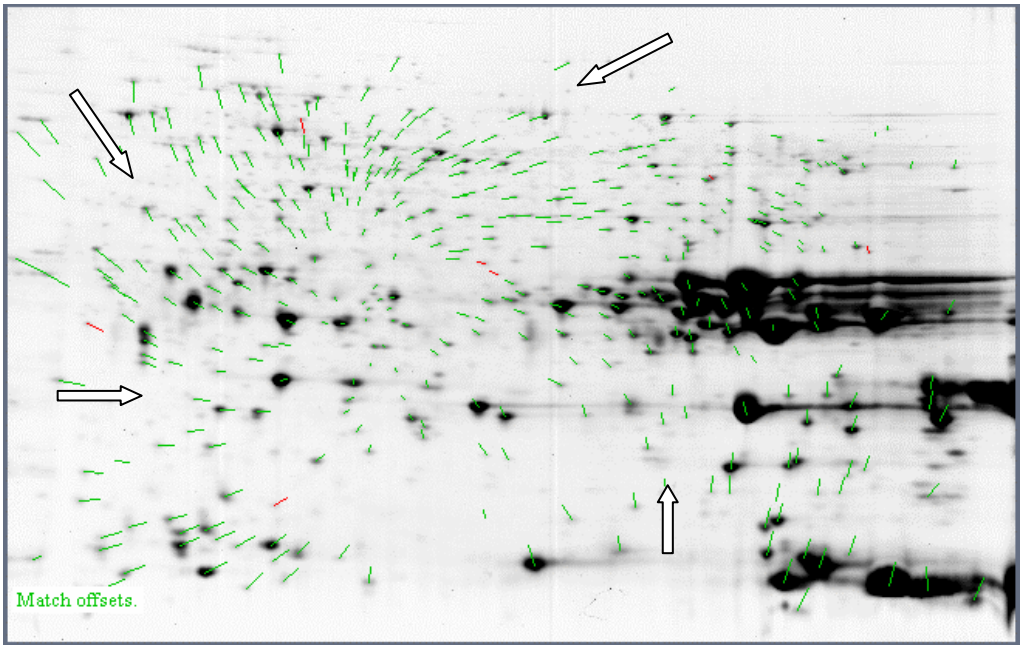


Figure 23 Effects of gel distortion: Offset vectors show gel-to-gel alignment of replicate gels 1 and 2 of extract 2. Arrows highlight distortion pattern.

Because the spot position variation was primarily caused by distortion of gel regions (Figure 23, arrows), an accurate matching of spots is possible by taking into account the spot positions of neighboring spots.

As mentioned above, although each gel had the same amount of protein loaded, a clear variation in total optical density and total spot quantity was observed between gels. Table 15 shows the total optical density and total spot quantity for each gel. The total spot quantity is the sum of all spot quantities detected in a gel, while the total optical density also includes the optical density of the background and streaks. Differences in total spot quantity between gels were particularly obvious between replicate gels 1 and 2 of extract 1 with 12,923 and 18,252 total spot quantity, respectively (Table 15).

Table 15 Total optical density and spot quantity of each gel

Extract	Gel	Total optical density (OD)	Total spot quantity (OD x spot area)
1	1	20,391	12,923
	2	27,985	18,252
	3	21,654	13,910
2	1	23,210	14,787
	2	25,850	16,271
	3	23,611	15,666
3	1	22,116	14,489
	2	27,655	16,775
	3	24,463	14,959

This variation can be caused by a number of factors, including sample loss during rehydration, focusing, equilibration, migration from IPG to PAGE, and inconsistent staining. To more accurately compare spot quantities between gels, method-related variations in spot quantity have to be compensated by normalization [318]. To accomplish this, the quantity of each gel spot is divided by the total quantity of all the gel spots. In order to obtain units of parts per million (ppm), the normalized quantity is multiplied by 10^6 as a scaling factor. Further analyses were performed with the normalized spot quantities.

The degree of analytical variation inherent to the 2DE process was assessed using only spots matched to all nine gels and with average spot quantities above the LOQ. A total of 254 spots (73% of spots present in all nine gels) met these requirements. The mean coefficient of variation (CV = SD/mean x 100) is a quantitative index for variation of quantities among matched spots and was computed for gel-to-gel variation and overall variation (Table 16). For the gel-to-gel variation, the mean CV was calculated after determining the average spot CV for each extract (3 replicates x 3 extracts), while the overall mean CV was calculated after averaging the individual spot CV of all nine gels.

Table 16 Results of the quantitative repeatability study

Type of variation	N	Spots	Mean CV (%)	Median CV (%)
Gel-to-gel	3 x 3	254	24.8 ± 18.5	19.6
Overall	9	254	26.2 ± 15.2	22.1

The mean CV of the matched spot quantities was found to be 24.8 ± 18.5% for the gel-to-gel and 26.2 ± 15.2% for the overall variability (Table 16). Similar quantitative variations were reported in the literature by Mahon et al. [254] (mean CV of 32%, median CV of 18%), Norbeck et al. [315] (mean CV of 17%), Blomberg et al. [226] (mean CVs of 20-28%), Molloy et al. [319] (18.7-26.4%), and Zhan et al. [239] (mean CV of 35.7% ± 20.8% (n=3). The small difference of only 1.4%-points between the gel-to-gel CV and the overall CV indicates that the extract-to-extract variation has only a minor impact on the quantitative repeatability of the 2DE method and the major contribution to analytical variation results from the 2DE procedure itself. The median CVs of spot quantities were much lower than their corresponding mean CVs with 19.6% and 22.1% for gel-to-gel and overall variability, respectively. This indicates that the mean CV value is affected by the presence of some spots

with very poor reproducibility. In order to assess the distribution of the CVs in greater detail, the spot CVs (254 spots, $N = 9$) are displayed as a histogram chart (Figure 24).

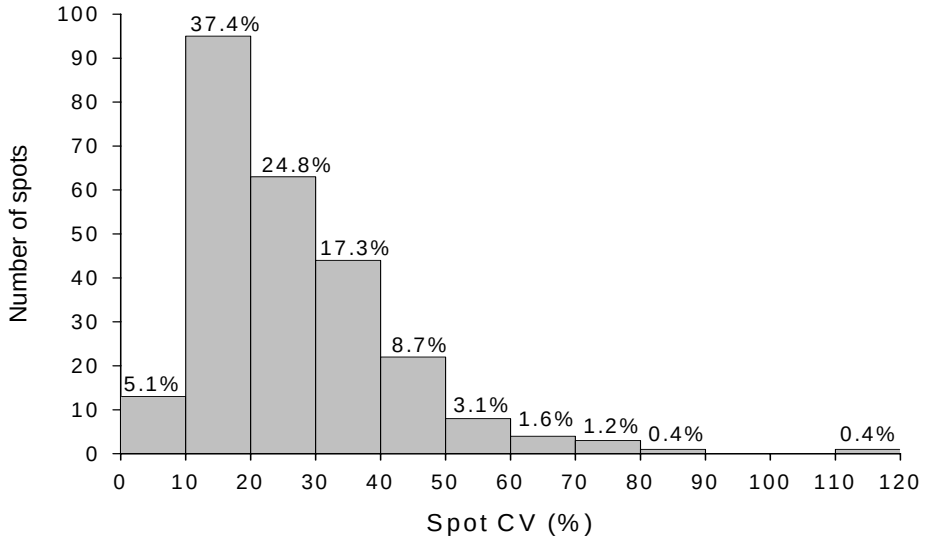


Figure 24 Distribution of the overall spot quantity coefficients of variation for the 254 spots detected in all nine gels and with mean spot quantities above LOQ.

The CV values ranged from 5.6% to up to 120%. Over two-thirds of all spots were found to have CVs below 30% and over 93% of the spots had CVs below 50%. Only 17 of the 254 spots analyzed exhibited a CV greater than 50% and one spot displayed a CV of 120%. All seventeen spots with CVs over 50% are not well-defined spots and were affected by streaking and/or neighboring spots (overlapping of spots) and may be inaccurately quantified.

In order to evaluate any relation between variation in spot quantity and spot position, the average spot quantity CV from all spots above the LOQ were related to their spot position in the gel (Figure 25). It is apparent that the degree of quantitative variation was evenly distributed in the dimension of isoelectric points (Figure 25A).

Unlike Norbeck and Blomberg [315], who did not see a dependence of spot quantity CV and protein position, a significant ($P < 0.001$) correlation between spot quantity CV and molecular weight was observed (Figure 25B). Higher molecular weight proteins showed larger quantitative variation. Also, high molecular weight proteins were more susceptible to horizontal streaking, making an accurate quantification of those spots more difficult. It is

known that the resolution of high molecular weight proteins with IPG-2DE is difficult because of size exclusion effects of the IPG gels [225, 255, 309]. However, Blomberg et al. [226] described the same tendency for higher Mw proteins but concluded that the concern for unpredictable transfer of large proteins from the first to the second dimension when IPG strips are used is unfounded.

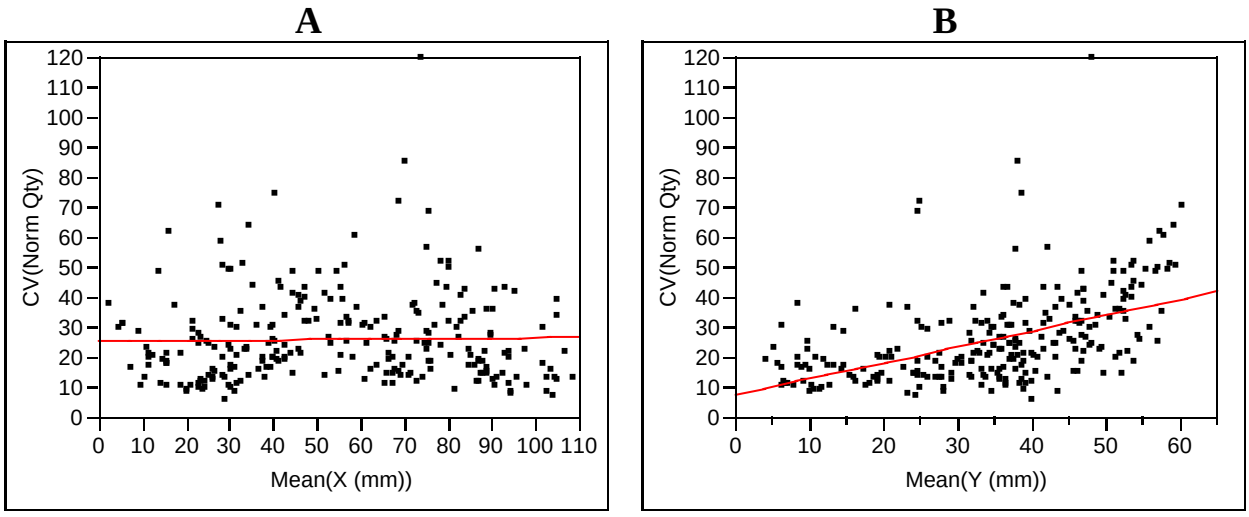


Figure 25 Correlation between spot quantity CV and spot position in the 2DE gel: (A) in x-position: acidic pH (10 mm = pI 4.5) to basic pH (100 mm = pI 7.8); (B) in y-position: low Mw (5 mm = 6 kD) to high Mw (60 mm = 86 kD). Calculations of CV and x and y-position values represent averages from all nine gels (all three extracts). Each graph indicates the best-fit line (or the least square regression line).

4.2.2 Estimation of replicate size

Determination of the optimal replicate size is crucial for experimental design. The use of too many replicates is a waste of resources, whereas too few replicates may not allow for sufficient statistical rigor. The optimal number of replicates depends on the precision of the 2DE method (standard deviation), the desired degree of reliability (significance level), confidence (power), and the desired detectable minimum difference (difference in spot quantity).

Table 17 shows the calculated replicate sizes needed to detect 2- and 3-fold differences in spot quantities between two test groups depending on the desired power. The power is defined as the probability of being able to detect the difference. Thus, a power of 90% says that with the given sample size, the difference in the means will be detected 90% of the time. The replicate size calculation is based on the overall averaged variance and the overall

averaged mean from the repeatability study (chapter 4.2.1), and it was computed with the statistic software program JMP.

Table 17 Replicate sizes based on the overall averaged variance and overall averaged mean from the repeatability study for different confident levels

Difference in spot quantity	2-fold difference			3-fold difference		
Power (%)	90	95	99	90	95	99
Replicate size	8	9	12	3	4	4

Test significance level (α) is 0.05

According to Table 17, three replicates have to be analyzed to detect a 3-fold quantitative difference in protein expression with a confidence of 90%. Almost three times more replicates would need to be analyzed in order to detect a 2-fold difference with the same confidence. These replicate sizes are based on the overall averaged variance of all 254 spots of the data set of the repeatability study and provide a good estimate for the number of replicates needed for the comparison studies. It does not mean that the detection of a 2-fold difference for an individual spot is not detectable with three replicates. The ability to detect a certain quantitative difference in spot quantity depends primarily on the variance (standard deviation) of this specific spot in the two samples. Therefore, the variance (here expressed as CV) for a protein spot needed to demonstrate a 2- or 3-fold difference can be computed for a given replicate size, desired power, and significance level (Table 18).

Table 18 Coefficient of variance (CV) threshold to demonstrate 2- or 3-fold difference in spot quantity

Difference in spot quantity	2-fold difference	3-fold difference
Test significance level	0.05	
Power (%)	90	
Replicate size	3	
CV (%)	28	55

According to this calculation (Table 18), with three replicates, a 2-fold change in protein expression will be detected with a 90% probability for all spots with CVs less than or equal to 28%, and a 3-fold change for all spots with CVs less than or equal to 55%. That means that 182 spots (72%) and 243 (96%) spots of Extract 1 of the repeatability study (4.2.1) could be tested for a 2-fold and a 3-fold change in spot quantity, respectively.

Spots with high quantity CVs are frequently irregularly shaped (Figure 26), saturated, or of poor quality. In these cases, the computer program will be unable to accurately determine the quantity of the spot. Therefore, high spot quantity CVs are also an indicator of inaccurate quantitation. Figure 26 provides an example for a spot with a high CV. Images Rep. 1 to 3 are magnified areas from three replicate gels of the same extract. The arrows indicate the spot of interest. The white line marks the contour of the spot and was used to manually define the spot for accurate quantitation.

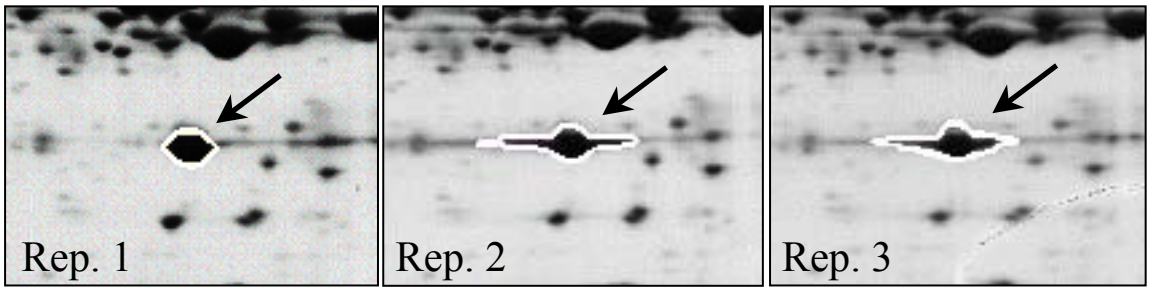


Figure 26 Example for irregularly shaped and manually defined spot.

The spot (marked with arrows in Figure 26) is well shaped in replicate gel 1 but irregularly shaped in the replicate gels 2 and 3. Table 19 contains the results from the quantitation of this spot in the three replicate gels and the CV for that spot.

Table 19 Quantitation of irregularly shaped spot - comparison of automatic and manual detection of spot boundaries

Gel	Spot quantitation with spot boundary defined automatically	manually
Rep.1 (ppm IOD)	51,000	59,000
Rep.2 (ppm IOD)	30,000	77,000
Rep.3 (ppm IOD)	24,000	60,000
Mean (ppm IOD)	34,000	63,000
CV (%)	41	11

IOD = integrated optical density (OD x spot area)

The spot quantity of the well-shaped spot in replicate gel 1 is 51,000 ppm IOD; this is more than double the spot quantity of the same irregularly shaped spot in replicate gel 3 with 24,000 ppm IOD. The software program PDQuest provides a boundary tool to manually define spot contours. After applying this tool to the spot in the three replicate gels (see white contour lines in Figure 26), the spot quantity CV was decreased from 41% to 11% (Table 19). For the well-shaped spot in replicate gel 1, the difference between the quantity of the

automatically defined spot and the quantity of the manually defined spot is less (<16%). More significant are the differences in spot quantity for the irregularly shaped spot in replicate gels 2 and 3 (157 and 150%; see Table 19). The quantities of the irregularly shaped spot of replicate gel 2 and 3 (77,000 and 60,000 ppm IOD) are more similar to the quantity of the well-shaped spot (59,000 ppm IOD) after manual tracing of the spot edges. Therefore, it appears that the manual definition of spot contour results in more accurate quantitation for irregularly shaped spots. In addition, the decrease of the CV from 41% to 11% would now allow conclusions to be made about a 2-fold difference to another sample with sufficient power (precision).

4.2.3 Sensitivity and linearity of response

Sensitivity and range of linear response are important method parameters. Sensitivity is the threshold amount of protein that gives a response (spot) clearly different from the background. Within the linear range the spot quantity will have a linear relationship to the protein amount. A linear response is important for a reliable quantitative comparison. The published sensitivities (detection limits) of colloidal Coomassie Brilliant Blue G-250 (CBB G-250) staining vary tremendously from less than 1 ng [258] to up to 100 ng [311]. Similar differences were found regarding published linear ranges of colloidal CBB. Berggren et al. [252] reported a linear dynamic range of 8-fold (30-250 ng) for colloidal CBB stain on 1D SDS-PAGE gels with various standard proteins, including bovine serum albumin (BSA; 66 kD). Mahon et al. [254] determined the linear range by using 2DE and demonstrate a linear relation between protein amount and spot quantity over a 20-fold range (e.g. BSA: 400 ng to 8 μg). The differences in reported sensitivity and linear ranges may be explained by differences in gel size, gel thickness, and gel type, as well as differences in staining protocols or duration. It is obvious that sensitivity and linear range are method specific parameters, and therefore, have to be determined individually for every developed method.

In order to estimate the absolute sensitivity of the developed 2DE method, the limit of detection was determined for two standard proteins. Bovine serum albumin (BSA) and β-lactoglobulin (β-LG) were chosen as external protein standards because they have different molecular weights, are well characterized, and are absent in *Arabidopsis* seeds. The linear relationship between amount of protein and the intensity of staining was examined for the two standard proteins BSA and β-LG, as well as for a subset of 20 seed protein spots. The

subset of 20 *Arabidopsis* proteins represents classes of proteins differing in pI, Mw, and relative abundance. Figure 27 shows the locations in the 2DE gel of the standard and the subset proteins.

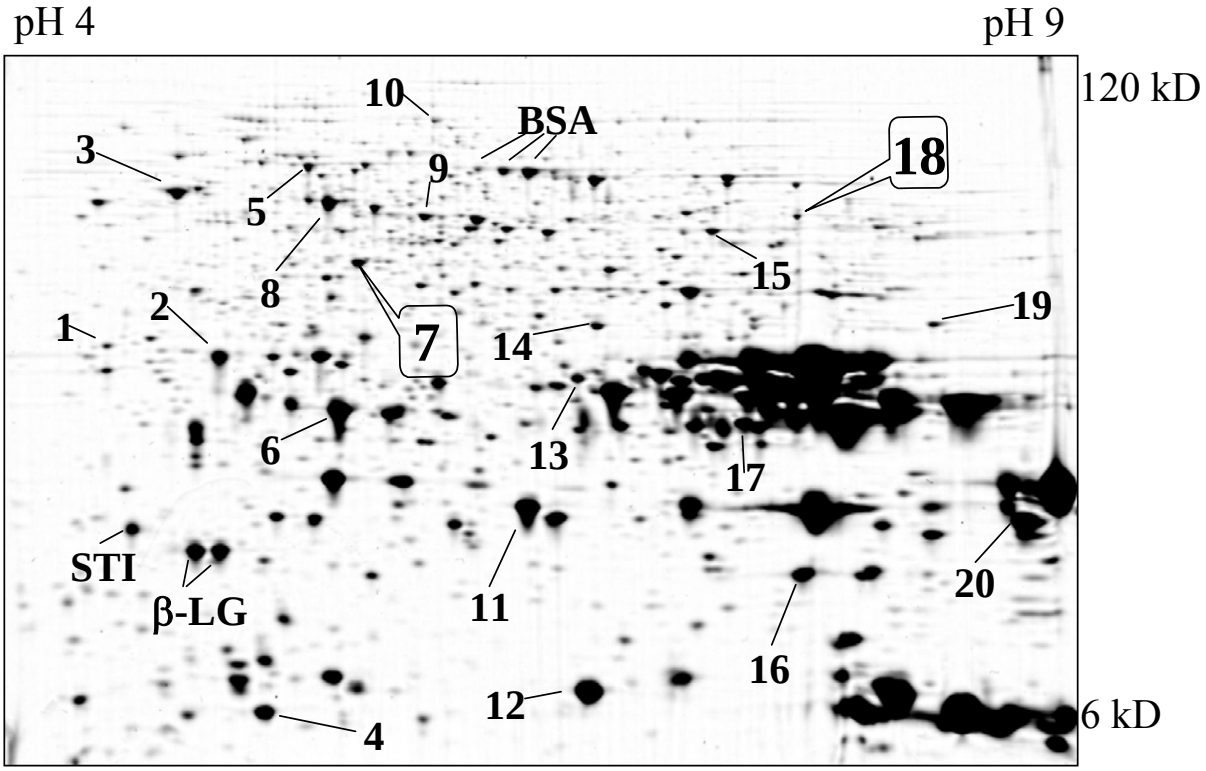


Figure 27 Filtered 2DE image of the 125 μg total protein loading (Figure 28). External protein standards and subset member spots are labeled. External protein standards: soybean trypsin inhibitor (STI), bovine serum albumin (BSA), β-lactoglobulin (β-LG). Calibration plots of Spot 7 and 18 are shown in Figure 33 on page 98.

An *Arabidopsis* seed extract was diluted to seven different protein levels spanning a 100-fold range of total protein load from 2.5 to 250 μg (Figure 28). The two protein standards were spiked into the seven dilutions in different amounts to span a 200-fold range (BSA: 5-1000 ng; B-LG: 4-820 ng; Figure 28). In order to normalize the spot quantities, soybean trypsin inhibitor (STI: pI 4.5, Mw 21.5 kD) was added to each dilution at the same level (150 μg for each gel) as a reference standard. Three replicates were performed for each dilution, and all of the gels were stained using the colloidal Coomassie Brilliant Blue G-250 (CBB G-250) method modified by Neuhoff et al. [258]. The gels were stained to equilibrium for three days at room temperature.

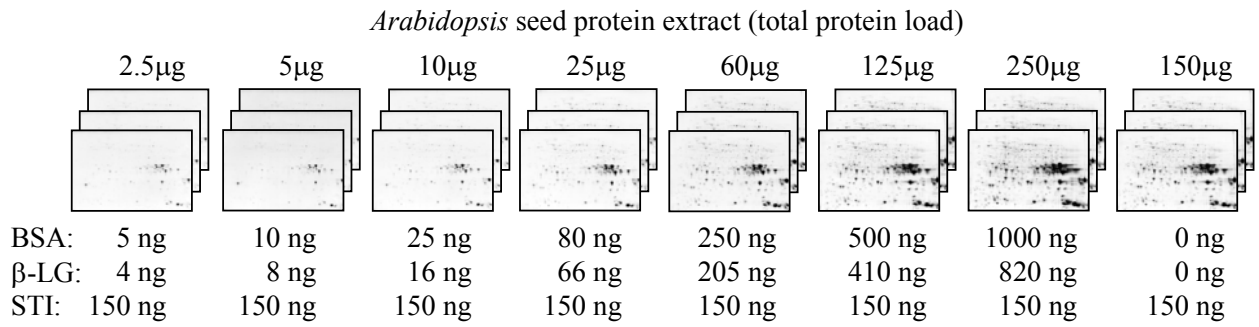


Figure 28 An *Arabidopsis* seed extract was diluted to seven different protein levels plus one control dilution. Two external protein standards (β-LG and BSA) were spiked into dilution 1 to 7 to span a 200-fold range. No standard was added to the control (150 µg). STI was added to all dilution in the same concentration as normalization standard. Three replicates were performed for each dilution.

4.2.3.1 Sensitivity

The limits of detection of BSA and β-LG were determined in order to estimate the absolute sensitivity of the developed 2DE method. Figure 29 shows the 2DE gel sectors of the two external protein standards for each dilution. Image brightness and contrast settings are the same for all images. Bovine serum albumin (BSA) separated into three protein spots (Figure 29 BSA-1, 2, 3) with the same Mw of 66.7 kD and pIs of 5.50, 5.55, and 5.60. β-lactoglobulin separated into two protein spots (Figure 29 β-LG-1, 2) with the same Mw of 20.1 kD and pIs of 4.6 and 4.7.

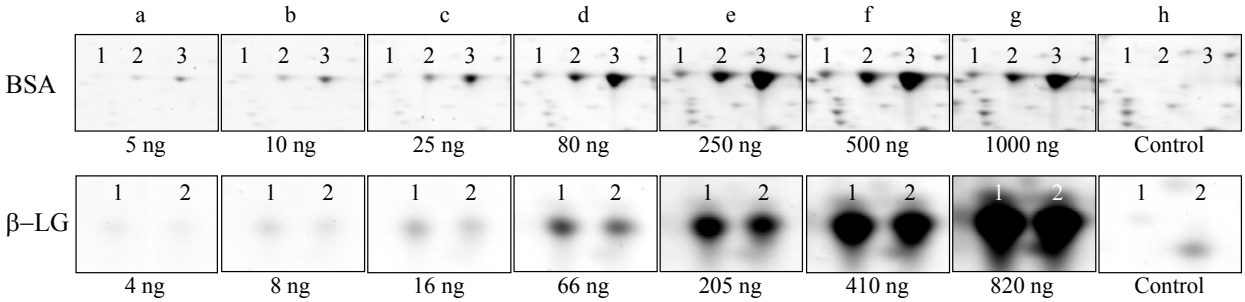


Figure 29 Bovine serum albumin (BSA) and β-lactoglobulin (β-LG) separated by 2DE. BSA appeared as two main spots (2 and 3) and one faint spot (1) with pIs of 5.50, 5.55, and 5.60, and a Mw of 66.7 kD. β-LG appeared as two spots (1 and 2) with pIs of 4.6 and 4.7, and a Mw of 20.1 kD. The absolute protein load of BSA and β-LG is indicated below corresponding image. The total *Arabidopsis* seed protein load was for dilution (a) 2.5 µg; (b) 5 µg; (c) 10 µg; (d) 25 µg; (e) 60 µg; (f) 125µg; (g) 250 µg; (h) 150 µg. Images of a dilution series are displayed with the same brightness and contrast settings.

The automatic spot detection tool of PDQuest (settings see Chapter 3.2Methods) detected BSA spots 2 and 3 down to 5 ng, the lowest tested nanogram level, and spot 1 down to 80 ng BSA. Both spots of β-LG were automatically detected down to 66 ng β-LG (dilution d). It seems that at similar nanogram levels (e.g., 10 ng BSA vs. 8 ng β-LG) BSA spots 2 and 3 gave a stronger response than spots 1 and 2 of β-LG (Figure 29). The difference in staining intensity may result from the fact that Coomassie Blue predominately binds to basic and sulfur-containing amino acids of proteins [320]; in addition, BSA has, by weight, more basic amino acids and cysteines than B-LG. In addition, low-mass spots have a higher diffusion coefficient than high-mass spots [321]. This leads to wider spots and, therefore, to a dilution of the response.

After adjusting image brightness and contrast for each image section individually, all the spots of BSA and β-LG were clearly distinguishable from the background, even at the lowest tested nanogram level of 5 and 4 ng for BSA and β-LG, respectively (Figure 30). Taking into account that the protein amount of 5 ng for BSA is divided by three spots with ratios of approximately 8, 25, and 67% of total spot quantities (quantities of all three spots summed up), approximately 1.3 ng BSA protein was detectable automatically with PDQuest and as little as 0.4 ng BSA protein was visually detectable. β-LG separated into two spots with approximate proportions of 43 and 57% of total spot amount; i.e. down to 1.7 and 2.3 ng β-LG were visually detectable by using 2DE combined with colloidal CBB.

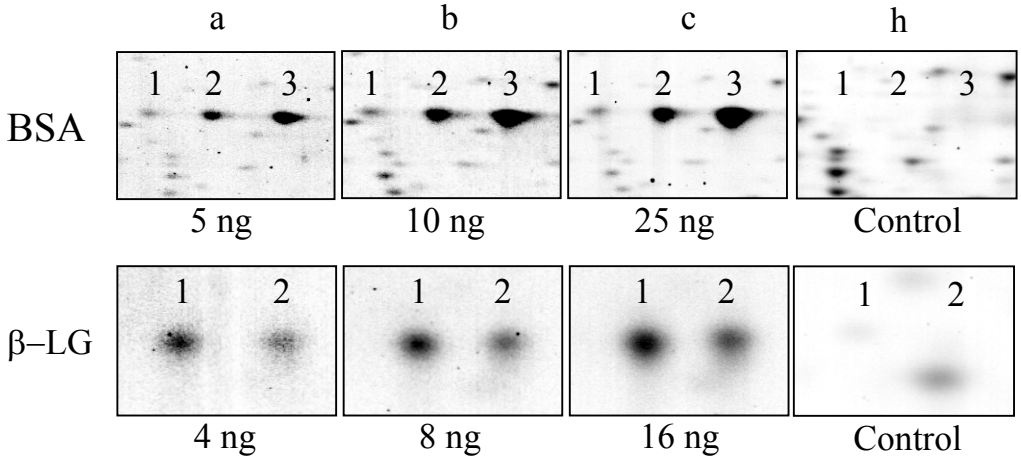


Figure 30 BSA and β-LG separated by 2DE. Absolute protein amount loaded on gel is indicated below images. The control is the seed protein extract without added external protein standards. Total protein loading of *Arabidopsis* seed protein was (a) 2.5 μg; (b) 5 μg; (c) 10 μg; (h) 150 μg. Image brightness and contrast are optimized for each image individually in order to improve distinction of background noise from real data.

The images of the control dilution h (not spiked with protein standards) verify the absence of any endogenous plant proteins migrating to the same position as the protein standards (Figure 30). Only one endogenous plant protein migrated slightly above the position of β-LG's spot 1. However, the spot quantity is small as seen in the control gel (Figure 30 h: 150 µg total *Arabidopsis* seed protein loading) and, therefore, can be neglected.

4.2.3.2 Linearity

The linear dynamic range of the method was determined for each of the two spiked standard proteins. The tested range was 200-fold (BSA: 5-1000 ng; B-LG: 4-820 ng; Figure 28). In order to evaluate the linear relationship between spot quantity and protein-loading amount, the total spot quantity (sum of individual spots of a protein standard) was plotted against protein load. A linear relationship was considered when the coefficient of determination (R^2) between spot quantity and the protein-loading amount was greater than 0.9 and the lack of fit was insignificant (P> 0.05). The absolute linear dynamic range of BSA and β-LG was found to be 5-250 ng (50 fold) and 8-820 ng (100 fold), respectively (Figure 31). A linear relationship over the same range was also seen when the quantities of the individual spots were plotted against protein amount.

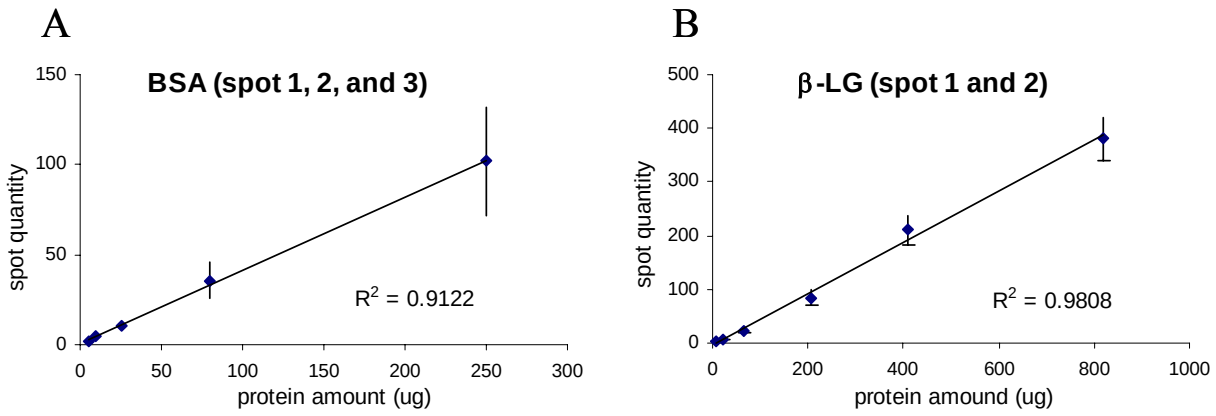


Figure 31 Calibration plots of BSA and β-LG with best-fit line. Spot quantities of individual spots of an external protein standard were summed to the total quantity of this standard (i.e., assume one spot). (A) BSA demonstrated a linear relationship (R^2>0.9) from 5 to 250 ng (50-fold range); (B) β-LG demonstrated a linear relationship (R^2>0.9) from 8 to 820 ng (100-fold range). Results are means ± standard deviation for three gels from each dilution.

BSA showed saturation effects at the two highest protein amounts (500 and 1000 ng). Thus, the 500 and 1000 ng spot quantities were excluded from the calibration plot (Figure 31A). For β-LG, the 4 ng spot was excluded from the linearity test because its average spot quantity was almost equal to the average spot quantity of the 8 ng spot, and its coefficient of variation was 86%; clearly larger than the average CV (11%) for the other spots (Figure 31B).

In order to investigate the linearity for *Arabidopsis* seed proteins, a subset of 20 seed proteins, representing a wide range of different pIs, Mws, and abundances, was chosen (#1-20, Figure 27). The linearity was tested over a 100-fold range. The total protein loading amounts for the dilution gels were 2.5, 5, 10, 25, 60, 125 and 250 µg. Figure 32 shows the PDQuest histograms of all 20 spots and spot quantities plotted against total protein loads for each spot.

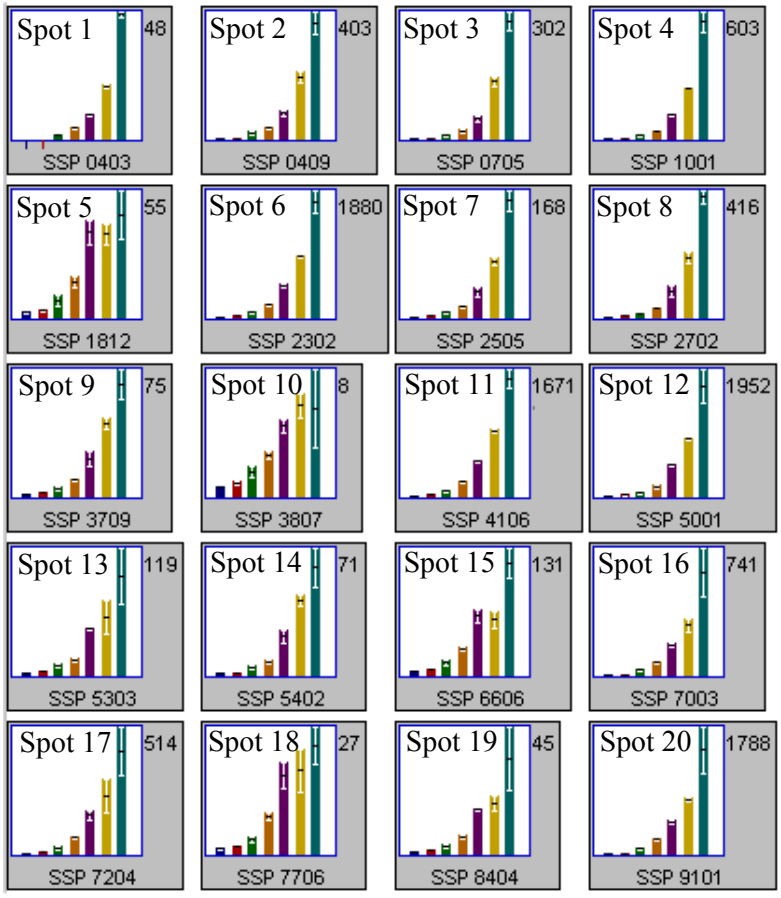


Figure 32 PDQuest histograms of the 20 selected spots. The average spot quantities (*N* = 3) of the seven protein levels are graphed from left to right in the following order: 2.5, 5, 10, 25, 60, 125, 250 µg. A linear response over the 100-fold range was observed for most of the spots.

Figure 33 shows a representative example of a linear relationship between spot quantity and protein loading (Figure 33A) and a representative example for a protein spot that shows saturation effects above a total protein load of 60 µg (Figure 33B).

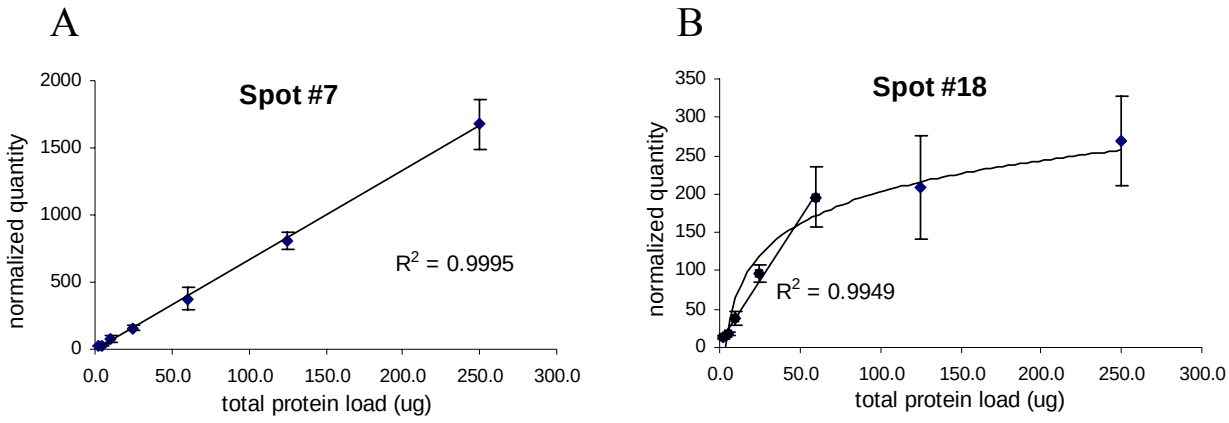


Figure 33 Examples of the relationships between spot quantity and total protein amount loaded (2.5-250 µg). (A) Linear relationship was demonstrated over the entire protein-load range; (B) staining saturation is shown above 60 µg of total protein load. Results are means ± standard deviation for three gels from each dilution.

The coefficient of determination (R^2) and lack of fit were calculated with the statistical software package JMP. A linear response ($R^2 > 0.9$ and lack of fit: $P > 0.05$) over the entire detected range was demonstrated for 16 spots (80%). Fifteen of these spots were linear over a 100-fold range (2.5-250 µg total protein) and one over a 25-fold range (10-250 µg total protein). In three cases (spot 1, 13, and 19), a Log transformation of the spot quantities was performed to remove the relationship between the mean and the variance. However, four spots (#5, #10, #15, #18) showed saturation effects and had a limited dynamic range (2.5 to 60 µg total protein). All four spots are located in the upper quarter of the 2DE gel (Mw > 50 kD) suggesting a relationship between Mw and the saturation effect. A possible explanation may be that high molecular weight proteins have a smaller diffusion coefficient than low molecular weight proteins and tend to be more concentrated in one spot. However, spots with similar molecular weights, like spot 3 (61.7 kD), spot 8 (58.2 kD), and spot 9 (53.9 kD), do not show such saturation effects. Thus, the linear range depends on the protein itself rather than on Mw. Among the subset of proteins from *Arabidopsis*, no relationship was seen between saturation effects and pI.

4.2.4 Summary

In the developed technique, the spot position is highly repeatable for the isoelectric point and molecular weight dimensions. The repeatability of the 2DE protein pattern (presence of spot and spot quantity) may be influenced by the extraction process (extract-to-extract variation) and/or by the 2DE method itself (gel-to-gel variation). The spot number variation was primarily affected by the 2DE method itself as seen by the large SD of replicate gels (SDs from 45 to 129) compared to the smaller standard deviation of extracts (SD of 20). The three extracts displayed the same protein pattern with no missing or unique spots present. The degree of quantitative variation was also mainly determined by the gel-to-gel variation, while the extract-to-extract variation was negligible. The average quantitative gel-to-gel variation was found to be 24.8% (median 19.6%). The degree of analytical variation due to the 2DE process establishes an important baseline for uncoupling biological variation from analytical variation.

Due to the negligible impact of extract-to-extract variation on repeatability of the 2DE protein patterns, a sample is considered as one extract and replicates are 2DE gels performed from this extract. A replicate size of three is sufficient to investigate a 3-fold difference in protein expression for almost all proteins and to investigate a 2-fold difference for the majority of proteins with good power (90%). If a protein spot of interest has a high CV because of focusing problems (irregularly shaped) and cannot be tested for a 2-fold difference, its quantity can be accurately determined by using the spot boundary tool of the PDQuest software program.

The optimized method was sensitive enough to detect BSA and β-LG amounts down to 5 ng and 66 ng, respectively. Taking into account that the BSA protein amount of 5 ng was shared by three protein spots in a 1:3:8-ratio, the lowest amount of protein detected by the software program PDQuest was less than 2 ng. Protein amounts less than 1 ng were visually detectable after optimizing image brightness and contrast.

The absolute linear range of the method was demonstrated for two protein standards. BSA and β-LG demonstrated a linear response from 5 to 250 ng protein (50-fold) and from 8 to 820 ng protein (100-fold), respectively. For all protein spots of a subset of the *Arabidopsis* seed proteins, a linear relationship between total protein loading amount and spot quantity was demonstrated for at least a 25-fold range. Many of the spots (75%) showed a linear

response over a 100-fold range. The demonstrated linear dynamic range of loaded protein amount and spot quantity is higher than the linear ranges published for colloidal Coomassie Brilliant Blue.

However, the differences between BSA and β-LG in their LOD and range of linearity demonstrate that proteins behave differently depending on their amino acid sequence and physicochemical properties. For example, saturation effects were more often found among high molecular weight proteins (Mw>50 kD).

4.3 Natural variation

Naturally occurring genetic variation is commonly found in all mammals, microbes, and plants, including *Arabidopsis*. Those random differences in an organism's genome are the basis of natural selection of a species. DNA polymorphism is the molecular basis of natural variability and results from various forms of mutations. Such heritable changes in the DNA directly affect the proteome unless the mutated codon codes for the same amino acid (silent mutation) due to the degeneracy of the genetic code. A point mutation (base-pair substitution) of the DNA, e.g., can result in (1) amino acid substitution in a protein (missense mutation) and can lead to phenotypic changes such as early flowering in *Arabidopsis* (Ler vs. Cvi) [322, 323], (2) change of a mRNA codon from one that specifies an amino acid to a chain-terminating (nonsense) codon with the result of premature chain termination and truncated proteins (nonsense mutation), and (3) change of a mRNA codon from a chain-terminating codon to one that specifies an amino acid resulting in an elongation of the resulting protein. An amino acid substitution may have secondary effects, including loss of phosphorylation or glycosylation sites or alteration of the degradation stability of the resulting protein. Addition or deletion of a base pair in a gene shifts the reading frame by one base and leads to a change in the amino acid sequence of the protein (frameshift mutation). All these mutations could affect the net charge (isoelectric point) and/or molecular weight of the resulting protein and, therefore, the protein's 2DE migration behavior (electrophoretic mobility). Protein polymorphism seen on 2DE can be divided into two categories. One is qualitative variation; these are a) position shifts of allelic protein spots in pH (horizontal shift), Mw (vertical shift), or both pH and Mw (diagonal shift) direction or b) separation of a protein into isoforms (spot series because of alternative posttranslational modification's sites). The second is quantitative variation; with either a) change in protein amount or b) a disappearance of a protein spot or appearance of a new protein spot. Quantitative variation may be a result of changes in the amino acid sequence (e.g. change of degradation stability) or mutations of noncoding DNA sequences (promoter regions) [324].

