Katholieke Universiteit Leuven
Faculteit Bio-ingenieurswetenschappen

DISSERTATIONES DE AGRICULTURA

Doctoraatsproefschrift nr. 850 aan de Faculteit
Bio-ingenieurswetenschappen van de K.U.Leuven

A proteomics approach to study core breakdown disorder in stored 'Conference' pears

Promotoren:
Prof. B. Nicolaï, K.U. Leuven
Prof. J. Vanderleyden, K.U. Leuven

Leden van de jury:
Prof. Maurice De Proft, voorzitter,
K.U. Leuven

Prof. J. Robben, K.U. Leuven
Prof. R. Swennen, K.U. Leuven
Prof. J. Lammertyn, K.U. Leuven
Dr. K. Lilley, University of Cambridge, UK
Dr. Maarten Hertog, K.U. Leuven

.

Proefschrift voorgedragen tot
het behalen van de graad van
Doctor in de
Bio-ingenieurswetenschappen

door

Romina Paola
Pedreschi Plasencia

APRIL 2009

Acknowledgments

To complete this PhD I had to work together with several research groups. Thus, I would like to thank all the people that were part of this team.

I am very grateful to Prof. Bart Nicolaï for trusting me from the beginning. Bart, I extremely appreciate the freedom you gave me to carry out my research and the chance to work with different research groups within and outside Belgium. My deep gratitude also goes to my co-promoter Prof. Jozef Vanderleyden.

Deep thanks to Dr. Maarten Hertog for always being so positive and always being there to answer most of my doubts. Maarten, It was great working under your supervision. You taught me to give the best out of me and value any result even the discouraging ones.

Thank you to Prof. Johan Robben and Dr. Jean Paul Noben. Johan thank you for your help, kindness and patience to introduce me into the world of protein mass spectrometry identification.

Thank you to Dr. Kathryn Lilley and her team at Cambridge Centre for Proteomics not only for opening the doors of your lab to carry out part of my research but also for the great moments I spent in Cambridge. Your good will and feedback is enormously appreciated.

Thank you to Prof. Jeroen Lammertyn for introducing me into the multivariate statistics world.

Thank you to all the members of the examination board for their critical reading of my manuscript in order to improve the quality of my PhD dissertation. Thank you to Prof. Maurice De Proft for facilitating my defense as chairmain of the jury.

Thank you to Prof. Rony Swennen for opening the doors of the Tropical Lab to carry out part of my research. Special thanks to Annelies and Sebastien for their help and great moments shared at conferences, in the lab and outside the lab.

Thank you to the International Relations Office of the K.U Leuven (IRO scholarship).

Thanks to all my lab colleagues. Special thanks to Sophie, Elfie, Bert Vanderbosch, Bert Verlinden, Jeroen Tirry for their help during these years. Deep thanks to Ms. Josee Verlaenen for your kindness and willingness to help. It is a pity that you left the group. We certainly miss you.

Thanks to my former promoter Dr. David Campos with whom I still collaborate and who introduced me for the first time in research.

My immense gratitude goes to my family: parents (Roberto and Maria Violeta), brothers (Roberto and Franco) and sisters (Fiorella and Tatiana) for their love and support coming from all parts of the world. Special thanks to my mom who always supported the idea of studying abroad and enrich myself most important than academically I must say personally. You certainly remain alive in my thoughts day to day.

'True friends are for life, no matter where and how you are'. Thanks to all those who shared with me moments of happiness, sadness, stress and bad mood. Special thanks to Giuliana , Wilfredo, David and Carla for being always unconditionally present and to special friends I made in Leuven: Dihn, Victor, Kate, Sven, Nico, Sandra, Gemma, Daniela, Arturo, Javier.

Gracias a todos ellos/ellas que hicieron y hacen que la estadia fuera de casa sea más que agradable, inolvidable!

Samenvatting

Consumenten hechten belang aan de beschikbaarheid van verse groenten en fruit doorheen het gehele jaar. Tuinbouwgewassen zijn echter bederfelijk van aard. Vandaar dat er diverse methoden zijn ontwikkeld voor het verhogen van de bewaarbaarheid. Doorgaans wordt bewaring onder een gewijzigde bewaaratmosfeer (Controlled Atmospheres, CA) toegepast om processen als ademhaling, ethyleenproductie, afrijping en veroudering tegen te gaan. Hierbij wordt de temperatuur verlaagd en de atmosfeersamenstelling aangepast door het zuurstofniveau te verlagen en het koolstofdioxideniveau te verhogen. Een dergelijke CA bewaring is product specifiek en niet per definitie succesvol. Het optreden van fysiologische gebreken is dan ook niet altijd te voorkomen.

Peren van de variëteit 'Conference' zijn gevoelig aan het optreden van bruinverkleuring en holtevorming indien deze niet correct worden bewaard. Deze bewaarafwijking is vermoedelijk het gevolg van een gebrekkige ademhaling die vervolgens een lawine van metabolische processen op gang brengt. Eerdere studies hebben zich voornamelijk gericht op de rol van gastransportprocessen en het antioxidantmetabolisme bij het induceren van deze bewaarafwijking, maar hebben daarbij belangrijke biochemische aspecten grotendeels verwaarloosd. In deze thesis wordt bruinverkleuring en holtevorming in Conference peer bestudeerd vanuit een breder perspectief middels een holistische studie van het proteoom. Het primaire doel van deze doctoraatsstudie is inzicht te verkrijgen in de fysiologische achtergrond van bruinverkleuring en holtevorming gebruik makend van eiwitanalysetechnieken.

Om bruinverkleuring en holtevorming in 'Conference' peer te bestuderen was het noodzakelijk geschikte technieken voor eiwitextractie, twee-

dimensionale gelelektroforese (2-DE) en data-analyse te optimaliseren. Eiwitextractieopbrengst werd geoptimaliseerd door toepassing van extractiebuffer met hoge zuurtegraad en hoge concentraties aan dithiothreitol. Een eerste 2-DE studie gericht op het karakteriseren van al dan niet aangetast weefsel toonde aan dat zowel de ademhaling als ook de eiwitsynthese, het antioxidantmetabolisme en de ethyleenbiosynthese was aangetast. In bruin weefsel was celdood aantoonbaar aanwezig.

Een terugkerend probleem bij gelgebaseerde eiwitdata is de aanwezigheid van ontbrekende waarden. Deze vormen een uitdaging voor een correcte statistische data-analyse en interpretatie. In deze thesis werd dit probleem succesvol opgelost door op grond van Bayesiaanse principale component analyse waarden toe te kennen aan de ontbrekende observaties.

Het lange termijneffect van CA bewaarcondities op het metabolisme van peer werd bestudeerd. De verschillende combinaties van zuurstof en koolstofdioxide induceerde verschillen op het niveau van zowel de ademhaling als ook de eiwitsynthese, het antioxidantmetabolisme en de ethyleenbiosynthese.

Tenslotte werd het korte termijneffect van CA bewaarcondities op het metabolisme van peer weefselstukjes bestudeerd. Deze resultaten bevestigde goeddeels de resultaten verkregen voor het lange termijneffect. Voor wat betreft het ademhalingsmetabolisme werden slechts subtiele effecten waargenomen, uitgezonderd de inductie van de pentose-fosfaat-route bij afwezigheid van zuurstof. Vergelijkbaar aan het lange-termijneffect van een verlaagd zuurstofgehalte werden diverse allergenen en andere beschermingseiwitten gereguleerd. Er werden echter geen significante verschillen waargenomen voor de glutathion-ascorbaat-route wat ons doet vermoeden dat bruinverkleuring pas zichtbaar wordt wanneer deze metabole route wordt aangetast. Op grond van de proteoomstudie in deze thesis werd het mogelijk een aantal kandidaatbiomerkers te selecteren voor een vroegtijdige signalering van het optreden van bruinverkleuring en holtevorming tijdens de bewaring van Conference peer.

Abstract

Consumers demand fresh fruits and vegetables throughout the whole year. Horticultural crops are very perishable. Thus, strategies to control and extend their storage life are mandatory. Commonly, controlled atmosphere (CA) conditions are applied to retard respiration, ethylene production, ripening and senescence. The temperature is reduced and the air atmosphere composition is modified by reducing the oxygen content and by increasing the carbon dioxide composition. This CA strategy is commodity specific; inappropriate conditions may cause physiological disorders.

When improperly stored, pears of the variety 'Conference' are very susceptible to develop core breakdown which is characterized by flesh browning and cavities. This disorder is supposed to be the consequence of an impaired respiration triggering a cascade of metabolic events. Previous studies have mainly focused on understanding this disorder by gas transport and antioxidant system studies, thereby ignoring many biochemical aspects. In this dissertation, core breakdown has been studied in a broader perspective by using a holistic proteomics approach. The main objective of this work was to understand the physiological background of core breakdown disorder by using proteomics tools.

To start the study of core breakdown in 'Conference' pears it was necessary to fine tune protein extraction, two-dimensional electrophoresis (2-DE) and data analysis protocols. Subsequently a complete 2-DE study aiming at characterizing tissue either or not affected by core breakdown revealed impairment of respiration, defense related and ethylene biosynthesis pathways. Cell death was also evident in brown tissue.

One common problem associated to gel based proteomics data is the

presence of missing values. Missing values present a challenge not only in terms of how to deal with them but also in the interpretation of the data. It was shown that missing values in gel-based proteomics data can be treated satisfactorily by means of the Bayesian Principal Component imputation method.

The effect of long term exposure to different CA conditions on the metabolism of stored pears was studied. Different combinations of oxygen and carbon dioxide concentrations induced differences at the level of respiration, defense related and ethylene pathways.

Finally, the short term effect of extreme CA conditions on the metabolism of pear slices was studied. Results partially confirmed the findings for the long term exposure of pears. Only subtle changes in respiration pathways were observed except for the pentose phosphate pathway which was activated under anoxia. Similarly to long term exposure of intact pears to anoxia, allergens and other defense related enzymes were regulated. No significant changes in the glutathione-ascorbate pathway were found leading us to suspect that browning development appears as soon as this pathway collapses. This proteomics approach allowed the selection of potential candidate markers for the early-stage detection and tracking of core breakdown during storage of pears.

Contents

Symbols and Abbreviations

ACC ox	ACC oxidase
ACC	1-aminocyclopropane-1-carboxylate
Aco	Aconitase
ANOVA	analysis of variance
AOX	cyanide resistant alternative oxidase
APX	ascorbate peroxidase
asinh	inverse hyperbole sine
BP	band pass
BPCA	Bayesian principal component analysis
CA	controlled atmosphere
CBB	coomassie brilliant blue
CHAPS	3-(3-cholamidopropyl)dimethylammonio-1-propane sulfonate
CoAT	acetyl CoA transferase
DA	discriminant analysis
2-DE	two-dimensional electrophoresis
DHAR	dehydroascorbate reductase
DIGE	differential in gel electrophoresis
DTT	dithiothreitol
EM	expectation-maximization like repetitive algorithm
emPAI	exponentially modified protein abundance index
Eno	enolase

ESI	electrospray ionization
EST	expressed sequence tag
Fum	fumarase
GR	glutathione reductase
HED	hydroxyethyldisulphide
HSP	heat shock protein
IEF	isoelectric focusing
IPG	immobilized pH gradient
KNN	k nearest neighbor
LC	liquid chromatography
log	logarithmic
LV	latent variable
M	peptide molecule
m/z	mass/charge ratio
MALDI	matrix assisted laser desorption ionization
MDAR	monodehydroascorbate reductase
MDH	malate dehydrogenase
ME	malic enzyme
MI	multiple imputation
MS	cobalamine independent synthase
MS	mass spectrometry
MS/MS	tandem mass spectrometry
Mudit	multidimensional protein identification technology
n^+	net charge of the ions
nH	number of protons attached to the molecule
NIPALS	non linear estimation by iterative partial least squares
PAGE	polyacrylamide gel electrophoresis
PC	principal component
PCA	principal component analysis
PCR	principal component regression

PDI	protein disulphide isomerase
PEB	protein extraction buffer
PGIP	polygalacturonase inhibiting protein
pI	isoelectric point
PLS	partial least squares
PLSR	partial least squares regression
PMF	peptide mass fingerprint
PMSF	phenyl methyl sulphonyl fluoride
PR	pathogenesis related
PPO	polyphenol oxidase
PVPP	polyvinylpolypyrrolidone
Q	quadrupole
ROS	reactive oxygen species
SD	standard deviation
SDS	sodium dodecyl sulphate
sqrt	square root
SVD	singular value decomposition
TBP	tributylphosphine
TCA	tricarboxylic acid
TK	transketolase
TOF	time of flight
TPI	triosephosphate isomerase
UDP-GP	UDP glucose phosphorylase
VIP	variable importance plot

Chapter 1

General Introduction

1.1 Rationale

The food industry is committed to deliver high quality, safe and nutritious products to the consumer. Achievement of this goal is not trivial and needs to consider the current contrasting world nutritional problems - over consumption and under nourishment (Hall *et al.*, 2008). Fruit and vegetables are important sources of many essential constituents. Therefore, their consumption is associated with health improvement. Fruits and vegetables are highly perishable with their postharvest life depending on the rate of stored food reserve usage and water loss.

Throughout the postharvest handling chain of fruits and vegetables, food losses can vary in magnitude and can occur at all stages from harvesting, during transport, handling, storage, processing, and marketing until they reach the end consumer. Fruits and vegetables are constantly exposed to biotic and abiotic stresses before and after harvesting. During harvesting and handling, a series of mechanical stresses are imposed which induce different cellular responses such as increased respiration rate, ethylene production, increased susceptibility to pathogen attack and wound responses, leading to a reduction of the quality and storability of the crops (Gomez-Galindo & Sjiholm, 2007). The perishability of fruits and vegetables after harvest demands special efforts to avoid rapid deterioration. Controlled atmosphere storage (CA) is commonly applied to fruits and vegetables to extend their

shelf life. It involves temperature reduction and modification of the air composition (oxygen reduction and carbon dioxide increase) in order to retard respiration, ethylene production and senescence (Kader, 2002). However, these CA conditions may impose additional abiotic stresses when the oxygen and carbon dioxide partial pressure are too low or too high, respectively. Different physiological disorders associated with improper CA management are chilling injury, failure to ripen, development of off flavors (Kuo & Parkin, 1989; Mohammed & Wickham, 1997) and browning disorders (Franck *et al.*, 2007).

Pome fruits, apple and pear, are native to most of Europe, the Near East and Asia. In terms of world production, they are considered among the major four fruit classes and the second major fruit in moderate climates (Knee, 1993). Pome fruits are generally prone to internal browning disorders when improperly stored. The pear variety 'Conference' is highly susceptible to internal browning and to subsequently develop core breakdown during CA storage. As these disorders have been shown to be related to the gas composition of the storage atmosphere (Lammertyn *et al.*, 2000), much attention has been paid to the mechanism of gas transport in the fruit (Lammertyn *et al.*, 2003b,a; Ho *et al.*, 2006b, 2008) and the microstructure of the intercellular space as the main transport path for metabolic gasses (Verboven *et al.*, 2008). Core breakdown in 'Conference' pears is believed to be the result of an imbalance between oxidative and reductive processes due to the formation of gas gradients in the fruit. This induces accumulation of reactive oxygen species leading to membrane damage that finally results in the enzymatic oxidation of polyphenols to brown colored compounds (Franck *et al.*, 2007; Pedreschi *et al.*, 2009a). Core breakdown is poorly understood from a physiological point of view as it is the result of disturbances of several biochemical processes. Enzymes (proteins) are key players in these biochemical processes. Thus, to improve our understanding, there is a need for a more holistic approach such as the use of proteomics. Knowledge on how horticultural crops respond to environmental changes and industrial manipulation is of key importance for quality assurance and process optimization in the food industry (Gomez-Galindo & Sjiholm, 2007).

1.2 Objectives and outline of the thesis

The main objective of this thesis is to understand the physiological implications of core breakdown disorder in 'Conference' pears by using proteomics tools. In order to achieve this goal, different subobjectives were defined.

- to identify at the proteome level characteristic proteins in brown affected and non-affected tissue,
- to apply proper statistical methods to gel based proteomics data,
- to evaluate at the proteome level the effect of different CA conditions on pear metabolism,
- to study the effect of short term exposure of pear tissue slices to extreme CA conditions.

The outline of this thesis is as follows. In *Chapter 2*, the literature on stress physiology and on core breakdown disorder in Conference pears will be reviewed as well the state-of-the-art of proteomics research.

Fine tuning of protein extraction and two dimensional electrophoresis (2-DE) protocols for pear parenchyma tissue will be presented in *Chapter 3*. For the first time, a quantitative proteomic analysis will be carried out comparing brown affected tissue with non-affected tissue. Proteins associated to core breakdown will be identified. The results presented in this chapter were published in Pedreschi *et al.* (2007).

Chapter 4 deals with the assessment of proper statistical tools to analyze gel-based proteomics data in a multivariate statistical context. The issue of how to handle missing data will be discussed. Different approaches to deal with missing data will be presented and some recommendations provided in order to draw sound conclusions from gel-based proteomics experiments. These results were published in Pedreschi *et al.* (2008b).

In *Chapter 5* the optimized protocols and statistical approaches will be used to evaluate the effect of different CA conditions on the metabolism of

long term stored pears. The results described in this chapter were described in Pedreschi *et al.* (2008a).

In *Chapter 6*, differential in gel electrophoresis (DIGE) will be used to study the effect of the short term exposure of pear slices to extreme gas compositions, thereby eliminating most of the barriers to gas transport. The results described in this chapter were published in Pedreschi *et al.* (2009c).

Finally, some general conclusions, on-going work and suggestions for future work will be formulated in *Chapter 7*.

Chapter 2

Literature review

2.1 Introduction

For consumers, quality attributes as well as food safety are of high importance. Postharvest storage of fruits and vegetables aims to preserve quality and reduce product losses. However, the appearance of storage disorders can not be completely ruled out. Many physiological disorders do not cause external symptoms, removal of affected fruit prior to commercialization is hence difficult. Such fruit, however, are poorly appreciated by consumers and may have negative consequences with respect to the commercial image of the product.

To increase the shelf life of the products, ripening as well as biochemical processes involved in quality deterioration need to be retarded. To achieve this, besides optimal harvest date, temperature reduction as well as a modification of the normal gas atmosphere are needed to retard respiration rate and thus to control the overall metabolic activity. All these interventions imply applying 'stresses'. The imposed stresses will disrupt homeostasis and adjustment of metabolic pathways referred as 'acclimation' will take place (Shulaev *et al.*, 2008). If the fruit or vegetable is not capable of coping with the series of applied stresses, then physiological disorders may develop.

To understand at a physiological level the different implications of the applied stresses on the alteration of the normal homeostasis, there is a need for holistic platforms such as genomics, transcriptomics, proteomics and metabolomics. These approaches generate huge amounts of data which must be properly processed in order to draw sound biological conclusions.

Pears are native to most of Europe, the Near East and Asia. Economically, they are considered among the major four fruit classes and the second major fruit in moderate climates (Knee, 1993). The pear variety 'Conference' is the most popular in Europe as well as in Belgium. 'Conference' pears if improperly stored are susceptible to develop core breakdown characterized by flesh browning and cavities at advanced stages (Franck *et al.*, 2007).

The next sections give a review on plant stresses in relation to postharvest physiology of Conference pears leading to core breakdown. Also, the state of the art of proteomics research including important aspects of data analysis will be discussed and have been partially published in Pedreschi *et al.* (2009b)[1].

2.2 Plant stress physiology

Any alteration in a physiological condition caused by a factor that alters homeostasis can be defined as 'stress' (Gaspar *et al.*, 2002). Acclimation or adjustment of metabolic pathways is needed to restore homeostasis. According to Gaspar *et al.* (2002), stress in a physiological context can be defined as the amount of environmental pressure for change placed on an organism's physiology. The environmental stresses can be classified as either biotic or abiotic. Biotic stresses can be induced by a wide range of plant pathogens (bacteria, fungi and viruses) as well as herbivorous animals (Gaspar *et al.*, 2002). Within the category of abiotic stresses, a sub-division can be made between physical stresses (drought, temperature, radiation,

[1]Pedreschi et al. (2009). Proteomics for the food industry: opportunities and challenges. Critical Reviews in Food Science and Nutrition, accepted.

flooding, wind, magnetic field, mechanical wounding) and chemical stresses (air pollution, heavy metals, pesticides, toxins, pH, salinity; Gaspar *et al.* 2002). For instance, after harvesting and during handling, fruits and vegetables are constantly exposed to physical stresses that induce different cellular responses (Gomez-Galindo & Sjiholm, 2007). The temperature reduction as well as the change in the atmosphere composition applied during storage to extend the shelf life can be considered as abiotic stresses.

2.3 Postharvest physiology

After harvest, fruits and vegetables need to get adapted to the new conditions. Thus, they will basically utilize their stored resources since their nutrient supply is cut off the moment they are harvested. The acquisition, storage and utilization of energy are key processes in central metabolism. Respiration involves the breakdown of carbohydrates to carbon dioxide and water in order to release energy. In this process, oxygen acts as the final electron acceptor. The enzymatic oxidation of carbohydrates takes place through different processes: glycolysis, Krebs cycle and electron transport chain.

In the glycolytic pathway, the oxidation of glucose to pyruvate takes place. Plant glycolysis is very unique in terms of its flexibility (Dennis *et al.*, 1998; Fernie *et al.*, 2004). Carbon flow in this pathway is not limited by the absence of 'an essential enzyme' such as pyruvate kinase (Gottlob-McHugh *et al.*, 1992) because the glycolytic pathway is duplicated in different organelles (e.g., cytosol and plastid) and these organelles communicate. In addition, several steps in this pathway can be catalyzed by different enzymes. For instance, the conversion of phosphoenolpyruvate to pyruvate can be catalyzed by cytosolic pyruvate kinase, plastid pyruvate kinase, phosphoenolpyruvate phosphatase and the subsequent reactions of phosphoenolpyruvate carboxylase and malic enzyme (Theodorou & Plaxton, 1995; Plaxton, 1996). The interconversion of fructose-6-phosphate and fructose- 1,6-biphosphate is catalyzed by three enzymes instead of two as

in most organisms (Dennis *et al.*, 1998). Even though still not totally understood, it is believed that the glycolytic pathway in plants is regulated at the level of conversion of fructose-6-phosphate to fructose-1,6-biphosphate and phosphoenolpyruvate to pyruvate steps (Dennis *et al.*, 1998). In the tricarboxylic acid cycle (TCA), phosphoenolpyruvate is converted in the cytosol to malate and/or pyruvate. Subsequently, these organic acids enter the mitochondria to produce energy and reducing power yielding 15 ATP equivalents per pyruvate molecule (Fernie *et al.*, 2004). The key regulatory point in this cycle is the reaction catalyzed by the mitochondrial pyruvate dehydrogenase that converts pyruvate into acetyl-CoA. In the mitochondrial electron transport chain, the generated reducing equivalents from the TCA are used for the synthesis of ATP. Interestingly, ascorbate biosynthesis seems to be coupled to the electron transport chain (Bartoli *et al.*, 2000; Millar *et al.*, 2003).

In anaerobic conditions, the plant responds by fermentation; decarboxylation of pyruvate to acetaldehyde by pyruvate decarboxylase followed by the reduction of acetaldehyde to alcohol. Under anaerobic conditions, the NADH generated by glyceraldehyde 3-phosphate dehydrogenase needs to be oxidized to NAD^+ again. The oxygen concentration at which fermentation starts is commodity, cultivar, maturity and temperature dependent (Wills *et al.*, 1998). An alternative respiratory pathway proposed to play a protective role against oxidative stress is the cyanide-resistant alternative oxidase (AOX) pathway as opposed to the conventional with cytochrome oxidase pathway. AOX plays a role in lowering mitochondrial ROS formation.

The oxidative pentose phosphate pathway is closely linked to glycolysis even though it is usually presented separated from glycolysis. Both pathways share common intermediates (glyceraldehyde-3-phosphate, fructose-6-phosphate and glucose-6-phosphate). The main function of the pathway is to generate reduced NADPH. Several intermediates are used for other important pathways like ribose for nucleic acid biosynthesis, erythrose-4-phosphate for aromatic acid biosynthesis and polyphenols and lignins derived from these aromatic amino acids (Dennis *et al.*, 1998).

2.3.1 Controlled atmosphere storage

The main aim of controlled atmosphere (CA) storage is to prolong shelf-life of the different commodities and reduce product losses. In addition, CA storage plays a key role to avoid the times of glut and to extend the periods of crop availability during off season production, facilitating overseas crop transport. The main principles used consist of: temperature reduction, oxygen reduction and carbon dioxide increase. The combination of high carbon dioxide and low oxygen concentrations reduces the respiration rate of the commodity (Kader, 1986; Ke *et al.*, 1993), reduces ethylene production (Gorny & Kader, 1996b; Salveit, 2003), delays ripening and prevents low temperature disorders (Levin *et al.*, 1995; Zhou *et al.*, 2000). A compromise needs to be established in terms of optimal storage conditions. Thus, temperature should be lowered to retard respiration, ethylene production and senescence while at the same time avoiding chilling injury disorders. The same principle is applied to the oxygen and carbon dioxide concentrations. For each commodity, the tolerance to a specific low level of oxygen and/or high carbon dioxide level can be evaluated by the onset of fermentation (Brecht *et al.*, 2003). Pears are susceptible to elevated carbon dioxide concentrations but not to freeze damage (Kadam *et al.*, 1995). CA storage has been succesfully applied to different commodities such as apples (Gorny & Kader, 1996b), peaches and nectarines (Lurie, 1992; Burmeister & Harman, 1998) and sweet cherries (Mattheis *et al.*, 1997). The CA storage conditions applied to optimize quality parameters can be mutually exclusive. For instance, high levels of carbon dioxide can control mold, reduce ethylene effects and reduce chlorophyll loss but promote fermentation. A low oxygen concentration will reduce respiration, ethylene synthesis and action but it can stimulate fermentation with the concomitant appearance of off flavors and possible microbial growth (Salveit, 2003).

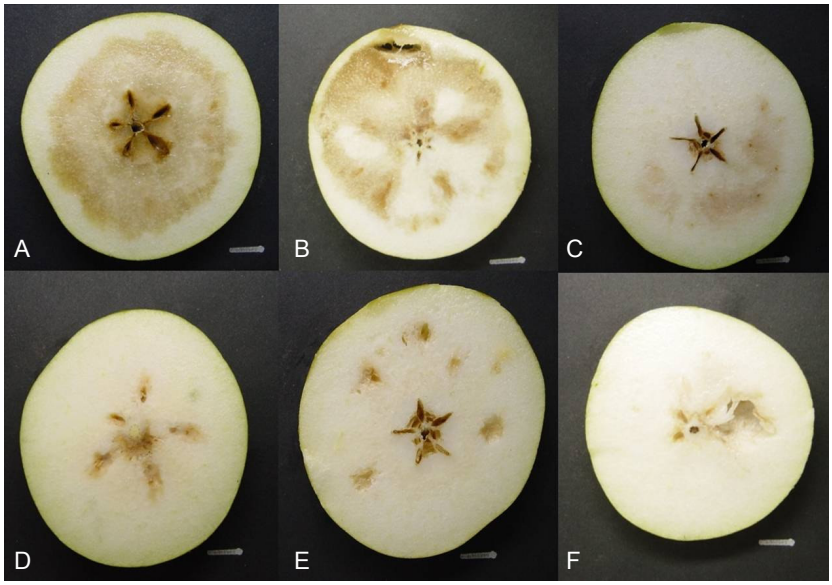

Figure 2.1: *Browning disorders in 'Conference' pears after 4 months in browning -inducing conditions (no cooling period, 1 % O_2, 10 % CO_2, $-1\,^{\circ}C$). The symptoms are divided in four categories. (A) Radial browning, (B) and (C) unequal or asymmetric browning, (D) brown and dry spots in between the extension of the five carpels, and (E) and (F) random cavities. Adapted from Franck* et al. *(2007).*

2.3.2 Postharvest storage disorders: core breakdown

Even though many efforts have been made to optimize CA storage conditions for different commodities, the appearance of postharvest storage disorders cannot be disregarded. Disorders associated with improper CA management include chilling injury, failure to ripen, development of off flavors (Kuo & Parkin, 1989; Mohammed & Wickham, 1997) and browning disorders (Bauchot *et al.*, 1999; Franck *et al.*, 2007; de Castro *et al.*, 2008). The pear variety Conference is highly susceptible to develop core breakdown when improperly stored. Improper storage involves but it is not limited to late harvested fruit, non-delayed controlled atmosphere conditions of 21 days and too low or too high oxygen and carbon dioxide concentrations.

Core breakdown is characterized by internal symptoms invisible from the outside (Figure 2.1). Still, there is a lack of standardization in the nomenclature of browning disorders of pears but basically these disorders can be

divided in three categories: (i) flesh browning, (ii) cavities and (iii) browning and cavities (Giraud *et al.*, 2001). It is well known that several pre-harvest factors favor the incidence of core breakdown. For instance, 'Conference' pears grown in cold areas are more susceptible to browning disorders than pears grown in warm areas (Zerbini *et al.*, 2002) which seems to be related to the amount of sunlight exposure which is positively correlated with the ascorbic acid content (Davey *et al.*, 2000). Reduced browning incidence has been noticed when boron is applied to the soil as the application of boron affects positively the ascorbic acid content of the fruit (Xuan *et al.*, 2001). Fruits from the top of the tree are more susceptible to browning (Franck *et al.*, 2003). Postharvest factors that affect the browning incidence are picking date, duration of the cooling period, the carbon dioxide and oxygen partial pressures, the storage temperature and time of storage (Lammertyn *et al.*, 2000). Late picked fruit is more susceptible to browning than early picked fruit. Delaying CA conditions for 21 days has been proved to be effective in reducing the incidence of core breakdown (Verlinden *et al.*, 2002).

The hypothesis behind the appearance of core breakdown starts from a low energy status that triggers a cascade of events eventually leading to membrane disruption and the enzymatic oxidation of polyphenols. Thus, when the CA conditions applied to Conference pears involve a too low oxygen concentration and a too high carbon dioxide concentration, formation of anoxic zones cannot be excluded. Previous studies have shown that the external CA conditions determine the intracellular oxygen available for maintaining an aerobic or fermentative metabolism (Figure 2.2) and the oxygen and carbon dioxide internal concentrations are related to the pear fruit size and ripening stage as illustrated for oxygen in Figure 2.2 (Ho, 2008). In general, larger fruit is more susceptible to browning as well as late picked fruit. As a consequence of this reduced energy status there is not enough energy produced to maintain cellular reactions such as membrane damage repair. In addition, the stressful conditions (e.g., temperature reduction and extreme oxygen and carbon dioxide concentrations) surpass the capacity of the antioxidant system of the pear, resulting in an overproduction of reactive

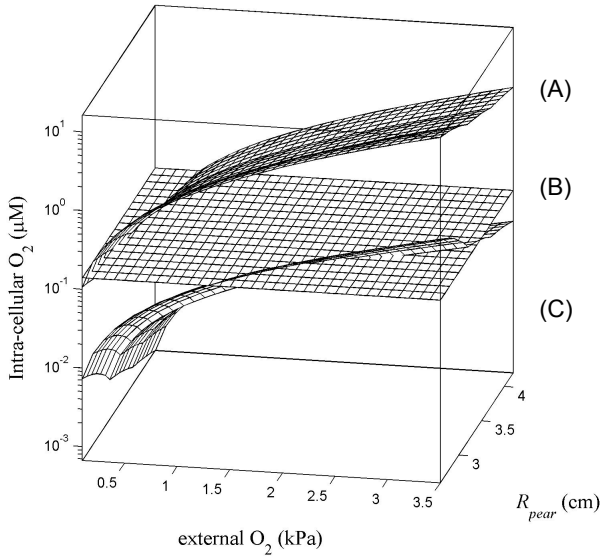

Figure 2.2: *The intra-cellular $O_2(\mu M)$ in the centre of the fruit as a function of O_2 partial pressure and pear radius. The storage temperature was $-1\,^{\circ}$C. (A) optimal harvested fruit (B) O_2 concentration where it is equal to the Michaelis Menten constant for cytochrome c oxidase $K_{m,O2}$ (0.14 μM) and (C) ripened fruit for 7 days in air at ambient temperature (20 $^{\circ}$C). Adapted from Ho (2008).*

oxygen species. As a result, lipid peroxidation starts, membrane disruption takes place, polyphenols and the enzyme polyphenol oxidase localized in different cellular compartments come together, and brown colored compounds are formed (Saquet *et al.*, 2003; Franck *et al.*, 2007). In this concept, core breakdown is the consequence of a series of biochemical processes. As previous studies focused on isolated events, there is a need for holistic approaches, such as proteomics, to better understand this disorder.

2.4 Plant functional genomics

The transformation of small and large molecules through the action of enzymes (proteins) is part of all cellular processes. In order to understand how plant cells function, it is necessary to elucidate the different performers involved in these metabolic processes, proteins and metabolites (Fridman & Pichersky, 2005). The different levels of control of cellular processes are the genome, the transcriptome, the metabolome and the proteome. The study of the different levels of cellular processes besides allowing a comprehensive and systematic functional analysis of genomes, have the potential to accelerate the rate of gene function prediction (Holtorf *et al.*, 2002).

The next section is dedicated to proteomics while this section focuses on the other OMICs platforms such as genomics, transcriptomics and metabolomics. Proteomics is only tackled in this section in relation to postharvest physiology.

Genomics is the large scale study of the genome of an organism. The genome holds the basic information which is the same in each cell. This information is independent of time and environmental conditions (Jacobs *et al.*, 2000). From a given DNA sequence, a potential function can be assigned, but this potential is not necessarily converted into an actual metabolic role (Cordwell, 1999). Thus, the information available in the DNA is not enough to predict if genes will be expressed and when they will do so, in what amounts the products will be present and how these products might be activated (Jacobs *et al.*, 2000).

Proteomics is the study of the whole set of proteins encoded by a genome. Proteomic studies related to postharvest physiology have focused on understanding the fruit ripening process at a proteomic level in tomato (Faurobert *et al.*, 2007) and grape berry (Giribaldi *et al.*, 2007). Such studies allowed the identification of potential markers for specific horticultural quality aspects (Lee *et al.*, 2006), for detecting optimum harvest maturity (Abdi *et al.*, 2002) and for detection of variations among genotypes (Rocco *et al.*, 2006). For a more extended review on the application of proteomics tools in postharvest

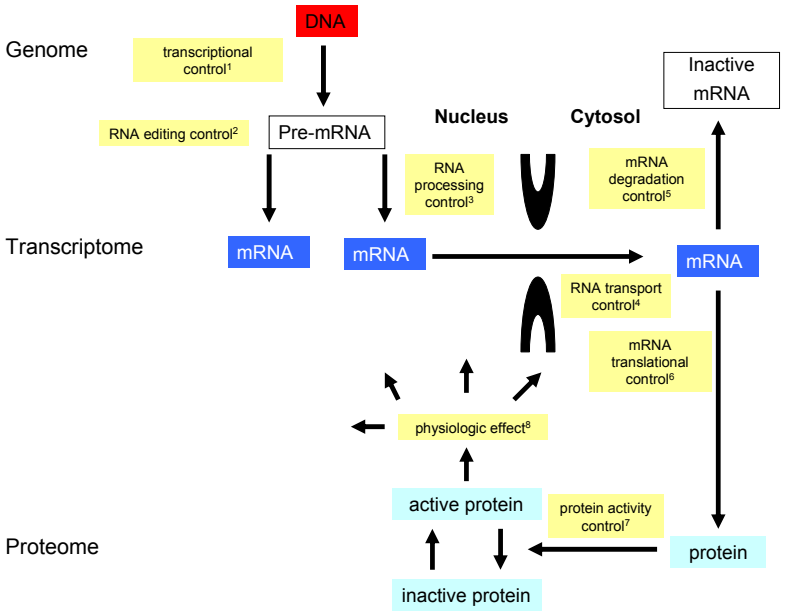

Figure 2.3: *Schematic representation of the transfer of information from the sequence in the genes to the functioning proteins of the cell with possible control mechanisms indicated. A gene (DNA) is transcribed (1) to pre-mRNA that might be edited (2) and then processed (3) to one mRNA or by alternative splicing to several forms of mRNA. mRNAs are transported (4) out of the nucleus to the cytosol where they might be degraded (5) or translated (6) into protein. Protein activity is controlled (7). Proteins might be synthesized as inactive forms or as active forms that later can be inactivated. Proteins are the operating molecules producing the physiological effect in a cell (8). Adapted from Honore* et al. *(2004).*

physiology, the reader is referred to Pedreschi *et al.* (2009b).

Transcriptomics is the study of the whole set of messenger RNA molecules or transcripts produced in a cell and provides a comprehensive view of all active genes at a certain time and condition. mRNA based approaches are high throughput and highly automated for screening thousands of genes in a massively paralleled manner. However, a transcript is only an intermediate, and there will be discrepancy between the transcript and the final protein encoded (Carpentier *et al.*, 2008b). From the presence of a certain mRNA, it is misleading to directly deduce how much protein will be present in the cell and if it will be active or not (Jacobs *et al.*, 2000) given

the poor correlation between mRNA and protein (Figure 2.3). Microarray analysis, cDNA fragment fingerprinting or serial analysis of gene expression (SAGE) are different methods used in transcriptomics (Kussmann *et al.*, 2006; Carpentier *et al.*, 2008a). In postharvest physiology, transcriptomic studies have been carried out in relation to the characterization of the role of ethylene during ripening in peaches (Tonutti *et al.*, 2008), to study biotic and abiotic stresses in citrus fruit (Gonzalez-Candelas *et al.*, 2005; Pons *et al.*, 2005), and to improve the quality of fresh cut produce (Granell *et al.*, 2007).

Metabolomics focuses on the study of the whole set of metabolites ($<$ 3000 m/z) present in a cellular system at a particular physiological state or process (Hollywood *et al.*, 2006). Considering that changes in the levels of individual proteins do not necessarily reflect the changes at the level of metabolite concentrations, metabolomics takes the OMICs approach one step further toward the level at which the changes become perceivable to the consumers. In postharvest physiology, metabolomic applications have been applied to discriminate between diseases (Vikram *et al.*, 2004) and to predict metabolic disorders (Pedreschi *et al.*, 2009a).

Information on proteomics is extensively provided below given that it is the platform chosen to study core breakdown. The decision on applying proteomics tools to study core breakdown was based on the fact that proteins go one step further by providing information on gene products, where, when and under which specific conditions proteins are being expressed and regulated. In addition, proteomics is the only platform that besides delivering biological markers also delivers targets of intervention. According to Kussmann *et al.* (2006), the only way to intervene in certain biological condition and to modulate its outcome is by interfering with the proteins involved. In addition, previous studies in our group focused on metabolic profiling and core breakdown disorder. Thus, the results from both approaches can effectively be combined for further validation of the results. Ideally, the data from the above mentioned platforms (transcriptomics, proteomics and metabolomics) should be integrated into a single systems biology approach, in order to understand the complexity of the disorder and to be able to build

a metabolic network.

2.5 Proteomics: state of the art

Proteomics is the study of the whole set of proteins encoded by a genome. It addresses three biological aspects: protein expression, protein structure and protein function (Kussmann *et al.*, 2006). The study of the whole set of proteins encoded by a genome at a certain time and under certain conditions encompasses different steps that rely on various technologies. A typical proteomics workflow consists of: (i) protein extraction (ii) protein or peptide separation and quantification, (iii) protein identification and (iv) data analysis and integration.

2.5.1 Protein extraction

Protein extraction can be difficult given that many foods are derived from plant organs largely composed of fibrous cell wall material. In addition, the watery content of the plant vacuole results in very low protein yields compared to bacteria or animal tissues. Several methods have been reported for protein extraction from plant materials that are able to cope with the presence of interfering compounds such as phenolic compounds, carbohydrates, proteolytic and oxidative enzymes, pigments, etc. A detailed list of protein extraction protocols for a series of species (banana, pear, apple, potato, maize, etc) has been given by Carpentier *et al.* (2008b). The extraction buffers used for plant tissues are usually characterized by a high pH (8.0 - 8.5) to inactivate proteases. The use of PVPP has been suggested due to its properties as an efficient proton acceptor able to interfere with the binding of proteins to polyphenols (Pierpoint, 2004; Laborde *et al.*, 2006). The use of DTT reduces polyphenols to form thioethers (Loomis & Battaile, 1966). For bacteria and animal tissues, which have higher protein yields, various protein solubilization buffers, including the use of chaotrophic agents, detergents, reducing agents, buffers and ampholites are used (Morzel *et al.*,

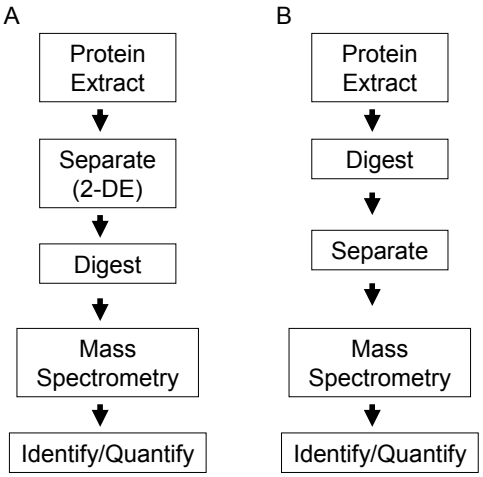

Figure 2.4: *Protein separation approaches: (A) gel-based vs (B) gel-free. In a gel-based approach, proteins are first separated by means of 2-DE and then digested for further mass spectrometry identification. In a gel free approach, LC-MS/MS is used to analyze protein digests of either unseparated protein mixtures and accurate quantification is possible by labeling the peptides with stable isotopes.*

2004; Fernandez *et al.*, 2008). Due to the complex chemical nature of proteins and their broad dynamic range, each known technique focuses on a particular subset of proteins. To circumvent the problem of dynamic range, prefractionation techniques such as organelle fractionation prior to 2-DE and LC-MS/MS are currently employed. Typically, these prefractionation techniques involve differential density ultracentrifugation (Ho *et al.*, 2006a).

2.5.2 Protein separation

Gel based and gel free approaches are complementary since they will focus on a specific subset of proteins. They differ in the way proteins or peptides are isolated, separated and detected (Figure 2.4). In a gel-based approach, proteins are first separated by means of 2-DE and then digested for further mass spectrometry identification. In a gel free approach, LC-MS/MS is used to analyze protein digests of unseparated protein mixtures and accurate quantification is possible by labelling the peptides with stable isotopes.

2.5.2.1 Gel based approach

The gel based proteomics approach relies on two dimensional electrophoresis (2-DE) for the separation of proteins based on two properties: isoelectric point (pI) and molecular weight (Figure 2.5). This technique was introduced in the 1970's by O'Farrell *et al.* (1977). Briefly, isoelectric focusing (IEF) separates proteins by their differences in electric charge taking advantage of the amphoteric character of proteins (the fact that proteins change charge depending on the pH of the environment). To accomplish the separation, an electric current is applied to an IPG strip (acrylamide gel matrix copolymerized with a pH gradient; commercially available). When a protein is in a pH region below its pI, it will be positively charged and it will migrate towards the cathode. When proteins reach their isoelectric point (pH at which the protein has no net charge), they stop migrating and are said to be 'focused'. For more information on IPG technology and IEF, the reader is referred to Righetti & Bossi (1997) and Görg *et al.* (1999, 2000). After completion of the IEF, the IPG strip with the separated proteins is used as starting point for the second dimension separation using SDS-PAGE. SDS polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins based on molecular weight. Before this second dimension can be carried out, proteins on the IPG strip need to be equilibrated in an excess of sodium dodecyl sulphate (SDS) to eliminate the intrinsic charges of the proteins. Due to this, the secondary and tertiary structure of proteins are eliminated and after reduction of the disulfide bridges between cysteines, the electrophoretic mobility of proteins is solely dependent on molecular weight. The acrylamide separating gel is composed of different particle sizes; thus, smaller molecules will move faster compared to large molecules which will be slowed down (Figure 2.5).

Once proteins are separated, visualization for further quantitative analysis proceeds. The most commonly used visualization strategies for quantitative analysis of gel separated proteins use colloidal coomassie blue (Neuhoff *et al.*, 1988), silver staining (Blum *et al.*, 1987), radiolabeling (Patton, 2002) and fluorescent staining (Chevalier *et al.*, 2004). For detailed information about the different protein stains the reader is referred to Miller *et al.* (2006).

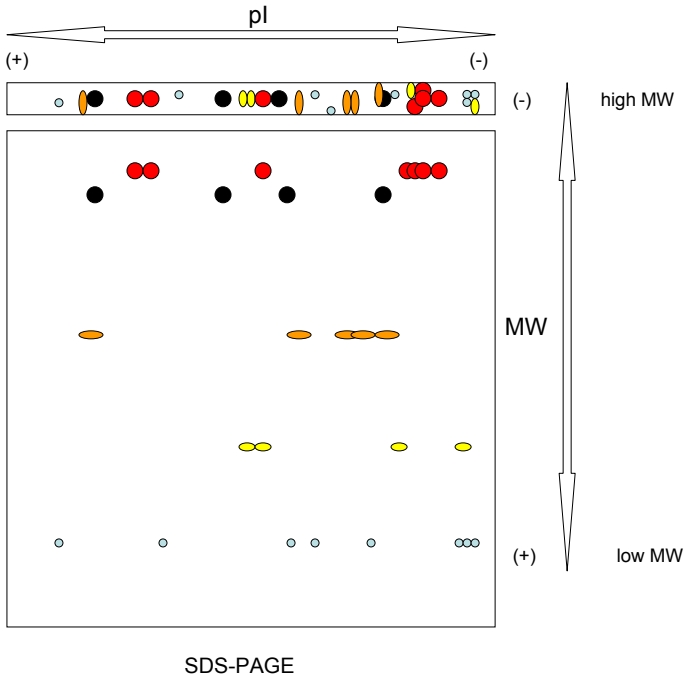

Figure 2.5: *Schematic representation of a two dimensional electrophoresis (2-DE) map. Proteins are first separated based on isoelectric point, pI (A) and then (B) based on molecular weight, MW. The resulting two dimensional map reveals proteins separated on the horizontal axis based on pI and on the vertical axis based on MW.*

The stains described above present certain limitations in terms of detection limit (e.g., colloidal coomassie blue), dynamic range and reproducibility (e.g., silver). One of the limitations of comparative 2-DE is the high gel to gel variation which renders the analysis difficult in terms of distinguishing biological variation from experimental variation. To overcome this issue, two dimensional in gel electrophoresis (DIGE) technology was developed (Alban *et al.*, 2003). In DIGE samples are labeled prior to the electrophoretic separation with spectrally resolvable dyes (Cy2, Cy3 and Cy5). Subsequently, samples are mixed prior to IEF and resolved on the same 2-DE gel (Ünlu *et al.*, 1997). DIGE increases the confidence in terms of detection and quantification of differences in protein abundance and also reduces the number of gels in an experiment. The primary advantage of multiplexing samples is that an internal standard (representing an average of all samples in an experiment) can be included to normalize protein abundance across multiple gels. Thus, each gel will contain an image with a highly similar spot pattern, improving the confidence of inter-gel spot matching and quantification (Figure 2.6). Basic or hydrophobic proteins are still difficult to separate under gel based 2-DE approaches in spite of the introduction of IPG strips up to pH 14 (Görg *et al.*, 1997). Substitution of the reducing agent DTT for TBP (tributylphosphine) or HED (hydroxyethyldisulphide, commercially known as Destreak^{TM}) (Hoving *et al.*, 2002; Olsson *et al.*, 2002) partially circumvents the problem of lack of resolution of basic proteins during the IEF run.

Gel based proteomics is the most powerful option for non-model organisms (e.g., pears; Carpentier *et al.* 2008b), where isoforms and post translational modifications can be studied. But some of the limitations of this approach are the unequal resolving power of 2-DE including the bias towards high abundant proteins, hydrophobic or very acidic proteins are not resolved, co-migration of proteins resulting in spots containing multiple proteins and the limited dynamic range covered besides the difficulty for automation.