Another important factor for protein expression are the environmental conditions. For example, drought stress affected 78 proteins in maize leaves [325], 38 proteins in maritime pine needles [326], and 42 proteins in rice leaves [327]. Bahrman et al. [328] demonstrated the effect of various nitrogen levels on the protein expression in wheat leaves. Almost 15% (76 spots) of the compared protein spots (524 spots) showed a significant *N* treatment effect.

In order to evaluate the biological significance of a difference between transgenic and its near isogenic parental line, it is important to have a comprehensive understanding of the natural variation of protein expression within and among the investigated plant species. It is important to assess whether a detected difference in protein expression, including "new" protein, increase, or decrease of a specific protein, can also be found in nature, i.e. under other environmental conditions or in other varieties of this plant species. This is particularly true for the assessment of the safety relevance of detected differences between transgenic and non-transgenic plants (Figure 34). Many international organizations [11, 13, 329, 330], suggest that further assessment (nutritional and toxicological) should only be required if the differences exceed natural variation in traditional food crops.

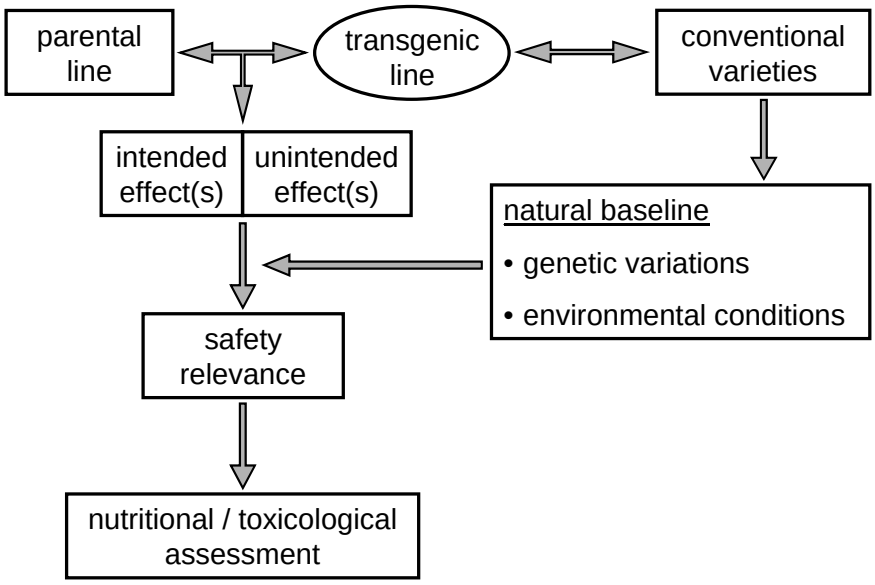


Figure 34 Sequential approach for safety assessment

4.3.1 Environmental

The influence of environmental conditions on the proteome has been studied as a response to abiotic stress. These include water deficit [325-327, 331], various nitrogen levels [328], heat stress [332], cold stress [333-336], anoxia [331], and ozone stress [337]. Seki et al. [338] monitored the expression profile of 7000 *Arabidopsis* genes under drought, cold, and high-salinity stresses using cDNA microarrays. They identified 277 drought-inducible, 53 cold-inducible, and 194 high-salinity stress-inducible genes. All these studies investigated the immediate response of the leaves to the environmental conditions but not the impact on the protein profile of the seeds. Monteiro et al. [339] studied the effect of environmental conditions during vegetative growth on proteins in mature grapes. The proteins from mature

grapes harvested in two consecutive years were analyzed by 2DE. There were tremendous differences. The protein patterns showed only 33% similarity when the presence and absence of protein spots were compared.

4.3.1.1 Plant-to-plant variability

4.3.1.1.1 Comparison of phenotypic traits

Six *Arabidopsis* plants (ecotype Col-0) were grown in the same growth chamber under environmentally controlled conditions. Three phenotypic traits were assayed and are summarized in Table 20. The first flowering day (FFD, Boyes growth stage 6.00 [340]) varied between 29 and 34 days. The rosette diameter (RD) and the seed yield of the individual plants were found to range from 6 to 8.5 cm and from 643 to 795 mg, respectively.

Table 20 Phenotypic measurement of the six individual *A. thaliana* (Col-0) plants

Sample	FFD[1)] (days)	RD[2)] (cm)	Seed yield[3)] (mg)
Plant 1	31	8	643
Plant 2	30	6	715
Plant 3	34	8.5	727
Plant 4	31	7	768
Plant 5	29	8	774
Plant 6	31	8	795
Mean±SD	31 ± 2	7.6 ± 0.9	737 ± 55
Mean±SD for >300 individual plants [340]	31.8 ± 3.6	8.2 ± 1.5	128 ± 53[4)]

[1)] FFD = number of days from the date of planting until the opening of the first flower
[2)] RD = rosette diameter at the time of first flowering
[3)] Amount of harvested seeds
[4)] desiccated

The mean and the range of FFD (31±2 days) and RD (7.6±0.9 cm) data are in accordance to the findings of Boyes et al. [340], who assayed the FFD (31.8±3.6 days) and RD (8.2±1.5 cm) of over 300 individual plants (Table 20). Boyes et al. measured 128±53 mg desiccated seed per plant. This is almost six times less than the 737±55 mg seed per plant found in the present study. The difference in seed treatment, desiccated (Boyes et al.) versus fresh weight (present study) seeds, is not important due to the low moisture content (on average 5%) of mature *Arabidopsis thaliana* seeds. A possible explanation for the seed yield difference between both studies may be differences in the seed harvesting methods. Boyes et al.

harvested the inflorescences 2 to 4 days after the completion of flowering and stored them in an envelope to complete the maturation process. In the present study, irrigation of the plants was continued until the plant reached complete senescence, and all the siliques were yellow. In order to prevent seed loss, the upper part of the plant was wrapped with a plastic tube that was fastened just above the rosette leaves (Figure 35A) and open at the other end (Figure 35B). This modified seed harvesting method prevents any seed loss due to premature siliques dehiscence and provides a more accurate way to measure the seed yield for *Arabidopsis* plants. Higher precision was obtained by using this method as it reduced variation of seed yields (CV = 7% present study versus CV = 41% [340]).

Figure 35 Pictures of seed collection apparatus on the *Arabidopsis* plants. Plants were wrapped into plastic tubes after first siliques were formed in order to prevent seed loss.

The observed variability in FFD and RD is in accordance with findings of Boyes et al. [340], who assayed the FFD and RD of more than 300 *A. thaliana* (Col-0) plants. While variation in seed yield cannot be compared because of different harvesting protocols, it demonstrates advantages in the seed collection method utilized in these studies.

4.3.1.1.2 Comparison of 2DE pattern

The seed protein profiles of six individual *Arabidopsis* Col-0 plants (Plant 1 to 6) and of their pooled seed samples (three independent extractions labeled as Mix 1 to 3) were compared. Table 21 summarizes the results from the spot detection and matching using PDQuest software and the manual verification. The number of detected spots varied from 477 ± 65 protein spots for Plant 6 to 616 ± 129 protein spots for Plant 1. The number of reproducible spots varied between 379 for Mix 1 and 481 for Plant 1. Three hundred forty six spots were matched between all 27 gels of the 9 samples, and 376 spots were matched between all 18 gels of the plant samples. All reproducibly present spots of one sample (Mix or Plant) were present in at least one replicate gel of all other samples; in other words, no reproducibly missing spot was detected for any of the samples.

Table 21 Results of spot detection and matching for six different *Arabidopsis* plants and three protein extractions of their pooled seed samples

Sample	Number of spots		Reproducibly	
	average	SD*	present spots	missing spots
Mix 1	517	117	379	0
Mix 2	543	45	419	0
Mix 3	557	129	403	0
Plant 1	616	129	481	0
Plant 2	530	89	416	0
Plant 3	556	25	458	0
Plant 4	528	26	456	0
Plant 5	520	96	395	0
Plant 6	477	65	388	0

* SD = standard deviation

The spot quantity ratio of the 376 spots present in all 18 2DE gels of the six plant samples were computed. Therefore, the highest spot quantity was divided by the lowest spot quantity for each spot. The statistical significance of all spots with a ratio greater than 2 was tested with the statistical software program JMP, using the Tukey-Kramer Test with a significant level of 0.05. Nine protein spots demonstrate a statistically significant difference in spot quantity greater than 2-fold (Table 22).

Table 22 Plant-to-plant variation of spot quantities: spots with statistically significant differences greater than 2-fold.

Spot	Compared plants	Spot quantity (IOD) Min.	Max.	Ratio	Power
SSP 2417	Plant 5 vs. Plant 1	631	1597	3	90%
SSP 8606	Plant 4 vs. Plant 3	478	1261	3	94%
SSP 8204	Plant 4 vs. Plant 3	548	1477	3	>99%
SSP 1206	Plant 4 vs. Plant 3	193	546	3	96%
SSP 1205	Plant 4 vs. Plant 6	97	294	3	87%
SSP 4203	Plant 6 vs. Plant 4	570	1891	3	>99%
SSP 8717	Plant 6 vs. Plant 3	152	541	4	87%
SSP 6507	Plant 6 vs. Plant 4	39	295	8	>99%
SSP 8203	Plant 6 vs. Plant 4	206	2804	14	>99%

It is apparent that Plant 4, Plant 6, and Plant 3 with 7, 5, and 4 out of 9 times contribute often to the plant-to-plant variability. The largest variation was observed for spot SSP 8203 between Plant 6 and Plant 4. Figure 36 shows the gel regions for Mix 1 to Mix 3 and Plant 1 to Plant 6 for spot SSP 8203. The gels and the chart illustrate the spot quantity variation of spot SSP 8203 among the six plants of one line and the three pooled samples. The line in the chart represents the overall mean. As expected, the values for the pooled seed samples (Mix 1 to 3) are arranged around the overall mean.

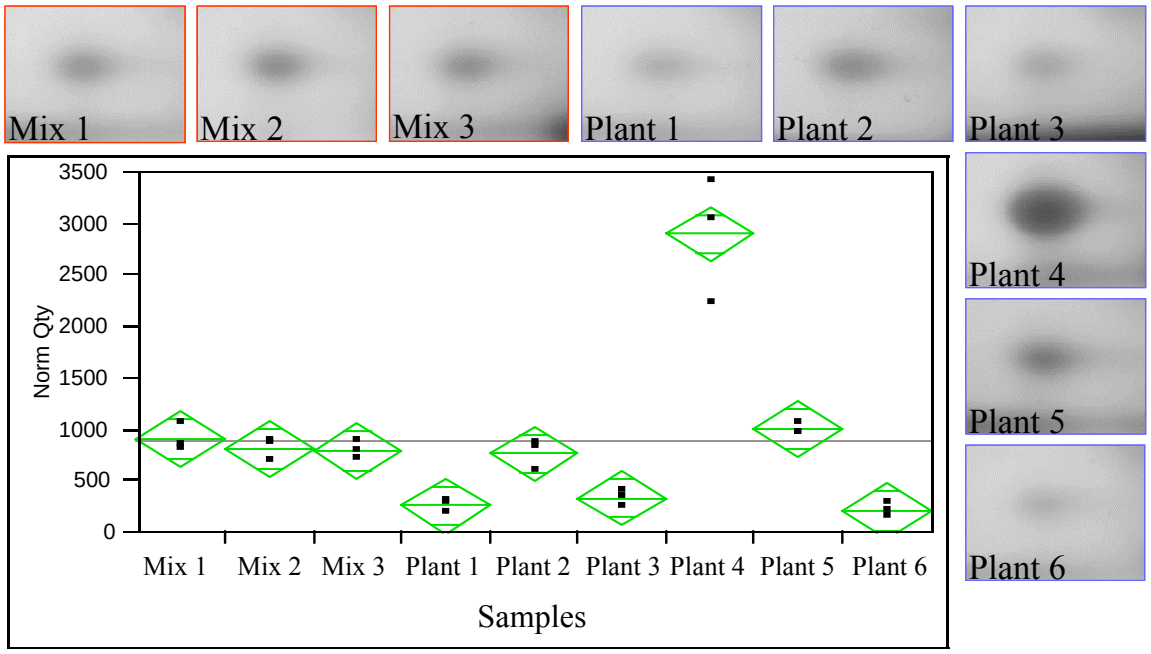


Figure 36 Spot SSP 8203 (Mw: 22.8 kD, pI: 6.7) as an example for the plant-to-plant variation in spot quantity. The diamonds illustrate the samples mean and the 95% confidence interval.

The observed differences in spot quantity may be explained either by slight differences in environmental conditions and/or differences in the genetics of the six plants. Although the

environmental conditions in the growth chamber were controlled as much as possible, environmental differences, such as light gradient, soil compression of the pots (water retention capacity), and effects due to neighboring plants, may affect the growth of the individual plants. This data is not sufficient to distinguish between environmental or genetic effects as the cause for the differences in spot quantity. Cooke [341] describes plant-to-plant variability within highly self-pollinating crops, such as wheat and other cereals produced by single plant descent techniques. The author calls them biotypes because they consist of more than one electrophoretically identifiable line. This study indicates that individual plant analysis may inaccurately reflect differences in protein profiles.

4.3.1.2 Effect of different environments

Arabidopsis plants from the same seed pool were grown in two different growth chambers (CG) in two consecutive years. Seeds from 6 to 8 plants from each growth chamber were pooled to one representative sample and analyzed by 2DE. Table 23 summarizes the environmental conditions in the two growth chambers.

Table 23 Environmental conditions of the growth chambers GC-1 and GC-2

	GC-1 (2002)	**GC-2** (2003)
Type	large GC with tables	small GC with racks
Lighting system	high intensity discharge lamps	fluorescence lamps
Light intensity ($\mu E/s/m^2$)	150 –200	150 –200
Light cycle (h)	16	16
Fertilizer	Peter's 20:20:20	Peter's 20:20:20
Fertilizer cycle	2/week	2/week
Fertilizer amount	100 ppm	100 ppm
Irrigation cycle	2/week	2/week
Relative humidity (%)	70	70
Temperature (°C)	20	20

Most of the growth chamber parameters in Table 23 are the same, such as light intensity, light cycle, fertilizer cycle/amount, irrigation cycle, humidity, and temperature. The major difference between the growth chambers is the lighting system. GC-1 uses high intensity discharge lamps, which have a high concentration of photon flux and provide a full spectrum of light with a stronger emphasis on the blue end of the spectrum. GC-2 instead uses fluorescence lamps, which provide a very balanced spectrum of light [William R Schuler (Monsanto), personal communication].

The results of the phenotypic comparison between plants grown in growth chamber GC-1 in 2002 (Col-0/GC-1) and in growth chamber GC-2 in 2003 (Col-0/GC-2) are listed in Table 24. Although only two data points for the first flowering date (FFD) and rosette diameter (RD) for Col-0/GC-2 are available, it seems that plants grown in growth chamber GC-2 with fluorescence lamps had an extended vegetative phase (FFD: 39 days in GC-2 vs. 31 days in GC-1) and also an increase in biomass (RD: 10.8 cm in GC-2 vs. 7.4 cm in GC-1). However, the seed yield is almost identical with 722 ± 92 mg for Col-0/GC-1 versus 721 ± 101 mg for Col-0/GC-2. Interestingly, the seed protein content measured as Protein (% fw) is on average 3% lower for seeds from plants grown in GC-2. Hence, growing the same *Arabidopsis* line (Col-0) in two different growth chambers had a clear impact on the phenotype.

Table 24 Phenotypic measurements, Col-0 grown in growth chamber GC-1 in 2002 (Col-0/GC-1) and in growth chamber GC-2 in 2003 (Col-0/GC-2)

	Col-0/GC-1	Col-0/GC-2
Sample size (N)	8	6
FFD (days)	31 ± 1	39 ± 0 (N=2)*
RD (cm)	7.4 ± 1.0	10.8 ± 0.4 (N=2)*
Seed yield (mg/plant)	722 ± 92	721 ± 101
Protein (% fw) (pooled seed; N=2)	25.7	22.7

* only two measurements were taken

Pooled seed samples were analyzed by 2DE. Figure 37 shows representative images of the 2DE seed protein profiles for (A), plants grown in growth chamber GC-1 in 2002, versus (B), plants grown in growth chamber GC-2 in 2003.

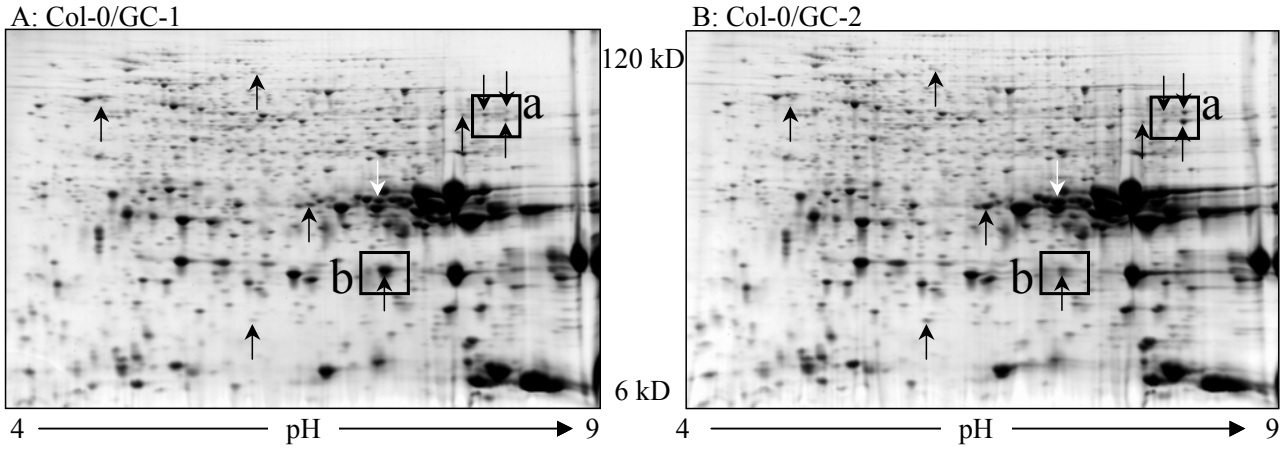


Figure 37 Representative 2DE gels of seed proteins from plants grown in (A) growth chamber GC-1 in 2002 and (B) growth chamber GC-2 in 2003. Boxes a and b correspond to gel regions enlarged in Figure 39 and Figure 41, respectively. Spots varied more than 2-fold in quantity are marked with arrows.

Based on the qualitative comparison of the 2DE protein profiles, the seed proteomes of plants grown in two different growth chambers are very similar. Table 25 summarizes the qualitative comparison of Col-0/GC-1 and Col-0/GC-2.

Table 25 Qualitative seed proteome comparison between Col-0/GC-1 and Col-0/GC-2

	Col-0/GC-1	Col-0/GC-2
Detected spots[1)]	726 ± 6	747 ± 9
Reproducible spots[2)]	707	708
Spots matched to both	702	702
Spot reproducibly missing[3)]	1 (SSP 8619)[4)]	0

[1)] false positive spots due to edge effects, focusing problems, or dust were excluded
[2)] spot must be detected in all three replicate gels
[3)] spot must be absent in all three replicate gels
[4)] very faint spot (below 100 ppm IOD) in 2DE gels of Col-0/GC-1

On average, 726±6 spots were detected for Col-0/GC-1 and 747±9 spots for Col-0/GC-2. The majority of them (707 spots for Col-0/GC-1 and 708 spots for Col-0/GC-2) were detected in all three replicate gels and therefore considered to be reproducibly present. The seed protein patterns of the plants grown in two different growth chambers in consecutive years were very similar based on the presence/absence of proteins. Seven hundred and two protein spots were matched to both samples. However, one protein spot (SSP 8619, Mw: 52.5; pI: 6.7) was reproducibly present in Col-0/GC-2 and reproducibly absent in Col-0/GC-1 (Figure 38). Although SSP 8619 is a very faint spot (IOD below 100 ppm), its presence in Col-0/GC-2 and its absence in Col-0/GC-1 is clearly demonstrated. The identity and function of this protein were not pursued since these were not the purpose of the study.

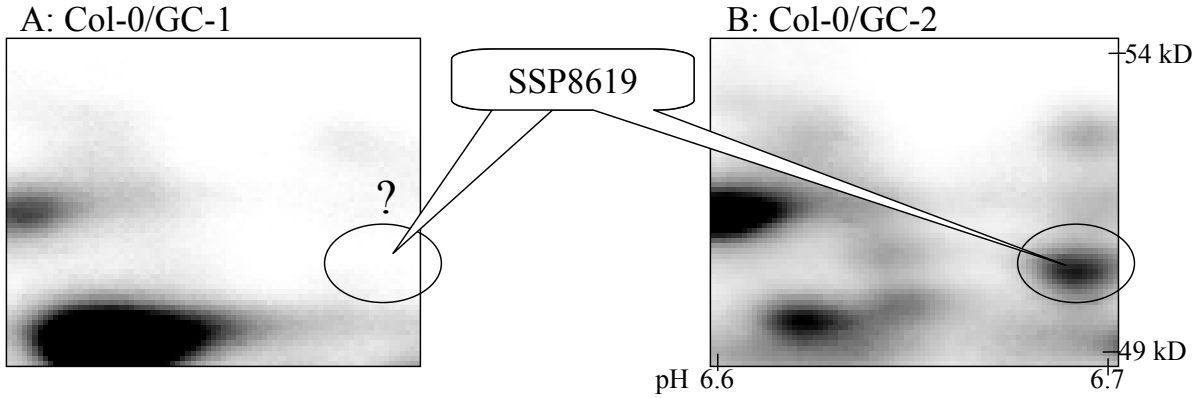


Figure 38 Enlarged regions of the 2DE gel prepared from (A) Col-0/GC-1 seeds and (B) Col-0/GC-2 seeds illustrating the missing spot in the seed proteome of the Col-0/GC-1 sample.

The seed protein expression profile of Col-0/GC-1 was compared to the expression profile of Col-0/GC-2 based on the quantity of the 702 spots reproducibly detected in both seed samples. Upon manual verification of the results, it became obvious that the quantities of protein spots with low values in average spot quality are frequently overestimated, as discussed in Chapter 4.3.2. To avoid overestimation of quantitative differences due to inaccurate quantification of poorly resolved protein spots, spots with the maximum value and an average spot quantity less than or equal to 40 were excluded from consideration in the comparison. In addition, the detected statistically significant difference (P<0.05) must have a power above 80%. Power is the probability to say correctly that the means in this ratio are different. The power takes the number of replicates and the variance of the measurements into account and is therefore, an indicator of false positive differences. All statistical analyses were performed with the statistical software package JMP. The results of the quantitative comparison of seed proteomes of Col-0/GC-1 and Col-0/GC-2 are summarized in Table 26. Of the 702 spots detected for both Col-0 seed samples grown in two different growth chambers, 52 spots (7.4%) were found to be statistically significantly (P<0.05) different in their quantity by two or more fold. Most of the significantly different spots (40 of the 52 spots) have a higher quantity in the seeds of Col-0 plants grown in growth chamber GC-2. Ten spots varied more than 2-fold in quantity with a maximum of 8-fold difference between Col-0/GC-1 and Col-0/GC-2. These spots are marked with arrows in Figure 37.

Table 26 Quantitative seed proteome comparison between Col-0/GC-1 and Col-0/GC-2

	Col-0/GC-1 vs. Col-0/GC-2
Number of 2-fold differences in spot quantity[1) 2)]	42
Number of 3-fold differences in spot quantity[1)]	5
Number of >3-fold differences in spot quantity[1)]	5
Maximum difference	8-fold

[1)] Power of difference must be > 80% and mean spot quality > 40
[2)] Not visually checked

Figure 39 shows the enlarged regions of the 2DE gels prepared from (A) Col-0/GC-1 seeds and (B) Col-0/GC-2 seeds to illustrate the differences in spot quantity detected in the seed profile of plants grown in two different growth chambers.

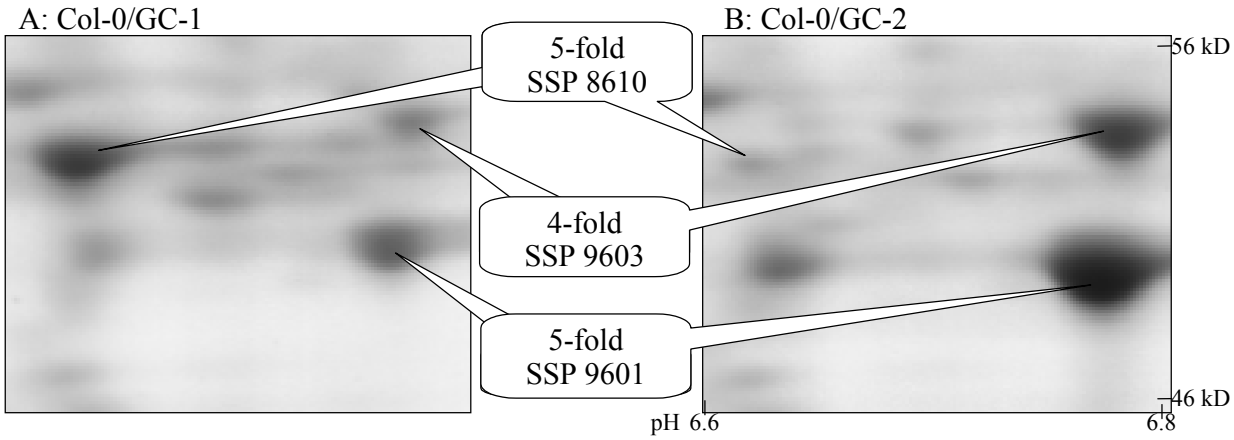


Figure 39 Examples for differences in spot quantity detected in the seed profiles of plants grown in two different growth chambers. Enlarged gel regions (a) from Figure 37.

The seed samples of Col-0/GC-1 were stored for 13 months prior to the comparison. Thus, effects of storage could have contributed to differences in protein spot quantities, although the storage conditions of the seeds (stored in freezer bags containing desiccant at 4°C) were carefully chosen to minimize any enzymatic activities. In order to exclude possible storage effects, the quantities of the 10 spots showing a 3- and more fold difference were analyzed in all 2DE gels run with these seeds over a time period of 13 months. At different time points, the same seeds were extracted and subjected to 2DE analysis. Table 27 displays the ID, growth chamber, harvesting date, extraction date, and the date of the 2DE run of the 2DE gels used for the comparison.

Table 27 Details about seed samples and 2DE gels investigated over time of seed storage

ID	Growth chamber	Harvesting	Extraction	2DE run
1 (2 months)	GC-1	May, 2002	07/30/02	07/31/02
2 (3 months)	GC-1	May, 2002	08/14/02	08/14/02
3 (3 months)	GC-1	May, 2002	08/14/02	08/21/02
4 (6 months)	GC-1	May, 2002	11/17/02	12/04/02
5 (8 months)	GC-1	May, 2002	01/17/03	01/19/03
6 (13 months)	GC-1	May, 2002	06/30/03	07/07/03
7 (<1 month)	GC-2	June, 2003	06/30/03	07/07/03

One spot (SSP 0714; 3-fold) was poorly resolved in most of the previously run 2DE gels and, therefore, was excluded from the comparison data set. Eight of the remaining nine spots did not demonstrate any significant changes in spot quantity over the 13 months in GC-1. Figure 40 shows, as an example, the spot quantity chart for spot SSP 9601 according to the age of the seeds and the growth chamber. The difference in spot quantity of spot SSP 9601 detected

between Col-0/GC-1 and Col-0/GC-2 is not due to storage effects as shown (Figure 40). The 2DE gel region with spot SSP 9601 is also shown in Figure 39.

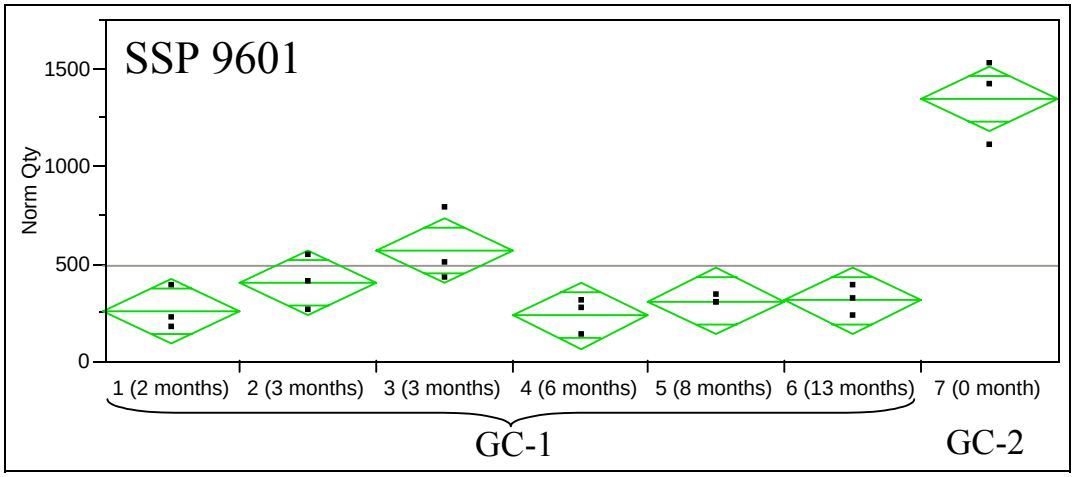


Figure 40 Spot quantity chart for spot SSP 9601 (Figure 39) over time of seed storage in GC-1.

The spot quantity of one of the nine spots increased continuously over 13 months as shown in Figure 41.

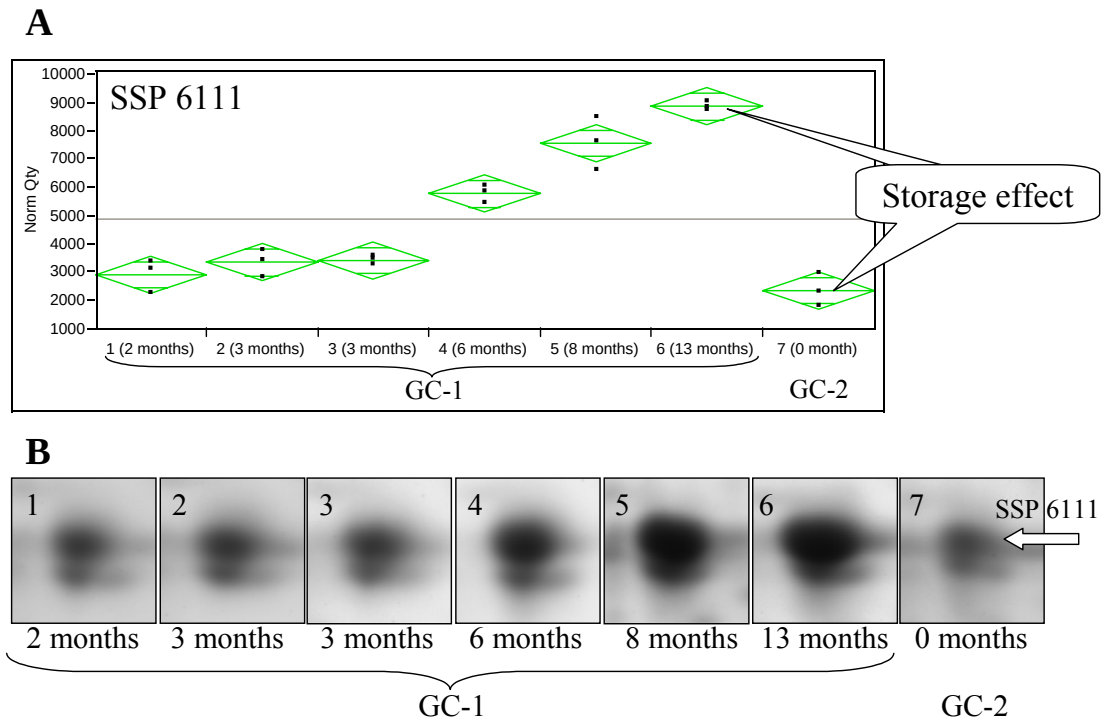


Figure 41 Spot quantity chart (A) and enlarged 2DE gel regions (B) for spot SSP 6111 (Mw: 22.6; pI: 6.1) as example for the impact of storage (2 months to 13 months) on the protein expression. Enlarged gel region (b) from Figure 37.

The increase in protein over time suggests that storage affected the concentration of spot SSP 6111 (Mw: 22.6; pI: 6.1). Murthy et al. [342-344] have demonstrated the importance of non-enzymatic protein modification of proteins through the Maillard reaction during seed storage. However, due to the specificity of the increase, a non-enzymatic protein modification by the Maillard reaction is deemed unlikely. Although the enzymatic activities were controlled by storing the seeds at 4°C under dry conditions, the enzymatic degradation of a precursor protein appears to be the reason for the protein increase over the 13 months of storage.

4.3.2 Genetic background

Genetic variation within a species has been investigated by 2DE for many crops, including barley [228, 345-347], maize [348-351], wheat [352-356], peanut [357], and rice [358]. Although *Arabidopsis* provides an extensive resource for natural genetic variation among ecotypes, only two studies have compared the 2DE profiles of various *Arabidopsis* ecotypes. Marques et al. [292] included the ecotypes Landsberg *erecta* and Columbia into an inter-species comparison within the Brassicaceae family and compared the 2DE protein profile of the aerial part of etiolated seedlings. In a recent study, Chevalier et al. [359] investigated natural variation in the root proteome among eight *Arabidopsis* ecotypes (Col-0, Col-4, Be-0, Ll-0, Rld-1, Cvi-0, Ws-1, and Ler-1). To date, no data has been published on the differences between seed protein profiles among *Arabidopsis thaliana* (*A. thaliana*) ecotypes. Therefore, the impact of the genetic background on the seed proteomes of various *A. thaliana* ecotypes was investigated in this study.

A. thaliana occurs naturally throughout temperate regions of the world including Europe, East Africa, Asia, Japan, North America, and Australia [360]. In order to cover a broad range of geographies, a set of 11 ecotypes from four different continents (10 countries) and one common laboratory line (Table 28) were selected. The 12 ecotypes represent a wide range of genetic diversity as assessed by restriction fragment length polymorphism (RFLP) [361], amplified fragment length polymorphism (AFLP) [360, 362, 363], and cleaved amplified polymorphic sequence (CAPS) [364].

Table 28 Geographic origin of selected *A. thaliana* ecotypes

Geographic region	Ecotype	Origin	TAIR* Stock #
Africa	Cvi-0	Cape Verde Islands	CS6675
	Mt-0	Martuba (Libya)	CS6799
Asia	Condara	Khurmatov (Tajikistan)	CS6175
	Tsu-0	Tsu (Japan)	CS6874
	Ws	Wassilewskija (Russia)	CS6891
Europe	Ll-0	Llagostera (Spain)	CS6781
	Nd-0	Niederzenz (Germany)	CS6803
	Ma-0	Marburg (Germany)	CS6789
	Mr-0	Monte (Italy)	CS6795
	Oy-0	Oystese (Norway)	CS6824
North America	Col-0	Columbia (USA)	
Laboratory line	C24		CS906

*TAIR = The Arabidopsis Information Resource (http://www.arabidopsis.org)

4.3.2.1 Comparison of phenotypic traits

The twelve *A. thaliana* ecotypes were grown side-by-side in an environmentally-controlled growth chamber. For each ecotype, twelve replicates were planted and randomly distributed in the growth chamber in order to limit the influence of environmental factors. Seeds were harvested after complete maturity of all seeds on a plant, i.e. harvesting of the seeds occurred at different times over a period of almost seven months. This was necessary to ensure complete seed development for all ecotypes as flowering time varied from 30 to 63 days depending on the ecotype. Four phenotypic traits (i.e. first flowering date (FFD), rosette diameter (RD), seed yield, and seed protein content) were assayed and are summarized in Table 29. Based on the measured phenotypic traits, the 12 ecotypes cover a wide spectrum of phenotypic diversity. The average FFD (Boyes' [340] growth stage 6.00) varied between 30 and 63 days for Mt-0/Ma-0 and Ws, respectively. The average RD was found to range from 7.4 cm for Col-0 to 13.5 cm for Ll-0. The chosen ecotypes also demonstrated an impressive variation in seed yield with average seed yields ranging from as low as 222 mg per plant for Cvi-0 to up to 1293 mg per plant for Ll-0. To determine the protein content of the seeds, seeds from 6-8 plants of an ecotype were pooled to one sample, and the pooled samples were then analyzed in replicates. Therefore, the standard deviation does not reflect the natural variation within an ecotype and was not specified in Table 29. The average protein content (%fw) varied between 25.7% (Col-0) and 31.0% (Cvi-0 and Ws). A correlation between seed yield and FFD, RD, or protein content of seeds was not found. This is not surprising as the measured parameters are not considered to be linked physiologically. These measurements

have shown that two ecotypes, Mt-0 from Libya (Africa) and Ma-0 from Germany (Europe), are very similar with respect to measured phenotypes and their leaf and stem morphology (visually assessed). Leaf and stem morphology was assessed by overall shape, length, thickness, and pubescence. All other ecotypes demonstrated a range of values in the measured parameters and their morphology.

Table 29 Phenotypic measurements of the selected *A. thaliana* ecotypes

Name	Continent (Country)	N	FFD[1] (days)	RD [2] (cm)	Seed yield[3] (mg)	Protein[4] (%)
Cvi-0	Africa (Cape Verde Isl.)	10	42 ± 6	10.4 ± 1.7	***222 ± 72***	***31.0***
Mt-0	Africa (Libya)	9	***30 ± 2***	8.0 ± 1.0	540 ± 218	25.8
Condara	Asia (Tajikistan)	12	35 ± 3	11.5 ± 1.1	875 ± 300	28.2
Tsu-0	Asia (Japan)	11	38 ± 3	12.3 ± 2.1	1106 ± 309	27.8
Ws	Asia (Russia)	7	***63 ± 8***	12.2 ± 0.8	720 ± 344	***31.0***
Ll-0	Europe (Spain)	8	54 ± 6	***13.5 ± 0.8***	***1293 ± 528***	28.8
Ma-0	Europe (Germany)	11	***30 ± 2***	7.6 ± 1.3	523 ± 71	26.9
Mr-0	Europe (Italy)	8	62 ± 6	11.3 ± 0.7	1137 ± 636	28.9
Nd-0	Europe (Germany)	12	34 ± 6	7.5 ± 0.6	455 ± 144	27.2
Oy-0	Europe (Norway)	8	38 ± 5	9.9 ± 1.4	972 ± 302	26.1
Col-0	North America (USA)	8	31 ± 1	***7.4 ± 1.0***	722 ± 92	***25.7***
C24	Laboratory line	11	37 ± 3	8.0 ± 0.7	590 ± 132	26.1

Values are means ± SD; extreme values are highlighted (italic and bold)
[1] FFD = number of days from the date of planting until the opening of the first flower
[2] RD = rosette diameter at the time of first flowering
[3] Amount of harvested seeds
[4] Protein content of pooled seed samples of six to eight plants; N=2

4.3.2.2 Comparison of 2DE patterns

To insure that all seeds of a plant reach maturity with minimal influence of drought, irrigation was continued until the plant reached complete senescence and all the siliques were yellow (Boyes' growth stage 9.70 [340]). This was crucial because environmental changes may induce changes in protein expression within a plant. For example, in maize leaves, the levels of 78 proteins were changed as a response to drought [325]. Seeds harvested from six to eight individual replicate plants were pooled to one representative ecotype sample to avoid the influence of plant-to-plant variation (Chapter 4.3.1.1) due to genetic and/or environmental factors. The pooled seed samples were extracted and the extracts were subjected to the optimized 2DE protocol. The software package PDQuest was used to analyze the gel images.

Spots detected and matched by the software program were manually verified. False positive and false negative spots were removed and added, respectively.

4.3.2.2.1 Qualitative comparison of 2DE patterns

Using the optimized 2DE technique, 573 (Mt-0) to 653 (Condara) seed proteins were reproducibly resolved per ecotype, with pI ranging from 4 to 9 and molecular weights ranging from 6 to 120 kD. A protein spot was considered to be reproducible when it is present or absent in all three replicate 2DE gels of an ecotype. A representative 2DE gel image of *Arabidopsis thaliana* seed proteome is displayed in Figure 42.

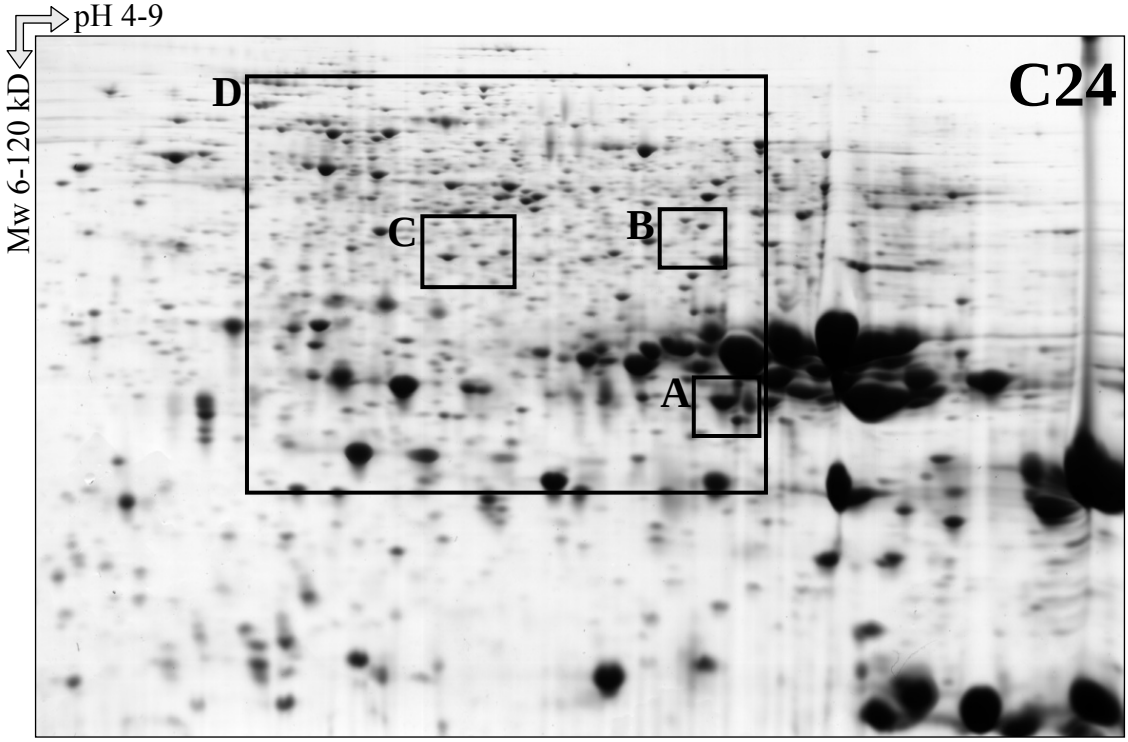


Figure 42 Representative seed proteome pattern (2DE) of *A. thaliana* (150 µg total seed protein of ecotype C24); boxes correspond to gel regions enlarged in Figure 43, Figure 45, and Figure 46.

The verification of a spot's presence or absence and the accurate matching of spots between ecotypes were constrained by large differences in their protein patterns, and, therefore, sometimes ambiguous. For instance, protein spots may remain unresolved due to overlapping effects of neighboring highly-abundant protein spots (Figure 43). For the Cvi-0 ecotype, the protein spot 6211 could not be detected due to the abundance of the neighboring SSP 6105 spot. However, in the Tsu-0 ecotype, the SSP 6211 spot is clearly visible as the SSP 6105 spot is absent.

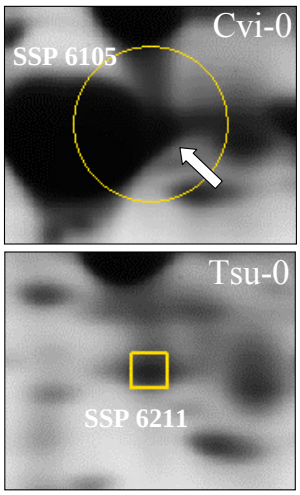


Figure 43 Example of an ambiguous spot. Spot SSP 6211 may be underneath (white arrow) the high-abundant spot SSP6105 in the 2DE gel of ecotype Cvi-0. Enlarged gel region A from Figure 42.