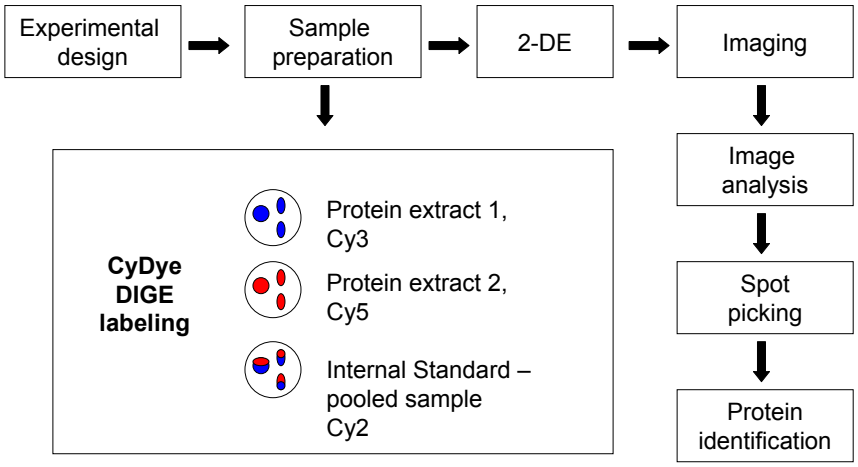

Figure 2.6: *Differential in gel electrophoresis - DIGE workflow. Prior to 2-DE separation, proteins are labeled with spectrally resolvable dyes (Cy2, Cy3 and Cy5). Then, the samples are mixed prior to IEF and resolved in the same 2-DE gel.*

2.5.2.2 Gel free approach

Gel free approaches in most of the cases use a bottom up strategy meaning that proteins are first proteolyzed and the obtained peptide mixture is then separated based on hydrophobicity via reverse-phase chromatography (Figure 2.4). Subsequently, the eluted peptides are sent to a mass spectrometer. All the tandem mass spectra gathered are then used to search databases and reconstruct the original proteins (Roe & Griffin, 2006). This approach is successful with simple protein mixtures and sequenced species. The problem of resolution was circumvented by the introduction of MudPit (multidimensional protein identification technology; Washburn *et al.* 2001). By the introduction of MudPit separation of membrane proteins as well as an increase in the ability to detect low abundant proteins was achieved (Roe & Griffin, 2006). The limitation of this approach is the lack of provision of quantitative information. This has been overcome by the use of stable isotope labeling and dilution strategies for the relative quantification of pro-

teins (Gygi *et al.*, 1999; Ross *et al.*, 2004; Roe & Griffin, 2006). Gel free approaches have the disadvantage that qualitative and quantitative information on protein isoforms and differential post-translational modifications are lost (Carpentier *et al.*, 2008b). Besides that, cross-species identification for poorly sequenced genomes, as for most crops, is not possible, as it relies on comparing peptides of the proteins of interest to orthologous proteins from other well characterized species.

2.5.3 Protein identification

Mass spectrometry is the technology most widely used for protein identification. Both with gel based and gel free approaches, proteins need to be digested before being introduced in a mass spectrometer. The unknown protein of interest is generally cleaved into smaller peptides by using, in most of the cases, a trypsin enzyme. This enzyme will specifically cleave proteins on the carboxy-terminal side of arginine and lysine residues (Steen & Mann, 2004). A mass spectrometer consists of an ion source (e.g., ESI or MALDI) to produce ions from the sample, one or more mass analyzers (e.g., quadrupole, TOF, ion trap, linear ion trap) to separate the ions based on their m/z ratios, a detector to register the number of ions coming from the last analyzer, and a computer to process the data and produce the mass spectra. In addition an inlet device is necessary to introduce the sample into the ion source (Lane, 2005).

2.5.3.1 Ionization techniques

The introduction of two soft ionization techniques for mass spectrometry for the analysis of proteins, matrix assisted laser desorption ionization (MALDI; Karas & Hillenkamp 1988) and electrospray ionization (ESI; Fenn *et al.* 1989) has revolutionized the proteomics platform enabling the high throughput identification of proteins. By using these soft ionization techniques, there is minimal fragmentation and mostly entire ions are formed. A schematic representation of the MALDI and ESI ionization processes is

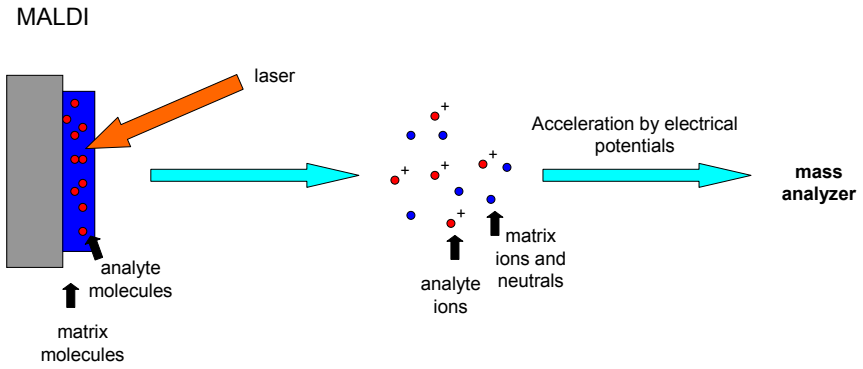

Figure 2.7: *Matrix assisted laser desorption ionization - MALDI. In MALDI, the analyte is mixed with a large excess of matrix. Both the analyte and the matrix are irradiated with a laser beam enabling the excess matrix molecules to sublime and transfer the embedded non-volatile analyte molecules into the gas phase. Mostly single protonated ions are formed and accelerated by electric potentials into a mass analyzer (Steen & Mann, 2004).*

given in Figure 2.7 and 2.8.

MALDI, relies on a laser beam which is fired at a sample plate containing a dried mixture of matrix (e.g., α-cyano-4-hydroxycinnamic acid) and sample to ionize the latter. The matrix will absorb the radiation from the laser resulting in excitation of the matrix molecules. As a result, a dense plume containing both matrix and analyte molecules is produced and analyte molecules interact with hydrogen atoms from the matrix to form mainly singly charged ions (Andersen & Roepstorff, 1996; Steen & Mann, 2004) entering the mass analyzer. During the ESI process, the sample is dissolved in a solvent mixture (e.g., acetonitrile-water) and then injected into a capillary held at a potential of 3-4 kV. As a result, a very fine spray of solvent droplets containing pre-formed ions of the forms $(M + nH)^{n+}$ (where M: the peptide molecule, nH: number of protons attached to the molecule and n^+: net charge of the ions) is formed. The multiply charged gas-phase ions

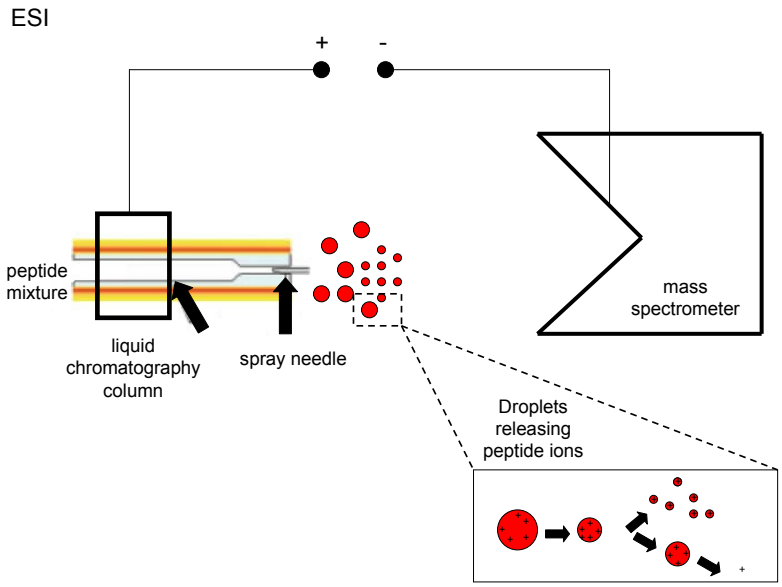

Figure 2.8: *Electrospray ionization - ESI. The sample is dissolved in a solvent mixture (e.g., acetonitrile-water) and then injected into a capillary held at a high electrical potential. A very fine spray of solvent droplets containing the pre-formed ions is obtained. The multiply charged gas-phase ions which are formed as a consequence of the desorption process due to evaporation of the solvent will subsequently enter the mass analyzer. Adapted from Steen & Mann (2004).*

are then formed as a consequence of the desorption process occurring due to evaporation of the solvent (Andersen & Roepstorff, 1996; Steen & Mann, 2004) and will subsequently enter the mass analyzer.

2.5.3.2 Mass analyzers

The formed ions are then separated according to their mass to charge ratio *(m/z)* in a mass analyzer. There are different mass analyzers, each one, with its strengths and weaknesses. A time of flight (TOF) analyzer uses an electric field to accelerate the ions at the same potential. Then, the time needed to reach the detector is measured. For particles with the same charge, their kinetic energies is the same, thus their velocities is solely dependent on their masses; lighter ions will reach the detector first (Wollnick, 1993).

Quadrupole (Q) analyzers use oscillating electrical fields to selectively stabilize or destabilize ions passing through a radio frequency quadrupole field (Lane, 2005). Ion traps or more specifically quadrupole ion traps, trap ions in a dynamic electric field and then sequentially eject them into the detector according to their m/z values (Steen & Mann, 2004). The low accuracy of the ion trap is its main disadvantage unless coupled to a high accuracy detection system such as found in Fourier transform and Orbitrap devices; however, it is robust, sensitive and inexpensive (Lane, 2005). A 'linear ion trap' has increased sensitivity, resolution and mass accuracy. It differs from a quadrupole ion trap in that it uses a two-dimensional quadrupole field instead of three.

2.5.3.3 MS vs tandem mass spectrometry (MS/MS)

Peptide mass fingerfrint (PMF) is typically performed by (MALDI-TOF) MS because a simple profile is produced providing one single peak per peptide (Carpentier *et al.*, 2008b). For PMF, individual proteins are first digested into smaller peptides, then the masses of these peptides are measured with a mass spectrometer. Independently, in the database search, each protein sequence is *in silico* digested. Then, the masses of the generated *in silico* peptides of each individual database entry (theoretical) are compared with the measured mass spectrum. To get a significant hit, only a subset of all peptides from the protein digest need to match. One of the pitfalls of PMF is the reduced power to identify proteins from non-model organisms (with a poorly characterized genome; Mathesius *et al.* 2002). Tandem mass spectrometry couples two stages of MS and sequence information about the proteins is obtained. In MS/MS, a particular peptide is isolated, energy is imparted by collision with an inert gas (e.g., nitrogen molecules or argon or helium atoms) and as a consequence this energy causes the peptide to fragment typically down the peptide bond. Thus, a mass spectrum of the resulting fragments is generated (Steen & Mann, 2004). The use of LC-ESI-MS/MS (Wilm & Mann, 1994) is well established as well as MALDI-Q-TOF (Shevchenko *et al.*, 2001), MALDI-TOF/TOF (Yergey *et al.*, 2002), or

MALDI- quadrupole ion trap/TOF instruments (Martin & Brancia, 2003). Reconstruction of an unknown peptide from MS/MS data through peptide sequencing is possible (referred as *de novo* peptide sequencing) but still some challenges exist. Proteins generally are made up of 20 different types of amino acids which most of them have different masses. Thus, different peptides will produce different spectra being possible to use the spectrum of a peptide to determine its sequence. In most of the cases, this is carried out by searching un-interpreted data using various algorithms such as MASCOT, SEQUEST, !XTandem and PHENYX. For non-model organisms, this approach offers great possibilities (Steen & Mann, 2004; Carpentier *et al.*, 2008b).

2.5.4 Data analysis

2.5.4.1 The importance of a good experimental design

Given the importance of a good experimental design and sampling to draw valid conclusions from a study, this section is dedicated to these topics. Sampling techniques as well as experimental design should be unbiased. A good experimental design will limit systematic errors, improve precision of subsequent statistical tests and reduce the number of false positive calls (Chich *et al.*, 2007).

Sources of variation can be classified into two classes: (i) technical variation and (ii) biological variation. Thus, technical replicates are repeated measures from the same biological sample while biological replicates are different replicate samples from the same treatment group (Karp *et al.*, 2005a; Hollywood *et al.*, 2006). The type of replicate will limit the statistical test to be used and the conclusions that can be drawn. To anticipate all sources of unknown variation, randomization is necessary in order to reduce systematic errors when treatments are compared and the precision of the results is estimated (Chich *et al.*, 2007). A good experimental design must follow and address certain key points: (i) state a null hypothesis (Ho: there is no difference between the treatments) and the alternative hypothesis (Ha:

there is a difference between the treatments), (ii) election of the most appropriate statistical test to test the hypothesis, (iii) specify a significance level and address the multiple testing problem, (iv) determine the sample size to have a sufficient power and (v) collect the data. A sample size that controls the risk of false positives is imperative. The power of the test is the probability that a true call will be detected as such and can be termed 'sensitivity' (Horgan, 2007). The power is dependent on the extent of the difference we are looking for, the random variation and the sample size per group.

2.5.4.2 Data preprocessing steps

Before data from 2-DE experiments can be statistically analyzed, certain preprocessing steps need to be carried out. Normalization of the spot volumes to remove effects of differential loading and staining is the first step (Meleth *et al.*, 2005). Many statistical methods are based on the normal distribution assumption. To fulfill this assumption, it is sometimes indispensable to apply some type of data transformation (e.g., logarithmic, sinh). The applied transformation will have a positive effect in stabilizing the variance (e.g., highly abundant proteins usually have larger variance than low abundant proteins) so that also the second assumption of parametric statistical methods 'homocedasticity' is fulfilled (Chang *et al.*, 2004; Jung *et al.*, 2005; Urfer *et al.*, 2006; Jung *et al.*, 2006). The presence of missing data from 2-DE experiments cannot be disregarded and must be handled with extreme care (Krogh *et al.*, 2007; Pedreschi *et al.*, 2008b).

2.5.4.3 Univariate data analysis

The most typical case in 2-DE studies is the comparison of a treatment group (e.g., diseased condition) *vs* a control group (e.g., healthy condition). There can be thousands of proteins to be compared. Thus, univariate statistical methods will compare protein by protein looking for significant changes in expression between the treatment and control group. The statistical ap-

proach includes a significance test for each protein to determine if the protein is differentially expressed or not. A significance level α is chosen and if the p value of the test is higher than the α value, then the null hypothesis is accepted.

The use of parametric univariate tests such as the t-test (to compare two treatments) and Analysis of Variance - ANOVA (to compare k treatments) and post-hoc tests (e.g., Duncan test, Tukey test) are powerful but require the above mentioned assumptions to be fulfilled. Given the number of proteins to be tested, the multiple testing issue should be tackled otherwise it can lead to a large number of false positive calls. Thus, it is important to control the number of false positive calls but still be able to detect true differences (power). In order to determine an appropiate sample size, the power of the statistical test needs to be specified on before hand. The power is also dependent on the variance of the data, on the significance level α, on the magnitude of the expression change to detect and on the type of test procedure (Urfer *et al.*, 2006). The false discovery rate and q-value methodology have been applied together with a proper experimental design and correct use of the statistical test to account for the number of false positive calls (Karp *et al.*, 2007). The q value is an adjusted p value for each test. It controls the number of false positives in those tests that were significant. It is less conservative than the Bonferroni approach and has greater ability to find truly significant results.

Instead of parametric tests, non-parametric tests do not assume any distribution of the data but have less power (Siegel, 1988). Still, they must fulfill certain assumptions such as independent sampling and continous ordinal data. The Kolmogorov-Smirnov test is the equivalent to the t-test (to compare two treatments or groups) and the Kruskal Wallis test is the equivalent to ANOVA (to compare k treatments). It is important to emphasize that mixing replicates can lead to an increased number of false positive calls if the statistical test chosen for analysis does not account for differentiation in the type of variability. When there is a mix of technical and biological replicates, a nested ANOVA is an alternative test. A nested ANOVA assesses whether the variance due to treatment is greater than the variance

among biological replicates within a treatment (Karp *et al.*, 2007).

For time course data in which not only the effect of treatment is assessed but also the effect of time and the time/treatment interactions, a two-way analysis of variance can be utilized (Jung *et al.*, 2005, 2006). The same approach can be utilized when more factors are to be studied (e.g., cultivar, storage time, disease type).

2.5.4.4 Multivariate data analysis

The 2-DE data is complex and high dimensional. Thus, multivariate statistical tools are useful in extracting as much information as possible in terms of similarities and correlations. These multivariate statistical techniques will consider the variables (proteins) as a group of correlated variables rather than focusing on one variable or protein at a time as in univariate statistics (Karp *et al.*, 2005b). When proteins are part of the same pathway or a complex interrelation of pathways, correlation among proteins will exist which can be extracted using multivariate techniques such as PCA and PLS.

Principal component analysis (PCA) is an unsupervised technique that forms new variables (principal components) that are linear combinations of the original variables thus capturing the essential data patterns of the original data in a reduced form. PCA is useful to examine datasets with multiple collinearity (e.g., proteins that act in concert with other proteins) and to get insight into certain patterns or trends (Wold *et al.*, 1987; Karp *et al.*, 2005b). The score plots obtained show the distribution of the objects (gels) and their distribution allowing the identification of outliers through the Hotelling T^2 ellipse. The loading plots obtained show the distribution between the different variables and their distribution. The further the variable from the origin, the more influential is the variable for explaining relationships in the dataset. The distances along the first components are more important because the first principal components explain more of the variation in the dataset. The superimposition of both score and loading plots result in a biplot in which directly the discrimination between the treatments and rel-

evant proteins involved in such discrimination can be assessed.

Partial least squares discriminant analysis (PLS-DA) is a supervised bilinear regression model to create prediction models of one or several responses from a set of factors (Wold *et al.*, 1987). PLS-DA will construct latent variables (LV) in such a way that a maximum separation is obtained among them. PLS-DA can be useful in addition to PCA to correlate variation in a dataset with class membership (Karp *et al.*, 2005b) and to select important variables involved in class distinction. As in PCA, score and loading plots are obtained and can be interpreted in the same way as in PCA (Figure 2.9). In addition, plots for variable importance (VIP), model coefficients, residuals, distances to model plots and validation plots are obtained (Danvind, 2002).

The VIP procedure identifies those proteins that are important for explaining the variance in the model response (Karp *et al.*, 2005b). The VIP coefficient of a protein is calculated as a weighed sum of the squared correlations between the PLS-DA components and the original variable. The weights correspond to the percentage variation explained by the PLS-DA component in the model. The number of terms in the sum depends on the number of PLS-DA components found to be significant in distinguishing the classes. Care must be taken when excluding variables from the model. If many important variables are excluded, important explanatory information may be lost as well (Danvind, 2002). For a detailed description about PLS and VIP the reader is referred to Norden *et al.* (2005).

Proteomics generates huge amounts of data and demands the integration of different research fields (e.g., biochemistry, bioinformatics, cell biology, etc). The following chapters, address different challenges encountered when analyzing gel based proteomics data as to draw sound biological conclusions to study core breakdown disorder.

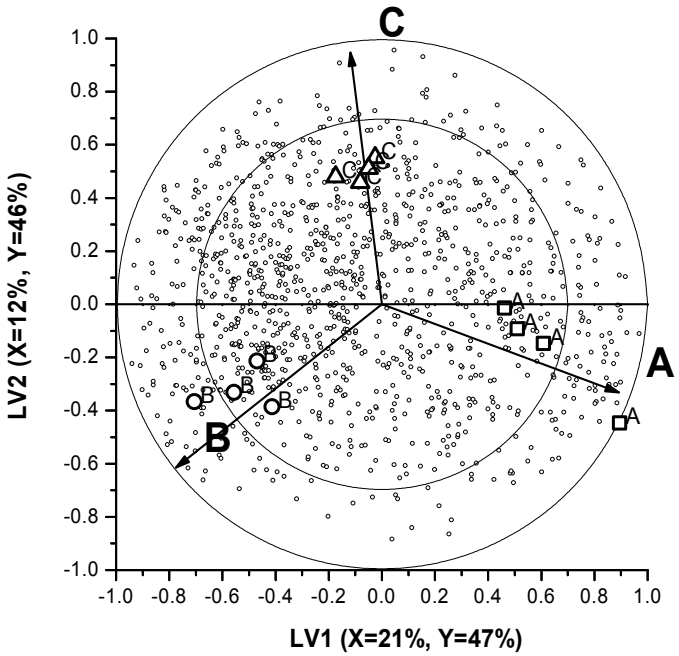

Figure 2.9: *Partial least squares discriminat analysis (PLS-DA) biplot of samples representing three different treatments (A, B and C). Clear discrimination among the treatments (Y-variables) can be observed. Sample scores (represented by circles, triangles and squares), X-loadings (represented by small black dots) and Y-loadings (represented by arrows) are superimposed. The percentage explained X and Y variation is presented on the axis. The analysis is based on the correlation matrix. LV stands for latent variable. The scores, representing each one a sample, already indicate, in this particular case, a very good separation among the treatments. Having determined that there is a clear pattern in the data, then the proteins that are important for separating the various treatments are characterized by having large loadings which describe the weighing coefficients for each protein. The further the protein (small black dots) from the origin, the more influential that particular protein is in the separation of the treatments. The closer the X-loadings to a specific Y-loading (arrow end point) indicates a high positive correlation to that specific treatment. The total explained X-variance with two LV components equals to 33 % and the total Y-variance with two LV components equals to 93 %.*

Chapter 3

Proteomic analysis of core breakdown

3.1 Introduction

To extend their storage life, apples and pears are generally stored under controlled atmosphere conditions. In such conditions, the oxygen partial pressure is reduced and the carbon dioxide partial pressure is increased in order to retard the respiratory metabolism and quality degradation reactions. However, if the oxygen partial pressure is too low or the carbon dioxide partial pressure too high, the metabolism may change from aerobic to anaerobic and this may cause fermentation related physiological disorders like core browning and accumulation of fermentation volatiles. Pears (*Pyrus communis* L.) of the cultivar 'Conference' are particularly susceptible to a physiological disorder called 'core breakdown' of which the symptoms are flesh browning and the formation of internal cavities during controlled atmosphere storage (Lammertyn *et al.*, 2000; Franck *et al.*, 2003; Veltman *et al.*, 2003). Additional factors that favor the incidence of core breakdown are pre-harvest factors such as harvest date (late harvested fruit is more susceptible to browning), orchard characteristics (tree and soil characteristics including but not limited to application of agro-chemicals, irrigation and

geographical position), seasonal variation which include fruit set, position of the fruit on the tree (fruit from the top of the tree is more susceptible to browning), weather conditions (pears grown in cold growing areas are more susceptible to browning), fruit size (large fruit is more susceptible to browning) and ascorbic acid content (Franck *et al.*, 2007). Since 'Conference' is commercially the most important cultivar in Europe, core breakdown may cause considerable economic losses when sub-optimally stored. As mentioned in Chapter 2, browning seems to start from the center of the fruit and it is highly correlated to the limited oxygen content available to carry out aerobic metabolism.

The biochemical mechanisms behind this physiological disorder are not fully understood. Flesh browning is the result of the enzymatic oxidation of diphenols by polyphenol oxidases, yielding quinones which may be further polymerized into brown colored substances which are called melanines. This reaction can only happen after membrane disintegration and cellular decompartmentalization since diphenols are localized in the vacuole and the enzyme polyphenol oxidase is localized in the plastids, chloroplast and cytoplasm. This enzyme has been characterized and compared across *Rosaceae* species including *Pyrus communis* L. (Haruta *et al.*, 1999). Generally, membrane disruption takes place because of an overproduction of reactive oxygen species (ROS) under stress conditions, which cannot be scavenged by the antioxidant system of the pear (Veltman *et al.*, 2000; Larrigaudiere *et al.*, 2001; Franck *et al.*, 2003). Furthermore, the oxygen and carbon dioxide gradients across the pear during controlled atmosphere conditions cause an impaired respiration (Lammertyn *et al.*, 2003b,a) which cannot provide sufficient energy generation for cell maintenance processes such as repair of membrane damage.

Two-dimensional polyacrylamide gel electrophoresis (2-DE), an important tool in proteomics, is widely used because it offers a comprehensive approach to study biochemical systems (Kjaersgard *et al.*, 2006). It is being extensively applied to unravel underlying physiological mechanisms of different diseases for instance by comparing protein maps of healthy and diseased subgroups (Marengo *et al.*, 2006; Casado-Vela *et al.*, 2005, 2006). In the

last years, the interest in plant proteomics focused on studying different aspects in fruits and vegetables has significantly increased (Rocco *et al.*, 2006; Hjern *et al.*, 2006). The classical way of finding differences in expression levels of proteins between groups of samples/treatments is by a univariate statistical approach like a Student's t-test or a Kolmogorov-Smirnov test (Salekdeh *et al.*, 2002; Shen *et al.*, 2003). However, the one by one comparison of individual protein spot volumes has its limitations when extraction of meaningful information from complex data is the objective (Jessen *et al.*, 2002). Multivariate statistical tools have the advantage of reducing the number of false positives and moreover allow for the identification of spots that are significantly altered in terms of correlated expression and not only in terms of absolute expression values (Marengo *et al.*, 2006). The use of multivariate statistical tools like principal component analysis (PCA), partial least squares regression (PLSR) and discriminant analysis (PLS-DA) to analyze proteomics data, hence, is strongly recommended (Marengo *et al.*, 2004; Karp *et al.*, 2005b; Kjaersgard *et al.*, 2006). Since, the pear proteome is poorly characterized, gel-based proteomics was employed as platform instead of a gel-free approach.

The objective of this chapter is to understand core breakdown by means of a holistic proteomics approach combined with univariate and multivariate statistics. To accomplish this objective, five main points will be covered: (i) to obtain suitable protein extraction and two-dimensional gel electrophoresis protocols for pear parenchyma tissue, (ii) to identify up-regulated proteins in brown tissue (iii) to assess for differentially expressed proteins between healthy and disordered pears using univariate and multivariate statistical tools, (iv) to identify the main differentially expressed proteins and (v) to interpret the results in terms of changes in the pear metabolism due to oxidative stress. The results of this chapter were published in Pedreschi *et al.* (2007)[2].

[2]Pedreschi et al. (2007). Proteomic analysis of core breakdown disorder in Conference pears (*Pyrus communis* L.). Proteomics, 7, 2083-2099.

3.2 Materials and methods

3.2.1 Plant material

Pears (*Pyrus communis* cv. Conference) were harvested in the orchard of the Centre for Fruit Culture in Rillaar (Belgium). All homogeneous size pears to be stored under optimal controlled atmosphere (CA) conditions were picked at the commercial harvest date on 8/9/2004 as determined by the Flanders Centre of Postharvest Technology (Belgium). These fruits were submitted to pre-cooling in air at $-1\,^{\circ}C$ for three weeks before applying controlled atmosphere conditions of 2.5 % O_2 and 0.7 % CO_2 according to commercial protocols. To induce core breakdown, pears were picked on 22/9/2004, two weeks after the commercial harvest date. Pears were immediately stored under 1.0 % O_2 and 10 % CO_2 with no pre-cooling period at $-1\,^{\circ}C$. After six months of CA storage, pears from both conditions were sampled. Pears were cut perpendicularly to the stem-calyx axis at 5 cm from the bottom of a pear. Tissue samples were taken from the equatorial region excluding the skin and core, immediately frozen in liquid nitrogen and kept at $-80\,^{\circ}C$ until further analysis. Tissue from optimally stored pears is referred to as healthy tissue and is composed of tissue coming from the inner and outer cortex to compensate for cell variation due to spatial position. Tissue from disordered or sub-optimally stored pears was divided in brown tissue (coming from the brown or affected area) and sound tissue (coming from the apparently healthy or not affected area; Figure 3.1).

Two biological replicates of tissue samples were prepared based on pooled tissue coming from 6 pears each. Each of the two pooled tissue samples was used to prepare two technical replicate protein extracts. In addition, the two tissue samples were pooled to create a third synthetic master sample. From this synthetic sample two technical replicate protein extracts were prepared as well. This resulted in a total of 6 replicate samples per tissue type that were used for the subsequent 2-DE gel analyses.

Figure 3.1: *Cross section of pears without (left) and with (right) core breakdown. (H) healthy tissue, (S) sound tissue and (B) brown tissue of sub-optimally stored pears.*

3.2.2 Protein extraction

Three different variants of the extensively applied phenol extraction, methanol ammonium acetate precipitation protocol for plant tissues were used (Saravanan & Rose, 2004; Carpentier *et al.*, 2005). Two hundred milligrams of powdered frozen pear tissue were homogenized in 500 μL of cold protein extraction buffer (PEB) for 30 min at 4 °C. The PEB consisted of 0.7 M sucrose, 100 mM KCl, 1 mM PMSF, 500 mM EDTA, Tris-HCl, DTT, and PVPP (insoluble polyvinylpolypyrrolidone) whose concentrations were varied as follows. Variant I contained 0.385 % DTT, 50 mM Tris-HCl (pH 7.2) and 1 % PVPP resulting in a final pH of 7.0. Variant II contained 50 mM Tris-HCl (pH 8.5) with a final pH of 8.0 and 1 % DTT and finally variant III contained 50 mM Tris-HCl (pH 7.2) and 1 % DTT with a final pH of the PEB of 7.0. Five hundred microliters of ice-cold Tris buffered phenol (pH 8.0) was added and the sample was vortexed (MS2, IKA Works Inc, Wilington, NC, USA) thoroughly for 5 min at 4 °C. After centrifugation (Sanyo Hawk 15/05, UK) of the sample at 7000 x g for 10 min at 4 °C, the phenol phase was collected and re-extracted with an equal volume of protein extraction buffer and an extra 50 μL of ice-cold Tris buffered phenol. The water phase containing the cellular debris was re-extracted with an equal volume of ice-cold Tris buffered phenol. Both phases were thoroughly vortexed and shaken for 5 min at 4 °C. After centrifugation at 7000 x g for 10 min at 4 °C, both phenol phases were joined and 900 μL of PEB and an extra 100 μL of ice-cold Tris buffered phenol were added. The mixture was vortexed for 5 min at −1 °C, and after centrifugation at 7000 x g for 10 min at 4 °C, the phenol phase was recovered and precipitated overnight with five volumes of 100 mM ammonium acetate in methanol at −20 °C. After centrifugation at 21 900 x g for 30 min, the supernatant was removed and the protein pellet washed two times with methanol and two times with acetone containing 0.1 % DTT. After washing, the pellet was allowed to dry in air until the remaining acetone was evaporated and stored at −80 °C until used.

3.2.3 Two dimensional electrophoresis - 2DE

Protein pellets were redissolved in rehydration buffer (7 M urea, 2 M thiourea, 4 % CHAPS, 0.2 % IPG buffer, 75 mM DTT and 0.002 % bromophenol blue) for 1 h and quantified by using a modified Bradford dye-binding procedure (Bradford, 1976). Thirty five, 70 and 100 μg protein sample for silver staining at pH 5-8, 3-6 and 7-10, respectively, or 350, 700 and 1000 μg protein sample for Coomassie blue staining at pH 3-5, 3-6 and 7-10, respectively were rehydrated in either 150 or 200 μL (silver or Coomassie blue staining) and applied via anodic cup loading. Strips of 24 cm of pH 3-6, 5-8 and 7-10 (Bio-Rad, Nazareth, Belgium) were rehydrated for at least 8 h in 460 μL rehydration buffer. Proteins were isoelectrically (IEF) focused in an IPGphorII System (Amersham Biosciences, Uppsala, Sweden) at 20 °C and with a 50 μA per strip limit. IEF was carried out in four steps. The first three steps were kept the same in all cases: 3 h at 300 V, 6 h at 1000 V, 3 h at 8000 V. The fourth step was varied depending on the pH range of the strip used and on the amount of protein loaded. Thus, in the case of silver staining, 32 000, 24 000 and 32 000 Vh at 8000 V were used for pH 3-6, 5-8 and 7-10 respectively. For Coomassie staining, 40 000, 32 000 and 40 000 Vh at 8000 V were used for pH 3-6, 5-8 and 7-10, respectively. DeStreak^{TM}(Amersham Biosciences, Uppsala, Sweden) was added when 7-10 pH strips were focused. After IEF completion, strips were equilibrated individually for 15 min in 8 mL equilibration buffer (6 M urea, 30 % glycerol, 2 % SDS, 0.002 % bromophenol blue, 50 mM Tris-HCl pH 8.8) containing 2 % (w/v) DTT, and subsequently for 15 min in 8 mL equilibration buffer containing 2.5 % iodoacetamide. Second dimension separation was performed in an Ettan DALT System (Amersham Biosciences) with lab cast 1.5 mm SDS polyacrylamide gels (12.5 %) for 45 min at 12 W and subsequently for 5 h at 100 W. Five or 6 replicate gels were run for every sample category.

3.2.4 Protein visualization and image analysis

Proteins were visualized by silver staining (Blum *et al.*, 1987) or by colloidal Coomassie Brilliant Blue (CBB) G-250 (Neuhoff *et al.*, 1988). For silver staining, gels were fixed overnight in 40 % ethanol and 10 % acetic acid, then washed for 20 min in 30 % ethanol, 20 min in 20 % ethanol and finally 20 min in Milli-Q water. Following washing, gels were sensitized in 0.02 % $Na_2S_2O_3$ for 1 min and incubated in silver stain (0.2 % $AgNO_3$ and 0.02 % formaldehyde) for 20 min. Gels were developed in 3 % Na_2CO_3 and 0.0005 % $Na_2S_2O_3$ for 3 min and immediately blocked in 11 % glycine. For colloidal CBB staining, gels were fixed for 1 h in a solution containing 1.3 % *o*-phosphoric acid and 20 % methanol and then stained overnight in a solution containing 20 % methanol, 0.1 % w/v CBB G-250, 1.6 % *o*-phosphoric acid and 8 % w/v ammonium sulphate. After staining, gels were neutralized with 0.1 M Tris-base titrated to pH 6.5 with *o*-phosphoric acid for 3 min and then washed in 25 % methanol for 1 min. Stained gels were scanned and calibrated with Labscan 5 software (Amersham Biosciences). Image analysis was performed with the Image Master 2-D platinum software 6.0 (GE). Spots were detected without spot editing and quantified as percentage volume. Silver stained gels were used as analytical gels and Coomassie stained gels as preparative gels for further MS protein identification.

3.2.5 Data analysis

The two-dimensional gel electrophoresis data, calculated as spot volume percentage, were analyzed statistically both with univariate and multivariate techniques to test for significant differences in protein expression profiles between the different tissue types. For both statistical approaches the analysis were repeated using two different reference gels. Switching to another reference is good practice to cross verify the results even though one must be aware that through transferring all the tissue specific spots to a synthetic master gel would have done the trick and would have limited the statistical analysis to one dataset and also reduced the number of spots sent for iden-

tification. First, the protein expression profile of healthy tissue was used as a reference and compared to the expression profiles of sound and brown tissue. Second, the analysis was repeated with the expression profile of sound tissue as a reference. For the characterization of brown tissue up-regulated proteins, brown tissue was used as reference gel and compared to sound and healthy tissue.

For the univariate statistical data analysis, the non-parametric two-sample Kolmogorov-Smirnov test was used, since the assumption of normality underlying a classical ANOVA was not fullfilled. Kolmogorov-Smirnov test allows pairwise comparison of the expression level of individual proteins, but is not suited to compare whole expression profiles as it ignores the correlation structure of the data. Proteins with a p value below 0.05 were considered to be significantly different. The univariate statistical analyses were performed with Image Master 2-D platinum software (Amersham Biosciences, Uppsala, Sweden).

For the multivariate data analysis, gels were matched to the reference gel and percentage spot volumes were calculated. The data were preprocessed by mean-centering, and all variables were weighed by their standard deviation to give them equal (unity) variance. A PCA analysis was carried out for first data exploration and outlier detection, but no significantly outlying gels were detected as defined by 95 % Hotelling's T^2 limit (Johnson & Wichern, 1998). A PLS-DA analysis was performed to cluster the individual gels according to similar protein expression profiles (Norden *et al.*, 2005). It is a partial least squares regression of a set of binary variables, describing the categories of the categorical variables (tissue type), on a set of predictor variables (percentage spot volume of different proteins). The variable importance plot (VIP) was used as a formal tool (Karp *et al.*, 2005b), based on the correlation loadings, to identify the most important proteins describing the difference in protein expression profiles. The most important proteins were identified through an iterative procedure. The proteins were sorted according to VIP coefficient and the PLS-DA was rerun on the 50 most important proteins. This was repeated twice selecting the 30 and 20 most important proteins, respectively. The final cutoff value in the VIP

selection is in a way arbitrary in where to set the limit, and related to the total cost of the proteins that need to be identified. This iterative process allowed selecting the 20 key proteins responsible for the clustering and, hence, for further identification by LC-ESI-MS/MS. PCA and PLS-DA must also hold the underlying normality assumption. Ninety five percent of the spots were normally disctributed. Cross-validation was applied to test the performance of the models since the number of observations is relatively small it is impossible to validate the models on an independent test set. The PCA and PLS-DA analyses were performed using The Unscrambler Version 9.0 (CAMO A/S, Trondheim, Norway).

3.2.6 Protein identification

Since the number of *Pyrus communis* sequences in the public databases is very limited, MS/MS data analysis and cross species ID were applied for protein identification. Selected spots were manually excised with a 1.5 mm diameter pipette tip and trypsin-digested according to the methodology described by Shevchenko *et al.* (1996). Protein identification was performed by LC-ESI-MS/MS on an LCQ classic (Thermo Electron) with search paramaters as described by Dumont *et al.* (2004). All MS/MS spectra were first searched using SEQUEST (Thermo Electron) against a customized protein database containing all *Pyrus*, *Malus* and *Prunus* GenBank protein sequences (3683 entries), and MASCOT (Matrix Sciences, London, UK) against the GenBank non-redundant Viridiplantae-specified protein sequences. In a second identification round, a Mascot search was performed against the *Malus x domestica* EST sequences from the UniGene database of 25 August 2006 (182.378 entries). Proteins assigned on the basis of two or more significantly scored, non-redundant peptides were considered as confidently identified. Protein identification was carried out in collaboration with the Biomedical Research Institute, Hasselt University and School of Life Sciences, Transnationale Universiteit Limburg, Diepenbeek, Belgium.

3.3 Results

As a first step in the study of core breakdown in 'Conference' pears, suitable protein extraction and two-dimensional electrophoresis protocols were established, since until now no such protocols have been reported in literature. Pears have been reported to contain significant amounts of phenolic compounds, oxidative enzymes and other interfering substances (Espin *et al.*, 1998; Cui *et al.*, 2005) which can combine with proteins by hydrogen bonding or irreversibly by oxidation followed by covalent binding. The presented protocols, pay special attention to the removal of these contaminants and, hence, avoid formation of streaks in two-dimensional gels which complicate the 2-DE analysis.

3.3.1 Protein extraction

Three variants of the phenol extraction, methanol ammonium acetate precipitation protein extraction method were evaluated for pear tissue. This method has already been successfully employed in the extraction of proteins from different plant tissues such as banana, tomato, apple, orange, grape berries and olive leaves (Abdi *et al.*, 2002; Wang *et al.*, 2003; Saravanan & Rose, 2004; Carpentier *et al.*, 2005; Vincent *et al.*, 2006). In general, protein yields for 'Conference' pears were very low. In order to get higher yields for subsequent 2-DE analysis, the amount of DTT, the pH of the protein extraction buffer (PEB) and the use of PVPP were altered. Variant I and III resulted in much lower yields compared to variant II (Table 1). Variant II showed to be significantly different from variant I and III ($p < 0.05$) giving more than double the yield of variant I and 1.6 times the yield of variant III. The higher pH of 8.0 of the PEB of variant II with a higher amount of DTT resulted in higher yields. PVPP in the protein extraction buffer has been reported in previous studies as an efficient proton acceptor able to interfere with the binding of proteins to polyphenols (Pierpoint, 2004; Laborde *et al.*, 2006). However, at the pH in our buffer (variant II), polyphenols are in an ionized form which explains why the use of PVPP did not improve protein

Table 3.1: *Protein yields obtained from pear tissue by three different variants of the phenol extraction/methanol ammonium acetate precipitation method. Percentage yield was calculated as (g protein/g fresh pear weight)*100. Results are mean of six replicates. Different letters stand for statistical differences in sample means ($p < 0.05$) as determined by a Tukey test.*

Variant	Component of the PEB varied	Average protein yield (%)	SD
I	0.328 % DTT, final pH 7.0 and 1 % insoluble PVPP	0.025^a	0.004
II	1 % DTT, final pH 8.0	0.053^b	0.005
III	1 % DTT, final pH 7.0	0.033^a	0.006

yields (Carpentier *et al.*, 2005). Moreover, when using the PEB with PVPP, two-dimensional maps showed streaking (data not shown), probably due to the fact that PVPP can not be completely washed away and is still present in the protein pellet. A high pH 8.0 of the PEB was kept in order to inactivate proteases. DTT is a powerful reducing agent that reduces polyphenols to form thioethers (Loomis & Battaile, 1966) and helped to obtain higher protein yields. In general, both variants II and III resulted in clean 2-DE maps. Because of the higher protein yields and clean 2-DE maps variant II was chosen for further experiments.

3.3.2 2-DE

The 2-DE maps for all the pH ranges analyzed are shown in Figure 3.2. It was observed that at 5-8 pH range the number of proteins in brown tissue was smaller than for healthy and sound tissue which had approximately the same amount of protein spots, as demonstrated with a 2 way ANOVA test ($p < 0.05$). For the other pH ranges, no significant differences in the number of protein spots between the three tissue classes were observed in Figure 3.3. Protein quantification in the different tissue types showed that brown tissue presented 78.2 $\pm$ 7.5 μg protein/200 mg fresh tissue; sound tissue 93 $\pm$ 9.1 μg protein/200 mg fresh tissue and healthy tissue 106.4 $\pm$ 10.9 μg protein/200 mg fresh tissue. The results can be interpreted by

reduced protein synthesis and/or protein degradation. We therefore decided to characterize only the most important up-regulated proteins in brown tissue by performing a multivariate PLS-DA analysis (Figure 3.4) for the 5-8 pH range. The following detailed analysis is restricted to sound and healthy tissue, as they might indicate how the core breakdown disorder is actually triggered. A better comprehension of the mechanisms behind the onset of core breakdown might lead to improved commercial controlled atmosphere storage.

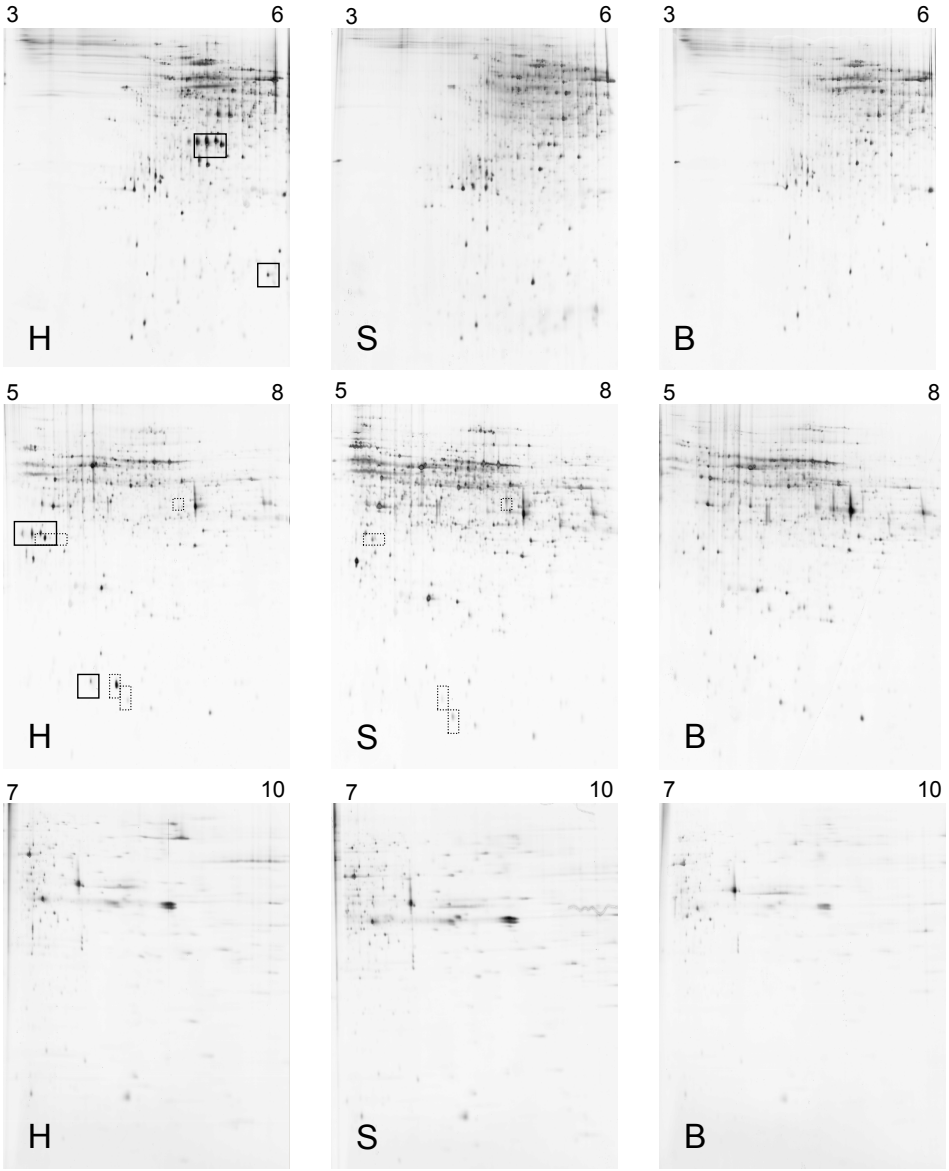

Figure 3.2: *2-DE maps for healthy (H), sound (S) and brown (B) pear tissues in the 3-6, 5-8 and 7-10 pH ranges. Gels were silver stained. Seventy, 35 and 100 μg protein was loaded for the 3-6, 5-8 and 7-10 pH ranges, respectively. Thick line squares refer to the same protein pairs in the different 3-6 and 5-8 pH ranges. Thin line squares refer to the same protein pairs within the 5-8 pH range.*

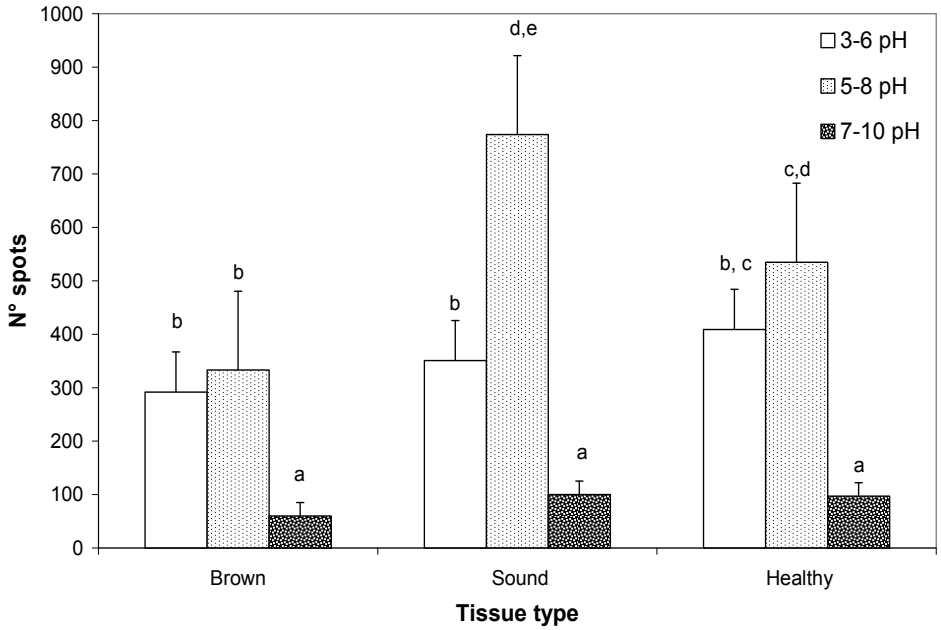

Figure 3.3: *Spot numbers for brown, sound and healthy tissue classes for the 3-6, 5-8 and 7-10 pH ranges. A two-way ANOVA was performed ($p < 0.05$). Tissue type and pH range were considered as factors. The response variable is the number of spots. The error bars indicate the standard deviation of the observations. Similar letters on top of the bars indicate no significant differences ($p < 0.05$).*

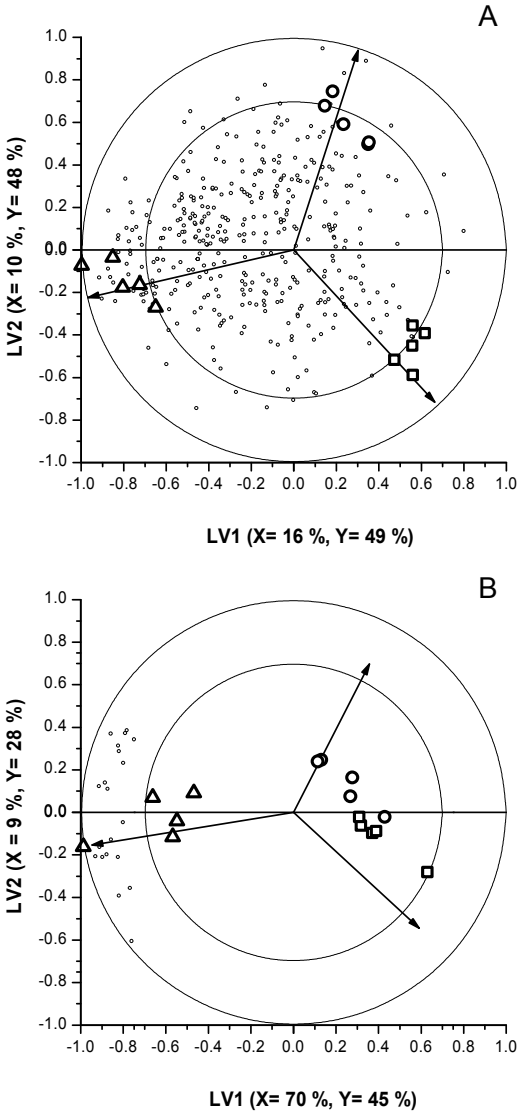

Figure 3.4: *PLS-DA analysis for the 2-DE data derived from the 5-8 pH range using brown tissue as reference gel. (A) Biplot showing the scores and correlation loadings for the three tissue classes before the VIP procedure and (B) biplot showing the scores and correlation loadings for the three tissue classes after the third VIP procedure. Triangles represent brown tissue, circles represent healthy tissue and squares represent sound tissue. The 20 selected spots were characteristic of brown tissue and correspond to the following spot numbers: 158, 160, 161, 163, 165, 167, 223, 252, 255, 334, 358, 366, 367, 369, 418, 421, 432, 447, 475 and 476.*