The total number of spots found in the twelve ecotypes was 931. Among these 931 spots, 334 spots (36%; blue bar, Figure 44) were present in all the ecotypes and 597 spots (64%) were variable, i.e., absent in at least one ecotype. Almost one-third (32%) of the spots were not detected in half the ecotypes. Twenty-seven per cent of all spots appeared to be either specifically present or absent for one ecotype. Figure 44 shows the entire spot distribution according to the number of ecotypes where they were detected.

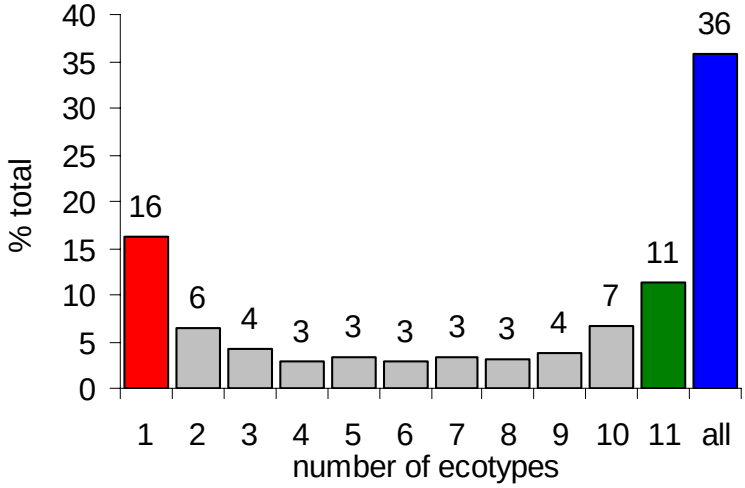


Figure 44 Distribution of 931 distinct spots detected among the 12 ecotypes according to the number of ecotypes where they were reproducible detected.

Each ecotype-specific protein spot (red and green bares, Figure 44) was visually inspected. Of the 150 spots (16%) present in only one ecotype (ecotype-specific), 33 spots were ambiguously absent in at least one other ecotype, i.e. the spot may be underneath another spot

or detected in less than three replicate gels due to edge or focusing effects and, therefore, not included in the data set. Of the 106 spots (11%) absent in one ecotype (ecotype-specific absent), 38 spots were ambiguous. Figure 45 shows the distribution of the ecotype-specific present spots (A) and absent spots (B) according to the ecotype in which they were detected. Examples for ecotype-specific present or absent protein spots are shown in Figure 45 C and D. The ecotypes Cvi-0, Mr-0, Condara and C24 accounted for 75% of these specifically present spots (Figure 45 A) and 74% of the specifically absent spots (Figure 45 B). For Ma-0 or Mt-0, no uniquely present or absent spots were identified.

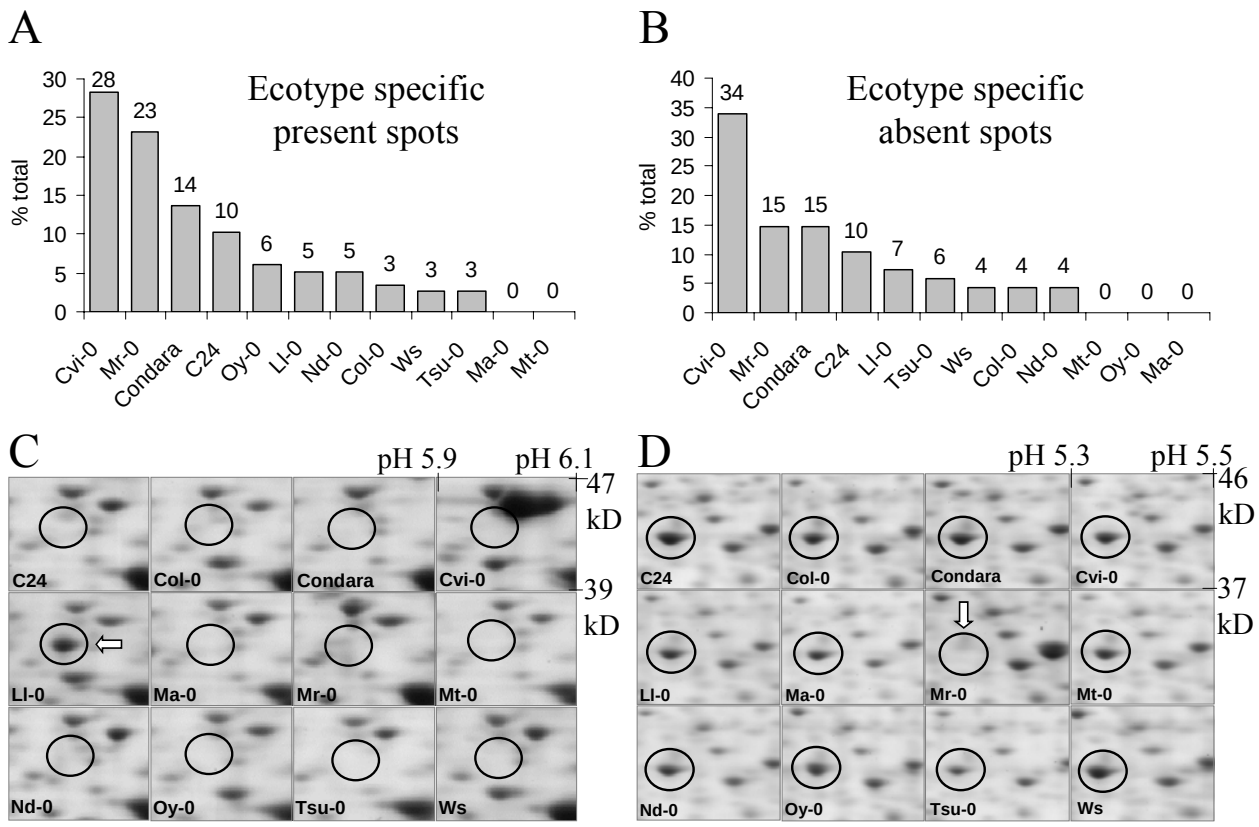


Figure 45 Ecotype-specific spots
A and B: Distribution of ecotype-specific present spots (A) and absent spots (B) according to the ecotypes.
C: Example of an ecotype-specific present protein spot (C). Enlarged gel region B from Figure 42.
D: Example of an ecotype-specific absent protein spot (D). Enlarged gel region C from Figure 42.

At this time, very little information is available about the variations in proteomes among *Arabidopsis* ecotypes. More information is available for crop plant species, such as barley [228, 345-347], maize [348-351], wheat [352-356], peanut [357], and rice [358]. All these studies demonstrate large variability in the proteomes of the studied varieties. Chevalier et al. (2004) [359] were the first group to assess the natural variation in the proteome of

Arabidopsis ecotypes. They investigated the natural variation in the root proteome among eight *Arabidopsis* ecotypes and resolved an average of 250 spots for each ecotype. The variability of the root proteomes is similar to the variability of the seed proteomes in this study as seen in Table 30. Although two different plant tissues were studied, the numbers such as variable spots and specific spots are very similar with 75% and 26% for the root proteome and 64% and 26% for the seed proteome.

Table 30 Comparison of published natural variability in the root proteome of eight *Arabidopsis* ecotypes [359] with the natural variability in the seed proteome of 12 *Arabidopsis* ecotypes found in this study.

	Natural variability of *Arabidopsis* root proteome [359]	seed proteome
Spots detected in all ecotypes	25%	36%
Variable spots	75%	64%
Specific present spots	10%	15% (13%*)
Specific absent spots	16%	11% (7%*)

* without ambiguous spots

The nature of the protein pattern variability was not investigated here and could be hypothesized to rely on post-translational modifications and on allelic variations for proteins identifying the same ecotype. Upon inspection of the ecotype-specific spots, apparent position shifts (PS) of proteins were observed. As shown in Figure 46, a comparison of Col-0 and C24 gels revealed the pI shift of many proteins.

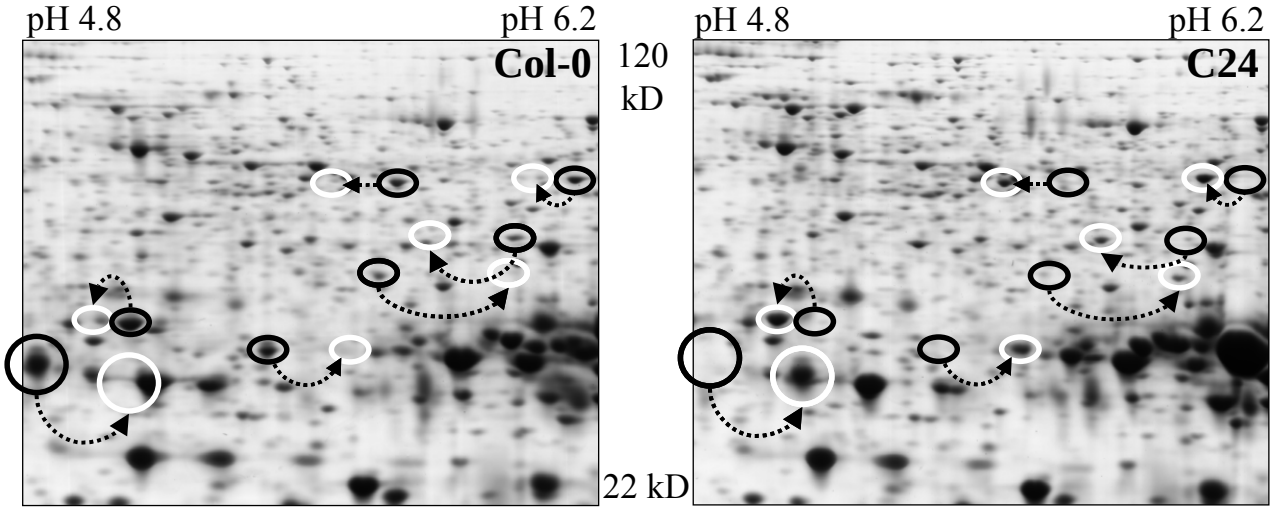


Figure 46 Examples of possible protein shifts of proteins. Enlarged gel region D from Figure 42.

These shifts are in both directions and range from small (pI difference < 0.1) to large (pI difference 0.6). In these examples (Figure 46), Mw shifts were not seen. Position shifts or allelic variations of a protein can be suspected when two spots, differing by their pI and/or molecular mass, are mutually exclusive in different ecotypes. Such allelic variation of a protein may be the result of a point mutation, frameshift, deletion, addition, or post-translational modification (PTM) as discussed in the introduction of this chapter. For example, Jungblut et al. [365] compared the proteomes of various *Helicobacter pylori* strains and demonstrated that a single amino acid change caused a change of pI of 0.05 units, which resulted in a clearly detectable shift in the 2DE pattern. Finnie et al. [347, 366] compared the protein patterns of a series of barley cultivars with different malting properties and demonstrated by MS that a single amino acid substitution is sufficient to explain the 0.1 pH-unit difference between two β-amylase spots. Also, Schlesier et al. [367] demonstrated that a single amino acid substitution resulted in a 0.45 pI shift of a germin-like protein found in the leaf proteome of two *Arabidopsis* ecotypes (Col-0 vs. Ws-2). This finding highlights the genetic basis for proteome differences and the power of 2DE to detect such differences. Anderson et al. [368] analyzed various wheat lines and found that charge modifications often occur in the major storage proteins of wheat and that mass modifications occur less frequently. Also in this study, more apparent horizontal position shifts were observed than apparent vertical position shifts. A possible explanation for this observation may be that either mutations leading to charge changes are more frequent or that horizontal position shifts are easier to detect than vertical position shifts due to a higher resolution power in this direction.

4.3.2.2.2 Quantitative comparison of 2DE patterns

In order to assess the natural variability of spot quantities, only proteins expressed in all ecotypes were considered. Figure 47 shows two examples for the natural variability in spot quantity. The spot quantity of spot SSP 4103 varies 5-fold (Condara vs. Nd-0) among the 12 ecotypes. More profound is the difference between the ecotype Ws and Col-0 for the spot SSP 8105 with 20-fold variation in spot quantity.

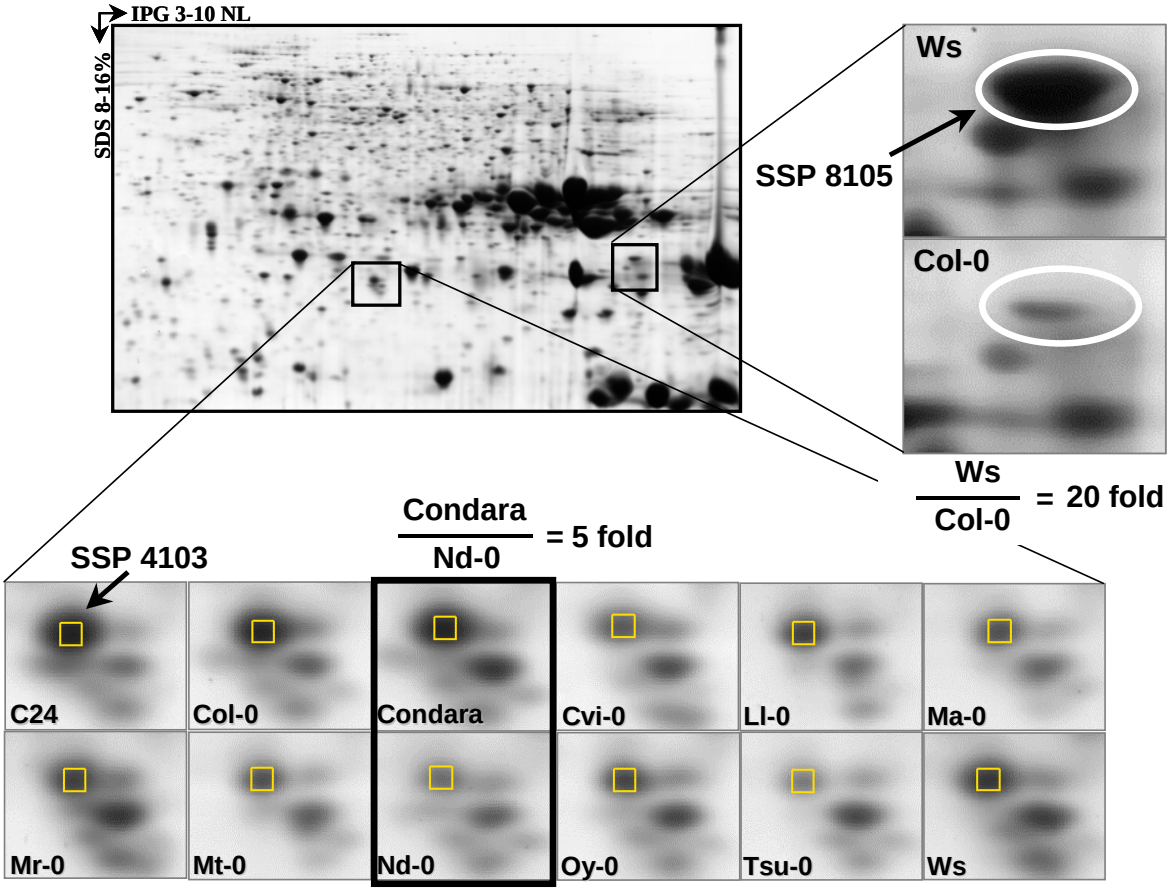


Figure 47 Example for natural variability in spot quantities.

Upon closer inspection of spots with very high ratios, it became apparent that spots affected by background, horizontal and vertical streaking, edge effects, and/or spots that are overlapped by neighboring spots are inaccurately quantified (Figure 48). As seen in Figure 48, the 25-fold difference calculated for spot SSP 4717 is misleading as streaking artificially increased the IOD for the C24 spot. This spot was excluded from the data set.

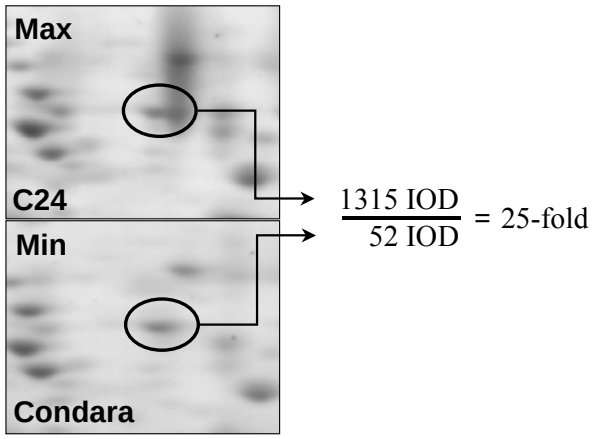


Figure 48 Ambiguous spot quantification: spot quantification is negatively affected by background staining.

The 2DE analysis software package PDQuest assigns a spot quality value to each spot ranging from 0 (very bad) to 100 (very good) that is calculated based on Gaussian fit, horizontal streaking, vertical streaking, overlapping, and linear range of scanner. Another indicator for ambiguous spot quantification is the coefficient of variance (CV). In order to avoid overestimation of natural variation, only spots with CVs below 55 % (maximal CV to be able to detect a 3-fold difference – see Chapter 4.2.2) and a preset spot-quality to compensate for streaking, overlap, etc, were taken into account. Four of the 334 spots had CVs and/or spot qualities below the threshold and were excluded from the data set. The range (ratio of the highest to the lowest spot quantity) computed for the remaining 330 protein spots is shown in Figure 49.

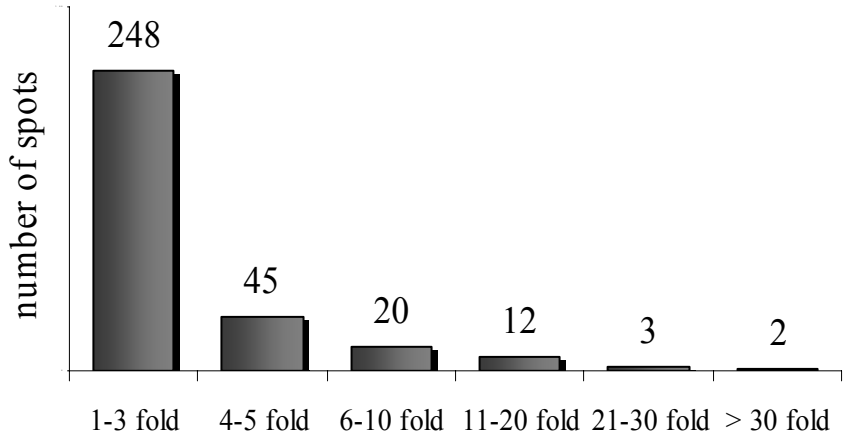


Figure 49 Natural variation of protein spots quantities among the 12 ecotypes considering the 330 spots detected in all ecotypes.

Spot quantities varied from 1- to 53-fold, with ranges higher than 3-fold for about 25 % of the spots. There is no data published regarding natural variability of protein abundance for *Arabidopsis*. However, Bustin et al. [349], analyzing 21 maize (*Zea mays* L.) inbred lines by 2DE, found the ratio of the highest to the lowest intensity in the 21 lines ranged from 1.4 to 26 for the 190 quantified spots. This range is similar to the range found in this study, even though Bustin et al. used silver staining, which may not allow for large dynamic ranges.

4.3.2.2.3 Quantification of the natural variation

The distribution charts of the ecotype-specific protein spots (Figure 45 A and B) suggest that four ecotypes (Cvi-0, Mr-0, Condara, and C24) have the most unique protein profiles compared to the other ecotypes but do not allow a conclusion about the overall relations between the ecotypes. In order to quantify and visualize the relationship (calculate the

distance) between the ecotypes, a phenetic tree was constructed according to Marques et al. [292] based on the pairwise comparison of the qualitative (presence / absence of spot) protein profiles of the ecotypes. These 66 pairwise comparisons were done by counting the number of spots present in both ecotypes (N_{AB}) and specifically present in one (N_{A0}) or the other (N_{0B}) of the two considered ecotypes (Table 31). The Jaccard or Dissimilarity index (Equation 4.3) was used to compute a dissimilarity matrix (Table 32). From this dissimilarity matrix, an unrooted phenetic tree (Figure 50) was calculated with the Neighbor Joining algorithm using the Phylip 3.6 software package [293]. An unrooted phenetic tree specifies the relationships among ecotypes and does not define the evolutionary path. One has to keep in mind that because two different proteins theoretically could migrate to the same position in a 2DE gel, the genetic dissimilarity between two genotypes is systematically underestimated [369].

Table 31 Spreadsheet for the Jaccard index calculation (for the entire comparison table, see Appendix 1)

Ecotype A	Ecotype B	Present only in ecotype A (N_{A0})	Present only in ecotype B (N_{0B})	Present in ecotype A and B (N_{AB})	Jaccard index
Ma-0	Mt-0	21	9	565	0.05
Oy-0	Ws	49	53	590	0.15
Ll-0	Ma-0	67	59	527	0.19
Nd-0	Oy-0	59	69	570	0.18
Oy-0	Tsu-0	96	34	543	0.19
Ll-0	Mt-0	78	58	516	0.21
:	:	:	:	:	:
Col-0	Cvi-0	126	120	481	0.34
Cvi-0	Ws	106	148	495	0.34
Cvi-0	Mr-0	114	143	487	0.35
Condara	Cvi-0	158	106	495	0.35
C24	Cvi-0	160	111	490	0.36

Jaccard index: $Dj = 1 - N_{AB} / (N_{AB} + N_{A0} + N_{0B})$ (Equation 4.3)

Table 32 Dissimilarity matrix between the ecotypes based on spot presence/absence

	C24	Col-0	Condara	Cvi-0	Ll-0	***Ma-0***	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0
Col-0	0.252	0									
Condara	0.242	0.228	0								
Cvi-0	***0.356***	0.338	0.348	0							
Ll-0	0.255	0.231	0.270	0.322	0						
Ma-0	0.266	0.233	0.279	0.338	0.193	0					
Mr-0	0.305	0.282	0.310	0.345	0.293	0.309	0				
Mt-0	0.286	0.245	0.289	0.340	0.209	***0.050***	0.321	0			
Nd-0	0.231	0.211	0.212	0.299	0.230	0.226	0.299	0.241	0		
Oy-0	0.234	0.217	0.230	0.318	0.218	0.240	0.294	0.242	0.183	0	
Tsu-0	0.269	0.243	0.265	0.317	0.247	0.243	0.317	0.232	0.213	0.193	0
Ws	0.231	0.244	0.217	0.339	0.248	0.254	0.312	0.264	0.211	0.147	0.242

Extreme values are highlighted (italic and bold)

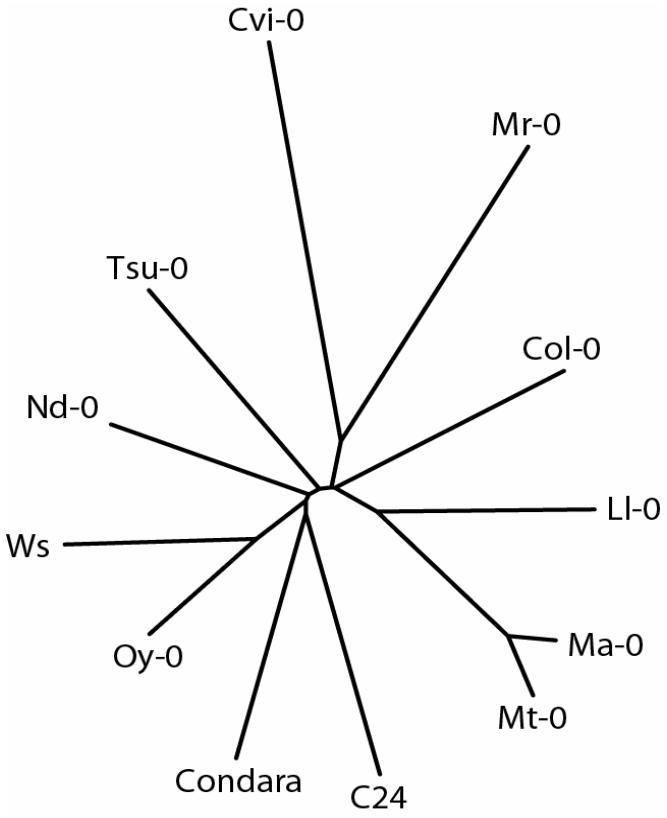


Figure 50 Unrooted phenetic tree built from the distance matrix calculated according to the Jaccard index on all the spots of the 12 ecotypes using the Neighbor-Joining algorithm.

The length of the branch (Figure 50) is proportional to the number of differences in the seed proteomes. The distance between two ecotypes is the sum of the length of all branches connecting them. The greatest distance was found between Cvi-0 and C24 with a total of 271

different spots and a Jaccard or Dissimilarity index of 0.356. Ma-0 and Mt-0 appeared to be very closely related with 30 different spots and a Jaccard index of 0.050. The next group of closely related ecotypes is Ws and Oy-0 with 102 different spots and a Jaccard index of 0.147. This type of analysis agrees with data for the two very closely related ecotypes Ma-0 and Mt-0 that had displayed very similar phenotypes (4.3.2.1 Comparison of phenotypic traits). Erschadi et al. [362], using 15 AFLP primer combinations, grouped 20 *Arabidopsis* ecotypes and Ma-0 and Mt-0 were also clustered into one group. Kliebenstein et al. [370] analyzed the glucosinolate profiles in the leaves and seeds of 39 *Arabidopsis* ecotypes and demonstrated extensive differences. However, Ma-0 and Mt-0 showed a very similar glucosinolate profile. Therefore, the phenetic tree built with the proteome data generated in these experiments provides ecotype relationships similar to AFLP primer combinations and metabolic profiles regarding the ecotypes Ma-0 and Mt-0. Due to differences in selected ecotypes and methods of data analysis, it is difficult to compare the present data with population genetics publications based on RFLP [361], AFLP [360, 362, 363], and CAPS [364]. However, it is known that *Arabidopsis* ecotypes are not easily grouped, i.e. do not conform to a bifurcating pattern of evolution, and there is no "ecotype phylogeny" [360]. The analysis of genetic relationship by amplified fragment length polymorphism (AFLP) among ecotypes revealed a star or bush-like dendrogram [360, 363]. The phenetic tree based on variation in the protein patterns illustrates the large genetic variability among the 12 ecotypes. The fact that most of the ecotype branches rise from the center of the phenetic tree suggests that the selected 12 ecotypes cover a large natural variability.

4.3.3 Summary

The seed protein expression patterns from individual plants grown in the same growth chamber or grown in two different growth chambers were similar. However, one protein spot was only detected in the seed protein profiles of plants grown in growth chamber GC-2. Differences in presence and absence of spots were not observed between the protein patterns of six individual plants grown in the same growth chamber. Overall, more differences were seen in spot quantity rather than in the presence/absence of new spots. For example, out of the 702 reproducible spots of the different growth chamber study, 52 spots were found to be statistically significantly different (P<0.05) in their quantity by two or more fold. In this study, it was also demonstrated that storage may have an impact on the seed protein profile as seen in the increase of spot SSP 6111 over 13 months. The nature of these differences was

not investigated here and could have been caused by differences in the lighting system and/or other environmental differences resulting from the different growth chambers. However, the data indicate the potential of environmental and storage impacts on the seed proteome even without extreme environmental conditions. The data also demonstrate the sensitivity of the optimized 2DE method to detect differences in seed proteomes due to small differences in the environmental conditions.

It was shown that the pooled seed samples represent the average seed proteome of the six individual seed proteomes. Therefore, by pooling the seeds from individual plants of a line, it is possible to obtain a characteristic overall protein profile for this ecotype, encompassing generation and environmental factors within a growth chamber.

Qualitative and quantitative differences in seed protein profiles of various *A. thaliana* ecotypes have been easily detected by the optimized 2DE method. It has been clearly shown that the natural variability of seed protein profiles due to genetic backgrounds is extensive among the 12 selected *A. thaliana* ecotypes. Almost half of the resolved spots (on average 615 reproducible spots) varied with respect to their presence/absence. In terms of quantitative differences, 25% of the common spots varied in quantity from 3 to 53 fold. The nature of the protein pattern variability was not investigated here and could be hypothesized to rely on post-translational modifications and on allelic variations for proteins identifying the same ecotype.

The distances between the ecotypes were quantified and visualized based on the protein profile (presence/absence of spots). The large impact of the genetic background on the protein profiles of *Arabidopsis* ecotypes was visualized by a phenetic tree. Most of the branches rise from the center of the phenetic tree, indicating that the chosen set represents a broad range of genetic variation. The star-like relationship between *Arabidopsis* ecotypes was also seen by AFLP-based dendrograms [360, 363]. The impact of the genetic background is most likely underestimated due to the possibility that two different proteins may share identical isoelectric points and identical molecular weights and thus migrate to exactly the same location in 2D gels.

The data compiled for the 931 reproducible spots may serve as a baseline for the head-to-head comparison of transgenic versus non-transgenic *Arabidopsis* lines in order to assess differences in the context of natural variability.

4.4 Genetically modified lines

The introduction of exogenous DNA sequences into the plant genome is a random process leading to physical disruption in the genome, and possible inactivation of endogenous genes. Activation of silent genes and the formation of fusion proteins by transcriptional read-through processes are also possible [95]. In addition, the introduced gene(s), the gene product, or the changed biochemical pathway may interact with the regulation of other genes or biochemical pathways. Since proteins are direct products of gene transcription and translation and involved in controlling many biochemical pathways, they are ideally suited for the detection of changes in the genome (e.g. insertional mutation), in gene regulation (pleiotropic effect), or changes in biochemical pathways (pleiotropic effect) of a genetically modified plant. The expected changes of the plant proteome are similar to the naturally occurring mutations described in Chapter 4.3.

In order to investigate insertional and pleiotropic effects due to genetic engineering, the seed proteomes of twelve transgenic (TG) *A. thaliana* lines were analyzed by 2DE and compared to the seed proteome of the wild type. The transgenic *A. thaliana* lines contain an inserted beta-glucuronidase (*gus*) gene, an inserted p-hydroxyphenylpyruvate dioxygenase (*hppd*) gene, or an inserted γ-tocopherol-methyltransferase (*gtmt*) gene. These lines were chosen because they represent two different strategies: (i) no change to a metabolic pathway (*gus* gene) and (ii) change of an endogenous metabolic pathway (*hppd* and *gtmt* gene). The data generated from the natural variability study (Chapter 4.3) was used to discuss the relevance of potential differences between the transgenic lines and their non-transgenic parental lines.

4.4.1 GUS-lines

The T3 generation of six transgenic *Arabidopsis thaliana* lines was grown side-by-side with the parental line (Col-0) under controlled environmental conditions. Twelve replicates were planted for each line and randomly distributed in the growth chamber in order to limit the influence of environmental factors. The transgenic *Arabidopsis* events were obtained by *Agrobacterium*-mediated transformation with a T-DNA containing the *β*-glucuronidase (*gus*) gene as well as the neomycin phosphotransferase II (*nptII*) gene for antibiotic selection (Figure 51). The GUS construct contains the enhanced 35S promoter, derived from cauliflower mosaic virus (E35S), and a transcription terminator sequence (E9). *β*-Glucuronidase is a hydrolase that catalyzes the cleavage of a wide variety of *β*-glucuronides and has not been found in any plant species [371, 372].

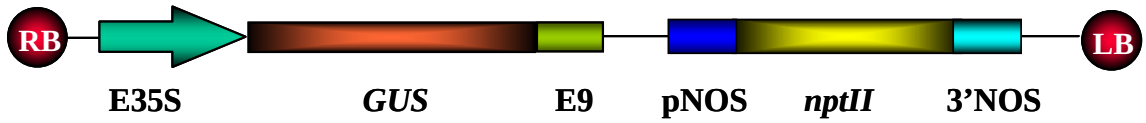


Figure 51 DNA cassette for the transgenic *Arabidopsis* lines. RB and LB marks the site of T-DNA integration into plant DNA.

The *nptII* gene is driven by a nopaline synthase promoter (pNOS) and is followed by the 3'non-translated region of the nopaline synthase gene from *Agrobacterium tumefaciens* (3'NOS). The *nptII* gene encodes the selectable marker enzyme neomycin phosphotransferase, which inactivates, by phosphorylation, a range of aminoglycoside antibiotics such as kanamycin, neomycin, geneticin, and paromomycin [373]. Kanamycin is very effective in inhibiting the growth of untransformed cells and therefore, frequently used as a selective agent. The *gus* gene and the *nptII* gene are prokaryotic genes and not normally found in plant tissues [371, 372, 374, 375]. A physiological effect of either of the introduced genes is therefore not anticipated.

The selected transgenic lines are homozygous for the transgene (*gus*), contain one copy of the gene, and have different transgene expression rates in leaf tissues (Figure 52). The different levels of transgene expression may be due to differences in the integration site of the introduced *gus* gene, referred to as position effect [376-378]. The expression rates were assayed by RT-PCR (reverse transcription-polymerase chain reaction) to determine the amount of E9-mRNA in relation to a transgenic control line with known expression rate (Figure 52). The relative levels of *GUS* expression ranged from 1.8 to 11.9 times of the levels in the control line with the highest levels seen in TG-6.

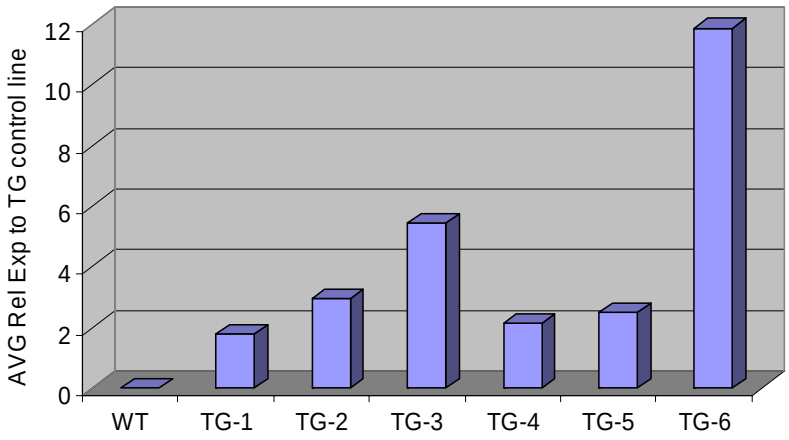


Figure 52 GUS expression measured in leaf tissues of T2 generation plants. Data expressed as average relative expression of E9-mRNA to a transgenic control line with known expression rate.

The lines were provided and characterized by Jon J. Schmuke (Monsanto Company, St. Louis, MO, USA).

4.4.1.1 Phenotypic comparison

Four phenotypic traits in the transgenic and wild-type *Arabidopsis* plants were assayed; the data are summarized in Table 33. The first flowering date (FFD, measured at Boyes's growth stage 6.00 [340]), seed yield, and protein content do not show significant differences (P>0.05) between WT Col-0 and the transgenic lines. A statistically significant difference (P<0.05) in rosette leaf diameter (RD) was observed between WT Col-0 and all transgenic lines except TG-2 and TG-3. The leaf rosettes of the transgenic lines TG-1, TG-4, TG-5, and TG-6 were on average 24% larger than the leaf rosettes of the WT. The line TG-6 had the largest RD (9.5 cm) compared to WT with a 7.4 cm RD. However, the differences in RD are in the range of natural variability (7.4 cm to 10.8 cm) observed when plants of WT Col-0 were grown in two different growth chambers (Chapter 4.3.1 Table 24).

Table 33 Phenotypic measurements of the WT Col-0 and the TG lines.

Name	N	FFD[1)] (days)	RD [2)] (cm)	Seed yield[3)] (mg)	Protein[4)] (%)
WT Col-0	8	31 ± 1	7.4 ± 1.0	722 ± 92	24.7
TG-1	10	31 ± 1	9.1 ± 0.8	711 ± 220	23.4
TG-2	10	31 ± 4	8.4 ± 0.5	641 ± 184	23.9
TG-3	12	32 ± 4	7.7 ± 0.8	583 ± 131	24.3
TG-4	11	31 ± 4	8.7 ± 0.9	593 ± 200	23.2
TG-5	11	31 ± 2	9.1 ± 0.7	596 ± 136	24.0
TG-6	12	32 ± 5	9.5 ± 1.4	540 ± 184	25.4

Values are means ± SD

[1)] FFD = number of days from the date of planting until the opening of the first flower
[2)] RD = rosette leaf diameter at the time of first flowering
[3)] Amount of harvested seeds
[4)] Protein content of pooled seed samples of six to eight plants; N=2

In addition, all the lines demonstrated very similar phenotypes with respect to their leaf and stem morphology. The leaf and stem morphology was visually assessed by overall shape, length, thickness, and pubescence.

4.4.1.2 Comparison of 2DE patterns

Seeds harvested from eight individual replicate plants were pooled to one representative ecotype sample in order to minimize the influence of environmental and genetic variation within a line. The pooled seed samples were extracted according to the optimized method and 2DE was performed in triplicate for each extract according to the optimized 2DE protocol. The software package PDQuest was used to analyze the gel images. Spots detected and matched by the software program were manually verified. False positive and false negative spots were removed and added, respectively.

4.4.1.2.1 Qualitative comparison of 2DE patterns

Using the optimized 2DE technique, the seed protein profiles of the six transgenic *Arabidopsis* lines were compared to the seed protein profile of the parental wild-type line Columbia (Col-0). Results of the qualitative comparisons are summarized in Table 34.

Table 34 Qualitative seed proteome comparison between WT line and TG lines

Line	Detected spots[1)]	Reproducible spots[2)]	Spots matched to WT Col-0	Spots reproducibly absent[3)]
WT Col-0	463 ± 10	429	429	1 (SSP 2814)
TG-1	472 ± 30	440	420	0
TG-2	465 ± 15	434	417	0
TG-3	472 ± 9	444	415	1 (SSP 2814)
TG-4	477 ± 6	447	421	0
TG-5	458 ± 13	426	416	0
TG-6	470 ± 19	442	419	0

[1)] false positive spots due to edge effects, focusing problems, or dust were excluded; Values are means ± SD
[2)] spot must be detected in all three replicate gels
[3)] spot must be absent in all three replicate gels

The number of detected spots varied from 458±13 protein spots for the line TG-5 to 477±6 protein spots for the line TG-4. The number of detected spots for the WT Col-0 fell into this range with 462±10 resolved spots. The majority of the spots found in all the lines were reproducible spots (resolved in all three replicate gels) and varied between 426 spots for TG-5 and 447 spots for TG-4. The protein profiles of the seven lines were very similar; 415 spots (TG-3) to 421 spots (TG-4) were matched to the wild-type Col-0 and 414 spots were matched to all lines. One reproducible spot (SSP 2814) resolved for the TG lines TG-1, TG-2, TG-4, TG-5, and TG-6 was consistently absent in the 2DE gels of the TG line TG-3 and the WT

line Col-0. Figure 53 shows representative 2DE gels for WT Col-0 and for the line TG-2 as an example of the head-to-head comparison.

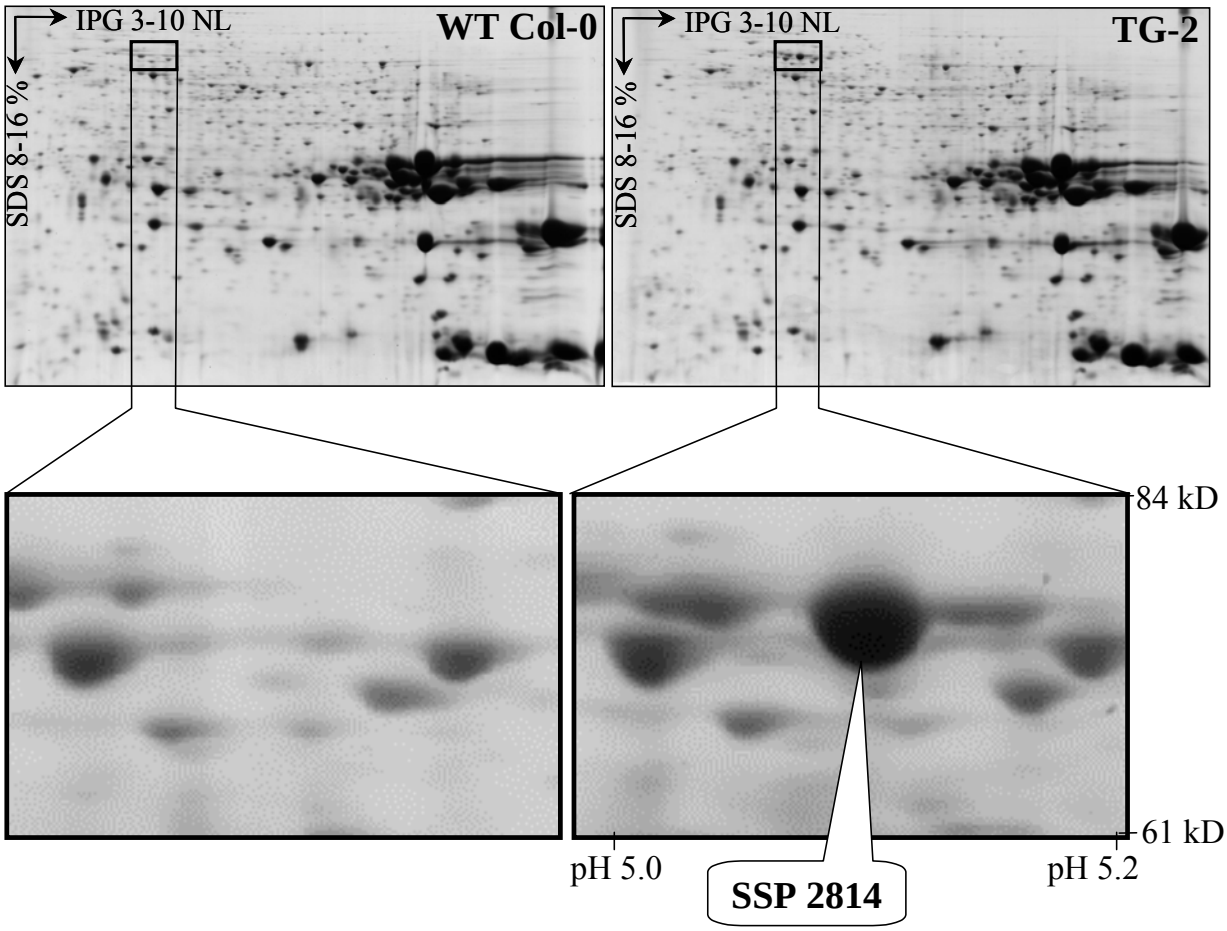


Figure 53 Head-to-head comparison between seed protein profiles of WT Col-0 and TG lines. Representative 2DE gels of WT Col-0 and TG line TG-2 are shown with enlargements of the gel regions of interest.

The gel region indicating the difference between the two lines is enlarged for both profiles. Protein spot SSP 2814 is absent in the 2DE pattern of WT Col-0 and present in the 2DE pattern of TG line TG-2. All other reproducibly present spots of one of the lines were present in at least one replicate gel of all other samples, i.e. no further reproducibly absent spot was detected for any of the lines.

4.4.1.2.2 Identification of protein spot SSP 2814

Protein spot SSP 2814 was excised and digested in-gel with trypsin. Trypsin cleaves proteins specifically on the carboxylic acid side of arginine and lysine residues. The tryptic peptides were eluted from the gel piece and analyzed by MALDI-TOF MS. The measured peptide

masses were searched against the NCBInr database using MS-Fit [288, 289] and against the Swiss-Prot/TrEMBL databases using PeptIdent [290, 291]. In order to consider a protein unambiguously identified, at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered [379]. The two programs (MS-Fit and PeptIdent) identified the protein spot as β-glucuronidase (GUS). Twenty-three peptides matched with the theoretically expected tryptic peptide masses for GUS (Table 35), covering 36% of the amino acid sequence of GUS (Figure 54).

Table 35 Peptide masses matched to β–glucuronidase (EC 3.2.1.31) from Escherichia coli by PeptIdent [291].

Mass measured	Mass calculated	Delta mass (Da)	Position	Sequence	Modification
500.4237	500.3555	-0.0681	84 - 87	IVLR	
536.3608	536.2827	-0.078	536 - 539	VFDR	
593.4262	593.3406	-0.0855	569 - 573	GIFTR	
601.4538	601.3668	-0.0869	394 - 398	ELIAR	
674.4454	674.3369	-0.1084	78 - 83	GWAGQR	
855.5422	855.3955	-0.1466	296 - 302	HEDADLR	
864.6277	864.4686	-0.159	569 - 575	GIFTRDR	
877.6509	877.5142	-0.1366	579 - 586	SAAFLLQK	
991.6019	991.4262	-0.1756	26 - 33	ENCGIDQR	Cys_CAM: 28
1016.702	1016.541	-0.1608	488 - 495	ELLAWQEK	
1037.673	1037.505	-0.168	88 - 96	FDAVTHYGK	
1267.847	1267.632	-0.2155	423 - 433	EYFAPLAEATR	
1270.815	1270.596	-0.2187	62 - 71	NYAGNVWYQR	
1291.825	1291.607	-0.2181	34 - 43	WWESALQESR	
1381.88	1381.642	-0.2387	258 - 268	SQTECDIYPLR	Cys_CAM: 262
1395.897	1395.6573	-0.2396	258 - 268	SQTECDIYPLR	Cys_PAM: 262
1395.897	1395.727	-0.1702	423 - 434	EYFAPLAEATRK	
1665.037	1664.785	-0.2523	588 - 602	WTGMNFGEKPQQGGK	
1902.177	1901.872	-0.3052	158 - 172	QSYFHDFFNYAGIHR	
1907.24	1906.929	-0.3104	44 - 61	AIAVPGSFNDQFADADIR	
2158.427	2158.087	-0.3403	278 - 295	GEQFLINHKPFYFTGFGR	
2358.549	2358.157	-0.3919	373 - 393	ELYSEEAVNGETQQAHLQAIK	
2478.604	2478.194	-0.4094	401 - 422	NHPSVVMWSIANEPDTRPQGAR	MSO: 407

```
  1 MVRPVETPTR EIKKLDGLWA FSLDRENCGI DQRWWESALQ ESRAIAVPGS FNDQFADADI
 61 RNYAGNVWYQ REVFIPKGWA GQRIVLRFDA VTHYGKVWVN NQEVMEHQGG YTPFEADVTP
121 YVIAGKSVRI TVCVNNELNW QTIPPGMVIT DENGKKKQSY FHDFFNYAGI HRSVMLYTTP
181 NTWVDDITVV THVAQDCNHA SVDWQVVANG DVSVELRDAD QQVVATGQGT SGTLQVVNPH
241 LWQPGEGYLY ELCVTAKSQT ECDIYPLRVG IRSVAVKGEQ FLINHKPFYF TGFGRHEDAD
301 LRGKGFDNVL MVHDHALMDW IGANSYRTSH YPYAEEMLDW ADEHGIVVID ETAAVGFNLS
361 LGIGFEAGNK PKELYSEEAV NGETQQAHLQ AIKELIARDK NHPSVVMWSI ANEPDTRPQG
421 AREYFAPLAE ATRKLDPTRP ITCVNVMFCD AHTDTISDLF DVLCLNRYYG WYVQSGDLET
481 AEKVLEKELL AWQEKLHQPI IITEYGVDTL AGLHSMYTDM WSEEYQCAWL DMYHRVFDRV
541 SAVVGEQVWN FADFATSQGI LRVGGNKKGI FTRDRKPKSA AFLLQKRWTG MNFGEKPQQG
601 GKQGSSDYKD DDDK
```

Figure 54 Coverage map for GUS. The amino acid sequence was deduced from the transgenic DNA sequence. Matched peptides appear in red.

The protein GUS has a theoretical isoelectric point (pI) of 5.13 and a theoretical molecular mass (Mw) of 70 kDa. The experimental pI and Mw were found to be 5.1 and 73 kDa, respectively. Hence, the protein spot SSP 2814 was identified as GUS according to its peptide mass fingerprint (PMF), pI, and Mw.

4.4.1.3 Quantitative comparison

4.4.1.3.1 Quantitative comparison of the novel protein

The transgenic protein GUS was detected in the 2DE gels of TG-1, TG-2, TG-4, TG-5, and TG-6 but not in WT Col-0 and TG-3 (Figure 55). The spot quantity of GUS varied among the transgenic lines, where GUS was reproducibly detected, and increased in the following order: TG-1 (<0.01% of total IOD), TG-5 (0.10% of total IOD), TG-4 (0.18% of total IOD), TG-6 (0.37% of total IOD), and TG-2 (0.44% of total IOD).

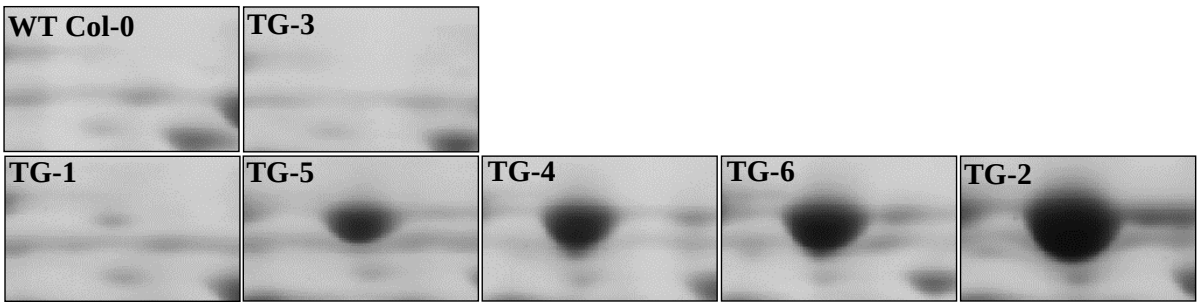


Figure 55 Differences in GUS expression among the six transgenic lines compared to WT.

Although the line TG-3 had the second highest expression of *gus* (E9-mRNA) in leaves of T2 generation plants (Figure 52), neither the GUS protein (Figure 55) nor the *gus* transcript (E9-mRNA) (Figure 56) were detected in the seed of the T3 generation of this line. The 2DE analysis of the progenitor (T2 generation) seeds confirmed the absence of GUS on the 2DE gels. These findings are surprising because the *gus* gene is driven by the CaMV 35S promoter that generally is considered to be a constitutive promoter [380]; this can be seen by the presence of GUS in leaf and seed tissue of the other transgenic lines (TG-1, -2, -4, and –5). A possible explanation is the difference in the insertion site that may have an impact on the expression of GUS in the seeds of the TG-3 line. It is also noteworthy that there is no linear correlation between the GUS mRNA in leaves or seeds and the actual GUS protein in seeds. The poor correlation between mRNA and protein has been demonstrated [127; in

yeast: 128-131] and is thought to be due to different turnover rates of mRNA and protein, alternative splicing, and post-translational modification.

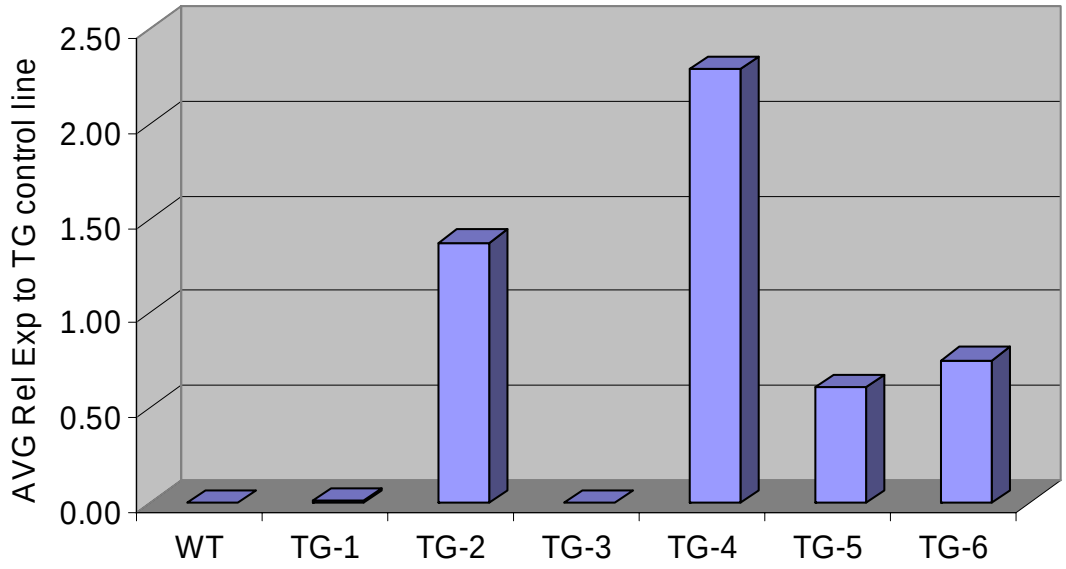


Figure 56 GUS expression measured in seed tissues of T3 generation plants. Data expressed as average relative expression of E9-mRNA to a transgenic control line with known expression rate.

4.4.1.3.2 Quantitative comparison of 2DE patterns

In order to investigate differences in spot quantities between the wild-type and the transgenic lines, the quantities of all spots reproducibly present in WT Col-0 and matched to the transgenic lines were examined. A statistical significance of 2-fold and more quantity differences was evaluated with the statistical software program JMP using the Tukey-Kramer simultaneous pairwise comparison procedure. This test protects the significance level ($\alpha = 0.05$) for all combinations of pairs. To avoid overestimation of quantitative differences due to inaccurate quantification of poorly resolved protein spots, all spots with the maximum value and an average spot quantity less than or equal to 40 were excluded from the data set. In addition, the detected statistically significant difference ($P<0.05$) must have a power above 80%, as described in Chapter 4.3. Table 36 summarizes the quantitative comparison of the seed proteomes of WT Col-0 and the six transgenic lines.

Table 36 Quantitative comparison of the seed proteomes of WT Col-0 and the six TG lines

Line vs. WT	Compared spots	Spots quantitatively different[1)]
TG-1	420	0
TG-2	417	3 (SSP 2807 6x↑; SSP 3803 4x↑; SSP 6409 2x↓)
TG-3	415	2 (SSP 2406 2x↑; SSP 6409 2x↓)
TG-4	421	3 (SSP 3210 2x↑; SSP 6409 2x↓; SSP 8103 3x↓)
TG-5	416	1 (SSP 6409 2x↓)
TG-6	419	4 (SSP 2807 6x↑; SSP 3803 4x↑; SSP 6409 2x↓; SSP 8103 2x↓)

[1)] Power of difference must be > 80% and mean spot quality > 40

Between 415 spots (WT vs. TG-3) and 421 spots (WT vs. TG-4) were compared. Less than 1% of the spots varied significantly (P<0.05) in spot quantity and the proteome comparison of WT and TG-1 did not reveal any differences in spot quantity. The transgenic line TG-6 had, with four spots, the highest number of spots significantly different in spot quantity. The differences in spot quantity varied between 2 to 6-fold. In order to evaluate any biological significance of those differences, the spot quantities of these spots were set in context with analytical (run-to-run), environmental (GC1 vs. GC2), and genetic variation (Table 37). Therefore, the data from the run-to-run comparison (Table 27), the two different growth chambers comparison, and the ecotypes comparison were subsequentially included into the data set for the significance test. Table 37 shows the ranges of spot quantity with respect to the type of variation, the spot quantity of the WT Col-0, and the significantly different transgenic lines.

Table 37 Quantities of the significantly different spots in the context of analytical and natural variation

SSP	Run-to-run	Environment GC1 vs. GC2	12 Ecotypes	WT Col-0	***TG-2***	TG-3	TG-4	TG-5	TG-6
2406	259 - 400	380 - 474	72 - 503	324		589			
2807	57 - 141	122 - 138	105 - 743	169	***1059***				823
3210	1436 - 2126	1557 - 1964	353 - 3768	1770			2741		
3803	11 - 97	63 - 67	41 - 202	83	***356***				304
6409	122 - 167	161	114 - 276	180	103	82	76	93	86
8103	361 - 796	374 - 480	117 - 2118	578			194		236

Values are the mean (N=3) normalized IOD of the spots.

Four of the thirteen significantly different spots fall in the range of analytical and/or natural variation. Eleven of the thirteen different spots are not significantly different to the extreme values found for natural variation. Figure 57 shows enlargements of 2DE gel regions containing the protein spot SSP 8103 as an example of the natural variation of spot quantity among 12 *Arabidopsis* ecotypes, the six transgenic lines, and the WT Col-0. The spot quantity varies from absent (Cvi-0) to 2118 IOD for WS. The lowest detected spot quantity is 117 IOD for Mt-0.

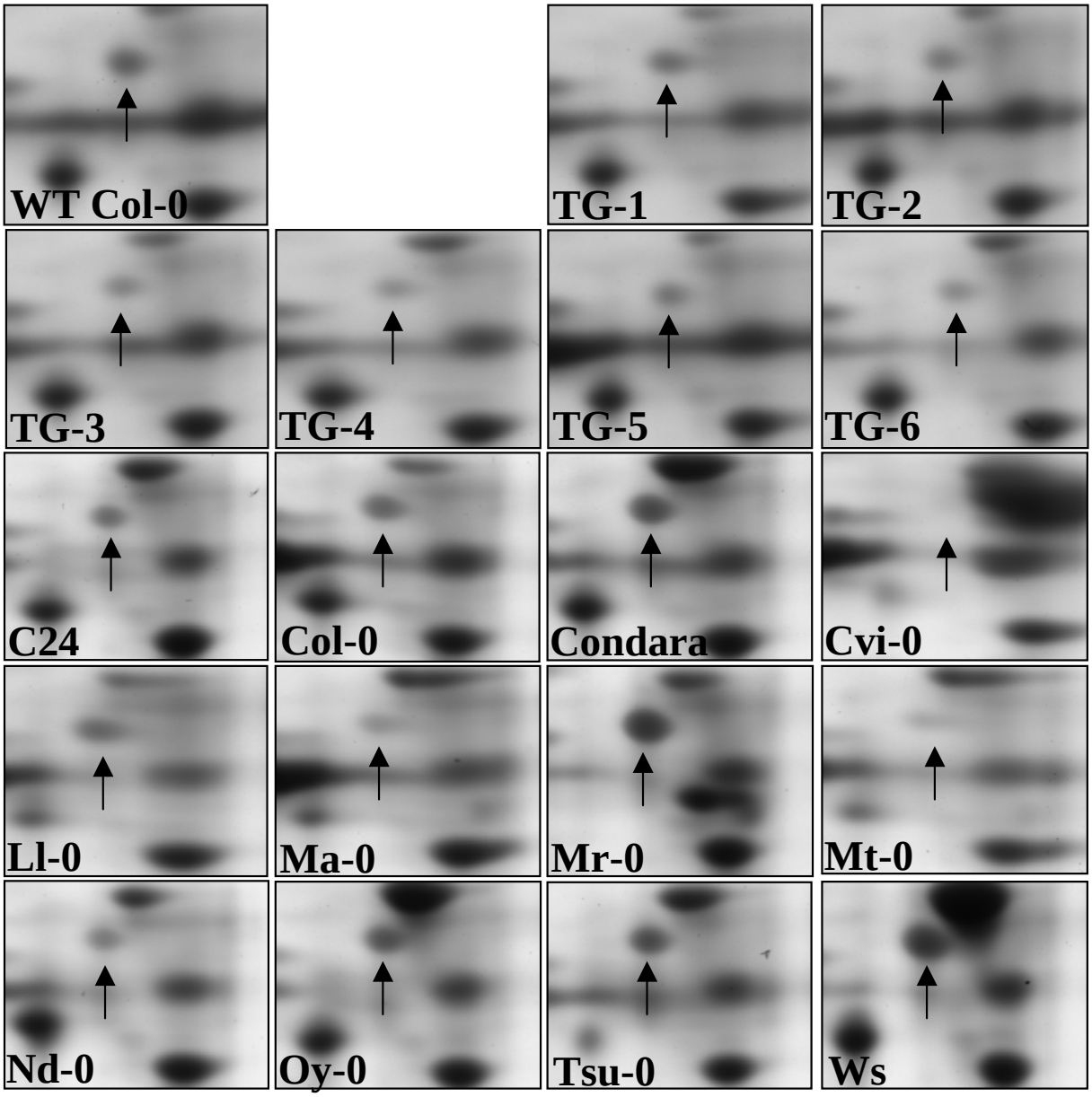


Figure 57 Spot quantity difference of protein spot SSP 8103 (pI=6.7; Mw=23 kD) compared to 12 *Arabidopsis* ecotypes.

Two of the three significant spots of TG-2 exceeded the range of analytical and natural variation. Both spots are located very closely to the GUS protein (pI=5.1; 73 kD) on the 2DE gel (Figure 58).

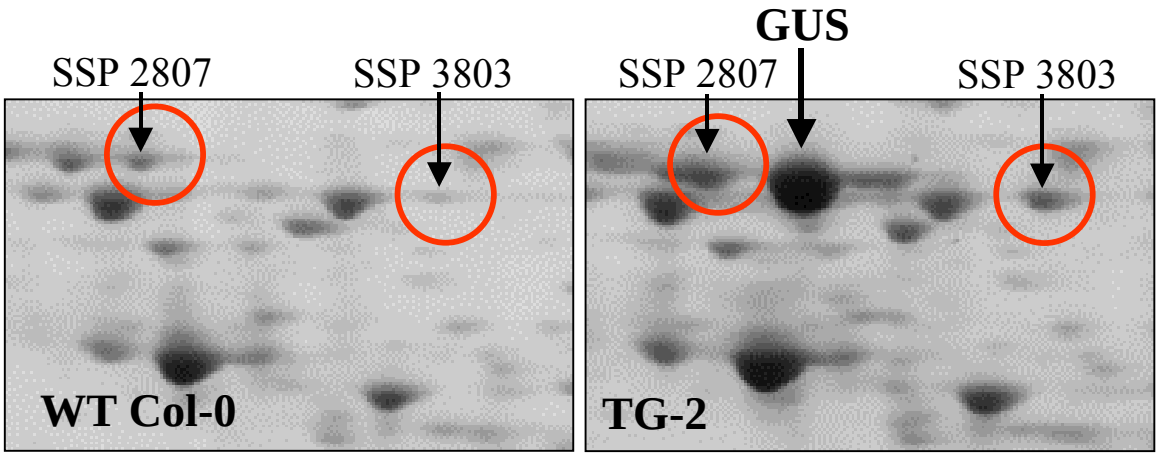


Figure 58 Enlargement of 2DE gel region showing the protein spots SSP 2807 (pI=5.1; Mw=76 kD) and SSP 3803 (pI=5.2; Mw=71 kD) of the transgenic line TG-2 in comparison to the WT Col-0.

The two protein spots SSP 2807 (pI=5.1; Mw=76 kD) and SSP 3803 (pI=5.2; Mw=71 kD) of TG-2 were subjected to MALDI-TOF MS analysis. The measured peptide masses of the protein spots were searched against Swiss-Prot databases using PeptIdent [290, 291]. Two different proteins were identified in spot SSP 2807; the *Arabidopsis* endogenous luminal binding protein (pI=5.08; Mw=71 kD) with 13 matched peptides and 21.2% sequence coverage and β-glucuronidase (pI=5.24; Mw=68 kD) with 11 matched peptides and a sequence coverage of 18.2%. Spot SSP 3803 also comprised more than one protein. β-glucuronidase (pI=5.24; Mw=68 kD) was identified with nine matched peptides and a sequence coverage of 16.7%. Five of the remaining masses match the computed tryptic peptide masses of *Arabidopsis* endogenous heat shock protein (HSP81) and cover 9.3% of the proteins sequence. Hence, the increase in spot quantity of spot SSP 2807 and SSP 3803 is due to fragments of GUS migrating to the same position as the endogenous proteins of *Arabidopsis*. This explanation is supported by the strong correlations between the spot quantity of the GUS spot and the spot quantities of spot SSP 2807 (correlation coefficient = 0.975, P=0.0046) and spot SSP 3803 (correlation coefficient = 0.973, P=0.0054). Figure 59

shows the increase of spot quantities of the spots SSP 2807 and SSP 3803 in conjunction with the increase of GUS.

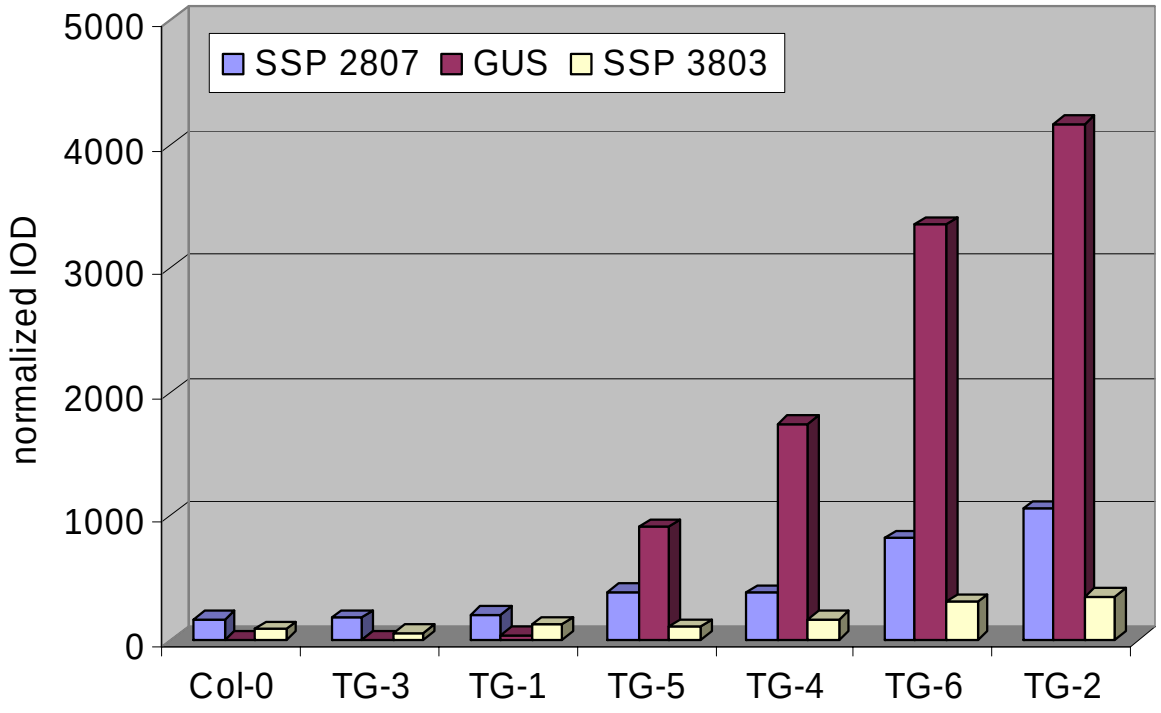


Figure 59 Correlation between the spot quantities of GUS, spot SSP 2807, and spot SSP 3803.

4.4.2 Enhanced tocopherol lines

In this study, the goal was to investigate changes in the protein pattern associated with transgenic lines with an altered endogenous metabolic pathway. Seeds of six transgenic *Arabidopsis* lines and their parental line (WT-P) were provided by Professor Dr. Dean DellaPenna from the Michigan State University, MI, USA. The transgenic lines have previously been described [281, 282] and were homozygous for the transgene [381]. Table 38 shows the lines and their descriptions.

Table 38 Description of control and transgenic lines

Line	Description
WT	Inter experimental control line, wild-type Columbia (Col-0)
WT-P	Parental line of TG lines, wild-type Columbia (Col-0)
35S:HPPD-2 35S:HPPD-3	p-Hydroxyphenylpyruvate dioxygenase (*hppd*) gene driven by CaMV 35S promoter (35S); kanamycin resistant; Ref: Tsegaye et al. [282]
DC3:HPPD-3 DC3:HPPD-8	p-Hydroxyphenylpyruvate dioxygenase (*hppd*) gene driven by seed-specific DC3 promoter (DC3); hygromycin resistant; Ref: Tsegaye et al. [282]
35S:gTMT-18 35S:gTMT-49	γ-tocopherol-methyltransferase (gTMT) gene driven by CaMV 35S promoter (35S); kanamycin resistant; Ref: Collakova and DellaPenna [281]

The WT line is the *Arabidopsis thaliana* ecotype Col-0 used in all previous experiments. Line WT-P is also *Arabidopsis thaliana* Col-0 but came from Dr. Dean DellaPenna's laboratory and is the parental line of the transgenic lines. Both wild-type lines are supposed to have the same genetic background, as both are *A. thaliana* Col-0. The six transgenic lines overexpress the *Arabidopsis* p-hydroxyphenylpyruvate dioxygenase (*hppd*) gene or the *Arabidopsis* γ-tocopherol-methyltransferase (*gtmt*) gene. Both genes encode enzymes of the tocopherol biosynthetic pathway. The enzyme HPPD (EC 1.13.11.27) catalyzes the conversion of p-hydroxyphenylpyruvate to homogentisic acid (HGA), the aromatic precursor for the biosynthesis of vitamin E (tocopherol) and plastoquinone. The expression of the transgene *hppd* is controlled in the TG lines 35S:HPPD-2 and –3 by the cauliflower mosaic virus (CaMV) 35S promoter and in the TG lines DC3:HPPD-3 and –8 by the seed-specific promoter DC3. As determined at the RNA, protein, and activity levels, *hppd* gene expression in these transgenic lines was at least 10-fold higher than that of wild-type plants. It resulted

in a maximum 28% increase (DC3:HPPD-3) in seed tocopherol levels relative to control plants [282].

The enzyme gTMT (2.1.1.95) catalyzes the methylation of γ- and δ-tocopherol to yield α- and β-tocopherol, respectively [382]. The expression of the transgene *gtmt* is controlled in both TG lines (35S:gTMT-18 and –49) by the CaMV 35S promoter. Overexpression of the *gtmt* gene altered the tocopherol composition of *Arabidopsis* seed but not the total tocopherol content. α- and β-tocopherol contents of the 35S:gTMT seeds were increased as much as 37-fold (35S:gTMT-18) due to the conversion of γ- and δ-tocopherol, respectively [381].

4.4.2.1 Phenotypic comparison

The two wild-type and the six transgenic *A. thaliana* lines were grown side-by-side in an environmentally-controlled growth chamber. Ten replicates were planted for each line and randomly distributed in the growth chamber in order to limit the influence of environmental factors. Seeds were harvested after complete maturity of all seeds on a plant. Four phenotypic traits (first flowering date (FFD), rosette leaf diameter (RD), seed yield, and protein seed content) were assayed; they are summarized in Table 39.

Table 39 Phenotypic measurements of the WT Col-0 and the TG lines.

Line	N	FFD[1)] (days)	RD [2)] (cm)	Seed yield[3)] (mg)	Protein[4)] (%)
WT	6	39 ± 0 (N=2)	10.8 ± 0.4 (N=2)	721 ± 101	22.7 ± 0.46
WT-P	10	36 ± 1	9.0 ± 1.0	567 ± 194	23.3 ± 0.38
35S:HPPD-2	10	35 ± 1	8.5 ± 1.0	532 ± 194	22.9 ± 0.21
35S:HPPD-3	10	36 ± 1	8.5 ± 1.0	581 ± 175	23.4 ± 0.49
DC3:HPPD-3	10	38 ± 2*	9.0 ± 1.0	572 ± 177	23.5 ± 0.03
DC3:HPPD-8	10	38 ± 2*	9.5 ± 1.0	553 ± 208	22.9 ± 0.28
35S:gTMT-18	10	37 ± 2	9.0 ± 1.5	633 ± 242	23.6 ± 0.22
35S:gTMT-49	10	37 ± 1	10.0 ± 1.0	562 ± 264	22.2 ± 0.09

Values are means ± SD

[1)] FFD = number of days from the date of planting until the opening of the first flower

[2)] RD = rosette leaf diameter at the time of first flowering

[3)] Amount of harvested seeds

[4)] Protein content of pooled seed samples of six to eight plants; N=2

* significantly different (P<0.05) to WT-P

The RD (measured at Boyes's growth stage 6.00 [340]), seed yield, and protein content do not show significant differences ($P>0.05$) between the transgenic lines and the parental line WT-P. A statistically significant difference ($P<0.05$) in first flowering date (FFD) was observed between WT-P and the transgenic lines DC3:HPPD-3 and DC3:HPPD-8. Plants of the transgenic lines DC3:HPPD-3 and DC3:HPPD-8 started to flower, on average, two days later than WT-P. Although this difference is statistically significant, it is within the range of natural variation of flowering time for the ecotype Col-0 as WT plants started to flower three days later than WT-P. Not observed were differences in leaf and stem morphology, assessed by overall shape, length, thickness, and pubescence.

In addition to these phenotypic traits, the total tocopherol contents and the tocopherol compositions were determined in seeds of the six transgenic lines and compared to the wild-type controls (Table 40).

Table 40 Total tocopherol contents and tocopherol compositions in seeds from WT Col-0 and the TG lines. Total tocopherol and composition are represented as the mean ± SD of four measurements of a pooled seed sample of six plants.[a)]

Line	Total tocopherol (ng/mg fw)	α-Tocopherol (ng/mg fw)	β-Tocopherol (ng/mg fw)	γ-Tocopherol (ng/mg fw)	δ-Tocopherol (ng/mg fw)
WT	378 ± 10	8.98 ± 0.19	< 1	350 ± 8	18.5 ± 1.0*
WT-P	370 ± 5	8.86 ± 0.85	< 1	338 ± 4	22.7 ± 1.1
35S:HPPD-2	420 ± 8*	9.38 ± 0.42	< 1	383 ± 6*	27.4 ± 1.6*
35S:HPPD-3	432 ± 6*	9.12 ± 0.39	< 1	394 ± 5*	28.6 ± 2.0*
DC3:HPPD-3	368 ± 4	8.38 ± 0.43	< 1	338 ± 5	22.2 ± 0.53
DC3:HPPD-8	404 ± 6*	8.77 ± 0.46	< 1	370 ± 5*	25.7 ± 1.2*
35S:gTMT-18	371 ± 4	315 ± 3*	20.3 ± 1.6	33.7 ± 0.46*	2.0 ± 0.08*
35S:gTMT-49	367 ± 7	315 ± 6*	16.6 ± 0.26	34.4 ± 0.78*	1.6 ± 0.04*

* significantly different ($P<0.05$) to WT-P

[a)]Tocopherol analysis of the seeds was conducted by Professor Dr. Dean DellaPenna, Michigan State University, MI, USA).

Total seed tocopherol levels were statistically significantly elevated in 3 of the six transgenic lines. The seed total tocopherol contents were increased 14% in 35S:HPPD-2, 17% in 35S:HPPD–3, and 9% in DC3:HPPD-8 above wild-type (WT-P) level (Table 40). The increases of total tocopherol in the transgenic lines over the WT-P are due to increases of γ- and δ-tocopherol (Table 40). The α- and β-tocopherol contents of these three transgenic lines were not significantly different ($P>0.05$) from the values for WT-P. The transgenic line

DC3:HPPD-3 did not show any changes in total tocopherol or tocopherol composition compared to WT-P. These findings are partly in alignment with the data of Tsegaye et al. [282] for the HPPD-overexpressing transgenic lines. They found an increase in tocopherol content of 11% and 10% for the transgenic lines 35:HPPD-2 and –3, respectively, compared to the wild-type line. The slightly higher increases (14% and 17%) in the present study may be attributed to variation in plant growth conditions. However, according to Tsegaye et al., the seed tocopherol levels for DC:HPPD-3 and –8 were 28% and 27% higher, respectively, than for the wild-type. In the present study, the tocopherol level of DC3:HPPD-8 was only 9% higher than the wild-type (WT-P) level and the tocopherol level of DC3:HPPD-3 was not significantly different from the wild-type (WT-P) level. Variations in plant growth and/or seed harvesting conditions may be possible explanations. The possibility that the lines may not be homozygous and segregation of the transgene may lead to lower tocopherol levels in the pooled seed samples seems unlikely (DellaPenna, personal communication [383]). The nature of this observation was not further investigated.

The total seed tocopherol levels of the transgenic lines 35S:gTMT-18 and –49 were not significantly different ($P>0.05$) from the wild-type (WT-P), while the α- and β-tocopherol levels were increased 36-fold and > 17-fold, respectively, at the expense of γ- and δ-tocopherol (Table 40). The observed changes in the seed tocopherol compositions of the gTMT-overexpressing lines were expected, as γ- and δ-tocopherol are substrates for gTMT [382]. Shintani and DellaPenna [382] also demonstrated that the overexpression of the *gtmt* gene altered the tocopherol composition of *Arabidopsis* seeds but not the total tocopherol content. The alteration of the tocopherol composition was due to the conversion of γ- and δ-tocopherol into α- and β-tocopherol [382].

4.4.2.2 Comparison of 2DE patterns

Seeds harvested from eight individual replicate plants were pooled to one representative ecotype sample in order to minimize the influence of environmental and genetic variation within a line. The pooled seed samples were extracted and four replicate 2DE were performed for each extract. The three best 2DE replicate gels with respect to resolution were subjected to image analysis. The software package PDQuest was used to analyze the gel images. Spots detected and matched by the software program were manually verified. False positive and false negative spots were removed and added, respectively.

4.4.2.2.1 Qualitative comparison of 2DE pattern

Using the optimized 2DE technique, the seed protein profiles of the six transgenic *Arabidopsis* lines were compared to the seed protein profiles of the parental wild-type line Columbia (Col-0). Results of the qualitative comparisons are summarized in Table 41.

Table 41 Qualitative comparison of the seed proteomes of WT lines and TG lines

Line	Detected spots[1)]	Reproducible spots[2)]	Spots matched to WT-P	Spots reproducibly absent[3)]
WT	747 ± 9	708	702	0
WT-P	736 ± 10	706	706	0
35S:HPPD-2	729 ± 10	702	700	0
35S:HPPD-3	737 ± 12	702	701	0
DC3:HPPD-3	747 ± 15	704	703	0[4)]
DC3:HPPD-8	750 ± 27	704	702	0
35S:gTMT-18	730 ± 13	700	699	0
35S:gTMT-49	733 ± 17	705	704	0

[1)] false positive spots due to edge effects, focusing problems, or dust were excluded; Values are means ± SD
[2)] spot must be detected in all three replicate gels
[3)] spot must be absent in all three replicate gels
[4)] spot SSP 9003 was absent in all three replicate gels but present in the forth (control) replicate gel

The number of detected spots varied from 729±10 protein spots for the TG line 35S:HPPD-2 to 750±27 protein spots for the TG line DC3:HPPD-8. The number of detected spots for the two wild-type lines WT and WT-P fell into this range with 747±9 and 736±10 resolved spots, respectively. The majority of the spots found in all lines were reproducible spots (resolved in all three replicate gels) and varied between 700 spots for 35S:gTMT-18 and 708 spots for WT. The protein profiles of the eight lines were very similar, as 99.1% (WT) to 99.9% (35S:gTMT-49) of the reproducible spots of a line were matched to the parental wild-type WT-P and more than 97% (686 spots) of the reproducible spots were matched to all the lines. One reproducible spot (SSP 9003) resolved for the WT-P was absent in the three 2DE replicate gels of the DC3:HPPD-3. However, the protein spot was not scored as reproducibly absent because it is a very faint spot that is negatively affected by focusing problems of the neighboring, very abundant spot SSP 8017. In addition, the spot was unambiguously resolved in the forth 2DE replicate gel of the TG line DC3:HPPD-3.

Figure 60 shows representative 2DE gels for WT-P and for the TG line 35S:HPPD-2 as an example for the head-to-head comparison.

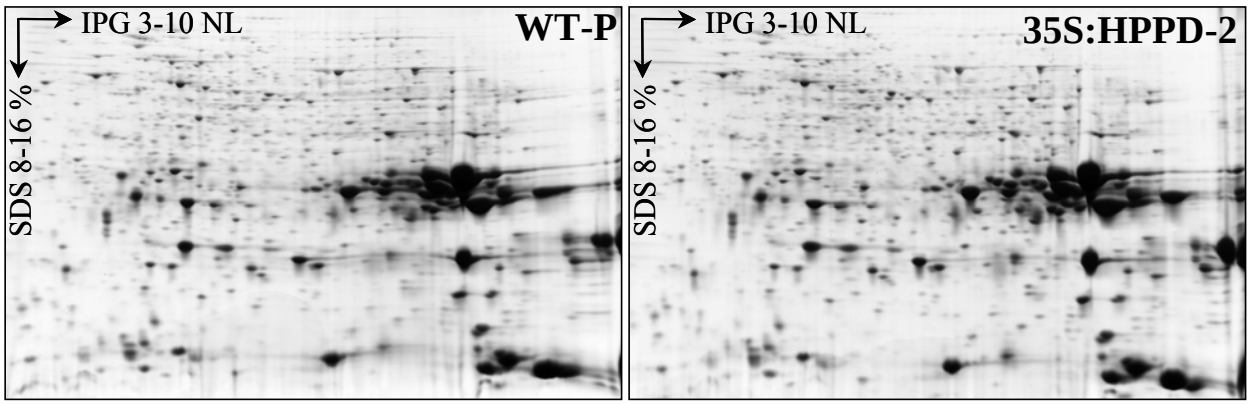


Figure 60 Representative 2DE gels of WT-P and TG line 35S:HPPD-2 are shown as example for the head-to-head comparison between seed protein profiles of WT and TG lines. No difference in presence or absence of spots was observed.

All reproducibly present spots of one of the lines were present in at least one replicate gel of all other samples, i.e. no reproducible qualitative difference was detected between any of the lines.

4.4.2.2.2 Quantitative comparison of 2DE pattern

In order to investigate differences in spot quantities between wild-type and transgenic lines, the quantities of all spots reproducibly present in at least one line and matched to WT-P were examined. A one-way analysis of variance (ANOVA) with protected least significant difference (LSD) was performed for each spot to identify significant differences in spot quantities of 2-fold and greater with a significance level of 0.05. To avoid overestimation of quantitative differences due to inaccurate quantification of poorly resolved protein spots, all spots with the maximum value and an average spot quantity less than or equal to 40 were excluded from the data set. In addition, the detected statistically significant difference (P<0.05) must have a power above 80%, as described in Chapter 4.3. Table 42 summarizes the quantitative comparison of the seed proteomes of WT-P and the six transgenic lines.

About 708 spots were compared. Less than 1% of the spots varied significantly (P<0.05) in spot quantity. The transgenic line 35S:HPPD-2 had, with six spots, the highest number of spots significantly different in spot quantity. The differences in spot quantity were smaller than 3-fold for all significantly different spots.

Table 42 Quantitative comparison of seed proteomes of WT Col-0 and the six TG lines

Line vs. WT	Compared spots	Spots quantitatively different[1)]	
		2-fold	≥ 3-fold
35S:HPPD-2	708	6	0
35S:HPPD-3	707	4	0
DC3:HPPD-3	707	3	0
DC3:HPPD-8	708	4	0
35S:gTMT-18	707	4	0
35S:gTMT-49	707	2	0

[1)] Power of difference must be > 80% and mean spot quality > 40

In order to evaluate the biological significance of those differences, the spot quantities of these spots were compared first to the spot quantities of the other wild-type (WT) grown side-by-side with the WT-P and the transgenic plants (Table 43) using a one-way ANOVA with a Tukey-Kramer post test comparison and then to the spot quantity ranges based on environmental variation (Table 44). Both wild-types, WT and WT-P, represent the same *Arabidopsis* ecotype, Columbia (Col-0), but come from two different laboratories. Differences in their proteome represent natural variation within the ecotype Col-0. Table 43 shows the significantly different spot quantities of WT, WT-P, and transgenic lines.

Table 43 Spot quantities of the significantly different spot for WT, WT-P, and the significantly different transgenic lines

SSP	WT (Col-0)	WT-P (Col-0)	35S: HPPD-2	35S: HPPD-3	DC3: HPPD-3	DC3: HPPD-8	35S: gTMT-18	35S: gTMT-49
0204	171	119				217	217	204
0302	666	470				889		
0308	122	76			142	141		
1106	3037	2414					3306	
4603	159	96	169	156	164		160	
6104	489	428		645				
8302	198	237	103	111		110		
8414	236	330	205					
8611	408	319	189				210	188
8618	183	88	45					
8708	196	232	102					

Values are the mean (N=3) normalized IOD of the spots.

Sixteen of the twenty-one significantly different spots are not significantly different to the spot quantities of WT. The spot quantities of the five remaining significantly different spots were set in context with environmental variation (GC1 vs. GC2). Therefore, the data from

the comparison of the two different growth chambers were included into the data set for the significance test. Table 44 shows the range of environmental variation in spot quantities and the significantly different transgenic lines.

Table 44 Quantity of the remaining significant different spots in the context of environmental variation

SSP	Environment GC1 vs. GC2	WT-P	35S: HPPD-2	35S: gTMT-18	35S: gTMT-49
8611	167 - 408	319	189	210	188
8618	24 - 183	88	45		
8708	183 - 196	232	102		

Values are the mean (N=3) normalized IOD of the spots.

Four of the five significantly different spots fall in the range of environmental variation, and the remaining spot (SSP 8708) is not significantly different from the lowest value found for the environmental variation. Hence it can be concluded that the differences between the transgenic lines and the parental wild-type WT-P did not exceed the range of natural variation.

Contrary to the transgenic protein GUS, which was easily detected in 5 of the 6 transgenic GUS-lines in the previous study, neither HPPD (pI: 5.74; Mw: 49kD) nor gTMT (pI: 5.81; Mw: 33kD) was detected in the 2DE patterns of the transgenic lines as a significant difference compared to the parental wild-type (WT-P). Possible explanations may be that (i) the transgenic proteins are not readily distinguishable from endogenous proteins in the 2DE protein pattern, (ii) low protein amount, and/or (iii) little extractability of the proteins. Unlike GUS, which has not been found in plant species [371, 372], HPPD and gTMT are endogenous *Arabidopsis* proteins, which were overexpressed in the transgenic lines. The proteins will most likely co-migrate to the same position in the 2DE gel as the non-transgenic proteins and not appear as additional protein spots in the 2DE protein pattern, such as GUS in the TG GUS-lines. Therefore, the transgenic proteins may only be recognizable due to changes in spot quantity. Although, Tsegaye et al. [282] demonstrated a HPPT protein increase in seeds of 3.5-fold (35S:HPPD-2) to 17-fold (DC3:HPPD-8) relative to wild-type by immuno-blot analysis, the protein amount of HPPD is probably still below the limit of quantification of the 2DE method (staining: colloidal CBB) due to the low abundance of HPPD in plant tissues [384]. No protein data has been published on gTMT of the transgenic lines used. However, unlike the cytosolic HPPD [384-386], gTMT is a membrane-bound protein [382, 387], and, therefore, most likely not readily extractable.

4.4.3 Summary

The developed and optimized differential 2DE method was used as a proteomics approach to investigate effects due to genetic engineering in the context of natural variation. Twelve transgenic *A. thaliana* lines were analyzed by 2DE and their seed proteomes were compared to that of the wild type line. The transgenic *A. thaliana* lines contain an inserted beta-glucuronidase (*gus*) gene, an overexpressed p-hydroxyphenylpyruvate dioxygenase (*hppd*) gene, or an overexpressed γ-tocopherol-methyltransferase (*gtmt*) gene. These lines were chosen as they represent different strategies: (i) no change to a existing metabolic pathway (*gus* gene) and (ii) change of an endogenous metabolic pathway (tocopherol pathway: *hppd* and *gtmt* gene). The data generated from the natural variability study was used to discuss the relevance of differences between the transgenic lines and their non-transgenic parental lines.

Differences in the phenotype between wild-type and transgenic plants were in the range of natural variation of phenotypes. An analysis of the tocopherol compositions of the seeds revealed differences in the levels of tocopherol (intended differences) due to the introduction of the transgenes *hppd* and *gtmt*. One line (DC3:HPPD-3) of the six tocopherol enhanced transgenic lines did not demonstrate the intended effect.

The qualitative comparison of the seed proteomes of the wild-type and the transgenic lines demonstrated almost identical 2DE protein patterns with respect to presence and absence of protein spots. One additional protein spot was resolved in five of the six transgenic lines expressing the *gus* gene. The protein spot was identified by MALDI-TOF MS analysis as the GUS protein. The transgene products for the HPPD and gTMT lines were not detected by 2DE analysis.

The quantitative comparison of the seed proteomes of the wild-type and the transgenic lines revealed some spots with significantly different spot quantities compared to wild-type. The quantities of two protein spots of one of the *gus* gene-expressing transgenic lines exceeded the range of analytical and natural variation. It was demonstrated that fragments of the GUS protein, which co-migrated to the same position in the 2DE gel as the endogenous proteins, caused the increase of spot quantity.