Table 3.2: *LC-ESI-MS/MS up-regulated identified proteins characteristic of brown tissue revealed by PLS-DA and VIP procedure. Cross-species matching with sequences in databases was used. n.d stands for non-detected. The spot numbers correspond to identified proteins found to be significant by the Kolmorogov Smirnov test ($p < 0.05$). % volume ± SD stands for the average volume of five or six replicates plus or minus standard deviation. PG stands for polygalacturonase. (H) healthy tissue, (S) sound tissue and (B) brown tissue. Significant non-redundant peptides[1] validated by Mascot/Sequest. The Mascot validation criteria used: expect value ≤ 0.05. The Sequest validation criteria used: cross correlation xC for 1^+, 2^+ and 3^+ charged ions ≥ 1.8, 2.5 and 3.5, respectively.*

spot number	Protein name	Protein ID		Theoretical		Experimental		% Volume ± SD			Peptides[1]
		Sequest/Mascot	Unigene	MW	pI	MW	pI	H	S	B	
160	Malic enzyme	NP_179580	Mdo.8158	64.3	6.3	68.5	6.15	0.44±0.18	0.35±0.08	0.67±0.12	3
161	Malic enzyme	NP_179580	Mdo.8158	64.3	6.3	68.5	6.30	0.42±0.14	0.36±0.13	0.59±0.06	4
163	Malic enzyme	NP_179580	Mdo.8158	64.3	6.3	68.5	6.40	0.70±0.19	0.45±0.13	1.04±0.12	6
165	Malic enzyme	NP_179580	Mdo.8158	64.3	6.3	68.5	6.45	0.46±0.05	0.31±0.05	0.60±0.16	4
167	Malic enzyme	NP_179580	Mdo.8158	64.3	6.3	68.5	6.50	0.37±0.06	0.22±0.01	0.48±0.09	4
252	ATP synthase β subunit	P19023	Mdo.940	59.1	6.0	56.2	5.40	0.36±0.25	0.23±0.03	0.82±0.27	6
255	ATP synthase β subunit	CAC35872	Mdo.490	59.7	6.2	56.2	5.45	0.90±0.26	0.74±0.14	1.59±0.35	14
358	Monodehydroascorbate reductase	AAM64531	Mdo.4862	46.5	6.4	41.5	7.30	0.05±0.02	0.08±0.00	0.16±0.04	4
367	PG inhibiting	AAP92910	Mdo.1066	36.5	6.2	40.5	6.60	0.23±0.06	0.18±0.05	1.04±0.54	3
418	Actin	BAD90938	Mdo.5936	38.4	5.5	33.0	5.80	n.d	0.03±0.00	0.08±0.03	9
432	Glyoxalase I	AAL84986	Mdo.2127	39.1	7.0	29.0	7.10	0.04±0.02	0.09±0.03	0.20±0.07	3
447	Malate dehydrogenase	P177783	Mdo.371	36.2	8.9	27.3	5.60	0.24±0.12	0.19±0.09	0.43±0.06	5
476	Nucleoside diphosphate kinase	NP_567346	Mdo.2732	18.8	8.4	16.3	7.00	0.33±0.08	0.28±0.08	0.73±0.20	5
876	PG inhibiting	AAP92910	Mdo.1066	36.5	6.2	40.7	5.90	0.20±0.06	0.11±0.04	1.31±0.53	3

3.3.2.1 Differentially expressed selected proteins by univariate statistical analysis

The classical approach of analyzing differences in protein expression levels is the use of univariate statistical tools which focus specifically on a pairwise comparison between types of tissue of individual protein spots (Salekdeh *et al.*, 2002; Shen *et al.*, 2003). Before performing the statistical analysis, gels were matched to a reference gel. A sound and a healthy reference gel were used interchangeably in order not to miss spots which are not present in the reference gel but obviously present in the other gels. When a sound gel was considered as reference gel, the Kolmogorov-Smirnov statistical test ($p < 0.05$) revealed the presence of 12 (pH 3-6), 28 (pH 5-8) and 5 (pH 7-10) differentially expressed spots when compared to a healthy gel. The number of differentially expressed proteins when a healthy gel was set as reference gel equalled 25 (pH 3-6), 17 (pH 5-8) and 3 (pH 7-10) spots when compared to a sound gel (Table 3.2). Differentially expressed spots are shown in Figure 3.5 for the 5-8 pH range and those identified spots are also shown in Figure 3.2 for all the pH ranges. There are many criticisms with respect to the use of univariate statistics to study protein differential expression (Karp *et al.*, 2005b; Marengo *et al.*, 2006). Univariate statistics are powerful for the analysis of individual proteins but do not allow comparing 2-DE gels as a whole. Given the complexity and high dimensionality of 2-DE gel based data, multivariate statistics offer more possibilities.

3.3.2.2 Differences in protein expression - multivariate statistical tools

As in the case of the univariate analysis described above, a sound and a healthy gel were used interchangeably as reference gels for this purpose. As mentioned earlier brown tissue was only used as reference gel for characterizing up-regulated brown tissue proteins. The identified up-regulated brown tissue characteristic proteins are presented in Table 3.2. PCA and PLS-DA were conducted for the 3-6, 5-8 and 7-10 pH ranges as to have a broad picture of proteins present in pear tissue. The complete analysis for the 5-8 pH

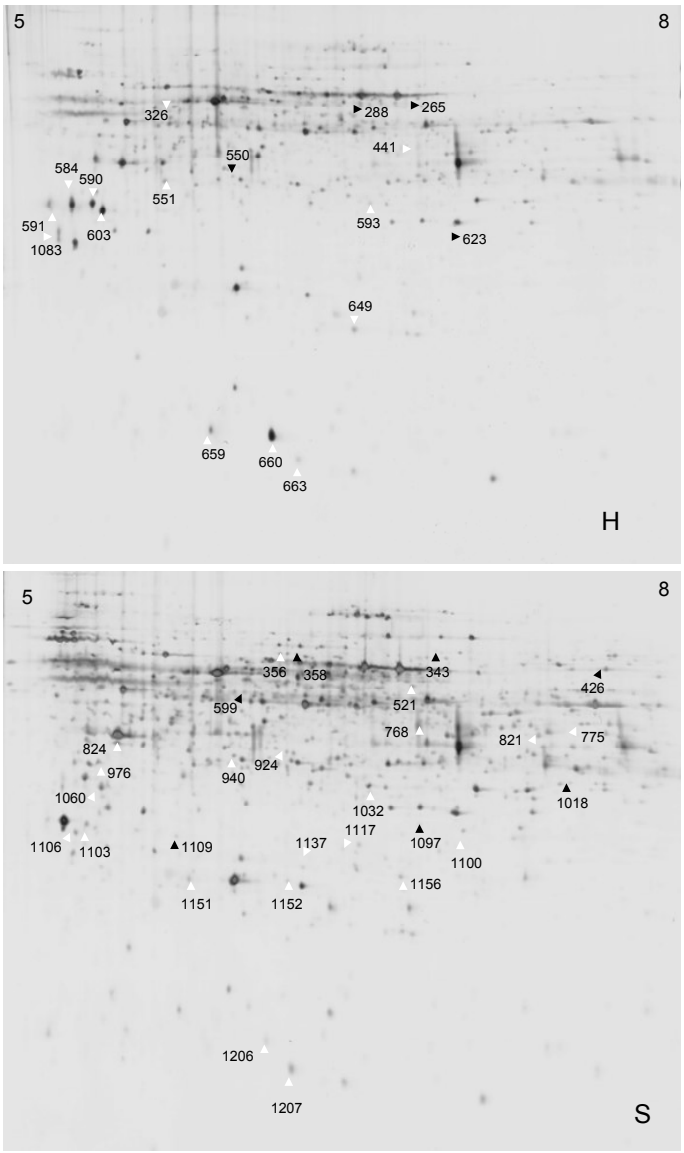

Figure 3.5: *Differentially expressed spots obtained by using a univariate (Kolmogorov-Smirnov test, $p < 0.05$) statistical approach. (H) Healthy tissue 2-DE map taken as reference gel and (S) Sound tissue 2-DE map taken as reference gel for the 5-8 pH range. White arrows indicate LC-ESI-MS/MS identified spots while black arrows refer to non-identified spots. Gels were silver stained and 35 μg protein was loaded.*

range data is presented below as more proteins were present in this range. For the 3-6 and 7-10 pH ranges only a brief description is given.

When 5-8 pH data were matched to a sound reference gel, an exploratory PCA analysis revealed a good clustering of the classes brown, sound and healthy along the PC1 axis. Thirty three percent of the variation in the samples was explained along PC1 and PC2. Similar or even lower values of explained variation were found for other proteomic data analyzed by PCA (Jessen *et al.*, 2002; Kjaersgard *et al.*, 2006). No significant outliers were detected on the PCA score plots. A PLS-DA with the purpose of discrimination of tissue types was then conducted. The score plot shows three separated clusters (Figure 3.6). The healthy gels were separated from the brown and sound gels along the LV2 axis and brown and sound gels were separated from each other along the LV1 (Figure 3.6). Proteins with a high positive or negative correlation with LV1 explain the discrimination and clustering of the brown tissue versus the non-brown tissue (healthy and sound). Similarly, proteins with a high correlation with LV2 are important in discriminating sound from healthy and brown tissue. After the VIP procedure, 19 of the 20 spots selected (Figure 3.6) were found characteristic of healthy and sound tissue and only one (spot 876) characteristic of brown tissue. The fewer spots that were detected in brown compared to sound and healthy tissue is also reflected in the correlation loading plot (Figure 3.6), where the number of proteins characteristic of brown tissue is much lower than that of sound tissue. This might be an indication of extensive proteolysis in brown tissue. This is in agreement with the significantly smaller amount of protein quantification in the different tissue types which also confirmed the presence of significantly lower amounts of protein in brown tissue. Therefore it was decided to continue the analysis focusing only on sound and healthy gels. Important selected spots obtained after performing the univariate and multivariate statistical analysis were sent for LC-ESI-MS/MS identification.

PCA and PLS-DA analysis considering sound (reference gel) and healthy gels were performed in order to directly compare the results with the ones obtained by the univariate statistical approach. PCA showed a good separation explaining with PC1 and PC2 40 % of the total variance. Both classes were

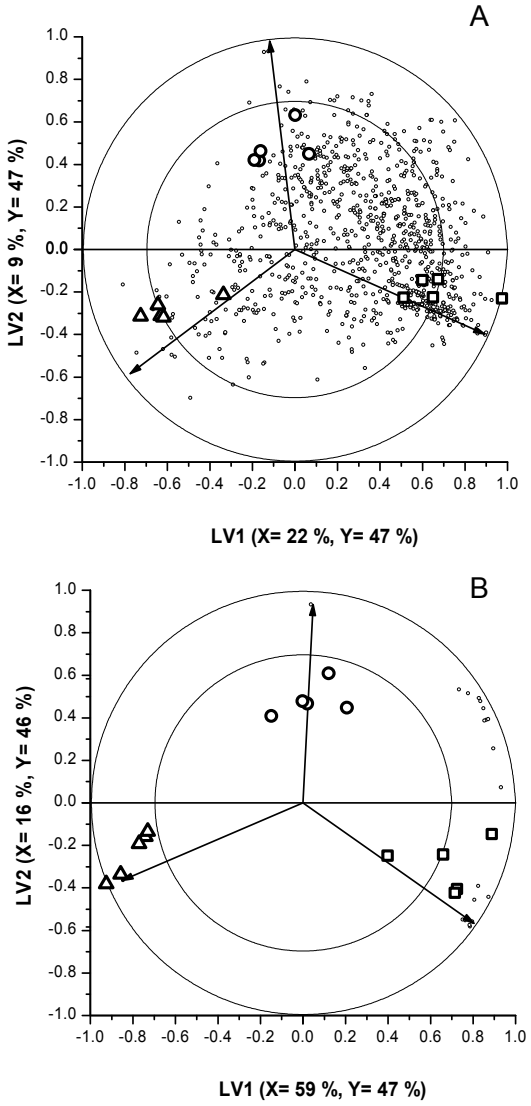

Figure 3.6: *PLS-DA analysis on 2-DE data for the 5-8 pH range with a sound gel taken as reference. (A) PLS-DA biplot before the VIP procedure showing the scores and loadings for the three tissue classes before the VIP procedure, (B) PLS-DA biplot after the third VIP procedure showing the scores and loadings for the three tissue classes after the VIP procedure. From the 20 spots selected, spot 876 was characteristic for brown tissue, spot 1060 was characteristic for healthy tissue and spots 358, 426, 521, 577, 914, 1109, and 1151 for sound tissue. The remaining spots 242, 367, 531, 604, 751, 792, 793, 799, 958, 1036 and 1060 were located between healthy and sound tissue and they were negatively correlated to brown tissue. Triangles represent (brown), circles (healthy) and squares (sound) tissues.*

mainly separated along PC1. The corresponding PLS-DA shows two nicely separated clusters along the LV1 axis (Figure 3.7). Twenty and fourteen percent of the total variance was explained by the first two LV components respectively. To identify the important protein spots involved in discriminating both groups, the analysis was focused on the correlation loading plot which represents the correlation between protein spots and class membership for all samples. The most important protein spots for classification were assessed by using the variance importance plot (VIP) iteratively. After the first VIP procedure where only the 50 most important proteins were kept, a better reconstructed model was obtained in which 70 % of the total variance was explained by LV1 and 9 % by LV2. A second VIP procedure which kept only 30 spots improved our model in such a way that 80 % and 3 % of the total variance was explained by LV1 and LV2, respectively. A third VIP procedure which kept the most important twenty spots resulted in a total variance of 86 % and 3 % being explained by LV1 and LV2 respectively (Figure 3.7). From these twenty spots, four of them (spots 1206, 1175, 1060 and 910) were characteristic for the healthy class while the remaining spots were characteristic for the sound class. From these twenty spots, sixteen were also found by using a univariate approach (Table 3.3).

The same analysis shown above with the three tissue types was used to analyze the data matched to a healthy reference gel. PCA analysis revealed a medium good separation among the classes: brown, sound and healthy. The total variance explained accounted for 36 %. PLS-DA analysis of the data explained 31 % of the X-explained variance with LV1 and LV2, obtaining nice clustering of the three different groups. In the following presented analysis brown tissue was kept out of the analysis for the reason mentioned above. PCA analysis explained 42 % of the total variance and did not reveal any outlier. However, PLS-DA revealed a good separation between healthy and sound gels along LV1. Eighteen and thirteen percent of the X-variance was explained by LV1 and LV2 respectively. In the correlation loading plot, the most important protein spots involved in class membership are located close to the respective classes. Finally, after the third VIP procedure when the most important 20 spots were kept, the model resulted in a good sepa-

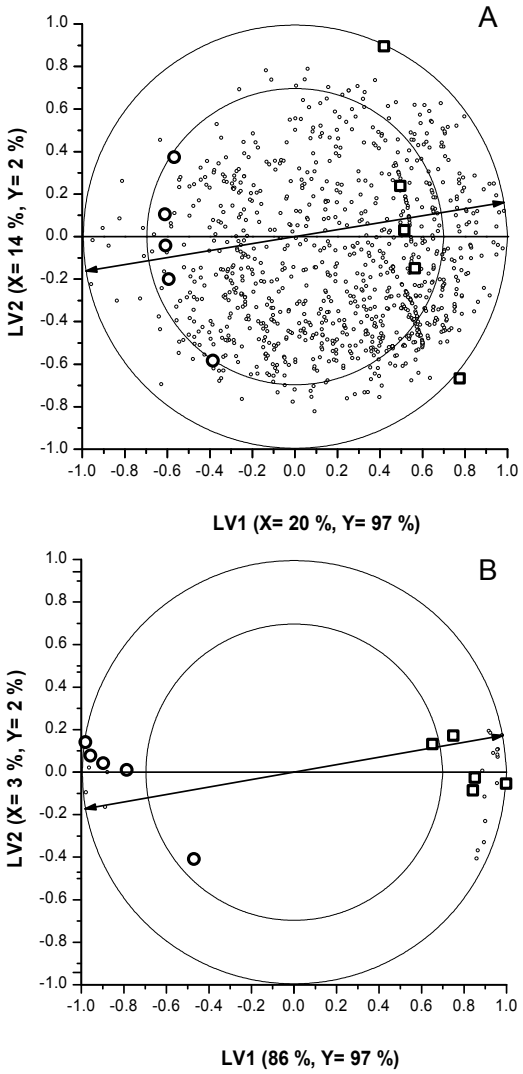

Figure 3.7: *PLS-DA analysis on 2-DE data for the 5-8 pH range with a sound gel taken as reference. (A) PLS-DA biplot before the VIP procedure showing the scores and loadings for only two tissue classes: healthy and sound before the VIP procedure and (B) PLS-DA biplot after the third VIP procedure showing the scores and loadings for only two classes: healthy and sound after the third VIP procedure. From the 20 selected spots, spots 910, 1060, 1175 and 1206 were characteristic for healthy tissue and the remaining spots 358, 521, 534, 685, 768, 775, 924, 940, 1017, 1032, 1097, 1137, 1151, 1152 and 1207 were characteristic for sound tissue. Circles represent (healthy) and squares (sound) tissues.*

ration of the two groups along LV1. Seventy nine and eleven percent of the X variance was explained by LV1 and LV2 respectively. Among these 20 protein spots, six were characteristic for the healthy class (spots 590, 584, 326, 660, 649, 655) and the remaining spots of the sound class. Nine out of the twenty spots selected were also obtained with the univariate analysis (Table 3.3). The univariate analysis revealed the presence of 17 significant spots. Several different spot IDs in different pH ranges (3-6 and 5-8) gels represent the same protein. For instance, the following pairs 1265 and 591, 1267 and 590, 1278 and 603 and 1392 and 659, represent the same protein as confirmed by protein identification. Within the 5-8 pH range, other confirmed pairs are 768 and 441, 1206 and 660, 1207 and 663, and 603 and 1060.

PLS-DA analysis of the 3-6 pH range data after the third VIP when only the 20 most important protein spots were kept showed an X-explained variance of 80 and 7 % with LV1 and LV2 respectively when data was matched to a sound gel and 80 and 13 % when data was matched to a healthy gel. The univariate statistical test revealed the presence of 12 and 25 significantly different spots when data matched to a sound or healthy gels of which 0 and 7 corresponded to the ones found by the multivariate statistical analysis (Table 3.3). This is in accordance with previous reports where also a higher number of differentially expressed spots were found when performing multivariate analysis as compared to univariate analysis (Karp *et al.*, 2005b; Marengo *et al.*, 2006). PLS-DA analysis of the 7-10 pH range data after the third VIP procedure, with 20 spots kept showed an X-explained variance of 27 and 44 % and 10 % and 75 % with LV1 and LV2, respectively when data matched to a sound or healthy gel respectively. The univariate statistical analysis revealed the presence of 5 and 3 significantly different spots when data matched to a sound or healthy gel of which four and one were also selected with the multivariate statistical analysis (Table 3.3).

3.3.3 Identification of relevant proteins

The statistically different spots obtained by the univariate and multivariate approaches were analyzed by LC-ESI-MS/MS. Thirty nine out of 90 spots (43 %) yielded a confident match with a pear or cross-species protein sequence from GenBank. An additional Mascot search against the apple EST database confirmed most of these identifications and additionally revealed 15 spots, increasing the identification rate to 60 %. This proved that Malus ESTs are a valuable data source for investigation of the poorly documented pear proteome. Confidently identified protein spots are listed in Table 3.3.

Table 3.3: *Identification of the significant spots revealed by univariate and multivariate approaches by means of LC-ESI-MS/MS and cross species matching with sequences in databases. n.d stands for non-detected. Protein spots marked with an * refer to proteins selected by both statistical approaches. The unmarked spot numbers correspond to identified proteins selected by Kolmorogov Smirnov test ($p < 0.05$). % volume ± SD stands for percentage volume average of five or six replicates plus or minus standard deviation. (H) healthy tissue, (S) sound tissue and (B) brown tissue. Proteins have been group according to biological function. Significant non-redundant peptides[1] validated by Mascot/Sequest. The Mascot validation criteria used: expect value ≤ 0.05. The Sequest validation criteria used: cross correlation xC for 1^+, 2^+ and 3^+ charged ions ≥ 1.8, 2.5 and 3.5, respectively.*

pH	Spot number	Protein name	Protein ID		Theoretical		Experimental		% Volume ± SD			Peptides[1]
			Sequest/Mascot	Unigene	MW	pI	MW	pI	H	S	B	
I.		*Energy and Metabolism*										
5-8	775*	Fumarase	NP_199908	Mdo.2315	54.1	6.0	37.1	7.8	0.03±0.00	0.07±0.01	n.d	10
5-8	1152*	Triosephosphate isomerase	P21820		27.2	5.5	21.4	6.3	n.d	0.02±0.00	n.d	3
5-8	1032*	Malate dehydrogenase cytoplasmatic	AAL11502	Mdo.3836	35.5	6.6	23.1	6.7	0.09±0.03	0.18±0.02	0.10±0.02	10
5-8	976	Adenosine kinase 2	AAU14833	Mdo.3056	37.9	5.2	23.1	5.4	0.05±0.01	0.09±0.02	n.d	5
II.		*Defense related*										
3-6	1392	Major allergen Pyrc1	ACC13315		17.6	5.6	20.6	5.7	0.36±0.06	0.20±0.03	0.16±0.03	7
5-8	356	stress induced sti1-like	AAM98143	Mdo.8227	60.8	6.1	61.0	6.2	0.03±0.01	0.05±0.00	n.d	2
5-8	768*	PG inhibiting	AAP92913		37.0	5.9	37.0	7.0	0.15±0.06	0.48±0.05	0.59±0.06	7
5-8	649*	Abscisic stress ripening	BAA96451		8.2	6.1	24.6	6.5	0.14±0.05	0.03±0.00	n.d	2
5-8	659	Major allergen Pyrc1	AAC13315		17.6	5.6	17.7	5.9	0.25±0.09	0.11±0.03	n.d	7
5-8	660*	Major allergen Pyrc1	AAC13315		17.6	5.6	17.6	6.1	1.72±0.46	0.06±0.04	n.d	11
5-8	663*	Superoxide dismutase	AAT66935	Mdo.597	15.1	5.8	16.5	6.3	0.11±0.01	0.27±0.05	0.32±0.10	4
5-8	821	Monodehydroascorbate reductase	BAA05408	Mdo.4862	47.5	5.3	31.4	7.6	0.04±0.00	0.09±0.03	n.d	14
5-8	441*	PG inhibiting	AAP92910		36.5	6.2	36.6	6.8	0.15±0.06	0.48±0.05	0.56±0.04	2
5-8	1100	ABA responsive		Mdo.8892			31.9	7.0	0.03±0.00	0.06±0.02	0.02±0.00	3
5-8	1103	L-ascorbate peroxidase	AAW49512	Mdo.7378	23.1	4.7	23.1	5.3	0.08±0.02	0.09±0.02	n.d	4