Based on the 2DE analysis, no difference, such as loss, shift, or quantity change of a protein spot, exceeding the natural variation was detected between the seed proteomes of parental and transgenic lines.

5 Summary

The applicability of proteomics to investigate differences in the plant proteome due to genetic engineering was explored using *Arabidopsis (A.) thaliana* as a model organism. Differences in the proteome were evaluated in the context of natural variability.

A proteomics method, based on two-dimensional gel electrophoresis (2DE), was established for the qualitative and quantitative analysis of the seed proteome of *A. thaliana* and validated for repeatability, sensitivity, and linearity. The developed 2DE method resolves proteins with isoelectric points between 4 and 9 and molecular weights of 6 to 120 kD and is sensitive enough to detect protein levels in the low nanogram range. Using this method, it has been possible to resolve and compare spot positions and quantities of > 700 protein spots on midsize gels. The linear relationship between protein amount and spot quantity was demonstrated for two spiked protein standards and 20 endogenous seed proteins representing a wide range of pIs, Mws, and concentrations. The protein pattern repeatability was found to be high and three replicate gels were sufficient to investigate a threefold difference in spot quantity for most resolved spots (~96%) and a twofold difference for the majority (~72%) of resolved spots.

The understanding of the natural variability of the proteome is crucial for the interpretation of biological and safety-relevant differences between transgenic and non-transgenic parental lines. Thus, the natural variation of protein profiles within the *A. thaliana* ecotype Columbia (Col-0) and among a set of *A. thaliana* ecotypes was determined by utilizing the validated 2DE method. The seed protein pattern within the ecotype Col-0 was rather constant. However, differences in spot quantity were observed between seed protein profiles of individual plants of Col-0 grown side-by-side or in two different growth chambers with the same settings. The data indicates the potential of small environmental changes on the seed proteome and highlights the sensitivity of the optimized 2DE method to detect such differences. In addition, it was shown that storage also impacted the seed proteome. The natural variability of seed protein profiles resulting from different genetic backgrounds was found to be extensive among a set of twelve *A. thaliana* ecotypes. Almost half of the resolved spots varied with respect of their presence/absence and one quarter of the spots, present in all ecotypes, varied in spot quantity (3- to 53-fold). These data have been used to ascertain the biological and genetic variation in protein profiles within and between *A. thaliana* ecotypes and as baseline a for the head-to-head comparison of transgenic versus

non-transgenic *Arabidopsis* lines in order to assess differences in the context of natural variability.

The 2DE method was applied to the comparison of transgenic versus non-transgenic *Arabidopsis* lines. Insertional and pleiotropic effects on the proteome due to genetic engineering were investigated using twelve transgenic *A. thaliana* lines containing an inserted β-glucuronidase (*gus*) gene, an overexpressed p-hydroxyphenylpyruvate dioxygenase (*hppd*) gene, or an overexpressed γ-tocopherol-methyltransferase (*gtmt*) gene. These lines were chosen as they present two different strategies: (i) a transgene (*gus*) that has no impact on an endogenous metabolic pathway and (ii) transgenes (*hppd* and *gtmt*) that have impact on an endogenous metabolic pathway (α-tocopherol biosynthesis). The genetic modification of the *Arabidopsis* lines using three different genes and three different promoters did not result in any phenotypic differences exceeding the natural variation other than the intended differences due to the introduction of the transgene. The same was found for the seed proteome. The process of transformation did not cause insertional or pleiotropic changes to the analyzed seed proteome. Differences in spot quantity between transgenic and non-transgenic lines fell in the range of natural variation or were part of the intended effect.

This work demonstrated that 2DE can be utilized to reliably analyze the seed proteome of *A. thaliana*. During the course of the study, it became evident that a critical data analysis needs to take into consideration the analytical and the natural variability of the proteome. The latter is essential for the evaluation of potential insertional and pleiotropic effects in comparing transgenic and non-transgenic plants. Thus, a proteome analysis should comprise the following steps: (i) method validation, (ii) generation of baseline data for the natural variation, and (iii) head-to-head comparison between transgenic and non-transgenic plants in the context of the established analytical and natural variation. The proteomics approach described for *Arabidopsis thaliana* promises to be useful for the analysis of the proteome in other plant species including crop plants in order to investigate effects due to genetic engineering.

6 Zusammenfassung

In der vorliegenden Arbeit wurde die Eignung von Proteomik zur Untersuchung von Auswirkungen gentechnischer Verfahren auf Pflanzenproteome überprüft. Als Modellorganismus wurde *Arabidopsis thaliana* gewählt. Veränderungen im Proteom wurden vor dem Hintergrund der natürlichen Schwankungsbreite bewertet.

Zur qualitativen und quantitativen Analyse des Samenproteoms von *A. thaliana* wurde eine auf zweidimensionaler Gelelektrophorese (2DE) basierende Methodik etabliert und hinsichtlich Wiederholbarkeit, Empfindlichkeit und Linearität validiert. Mit der entwickelten 2DE-Methode war es möglich, Proteine mit isoelektrischen Punkten (pI) zwischen 4 und 9 und Molekulargewichten (Mw) von 6 bis 120 kD aufzutrennen und im Nanogramm-Bereich nachzuweisen. Auf einem Mittelformat-Gel konnten mehr als 700 Proteinspots aufgetrennt und hinsichtlich ihrer Position und Quantität verglichen werden. Der lineare Zusammenhang zwischen Proteinmenge und Spotquantität wurde am Beispiel von zwei dotierten Proteinstandards und anhand von 20 Samenproteinen, die einen weiten Bereich an pIs, Mws und Konzentrationen abdeckten, aufgezeigt. Die Reproduzierbarkeit der Proteinmuster war sehr gut; drei Replikate waren ausreichend, um einen 3-fachen Unterschied in Spotquantität für fast alle Proteinspots (~96%) und einen 2-fachen Unterschied für die Mehrzahl (~72%) der aufgetrennten Spots zu ermitteln.

Die validierte 2DE Methode wurde eingesetzt, um natürliche Schwankungen von Samenproteinprofilen für den *A. thaliana* Ökotyp Columbia (Col-0) und innerhalb einer Reihe von *A. thaliana* Ökotypen zu bestimmen. Trotz standardisierter Umweltbedingungen konnten Unterschiede in den Spotquantitäten individueller Col-0 Pflanzen beobachtet werden. Das qualitative Samenproteinmuster war innerhalb individueller Pflanzen des Ökotypes Col-0 relativ konstant. Die Daten zeigen das Einflusspotential schon geringer Änderungen in den Umweltbedingungen auf das Samenproteom und verdeutlichen die Nachweisempfindlichkeit der optimierten 2DE-Methode. Auch die Lagerung der geernteten Samen beeinflusste das Samenproteom und führte zu Änderungen in der Quantität von Proteinspots. Für die zwölf untersuchten *A. thaliana* Ökotypen erwies sich die aus den unterschiedlichen genetischen Hintergründen resultierende natürliche Schwankungsbreite der Samenproteinprofile als bemerkenswert groß. Fast die Hälfte der aufgelösten Spots variierten hinsichtlich An-/Abwesenheit und 25% der Spots, die in allen Ökotypen anwesend waren, variierten in Spotquantität (3- bis 53-fach).

Die ermittelten Schwankungsbreiten des Proteinprofils innerhalb und zwischen *A. thaliana* Ökotypen wurden als Grundlage herangezogen, um die in einem direkten Vergleich von transgenen und nicht-transgenen *Arabidopsis*-Linien ermittelten Unterschiede hinsichtlich ihrer biologischen Relevanz zu beurteilen. Mit Hilfe der 2DE-Methode wurde aus den gentechnischen Veränderungen möglicherweise resultierende Insertionseffekte oder pleiotrope Effekte in zwölf transgenen *A. thaliana* Linien, die entweder ein neu eingefügtes *β*-Glucuronidase (*gus*) Gen, ein überexprimiertes p-Hydroxyphenylpyruvat-Dioxygenase (*hppd*) Gen oder ein überexprimiertes γ-Tocopherol-methyltransferase (*gtmt*) Gen enthielten, untersucht. Diese Linien wurden ausgewählt, da sie zwei unterschiedliche Strategien repräsentieren: (i) ein Transgen (*gus*), das keinen Einfluss auf einen endogenen Stoffwechselweg hat und (ii) Transgene (*hppd* und *gtmt*), die einen Einfluss auf einen endogenen Stoffwechselweg (α-Tocopherol-Biosynthese) haben. Die gentechnisch veränderten *Arabidopsis*-Linien zeigten keine Veränderungen im Phänotyp, die über die natürlichen Schwankungen hinausgingen. Auch die durch Proteomanalytik aufgezeigten Unterschiede in Spotquantitäten zwischen transgenen und nicht-transgenen Linien lagen im Rahmen der natürlichen Schwankungsbreiten oder waren das Ergebnis der beabsichtigten Veränderungen.

Die erarbeiteten Daten zeigen, dass 2DE eine zuverlässige Methode zur Untersuchung des Samenproteoms von *A. thaliana* darstellt. Es wurde deutlich, dass für eine kritische Datenanalyse und eine Beurteilung möglicher Insertionseffekte oder pleiotroper Effekte die Berücksichtigung der analytischen und natürlichen Schwankungsbreite des Proteoms essentiell ist. Eine Proteomuntersuchung sollte daher folgende Schritte beinhalten: (i) Methodenvalidierung, (ii) Erarbeitung von Basisdaten zu natürlichen Schwankungsbreiten, und (iii) direkter Vergleich von transgenen und nicht-transgenen Pflanzen unter Berücksichtigung der etablierten analytischen und natürlichen Schwankungsbreiten. Die hier zur Analyse des Proteoms von *Arabidopsis thaliana* vorgestellte Methode verspricht auch zur Untersuchung möglicher Auswirkungen gentechnischer Verfahren auf das Proteom anderer Pflanzen, z. B. Nutzpflanzen, anwendbar zu sein.

7 Bibliography

1. James, C. (2003) Preview: Global Status of Commercialized Transgenic Crops: 2003. International Service for the Acquisition of Agri-biotech Applications (ISAAA): Ithaca, NY.
2. OECD (1993) Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles. Organization for Economic Cooperation and Development (OECD), Paris.
3. Kleter, G. A. and Kuiper, H. A. (2003) Safety assessment of foods derived from genetically modified crops. *AgBiotechNet*, **5**: ABN114.
4. Astwood, J. D. and Fuchs, R. L. (2001), Status and Safety of Biotech Crop, in *Agrochemical Discovery - Insect, Weed, and Fungal Control*, D.R. Baker and N.K. Umetsu, Editors. American Chemical Society: Washington, DC. 152-64.
5. GM Science Review Panel - UK (2004) GM Science Review Panel: Second Report. An open review of the science relevant to GM crops and food based on interests and concerns of the public. *online publication:* http://www.gmsciencedebate.org.uk/report/, ed. D. Sir King and H. Dalton, 22 January 2004.
6. Society of Toxicology (2003) The Safety of Genetically Modified Foods Produced through Biotechnology. *Toxicol. Sci.*, **71**(1): 2-8.
7. EC Scientific Steering Committee (2000) Risk assessment in a rapidly evolving field: the case of genetically modified plants (GMP), ed. Scientific Steering Committee. European Commission, Brussels.
8. Kuiper, H. A., Kok, E. J., and Engel, K. H. (2003) Exploitation of molecular profiling techniques for GM food safety assessment. *Curr Opin Biotechnol*, **14**(2): 238-43.
9. Potrykus, I. (2001) Golden rice and beyond. *Plant Physiol*, **125**(3): 1157-61.
10. Kok, E. J. and Kuiper, H. A. (2003) Comparative safety assessment for biotech crops. *Trends Biotechnol*, **21**(10): 439-44.
11. FAO/WHO (2000) Safety aspects of genetically modified foods of plant origin. Joint FAO/WHO consultation on foods derived from biotechnology, Geneva, Switzerland 29 May - 2 June 2000. Food and Agriculture Organization of the United Nations and World Health Organization.
12. Kuiper, H. A., Kleter, G. A., Noteborn, H. P. J. M., and Kok, E. J. (2001) Assessment of the food safety issues related to genetically modified foods. *Plant Journal*, **27**(6): 503-28.
13. EC Scientific Steering Committee (2003) Guidance Document for the risk assessment of genetically modified plants and derived food and feed (6-7 March 2003), The Joint Working Group on Novel Foods and GMOs, Editor. European Commission: Brussels. http://europa.eu.int/comm/food/fs/sc/ssc/out327_en.pdf.
14. OECD (1998) Report of the OECD Workshop on the Toxicological and Nutritional Testing of Novel Foods, Aussois, France, 5-8 March, 1997. Organization for Economic Cooperation and Development (OECD), Paris.
15. Ereky, K. (1919) Biotechnologie der Fleisch-, Fett- und Milcherzeugung im landwirtschaftlichen Grossbetrieb. Verlag Paul Parey, Berlin.
16. United Nations (1992) Convention on Biological Diversity: Rio de Janeiro, Brazil. http://www.biodiv.org/doc/legal/cbd-en.pdf.
17. Allard, R. W. (1999) History of plant population genetics. *Annu Rev Genet*, **33**: 1-27.
18. Lukaszewski, A. J. and Gustafson, J. P. (1987) Cytogenetics of triticale. *Plant Breeding*, **5**: 41-93.
19. Cohen, S. N., Chang, A. C., Boyer, H. W., and Helling, R. B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A*, **70**(11): 3240-4.

20. Russo, E. (2003) The birth of biotechnology. *Nature*, **421**(6921): 456-7.

21. Zambryski, P. C. (1992) Chronicles from the *Agrobacterium*-Plant Cell DNA Transfer Story. *Annual Review of Plant Physiology and Plant Molecular Biology*, **43**: 465-90.

22. Gelvin, S. B. (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev*, **67**(1): 16-37, table of contents.

23. Klein, T. M., Wolf, E. D., Wu, R., and Sanford, J. C. (1987) High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, **327**: 70-3.

24. U.S. Regulatory Agencies (2004) U.S. Database of Completed Regulatory Agency Reviews, in http://usbiotechreg.nbii.gov/database_pub.asp. as of February 2004.

25. Puonti-Kaerlas, J., Eridsson, T., and Engstrom, P. (1990) Production of transgenic pea (*Pisum sativum* L.) plants by *Agrobacterium tumefaciens*-mediated gene transfer. *Theoretical and Applied Genetics*, **80**: 246-52.

26. Schroeder, H. E., Schotz, A. H., Wardley-Richardson, T., Spencer, D., and Higgins, T. (1993) Transformation and Regeneration of Two Cultivars of Pea (*Pisum sativum* L.). *Plant Physiol*, **101**(3): 751-7.

27. Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T., and Kumashiro, T. (1996) High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. *Nat Biotechnol*, **14**(6): 745-50.

28. Hiei, Y., Komari, T., and Kubo, T. (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Mol Biol*, **35**(1-2): 205-18.

29. Cheng, M., Fry, J. E., Pang, S., Zhou, H., Hironaka, C. M., Duncan, D. R.*et al.* (1997) Genetic Transformation of Wheat Mediated by *Agrobacterium tumefaciens*. *Plant Physiol*, **115**(3): 971-80.

30. Gordon-Kamm, W. J., Spencer, T. M., Mangano, M. L., Adams, T. R., Daines, R. J., Start, W. G.*et al.* (1990) Transformation of Maize Cells and Regeneration of Fertile Transgenic Plants. *Plant Cell*, **2**(7): 603-18.

31. Ignacimuthu, S., Arockiasamy, S., and Terada, R. (2000) Genetic transformation of rice: Current status and future prospects. *Current Science*, **79**(2): 186-95.

32. Wan, Y. and Lemaux, P. G. (1994) Generation of Large Numbers of Independently Transformed Fertile Barley Plants. *Plant Physiol*, **104**(1): 37-48.

33. Janakiraman, V., Steinau, M., McCoy, S., and Trick, H. (2002) Recent advances in wheat transformation. *In Vitro Cellular & Developmental Biology-Plant*, **38**(5): 404-14.

34. U.S. Food and Drug Administration (FDA) (1994) Summary of consultation with Calgene, Inc., concerning FLAVR SAVR(TM) tomatoes, Memorandum. Food and Drug Administration (FDA).

35. AgBios (2004) GM Crop Database: Information on GM Approved Products, in http://www.agbios.com/dbase.php. as of February 2004.

36. Munkvold, G. P. (2003) Cultural and genetic approaches to managing mycotoxins in maize. *Annu Rev Phytopathol*, **41**: 99-116.

37. Beyer, P., Al-Babili, S., Ye, X., Lucca, P., Schaub, P., Welsch, R.*et al.* (2002) Golden Rice: introducing the beta-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. *J Nutr*, **132**(3): 506S-10S.

38. Galili, G., Galili, S., Lewinsohn, E., and Tadmor, Y. (2002) Genetic, Molecular, and Genomic Approaches to Improve the Value of Plant Foods and Feeds. *Critical Reviews in Plant Sciences*, **21**(3): 167-204.

39. Humphrey, J. H., West, K. P., Jr., and Sommer, A. (1992) Vitamin A deficiency and attributable mortality among under-5-year-olds. *Bull World Health Organ*, **70**(2): 225-32.

40. Hoa, T. T., Al-Babili, S., Schaub, P., Potrykus, I., and Beyer, P. (2003) Golden Indica and Japonica rice lines amenable to deregulation. *Plant Physiol*, **133**(1): 161-9.

41. Monsanto Company public website (2004) Nutritionally enhanced products being developed via biotechnology, in http://www.monsanto.com/monsanto/content/sci_tech/literature/techpubs/nutrition.pdf. as of February 2004.

42. Shewmaker, C. K., Sheehy, J. A., Daley, M., Colburn, S., and Ke, D. Y. (1999) Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *Plant J*, **20**(4): 401-12X.

43. Dhawan, V. (2001) Biotechnology and the promise for control of vitamin A deficiency, in Presentation at the XX meeting of the International Vitamin A Consultative Group, February 15, 2001 Vietnam.

44. Goto, F., Yoshihara, T., Shigemoto, N., Toki, S., and Takaiwa, F. (1999) Iron fortification of rice seed by the soybean ferritin gene. *Nat Biotechnol*, **17**(3): 282-6.

45. Voelker, T. (1997), Transgenic manipulation of edible oilseeds, in *Functionality of Food Phytochemicals*, T. Johns and J.T. Romeo, Editors. Plenum Publ Corp.: New York, NY.

46. Kinney, A. J. (1997), Genetic engineering of oilseeds for desired traits, in *Genetic Engineering*, J.K. Setlow, Editor. Plenum Press: New York. 149-66.

47. List, G. R. and al., e. (1996) Potential margarine oils from genetically modified soybeans. *Journal of the American Oil Chemists Society (JAOCS)*, **73**: 729-32.

48. Froman, B. and Ursin, V. (2002) Genetic modification of oils for improved health benefits: Production of long chain omega-3 fatty acids in plants, in Abstr Pap Am Chem Soc 223: 031-AGFD Part 1.

49. Coghlan, A. (2003) "Potato" to feed India's poor. *New Scientist*, **177**(2376): 7.

50. Katsube, T., Kurisaka, N., Ogawa, M., Maruyama, N., Ohtsuka, R., Utsumi, S.*et al.* (1999) Accumulation of Soybean Glycinin and Its Assembly with the Glutelins in Rice. *Plant Physiol.*, **120**(4): 1063-74.

51. Momma, K., Hashimoto, W., Ozawa, S., Kawai, S., Katsube, T., Takaiwa, F.*et al.* (1999) Quality and safety evaluation of genetically engineered rice with soybean glycinin: analyses of the grain composition and digestibility of glycinin in transgenic rice. *Biosci Biotechnol Biochem*, **63**(2): 314-8.

52. Momma, K., Hashimoto, W., Yoon, H. J., Ozawa, S., Fukuda, Y., Kawai, S.*et al.* (2000) Safety assessment of rice genetically modified with soybean glycinin by feeding studies on rats. *Biosci Biotechnol Biochem*, **64**(9): 1881-6.

53. Jung, R. and Kinney, A. (2001) Hypoallergenic Transgenic Soybeans, in World Intellectual Property Organization. WO 01/6887 A2.

54. Robinson, S., Scott, N., and Gackle, A. (2000) Gene technology and future foods. *Asia Pac J Clin Nutr*, **9**(s1): S113-S8.

55. CSIRO (2003) The not-so-humble spud, in http://www.csiro.au/.

56. Kerr, P. S. and al., e. (1997) Nucleotide sequences of galactinol synthase from zucchini and soybean, in Official Gazette of the United States Patent and Trademark Office.

57. O'Quinn, P. R., Nelssen, J. L., Goodband, R. D., Knabe, D. A., Woodworth, J. C., Tokach, M. D.*et al.* (2000) Nutritional value of a genetically improved high-lysine, high-oil corn for young pigs. *J Anim Sci*, **78**(8): 2144-9.

58. Ma, J. K., Drake, P. M., and Christou, P. (2003) The production of recombinant pharmaceutical proteins in plants. *Nat Rev Genet*, **4**(10): 794-805.

59. Rowlandson, K. and Tackaberry, E. (2003) Edible vaccines: alternatives to conventional immunization. *AgBiotechNet*, **5**: ABN 115.

60. Hood, E. E., Kusnadi, A., Nikolov, Z., and Howard, J. A. (1999) Molecular farming of industrial proteins from transgenic maize. *Adv Exp Med Biol*, **464**: 127-47.

61. MacKenzie, D. J. (2000) International Comparison of Regulatory Frameworks for Food Products of Biotechnology. Agriculture and Biotechnology Strategies: Canada. 62.

62. U.S. Office of Science and Technology Policy (OSTP) (1986) Coordinated framework for regulation of biotechnology; announcement of policy; notice for public comment. *Federal Register*, **51**(123): 23302-50.

63. Abbott, A. (1999) A post-genomic challenge: learning to read patterns of protein synthesis. *Nature*, **402**(6763): 715-20.

64. Hails, R. and Kinderlerer, J. (2003) The GM public debate: context and communication strategies. *Nat Rev Genet*, **4**(10): 819-25.

65. GATT (1994) The General Agreement on Tariffs and Trade, Uruguay Round of Multilateral Trade Negotiations, Marrakesh on 15 April 1994. *WTO Agreements*. GATT.

66. SPS Agreement (1994) Agreement on the Application of Sanitary and Phytosanitary Measures, Uruguay Round of Multilateral Trade Negotiations, Marrakesh on 15 April 1994. *WTO Agreement*. Committee on Sanitary and Phytosanitary Measures.

67. Cartagena Protocol (2000) Cartagena Protocol on Biosafety to the Convention on Biological Diversity: text and annexes. Secretariat of the Convention on Biological Diversity, Montreal.

68. EC (2001) Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms. *Official Journal of the European Union*, **L 106**(17/04/2001): 0001 - 39.

69. EEC (1990) Council Directive 90/220/EEC of 23 April 1990 on the deliberate release into the environment of genetically modified organisms. *EUR-Lex Official Journal*, **L 117**(08/05/1990): 0015 - 27.

70. EC MEMO/04/17 (2004) State of play on GMO authorisations under EU law. European Commission, Brussels; January 28, 2004.

71. JRC (2004) Deliberate releases and placing on the EU market of Genetically Modified Organisms (GMOs). *online database*. Joint Research Centre of the European Commision.

72. EC (2003) Regulation (EC) 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. *Official Journal of the European Union*, **L 268**(18/10/2003): 0001 - 23.

73. EC (1997) Regulation (EC) No 258/97 of the European Parliament and of the Council of 27 January 1997 concerning novel foods and novel food ingredients. *Official Journal of the European Union*, **L 043**(14/02/1997): 0001 - 6.

74. EC (2003) Regulation (EC) 1830/2003 of the European Parliament and of the Council of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms. *Official Journal of the European Union*, **L 268**(18/10/2003): 0024 - 8.

75. FAO/WHO (1996) Biotechnology and food safety. Report of a Joint FAO/WHO Consultation, Rome, Italy, 30 September-4 October 1996. *FAO Food and Nutrition Paper*, **No. 61**.

76. FAO/WHO (2001) Safety assessment of foods derived from genetically modified microorganisms - joint FAO/WHO expert consultation on foods derived from biotechnology, Geneva, Switzerland, 24 to 28 September 2001. Food and Agriculture Organization of the United Nations and World Health Organization.

77. FAO/WHO (2002) Report of the third session of the Codex ad hoc Intergovernmental Task Force on Foods Derived from Biotechnology. Yokohama, Japan, 4-8 March 2002. *ALINORM 03/34*. CODEX ALIMENTARIUS COMMISSION.

78. IFBiC (1990) Biotechnologies and Food: Assuring the Safety of Foods Produced by Genetic Modification. International Food Biotechnology Council. *Regulatory Toxicology and Pharmacology*. Vol. 12, No. 3, Part 2. Academic Press, San Diego, CA.

79. ILSI Europe (1997) Novel Foods Task Force: The safety assessment of novel foods. *Food Chem Toxicol*, **34**: 931-40.

80. OECD (1996) Food Safety Evaluation. Organization for Economic Cooperation and Development (OECD), Paris.

81. OECD (2002) Report of the OECD Workshop on Nutritional Assessment of Novel Foods and Feeds in Ottawa, Canada, February 2001. *Series on the Safety of Novel Foods and Feeds, No. 5*. Organization for Economic Cooperation and Development (OECD), Paris.

82. Holm, S., Alexander, J., Andersson, C., Aune, T., Gry, J., Phu Lich, N.*et al.* (1998) Nettox list of food plants prioritised for inclusion in a future European database. Danish Vet. Food Adm. Report no. 6.

83. Novak, W. K. and Haslberger, A. G. (2000) Substantial equivalence of antinutrients and inherent plant toxins in genetically modified novel foods. *Food Chem Toxicol*, **38**(6): 473-83.

84. FAO/WHO (2003) Report of the fourth session of the Codex ad hoc Intergovernmental Task Force on Foods Derived from Biotechnology. Yokohama, Japan, 11 - 14 March 2003. *ALINORM 03/34A*. CODEX ALIMENTARIUS COMMISSION.

85. 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, Consensus Documents*. Organization for Economic Cooperation and Development (OECD), Paris, France.

86. OECD (2002) 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, Consensus Documents*. Organization for Economic Cooperation and Development (OECD), Paris, France.

87. OECD (2002) 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, Consensus Documents*. Organization for Economic Cooperation and Development (OECD), Paris, France.

88. OECD (2002) 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, Consensus Documents*. Organization for Economic Cooperation and Development (OECD), Paris, France.

89. OECD (2001) 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, Consensus Documents*. Organization for Economic Cooperation and Development (OECD), Paris, France.

90. OECD (2001) Consensus Document on Key Nutrients and Key Toxicants in Low Erucic Acid Rapeseed (Canola). *Series on the Safety of Novel Foods and Feeds No. 1, Consensus Documents*. Organization for Economic Cooperation and Development (OECD), Paris, France.

91. OECD (1995) OECD Guidelines for the Testing of Chemicals. Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents (Adopted 27th July 1995). Organization for Economic Cooperation and Development (OECD), Paris.

92. Metcalfe, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor, S. L., and Fuchs, R. L. (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Crit Rev Food Sci Nutr*, **36 Suppl**: S165-86.

93. FAO/WHO (2001) Allergenicity of genetically modified foods - Joint FAO/WHO consultation on foods derived from biotechnology, Rome, Italy, 22-25 January 2001. Food and Agriculture Organization of the United Nations and World Health Organization.

94. Conner, A. J. (1993) Food safety issues relating to genetic engineering of crop plants. *Agricultural Science*, **6**: 36-41.

95. Conner, A. J. and Jacobs, J. M. E. (1999) Genetic engineering of crops as potential source of genetic hazard in the human diet. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, **443**(1-2): 223-34.

96. Conner, A. J. and Jacobs, J. M. E. (2000) Food risks from transgenic crops in perspective. *Nutrition*, **16**(7-8): 709-11.

97. Pedersen, S., Eriksen, F. D., and Knudsen, I. (2001), Toxicity and Food Safety of Genetically Engineered Crops, in *Safety of Genetically Engineered Crops*, R. Custers, Editor. Jo Bury, VIB: Zwijnaarde. 27-60.

98. van Gelder, W. M. J. and Scheffer, J. J. C. (1991) Transmission of steroidal glycoalkaloids from Solanum vernei to the cultivated potato. *Phytochemistry*, **30**(1): 165-8.

99. Jany, K.-D. and Greiner, R. (1998) Gentechnik und Lebensmittel. *Berichte der Bundesforschungsanstalt fuer Ernaehrung (BFE-R--98-1)*. Bundesforschungsanstalt fuer Ernaehrung, Karlsruhe.

100. Cellini, F., Chesson, A., Colquhoun, I., Constable, A., Davies, H. V., Engel, K. H.*et al.* (2004) Unintended effects and their detection in genetically modified crops. *Food Chem Toxicol*, **42**(7): 1089-125.

101. Thomas, W. T. B., Baird, E., Fuller, J. D., Lawrence, P., Young, G. R., Russell, J.*et al.* (1998) Identification of a QTL decreasing yield in barley linked to Mlo powdery mildew resistance, in Molecular Breeding. 381-93.

102. Beier, R. C. (1990) Natural pesticides and bioactive components in foods. *Rev Environ Contam Toxicol*, **113**: 47-137.

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

104. Harvey, M. H., McMillan, M., Morgan, M. R., and Chan, H. W. (1985) Solanidine is present in sera of healthy individuals and in amounts dependent on their dietary potato consumption. *Hum Toxicol*, **4**(2): 187-94.

105. Coulston, F. and Kolbye, A. C. (1990) Biotechnologies and food: assuring the safety of foods produced by genetic modification. *Regulatory Toxicology and Pharmacology*, **12**: 1-196.

106. Engel, K.-H., Gerstner, G., and 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. 197-209.

107. Hashimoto, W., Momma, K., Katsube, T., Ohkawa, Y., Ishige, T., Kito, M.*et al.* (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, in Journal of the Science of Food and Agriculture. 1607-12.

108. Turk, S. C. H. J. and Smeekens, S. C. M. (1999), Genetic modification of plant carbohydrate metabolism, in *Applied Plant Biotechnology*, V.L. Chopra, V.S. Malik, and S.R. Bhat, Editors. Science Publishers: Enfield. 71-100.

109. Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P.*et al.* (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, **287**(5451): 303-5.

110. Delhaize, E., Hebb, D. M., Richards, K. D., Lin, J. M., Ryan, P. R., and Gardner, R. C. (1999) Cloning and expression of a wheat (*Triticum aestivum* L.) phosphatidylserine synthase cDNA. Overexpression in plants alters the composition of phospholipids. *J Biol Chem*, **274**(11): 7082-8.

111. ILSI (2003) Crop Composition Database. International Life Sciences Institute.

112. Cockburn, A. (2002) Assuring the safety of genetically modified (GM) foods: the importance of an holistic, integrative approach. *J Biotechnol*, **98**(1): 79-106.

113. OECD (2003), Consensus Document for the Work on the Safety of Novel Foods and Feeds. Organization for Economic Cooperation and Development (OECD): Paris, France. URL: http://www.oecd.org/document/9/0,2340,en_649_34387_1812041_1_1_1_1,00.html, as of February 2004.

114. Lockhart, D. J. and Winzeler, E. A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**(6788): 827-36.

115. Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*, **74**(12): 5350-4.

116. Liang, P. and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**(5072): 967-71.

117. 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. Workshop*. Ministry of Economic Affairs: The Hague, the Netherlands. 37-49.

118. Welsh, J., Chada, K., Dalal, S. S., Cheng, R., Ralph, D., and McClelland, M. (1992) Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res*, **20**(19): 4965-70.

119. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science*, **270**(5235): 484-7.

120. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D.*et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*, **18**(6): 630-4.

121. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**(5235): 467-70.

122. NNT (2003) Use of the cDNA microarray technology in the safety assessment of GM food plants, ed. Nordic Working Group on Food Toxicology and Risk Evaluation (NNT). Nordic Council of Ministers, Copenhagen.

123. Kuiper, H. A., Kok, E. J., and Noteborn, H. P. (2000), Topic 5: Profling techniques to identify differences between foods derived from biotechnology and their Counterparts, in *Joint FAO/WHO expert consultation on foods derived from biotechnology*. FAO/WHA: Geneva.

124. Richmond, T. and Somerville, S. (2000) Chasing the dream: plant EST microarrays. *Curr Opin Plant Biol*, **3**(2): 108-16.

125. Brown, P. O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet*, **21**(1 Suppl): 33-7.

126. Richmond, C. S., Glasner, J. D., Mau, R., Jin, H., and Blattner, F. R. (1999) Genome-wide expression profiling in Escherichia coli K-12. *Nucleic Acids Res*, **27**(19): 3821-35.

127. Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, **18**(3-4): 533-7.

128. Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C.*et al.* (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **100**(6): 3107-12.

129. Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L.*et al.* (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol Cell Proteomics*, **1**(4): 323-33.

130. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*, **19**(3): 1720-30.

131. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I. (1999) A sampling of the yeast proteome. *Mol Cell Biol*, **19**(11): 7357-68.

132. Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L.*et al.* (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep*, **4**(10): 989-93.

133. Trethewey, R. N., Fernie, A. R., Bachmann, A., Fleischer-Notter, H., Geigenberger, P., and Willmitzer, L. (2001) Expression of a bacterial sucrose phosphorylase in potato tubers results in a glucose-independent induction of glycolysis. *Plant Cell Environ*, **24**(3): 357-65.

134. Sonnewald, U., Hajirezaei, M. R., Kossmann, J., Heyer, A., Trethewey, R. N., and Willmitzer, L. (1997) Increased potato tuber size resulting from apoplastic expression of a yeast invertase. *Nat Biotechnol*, **15**(8): 794-7.

135. Fadiel, A. and Naftolin, F. (2003) Microarray applications and challenges: a vast array of possibilities. *International Archives of Bioscience*: 1111-21.

136. Aharoni, A. and Vorst, O. (2002) DNA microarrays for functional plant genomics. *Plant Mol Biol*, **48**(1-2): 99-118.

137. Anderson, N. L. and Anderson, N. G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, **19**(11): 1853-61.

138. Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F.*et al.* (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*, **13**: 19-50.

139. Zivy, M. and de Vienne, D. (2000) Proteomics: a link between genomics, genetics and physiology. *Plant Molecular Biology*, **44**(5): 575-80.

140. Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*, **2**(1): 3-10.

141. Dowsey, A. W., Dunn, M. J., and Yang, G. Z. (2003) The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics*, **3**(8): 1567-96.

142. Gallardo, K., Job, C., Groot, S. P., Puype, M., Demol, H., Vandekerckhove, J.*et al.* (2001) Proteomic analysis of *Arabidopsis* seed germination and priming. *Plant Physiol*, **126**(2): 835-48.

143. Gallardo, K., Job, C., Groot, S. P., Puype, M., Demol, H., Vandekerckhove, J.*et al.* (2002) Proteomics of *Arabidopsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol*, **129**(2): 823-37.

144. Leymarie, J., Damerval, C., Marcotte, L., Combes, V., and Vartanian, N. (1996) Two-dimensional protein patterns of Arabidopsis wild-type and auxin insensitive mutants, axr1, axr2, reveal interactions between drought and hormonal responses. *Plant Cell Physiol*, **37**(7): 966-75.

145. Santoni, V., Delarue, M., Caboche, M., and Bellini, C. (1997) A comparison of two-dimensional electrophoresis data with phenotypical traits in Arabidopsis leads to the identification of a mutant (cri1) that accumulates cytokinins. *Planta*, **202**(1): 62-9.

146. Santoni, V., Rouquie, D., Doumas, P., Mansion, M., Boutry, M., Degand, H.*et al.* (1998) Use of a proteome strategy for tagging proteins present at the plasma membrane. *Plant J*, **16**(5): 633-41.

147. Santoni, V., Rabilloud, T., Doumas, P., Rouquie, D., Mansion, M., Kieffer, S.*et al.* (1999) Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. *Electrophoresis*, **20**(4-5): 705-11.

148. Kehr, J. (2001) High resolution spatial analysis of plant systems. *Curr Opin Plant Biol*, **4**(3): 197-201.

149. Jacobs, D. I., van der Heijden, R., and Verpoorte, R. (2000) Proteomics in plant biotechnology and secondary metabolism research. *Phytochemical Analysis*, **11**(5): 277-87.

150. Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossignol, M., Santoni, V.*et al.* (1999) Proteomics for genetic and physiological studies in plants. *Electrophoresis*, **20**(10): 2013-26.

151. Damerval, C. and Le Guilloux, M. (1998) Characterization of novel proteins affected by the o2 mutation and expressed during maize endosperm development. *Mol Gen Genet*, **257**(3): 354-61.

152. Gottlieb, L. D. and de Vienne, D. (1988) Assessment of Pleiotropic Effects of a Gene Substitution in Pea by Two-Dimensional Polyacrylamide Gel Electrophoresis. *Genetics*, **119**(3): 705-10.

153. Thiellement, H., Zivy, M., and Plomion, C. (2002) Combining proteomic and genetic studies in plants. *J Chromatogr B Analyt Technol Biomed Life Sci*, **782**(1-2): 137-49.

154. Corpillo, D., Gardini, G., Vaira, A. M., Basso, M., Aime, S., Accotto, G. P.*et al.* (2004) Proteomics as a tool to improve investigation of substantial equivalence in genetically modified organisms: The case of a virus-resistant tomato. *Proteomics*, **4**(1): 193-200.

155. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R.*et al.* (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, **17**(7): 676-82.

156. Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., and Yates, J. R., 3rd (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem*, **74**(7): 1650-7.

157. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, **19**(3): 242-7.

158. Washburn, M. P. and Yates, J. R., 3rd (2000) Analysis of the microbial proteome. *Curr Opin Microbiol*, **3**(3): 292-7.

159. Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*, **73**(23): 5683-90.

160. Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A.*et al.* (2002) Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci U S A*, **99**(18): 11969-74.

161. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, **17**(10): 994-9.

162. Islam, N., Tsujimoto, H., and Hirano, H. (2003) Wheat proteomics: relationship between fine chromosome deletion and protein expression. *Proteomics*, **3**(3): 307-16.

163. Smolka, M., Zhou, H., and Aebersold, R. (2002) Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry. *Mol Cell Proteomics*, **1**(1): 19-29.

164. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K.*et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, **292**(5518): 929-34.

165. Roberts, J. K. (2002) Proteomics and a future generation of plant molecular biologists. *Plant Mol Biol*, **48**(1-2): 143-54.

166. Schmidt, F., Donahoe, S., Hagens, K., Mattow, J., Schaible, U. E., Kaufmann, S. H. E.*et al.* (2004) Complementary Analysis of the Mycobacterium tuberculosis Proteome by Two-dimensional Electrophoresis and Isotope-coded Affinity Tag Technology. *Mol Cell Proteomics*, **3**(1): 24-42.

167. Quadroni, M., Ducret, A., and Stocklin, R. (2004) Quantify this! Report on a round table discussion on quantitative mass spectrometry in proteomics. *Proteomics*, **4**(8): 2211-5.

168. Tweeddale, H., Notley-McRobb, L., and Ferenci, T. (1998) Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis. *J Bacteriol*, **180**(19): 5109-16.

169. Weckwerth, W. (2003) Metabolomics in systems biology. *Annu Rev Plant Biol*, **54**: 669-89.

170. Fiehn, O., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal Chem*, **72**(15): 3573-80.

171. Fiehn, O. and Weckwerth, W. (2003) Deciphering metabolic networks. *Eur J Biochem*, **270**(4): 579-88.

172. Sumner, L. W., Mendes, P., and Dixon, R. A. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **62**(6): 817-36.

173. Weckwerth, W., Wenzel, K., and Fiehn, O. (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*, **4**(1): 78-83.

174. Phelps, T. J., Palumbo, A. V., and Beliaev, A. S. (2002) Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Curr Opin Biotechnol*, **13**(1): 20-4.

175. Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L.*et al.* (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, **13**(1): 11-29.

176. Fiehn, O. (2002) Metabolomics: the link between genotypes and phenotypes. *Plant Mol Biol*, **48**(1-2): 155-71.

177. Fiehn, O. (2003) Metabolic networks of Cucurbita maxima phloem. *Phytochemistry*, **62**(6): 875-86.

178. Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., and Willmitzer, L. (2000) Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J*, **23**(1): 131-42.

179. Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol*, **18**(11): 1157-61.

180. Frenzel, T., Miller, A., and Engel, K.-H. (2002) Metabolite Profiling-A Fractionation Method for Analysis of Major and Minor Compounds in Rice Grains. *Cereal Chemistry*, **79**(2): 215-21.

181. Ward, J. L., Harris, C., Lewis, J., and Beale, M. H. (2003) Assessment of 1H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry*, **62**(6): 949-57.

182. Noteborn, H. P., Lommen, A., van der Jagt, R. C., and Weseman, J. M. (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J Biotechnol*, **77**(1): 103-14.

183. Le Gall, G., Colquhoun, I. J., Davis, A. L., Collins, G. J., and Verhoeyen, M. E. (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using 1H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J Agric Food Chem*, **51**(9): 2447-56.

184. ENTRANSFOOD (2004) GMOCARE (QLK1-1999-00765), in http://www.entransfood.com.

185. UK Foods Standards Agency (2002) Safety Assessment of GM Foods research programme (G02), in http://www.food.gov.uk/.

186. Klein, E., Klein, J. B., and Thongboonkerd, V. (2004) Two-dimensional gel electrophoresis: a fundamental tool for expression proteomics studies. *Contrib Nephrol*, **141**: 25-39.

187. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*, **250**(10): 4007-21.

188. Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, **26**(3): 231-43.

189. Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R.*et al.* (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods*, **6**(4): 317-39.

190. Gorg, A., Postel, W., and Gunther, S. (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, **9**(9): 531-46.

191. Patterson, S. D. and Aebersold, R. (1995) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis*, **16**(10): 1791-814.

192. Keough, T., Youngquist, R. S., and Lacey, M. P. (1999) A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc Natl Acad Sci U S A*, **96**(13): 7131-6.

193. Swiderek, K. M., Davis, M. T., and Lee, T. D. (1998) The identification of peptide modifications derived from gel-separated proteins using electrospray triple quadrupole and ion trap analyses. *Electrophoresis*, **19**(6): 989-97.

194. Molloy, M. P. (2000) Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. *Anal Biochem*, **280**(1): 1-10.

195. Righetti, P. G. and Castagna, A. (2003) Recent trends in proteome analysis. *Adv Chromatogr*, **42**: 269-321.

196. Figeys, D. (2003) Proteomics in 2002: a year of technical development and wide-ranging applications. *Anal Chem*, **75**(12): 2891-905.

197. Canovas, F. M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H. P., and Rossignol, M. (2004) Plant proteome analysis. *Proteomics*, **4**(2): 285-98.

198. Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature*, **405**(6788): 837-46.

199. Thongboonkerd, V., Klein, E., and Klein, J. B. (2004) Sample preparation for 2-D proteomic analysis. *Contrib Nephrol*, **141**: 11-24.

200. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R.*et al.* (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, **21**(6): 1037-53.

201. Rabilloud, T. (1996) Solubilization of proteins for electrophoretic analyses. *Electrophoresis*, **17**(5): 813-29.

202. Righetti, P. G., Gelfi, C., and Bossi, M. L. (1987) Hydrophobic interaction between alkaline immobilines and ferritin during isoelectric focusing in immobilized pH gradients. *J Chromatogr*, **392**: 123-32.

203. Kopchick, J. J., List, E. O., Kohn, D. T., Keidan, G. M. O., Qiu, L., and Okada, S. (2002) Perspective: Proteomics--See "Spots" Run. *Endocrinology*, **143**(6): 1990-4.

204. Shaw, M. M. and Riederer, B. M. (2003) Sample preparation for two-dimensional gel electrophoresis. *Proteomics*, **3**(8): 1408-17.

205. Granier, F. (1988) Extraction of plant proteins for two-dimensional electrophoresis. *Electrophoresis*, **9**(11): 712-8.

206. Rabilloud, T., Adessi, C., Giraudel, A., and Lunardi, J. (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, **18**(3-4): 307-16.

207. Henningsen, R., Gale, B. L., Straub, K. M., and DeNagel, D. C. (2002) Application of zwitterionic detergents to the solubilization of integral membrane proteins for two-dimensional gel electrophoresis and mass spectrometry. *Proteomics*, **2**(11): 1479-88.