continued on next page

continued from previous page

pH	Spot number	Protein name	Protein ID		Theoretical		Experimental		% Volume ± SD			Peptides[1]
			Sequest/Mascot	Unigene	MW	pI	MW	pI	H	S	B	
5-8	1137*	Glutathione S transferase	BI977745	Mdo.2626			21.3	6.3	n.d	0.03±0.00	n.d	5
5-8	1156	L-ascorbate peroxidase	NP_849607	Mdo.1891	27.8	5.7	21.3	6.9	n.d	0.04±0.02	n.d	3
		Major allergen Pyrc1	AAC13315		17.6	5.6						2
5-8	1206*	Major allergen Pyrc1	ACC13315		17.6	5.6	17.7	6.2	1.72±0.46	0.08±0.01	0.07±0.01	10
5-8	1207*	Cu/Zn superoxide dismutase	CAA51654	Mdo.597	15.2	5.8	16.5	6.3	0.11±0.01	0.27±0.05	0.36±0.10	3
III.		*Ethylene biosynthesis related*										
3-6	1230	ACC oxidase	CAA60576		35.3	5.2	38.4	4.6	0.19±0.01	0.09±0.02	0.07±0.01	3
3-6	1265*	ACC oxidase	CAA60576		35.3	5.2	35.2	4.9	1.94±0.79	n.d	n.d	9
3-6	1267*	ACC oxidase	CAA60576		35.3	5.2	35.6	5.1	1.86±0.57	0.62±0.10	0.55±0.13	15
5-8	584*	ACC oxidase	CAA60576		35.3	5.2	36.0	5.2	1.63±0.35	0.07±0.00	n.d	13
5-8	590*	ACC oxidase	CAA60576		35.3	5.2	38.4	5.3	1.24±0.21	0.26±0.10	0.15±0.09	7
5-8	591	ACC oxidase	CAA60576		35.3	5.2	38.7	5.1	0.35±0.18	0.02±0.00	0.03±0.01	10
IV.		*Programmed cell death*										
3-6	1381	Proteasome subunit PAE2	NP_175788		26.1	4.7	25.0	4.4	0.09±0.02	0.22±0.16	0.46±0.17	5
3-6	1395	Proteasome β subunit PBC2	CAC43324	Mdo.4838	20.1	5.1	21.6	5.1	0.08±0.01	0.13±0.01	0.09±0.04	5
		Glyoxalase I		Mdo.10360								2
5-8	1151*	Proteasome subunit α type 2	NP_178042	Mdo.3411	25.7	5.5	21.4	5.8	n.d	0.09±0.02	n.d	8
V.		*Organization of the cytoskeleton*										
3-6	1096	β tubulin like	BAE47135		43.7	6.2	43.7	5.4	0.42±0.11	0.23±0.07	0.24±0.06	24
3-6	1108	β tubulin like	BAE47135		43.7	6.2	43.7	4.7	0.22±0.07	0.11±0.05	0.08±0.04	13
3-6	1112	β tubulin like	BAE47135		43.7	6.2	43.7	4.8	0.32±0.08	0.11±0.05	0.09±0.005	9
3-6	1124*	α tubulin	CAA47635		49.5	5.2	43.6	5.0	0.21±0.05	n.d	n.d	7
		high MW HSP	AAF34134	Mdo.440	71.2	5.2						4
3-6	1127	β tubulin like	BAE47135		43.7	6.2	47.6	4.8	0.22±0.07	0.11±0.05	0.10±0.005	3
5-8	593	Actin	BAD90938	Mdo.5972	38.5	5.5	38.6	6.6	0.10±0.04	0.02±0.00	0.08±0.03	2
5-8	824	Actin	BAD90938		38.4	5.5	30.9	5.5	1.40±0.00	1.39±0.26	1.30±0.26	13
5-8	924*	Actin	BAD90938		38.4	5.5	23.1	6.3	n.d	0.03±0.01	0.05±0.02	19
		β tubulin like	BAE47135		43.7	6.2						11

continued on next page

continued from previous page

pH	Spot number	Protein name	Protein ID		Theoretical		Experimental		% Volume ± SD			Peptides[1]
			Sequest/Mascot	Unigene	MW	pI	MW	pI	H	S	B	
VI.		*Other proteins*										
5-8	326*	Xylose isomerase	CAA10981	Mdo.14175	53.9	5.7	45.4	5.6	0.19±0.06	0.10±0.02	0.06±0.02	2
5-8	940*	Stearoyl-acyl carrier protein desaturase	AAM64846	Mdo.8239	45.5	5.8	23.1	6.0	0.04±0.00	0.06±0.00	0.04±0.00	2
		Glutamine synthetase	P32289	Mdo.8239	39.3	5.7	23.1	6.0				3
3-6	985*	Protein disulphide isomerase	CAA77575	Mdo.1921	57.3	5.0	57.3	4.4	0.34±0.07	0.19±0.05	0.20±0.04	3
3-6	1008*	HSP60		Mdo.5279			56.7	4.6	0.20±0.04	0.11±0.04	0.09±0.02	2
3-6	1238	O-glycosyl hydrolase		Mdo.15866			37.4	4.7	0.14±0.04	n.d	n.d	4
3-6	1278*	Copper amine oxidase like		Mdo.5972			34.8	5.2	1.72±0.30	0.84±0.20	1.02±0.07	8
3-6	1288	Unknown protein	XP_506643	Mdo.14102	24.3	4.8	35.0	4.8	0.09±0.02	0.15±0.02	0.13±0.02	3
3-6	1402	Peroxiredoxin	AAL90751	Mdo.3106	17.4	5.6	17.6	5.2	0.09±0.02	0.18±0.08	0.15±0.05	7
3-6	1408	Unknown protein		Mdo.3106			16.1	4.4	0.31±0.03	0.27±0.08	0.16±0.02	2
5-8	521*	Putative T-complex		Mdo.7988			49.0	6.9	0.03±0.00	0.07±0.01	n.d	3
5-8	551	β cyanoalanine synthase		Mdo.461			51.5	5.6	0.16±0.08	0.05±0.02	0.15±0.14	13
5-8	603	Copper amine oxidase like		Mdo.5972			34.7	5.3	1.15±0.48	0.26±0.03	0.30±0.14	7
5-8	1060*	Copper amine oxidase like	Mdo.5972				23.1	6.7	1.35±0.13	0.26±0.03	0.44±0.14	3
		Mercaptopyruvate sulfurtransferase	NP565203	Mdo.2858	42.2	6.0						2
5-8	1083	Unknown protein		Mdo.7844			32.6	5.1	0.20±0.09	0.05±0.00	n.d	4
5-8	1106	Soluble NSF attachment		Mdo.14701			23.0	5.2	n.d	0.11±0.09	n.d	2
5-8	1117	Unknown protein		Mdo.3751			22.7	6.6	n.d	0.06±0.02	n.d	4
3-6	1375	Adenine P-rybosyl transferase (APRT)	NP_193988	Mdo.2640	20.1	5.8	20.1	4.7	0.39±0.08	0.05±0.02	0.07±0.01	10
7-10	223*	O-glycosyl hydrolase		Mdo.15866			32.5	7.3	0.35±0.09	0.21±0.06	0.48±0.27	2

3.4 Discussion

Core breakdown disorder in Conference pears appears as a consequence of an altered metabolism due to stressing storage conditions. The change in expression of certain proteins in response to different abiotic stresses (ozone, heat, metals, hypoxia, wounding, drought, salinity and cold) has been extensively reported in plant tissues (Salekdeh *et al.*, 2002; Shen *et al.*, 2003; Cui *et al.*, 2005; Yan *et al.*, 2005; Hajheidari *et al.*, 2005; Askari *et al.*, 2006; Mittler, 2006). As a result of abiotic stresses, enzyme activities are altered which can lead to a disturbance of basic metabolism. In order to maintain homeostasis under stress conditions, plants need to fortify resistance mechanisms such as ion transport, reactive oxygen species (ROS) scavenging and osmolyte synthesis which are characteristic events during plant stress (Majoul *et al.*, 2003). In order to identify the cellular processes involved in core breakdown disorder, proteins and enzymes identified in this study have been grouped and will be discussed according to their function or metabolic pathway. All comparisons refer to healthy conditions.

3.4.1 Core breakdown is associated with an altered energy metabolism

This first group comprises enzymes involved in energy metabolism such as triosephosphate isomerase (spot 1152), cytoplasmatic malate dehydrogenase (spot 1032), fumarase (spot 775) and mitochondrial malate dehydrogenase (spot 447). Triosephosphate isomerase plays an important role in several metabolic pathways and is essential for efficient energy production. Its deficiency has been pointed out as the most severe disorder of glycolysis (Mande *et al.*, 1994). It was up-regulated in gels from sound tissue. It was almost absent in gels from healthy tissue and totally absent in gels from brown tissue. In rice roots exposed to salt-stress, triosephosphate isomerase has been reported to be up-regulated (Yan *et al.*, 2005). It has also been pointed out that when a plant is submitted to a stress, the resistance mechanisms that need to be activated in order to respond to stress require an extra energy

supply. The up-regulation of cytoplasmatic malate dehydrogenase in sound tissue might be related to the increased glycolysis as evidenced through up-regulation of triosephosphate isomerase as a way to translocate electrons into the mitochondria so that ATP can be synthesized. Fumarase (spot 775) is an enzyme involved in the hydration of fumarate to form L-malate. This enzyme was up-regulated in sound tissue. Fumarate has been shown to act as an activator of malate dehydrogenase (Grisson *et al.*, 1983). A much higher malic acid content has been reported in sound tissue than in brown (Pedreschi *et al.*, 2009a) and seems to be related to an up-regulation of fumarase in sound tissue. Degradation of fumarase in *Arabidopsis* cell cultures submitted to induced oxidative stress has been reported (Sweetlove *et al.*, 2002). Brown tissue showed complete disappereance of fumarase. The metabolic role of malate dehydrogenase has been reported to be consistent with the cooperative behaviour of its subtrate malate and the allosteric regulation of fumarate as an activator (Yang *et al.*, 2000). Induction of expression of a malate dehydrogenase gene in pea plant exposed to a Cd^{2+} stress has been reported (Savenstrand & Strid, 2004).

Five isoforms of malic enzymes (spots 160, 161, 163, 165 and 167) were up-regulated in brown tissue. This enzyme catalyzes the oxidative decarboxilation of L-malate to produce pyruvate, CO_2 and NADPH. This enzyme is considered to act as a housekeeping enzyme because it is involved in many functions in plants: fruit ripening, anabolic functions to provide NADPH and pyruvate for biosynthesis, catabolic functions to provide NADPH and pyruvate for energy production by respiration and the maintenance of intracellular pH (Edwards & Andreo, 1992). An increase in the level of NADP malic enzyme has been suggested to be related to defense plant mechanisms through providing building blocks and also energy for the biosynthesis of defense compounds (Casati *et al.*, 1999). For instance, when this enzyme was assayed in healthy and disordered infected cotyledons of marrow plants (*Cucurbita pepo* L.), the activity within the lesion was much higher compared to the activity outside the lesion increasing about 2 fold compared to the healthy tissue. The results were associated with the fact that infected tissues needed to increase biosynthetic capacity (Tecsi *et al.*, 1996).

It seems that malic enzyme in stressed tissues increases activity due to *de novo* synthesis and that the reducing power and pyruvate generated by this enzyme can be used not only for respiration in cellular repair processes but as substrates for fatty acid synthesis in order to repair membranes (Casati *et al.*, 1999). This might be the case for brown tissue. A possible role of this enzyme in the TCA cycle flux control under specific stresses is not discarded (Jenner *et al.*, 2001).

Two ATP synthase beta subunit isoforms (spots 252 and 253) were also found to be up-regulated in brown tissue. ATP synthases in general can synthesize adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate. In rice leaves exposed to cold stress at different temperatures, up-regulation of ATP synthase subunits was revealed (Cui *et al.*, 2005). From the results obtained, it seems that more energy is needed as to try to cope with the applied stress. Nucleoside diphosphate kinase (spot 476) was also up-regulated in brown tissue. This enzyme participates in the exchange of phosphate groups between the different nucleoside diphosphates. An enhanced expression of a gene encoding for nucleoside diphosphate kinase in rice plants exposed to bacterial pathogen infections has been reported (Cho *et al.*, 2004).

3.4.2 Defense related proteins involved in core breakdown

A second group of identified differentially expressed proteins is involved in defense mechanisms. The major allergen Pyrc1 (spots 659 and 660) was down-regulated in sound tissue and almost absent in brown tissue. In general, pathogenesis related (PR) proteins are considered to be involved in conferring some kind of protection to the plant during periods of stress (Abdi *et al.*, 2002). The polygalacturonase-inhibiting protein (spot 768/441) is a plant cell wall protein that protects plants basically from fungi invasion. They interact with endopolygalacturonases secreted by fungi and inhibit their enzyme activity favoring the accumulation of galacturonides which activate plant defense responses (Wayne, 2005). It has been observed in cantaloupes that as the fruit matured, the levels of this enzyme decreased

in the fruit (Hunt *et al.*, 2004). Polygalacturonase-inhibiting protein was up-regulated in sound and brown tissue compared to healthy tissue. The reason remains unknown.

The adenosine kinase isoform 2 protein is involved in catalyzing the transfer of a phosphate group from ATP to NAD to form NADP (Zorb *et al.*, 2004; Kwade *et al.*, 2005). Adenosine kinases are well characterized in plants for their requirement for the calcium binding protein calmodulin, thus its regulation is linked to stress-induced intracellular calcium release (Kwade *et al.*, 2005). This protein (spot 976) was up-regulated in sound tissue. A potential role of adenosine kinases in stress signaling has been highlighted (Kwade *et al.*, 2005). Evidence from several plant species revealed that adenosine kinase activity is increased in response to abiotic stress conditions (Delumeau *et al.*, 2000; Kwade *et al.*, 2005). As an example, adenosine kinase activity levels were higher in the salt tolerant tomato species *Lycopersicon pimpinellifolium* than in the salt sensitive one (Delumeau *et al.*, 2000). A possible role of adenosine kinases in the regulation of reactive oxygen species (ROS) removal has been postulated (Kwade *et al.*, 2005). Given that the only way of synthesizing new NADPH is through the reduction of NAD^+ synthesized by NAD kinases or through the action of a NADH kinase, the activation of adenosine kinases might be regarded as an important first step towards ROS synthesis.

Another enzyme involved in the plant antioxidant system is ascorbate peroxidase. Cytosolic ascorbate peroxidase, APX (spot 1103 and 1156) utilizes ascorbic acid and its specific electron donor to reduce hydrogen peroxide to water with the generation of dehydroascorbate which is spontaneously disproportionated into ascorbic acid and dehydroascorbate. Monodehydroascorbate is directly reduced to ascorbic acid by NAD(P)H-dependent monodehydroascorbate reductase. Dehydroascorbate reductase utilizes glutathione for reducing dehydroascorbate into ascorbic acid. The oxidized glutathione is regenerated by glutathione reductase utilizing the reducing equivalents from NAD(P)H (Davletova *et al.*, 2005). In this way, APX together with the ascorbate-glutathione cycle works as a mechanism to prevent the accumulation of toxic levels of hydrogen peroxide in plants. It has

been pointed out in previous studies that in absence of cytosolic ascorbate peroxidase, the entire antioxidant system of the plant collapses. As an example, Davletova *et al.* (2005) found that when this enzyme was absent, the entire hydrogen peroxide scavenging system collapsed having as consequences high levels of hydrogen peroxide and protein oxidation in *Arabidopsis thaliana*. The induction of protective antioxidant enzymes when abiotic stresses are applied to plants have been widely reported (Granier *et al.*, 2000; Van Breusegem *et al.*, 2001). Our results showed that APX was induced in sound tissue. In brown tissue, spot 1156 appeared but in significant smaller amounts compared to sound gels. This might be the result of cell death and protein degradation already taking place in brown tissue whereas sound tissue is still capable of withstanding the high ROS burst. The reduction or stop of plant growth under environmental stress conditions has been described before (Van Breusegem *et al.*, 2001). Under several stress conditions, a decrease in the number of cells, cell division rates has been observed in previous studies in leaves and roots (Granier *et al.*, 2000; Alscher, 1989). In a proteomic study of the disorder called 'blossom-end rot' in tomato fruits, Casado-Vela *et al.* (2005) found that antioxidant enzymes where up-regulated in the apparently healthy part of a disordered tomato. They also found that in the necrotized or affected part of the tomato fruit the number of spots detected was clearly lower than in the healthy half part or in the healthy tissue, thus indicating extensive degradation of proteins. The much lower number of spots found in brown tissue in our study might be the result of the same degradation process. Glutathione-S-transferase (spot 1137) appeared only in sound tissue and this enzyme has been pointed out as a key enzyme in protection against free radicals (Moran *et al.*, 1994). The level of this enzyme has also been pointed out to increase under various oxidative stress conditions (Moran *et al.*, 1994). Monodehydroascorbate reductase (spot 358) was also found to be up-regulated in brown tissue. Monodehydroascorbate reductase is the component in the ascorbate-glutathione cycle in charge of regenerating reduced ascorbate. In *Pisum sativum* exposed to drought a decreased activity of this enzyme was found which resulted in adequate hydrogen peroxide removal in such a way that it did not accumulate in the drought stressed leaves (Kumar Yadav *et al.*, 2005).

Glyoxalase (spot 432) was up-regulated in brown tissue. It showed to be present in higher amounts in brown followed by sound and healthy tissue. It has been shown in a previous study on transgenic tobacco plants that overexpression of this enzyme leads to an increased salt tolerance (Sommer *et al.*, 2001). The Glyoxalase system is in charge of the enzymatic desintoxication of glyoxal, methylglyoxal and other α-oxoaldehydes formed by lipid peroxidation, glycation and degradation of glycolytic intermediates. Its overexpression at elevated levels of oxidative stress has been reported (Fodoroff, 2006). Superoxide dismutase (spot 663/1207), another important antioxidant enzyme, catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide, was significantly up-regulated in sound tissue. Another isoform of this enzyme appeared up-regulated in brown tissue (spot 664). This enzyme is involved in the detoxification of ROS through the ascorbate-glutathione cycle. Up-regulation of this enzyme in areas affected with blossom-end rot in tomato fruits has been reported (Casado-Vela *et al.*, 2005).

Chaperones including protein disulfide isomerase PDI (spot 985) and heat shock protein HSP (spot 1008) both down regulated in sound and brown tissue, are involved in the process of protein biosynthesis and correct folding as well as in the transfer of proteins between organelles. In plants, there is evidence of clear linkage with stress responses (Fodoroff, 2006).

3.4.3 Proteins involved in ethylene biosynthesis

A third group of identified proteins can be classified as proteins involved in ethylene biosynthesis. The hormone ethylene is a modulator of plant growth and development. It is involved in many aspects of plant life cycle including seed germination, root hair development, senescence, abscission and fruit ripening (Capitani *et al.*, 1999). Ethylene is tightly regulated by internal signals during development and in response to biotic (e.g., pathogen attack) and abiotic stresses such as wounding, hypoxia, ozone, chilling, and freezing (Lougheed, 1987). Low oxygen and high carbon dioxide concentrations have been reported as controllers of ethylene biosynthesis and its action

(Lougheed, 1987; Gorny & Kader, 1996a; Capitani *et al.*, 1999). Three isoforms of the 1-aminocyclopropane-1-carboxylate (ACC) oxidase enzyme (spots 584, 590/1267 and 591/1265) were found to be down-regulated in sound and brown tissue. ACC oxidase catalyses the final step in the biosynthesis of ethylene. Brown and sound tissues were stored at very low oxygen levels and very high carbon dioxide levels. Gorny & Kader (1996a) showed for 'Golden delicious' apples that reduced oxygen and elevated carbon dioxide concentrations impeded ethylene biosynthesis, directly by reducing ACC oxidase catalytic activity to convert ACC to ethylene, and indirectly by blocking regulation of the genes encoding ACC oxidase. The abundance of active ACC oxidase was reduced. ACC oxidase needs oxygen as co-substrate and ascorbate acts as co-factor. Inactivation of this enzyme has been related to high amounts of hydrogen peroxide (Vranova *et al.*, 2002). In our case, we observed a decrease in the expression level of the enzyme in sound and brown tissue, which might be correlated to a certain point with the described finding of inactivation and decreased activity. High hydrogen peroxide concentrations could be present in disordered pears and result in ethylene biosynthesis inactivation. It has also been suggested in previous studies that the way plant growth and metabolism are affected by stress conditions is through the interference of ROS with other signaling pathways or molecules. There is evidence that plant hormones are influenced by ROS signals (Burssens *et al.*, 2000; Vranova *et al.*, 2002).

3.4.4 Proteasomes involved in core breakdown

Three different proteasome subunit proteins were found to be up-regulated in sound and brown tissues: α type 2 (1151), α type 5 (1381), and β type 3 (1395). Proteasomes are involved in ATP/ubiquitin-dependent non-lysosomal proteolysis of poorly folded and damaged proteins. They play a key role in regulated proteolysis throughout the life cycle of plants (Sullivan *et al.*, 2003). Proteasomes have been shown in previous studies to play a key role in programmed cell death in plants but still more clear evidence is needed (Kim *et al.*, 2003; Tsunezuka *et al.*, 2005). An increase in

proteasome abundance has been reported by Dawson *et al.* (1995) in intersegmental muscle cells undergoing programmed cell death. The cell death mechanism involved in core breakdown disorder is something that should be further investigated.

3.5 Conclusions

In this work a novel approach to study the physiological disorder core breakdown, that can affect Conference pears during CA storage, was employed. Core breakdown can be considered as a model system for controlled atmosphere related physiological disorders. Proteins extracted from tissue from healthy and disordered pears were subjected to a proteomic analysis employing two-dimensional electrophoresis and protein identification via LC-ESI-MS/MS. As a first step suitable protein extraction/precipitation and 2-DE protocols were developed. A variant of the classical phenol extraction/methanol ammonium acetate precipitation, with a pH of 8.0, 1 % DTT and no PVPP added, proved to be the best. Optimal 2-DE protocols for pH 3-6, 5-8 and 7-10 and different sample loads were also established. Univariate and multivariate statistical techniques to analyze protein differences among the different tissue classes involved in core breakdown disorder (sound, brown, healthy) showed to be effective and consequent in the results. It is suggested that both approaches should be used together to account for the proteins that are really differentially expressed and not the result of mere false positives. Spots recognized to be significantly different by both approaches should receive strong attention. In general, multivariate statistical tools such as PCA and PLS-DA are recommended when a great amount of data is generated. This proteomic approach combined with univariate and multivariate tools and LC-ESI-MS/MS identification of some differentially expressed proteins was effective in giving a general snapshot of the biochemical mechanisms involved in core breakdown. Proteins involved in energy metabolism and antioxidant system were clearly involved in the causes of the disorder, as well as ethylene biosynthesis. This study, for the first time, provides a general overview of the biochemical mechanisms

involved in core breakdown disorder in ‘Conference’ pears.

Chapter 4

Treatment of missing data for multivariate statistical analysis of gel-based proteomics data

4.1 Introduction

Two dimensional electrophoresis (2-DE) requires proper data analysis techniques to avoid misleading conclusions. The use of post run protein stains for quantitative analysis is currently being questioned due to its limited power in terms of dynamic range, sensitivity and variability (Miller *et al.*, 2006). The improved power of the DIGE approach arises from the use of an internal standard (Tonge *et al.*, 2001) which is used to calculate a standardized abundance of each spot and to match the spots across the gels. The classical post run dyes are however still useful as long as the technical variance is kept low and the number of replicates is high enough. The use of appropriate statistical tools to interpret the data is a must, either with classical dyes or with DIGE. The simplest statistical analysis commonly involves pairwise comparison using parametric or non parametric tests while

more complicated statistical analysis involves the use of multivariate statistics and multiple comparison tests (Fuji *et al.*, 2005; Tuomainen *et al.*, 2006; Pedreschi *et al.*, 2007). Before applying a statistical test, its assumptions need to be fulfilled and some data pre-processing might be required depending on the experimental data. If a parametric test is used, for every protein, normality and homoscedasticity should be tested. For a small number of replicates (3-6 in most proteomics studies), the Shapiro-Wilk test is the most reliable test for normality (Shapiro & Wilk, 1965). It has been shown that low intensity spots exhibit a smaller variance between replicate gels as compared to high intensity spots (Grove *et al.*, 2006). As the data should be homoscedastic (show equal variances), some form of data transformation (log, asinh, square root) is required (Karp *et al.*, 2005b). Another important issue is that samples should be independent to prevent false positives (Karp *et al.*, 2007).

Proteomics data always contain missing values; being a spot detected in the reference or master gel but not in the sample gel. The main causes for the occurrence of missing values are (i) spots below a threshold or detection limit; (ii) mismatches caused by distortions in the protein pattern (iii) absent spots due to bad transfer from the first to the second dimension or (iv) truly absent spots from the samples. Two-dimensional data can have around 50 % missing values (Wood *et al.*, 2004; Jung *et al.*, 2005; Krogh *et al.*, 2007). However, there are no straightforward rules how to deal with missing values. It has been demonstrated that the deletion of variables containing missing values assumes that the number of missing values is relatively small and completely at random (Little & Rubin, 1987). But, in gel-based proteomics, the number of missing data is often considerable and not at random but for instance correlated to the staining procedure or the mean volume percent of the matched spots (Grove *et al.*, 2006). If variables with missing values are just discarded or ignored, a substantial bias can be introduced because information is simply lost. One other possibility is filling the missing values with zeroes or some lower threshold value. When a missing value is the result of a spot being below the detection limit, a threshold or zero value can be justifiable. However, whenever a value is missing due to mismatching,

this would lead to wrong interpretation of the results (Little & Rubin, 1987; Wood *et al.*, 2004). Several methods have been suggested to impute missing values such as: the row average method, k-nearest neighbor (KNN), singular value decomposition (SVD) impute algorithm (Troyanskaya *et al.*, 2001; Jung *et al.*, 2006), Bayesian Principal Component Analysis (BPCA) missing value estimation method (Oba *et al.*, 2003) and the Maximum Likelihood algorithm (Krogh *et al.*, 2007).