208. Molloy, M. P., Herbert, B. R., Walsh, B. J., Tyler, M. I., Traini, M., Sanchez, J. C.*et al.* (1998) Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis*, **19**(5): 837-44.

209. Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S.*et al.* (1998) New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis*, **19**(11): 1901-9.

210. Santoni, V., Kieffer, S., Desclaux, D., Masson, F., and Rabilloud, T. (2000) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis*, **21**(16): 3329-44.

211. Herbert, B. R., Molloy, M. P., Gooley, A. A., Walsh, B. J., Bryson, W. G., and Williams, K. L. (1998) Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. *Electrophoresis*, **19**(5): 845-51.

212. Righetti, P. G., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N.*et al.* (2003) The proteome: anno Domini 2002. *Clin Chem Lab Med*, **41**(4): 425-38.

213. Herbert, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis*, **20**(4-5): 660-3.

214. Damerval, C., de Vienne, D., Zivy, M., and Thiellement, H. (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling protein. *Electrophoresis*, **7**: 52-4.

215. Giavalisco, P., Nordhoff, E., Lehrach, H., Gobom, J., and Klose, J. (2003) Extraction of proteins from plant tissues for two-dimensional electrophoresis analysis. *Electrophoresis*, **24**(1-2): 207-16.

216. Carroll, K., Ray, K., Helm, B., and Carey, E. (2000) Two-dimensional electrophoresis reveals differential protein expression in high- and low-secreting variants of the rat basophilic leukaemia cell line. *Electrophoresis*, **21**(12): 2476-86.

217. Storrie, B. and Madden, E. A. (1990) Isolation of subcellular organelles. *Methods Enzymol*, **182**: 203-25.

218. Lopez, M. F. (2000) Better approaches to finding the needle in a haystack: optimizing proteome analysis through automation. *Electrophoresis*, **21**(6): 1082-93.

219. Righetti, P. G., Castagna, A., Herbert, B., Reymond, F., and Rossier, J. S. (2003) Prefractionation techniques in proteome analysis. *Proteomics*, **3**(8): 1397-407.

220. Lilley, K. S., Razzaq, A., and Dupree, P. (2002) Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol*, **6**(1): 46-50.

221. Issaq, H. J., Conrads, T. P., Janini, G. M., and Veenstra, T. D. (2002) Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis*, **23**(17): 3048-61.

222. Gygi, S. P., Rist, B., and Aebersold, R. (2000) Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol*, **11**(4): 396-401.

223. Jacobs, D. I., van Rijssen, M. S., van der Heijden, R., and Verpoorte, R. (2001) Sequential solubilization of proteins precipitated with trichloroacetic acid in acetone from cultured Catharanthus roseus cells yields 52% more spots after two-dimensional electrophoresis. *Proteomics*, **1**(11): 1345-50.

224. Peng, J. and Gygi, S. P. (2001) Proteomics: the move to mixtures. *J Mass Spectrom*, **36**(10): 1083-91.

225. Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., and Sanchez, J. C. (2000) The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis*, **21**(6): 1104-15.

226. Blomberg, A., Blomberg, L., Norbeck, J., Fey, S. J., Larsen, P. M., Larsen, M.*et al.* (1995) Interlaboratory reproducibility of yeast protein patterns analyzed by immobilized pH gradient two-dimensional gel electrophoresis. *Electrophoresis*, **16**(10): 1935-45.

227. Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J. C., and Hochstrasser, D. (1993) A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis*, **14**(12): 1357-65.

228. Gorg, A., Postel, W., Domscheit, A., and Gunther, S. (1988) Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (Hordeum vulgare): method, reproducibility and genetic aspects. *Electrophoresis*, **9**(11): 681-92.

229. Corbett, J. M., Dunn, M. J., Posch, A., and Gorg, A. (1994) Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: an interlaboratory comparison. *Electrophoresis*, **15**(8-9): 1205-11.

230. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A*, **97**(17): 9390-5.

231. Gorg, A., Obermaier, C., Boguth, G., and 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**(4-5): 712-7.

232. Amersham Biosciences (2004) Catalog, in http://www.amershambiosciences.com/.

233. Rabilloud, T., Valette, C., and Lawrence, J. J. (1994) Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis*, **15**(12): 1552-8.

234. Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M.*et al.* (1997) Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. *Electrophoresis*, **18**(3-4): 324-7.

235. Barry, R. C., Alsaker, B. L., Robison-Cox, J. F., and Dratz, E. A. (2003) Quantitative evaluation of sample application methods for semipreparative separations of basic proteins by two-dimensional gel electrophoresis. *Electrophoresis*, **24**(19-20): 3390-404.

236. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**(259): 680-5.

237. Schagger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*, **166**(2): 368-79.

238. Gorg, A., Boguth, G., Obermaier, C., Posch, A., and Weiss, W. (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): the state of the art and the controversy of vertical versus horizontal systems. *Electrophoresis*, **16**(7): 1079-86.

239. Zhan, X. and Desiderio, D. M. (2003) Differences in the spatial and quantitative reproducibility between two second-dimensional gel electrophoresis systems. *Electrophoresis*, **24**(11): 1834-46.

240. Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., MacCarthy, J., Pedersen, S.*et al.* (2001) Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when, and how? *Electrophoresis*, **22**(10): 2046-57.

241. Herbert, B., Hopwood, F., Oxley, D., McCarthy, J., Laver, M., Grinyer, J.*et al.* (2003) Beta-elimination: an unexpected artefact in proteome analysis. *Proteomics*, **3**(6): 826-31.

242. Galvani, M., Rovatti, L., Hamdan, M., Herbert, B., and Righetti, P. G. (2001) Protein alkylation in the presence/absence of thiourea in proteome analysis: a matrix assisted laser desorption/ionization-time of flight-mass spectrometry investigation. *Electrophoresis*, **22**(10): 2066-74.

243. Galvani, M., Hamdan, M., Herbert, B., and Righetti, P. G. (2001) Alkylation kinetics of proteins in preparation for two-dimensional maps: a matrix assisted laser desorption/ionization-mass spectrometry investigation. *Electrophoresis*, **22**(10): 2058-65.

244. Pradet-Balade, B., Boulme, F., Beug, H., Mullner, E. W., and Garcia-Sanz, J. A. (2001) Translation control: bridging the gap between genomics and proteomics? *Trends Biochem Sci*, **26**(4): 225-9.

245. Wirth, P. J. and Romano, A. (1995) Staining methods in gel electrophoresis, including the use of multiple detection methods. *J Chromatogr A*, **698**(1-2): 123-43.

246. Li, K. W., Geraerts, W. P., van Elk, R., and Joosse, J. (1989) Quantification of proteins in the subnanogram and nanogram range: comparison of the AuroDye, FerriDye, and India ink staining methods. *Anal Biochem*, **182**(1): 44-7.

247. Leimgruber, R. M., Malone, J. P., Radabaugh, M. R., LaPorte, M. L., Violand, B. N., and Monahan, J. B. (2002) Development of improved cell lysis, solubilization and imaging approaches for proteomic analyses. *Proteomics*, **2**(2): 135-44.

248. Amersham Biosciences (2002) 2-D Electrophoresis using immobilized pH gradients: Principles and Methods, ed. Handbook. http://www.amershambiosciences.com.

249. Nishihara, J. C. and Champion, K. M. (2002) Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis*, **23**(14): 2203-15.

250. Patton, W. F. (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis*, **21**(6): 1123-44.

251. Rabilloud, T., Vuillard, L., Gilly, C., and Lawrence, J. J. (1994) Silver-staining of proteins in polyacrylamide gels: a general overview. *Cell Mol Biol (Noisy-le-grand)*, **40**(1): 57-75.

252. Berggren, K., Chernokalskaya, E., Steinberg, T. H., Kemper, C., Lopez, M. F., Diwu, Z.*et al.* (2000) Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis*, **21**(12): 2509-21.

253. Wilkins, M. R., Sanchez, J. C., Williams, K. L., and Hochstrasser, D. F. (1996) Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis*, **17**(5): 830-8.

254. Mahon, P. and Dupree, P. (2001) Quantitative and reproducible two-dimensional gel analysis using Phoretix 2D Full. *Electrophoresis*, **22**(10): 2075-85.

255. Ong, S. E. and Pandey, A. (2001) An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol Eng*, **18**(5): 195-205.

256. Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T. (2004) About the mechanism of interference of silver staining with peptide mass spectrometry. *Proteomics*, **4**(4): 909-16.

257. Neuhoff, V., Stamm, R., and Eibl, H. (1985) Clear background and highly sensitive protein staining with Coomassie blue dyes in polyacrylamide gels: a systematic analysis. *Electrophoresis*, **6**: 427- 48.

258. Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, **9**(6): 255-62.

259. Neuhoff, V., Stamm, R., Pardowitz, I., Arold, N., Ehrhardt, W., and Taube, D. (1990) Essential problems in quantification of proteins following colloidal staining with coomassie brilliant blue dyes in polyacrylamide gels, and their solution. *Electrophoresis*, **11**(2): 101-17.

260. Lauber, W. M., Carroll, J. A., and Duffin, K. (1998) Proteomics: Integrating Protein-Based Tools and Applications for Drug Discovery. in *International Business Communications*. Southborough, MA.

261. Steinberg, T. H., Haugland, R. P., and Singer, V. L. (1996) Applications of SYPRO Orange and SYPRO Red Protein Gel Stains. *Analytical Biochemistry*, **239**(2): 238-45.

262. Steinberg, T. H., Jones, L. J., Haugland, R. P., and 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. *Anal Biochem*, **239**(2): 223-37.

263. Lauber, W. M., Carroll, J. A., Dufield, D. R., Kiesel, J. R., Radabaugh, M. R., and Malone, J. P. (2001) Mass spectrometry compatibility of two-dimensional gel protein stains. *Electrophoresis*, **22**(5): 906-18.

264. Mackintosh, J. A., Choi, H. Y., Bae, S. H., Veal, D. A., Bell, P. J., Ferrari, B. C.*et al.* (2003) A fluorescent natural product for ultra sensitive detection of proteins in one-dimensional and two-dimensional gel electrophoresis. *Proteomics*, **3**(12): 2273-88.

265. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J.*et al.* (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, **1**(3): 377-96.

266. Shaw, J., Rowlinson, R., Nickson, J., Stone, T., Sweet, A., Williams, K.*et al.* (2003) Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics*, **3**(7): 1181-95.

267. Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., and Fowler, S. (2002) Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. *Proteomics*, **2**(12): 1682-98.

268. Miller, M. D., Jr., Acey, R. A., Lee, L. Y., and Edwards, A. J. (2001) Digital imaging considerations for gel electrophoresis analysis systems. *Electrophoresis*, **22**(5): 791-800.

269. Miura, K. (2001) Imaging and detection technologies for image analysis in electrophoresis. *Electrophoresis*, **22**(5): 801-13.

270. Raman, B., Cheung, A., and Marten, M. R. (2002) Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. *Electrophoresis*, **23**(14): 2194-202.

271. Rogers, M., Graham, J., and Tonge, R. P. (2003) Using statistical image models for objective evaluation of spot detection in two-dimensional gels. *Proteomics*, **3**(6): 879-86.

272. Westermeier, R. (2001) Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations. third edition ed. WILEY-VCH Verlag, Weinheim.

273. Beavis, R. C. and Chait, B. T. (1996) Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods Enzymol*, **270**: 519-51.

274. Jungblut, P. and Thiede, B. (1997) Protein identification from 2-DE gels by MALDI mass spectrometry. *Mass Spectrom Rev*, **16**(3): 145-62.

275. Karas, M., Gluckmann, M., and Schafer, J. (2000) Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom*, **35**(1): 1-12.

276. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A*, **90**(11): 5011-5.

277. Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O.*et al.* (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci U S A*, **93**(25): 14440-5.

278. Banks, J. F., Jr. and Whitehouse, C. M. (1996) Electrospray ionization mass spectrometry. *Methods Enzymol*, **270**: 486-519.

279. Gillece-Castro, B. L. and Stults, J. T. (1996) Peptide characterization by mass spectrometry. *Methods Enzymol*, **271**: 427-48.

280. Scheible, W.-R., Richmond, T. A., Wilson, I. W., and Somerville, C. R. (2001), *Arabidopsis* genetics and functional genomics in the post-genome era, in *Advances in photosynthesis and respiration: Regulation of photosynthesis (Volume 11)*, E.-M. Aro and B. Andersson, Editors. Kluwer Academic Publishers: Dordrecht; Boston. 563-92.

281. Collakova, E. and DellaPenna, D. (2003) Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in *Arabidopsis*. *Plant Physiology*, **131**(2): 632-42.

282. Tsegaye, Y., Shintani, D., and DellaPenna, D. (2002) Overexpression of the enzyme p-hydroxyphenolpyruvate dioxygenase in Arabidopsis and its relation to tocopherol biosynthesis. *PLANT PHYSIOLOGY AND BIOCHEMISTRY*, **40**(11): 913-20.

283. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**: 248-54.

284. Malone, J. P., Radabaugh, M. R., Leimgruber, R. M., and Gerstenecker, G. S. (2001) Practical aspects of fluorescent staining for proteomic applications. *Electrophoresis*, **22**(5): 919-32.

285. Miller, J. N. and Miller, J. C. (2000) Statistics and Chemometrics for Analytical Chemistry. Fourth ed. Pearson Education Limited, London.

286. Fajgelj, A. and Ambrus, A. (2000) Principles and Practices of Method Validation. Vol. No. 256. The Royal Society of Chemistry, Cambridge, UK.

287. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem*, **68**(5): 850-8.

288. Clauser, K. R., Baker, P., and Burlingame, A. L. (1999) Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem*, **71**(14): 2871-82.

289. UCSF Mass Spectrometry Facility (2004) MS-Fit, in http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm. Sept. 2004.

290. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, **31**(13): 3784-8.

291. ExPASy - Expert Protein Analysis System (2004) Proteomics tools: PeptIdent, in http://www.expasy.org/tools/peptident.html. Sept. 2004.

292. Marques, K., Sarazin, B., Chane-Favre, L., Zivy, M., and Thiellement, H. (2001) Comparative proteomics to establish genetic relationships in the Brassicaceae family. *Proteomics*, **1**(11): 1457-62.

293. Felsenstein, J. (2002) Phylip - Phylogeny Inference Package Version 3.6 (alpha3), in http://evolution.genetics.washington.edu/phylip.html. July, 2002.

294. Meyerowitz, E. M. and Somerville, C. R. (1994) Arabidopsis. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

295. Kamo, M., Kawakami, T., Miyatake, N., and Tsugita, A. (1995) Separation and characterization of *Arabidopsis thaliana* proteins by two-dimensional gel electrophoresis. *Electrophoresis*, **16**(3): 423-30.

296. Heath, J. D., Weldon, R., Monnot, C., and Meinke, D. W. (1986) Analysis of storage proteins in normal and aborted seeds from embryo-lethal mutants of *Arabidopsis thaliana*. *Planta*, **169**: 304-12.

297. Damerval, C., Zivy, M., Granier, F., and de Vienne, D. (1988), Two-dimensional electrophoresis in plant biology, in *Advances in electrophoresis*, A. Chrambach, M.J. Dunn, and B.J. Radola, Editors. Vch: Weinheim (Federal Republic of Germany); New York, NY. 7 v.

298. Gruis, D. F., Selinger, D. A., Curran, J. M., and Jung, R. (2002) Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. *Plant Cell*, **14**(11): 2863-82.

299. Tsugita, A., Kamo, M., Kawakami, T., and Ohki, Y. (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. *Electrophoresis*, **17**(5): 855-65.

300. Tsugita, A. and Kamo, M. (1999) 2-D electrophoresis of plant proteins. *Methods Mol Biol*, **112**: 95-7.

301. Gallardo, K., Job, C., Groot, S. P., Puype, M., Demol, H., Vandekerckhove, J.*et al.* (2002) Importance of methionine biosynthesis for *Arabidopsis* seed germination and seedling growth. *Physiol Plant*, **116**(2): 238-47.

302. Perdew, G. H., Schaup, H. W., and Selivonchick, D. P. (1983) The use of a zwitterionic detergent in two-dimensional gel electrophoresis of trout liver microsomes. *Anal Biochem*, **135**(2): 453-5.

303. Luche, S., Santoni, V., and Rabilloud, T. (2003) Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics*, **3**(3): 249-53.

304. Gorg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J. J., and Madjar, J. J. (1997) Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis*, **18**(3-4): 328-37.

305. Hoving, S., Gerrits, B., Voshol, H., Muller, D., Roberts, R. C., and van Oostrum, J. (2002) Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients. *Proteomics*, **2**(2): 127-34.

306. Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R. D., and Burstin, J. (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol*, **133**(2): 664-82.

307. Mechin, V., Consoli, L., Le Guilloux, M., and Damerval, C. (2003) An efficient solubilization buffer for plant proteins focused in immobilized pH gradients. *Proteomics*, **3**(7): 1299-302.

308. Pietrogrande, M. C., Marchetti, N., Dondi, F., and Righetti, P. G. (2002) Spot overlapping in two-dimensional polyacrylamide gel electrophoresis separations: a statistical study of complex protein maps. *Electrophoresis*, **23**(2): 283-91.

309. Oh-Ishi, M., Satoh, M., and Maeda, T. (2000) Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins. *Electrophoresis*, **21**(9): 1653-69.

310. Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., and Patton, W. F. (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis*, **21**(17): 3673-83.

311. Rabilloud, T. (2000) Detecting proteins separated by 2-D gel electrophoresis. *Anal Chem*, **72**(1): 48A-55A.

312. Bio-Rad 2-D Electrophoresis for Proteomics: A Methods and Product Manual. Vol. Bulletin 2651 US/EG Rev B. Bio-Rad.

313. Molecular Probes (2004) SYPRO(R) Ruby Protein Gel Stain, Product Information. 4.

314. ISO 3534-1 (1993) Statistics - Vocabulary and Symbols.

315. Norbeck, J. and Blomberg, A. (1997) Two-dimensional electrophoretic separation of yeast proteins using a non-linear wide range (pH 3-10) immobilized pH gradient in the first dimension; reproducibility and evidence for isoelectric focusing of alkaline (pI > 7) proteins. *Yeast*, **13**(16): 1519-34.

316. Li, C., Chen, Z., Xiao, Z., Wu, X., Zhan, X., Zhang, X.*et al.* (2003) Comparative proteomics analysis of human lung squamous carcinoma. *Biochem Biophys Res Commun*, **309**(1): 253-60.

317. Zhan, X. and Desiderio, D. M. (2003) A reference map of a human pituitary adenoma proteome. *Proteomics*, **3**(5): 699-713.

318. Burstin, J., Zivy, M., de Vienne, D., and Damerval, C. (1993) Analysis of scaling methods to minimize experimental variations in two-dimensional electrophoresis quantitative data: application to the comparison of maize inbred lines. *Electrophoresis*, **14**(10): 1067-73.

319. Molloy, M. P., Brzezinski, E. E., Hang, J., McDowell, M. T., and VanBogelen, R. A. (2003) Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics*, **3**(10): 1912-9.

320. Sapan, C. V., Lundblad, R. L., and Price, N. C. (1999) Colorimetric protein assay techniques. *Biotechnol Appl Biochem*, **29 (Pt 2)**: 99-108.

321. Appel, R. D., Palagi, P. M., Walther, D., Vargas, J. R., Sanchez, J. C., Ravier, F.*et al.* (1997) Melanie II--a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface. *Electrophoresis*, **18**(15): 2724-34.

322. El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A. J., Raz, V., and Koornneef, M. (2001) A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. *Nat Genet*, **29**(4): 435-40.

323. Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004) Naturally occurring genetic variation in *Arabidopsis thaliana. Annual Review of Plant Biology*, **55**(1): 141-72.

324. Klose, J., Nock, C., Herrmann, M., Stuhler, K., Marcus, K., Bluggel, M.*et al.* (2002) Genetic analysis of the mouse brain proteome. *Nat Genet*, **30**(4): 385-93.

325. Riccardi, F., Gazeau, P., de Vienne, D., and Zivy, M. (1998) Protein changes in response to progressive water deficit in maize - Quantitative variation and polypeptide identification. *Plant Physiology*, **117**(4): 1253-63.

326. Costa, P., Bahrman, N., Frigerio, J. M., Kremer, A., and Plomion, C. (1998) Water-deficit-responsive proteins in maritime pine. *Plant Mol Biol*, **38**(4): 587-96.

327. Salekdeh, G. H., Siopongco, J., Wade, L. J., Ghareyazie, B., and Bennett, J. (2002) Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics*, **2**(9): 1131-45.

328. Bahrman, N., Le Gouis, J., Negroni, L., Amilhat, L., Leroy, P., Laine, A. L.*et al.* (2004) Differential protein expression assessed by two-dimensional gel electrophoresis for two wheat varieties grown at four nitrogen levels. *Proteomics*, **4**(3): 709-19.

329. European Food Safety Authority (2004) Draft guidance document for the risk assessment of genetically modified plants and derived food and feed, S.P.o.G.M.O.o.t.E.F.S. Authority, Editor. European Food Safety Authority (EFSA). http://www.efsa.eu.int/consultation/372/consultation_guidance_gmo_01_en1.pdf.

330. ILSI International Food Biotechnology Committee (2004) Nutritional and Safety Assessments of Foods and Feeds Nutritionally Improved through Biotechnology. *Comprehensive Reviews in Food Science and Food Safety*. Vol. 3. Institute of Food Technologists.

331. Kawasaki, S., Miyake, C., Kohchi, T., Fujii, S., Uchida, M., and Yokota, A. (2000) Responses of wild watermelon to drought stress: accumulation of an ArgE homologue and citrulline in leaves during water deficits. *Plant Cell Physiol*, **41**(7): 864-73.

332. Majoul, T., Bancel, E., Triboi, E., Ben Hamida, J., and Branlard, G. (2003) Proteomic analysis of the effect of heat stress on hexaploid wheat grain: Characterization of heat-responsive proteins from total endosperm. *Proteomics*, **3**(2): 175-83.

333. Renaut, J., Lutts, S., Hoffmann, L., and Hausman, J. F. (2004) Responses of poplar to chilling temperatures: proteomic and physiological aspects. *Plant Biol (Stuttg)*, **6**(1): 81-90.

334. Tafforeau, M., Verdus, M. C., Charlionet, R., Cabin-Flaman, A., and Ripoll, C. (2002) Two-dimensional electrophoresis investigation of short-term response of flax seedlings to a cold shock. *Electrophoresis*, **23**(15): 2534-40.

335. Kollipara, K. P., Saab, I. N., Wych, R. D., Lauer, M. J., and Singletary, G. W. (2002) Expression profiling of reciprocal maize hybrids divergent for cold germination and desiccation tolerance. *Plant Physiol*, **129**(3): 974-92.

336. Bae, M. S., Cho, E. J., Choi, E. Y., and Park, O. K. (2003) Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. *Plant J*, **36**(5): 652-63.

337. Agrawal, G. K., Rakwal, R., Yonekura, M., Kubo, A., and Saji, H. (2002) Proteome analysis of differentially displayed proteins as a tool for investigating ozone stress in rice (*Oryza sativa* L.) seedlings. *Proteomics*, **2**(8): 947-59.

338. Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y.*et al.* (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J*, **31**(3): 279-92.

339. Monteiro, S., Picarra-Pereira, M. A., Teixeira, A. R., Loureiro, V. B., and Ferreira, R. B. (2003) Environmental Conditions during Vegetative Growth Determine the Major Proteins That Accumulate in Mature Grapes. *J Agric Food Chem*, **51**(14): 4046-53.

340. Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, A. J., Hoffman, N. E., Davis, K. R.*et al.* (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell*, **13**(7): 1499-510.

341. Cooke, R. J. (1995) Gel electrophoresis for the identification of plant varieties. *Journal of Chromatography A*, **698**(1-2): 281-99.

342. Murthy, U. M., Kumar, P. P., and Sun, W. Q. (2003) Mechanisms of seed ageing under different storage conditions for Vigna radiata (L.) Wilczek: lipid peroxidation, sugar hydrolysis, Maillard reactions and their relationship to glass state transition. *J Exp Bot*, **54**(384): 1057-67.

343. Murthy, U. M., Liang, Y., Kumar, P. P., and Sun, W. Q. (2002) Non-enzymatic protein modification by the Maillard reaction reduces the activities of scavenging enzymes in Vigna radiata. *Physiol Plant*, **115**(2): 213-20.

344. Murthy, U. M. and Sun, W. Q. (2000) Protein modification by Amadori and Maillard reactions during seed storage: roles of sugar hydrolysis and lipid peroxidation. *J Exp Bot*, **51**(348): 1221-8.

345. Gorg, A., Postel, W., Baumer, M., and Weiss, W. (1992) Two-dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: discrimination of cultivars with different malting grades. *Electrophoresis*, **13**(4): 192-203.

346. Ostergaard, O., Melchior, S., Roepstorff, P., and Svensson, B. (2002) Initial proteome analysis of mature barley seeds and malt. *Proteomics*, **2**(6): 733-9.

347. Finnie, C., Melchior, S., Roepstorff, P., and Svensson, B. (2002) Proteome analysis of grain filling and seed maturation in barley. *Plant Physiol*, **129**(3): 1308-19.

348. Higginbotham, J. W., Smith, J. S., and Smith, O. S. (1991) Quantitative analysis of two-dimensional protein profiles of inbred lines of maize (*Zea mays* L.). *Electrophoresis*, **12**(6): 425-31.

349. Burstin, J., Devienne, D., Dubreuil, P., and Damerval, C. (1994) Molecular Markers and Protein Quantities as Genetic Descriptors in Maize.1. Genetic Diversity among 21 Inbred Lines. *Theoretical and Applied Genetics*, **89**(7-8): 943-50.

350. de Vienne, D., Bost, B., Fievet, J., Zivy, M., and Dillmann, C. (2001) Genetic variability of proteome expression and metabolic control. *Plant Physiology and Biochemistry*, **39**(3-4): 271-83.

351. Consoli, L. and Damerval, C. (2001) Quantification of individual zein isoforms resolved by two-dimensional electrophoresis: genetic variability in 45 maize inbred lines. *Electrophoresis*, **22**(14): 2983-9.

352. Zivy, M., Thiellement, H., de Vienne, D., and Hofmann, J.-P. (1983) Study on nuclear and cytoplasmic genom expression in wheat by two-dimensional gel electrophoresis. 1.First results on 18 alloplasmic lines. *Theoretical and Applied Genetics*, **66**: 1-7.

353. Marlow, G. C., Wurst, D. E., and Loschke, D. C. (1988) The use of ultrathin-layer polyacrylamide gel isoelectric focusing in two-dimensional analysis of plant and fungal proteins. *Electrophoresis*, **9**(11): 693-704.

354. Weiss, W., Huber, G., Engel, K. H., Pethran, A., Dunn, M. J., Gooley, A. A.*et al.* (1997) Identification and characterization of wheat grain albumin/globulin allergens. *Electrophoresis*, **18**(5): 826-33.

355. David, J. L., Zivy, M., Cardin, M. L., and Brabant, P. (1997) Protein evolution in dynamically managed populations of wheat: adaptive responses to macro-environmental conditions. *Theoretical and Applied Genetics*, **95**(5-6): 932-41.

356. Jacobsen, S., Nesic, L., Petersen, M., and Sondergaard, I. (2001) Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization-time of flight-mass spectrometry and an artificial neural network. *Electrophoresis*, **22**(6): 1242-5.

357. Basha, S. M. M. (1979) Identification of cultivar differences in seed polypeptide composition of peanuts (*Arachis hypogaea* L.) by two-dimensional polyacrylamide gel electrophoresis. *Plant Physiol*, **63**: 301-6.

358. Abe, T., Gusti, R. S., Ono, M., and Sasahara, T. (1996) Variations in glutelin and high molecular weight endosperm proteins among subspecies of rice (*Oryza sativa* L.) detected by two-dimensional gel electrophoresis. *Genes Genet Syst*, **71**(2): 63-8.

359. Chevalier, F., Martin, O., Rofidal, V., Devauchelle, A. D., Barteau, S., Sommerer, N.*et al.* (2004) Proteomic investigation of natural variation between *Arabidopsis* ecotypes. *Proteomics*, **4**(5): 1372-81.

360. Sharbel, T. F., Haubold, B., and Mitchell-Olds, T. (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol Ecol*, **9**(12): 2109-18.

361. King, G., Nienhuis, J., and Hussey, C. (1993) Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphism. *Theoretical and Applied Genetics*, **86**: 1028-32.

362. Erschadi, S., Haberer, G., Schoniger, M., and Torres-Ruiz, R. (2000) Estimating genetic diversity of *Arabidopsis thaliana* ecotypes with amplified fragment length polymorphisms (AFLP). *Theoretical and Applied Genetics*, **100**(3-4): 633-40.

363. Miyashita, N. T., Kawabe, A., and Innan, H. (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics*, **152**(4): 1723-31.

364. Baumbusch, L., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R.*et al.* (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Research*, **29**(21): 4319-33.

365. Jungblut, P. R., Bumann, D., Haas, G., Zimny-Arndt, U., Holland, P., Lamer, S.*et al.* (2000) Comparative proteome analysis of *Helicobacter pylori*. *Mol Microbiol*, **36**(3): 710-25.

366. Finnie, C., Maeda, K., O, O. S., Bak-Jensen, K. S., Larsen, J., and Svensson, B. (2004) Aspects of the barley seed proteome during development and germination. *Biochem Soc Trans*, **32**(Pt3): 517-9.

367. Schlesier, B., Berna, A., Bernier, F., and Mock, H.-P. (2004) Proteome analysis differentiates between two highly homologues germin-like proteins in *Arabidopsis thaliana* ecotypes Col-0 and Ws-2. *Phytochemistry*, **65**(11): 1565-74.

368. Anderson, N. G., Tollaksen, S. L., Pascoe, F., H., and Anderson, L. (1985) Two-dimensional electrophoretic analysis of wheat seed proteins. *Crop Science*, **25**: 667-74.

369. Bahrman, N., Zivy, M., and Thiellement, H. (1988) Genetic relationships in the Sitopsis section of *Triticum* and the origin of the B genome of polyploid wheats. *Heredity*, **61**: 473-80.

370. Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J.*et al.* (2001) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol*, **126**(2): 811-25.

371. Jefferson, R. A. (1987) Assaying Chimeric Genes in Plants: The GUS Gene Fusion System. *Plant Molecular Biology Reporter*, **5**(4): 387-405.

372. Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *Embo J*, **6**(13): 3901-7.

373. Pridmore, R. D. (1987) New and versatile cloning vectors with kanamycin-resistance marker. *Gene*, **56**(2-3): 309-12.

374. Bevan, M., Barnes, W. M., and Chilton, M. D. (1983) Structure and transcription of the nopaline synthase gene region of T-DNA. *Nucleic Acids Res*, **11**(2): 369-85.

375. Fraley, R. T., Rogers, S. G., Horsch, R. B., Sanders, P. R., Flick, J. S., Adams, S. P.*et al.* (1983) Expression of bacterial genes in plant cells. *Proc Natl Acad Sci U S A*, **80**(15): 4803-7.

376. Peach, C. and Velten, J. (1991) Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. *Plant Mol Biol*, **17**(1): 49-60.

377. Dean, C., Jones, J., Favreau, M., Dunsmuir, P., and Bedbrook, J. (1988) Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. *Nucleic Acids Res*, **16**(19): 9267-83.

378. Van Leeuwen, W., Mlynarova, L., Nap, J. P., van der Plas, L. H., and van der Krol, A. R. (2001) The effect of MAR elements on variation in spatial and temporal regulation of transgene expression. *Plant Mol Biol*, **47**(4): 543-54.

379. Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem*, **70**: 437-73.

380. Odell, J. T., Nagy, F., and Chua, N. H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature*, **313**(6005): 810-2.

381. DellaPenna, D. (2003) Personal communication: Description to tocopherol enhanced transgenic *Arabidopsis* lines.

382. Shintani, D. and DellaPenna, D. (1998) Elevating the vitamin E content of plants through metabolic engineering. *Science*, **282**(5396): 2098-100.

383. DellaPenna, D. (2003) Personal Communication: Tocopherol composition and content data.

384. Garcia, I., Rodgers, M., Lenne, C., Rolland, A., Sailland, A., and Matringe, M. (1997) Subcellular localization and purification of a p-hydroxyphenylpyruvate dioxygenase from cultured carrot cells and characterization of the corresponding cDNA. *Biochemical Journal*, **325**: 761-9.

385. Norris, S., Barrette, T., and DellaPenna, D. (1995) Genetic dissection of carotenoid synthesis in arabidopsis defines plastoquinone as an essential component of phytoene desaturation. *Plant Cell*, **7**(12): 2139-49.

386. Garcia, I., Rodgers, M., Pepin, R., Hsieh, T., and Matringe, M. (1999) Characterization and subcellular compartmentation of recombinant 4-hydroxyphenylpyruvate dioxygenase from *Arabidopsis* in transgenic tobacco. *Plant Physiology*, **119**(4): 1507-16.

387. Arango, Y. and Heise, K. P. (1998) Tocopherol synthesis from homogentisate in *Capsicum anuum* L. (yellow pepper) chromoplast membranes: evidence for tocopherol cyclase. *Biochem J*, **336 (Pt 3)**: 531-3.

Appendix 1: Spreadsheet for the Jaccard index calculation

Ecotype A	Ecotype B	Present only in ecotype A (N_{A0})	Present only in ecotype B (N_{0B})	Present in ecotype A and B (N_{AB})	Jaccard index
Ma-0	Mt-0	21	9	565	0.05
Oy-0	Ws	49	53	590	0.15
Nd-0	Oy-0	59	69	570	0.18
Ll-0	Ma-0	67	59	527	0.19
Oy-0	Tsu-0	96	34	543	0.19
Ll-0	Mt-0	78	58	516	0.21
Nd-0	Ws	68	82	561	0.21
Col-0	Nd-0	62	84	545	0.21
Condara	Nd-0	88	64	565	0.21
Nd-0	Tsu-0	98	46	531	0.21
Condara	Ws	84	74	569	0.22
Col-0	Oy-0	60	92	547	0.22
Ll-0	Oy-0	53	98	541	0.22
Ma-0	Nd-0	56	99	530	0.23
Col-0	Condara	58	104	549	0.23
Ll-0	Nd-0	62	97	532	0.23
Condara	Oy-0	91	77	562	0.23
C24	Nd-0	94	73	556	0.23
C24	Ws	88	81	562	0.23
Col-0	Ll-0	85	72	522	0.23
Mt-0	Tsu-0	74	77	500	0.23
Col-0	Ma-0	89	68	518	0.23
C24	Oy-0	91	80	559	0.23
Ma-0	Oy-0	57	110	529	0.24
Mt-0	Nd-0	55	110	519	0.24
C24	Condara	88	91	562	0.24
Mt-0	Oy-0	51	116	523	0.24
Tsu-0	Ws	51	117	526	0.24
Ma-0	Tsu-0	85	76	501	0.24
Col-0	Tsu-0	97	67	510	0.24
Col-0	Ws	69	105	538	0.24
Col-0	Mt-0	99	66	508	0.25
Ll-0	Tsu-0	91	74	503	0.25
Ll-0	Ws	63	112	531	0.25
C24	Col-0	112	69	538	0.25
Ma-0	Ws	61	118	525	0.25
C24	Ll-0	119	63	531	0.26
Mt-0	Ws	58	127	516	0.26
Condara	Tsu-0	132	56	521	0.27
C24	Ma-0	127	63	523	0.27
C24	Tsu-0	132	59	518	0.27
Condara	Ll-0	127	68	526	0.27
Condara	Ma-0	134	67	519	0.28

Ecotype A	Ecotype B	Present only in ecotype A (N_{A0})	Present only in ecotype B (N_{0B})	Present in ecotype A and B (N_{AB})	Jaccard index
Col-0	Mr-0	90	113	517	0.28
C24	Mt-0	140	64	510	0.29
Condara	Mt-0	143	64	510	0.29
Ll-0	Mr-0	87	123	507	0.29
Mr-0	Oy-0	105	114	525	0.29
Mr-0	Nd-0	111	110	519	0.30
Cvi-0	Nd-0	94	122	507	0.30
C24	Mr-0	125	105	525	0.30
Ma-0	Mr-0	89	133	497	0.31
Condara	Mr-0	129	106	524	0.31
Mr-0	Ws	111	124	519	0.31
Mr-0	Tsu-0	140	87	490	0.32
Cvi-0	Tsu-0	123	99	478	0.32
Cvi-0	Oy-0	98	136	503	0.32
Mr-0	Mt-0	143	87	487	0.32
Cvi-0	Ll-0	118	111	483	0.32
Cvi-0	Ma-0	128	113	473	0.34
Col-0	Cvi-0	126	120	481	0.34
Cvi-0	Ws	106	148	495	0.34
Cvi-0	Mt-0	134	107	467	0.34
Cvi-0	Mr-0	114	143	487	0.35
Condara	Cvi-0	158	106	495	0.35
C24	Cvi-0	160	111	490	0.36

Appendix 2: Ecotype study - Quantitative data (IOD) (931 reproducible spots)

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 0007	17.5	< 4.5	476	612	433	590	461	678	434	737	553	466	458	621
SSP 0008	13.8	< 4.5	248	362	.	248	305	377	.	424	388	193	274	258
SSP 0009	21.9	< 4.5	251	180	148	239	141	161	60	207	173	114	144	185
SSP 0010	7.5	< 4.5	953	1239	811	1088	665	1211	967	1219	1181	650	724	826
SSP 0013	18.2	< 4.5	328	481	94	177	604	142	367	188	255	362	527	386
SSP 0015	18.2	< 4.5	269	278	215	71	219	267	235	256	271	188	260	294
SSP 0018	15.4	< 4.5	437	318	268	292	171	260	154	283	302	382	243	314
SSP 0019	18.1	4.5	236	376	281	524	189	448	638	301	288	503	130	423
SSP 0020	6.8	4.7	383	835	926	281	.	725	295	565	927	398	635	669
SSP 0022	5.5	< 4.5	524	.	.	.	.	.	.	.	.	.	.	.
SSP 0025	12.8	< 4.5	.	.	443	.	.	.	580	.	.	.	.	.
SSP 0027	15.1	4.6	54	78	115	145	93	144	147	155	141	111	72	117
SSP 0102	21.9	< 4.5	175	308	221	299	92	186	253	191	244	150	107	169
SSP 0103	23.8	< 4.5	492	656	563	653	440	556	511	599	528	447	457	601
SSP 0104	26.9	< 4.5	118	180	125	186	105	126	107	148	121	94	98	129
SSP 0105	22.0	4.6	156	260	241	204	166	137	193	176	190	148	142	185
SSP 0106	26.8	4.6	.	.	.	6148	.	.	.	.	.	.	.	.
SSP 0107	25.6	4.6	.	.	.	1284	.	.	.	.	.	.	.	.
SSP 0207	28.9	< 4.5	121	156	102	126	74	130	114	133	182	52	98	110
SSP 0208	31.8	< 4.5	.	109	95	110	54	109	89	122	104	.	.	110
SSP 0209	32.0	< 4.5	600	776	543	796	481	580	677	679	635	558	556	799
SSP 0211	28.1	< 4.5	35	73	61	54	47	53	36	57	88	80	45	51
SSP 0213	27.4	< 4.5	30	51	39	63	93	29	53	34	40	48	37	50
SSP 0214	31.9	4.5	160	187	187	163	105	137	159	129	130	131	82	171
SSP 0215	30.9	4.6	161	337	252	320	222	237	288	223	242	197	173	264
SSP 0216	28.8	4.6	41	72	41	39	32	38	27	34	40	31	20	48
SSP 0218	30.2	4.6	89	94	69	115	83	55	120	59	60	45	53	67
SSP 0220	28.8	4.7	62	64	52	55	36	44	47	31	54	39	21	68
SSP 0221	31.3	4.6	102	.	.	.	.	.	.	.	.	.	.	.
SSP 0223	27.8	4.6	62	167	55	569	206	130	402	139	130	114	237	120
SSP 0226	31.8	< 4.5	.	.	.	.	.	.	60	.	.	.	.	.
SSP 0302	34.6	< 4.5	70	80	61	90	48	57	74	79	73	53	59	51
SSP 0307	34.0	< 4.5	412	486	399	495	336	409	520	411	398	340	380	481
SSP 0309	36.7	< 4.5	91	170	.	.	72	116	.	126	.	153	.	.
SSP 0311	34.3	< 4.5	48	66	73	85	38	47	58	50	46	45	96	68
SSP 0312	32.4	4.5	71	168	93	139	105	121	94	119	118	76	90	107
SSP 0313	34.7	4.5	.	521	428	403	308	399	385	405	275	328	289	460
SSP 0315	36.9	4.5	76	142	120	157	64	72	101	76	101	68	80	101
SSP 0316	33.1	4.6	81	140	76	83	94	102	90	98	102	68	53	101
SSP 0318	33.0	4.7	210	262	185	305	174	211	241	182	153	133	516	216
SSP 0319	34.7	4.6	278	.	.	.	.	.	.	.	.	.	.	.
SSP 0321	36.9	4.5	.	.	113	145	.	.	93	.	127	44	92	137
SSP 0323	34.9	< 4.5	117	208	40	55	132	31	56	37	130	23	24	38
SSP 0324	35.1	4.7	60	115	55	64	38	50	59	54	.	52	.	62
SSP 0325	34.8	4.7	31	.	53	67	41	44	136	44	85	51	56	79
SSP 0326	35.2	4.6	.	.	33	63	51	40	.	49	57	.	.	60
SSP 0327	32.5	< 4.5	145	297	165	291	140	291	193	264	236	62	210	201

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 0328	37.5	4.7	.	.	64	.	.	.	41	.	.	.	38	.
SSP 0329	32.5	4.6	55	156	92	53	76	87	32	93	105	66	37	100
SSP 0401	40.9	< 4.5	51	125	57	108	47	82	87	100	66	58	68	60
SSP 0402	37.1	< 4.5	58	80	52	56	26	70	50	77	44	37	37	55
SSP 0403	41.8	< 4.5	169	152	137	197	121	169	112	187	163	109	133	133
SSP 0404	39.7	< 4.5	.	67	.	.	.	.	.	.	.	.	.	.
SSP 0409	40.5	4.6	519	.	.	542	607	637	669	692	.	288	.	753
SSP 0412	42.8	4.6	.	.	.	.	.	.	126	.	.	.	.	.
SSP 0415	38.4	4.7	.	.	597	.	.	.	.	.	563	225	440	.
SSP 0416	41.7	4.6	.	.	26	.	37	.	.	.	.	.	.	.
SSP 0418	38.9	4.5	81	.	136	195	.	.	67	.	128	49	143	61
SSP 0419	40.5	4.5	.	.	34	.	.	.	.	.	.	.	.	.
SSP 0420	40.7	< 4.5	.	.	.	39	.	.	41	.	.	.	.	.
SSP 0421	39.2	4.6	.	171	61	.	.	.	132	.	.	.	.	.
SSP 0503	47.9	4.5	208	255	217	276	164	211	185	230	192	146	147	195
SSP 0504	42.7	4.7	58	80	45	71	52	47	41	49	44	44	28	54
SSP 0506	45.5	< 4.5	20	29	18	26	16	21	20	20	21	11	15	21
SSP 0602	51.9	< 4.5	290	330	394	552	301	333	424	363	344	336	292	459
SSP 0603	59.6	< 4.5	1071	1322	1162	1248	1064	1152	1153	1311	1166	1098	859	1436
SSP 0604	50.7	< 4.5	126	156	123	169	96	119	128	142	125	90	103	122
SSP 0607	52.0	< 4.5	.	92	63	.	.	83	87	71	.	.	.	102
SSP 0608	57.9	< 4.5	24	30	17	35	21	22	29	24	.	12	21	18
SSP 0616	52.0	< 4.5	75	.	.	113	.	.	.	.	71	.	.	.
SSP 0702	62.2	< 4.5	46	82	47	.	45	61	69	64	66	47	52	51
SSP 0708	64.1	4.6	38	53	36	62	41	38	59	31	30	23	19	37
SSP 0709	61.7	4.6	2592	3181	2537	752	3256	3383	3121	3768	3397	3173	2956	3397
SSP 0711	63.8	4.5	119	52	91	84	.	75	.	95	122	94	63	96
SSP 0715	61.7	4.6	.	.	.	4540	.	.	.	.	.	.	.	.
SSP 0717	63.7	4.7	81	73	76	100	67	58	87	71	89	93	73	154
SSP 0720	60.5	< 4.5	26	39	34	.	23	25	32	33	36	27	24	41
SSP 0802	105.0	< 4.5	239	227	153	358	184	146	283	163	261	132	118	175
SSP 0806	76.5	4.6	305	493	386	366	355	276	525	393	315	278	270	467
SSP 0808	96.7	4.7	54	81	51	73	56	55	102	45	53	38	36	79
SSP 0809	78.3	4.7	.	46	.	.	.	.	.	.	.	.	.	.
SSP 1003	19.6	4.8	256	336	346	347	346	236	320	269	378	449	201	623
SSP 1004	13.0	4.8	683	758	688	833	744	1003	792	982	765	654	948	761
SSP 1005	11.2	4.8	1010	1476	1142	941	1097	1451	1701	1453	1218	1112	1095	1164
SSP 1007	9.3	4.8	1950	2596	2742	1755	2876	2694	8011	2890	2866	1513	1518	2177
SSP 1009	20.9	4.9	66	126	93	107	99	95	83	119	84	83	84	111
SSP 1011	21.9	4.9	.	141	66	.	.	207	.	220	.	51	64	99
SSP 1013	11.7	4.9	1728	2034	1606	1034	2200	2276	.	2266	1943	1551	1500	1813
SSP 1014	7.2	4.9	1606	4761	3835	2322	1303	2996	1587	3442	3827	1862	3358	2718
SSP 1015	19.7	4.9	433	678	616	537	614	555	511	596	546	710	366	998
SSP 1101	26.6	4.7	3498	4025	2943	.	2896	4179	3319	4046	2595	3024	2406	3971
SSP 1102	25.3	4.7	772	865	601	.	553	879	713	882	564	593	554	643
SSP 1103	24.9	4.7	681	628	380	71	474	599	589	566	536	513	473	608
SSP 1104	26.0	4.7	1976	1516	1290	.	1566	1705	1651	1852	1821	1041	917	1077
SSP 1107	25.0	4.8	132	231	213	274	180	162	208	166	174	209	204	266
SSP 1109	22.0	4.9	103	137	133	155	90	114	211	123	131	83	121	111
SSP 1110	26.1	4.9	215	278	255	161	77	307	147	397	445	241	126	268

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 1111	24.9	4.9	546	837	596	233	240	659	407	774	1118	474	351	366
SSP 1112	22.0	4.9	1919	1873	1787	562	1066	785	737	771	683	1224	.	1605
SSP 1114	27.1	4.8	.	.	.	624	.	.	.	.	.	.	.	.
SSP 1115	27.1	4.9	.	.	.	500	.	.	.	.	.	.	.	.
SSP 1116	23.6	4.8	.	.	90	.	.	.	.	.	.	.	.	.
SSP 1117	21.9	4.9	.	.	.	87	62	.	.	.	92	.	.	.
SSP 1118	23.9	4.8	.	.	86	.	.	.	.	.	.	.	.	.
SSP 1119	24.0	5.0	59	83	.	108	56	72	100	75	83	66	92	75
SSP 1204	29.7	4.9	.	7826	.	.	.	.	.	.	.	4501	.	.
SSP 1208	31.9	4.9	338	398	187	.	252	242	441	237	301	239	192	424
SSP 1209	29.3	4.9	.	610	.	10219	.	.	94	.	.	187	.	.
SSP 1211	31.8	4.8	.	.	.	198	.	.	.	.	.	.	.	.
SSP 1212	29.6	4.9	.	.	.	409	5852	7954	.	8960	.	.	6764	.
SSP 1213	28.1	4.9	.	.	.	.	.	.	.	.	784	.	.	.
SSP 1214	28.8	5.0	1331	2540	2136	777	2521	1612	1075	2308	1073	1745	880	4153
SSP 1303	32.8	4.8	3963	4769	4809	4667	4301	4165	4143	5141	4813	4531	4269	5243
SSP 1305	35.3	4.8	111	123	171	95	76	118	77	144	133	77	43	113
SSP 1312	36.1	4.9	254	205	214	207	144	178	243	260	192	120	162	205
SSP 1314	32.7	4.9	1445	1449	1261	1595	1032	1242	1003	1309	1222	1273	880	1670
SSP 1317	37.3	4.9	.	.	.	258	.	.	.	.	.	.	.	.
SSP 1319	34.0	4.9	.	.	.	.	.	.	1408	.	.	.	.	.
SSP 1320	34.7	4.9	154	124	106	141	158	146	104	162	143	72	103	275
SSP 1401	39.2	4.7	.	929	72	110	60	79	.	75	93	65	80	76
SSP 1402	37.9	4.7	42	78	185	40	48	.	124	.	40	245	145	52
SSP 1403	38.0	4.8	70	107	176	159	62	94	87	98	.	69	.	94
SSP 1404	37.0	4.8	100	122	24	97	55	91	95	88	96	68	48	95
SSP 1405	37.8	4.8	116	194	99	232	129	120	.	113	163	119	.	139
SSP 1406	37.6	4.8	126	147	118	.	91	127	112	130	140	110	109	141
SSP 1408	38.4	4.9	363	304	333	309	205	341	325	277	340	278	137	337
SSP 1413	38.4	4.9	327	356	376	397	204	315	336	279	357	258	143	404
SSP 1414	38.9	4.8	.	.	41	64	.	.	63	.	.	.	.	.
SSP 1415	41.7	4.9	.	.	.	66	.	.	95	.	.	.	.	.
SSP 1416	39.0	4.9	.	.	.	.	.	.	.	.	.	.	190	.
SSP 1417	39.1	5.0	.	.	240	66	.	.	343	.	.	.	.	.
SSP 1418	38.4	4.9	.	.	.	.	.	.	.	.	.	.	.	213
SSP 1419	38.3	4.8	.	.	.	.	.	.	.	.	.	.	.	113
SSP 1501	42.6	4.7	88	111	86	117	.	82	92	93	80	68	64	90
SSP 1502	46.4	4.8	32	37	35	34	25	32	36	32	29	25	29	33
SSP 1503	42.3	4.8	61	85	83	164	21	142	39	132	.	79	.	183
SSP 1504	47.8	4.8	144	172	141	201	117	130	151	162	141	114	122	125
SSP 1505	42.5	4.9	104	108	79	90	61	125	63	107	85	32	67	.
SSP 1507	49.7	4.9	202	172	176	296	114	94	139	98	140	96	90	134
SSP 1509	45.3	4.9	.	.	.	.	.	.	.	.	76	.	.	.
SSP 1510	43.2	4.8	.	.	.	.	.	.	90	.	.	.	.	.
SSP 1511	45.3	4.9	33	35	28	.	72	.	47	.	28	17	15	34
SSP 1601	51.8	4.7	67	102	73	96	63	58	62	77	61	197	.	208
SSP 1602	60.2	4.7	54	66	.	53	37	48	70	50	43	50	24	73
SSP 1603	51.6	4.8	145	144	108	338	86	88	115	.	85	.	95	.
SSP 1605	51.8	4.8	207	145	195	.	142	171	205	168	.	137	150	.
SSP 1609	51.4	4.8	218	195	225	299	175	153	230	153	179	155	154	186

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 1611	54.6	4.9	.	427	.	.	.	.	.	.	.	128	.	.
SSP 1614	55.0	4.9	191	225	210	392	151	196	238	191	172	141	145	179
SSP 1617	54.7	4.9	49	76	110	72	40	44	71	.	74	111	.	56
SSP 1618	60.2	4.9	48	91	48	59	57	61	60	40	57	18	57	58
SSP 1620	60.0	4.9	27	19	27	559	25	19	39	16	21	15	.	27
SSP 1621	56.5	4.9	.	.	.	.	247	340	.	506	.	.	.	.
SSP 1622	53.1	4.8	193	124	103	302	109	99	171	135	242	93	105	339
SSP 1626	52.5	4.9	52	.	30	.	35	.	.	.	46	.	.	49
SSP 1627	60.2	5.0	2934	.	3491	4014	2368	2640	4936	3030	2342	2467	2363	3321
SSP 1628	52.7	4.9	.	.	.	.	.	.	45	.	.	.	.	.
SSP 1629	51.9	4.9	.	.	.	.	.	44	.	54	.	.	.	.
SSP 1701	62.4	4.7	420	618	535	724	.	500	672	627	439	524	460	827
SSP 1703	63.6	4.7	148	170	205	195	145	128	198	174	188	201	187	383
SSP 1704	61.6	4.7	152	230	175	253	134	165	232	154	155	164	134	218
SSP 1705	68.1	4.8	36	54	35	.	26	.	.	.	37	23	22	36
SSP 1707	72.2	4.9	122	117	360	172	110	132	180	104	162	78	94	198
SSP 1708	62.8	4.9	90	112	85	88	69	86	87	82	84	55	56	80
SSP 1709	71.5	4.9	309	208	.	335	173	222	397	200	307	173	154	334
SSP 1711	62.4	4.9	136	133	127	112	103	97	232	124	113	102	81	147
SSP 1713	61.5	4.9	85	.	.	89	.	.	.	.	.	.	.	.
SSP 1714	61.3	5.0	659	.	653	871	510	511	978	641	456	560	563	664
SSP 1717	68.3	4.8	.	.	.	37	.	32	38	134	.	.	.	.
SSP 1802	78.3	4.7	51	100	66	90	69	56	111	66	56	55	50	70
SSP 1805	122.6	4.8	188	191	177	262	158	155	239	190	133	183	.	.
SSP 1806	92.9	4.9	214	205	281	170	205	226	181	162	127	96	139	486
SSP 1808	119.8	4.9	274	196	216	439	164	171	359	193	141	182	153	240
SSP 1809	104.9	4.9	81	43	69	61	32	36	.	39	32	34	29	53
SSP 1811	117.4	4.9	322	219	208	336	221	164	516	178	172	188	212	266
SSP 1814	104.9	4.9	79	57	55	55	.	46	31	50	35	38	37	65
SSP 1821	104.9	4.9	76	65	62	76	42	56	68	54	51	45	43	79
SSP 1822	75.2	4.9	90	80	68	98	52	59	102	67	70	60	78	95
SSP 1824	77.8	4.8	94	.	111	128	107	94	105	109	90	100	87	170
SSP 1825	118.2	4.9	97	61	67	89	72	58	179	67	40	62	67	.
SSP 1826	112.2	4.8	.	.	.	169	.	.	.	.	.	.	.	.
SSP 1827	140.6	4.9	51	60	49	65	47	38	71	44	39	48	54	69
SSP 1828	139.8	4.9	64	49	76	78	38	40	92	63	46	52	48	94
SSP 1829	139.3	5.0	.	.	59	.	.	.	.	.	59	.	.	.
SSP 1901	118.6	4.8	141	.	117	.	.	.	.	.	92	102	.	331
SSP 2001	16.6	4.9	1312	1378	1125	1120	1660	1709	.	1743	1271	1209	1310	1456
SSP 2002	9.7	5.1	3798	4811	3782	3300	4507	4874	3650	4938	4779	4138	4365	4574
SSP 2003	19.4	5.1	120	182	193	197	145	155	157	176	111	131	151	175
SSP 2005	8.2	5.2	1122	1407	1624	1205	1811	1453	3420	1845	1706	1351	1637	1571
SSP 2007	9.2	5.0	.	233	.	.	.	.	.	.	.	.	312	.
SSP 2008	15.6	5.1	88	155	148	230	83	94	133	109	136	142	157	138
SSP 2009	21.9	5.0	.	.	.	.	338	.	.	.	.	.	.	.
SSP 2010	12.5	5.2	.	.	.	.	.	.	2375	.	.	.	.	.
SSP 2011	20.4	5.0	144	.	.	.	.	.	.	.	.	.	.	.
SSP 2102	25.4	5.0	148	193	207	129	128	157	167	161	186	149	123	176
SSP 2104	22.0	5.0	1287	2219	2357	1049	989	1912	198	1961	1860	1399	1698	2094
SSP 2105	26.5	5.1	381	330	458	326	352	260	289	287	359	321	362	411

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 2106	24.5	5.1	12199	13712	13886	9026	9130	14492	12651	15198	15802	12490	11907	11237
SSP 2110	26.2	5.2	.	.	474	.	.	.	.	.	.	.	.	.
SSP 2113	22.0	5.2	.	.	.	.	.	.	545	.	.	.	.	.
SSP 2114	22.0	5.1	.	.	.	.	.	.	.	.	.	.	567	.
SSP 2115	26.8	5.2	245	1150	494	311	212	2221	.	2127	272	525	351	560
SSP 2117	23.9	5.2	.	.	89	.	.	35	.	37	.	.	.	23
SSP 2118	26.9	5.2	727	.	169	.	.	25	7596	42	150	.	.	.
SSP 2202	32.1	5.0	1423	1334	1328	1477	845	914	1514	800	963	1109	976	1510
SSP 2209	28.0	5.1	928	13866	14504	1283	1069	15395	758	17133	4548	8963	787	14376
SSP 2211	28.5	5.0	8666	.	.	.	.	.	.	.	.	.	.	.
SSP 2212	31.1	5.0	.	.	174	.	.	.	.	.	.	.	.	.
SSP 2213	29.6	5.2	.	.	.	443	.	.	.	.	.	.	.	.
SSP 2214	29.3	5.1	.	.	.	.	.	.	.	.	5852	.	.	.
SSP 2215	28.8	5.2	.	.	8480	.	.	.	.	.	368	4416	.	9515
SSP 2217	27.5	5.0	.	.	.	.	.	.	.	.	16949	.	.	.
SSP 2218	29.4	5.0	.	.	.	424	242	750	.	564	.	.	629	.
SSP 2219	27.4	5.2	13770	.	.	17425	14332	.	10856	.	188	10085	13181	.
SSP 2301	36.1	4.9	302	228	250	223	187	308	176	287	495	350	209	363
SSP 2302	33.0	5.0	3284	824	598	6197	547	.	3925	.	.	496	.	675
SSP 2305	35.9	5.0	1412	799	923	998	505	668	1321	1097	793	851	597	966
SSP 2306	32.7	5.1	.	3753	4102	.	3108	.	742	.	.	3395	672	4347
SSP 2308	32.6	5.1	.	1292	517	.	584	2475	.	3004	3493	593	2976	821
SSP 2311	35.5	5.2	2554	1239	1791	2077	1119	889	2160	1835	2275	1932	1290	2033
SSP 2313	35.1	5.1	182	.	259	473	284	210	.	270	.	234	216	190
SSP 2317	35.3	5.0	.	.	.	.	.	.	512	.	866	.	.	.
SSP 2318	32.7	5.2	.	.	.	.	.	225	109	337	304	.	583	.
SSP 2319	32.2	5.1	113	.	.	.	.	126	890	.	.	.	.	.
SSP 2401	41.3	5.0	175	193	137	158	123	.	172	.	149	72	.	.
SSP 2402	39.8	5.0	75	56	.	.	26	61	.	66	.	.	.	31
SSP 2406	41.1	5.1	490	631	430	451	390	525	550	590	525	369	304	421
SSP 2408	38.6	5.1	75	503	221	156	357	72	136	80	117	87	86	89
SSP 2410	41.3	5.2	158	261	251	386	174	180	257	205	234	212	182	230
SSP 2412	37.3	5.2	78	96	101	73	53	85	92	96	90	66	73	91
SSP 2413	38.6	5.0	193	.	.	.	.	269	.	310	419	.	.	.
SSP 2414	42.3	5.1	70	.	.	117	.	.	87	.	77	63	43	69
SSP 2420	39.3	5.2	.	.	95	.	.	.	82	.	.	.	.	88
SSP 2502	44.5	5.0	109	115	96	111	.	100	101	92	95	78	79	103
SSP 2503	43.1	5.0	170	176	188	182	169	139	179	115	137	205	156	293
SSP 2504	47.8	5.0	70	54	98	56	46	65	78	52	59	38	32	72
SSP 2505	45.7	5.0	75	78	91	70	.	66	68	71	.	55	59	75
SSP 2511	43.4	5.2	1350	1262	1138	1336	942	1251	1469	1346	1372	1015	1039	1252
SSP 2513	44.1	5.0	71	.	57	52	61	41	58	36	90	.	55	72
SSP 2517	42.9	5.1	61	77	.	88	.	.	85	.	58	74	60	86
SSP 2518	48.2	5.2	.	.	.	108	.	.	142	.	.	.	.	133
SSP 2520	47.4	5.0	.	.	.	.	.	.	48	.	.	.	.	.
SSP 2521	48.8	5.2	.	.	483	.	.	.	376	.	.	.	.	592
SSP 2522	47.5	5.1	38	96	99	69	41	45	.	49	42	42	.	70
SSP 2601	50.8	5.0	183	190	132	240	138	133	164	147	147	137	142	176
SSP 2602	55.3	5.0	.	48	97	217	.	.	.	.	120	107	124	.
SSP 2606	60.0	5.1	256	3547	335	322	211	126	1949	109	212	276	.	379

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 2610	50.0	5.2	362	412	544	597	318	351	270	399	451	373	419	860
SSP 2616	52.2	5.1	468	.	.	.	.	.	.	.	.	.	.	.
SSP 2617	56.2	5.2	137	.	43	136	.	.	.	.	51	.	.	.
SSP 2618	58.2	5.2	1391	.	1316	2458	960	.	302	.	1068	1131	.	1574
SSP 2619	50.4	5.2	163	.	210	.	.	.	141	.	.	134	.	180
SSP 2621	56.6	5.1	.	.	158	150	.	.	.	.	109	.	.	201
SSP 2622	53.9	5.1	.	.	123	145	90	96	.	107	86	69	90	.
SSP 2704	69.4	5.0	1468	683	926	1566	823	765	1062	723	1061	1036	714	1714
SSP 2713	65.3	5.2	114	85	110	141	50	.	71	.	118	49	46	99
SSP 2716	62.4	5.2	252	231	287	280	178	218	355	238	265	251	230	332
SSP 2717	67.3	5.2	459	314	393	400	332	324	677	337	383	301	291	496
SSP 2718	65.3	5.1	296	306	309	344	232	279	341	292	260	220	214	291
SSP 2719	70.9	5.0	242	171	146	229	203	153	161	155	215	181	171	284
SSP 2722	61.1	5.1	128	.	.	.	.	.	.	.	.	.	.	.
SSP 2724	60.8	5.0	.	697	.	.	.	.	.	.	.	.	.	.
SSP 2801	79.3	4.9	157	150	97	.	.	100	.	109	96	78	85	122
SSP 2812	83.4	5.0	84	71	107	148	53	77	.	73	42	37	.	99
SSP 2816	93.7	5.1	46	36	39	34	31	28	48	27	30	25	22	43
SSP 2817	114.7	5.1	186	108	145	174	143	141	181	144	175	142	144	252
SSP 2824	141.9	5.1	96	.	.	114	.	.	.	.	.	.	.	.
SSP 2826	77.8	5.0	349	.	225	291	140	143	309	175	.	159	.	.
SSP 2827	77.3	5.0	436	.	331	363	196	225	384	192	346	396	218	826
SSP 2828	77.5	5.1	154	.	121	202	182	.	.	105	379	246	287	743
SSP 2829	80.2	5.0	.	.	.	.	131	.	.	.	.	.	.	.
SSP 2830	82.7	5.0	.	.	.	.	.	.	155	.	.	.	.	.
SSP 2831	83.7	5.1	.	.	.	.	.	.	.	.	.	45	.	.
SSP 2833	98.7	5.1	21	18	31	.	23	17	52	11	30	21	17	42
SSP 2834	80.2	5.2	.	.	.	.	.	179	748	212	.	.	.	.
SSP 2906	154.8	5.0	50	.	34	40	24	.	48	22	26	25	15	47
SSP 2907	154.1	5.0	74	35	55	69	40	47	73	34	51	38	36	77
SSP 2909	138.3	5.0	.	.	104	.	41	47	67	51	104	61	85	69
SSP 2911	138.6	5.0	.	.	.	51	86	128	127	100	.	54	.	146
SSP 2912	142.7	5.1	164	.	.	.	.	.	.	.	.	.	.	.
SSP 2913	143.8	5.2	67	.	34	37	32	27	33	18	34	26	23	77
SSP 2914	134.0	5.2	.	.	.	.	.	.	.	.	.	.	.	152
SSP 3001	18.8	5.2	731	648	561	676	588	705	734	855	663	630	833	737
SSP 3002	21.3	5.3	229	251	249	289	212	199	261	306	248	192	238	244
SSP 3003	19.7	5.3	286	286	349	421	202	305	365	424	422	213	336	199
SSP 3008	18.3	5.4	46	96	87	.	154	40	148	49	36	74	.	44
SSP 3009	21.8	5.3	.	.	.	.	.	.	.	.	.	115	.	382
SSP 3010	20.7	5.3	.	.	.	102	.	.	.	.	.	.	.	.
SSP 3011	7.3	5.4	.	.	.	403	.	.	.	.	.	.	.	.
SSP 3012	6.4	5.3	.	1219	321	477	210	456	691	440	559	200	375	489
SSP 3102	22.0	5.3	210	262	285	.	.	97	.	150	214	225	195	173
SSP 3103	22.2	5.3	748	678	954	818	812	.	802	.	740	787	.	916
SSP 3104	24.5	5.3	4980	5920	6847	2812	4735	5566	7108	7084	5644	5392	5986	6249
SSP 3105	26.4	5.3	144	282	223	309	200	170	229	239	204	232	287	249
SSP 3107	25.3	5.3	67	71	63	.	60	70	85	80	69	27	63	65
SSP 3108	26.4	5.3	372	529	342	.	415	457	427	502	503	447	565	371
SSP 3109	26.3	5.3	493	439	.	469	430	405	520	681	397	407	548	507

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 3112	25.2	5.4	205	121	75	314	102	169	84	107	153	157	177	191
SSP 3115	22.0	5.2	.	.	.	.	.	.	713	.	.	.	.	.
SSP 3116	25.9	5.3	.	.	.	.	.	.	291	.	.	.	.	.
SSP 3117	26.4	5.4	209	250	308	265	204	424	.	483	216	271	213	336
SSP 3201	31.5	5.2	223	197	411	348	193	195	364	212	256	334	275	428
SSP 3205	30.9	5.2	276	242	274	375	204	190	249	293	249	175	218	207
SSP 3210	30.7	5.3	248	256	63	373	178	221	194	270	274	254	250	298
SSP 3211	29.8	5.3	180	153	285	215	117	130	244	187	181	156	175	256
SSP 3213	32.3	5.3	294	325	183	244	127	211	102	207	220	173	198	200
SSP 3214	31.0	5.3	250	203	241	239	139	156	219	192	198	177	197	225
SSP 3216	28.4	5.4	.	538	219	.	.	.	.	.	371	.	.	183
SSP 3219	31.4	5.4	.	2477	353	3768	.	2265	.	2718	2310	.	.	.
SSP 3221	28.5	5.3	248	.	.	.	162	.	436	.	171	.	.	.
SSP 3223	27.6	5.3	.	.	.	.	.	.	.	.	133	.	.	.
SSP 3224	29.9	5.4	259	433	389	408	221	292	360	342	385	293	204	337
SSP 3226	29.7	5.3	.	.	.	708	.	.	.	.	.	.	.	.
SSP 3227	31.5	5.3	162	192	186	155	563	152	177	140	182	177	164	232
SSP 3233	28.0	5.3	.	5593	2626	.	.	3114	.	3817	.	.	.	2263
SSP 3310	34.9	5.3	943	631	727	603	409	348	976	658	705	698	541	877
SSP 3311	32.3	5.2	716	.	742	677	1006	627	.	881	841	553	1011	893
SSP 3312	35.9	5.3	.	.	384	.	.	.	.	.	.	182	76	340
SSP 3313	36.4	5.4	.	.	.	.	.	147	.	119	.	43	105	.
SSP 3314	37.1	5.3	115	.	129	131	109	64	116	99	101	.	86	104
SSP 3315	37.2	5.4	62	58	42	67	64	53	61	60	41	48	34	58
SSP 3402	39.3	5.2	114	65	70	111	74	54	131	60	74	71	73	92
SSP 3405	41.8	5.3	60	77	83	.	63	.	82	66	305	50	48	62
SSP 3407	37.0	5.3	.	133	.	.	.	.	.	.	.	.	.	.
SSP 3408	37.7	5.3	125	127	166	157	90	62	129	92	86	91	100	121
SSP 3409	39.9	5.3	824	604	757	758	496	579	.	593	646	676	383	810
SSP 3414	38.0	5.3	104	.	.	151	.	.	59	.	.	31	109	36
SSP 3415	39.7	5.3	.	.	206	.	.	.	.	.	.	.	.	229
SSP 3416	42.2	5.3	99	.	91	.	.	.	.	.	67	67	46	89
SSP 3417	39.4	5.3	.	.	.	165	.	.	.	.	.	.	.	.
SSP 3420	37.3	5.3	57	63	68	100	51	46	130	75	47	65	62	64
SSP 3422	38.3	5.2	.	55	65	70	45	37	51	61	61	55	38	69
SSP 3424	42.3	5.4	.	.	.	77	.	.	.	.	.	.	.	.
SSP 3505	42.5	5.2	132	328	58	130	72	78	129	117	115	85	.	95
SSP 3510	45.2	5.3	154	134	217	229	133	103	234	91	156	209	170	327
SSP 3511	47.8	5.3	235	275	233	265	219	209	473	221	210	182	144	244
SSP 3513	49.4	5.3	166	176	168	225	130	122	160	127	153	122	108	122
SSP 3518	48.2	5.4	322	349	386	312	267	313	421	234	438	280	228	562
SSP 3519	43.7	5.4	262	288	229	229	176	216	286	184	295	222	191	270
SSP 3521	47.9	5.3	576	254	301	670	354	260	558	174	232	376	321	336
SSP 3526	44.1	5.3	.	.	175	.	.	.	.	.	.	.	.	.
SSP 3527	49.8	5.4	.	.	135	.	.	.	.	.	.	.	.	.
SSP 3528	45.9	5.3	.	.	.	106	.	.	.	.	69	.	.	.
SSP 3529	49.7	5.2	.	.	.	.	.	.	368	.	.	209	.	.
SSP 3530	47.0	5.2	83	.	.	.	.	.	.	.	261	.	.	.
SSP 3531	49.6	5.3	119	122	.	112	81	71	106	75	102	79	80	143
SSP 3532	43.9	5.3	150	327	152	142	190	188	199	147	148	160	181	248

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 3533	47.9	5.3	185	.	385	.	.	.	.	277	.	334	119	781
SSP 3603	58.5	5.2	90	977	106	224	115	778	158	1328	111	130	815	239
SSP 3605	50.1	5.3	165	210	194	207	154	162	201	179	206	96	160	.
SSP 3609	60.2	5.3	262	261	161	33	174	232	.	246	207	73	147	30
SSP 3610	52.8	5.3	.	172	233	489	128	122	161	209	136	167	150	273
SSP 3611	53.0	5.3	973	1023	1342	1509	783	715	1507	885	1046	959	864	1168
SSP 3613	51.3	5.4	172	160	89	117	167	138	1194	134	95	114	116	198
SSP 3615	51.6	5.3	.	.	.	146	.	.	331	.	.	.	.	.
SSP 3616	54.4	5.3	355	332	153	198	291	321	322	327	372	317	.	367
SSP 3617	51.8	5.3	.	.	.	78	.	.	.	.	.	50	.	.
SSP 3702	70.9	5.2	1198	226	736	214	253	148	367	151	348	359	176	1215
SSP 3708	62.3	5.3	68	128	136	161	101	81	164	46	113	138	130	179
SSP 3711	62.4	5.3	104	87	.	107	72	95	108	102	97	76	90	103
SSP 3713	63.5	5.3	.	165	160	.	117	105	141	96	86	82	86	144
SSP 3714	61.2	5.3	159	114	154	.	111	130	265	111	141	101	89	144
SSP 3716	60.8	5.4	70	157	93	156	.	.	76	.	99	.	.	54
SSP 3719	65.1	5.2	.	.	.	.	.	61	.	67	.	51	.	.
SSP 3721	61.3	5.3	108	156	119	120	72	99	.	71	.	60	68	132
SSP 3722	63.5	5.4	107	.	.	.	.	.	.	.	.	.	.	.
SSP 3723	64.1	5.3	45	.	.	.	.	.	.	.	.	.	.	.
SSP 3725	61.7	5.3	.	.	73	.	52	.	71	.	.	.	.	67
SSP 3727	63.6	5.3	192	.	.	.	.	.	.	.	.	.	.	.
SSP 3728	63.4	5.3	.	.	.	129	.	.	.	.	.	.	.	.
SSP 3729	63.5	5.2	89	.	82	.	.	.	.	.	.	45	.	79
SSP 3801	90.8	5.2	166	76	108	76	41	87	54	68	129	72	58	124
SSP 3809	109.7	5.3	97	52	116	99	44	51	81	41	84	68	45	82
SSP 3810	91.1	5.3	56	83	56	74	38	42	72	64	51	48	33	95
SSP 3811	81.5	5.3	288	314	324	383	211	247	514	213	230	290	142	513
SSP 3812	119.1	5.3	28	51	56	46	.	.	40	.	.	48	36	77
SSP 3813	110.7	5.3	73	27	29	.	21	29	30	21	35	45	20	85
SSP 3816	94.1	5.4	138	44	46	62	34	45	53	27	51	.	.	.
SSP 3817	114.4	5.4	78	188	245	184	66	68	70	38	49	185	135	299
SSP 3819	80.0	5.4	134	112	117	128	110	130	99	80	99	85	75	138
SSP 3822	72.7	5.3	135	.	153	.	84	.	.	.	60	71	41	202
SSP 3823	82.5	5.3	290	564	198	490	509	660	2868	1116	470	318	229	311
SSP 3826	91.1	5.2	.	.	.	.	76	.	.	.	.	.	.	.
SSP 3827	101.1	5.4	.	.	.	.	.	.	.	.	.	.	.	70
SSP 3828	81.0	5.3	1201	.	.	.	.	.	.	.	.	.	.	.
SSP 3829	98.0	5.4	.	.	.	.	.	.	.	.	.	36	.	130
SSP 3830	94.4	5.3	44	.	.	.	.	.	.	.	.	.	.	31
SSP 3832	106.8	5.4	42	33	32	31	42	22	34	16	22	32	24	.
SSP 3833	88.6	5.4	.	.	.	.	.	.	38	.	.	.	.	.
SSP 4002	21.0	5.4	480	419	559	596	305	.	723	.	400	396	572	791
SSP 4003	21.8	5.5	444	558	1378	1033	621	780	1092	997	696	913	1250	1149
SSP 4004	7.8	5.5	.	.	321	.	.	.	.	.	.	.	.	.
SSP 4103	22.0	5.4	2667	2402	2693	1381	1371	1313	1160	1127	587	1095	669	1505
SSP 4104	24.6	5.5	1269	1597	2275	847	1516	1744	2334	1582	1305	1396	1614	2034
SSP 4105	26.1	5.5	108	147	129	158	105	128	123	147	135	114	108	116
SSP 4106	25.2	5.5	442	494	552	551	429	499	441	554	639	408	461	448
SSP 4107	26.8	5.5	423	496	544	299	526	439	588	519	435	576	493	781

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 4110	22.1	5.6	12399	12969	11502	14600	13031	13024	10137	14357	15685	14108	12562	13028
SSP 4112	25.2	5.6	.	99	131	121	76	140	84	96	108	78	76	85
SSP 4113	26.5	5.5	.	.	.	439	.	.	.	.	.	.	.	.
SSP 4114	22.1	5.5	.	.	.	334	.	.	.	274	180	322	.	299
SSP 4117	26.5	5.6	55	.	63	.	53	73	.	72	.	.	.	.
SSP 4203	28.0	5.4	2922	1363	1272	4376	3654	1098	5649	1360	1379	2469	2756	1372
SSP 4207	29.5	5.6	137	161	162	193	117	143	.	133	137	191	132	265
SSP 4208	27.6	5.6	602	979	621	1605	478	743	796	832	797	459	828	543
SSP 4210	32.2	5.5	425	.	2671	133	2195	.	1922	.	.	1851	1910	1279
SSP 4212	28.2	5.5	.	.	.	2712	.	.	.	.	.	.	.	.
SSP 4213	28.7	5.5	.	.	.	.	690	1804	.	1140	.	.	996	.
SSP 4214	31.3	5.5	.	.	.	.	447	.	.	.	.	.	.	.
SSP 4215	29.1	5.6	.	.	137	.	.	.	.	.	.	.	.	134
SSP 4216	30.3	5.6	197	.	.	.	201	.	.	.	.	.	.	.
SSP 4217	29.8	5.5	153	.	164	96	86	.	.	.	239	107	128	189
SSP 4307	36.6	5.5	218	257	153	202	227	184	172	227	225	234	249	321
SSP 4310	35.5	5.5	141	122	106	120	102	122	137	97	122	121	91	124
SSP 4311	36.3	5.6	177	155	154	306	261	88	340	90	186	167	62	200
SSP 4317	32.3	5.6	2115	198	1037	.	107	115	.	126	753	.	.	.
SSP 4320	32.5	5.5	77	.	.	511	2856	.	1178	.	.	.	101	.
SSP 4321	33.1	5.5	.	.	.	.	.	.	250	.	.	.	.	.
SSP 4322	32.5	5.6	.	.	.	5901	.	.	4640	.	.	.	.	.
SSP 4323	34.3	5.6	.	.	.	.	.	793	.	646	.	.	.	.
SSP 4324	36.4	5.4	.	.	.	.	.	.	198	.	.	.	.	.
SSP 4325	32.3	5.5	.	.	.	.	.	.	.	.	.	2525	.	3290
SSP 4401	41.7	5.4	180	191	204	243	131	174	216	172	208	162	168	184
SSP 4402	39.2	5.4	363	470	445	541	348	375	538	444	436	438	349	472
SSP 4405	37.5	5.4	188	266	229	226	185	137	232	133	193	258	158	284
SSP 4406	40.8	5.5	471	549	598	484	316	510	1022	465	516	471	390	685
SSP 4407	38.2	5.5	123	134	104	114	79	110	132	104	130	59	68	.
SSP 4412	37.6	5.6	73	82	89	.	68	67	66	56	66	66	55	93
SSP 4418	39.9	5.6	340	471	322	416	298	336	403	323	417	373	285	484
SSP 4420	41.0	5.6	.	.	513	.	.	.	.	.	.	.	.	.
SSP 4421	39.2	5.5	.	.	66	155	.	.	69	.	56	56	.	.
SSP 4422	39.3	5.6	92	94	.	.	.	72	65	75	85	63	43	100
SSP 4504	47.8	5.4	327	397	380	500	274	315	.	290	306	302	199	419
SSP 4506	49.0	5.4	191	187	194	177	127	180	207	101	189	158	102	244
SSP 4507	46.0	5.5	62	51	42	134	34	33	33	38	47	52	26	56
SSP 4514	45.6	5.5	.	65	70	87	48	64	64	50	82	60	51	78
SSP 4515	48.9	5.6	518	391	489	437	290	348	573	227	451	398	368	618
SSP 4516	43.6	5.6	123	193	228	208	126	148	115	123	186	178	152	219
SSP 4517	44.6	5.6	127	143	147	119	89	121	155	94	112	60	71	158
SSP 4521	44.5	5.5	69	116	88	68	55	104	82	56	64	40	44	61
SSP 4522	49.7	5.6	76	.	127	151	67	61	.	36	82	109	63	120
SSP 4601	50.0	5.4	200	141	110	121	103	123	208	107	129	133	100	194
SSP 4606	54.7	5.4	208	233	174	235	160	227	311	128	191	133	183	231
SSP 4607	51.1	5.5	734	678	693	714	502	611	575	604	885	600	468	817
SSP 4608	59.9	5.5	71	56	56	49	33	62	60	27	65	52	36	84
SSP 4610	53.0	5.5	1664	1792	1986	2297	1119	1563	2071	1513	1885	1693	1359	1912
SSP 4614	52.1	5.5	284	274	268	293	47	237	298	206	213	240	272	318

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 4616	51.3	5.6	961	804	879	734	458	678	.	456	1152	437	677	1134
SSP 4619	50.5	5.5	876	58	.	.	401	.	71	.	.	.	697	.
SSP 4620	50.8	5.5	.	.	.	.	3051	140	.	123	195	158	.	1226
SSP 4622	54.0	5.6	187	.	176	192	124	209	203	117	207	187	112	.
SSP 4702	61.6	5.4	166	222	299	313	170	123	238	119	205	166	183	222
SSP 4703	65.9	5.4	308	294	353	69	177	231	436	208	269	216	223	284
SSP 4705	63.7	5.4	165	187	213	251	138	.	181	.	200	136	122	203
SSP 4706	63.5	5.5	.	148	.	.	.	.	.	.	.	65	.	.
SSP 4707	68.1	5.5	169	176	88	136	120	64	209	37	154	27	91	85
SSP 4710	65.9	5.5	62	68	67	.	38	54	67	47	78	38	49	57
SSP 4711	62.6	5.5	176	116	165	75	101	126	139	82	139	102	83	142
SSP 4712	63.8	5.5	107	118	128	415	103	.	158	.	129	106	93	137
SSP 4713	61.9	5.5	110	112	154	143	90	109	152	86	144	158	155	247
SSP 4717	63.2	5.6	1315	85	120	121	72	75	118	51	104	106	81	157
SSP 4718	67.3	5.6	207	68	106	.	111	72	101	52	99	55	61	126
SSP 4721	62.7	5.4	56	.	48	.	29	37	44	.	32	28	.	40
SSP 4724	63.6	5.4	.	.	.	.	.	196	.	193	.	.	.	.
SSP 4725	63.6	5.5	.	.	.	.	.	127	.	114	.	.	.	.
SSP 4726	65.3	5.5	.	.	.	245	.	.	.	.	.	.	.	.
SSP 4728	64.3	5.6	.	.	.	.	.	.	.	.	33	.	.	.
SSP 4729	69.3	5.6	.	.	.	.	.	.	.	.	.	35	.	.
SSP 4731	63.2	5.5	311	38	267	259	56	90	60	94	52	44	84	62
SSP 4802	87.6	5.4	40	43	23	52	.	.	42	.	32	28	25	38
SSP 4804	106.8	5.4	124	87	105	121	74	82	105	62	110	99	80	170
SSP 4805	93.7	5.4	96	120	126	107	95	103	145	74	119	89	82	145
SSP 4807	89.6	5.4	.	48	.	96	.	.	.	.	.	.	.	.
SSP 4808	87.6	5.5	79	59	39	.	.	.	53	.	58	34	35	56
SSP 4811	75.5	5.5	16	20	21	.	.	.	.	.	.	.	.	.
SSP 4813	95.4	5.5	36	19	25	.	18	26	16	.	34	16	.	32
SSP 4814	79.3	5.6	.	27	.	.	.	.	.	.	.	24	.	84
SSP 4815	81.1	5.6	130	92	64	121	116	70	117	38	84	70	63	137
SSP 4816	132.0	5.6	52	27	.	.	24	.	63	.	.	.	.	.
SSP 4818	98.5	5.4	.	62	.	65	61	63	.	36	92	37	38	.
SSP 4819	89.4	5.5	43	.	47	43	.	.	.	.	.	.	.	.
SSP 4820	115.0	5.4	248	.	.	.	117	165	215	117	147	.	.	.
SSP 4821	88.1	5.6	53	.	45	.	29	28	34	10	30	.	.	46
SSP 4822	80.7	5.5	31	.	.	.	69	.	35	.	.	.	.	.
SSP 4823	100.6	5.5	.	.	44	.	.	.	.	.	.	.	.	.
SSP 4825	102.5	5.5	55	19	20	17	25	37	.	.	40	20	.	21
SSP 4826	102.4	5.4	.	.	.	.	.	.	35	.	.	.	.	.
SSP 4827	80.4	5.4	187	.	.	.	.	.	.	.	.	.	.	.
SSP 4828	98.4	5.6	.	.	24	.	.	.	69	.	.	.	.	.
SSP 4829	121.6	5.5	41	30	25	.	.	35	32	.	28	.	.	.
SSP 4831	143.0	5.6	14	.	.	.	.	.	26	.	.	.	.	.
SSP 4832	99.2	5.6	.	13	.	.	.	.	29	.	.	.	.	.
SSP 4833	71.5	5.5	34	39	.	38	.	37	51	35	26	24	17	44
SSP 4910	198.0	5.5	135	74	90	70	44	71	77	39	105	62	74	120
SSP 4911	199.1	5.5	133	77	104	84	67	85	104	47	99	65	70	218
SSP 5001	19.9	5.6	180	176	144	223	126	159	169	167	193	176	154	176
SSP 5002	19.4	5.7	140	157	138	167	125	137	58	145	186	119	166	97

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 5003	8.0	5.7	18825	15524	16065	28949	19484	16116	21632	17179	7246	21368	10667	26904
SSP 5004	15.0	5.9	418	474	286	1089	537	409	358	465	481	513	656	424
SSP 5005	20.7	5.9	443	.	343	.	.	.	.	.	330	486	.	288
SSP 5006	19.8	5.9	468	.	569	.	.	.	.	.	366	496	.	302
SSP 5010	18.2	6.0	131	156	59	.	217	213	310	43	88	99	.	67
SSP 5103	22.0	5.7	2583	5805	5755	5962	6003	5007	3086	3466	4855	3807	4695	4774
SSP 5109	24.5	5.9	786	697	556	968	753	848	770	964	800	849	863	794
SSP 5111	25.8	5.9	208	215	170	208	313	220	280	269	269	236	216	236
SSP 5112	26.7	5.6	121	88	65	134	125	164	.	145	153	88	104	96
SSP 5113	22.0	5.7	.	.	260	.	.	.	.	.	.	.	.	.
SSP 5118	25.9	5.8	71	86	72	82	72	84	84	81	62	71	73	71
SSP 5119	25.5	5.8	.	.	.	.	.	.	.	.	.	48	.	119
SSP 5203	31.4	5.6	885	1359	1092	475	3010	686	1929	670	585	1312	4316	1302
SSP 5207	31.6	5.7	6402	1827	1945	262	4321	1458	2777	1262	1533	4599	.	3297
SSP 5210	32.2	5.7	2338	1154	1358	1342	.	1656	1181	825	2817	2336	1423	6685
SSP 5211	28.1	5.7	6142	3895	5907	20543	1654	3129	2785	2931	4743	2491	4411	2856
SSP 5214	30.7	5.9	14294	16005	21010	5381	22162	13145	16380	12192	9790	15720	1060	12348
SSP 5215	27.4	5.9	1725	2823	2471	3221	2142	2779	1467	2790	2509	2004	2435	2006
SSP 5216	27.2	5.7	4733	794	4539	.	3620	1006	.	669	714	4319	593	4053
SSP 5217	30.0	5.7	542	185	509	776	117	599	348	741	385	326	1012	686
SSP 5218	31.2	5.7	.	3246	141	.	4062	.	2657	.	.	.	.	.
SSP 5219	27.6	5.6	943	726	505	4516	.	.	787	.	3744	.	.	.
SSP 5220	32.0	5.9	8133	.	2463	404	.	.	.	.	2017	4312	6361	3276
SSP 5221	27.6	5.9	.	.	.	.	.	.	3896	.	.	.	.	.
SSP 5222	31.7	5.7	2162	.	.	2930	.	1426	.	1284	.	.	2593	3075
SSP 5223	29.4	5.9	.	.	.	.	.	.	1666	.	.	.	.	.
SSP 5224	32.2	5.6	402	.	.	.	.	.	6731	.	.	346	.	528
SSP 5226	31.1	5.9	.	.	.	8000	.	4651	.	5336	3153	7591	4382	4177
SSP 5227	29.8	5.6	.	.	.	.	.	.	1340	.	.	.	.	.
SSP 5228	30.6	5.7	.	.	1150	.	.	.	.	.	.	.	.	.
SSP 5229	29.5	5.9	213	2652	3981	3428	.	307	.	443	318	2777	487	2996
SSP 5302	36.3	5.6	.	207	195	290	149	177	202	166	219	166	191	245
SSP 5303	34.3	5.6	89	106	102	.	72	.	108	76	103	88	95	191
SSP 5304	32.7	5.7	796	673	822	750	721	738	363	688	971	726	817	404
SSP 5306	35.1	5.7	284	316	315	316	203	284	286	227	295	244	264	319
SSP 5311	36.5	5.8	803	1018	1200	1360	909	954	1003	988	978	1116	1079	1297
SSP 5316	34.8	5.8	314	.	451	.	.	.	.	.	.	403	377	989
SSP 5317	36.3	5.7	.	.	103	.	.	.	.	.	.	.	.	.
SSP 5319	37.2	5.9	.	.	.	.	318	.	.	.	.	.	.	.
SSP 5320	32.2	5.9	.	.	.	.	.	.	.	.	.	.	290	.
SSP 5321	36.2	5.7	81	62	.	57	63	58	.	44	81	63	56	.
SSP 5322	32.3	5.8	530	820	1173	518	3135	675	826	583	1189	950	428	1031
SSP 5402	37.3	5.6	99	524	106	156	76	406	158	408	367	354	102	498
SSP 5403	38.7	5.6	171	186	177	138	136	138	159	152	262	215	112	222
SSP 5406	39.9	5.7	440	600	683	778	372	491	599	325	579	640	539	949
SSP 5408	40.5	5.7	150	238	193	151	163	125	179	232	119	175	539	376
SSP 5409	37.5	5.7	83	109	117	102	52	109	90	74	99	84	96	61
SSP 5413	40.4	5.8	118	396	278	135	182	.	181	.	138	161	182	269
SSP 5414	39.1	5.9	177	204	126	.	.	173	108	187	.	.	59	.
SSP 5415	40.3	5.9	188	188	212	210	164	181	237	190	96	140	170	202