Multivariate statistical packages such as Unscrambler (CAMO, Trondheim, Norway), Decyder EDA (GE Healthcare, Upsula, Sweden) and SIMCA-P (Unimetrics AB, Sweden) can deal with missing values during multivariate analysis (PCA, PLS-DA) avoiding the need to impute them. They rely on the NIPALS algorithm to set the residuals for the missing values to zero during the calculations of the principal components or latent variables. This flexibility for the user to perform the analysis when missing data is present can represent a serious problem if the amount of missing data is substantial. Moreover, the amount of missing data that is considered to be substantial to distort the results is debatable. Currently, the all gels against all gels matching approach introduced by some image analysis packages (e.g., Progenesis Same Spots), theoretically generates complete datasets suitable for multivariate statistical analysis after proper data standardization. However, technical issues intrinsically associated with 2-DE and image analysis such as: gel distortions, missing spots due to bad transfer from first to second dimension, incorrect spot merging or splitting, are ignored introducing misleading values that generate a bias (Karp *et al.*, 2008). Considering all the possibilities available we believe it is crucial to be aware of the importance of how missing values are faced. Whatever approach is taken in the end, one must consider the structure of the data and a compromise should be found between a sound statistical and biological interpretation of the data.

Multivariate statistics have a key role to play in 'systems biology' because much more information can be extracted than by a simple univariate test. Therefore there is an urgent need to handle missing values in an accurate way to draw realistic conclusions. When univariate statistical tests are performed (e.g., t-test) it might be argued that missing values can be

ignored analyzing only the available data. The reduced number of replicates due to missing values would result in a reduced power and possibly also in wrong conclusions. In both univariate and multivariate statistical analysis, missing values represent a problem. The univariate statistical analysis in presence of missing data is out of the scope of this study.

The main objective of this chapter is to focus attention on different techniques to handle missing values for multivariate statistical analysis and the subsequent possible impact on the interpretation of the results. Datasets generated with classical stained dyes as well as DIGE will be utilized to draw the recurring problem of missing data in gel-based proteomics data. Different methods to handle missing values will be evaluated and their influence on the selection of important proteins through multivariate techniques will be discussed. The results of this chapter were published in Pedreschi *et al.* (2008b)[3].

4.2 Materials and Methods

4.2.1 Proteomics data

Proteomics datasets from pear and banana were used as case studies. Technical details for pear and banana proteomics can be found in respectively Pedreschi *et al.* (2008a) and Carpentier *et al.* (2007). For this reason the experimental background of these datasets is only described in brief. The pear dataset contains data from six independent biological replicate samples for each of four treatments (different storage gas conditions). Proteins were visualized by silver staining (Blum *et al.*, 1987). Image analysis was performed with the Image Master 2-D Platinum software 6.0 (GE Healthcare). Spots were detected without spot editing and quantified as percentage volume. The banana data set contains data from three replicate gels for each of four treatments (different sample dates; 2, 4, 8 and 14 days). Samples

[3]Pedreschi et al. (2008). Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. Proteomics, 8, 1371-1383.

were labeled using the fluorescent Cyanine dyes developed for DIGE (GE Healthcare) according to the manufacturer's recommendations. In order to anticipate any dye specific effect, the samples were labeled at random with Cy3 and Cy5 and randomized over the gels. The internal standard was a mixture of all analyzed samples and was labeled with Cy2. Labeled proteins were visualized using a Typhoon^{TM} imager (GE Healthcare) and the gels were analyzed using the Decyder EDA software. The data pre-processing with DIGE occurs automatically in the DECYDER^{TM} software: the data is normalized using a ratiometric approach and a log_{10} transformation is used on the standard abundance to stabilize the variance.

4.2.2 Handling of missing values

Three methods to handle missing values were tested which consisted of two imputation techniques preceding the multivariate analysis (KNN and BPCA) and simply dealing with the missing values during the multivariate analysis (referred to as 'NIPALS').

4.2.2.1 *k*-Nearest neighbor (KNN)

The KNN method assumes a relationship between spot volume patterns of groups of proteins. The KNN method selects spots showing spot volume patterns similar to the spot of interest for which to impute missing values (Jung *et al.*, 2006). A weighted average of values from the k most similar spots is used as an estimate for the missing value under concern. The contribution of each spot is weighted by its similarity determined as the Euclidean distance. The optimum number of k-neighbors has to be determined empirically. The KNN imputation procedure was implemented in Matlab (The MathWorks, Inc., Natick, MA, USA) by Jorsten *et al.* (2005) and applied in this chapter using k=20. KNN has no theoretical criteria for selecting the best k values. This number has to be empirically determined. In this chapter a k=20 value was chosen because previous studies showed to be the optimal number (Oba *et al.*, 2003; Brock *et al.*, 2008).

4.2.2.2 Bayesian Principal Component Analysis (BPCA)

In BPCA the missing values are estimated from the known spot volumes using principal component regression (PCR). The principal components are estimated simultaneously with the regression coefficients of the PCR model using a variational Bayes algorithm. After convergence of the algorithm missing values are imputed. The BPCA imputation procedure was implemented in Matlab (The MathWorks, Inc., Natick, MA, USA) by Oba *et al.* (2003). BPCA consists of three processes as described above: (i) principal component regression, (ii) Bayesian estimation and (iii) Expectation-maximization (EM) like repetitive algorithm. For a detailed explanation the reader is referred to Oba *et al.* (2003). BPCA does not require parameter optimization, since the parameter settings are predetermined on before hand (Brock *et al.*, 2008). BPCA depends on the number of principal axes or eigenvectors. In BPCA the number of samples minus 1 is chosen as the number of principal axes (Yoon *et al.*, 2007). In this approach, the algorithm automatically screens for those axes that are the most relevant. Scores and loadings obtained with BPCA are slightly different from those obtained with conventional PCA because BPCA has been specially developed for missing value estimation.

4.2.2.3 Nonlinear Estimation by Iterative Partial Least Squares (NIPALS)

Both Unscrambler and Decyder EDA softwares are able to perform multivariate analysis in the presence of missing data using the NIPALS algorithm. In every iteration, during calculation of the principal components or latent variables, the residuals for the missing elements in the least square function are set to zero or the missing values are replaced by their minimum distance projections onto the current estimate of the loading and score vector (Nelson *et al.*, 1996). This method is generally used in chemometrics and proteomics (Grung & Manne, 1998) and is tolerant to small amounts of missing data (up to 5-20 %).

4.2.3 Performance of handling missing values

The performance of handling missing values was tested on a subset of the DIGE dataset referred to as 'complete DIGE' dataset containing 542 proteins matched across all the gels without missing values. The experimental set-up is described in Figure 4.1. From this 'complete DIGE' dataset thirty percent of the data was randomly removed. Using this dataset with artificially induced missing values, the various methods for handling missing values described above were tested. Since the underlying normality and equal variance assumptions are supposed to be met with DIGE data after Decyder analysis (Karp & Lilley, 2005), transformation of the data was not required. Outliers were detected using the 95 % Hotelling's T^2 limit (Johnson & Wichern, 1998). The multivariate data analysis involved PLS-DA analysis to discriminate the individual gels according to similar protein expression profiles. Cross-validation was applied to test the performance of the models since the number of observations is too small to validate the models on an independent test set. The VIP procedure was used to identify the 50 most important proteins describing the difference in protein expression profiles. These selected proteins were compared between the different approaches of handling missing values using the 'complete DIGE' dataset as a reference. This procedure, starting from the induction of random missing values, was repeated 10 times to evaluate its consistency. A method is considered to be 'consistent' if by repeating several times (10 in this particular case), the obtained proteins are the same as the 'real ones' (obtained when no missing values are present). PLS-DA and VIP analyses were performed using The Unscrambler Version 9.1 (CAMO A/S, Trondheim, Norway).

4.2.4 Impact of missing values handling techniques on VIP selection using incomplete DIGE data

To test the impact of different missing values handling techniques on the final VIP selection, the original incomplete DIGE data (covering 1462 proteins, containing missing values) was used. As the normality and equal

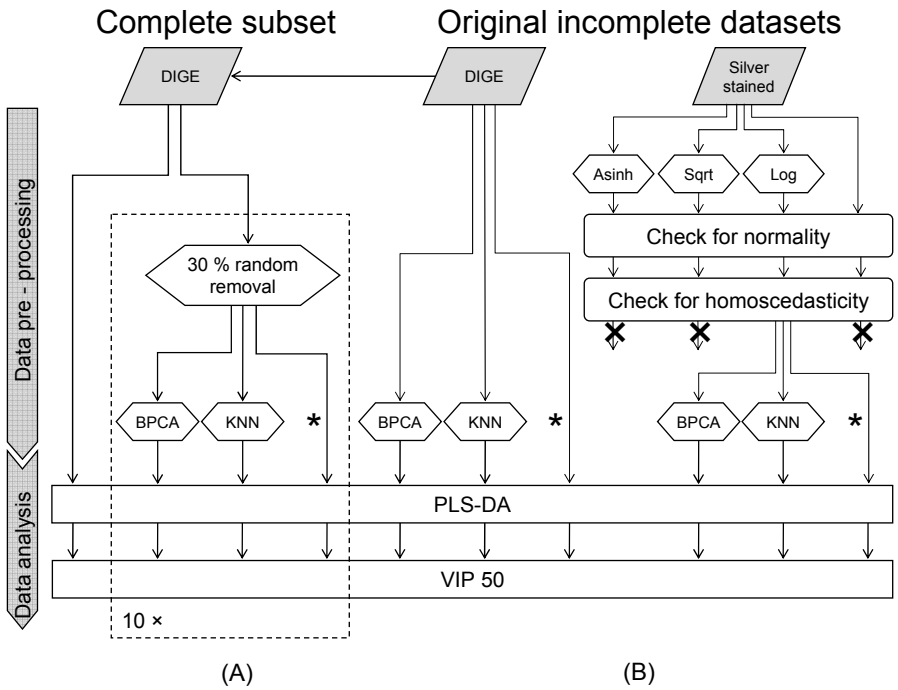

Figure 4.1: *Flow chart detailing the procedure followed for (A) testing the performance of handling missing values using the 'complete dataset' composed of 542 totally matched proteins, ten times (10x) and (B) testing the impact of missing values handling techniques on VIP selection using 'incomplete datasets': DIGE and classical dyes. The asterisk indicates that missing values were not imputed during pre-processing but were handled during the multivariate analysis through the NIPALS algorithm.*

variance assumptions were assumed to be met, transformation of the data was not required. Missing values were handled either during the multivariate analysis (NIPALS) or by imputing them on beforehand using either the KNN or BPCA method. PLS-DA and the VIP procedure were used to build models able to explain the variance in the dataset. The followed procedure is described in Figure 4.1.

4.2.5 Impact of missing values handling techniques on VIP selection using classical dyes data

Normality was checked with the Shapiro and Wilk test. To meet the equal variance assumption, different transformations were tested: no transformation, a logarithmic (log), inverse hyperbolic sine (asinh) and square root transformation. Handling missing values during the multivariate analysis (NIPALS) was compared to imputing them on beforehand using either the KNN or BPCA method. If for a particular protein in one of the treatments all replicates presented missing values but were clearly present in the other treatments, they were treated as threshold values. Before performing PLS-DA and the VIP procedure to select the fifty most important proteins involved in class distinction, PCA outlier detection through the Hotelling T^2 ellipse was performed.

4.3 Results

4.3.1 Matching of the data and estimation of missing values

The percentage of missing values in either the DIGE or classical dyes datasets was 24 % and 29 % respectively (Tables 4.1 and 4.2). Despite the use of an internal standard and the co-detection algorithm with the DIGE, the individual gels still need to be matched resulting in substantial amounts of missing values (Table 4.1). The total number of spots fully matched across all samples of the DIGE dataset was 542.

Table 4.1: *Matching results for the incomplete DIGE dataset.*

Gel (treatments: Cy3, Cy5, Cy2)	Detected spots	% Spots matched to master gel 3
Gel 1	1601	75
Gel 2	1532	67
Gel 3	1692	100
Gel 4	1412	67
Gel 5	1548	78
Gel 6	1256	69

Table 4.2: *Matching results for classical dyes dataset.*

Treatments:	Average detected spots ($n = 6$)	% Spots matched to reference gel
Condition 1	733	63
Condition 2	520	64
Condition 3	622	69
Condition 4	609	63

4.3.2 Performance of handling missing values

The 'complete DIGE' dataset (542 proteins) was used to evaluate the performance of different methods to handle missing values after random removal of 30 % of the data (Figure 4.2). Based on the score plots, none of the methods clearly outperformed the others in terms of quality of the separation (Figure 4.3). The score plots are a useful visualization tool to inspect if the real variance from the 'complete dataset' is being masked or not by the tested methods to handle missing values in the derived datasets with artificially induced missing values. Particularly, since we have the 'complete dataset' a direct comparison can be made. However, looking at the proteins involved in the classification, quantitative differences are observed. Depending on how missing values were handled, in average only 34 % to 63 % of the selected proteins were identical to the fifty selected proteins obtained from the 'complete DIGE' dataset (Figure 4.4). The number of imputed missing values in these fifty selected proteins for all the methods tested did not dif-

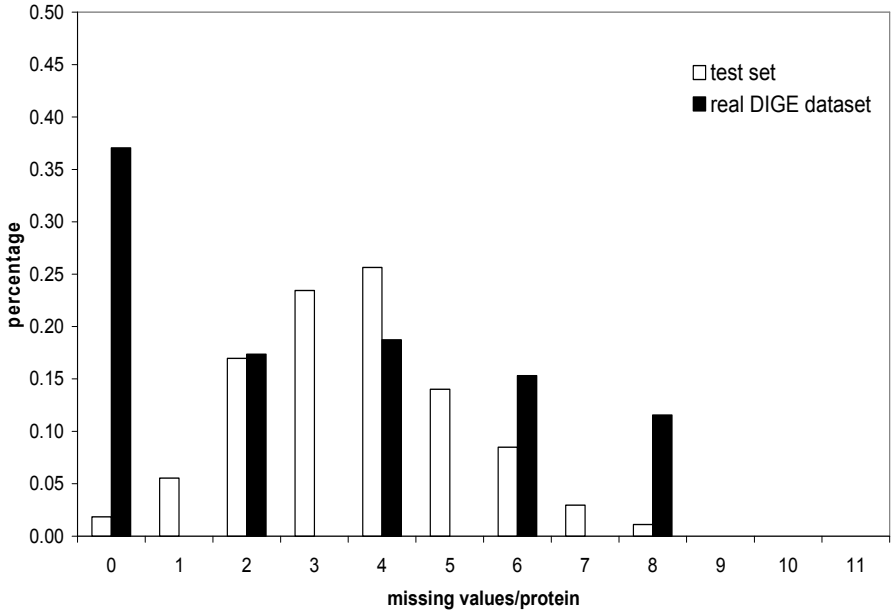

Figure 4.2: *Distribution of missing values/protein for (a) random removal in complete DIGE dataset (test dataset containing 542 proteins matched across all gels) and (b) incomplete DIGE dataset (containing 1462 proteins).*

fer extensively. In addition, the BPCA imputed data seems to be closer to the original data (Figure 4.5) as compared to the KNN imputed data. The calculated correlation coefficients for the real data *vs* BPCA imputed data and real data *vs* KNN imputed data were 0.85 and 0.65, respectively. These coefficients clearly show that BPCA provides more accurate estimates of the missing values than KNN. The selection of proteins for the KNN also varied extensively during the ten simulations (34 % $\pm$ 17 %; Figure 4.5). From these results, BPCA showed to be the most consistent method in terms of selecting those proteins that would have been selected if there would have been no missing values in the dataset.

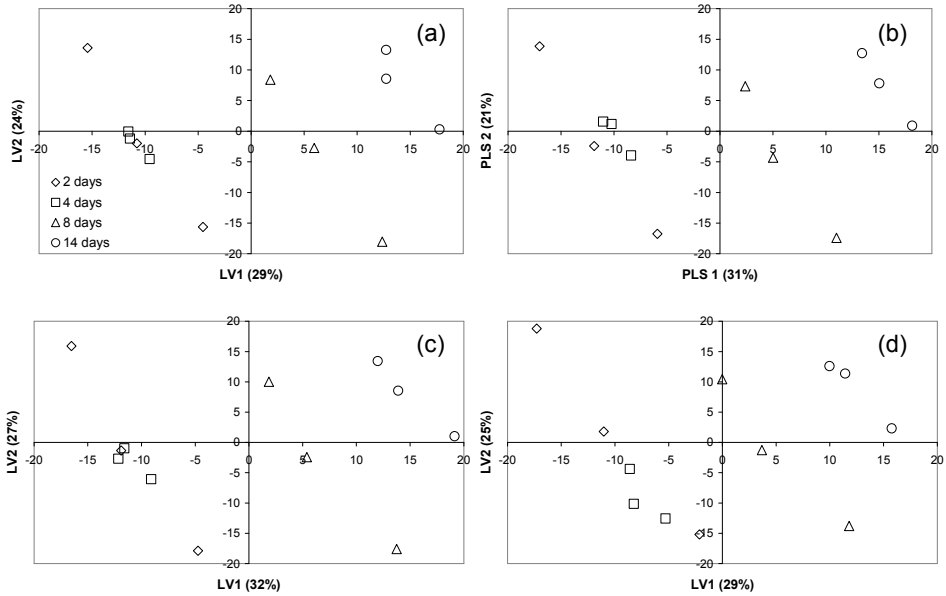

Figure 4.3: *PLS-DA score plots for the (a) complete DIGE dataset (542 proteins matched across all gels), (b) after random removal of 30% of the data and treated with Unscrambler (NIPALS algorithm), or imputed with (c) BPCA or (d) KNN.*

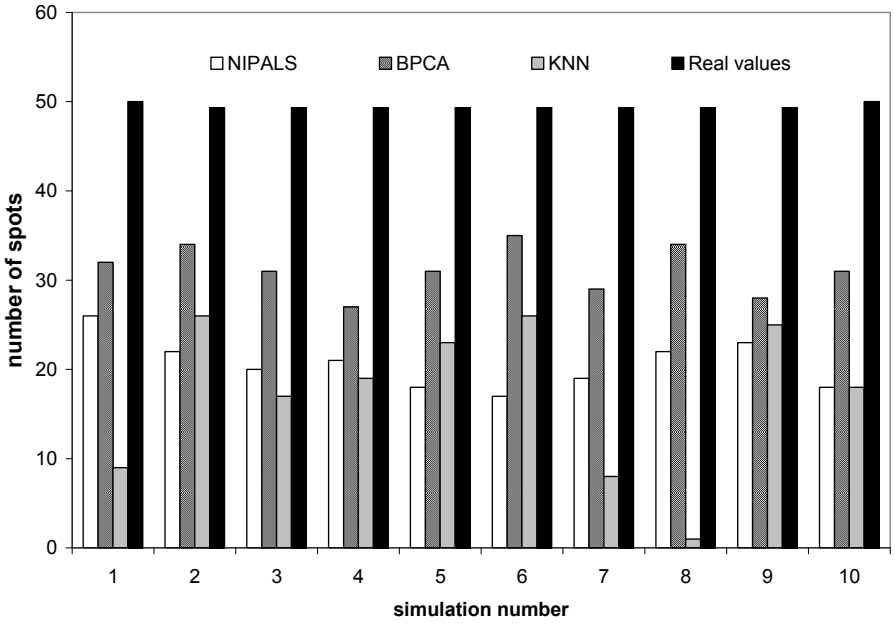

Figure 4.4: *Number of proteins selected through VIP50*, coinciding with the 50 proteins selected on the original complete DIGE dataset. Artificial datasets were created by randomly removing 30 % of the data from the 'complete DIGE dataset'. Subsequently, these missing values were handled using NIPALS, BPCA or KNN. The procedure was repeated 10 times.*

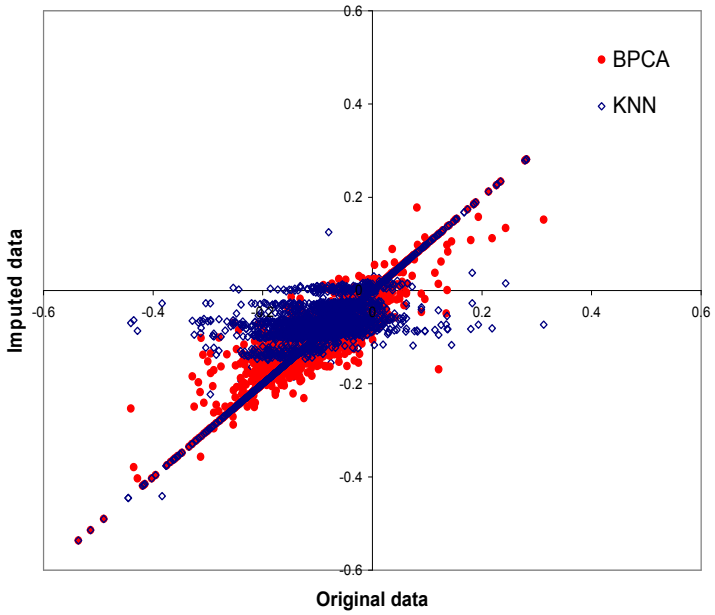

Figure 4.5: *Complete DIGE dataset versus imputed data with BPCA and KNN.*

4.3.3 Impact of missing values handling techniques on VIP selection using incomplete DIGE data

Depending on how missing values were handled different selections of 50 proteins were obtained for the original incomplete DIGE data (covering 1462 proteins, containing 24 % missing values). Between KNN and BPCA 30 out of the 50 selected proteins were the same. When the missing data was handled during the multivariate analysis (NIPALS), only one out of the fifty proteins was the same when compared to the BPCA method which in the previous section was shown to perform best (Figure 4.6). Most of the proteins selected based on the BPCA imputed data contained no missing values while the proteins selected when missing values were handled during the multivariate analysis (NIPALS) contained large numbers of missing values (Figure 4.7). The score plots and explained variances do not differ significantly for the BPCA and KNN methods (Figure 4.8). But when missing data was handled during the multivariate analysis (NIPALS), the variance within each group seems to be artificially reduced (Figure 4.8) which was

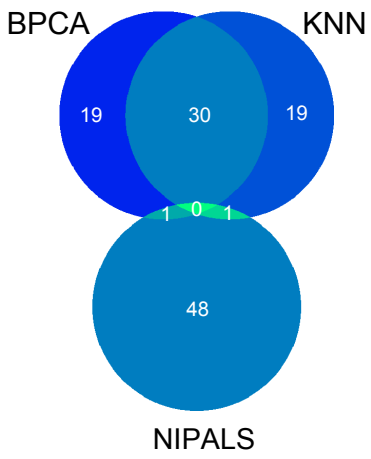

Figure 4.6: *Venn diagram showing the overlap of the selected proteins through PLS-DA and VIP50* for the incomplete DIGE dataset. VIP50* is defined as the 50 most important proteins selected by PLS-DA and VIP analysis.*

not observed with the 'complete DIGE' dataset (Figure 4.3). By handling missing data during the multivariate analysis or prior application of BPCA and KNN, PLS-DA was able to explain 83 %, 86 % and 84 % of the total variance when only the 50 most important proteins were kept although the final selection of these proteins clearly differed (Figure 4.6).