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 5416	38.7	5.8	98	73	.	.	.	56	102	46	.	.	.	.
SSP 5418	40.9	5.8	644	.	.	.	.	628	731	600	802	.	485	.
SSP 5419	42.9	5.9	983	.	.	.	.	1023	.	880	.	.	.	.
SSP 5420	38.5	5.7	.	.	.	203	.	.	.	.	.	.	.	.
SSP 5423	39.5	5.8	.	.	.	.	.	333	.	277	.	.	.	.
SSP 5425	42.2	5.8	45	.	53	.	.	53	.	.	41	.	.	72
SSP 5426	41.8	5.7	19	33	70	52	35	33	38	23	51	37	35	59
SSP 5503	45.4	5.6	143	127	115	144	94	190	146	183	180	158	89	201
SSP 5505	49.2	5.6	.	96	.	.	.	.	.	.	.	.	.	.
SSP 5506	46.5	5.7	57	34	50	436	30	40	47	17	70	45	32	52
SSP 5507	43.5	5.7	182	301	250	328	215	187	314	140	158	246	174	232
SSP 5508	49.3	5.7	91	144	180	167	117	106	168	89	137	126	105	144
SSP 5509	46.2	5.7	108	79	101	.	68	73	88	47	93	83	55	100
SSP 5511	42.4	5.7	.	55	106	.	.	.	82	.	.	.	.	.
SSP 5513	45.7	5.9	232	99	127	383	113	158	152	150	90	133	133	157
SSP 5514	43.3	5.9	.	1063	.	1017	91	.	1978	.	.	57	.	.
SSP 5516	48.0	5.9	.	59	68	.	.	.	.	.	58	.	.	69
SSP 5518	45.8	5.9	128	128	121	.	115	120	144	98	148	114	111	129
SSP 5519	49.3	5.7	108	.	297	210	137	.	.	.	.	172	93	180
SSP 5520	47.1	5.9	62	61	.	.	58	48	97	44	61	56	56	65
SSP 5521	45.2	5.7	61	50	57	77	31	44	40	50	62	.	45	.
SSP 5523	47.9	5.7	124	.	.	130	56	95	120	62	.	.	.	.
SSP 5524	44.8	5.9	62	.	72	.	90	.	.	.	.	62	.	86
SSP 5525	43.5	5.7	136	.	677	123	.	.	79	.	930	971	866	1456
SSP 5526	44.5	5.6	60	.	98	.	.	.	.	29	.	.	.	.
SSP 5527	48.7	5.7	36	.	.	.	.	.	.	.	.	.	.	.
SSP 5528	43.7	5.6	30	81	.	78	72	30	95	25	97	138	159	231
SSP 5529	45.8	5.9	.	.	.	2111	.	.	.	.	.	.	.	.
SSP 5530	48.1	5.8	.	.	.	.	.	.	.	.	.	46	47	.
SSP 5533	46.2	5.8	64	67	80	91	45	70	69	56	72	.	.	65
SSP 5601	51.6	5.6	.	177	297	156	.	.	131	.	.	.	194	.
SSP 5602	55.5	5.6	177	144	156	204	177	.	.	.	209	129	151	158
SSP 5603	52.7	5.6	264	174	178	167	133	135	172	111	189	155	115	220
SSP 5605	53.0	5.7	558	323	322	685	351	310	392	214	506	566	367	735
SSP 5608	52.8	5.7	66	96	69	.	.	.	59	.	75	69	.	61
SSP 5609	51.0	5.8	148	211	181	.	143	172	202	147	214	194	155	231
SSP 5613	50.8	5.9	.	112	102	84	.	.	.	75	133	116	.	222
SSP 5616	52.1	5.9	82	26	.	.	.	.	.	.	.	.	.	54
SSP 5617	56.4	5.7	56	38	52	61	31	48	44	25	42	41	34	45
SSP 5618	55.8	5.9	107	127	145	175	114	223	167	230	120	133	137	149
SSP 5619	50.4	5.7	162	1096	922	946	90	724	790	735	1303	1479	169	1420
SSP 5620	51.1	5.6	.	.	.	244	.	.	.	.	.	.	.	.
SSP 5621	52.8	5.9	94	45	35	79	16	.	54	.	57	38	.	47
SSP 5702	61.0	5.6	146	120	132	136	88	102	155	68	152	112	103	177
SSP 5705	63.3	5.6	85	51	74	.	49	44	.	.	146	39	.	.
SSP 5707	65.9	5.7	.	176	218	528	330	213	208	177	347	277	119	205
SSP 5709	65.7	5.7	.	305	417	500	.	283	460	232	340	417	323	643
SSP 5712	64.1	5.8	405	1507	2540	2907	175	1628	3104	1553	1940	2028	1365	2256
SSP 5719	65.5	5.9	2390	239	191	164	.	153	275	151	135	221	132	356
SSP 5720	61.2	5.9	82	51	54	49	35	54	55	37	72	60	42	179

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 5721	61.5	5.9	201	167	203	187	127	152	201	115	215	180	160	244
SSP 5722	62.3	5.7	54	55	170	49	116	42	68	32	53	104	36	196
SSP 5723	63.1	5.8	117	78	105	123	64	.	.	.	118	114	113	223
SSP 5725	63.1	5.9	.	.	.	.	.	56	.	49	.	.	.	.
SSP 5729	63.3	5.6	48	.	.	.	.	.	.	.	.	.	.	.
SSP 5730	62.6	5.6	36	30	20	25	.	29	39	18	133	26	.	51
SSP 5732	64.5	5.7	413	267	168	102	1609	162	121	96	175	197	584	.
SSP 5733	62.1	5.9	132	116	.	206	81	121	150	106	175	128	119	154
SSP 5734	61.9	5.7	44	27	86	52	58	41	62	24	102	61	57	157
SSP 5802	79.3	5.6	180	136	122	205	29	118	232	75	149	116	119	217
SSP 5803	143.0	5.6	38	22	26	.	.	22	.	9	33	.	17	.
SSP 5804	121.9	5.6	43	22	19	.	.	.	33	.	.	.	.	29
SSP 5805	88.0	5.6	.	26	24	132	.	.	30	.	141	70	.	.
SSP 5806	80.0	5.7	61	57	.	60	43	56	96	36	68	43	35	60
SSP 5810	119.8	5.8	84	90	76	53	47	.	112	.	76	88	61	158
SSP 5811	89.6	5.9	32	20	30	25	16	20	17	16	29	22	14	56
SSP 5812	117.4	5.9	202	168	155	140	86	24	225	20	129	171	117	247
SSP 5813	84.1	5.9	.	112	.	.	70	85	.	63	.	.	94	.
SSP 5814	117.4	5.6	.	15	.	.	.	.	.	.	.	.	.	25
SSP 5815	70.5	5.7	.	.	61	.	.	108	.	93	.	.	.	194
SSP 5816	83.9	5.7	.	.	128	.	.	.	118	.	.	.	.	.
SSP 5817	71.3	5.9	18	.	84	.	78	.	.	.	.	.	.	.
SSP 5818	72.7	5.7	.	.	.	.	.	.	87	.	.	.	.	.
SSP 5820	129.4	5.6	.	.	37	33	.	31	.	16	47	30	32	77
SSP 5821	145.3	5.6	.	.	.	.	.	.	.	.	.	30	.	30
SSP 5822	97.5	5.9	32	.	.	.	.	27	21	.	24	.	.	29
SSP 5823	84.9	5.8	124	.	.	.	22	.	.	.	35	62	.	238
SSP 5824	96.9	5.8	38	.	.	.	.	.	.	.	.	.	.	.
SSP 5902	121.9	5.7	34	35	28	.	.	.	55	.	23	26	.	57
SSP 5903	139.9	5.7	.	.	.	18	26	.	12	.	.	.	.	.
SSP 6001	9.5	6.0	3125	2931	1982	2111	2300	2795	2963	3226	3980	2889	3698	2258
SSP 6002	16.2	6.1	583	484	522	500	505	556	424	640	577	442	530	505
SSP 6005	20.3	6.1	.	689	472	.	531	308	197	354	.	.	.	.
SSP 6007	17.9	6.2	321	286	294	363	272	316	343	359	310	282	350	311
SSP 6008	21.8	6.1	177	173	201	249	205	213	259	196	237	212	233	254
SSP 6010	20.5	6.2	.	.	.	182	.	1003	.	1071	458	86	1625	224
SSP 6012	19.1	6.2	.	.	.	.	.	.	.	.	.	3233	.	.
SSP 6101	22.0	5.9	880	635	.	387	599	882	807	781	855	832	613	609
SSP 6102	26.8	5.9	136	126	88	188	113	113	316	100	117	109	102	102
SSP 6104	25.5	6.0	382	286	342	.	514	608	378	450	409	399	661	485
SSP 6105	27.0	6.0	516	657	856	27191	613	630	694	819	2567	566	629	793
SSP 6108	25.7	6.0	276	295	318	833	270	265	1733	1174	306	298	357	359
SSP 6109	22.1	6.1	12036	10700	10409	14413	9486	9589	10993	6845	9577	10204	9841	10221
SSP 6113	26.2	6.2	2166	2136	1595	1388	1884	2173	1836	2322	2036	2041	1971	2135
SSP 6115	27.0	6.2	4303	5370	4196	4682	3680	4388	3127	5182	4642	3852	4534	3241
SSP 6125	22.0	6.2	214	216	182	308	197	216	260	263	208	168	183	209
SSP 6126	26.4	6.0	.	.	.	.	.	.	1117	.	.	.	.	.
SSP 6127	25.2	6.1	165	65	.	67	73	66	87	69	60	282	61	118
SSP 6128	24.9	6.2	111	108	230	149	86	121	169	165	120	.	110	.
SSP 6130	26.6	6.1	59	179	1202	2933	153	157	171	178	218	178	611	773

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 6205	32.2	6.0	9811	6698	6149	.	6290	5817	.	5706	7543	11499	4934	9224
SSP 6206	29.2	6.0	991	662	543	978	219	707	4032	786	936	943	1137	1310
SSP 6208	30.1	6.0	5356	10116	11699	.	17525	13662	9105	13127	9277	8953	13118	9952
SSP 6210	31.8	6.1	2274	4639	732	.	8026	1964	6769	2875	1492	2610	.	5416
SSP 6211	27.2	6.1	8919	4254	8660	.	3755	3802	3089	4075	3697	3027	3625	3894
SSP 6212	32.0	6.1	.	1554	.	9861	14550	5256	2465	6430	1618	8097	4756	9091
SSP 6216	31.1	6.2	.	5873	.	.	.	.	8572	.	.	.	.	.
SSP 6217	31.8	6.2	15879	8970	14352	9480	11831	9689	4052	9642	17444	17374	17818	17313
SSP 6220	28.6	6.2	2150	3408	1257	.	3226	2558	.	2341	1979	3081	4762	1644
SSP 6221	30.2	6.1	.	.	.	14485	.	.	.	.	.	.	.	.
SSP 6301	32.5	5.9	2758	3178	4696	4262	2496	3789	2289	3763	4699	4905	3101	8040
SSP 6304	35.3	6.1	971	522	1505	776	853	907	1122	1014	1136	757	837	936
SSP 6305	33.0	6.1	10520	9093	10549	7064	10893	7818	7270	9300	8494	10134	11831	13201
SSP 6309	36.2	6.0	142	.	.	.	.	.	36	.	.	.	.	.
SSP 6310	37.2	6.0	434	.	.	604	.	.	.	.	.	.	200	.
SSP 6311	35.2	6.0	569	.	773	714	427	501	751	557	638	737	.	803
SSP 6312	33.3	6.0	.	.	.	561	.	.	.	.	.	.	.	.
SSP 6313	37.2	6.2	.	.	.	119	.	.	486	.	.	.	.	.
SSP 6315	32.6	6.2	.	.	.	3012	4415	.	.	.	.	.	.	.
SSP 6316	35.0	6.0	.	.	.	.	.	.	.	.	370	.	338	.
SSP 6401	38.6	5.9	54	61	68	87	48	110	76	63	70	48	36	45
SSP 6404	42.0	5.9	109	86	110	135	65	76	151	103	99	121	116	132
SSP 6405	39.5	6.0	88	191	176	.	134	.	223	.	142	147	147	189
SSP 6407	40.8	6.0	80	450	85	560	448	74	67	76	81	487	76	515
SSP 6408	38.2	6.0	585	682	635	778	583	588	700	658	699	673	539	743
SSP 6411	39.3	6.1	3182	3793	3662	2640	2306	2187	3244	1920	2651	2386	2157	3709
SSP 6414	38.0	6.1	125	111	463	444	84	93	516	88	100	93	79	146
SSP 6415	39.0	6.2	191	183	136	.	225	140	251	114	.	213	229	276
SSP 6418	39.0	6.0	.	.	.	168	141	.	23	.	191	150	.	161
SSP 6419	43.1	6.0	.	.	.	.	613	.	.	.	.	.	.	.
SSP 6420	37.8	6.2	.	.	476	347	.	.	580	.	.	.	.	.
SSP 6422	41.2	5.9	.	.	.	179	.	.	.	.	.	.	.	.
SSP 6423	42.0	6.1	136	129	129	164	140	106	190	100	183	98	119	222
SSP 6502	49.2	6.0	.	34	69	.	.	.	.	.	.	.	.	62
SSP 6503	46.4	6.0	321	310	402	399	222	270	511	258	308	272	274	317
SSP 6506	43.0	6.0	87	61	80	.	70	84	.	91	72	85	.	77
SSP 6508	45.1	6.0	472	344	45	1704	311	367	391	307	478	374	379	498
SSP 6509	49.8	6.1	.	181	210	.	202	.	.	.	195	193	84	202
SSP 6510	48.1	6.1	98	85	102	76	136	81	.	81	103	165	158	266
SSP 6513	47.5	6.2	504	182	354	87	400	329	210	264	426	736	577	1354
SSP 6518	47.3	6.1	.	.	132	.	.	.	.	.	.	.	.	.
SSP 6519	45.1	6.0	.	.	.	5706	.	.	.	.	.	.	.	.
SSP 6520	47.9	6.1	.	.	.	.	.	.	.	.	.	.	.	223
SSP 6521	48.1	6.0	.	.	.	.	.	.	250	.	.	.	.	.
SSP 6522	44.8	6.2	105	118	168	473	81	116	157	92	106	146	104	179
SSP 6523	43.3	6.2	53	36	70	.	.	43	.	35	62	.	.	.
SSP 6606	53.9	6.0	.	235	324	368	229	224	435	223	266	241	280	309
SSP 6608	52.9	6.0	124	75	81	.	64	107	91	92	108	114	73	111
SSP 6609	51.0	6.1	1650	121	129	.	218	187	.	172	166	229	159	241
SSP 6610	55.3	6.1	666	534	745	689	432	490	835	439	587	505	557	664

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 6612	52.0	6.2	178	94	249	219	42	101	99	93	165	171	89	237
SSP 6614	50.4	6.2	.	1197	1202	.	1128	1867	.	1708	1948	2102	1478	1855
SSP 6615	55.8	6.2	.	157	237	266	.	.	204	.	184	70	144	.
SSP 6616	59.4	6.2	.	179	216	170	.	.	223	.	208	88	152	.
SSP 6618	56.1	6.0	79	.	.	.	.	.	.	.	.	.	.	.
SSP 6622	50.5	5.9	321	.	.	.	47	217	220	210	.	.	.	.
SSP 6623	51.7	6.0	155	.	117	.	.	.	.	.	.	.	.	124
SSP 6625	50.2	6.0	43	.	141	137	.	.	.	.	.	.	.	158
SSP 6703	61.6	6.0	93	71	72	100	52	64	100	53	.	79	70	106
SSP 6706	62.1	6.0	55	49	946	83	37	.	65	.	.	71	58	92
SSP 6708	66.5	6.0	56	70	85	38	89	78	133	63	105	149	109	219
SSP 6709	61.7	6.0	56	43	96	.	29	45	60	31	63	38	44	.
SSP 6711	65.9	6.1	.	47	.	.	.	.	.	.	78	.	.	.
SSP 6712	62.3	6.1	89	59	79	71	42	50	89	38	84	56	61	73
SSP 6713	63.6	6.1	111	75	102	67	81	77	141	59	116	159	97	229
SSP 6714	65.3	6.1	197	197	213	116	193	197	346	158	283	339	221	613
SSP 6719	65.1	6.0	43	32	33	43	.	.	44	.	27	.	.	.
SSP 6722	67.5	6.2	55	.	.	.	.	69	.	59	.	.	.	.
SSP 6723	62.1	6.1	34	24	38	37	.	.	.	.	33	36	.	59
SSP 6724	61.1	6.2	.	.	117	.	.	.	.	.	.	.	.	.
SSP 6726	62.5	6.0	.	.	.	.	.	.	43	.	.	.	.	.
SSP 6727	58.5	6.1	165	95	128	185	65	88	135	77	166	103	114	232
SSP 6728	60.8	6.2	.	.	.	.	.	.	.	.	.	.	70	.
SSP 6729	63.1	6.2	73	51	58	57	53	45	.	.	67	63	.	171
SSP 6732	64.2	5.9	.	42	50	36	37	79	46	27	.	61	44	91
SSP 6733	62.6	6.0	.	.	.	51	.	.	.	.	.	.	.	.
SSP 6801	121.9	6.0	85	60	185	50	45	53	91	34	71	58	41	150
SSP 6802	132.8	6.0	31	23	24	.	18	25	26	.	24	22	.	43
SSP 6807	119.1	6.0	168	120	25	114	77	92	188	51	142	112	116	406
SSP 6810	99.3	6.1	87	163	131	88	75	104	106	91	125	141	118	239
SSP 6811	82.2	6.1	48	26	29	30	19	31	44	18	41	35	30	62
SSP 6816	75.5	6.2	43	28	31	26	.	.	.	.	31	.	19	.
SSP 6818	113.2	6.2	88	58	65	43	69	92	85	33	130	104	103	118
SSP 6819	100.7	6.0	41	55	41	33	31	28	36	16	28	40	39	77
SSP 6820	85.2	6.2	.	.	.	124	32	.	.	.	35	.	.	68
SSP 6821	87.3	6.1	.	.	.	.	.	.	109	.	.	.	.	.
SSP 6822	83.1	6.0	25	.	15	.	.	.	25	.	21	18	.	35
SSP 6823	75.8	6.0	17	11	.	31	15	13	22	.	17	19	.	30
SSP 6824	80.4	6.1	.	.	.	.	.	.	.	.	.	.	.	29
SSP 6825	68.3	6.1	38	27	.	30	.	.	.	.	.	.	.	77
SSP 6826	68.9	6.0	.	.	.	.	.	.	.	.	.	48	.	.
SSP 6827	66.0	6.0	170	.	.	.	.	.	.	.	.	.	.	.
SSP 6828	69.3	6.0	85	52	103	83	26	57	63	40	81	48	60	131
SSP 6831	100.5	6.2	19	.	.	.	.	.	.	.	.	.	.	.
SSP 6903	115.5	6.0	.	.	.	.	.	125	.	88	.	.	.	.
SSP 6904	119.0	6.0	.	.	.	.	.	.	.	.	.	109	.	.
SSP 6905	142.3	6.0	.	.	.	.	.	.	.	.	.	.	.	32
SSP 7001	21.1	6.2	216	235	267	289	132	.	238	.	309	188	.	252
SSP 7003	18.4	6.4	6206	6638	6710	7044	4016	7253	4282	6824	8556	2521	5076	5070
SSP 7006	8.5	6.5	1338	2899	2592	1656	1722	2003	1649	1602	3686	1909	3299	2052

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 7007	13.3	6.6	3631	6254	4886	4062	4455	3813	8217	3706	3866	3596	4340	5782
SSP 7014	19.9	6.4	.	.	647	.	.	.	.	.	.	.	.	574
SSP 7015	11.2	6.5	.	.	917	.	.	.	.	.	.	.	.	.
SSP 7016	7.7	6.5	.	.	1585	.	.	.	.	.	.	.	.	.
SSP 7017	17.8	6.3	.	.	.	126	120	.	217	.	.	.	.	.
SSP 7021	19.9	6.3	.	.	.	.	.	.	.	.	340	.	827	.
SSP 7025	17.9	6.3	269	181	.	.	180	.	324	.	191	146	72	117
SSP 7027	14.1	6.3	78	55	133	.	74	171	.	311	106	129	149	174
SSP 7102	27.0	6.3	4445	4066	4938	5819	4319	3913	11460	4369	4195	4181	4176	4492
SSP 7104	26.9	6.3	.	5181	.	12127	4423	7178	12050	6933	5618	5511	5872	5256
SSP 7105	26.2	6.3	701	1366	940	2056	1103	1349	733	1338	1425	1599	1430	1605
SSP 7107	24.0	6.4	153	212	315	.	270	204	293	128	208	150	382	364
SSP 7111	27.0	6.4	3281	6937	8950	.	4963	6120	.	5900	6193	4030	3636	4556
SSP 7114	26.8	6.5	29538	34054	25920	.	32741	26809	19646	26885	18861	27887	29220	16457
SSP 7117	22.1	6.3	.	.	.	.	.	.	5453	.	.	.	.	.
SSP 7118	26.3	6.4	225	.	.	176	.	64	.	44	155	172	139	181
SSP 7119	25.8	6.5	.	.	.	148	.	.	.	.	171	.	.	146
SSP 7120	26.2	6.2	.	.	.	.	468	.	856	.	.	.	.	.
SSP 7123	24.3	6.3	.	.	.	10216	.	.	.	.	.	.	.	361
SSP 7125	23.9	6.2	245	280	162	.	241	280	437	160	300	149	.	114
SSP 7126	26.6	6.4	244	168	182	277	225	177	.	164	366	416	.	200
SSP 7142	22.0	6.4	23374	23256	30301	32867	29439	28530	29288	37929	41712	40424	44652	33283
SSP 7206	29.1	6.3	13421	12370	12829	9889	14463	16057	10559	17180	14490	17633	10661	16260
SSP 7207	27.6	6.3	10545	12712	11883	.	15319	10302	.	13538	11811	12990	17383	14369
SSP 7208	30.4	6.3	.	13044	17339	25912	14663	14882	10993	17024	14568	18651	18682	18410
SSP 7209	29.5	6.4	13028	15517	12589	.	13509	13908	10403	12208	16530	12897	.	.
SSP 7210	28.4	6.5	14069	14000	14210	20334	18932	16819	22116	16668	15908	14882	18747	13985
SSP 7211	27.9	6.2	.	.	.	.	324	.	.	.	.	.	.	.
SSP 7212	28.7	6.3	.	.	.	.	.	.	.	.	.	.	9404	.
SSP 7213	30.9	6.3	.	.	.	15220	.	.	9466	.	.	.	.	.
SSP 7302	34.9	6.2	.	433	.	.	.	.	.	.	.	.	.	.
SSP 7305	35.6	6.3	331	336	223	490	266	316	433	260	368	.	466	1004
SSP 7311	32.6	6.3	33079	20852	36396	17012	25457	24280	31486	20585	29634	24358	22665	14858
SSP 7312	36.4	6.3	468	383	451	1047	567	358	767	367	545	437	525	1105
SSP 7313	35.9	6.2	84	.	120	89	51	72	97	55	.	244	63	412
SSP 7317	34.6	6.4	.	.	.	222	.	.	.	.	.	.	.	.
SSP 7319	36.8	6.2	.	.	.	.	.	.	.	.	.	328	.	.
SSP 7320	36.1	6.1	.	.	.	.	.	.	.	.	.	157	.	.
SSP 7321	36.1	6.5	.	.	121	90	121	143	78	114	202	.	.	.
SSP 7322	32.5	6.4	36063	21039	16090	19301	19049	47933	14920	20104	19253	37844	21249	16374
SSP 7323	37.5	6.4	354	292	.	.	235	354	.	213	324	285	309	322
SSP 7401	38.1	6.2	191	188	117	280	125	155	209	165	213	197	172	229
SSP 7404	41.8	6.3	644	520	413	448	324	383	1031	337	597	616	500	797
SSP 7405	37.3	6.3	184	279	281	301	222	166	331	181	295	315	291	304
SSP 7406	39.1	6.3	257	248	258	371	265	223	344	189	243	359	302	456
SSP 7408	38.9	6.3	318	214	350	270	232	155	363	86	291	368	204	284
SSP 7409	41.5	6.3	81	93	.	.	.	.	.	.	.	.	.	107
SSP 7410	39.5	6.3	109	119	121	91	63	81	143	55	102	114	94	152
SSP 7415	41.9	6.4	180	153	207	134	284	146	230	110	344	293	208	337
SSP 7416	38.8	6.5	2919	796	1182	1383	632	2637	992	600	3138	1082	910	5481

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 7418	38.5	6.5	1664	1235	1805	1434	941	849	1706	701	1664	1888	1650	1218
SSP 7419	39.4	6.5	.	245	.	165	275	302	162	212	142	195	359	105
SSP 7422	40.7	6.3	128	.	.	.	.	.	.	.	.	133	.	125
SSP 7423	39.8	6.4	340	.	251	192	.	172	.	.	288	437	217	442
SSP 7503	47.6	6.3	383	307	472	555	316	345	571	364	504	448	417	527
SSP 7506	46.8	6.3	559	412	740	.	448	704	.	524	1294	1541	1139	1655
SSP 7507	49.4	6.3	59	44	.	.	.	.	1196	.	.	.	.	.
SSP 7509	47.2	6.4	4915	380	3606	337	4587	2289	589	2053	2367	2506	2634	5797
SSP 7513	44.4	6.5	211	180	106	166	.	150	211	125	100	.	169	.
SSP 7521	47.8	6.4	332	.	177	.	.	.	409	.	279	710	215	421
SSP 7522	46.9	6.5	106	.	133	.	172	125	.	.	.	216	198	367
SSP 7523	48.3	6.6	617	.	186	.	.	.	204	.	468	1762	459	558
SSP 7525	45.3	6.3	.	.	.	.	.	.	123	.	.	.	.	.
SSP 7527	47.5	6.2	.	.	.	.	.	.	1017	.	.	.	.	.
SSP 7528	45.3	6.2	.	.	.	408	.	.	.	.	.	.	.	.
SSP 7529	47.1	6.3	.	.	.	.	.	.	3020	.	.	.	.	.
SSP 7530	49.9	6.3	.	.	.	231	.	.	.	.	.	.	.	.
SSP 7532	44.9	6.2	103	40	534	.	62	43	119	25	112	105	54	158
SSP 7533	45.6	6.6	.	.	.	95	.	.	.	.	.	.	.	.
SSP 7534	43.6	6.6	.	.	.	.	.	.	.	.	.	113	.	.
SSP 7602	60.1	6.2	169	151	158	140	110	132	197	103	159	140	115	248
SSP 7603	50.6	6.3	94	312	512	236	275	346	254	279	315	520	291	637
SSP 7608	58.5	6.4	61	33	41	40	64	66	72	.	54	55	40	86
SSP 7609	50.2	6.4	.	191	.	.	237	.	.	.	.	.	.	.
SSP 7610	52.2	6.4	1060	520	1784	.	550	572	.	553	1605	1991	2135	3007
SSP 7617	58.8	6.3	224	.	.	.	157	165	.	116	.	70	.	224
SSP 7618	55.2	6.3	251	.	.	.	149	176	55	102	.	255	.	385
SSP 7619	52.2	6.3	358	.	.	1530	.	.	166	.	.	.	.	.
SSP 7620	60.1	6.5	82	.	.	.	83	.	.	.	.	137	203	162
SSP 7622	53.9	6.4	77	40	71	49	27	39	57	.	75	95	78	132
SSP 7623	50.5	6.3	.	.	.	497	.	.	.	.	.	.	.	.
SSP 7625	52.3	6.4	.	.	.	.	27	.	.	.	107	.	.	.
SSP 7627	51.0	6.4	258	.	.	306	.	.	1248	.	.	.	.	.
SSP 7701	63.1	6.2	327	338	364	290	296	357	414	286	462	584	388	796
SSP 7702	64.0	6.2	1426	1392	1888	778	856	1409	2498	816	2375	2423	1261	3698
SSP 7703	61.0	6.3	135	113	62	111	65	96	112	85	143	49	38	94
SSP 7704	67.1	6.3	.	57	45	48	43	.	61	.	72	42	62	64
SSP 7706	64.8	6.3	148	160	160	83	114	178	225	.	232	228	132	431
SSP 7707	61.9	6.4	.	57	70	105	88	87	80	61	67	72	129	112
SSP 7708	62.7	6.4	607	364	.	69	88	56	131	37	50	329	372	185
SSP 7715	61.9	6.3	133	.	71	.	.	43	.	34	53	105	.	113
SSP 7716	63.6	6.5	96	27	105	91	33	58	61	41	72	108	71	195
SSP 7717	63.1	6.6	.	.	132	429	228	274	589	221	284	211	.	538
SSP 7718	61.3	6.3	.	.	.	122	.	.	.	.	.	.	.	.
SSP 7719	63.0	6.3	.	.	.	112	.	118	132	108	151	133	.	264
SSP 7720	62.1	6.5	.	.	.	72	.	.	.	.	.	.	.	.
SSP 7802	84.5	6.2	33	26	31	.	.	.	56	.	.	.	.	.
SSP 7804	84.9	6.3	99	84	83	108	48	121	92	57	122	123	103	184
SSP 7806	101.1	6.3	.	14	.	.	.	.	.	.	22	.	.	.
SSP 7810	89.2	6.5	14	16	13	12	.	.	.	.	15	.	.	.

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 7811	81.1	6.5	20	17	20	15	16	.	.	.	20	.	.	.
SSP 7813	73.7	6.3	.	.	.	.	27	53	40	.	.	.	.	64
SSP 7814	100.6	6.4	29	.	22	.	.	.	.	.	.	.	.	28
SSP 7815	79.0	6.3	.	.	.	.	.	.	.	.	.	50	.	86
SSP 7816	99.4	6.5	.	.	.	20	.	.	.	.	.	.	.	.
SSP 7902	143.8	6.3	15	18	15	.	.	.	16	.	.	.	.	30
SSP 8001	18.3	6.6	791	1162	786	405	624	1076	.	1281	1697	605	1274	917
SSP 8002	6.4	6.6	31391	31364	27746	26469	28875	28847	27883	36853	27386	28176	23794	26843
SSP 8003	18.4	6.6	2902	3149	.	2346	2429	3646	1982	3474	4423	1922	2005	.
SSP 8005	21.4	6.7	1526	2238	2360	911	863	909	.	1022	3343	2731	780	3124
SSP 8007	7.3	6.7	38202	42215	42830	50050	46239	53635	37969	40200	41628	43001	49123	46897
SSP 8012	20.4	6.8	3799	3207	3814	2921	3200	3243	3663	3122	3701	3575	4136	3904
SSP 8015	6.4	7.1	70205	76422	64749	73866	65852	49066	62519	96058	92027	84772	79436	68028
SSP 8016	5.9	7.8	19448	30852	22854	18104	23540	21890	16935	19555	29282	22130	25440	19268
SSP 8019	14.8	8.1	310	694	453	.	.	377	358	477	314	260	287	372
SSP 8023	21.9	7.9	7513	5239	6295	.	7638	6519	8568	6223	8031	5824	11621	6697
SSP 8024	18.1	6.8	210	143	192	178	.	.	306	.	154	209	168	169
SSP 8027	18.2	6.8	.	152	136	155	66	.	.	.	117	137	139	170
SSP 8030	8.7	7.9	.	.	.	428	.	.	.	.	.	.	.	.
SSP 8032	19.4	6.7	.	.	.	.	.	.	271	.	.	.	.	.
SSP 8033	21.8	6.8	.	.	.	.	.	.	2216	.	.	.	.	.
SSP 8035	19.4	6.8	.	.	.	.	.	.	652	.	.	.	.	.
SSP 8039	18.3	6.8	299	471	431	574	422	646	514	701	597	695	473	471
SSP 8043	7.8	6.8	151	213	183	.	.	.	162	.	175	158	99	165
SSP 8104	22.5	6.7	437	746	1321	.	991	281	1867	117	481	1134	1646	2118
SSP 8105	24.2	6.7	1790	564	5177	.	910	1492	1476	1381	1306	7026	4373	11179
SSP 8107	26.0	6.8	132	146	146	181	144	202	111	243	150	130	138	135
SSP 8108	22.0	6.8	1924	2331	2245	4407	2556	.	2795	.	2228	3036	3239	3129
SSP 8115	22.1	7.9	11167	10945	12547	.	14807	12718	9823	11269	14399	13841	16135	17568
SSP 8117	22.1	6.6	390	618	503	794	461	762	420	748	418	480	473	347
SSP 8118	26.8	7.9	461	346	514	.	1608	760	1172	559	504	516	493	286
SSP 8119	26.8	6.8	492	.	.	.	.	.	6858	.	167	205	221	333
SSP 8120	22.7	6.9	.	.	.	16619	.	.	.	.	.	.	.	.
SSP 8121	24.5	6.6	.	.	.	.	.	.	215	.	.	.	.	.
SSP 8124	25.2	7.6	.	.	.	.	.	.	.	.	.	.	.	145
SSP 8125	27.0	6.7	10837	16127	12569	7661	14760	14718	12093	15943	13513	11357	18094	12330
SSP 8126	27.1	7.3	.	.	13582	4279	.	.	.	.	.	.	2720	.
SSP 8127	26.0	6.7	.	.	.	429	.	.	.	.	.	.	.	.
SSP 8207	31.0	6.8	1217	343	381	.	.	.	46	.	391	2208	516	3403
SSP 8211	28.4	6.7	14373	12532	11729	14215	13460	14911	10510	14659	19814	22871	22501	12700
SSP 8215	29.4	7.7	62	.	.	.	.	.	.	.	.	.	.	.
SSP 8217	30.9	6.7	.	.	1488	.	.	.	.	.	1525	942	2534	.
SSP 8219	29.2	6.8	.	.	.	.	.	.	9126	.	.	532	.	1477
SSP 8222	27.5	7.3	23226	32043	13136	7813	36248	22765	41300	20407	21552	19254	8491	15550
SSP 8225	30.4	6.6	8826	10332	10053	15866	8016	10909	6188	12265	12503	9651	7941	10868
SSP 8308	35.4	6.8	577	490	635	217	371	253	573	216	619	579	518	725
SSP 8314	35.2	6.7	74	96	90	61	60	71	117	46	108	102	79	102
SSP 8401	37.1	6.6	96	97	119	103	67	79	99	65	118	107	97	114
SSP 8402	39.9	6.6	200	173	174	142	62	164	206	146	195	218	298	296
SSP 8421	42.2	6.8	.	.	.	.	.	.	.	.	240	.	.	.

SSP	Mw	pI	C24	Col-0	Condara	Cvi-0	Ll-0	Ma-0	Mr-0	Mt-0	Nd-0	Oy-0	Tsu-0	Ws
SSP 8423	38.1	6.8	.	.	.	.	.	.	.	.	.	222	.	.
SSP 8424	39.1	6.7	.	.	.	102	.	.	.	.	.	.	122	.
SSP 8427	41.3	6.6	104	61	120	.	.	70	.	.	.	.	.	174
SSP 8502	42.7	6.6	65	49	215	48	165	49	72	.	77	171	72	366
SSP 8508	47.3	6.8	.	453	.	1310	2278	373	3397	334	.	1481	.	3765
SSP 8517	48.2	7.0	158	190	174	179	2567	70	3263	74	97	232	96	1282
SSP 8518	48.8	6.7	.	80	3784	.	.	.	.	.	.	.	.	.
SSP 8520	48.3	7.8	58	118	70	.	.	.	332	84	.	117	49	286
SSP 8522	47.8	6.7	933	150	362	307	775	346	1051	291	.	1551	2743	2772
SSP 8523	47.2	7.3	.	.	.	1017	.	.	.	.	230	.	.	.
SSP 8524	48.6	6.7	148	67	344	58	.	.	780	.	92	132	78	955
SSP 8525	49.4	6.7	.	.	181	.	.	.	.	.	.	.	.	.
SSP 8606	51.4	6.7	240	106	172	214	123	143	399	119	236	438	358	.
SSP 8611	57.6	6.7	.	61	.	.	275	30	88	26	.	69	.	289
SSP 8615	50.4	6.8	936	339	923	935	755	483	1941	414	758	2301	853	5128
SSP 8620	51.1	6.6	386	.	429	403	.	.	.	.	.	.	456	770
SSP 8621	50.7	6.7	199	.	156	.	.	.	.	.	.	458	.	.
SSP 8622	60.1	6.6	153	.	.	.	136	.	.	.	.	186	195	487
SSP 8623	49.5	6.6	.	.	.	.	.	.	443	.	.	.	.	.
SSP 8624	50.1	6.7	.	354	341	1487	856	214	1063	209	201	662	.	1806
SSP 8712	66.5	8.2	43	33	22	.	.	.	.	.	49	46	.	77
SSP 8713	61.1	6.6	.	107	168	117	61	97	.	61	234	119	160	.
SSP 8715	61.2	7.5	42	35	23	26	13	31	39	21	37	34	36	51
SSP 8716	64.0	6.7	20	10	14	10	.	.	10	.	17	16	9	26
SSP 9013	6.8	> 8.3	52322	54809	51726	23512	57999	56644	46316	55240	13120	.	29913	6818
SSP 9019	21.1	8.3	26612	13617	26030	.	13813	20940	12561	23945	29406	33998	38898	14071
SSP 9020	5.8	> 8.3	5287	11372	6649	6834	5592	10159	8891	8134	6673	7012	9553	6728
SSP 9030	5.9	> 8.3	9528	11900	5344	3530	6191	10326	5756	6622	7009	4471	7340	3583
SSP 9031	18.5	> 8.3	741	2691	936	.	913	2939	5130	958	659	329	1062	644
SSP 9032	19.2	> 8.3	.	.	.	.	.	.	3184	.	.	.	.	.
SSP 9108	22.0	> 8.3	16679	17332	16143	.	19089	19227	14292	20125	18725	17625	20582	16192
SSP 9109	22.1	> 8.3	10740	9230	8865	40547	9242	11721	8563	10043	9316	8115	11075	8923
SSP 9210	27.9	> 8.3	.	.	.	19035	.	11188	1806	11038	15903	2191	16633	6780
SSP 9212	31.0	8.2	366	234	571	247	524	448	190	338	427	682	752	1114

Curriculum Vitae

Name: Martin Christian Rübelt
Accredited Food Chemist

Date of Birth: 05/25/1971 in Saarbrücken (Germany)

EDUCATION

2001 – Present	**Ph.D. Student** Technical University Munich (Germany) – Professor Dr. Karl-Heinz Engel in collaboration with Monsanto Co., MO (USA) and Federal Research Centre for Nutrition (Germany) – Professor Dr. Klaus-Dieter Jany. Thesis: Applicability of proteomics to assess effects due to genetic engineering in the context of natural variability using *Arabidopsis thaliana* as a model organism.
1997	**Accredited Food Chemist** 2nd STATE EXAMINATION; State Institute of Health and Environment, Saarbrücken (Germany)
1996 – 1997	State Institute of Health and Environment, Saarbrücken (Germany) Conducted analyses of food products and drugs and provided evaluation reports for the principal investigator regarding their safety for human consumption and their compliance to the German Food & Commodities Law (LMBG).
1995	**Food Chemist** 1st STATE EXAMINATION (University)
1991 – 1995	Food chemistry studies at the University of Kaiserslautern (Germany) Department of Chemistry, Section of Food Chemistry and Environmental Toxicology Thesis: "Evaluation of two synthesis strategies for 3-[1-Methyl-4-(2-alkoxyacetamido)pyrrol-2-carboxamido]-1-N,N-dimethylaminopropan and investigation of its cytotoxicity."
1991	**Matriculation** Hoffenfels Gymnasium Zweibrücken (Germany)

PROFESSIONAL WORK

2001 – Present	**Research Scientist –** Monsanto Company, Regulatory Science, St. Louis (USA) Validated and conducted differential protein expression studies to investigate effects due to genetic engineering.
1997 – 2001	**Laboratory manager –** Laboratory Dr. Ruebelt, Homburg (Germany) Laboratory for environmental and food analyses and evaluation, certified as second opinion laboratory according to Section 42 of the German Food & Commodities Law (LMBG)
1998 – 2001	**Consultant –** R&H Hygiene Consulting, Homburg (Germany) Organized and conducted seminars about Hazard Analysis and Critical Control Point (HACCP) concept, food safety, and food hygiene for food industry. Counseled food companies in implementing HACCP systems.
1999 – 2001	**Associate teacher –** University Hospital of Saarland, Homburg (Germany) Conducted lectures in food safety, microbiology, and food toxicology.

PRESENTATIONS

November 2004	University of Missouri-Columbia, Special Seminar series: "Differential 2DE as an analytical tool to investigate effects due to genetic engineering in the context of natural variability." (Oral presentation)
June 2004	Donald Danforth Plant Science Center, Special Seminar series: "Differential proteomics in biotechnology: The need for standardization and baseline studies." (Oral presentation)
April 2004	Analytical Environmental Immunochemical Consortium (AEIC) Spring Meeting, Memphis (TN): "Proteomics: A tool to investigate effects due to genetic engineering in the context of natural variability." (Oral presentation)
October 2003	First Annual MONSANTO and University of Missouri Proteomics Symposium, Columbia, USA: "Validation and feasibility study for the proteome analysis of *Arabidopsis thaliana* seeds. (Poster) "Proteomics as a tool to investigate effects due to genetic engineering in the context of natural variability: A model study using *Arabidopsis thaliana*." (Poster)
September 2003	Monsanto Brussels, Belgium: "Proteomics as an analytical tool to investigate effects due to genetic engineering in the context of natural variability: a model study using *Arabidopsis thaliana*." (Oral presentation)
September 2003	EURO FOOD CHEM XII, Strategies for safe food: Challenges in Organization and Communication, Brugge, Belgium: "Novel foods – safety assessment: Method development for proteome analysis of *Arabidopsis* seeds produced by different ecotypes (accessions) and by transgenic events" (Oral presentation)
September 2003	International Meeting on Proteome Analysis, Munich, Germany: "Proteomics as a tool to investigate effects due to genetic engineering in the context of natural variability: A model study using *Arabidopsis thaliana*." (Poster)
July 2003	Plant Biology 2003, Honolulu, Hawaii: "Quantification of natural variability in the seed proteome of *Arabidopsis thaliana* ecotypes." (Poster)
June 2003	Protein Characterization Center, Monsanto Co., St. Louis, USA: "Proteomics - A tool to investigate effects due to genetic engineering in the context of natural variability: a model study using *Arabidopsis thaliana*." (Oral presentation)
June 2003	TCM Ag Tech, Monsanto Co., St. Louis, USA: "Proteomics as a tool to investigate unintended effects due to biotechnological engineering in the context of natural variability: a model study using *Arabidopsis thaliana*." (Oral presentation)
June 2002	TCM Ag Tech, Monsanto Co., St. Louis, USA: "Method Development for Proteome Analysis of *Arabidopsis* Seeds Produced by Different Ecotypes (Accessions) and by Transgenic Events." (Poster)