4.3.4 Impact of missing values handling techniques on VIP selection using classical dyes data

According to the Shapiro and Wilk test, approximately 5 % of the spots failed normality. Applying different transformations did not reduce this percentage but mainly stabilized the variances (data not shown). The log transformation improved homoscedasticity since the standard deviation was no longer correlated with the mean percentage spot volume. Thus, the log transformation was applied for further processing. On average the fifty

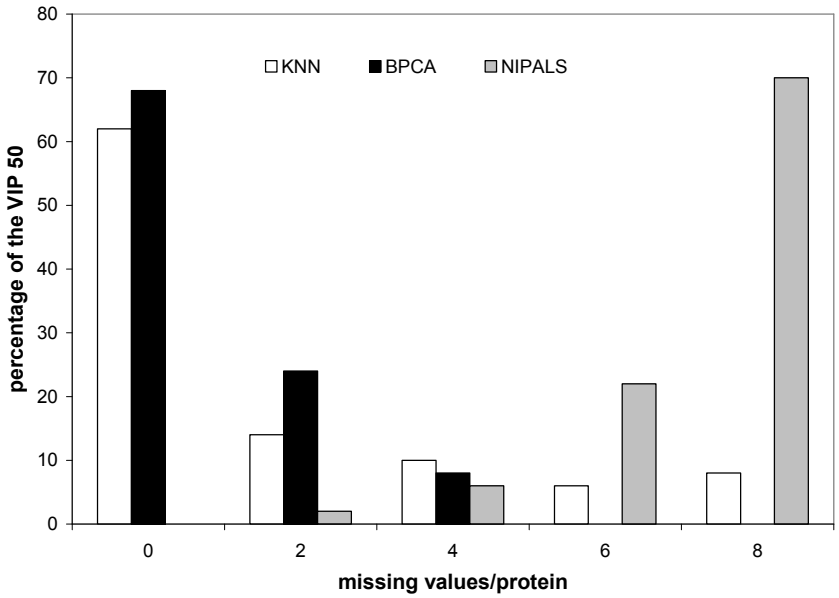

Figure 4.7: *Percentage proteins from the VIP50* of the original incomplete DIGE dataset containing 0-8 missing values depending on the missing value handling technique applied. VIP50* is defined as the 50 most important proteins selected by PLS-DA and VIP analysis. The maximum number of missing values in this dataset would be 10 out of 12 because of the DIGE set-up (three dye approach).*

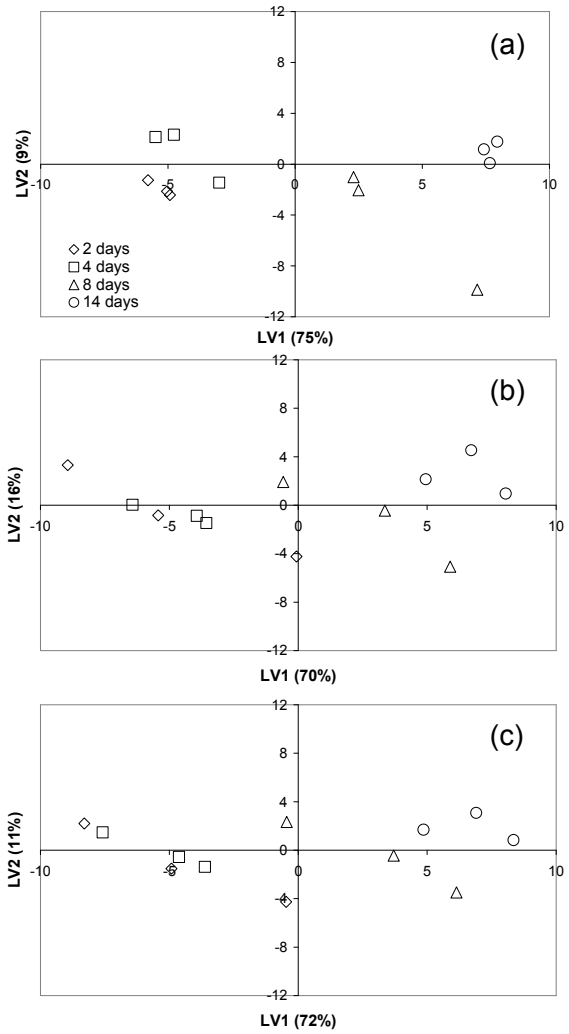

Figure 4.8: *Score plots (PLS) after the VIP50* procedure for the incomplete dataset (a) missing values handled during the calculations NIPALS, (b) BPCA imputed, and (c) KNN imputed. VIP50* is defined as the 50 most important proteins selected by PLS-DA and VIP analysis.*

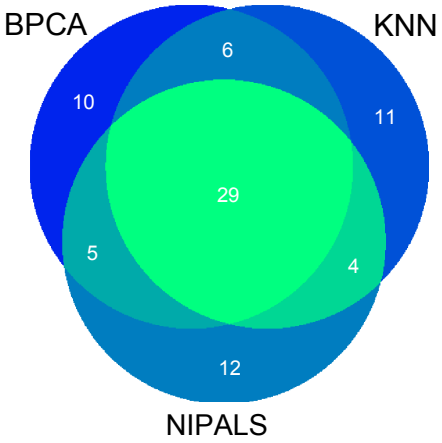

Figure 4.9: *Venn diagram showing the overlap of the selected proteins through PLS-DA and VIP50* when using the different missing value handling techniques for the incomplete classical dyes dataset. VIP50* is defined as the 50 most important proteins selected by PLS-DA and VIP analysis.*

selected proteins obtained by handling the missing values during the multivariate analysis (NIPALS) contained on average 8 missing values out of 24 values while after prior application of BPCA and KNN the fifty selected proteins contained only 6 missing values (Figure 4.10). In addition, the score plots obtained after the treatment of missing values and the final selection of the 50 most important proteins according to the VIP procedure and amount of explained variance are shown in Figure 4.11.

4.4 Discussion

Missing values are often present in classical stained and DIGE gels and must be treated appropriately. In general, less intense spots are more susceptible to be missing; nonetheless, these proteins might represent an important class responsible for regulation and signaling (Wood *et al.*, 2004; Krogh *et al.*,

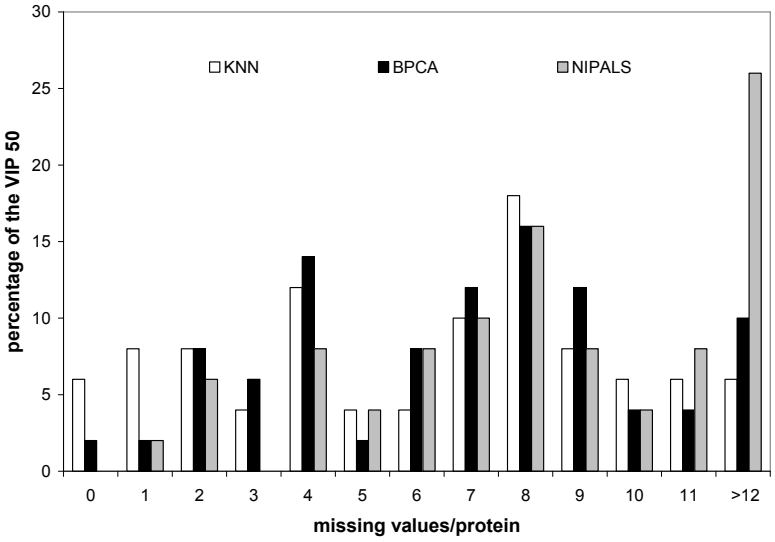

Figure 4.10: *Percentage proteins from the VIP50* containing 0 to more than 12 missing values depending on the missing value handling strategy applied for the incomplete classical dyes dataset. VIP50* is defined as the 50 most important proteins selected by PLS-DA and VIP analysis. The maximum number of missing values in this dataset would be 23 out of 24.*

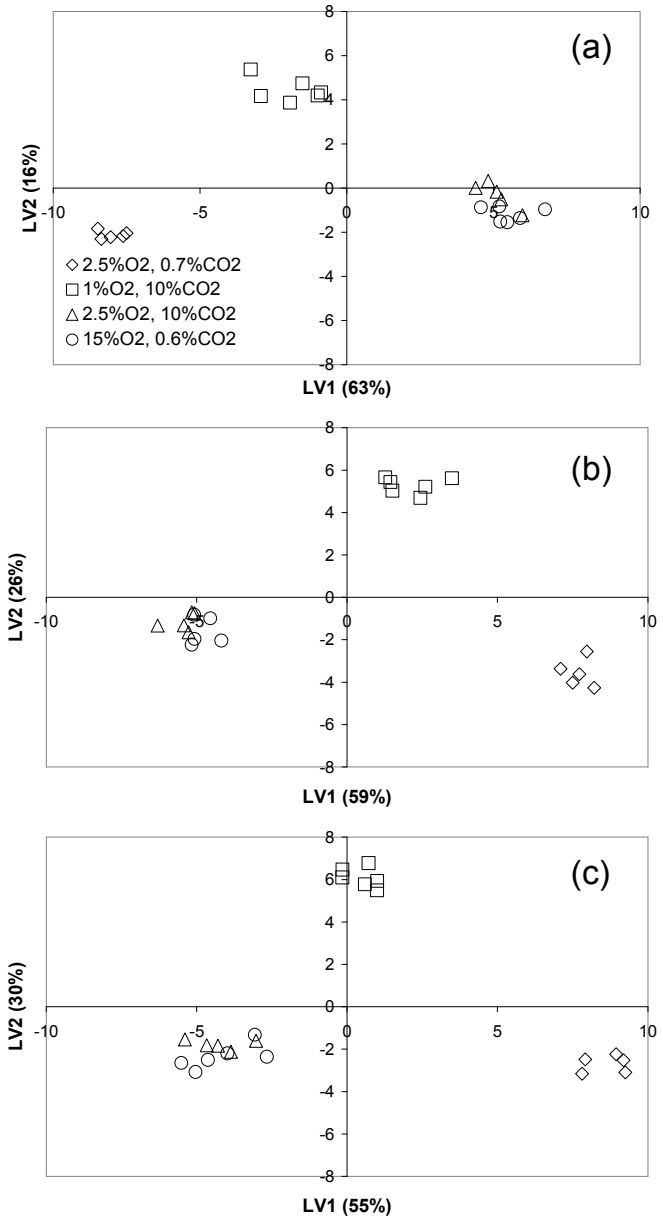

Figure 4.11: *Score plots (PLS-DA) after the VIP50* procedure for the classical dyes dataset (563 proteins containing 29 % missing values) (a) missing values ignored during the calculations, (b) BPCA imputed and (c) KNN imputed. VIP50* is defined as the 50 most important proteins selected by PLS-DA and VIP analysis.*

2007). The introduction of more and more sensitive mass spectrometric techniques, allow the identification of this low abundant class of proteins. In addition, currently many diagnostic studies rely on data mining techniques to assign samples to a certain group, thus the low abundant fraction proteins is essential (Karp *et al.*, 2008). Discarding such proteins would result in enormous loss of valuable biological information. The BPCA method showed to be the most consistent in terms of selecting most of the proteins that would have been selected if there were no missing values in the data while KNN tended to distort the structure of the original data (Figure 4.5). This was confirmed by the calculated correlation coefficients.

When evaluating the three methods to handle missing values on the original DIGE dataset (1462 variables, 24 % missing values), the fifty most important proteins selected with PLS-DA by handling the missing values during multivariate analysis was completely different from the results obtained after imputation by BPCA or KNN (Figure 4.6). An explanation for this is that missing values for proteomics data are not just the result of completely random events. This can be clearly seen in Figure 4.2 in which the distribution of missing values is plotted for the artificial dataset based on the 'complete DIGE' dataset and for the original incomplete DIGE dataset. By just discarding the missing dimensions, Eisen *et al.* (1998) found cluster of genes with many missing values when carrying out a cluster analysis on gene expression profiles. This finding was caused by ignoring the missing values which is similar to assume that the expression levels are the same within an experimental group. The presence of missing data in the multivariate analysis thus caused a bias towards the selection of proteins containing 60 % missing values (Figure 4.7). It has been shown that NIPALS tends to cause loss of robustness as the amount of missing values increases to 20% (Grung & Manne, 1998) compared to other algorithms such as BPCA (Oba *et al.*, 2003) or Multiple imputation (MI; Allison 2000). It is worth to mention here that not only the total amount of missing data in the dataset (24 %) is important but how it is distributed among the different proteins. For instance, in the 'incomplete DIGE' dataset, 27 % of the total number of proteins containing missing values showed to have missing values equal or

higher than 50 %.

From the original datasets false positives or negatives cannot be recognized, but imputation of missing values by BPCA is more appropriate than just handling them during the multivariate analysis. In contrast to the BPCA method that includes maximum likelihood estimation, the other two methods do not take into account the uncertainty associated with the prediction of the missing values. In addition the maximum likelihood algorithm does not assume the existence of missing values completely at random across all the observations but only at random within one or more subgroups (e.g., missing more among low abundant proteins than high abundant proteins, but within this low abundant category they are missing at random) which is an advantage. However, the total uncertainty associated with the prediction is not included and some other features such as the dependency of missing values on the characteristics (e.g., abundance, hydrophobicity, etc) of the proteins might be disregarded.

For the classical dyes dataset, the normality and equal variance assumptions were tested before performing the statistical analysis. The use of different transformations to stabilize the variance has been described before for proteomics data (Jung *et al.*, 2005; Hunt *et al.*, 2005; Tuomainen *et al.*, 2006; Grove *et al.*, 2006). For the classical dyes dataset it was shown that applying a log transformation is only needed to stabilize the variance but not to turn the data normal as 95 % of the data was already normally distributed regardless the transformation applied. For the different ways to handle missing data in the classical dyes dataset, 60 % homology in terms of the same selected 50 most important proteins remains (Figure 4.9). It has been shown in a previous study with gene expression data by Bras and Menezes (Bras & Menezes, 2006) that PLS based imputation methods performed better when the correlation structure of the data is weak (e.g., non time series experiments), as this experiment. However, with all the datasets tested (time series, non-time series and mixed experiments) BPCA in most of the cases outperformed the PLS based estimation methods. The fact that the three of them yielded more or less the same results is encouraging in terms of robustness for a biological interpretation of the data, given that

a choice has to be taken. Some examples of how the imputation methods are affecting the inclusion of particular proteins in the final VIP selection for the 'pear dataset' are given in Figure 4.12. All these proteins were visually inspected and confirmed as real spots. The figure shows both the imputed and original non-missing observations. In case of BPCA and KNN imputed data the VIP selection is based on the combination of the original non-missing observations with the respective imputed values. In case of the NIPALS dataset, the VIP selection is based on the original non-missing observations only. A typical protein included in all final VIP selections after each of the three methods used to deal with missing data (Figure 4.12 (A)) showed imputed values similar to the original non-missing spot volumes, suggesting accurate imputations. The protein selected by the three methods showed to be involved in a physiological disorder in pears which confirms what was found in our previous study (Pedreschi *et al.*, 2007). Whenever a protein was not selected after one missing values handling method but was selected by the remaining two missing values handling methods (Figure 4.12 (B-D)) this was due to the fact that the imputed values were clearly different from each other and the original non-missing values. However, one needs to be careful in interpreting data of individual proteins (an implicit univariate approach) as the selected proteins were identified within their original multivariate context. One possible argument, for the disagreement in performance of the NIPALS algorithm between this dataset and the 'incomplete DIGE dataset' might be related to the total percentage of individual proteins containing huge amounts of missing data. Even when this classical dyes dataset presents a higher total amount of missing values (29 %) than the 'incomplete DIGE' dataset (24 %), the classical dyes dataset only presented 13 % of the total proteins containing missing values with 50 % or more missing values. This feature leads to a better performance of the NIPALS algorithm for this particular dataset. It might be argued that a 'preliminary filtering' of proteins, in terms of the maximum amount of missing values allowed within each protein would be good practice but would still be subjective in where to set the maximum.

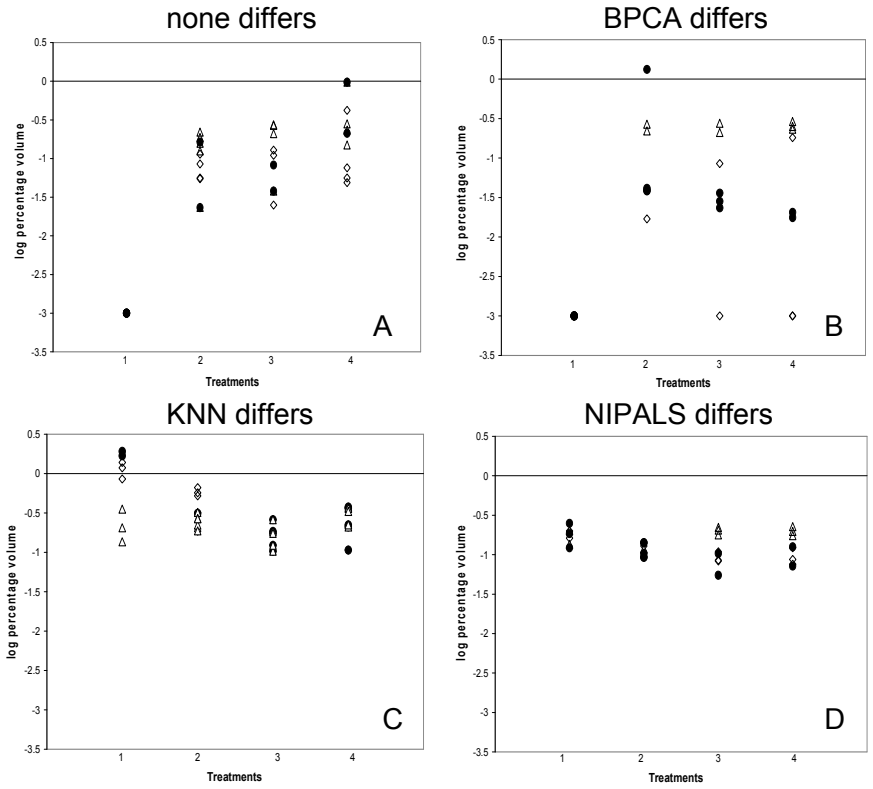

Figure 4.12: *Observed and imputed spot volume values for 4 selected proteins (plot A-D) from the 'classical dyes dataset'. Treatments (1-4) stand for the different storage conditions used. The open symbols represent the imputed values using either BPCA (diamonds) or KNN (triangles) imputation. The closed symbols (circles) represent the original non-missing observations making up the NIPALS dataset. Plot A, 'None differs' shows data for a protein (439) that was included in the VIP selection for all three missing values handling methods (either imputed during preprocessing, by BPCA or KNN imputation, or handled during the multivariate analysis through the NIPALS algorithm). The other plots (B-D) show data for proteins (respectively 401, 589 and 348) that were NOT selected after the missing values handling method referred to in the heading of the plot, but were selected by the other two missing values handling methods.*

4.5 Conclusions

Data pre-processing steps have a large impact on the final selection of the most important proteins when using multivariate statistical tools such as PLS and VIP and heavily rely on how missing values are treated. There is no absolute truth in terms of which is the most appropriate way to deal with missing data, however, from the ones studied, BPCA gave the best result. We recommend: (1) not to discard proteins containing missing values from the start, (2) estimate the amount of missing values in the dataset and within each individual protein, (3) based on the amount of missing values make a choice to impute missing values with an appropriate available method (we recommend BPCA in our case), (4) go back to the gels to check whether those selected proteins are real spots and not just artifacts or threshold values.

Chapter 5

Physiological implications of controlled atmosphere storage

5.1 Introduction

Fruits are often stored under controlled atmosphere (CA) conditions to extend their storage shelf life. For any given commodity, optimal oxygen and carbon dioxide concentrations must be determined in order to reduce respiration, ethylene production rates and action, delay ripening and senescence as well as to reduce the growth of pathogens (Kader, 2002). However, even when the applied external gas concentrations are relatively high, the oxygen concentration across plant tissues may fall because of the large diffusion gradients that are required to direct oxygen across the tissue at a rate fast enough to maintain the rate of oxygen consumption (Geigenberger, 2003; Ho *et al.*, 2006b, 2008). The oxygen level plays a key role in cellular function and metabolism. Low levels of oxygen, for instance, can trigger cellular damage leading to cell injury and death. There should be sufficient oxygen to maintain mitochondrial activity in order to provide enough ATP and essential substances for normal metabolism (Grinberg *et al.*, 1998).

The hypothesis on the formation of anoxic zones when Conference pears are stored under reduced oxygen concentrations has been extensively explained in a diffusion context (Lammertyn *et al.*, 2003b,a; Ho *et al.*, 2006b). Depending on the size of the fruit and ripening stage, toward the center of the fruit, local anaerobic conditions may exist (Ho, 2008) given certain CA conditions. But the physiological interpretation and consequences have so far been largely ignored. The use of proteomic approaches in the area of fruit and vegetable physiology has been increasing over the last years (Hjern *et al.*, 2006; Rocco *et al.*, 2006). However, in the area of postharvest physiology, proteomics is a fairly new approach. Research conducted on other commodities and physiological disorders has been limited to isolated assays trying to find explanations for the disorders under study (Burmeister & Dilley, 1995; Alferez *et al.*, 2005; Lurie & Crisosto, 2005; Sala *et al.*, 2005). Only a limited number of comprehensive studies aiming at understanding the physiology behind postharvest disorders have been reported so far (Casado-Vela *et al.*, 2005; Pedreschi *et al.*, 2007, 2009a).

In Chapter 3, focus was given to the characterization of browning related proteins (Pedreschi *et al.*, 2007) in stored pears. However, a specific focus on the controlled atmosphere conditions applied at an early stage, before browning is evident has not been conducted so far. In addition, as shown in Chapter 4, the gel-based proteomics approach demands extensive data preprocessing in order to apply correctly statistical tests and be able to draw sound biological conclusions. Thus, the objective of this Chapter is to study the effects of four different controlled atmosphere conditions on the protein levels to better understand the physiological effects of controlled atmosphere storage on Conference pears applying all data pre-processing steps discussed in Chapter 4. The results of this chapter were published in Pedreschi *et al.* (2008a)[4].

[4]Pedreschi et al. (2008). Physiological implications of controlled atmosphere storage of Conference pears (*Pyrus communis* L.): a proteomic approach. Postharvest Biology and Technology, 50, 110-116.

5.2 Materials and Methods

5.2.1 Fruit material

Pears (*Pyrus communis* L cv. Conference) were harvested in the orchard of the Centre for Fruit Culture in Rillaar (Belgium). The trees in this orchard are planted along the north-south axis and hence, with exception of the corner trees, have a west and east side. Homogeneous size pears were picked randomly from the east side at a height of 2 m. One batch of pears was picked at their commercial harvest date (September 8^{th}, 2004) as determined by the Flanders Centre of Postharvest Technology (Belgium). These fruit were submitted to pre-cooling in air at $-1\,^{\circ}$C for three weeks before applying controlled atmosphere conditions of either 2.5 % O_2 and 0.7 % CO_2 (commercial condition), 15 % O_2 and 0.6 % CO_2 (high O_2 condition) and 2.5 % O_2 and 10 % CO_2 (high CO_2 condition). Pears from these CA conditions did not present any visual symptom of disorder. A second batch of pears was picked two weeks after the commercial harvest date (September 22^{nd}, 2004) from the same trees and immediately stored under 1.0 % O_2 and 10 % CO_2 at $-1\,^{\circ}$C with no pre-cooling period to induce the browning disorder. After six months of CA storage, pears from the different conditions were sampled. Pears were cut perpendicularly to the stem-calyx axis at 5 cm from the bottom of a pear. Tissue samples were taken from the equatorial region excluding the skin and core. From all pears only sound (non-brown) tissue was sampled. Samples were immediately frozen in liquid nitrogen and kept at $-80\,^{\circ}$C until further analysis. Six independent biological replicates of tissue samples were prepared based on pooled tissue coming from 6 pears each. Harvest time and pre-cooling regime were not used as real independent treatment factors but were just part of the standard protocol to induce browning disorder in Conference pears. The aim was to use pears from the browning inducing conditions (that would definitely develop browning) as a reference for those fruit stored under the three other CA conditions where no visible browning was induced but where protein changes would be triggered to various degrees.

5.2.2 Protein extraction

Proteins were extracted with a modified phenol extraction, methanol ammonium acetate precipitation method optimized for pear tissue by Pedreschi *et al.* (2007) and detailed in Material and Methods of Chapter 3.

5.2.3 2-DE

The same protocol described in Material and Methods of Chapter 3 was used for protein separation. IPG strips of a 5-8 pH range were used.

5.2.4 Protein visualization and image analysis

Proteins were visualized by silver staining (Blum *et al.*, 1987) or by colloidal Coomassie Brilliant Blue (CBB) G-250 (Neuhoff *et al.*, 1988). Image analysis was performed with the Image Master 2-D platinum software 6.0 (Amersham Biosciences). Spots were detected without spot editing and quantified as percentage volume. Silver stained gels were used as analytical gels and Coomassie stained gels as preparative gels for further LC-ESI-MS/MS protein identification.

5.2.5 Statistical data analysis

The 2-DE data calculated as spot percentage volume were pre-processed before multivariate statistical analysis. Gels were matched to a reference gel composed of equal amount of samples from all the different treatments. Missing values were imputed with the BPCA (Bayesian Principal Component Analysis) method described by Oba *et al.* (2003) as detailed in Chapter 4. Data was log transformed to stabilize variance and outlying gels were removed with the 95 % Hotelling's T^2 limit (Johnson & Wichern, 1998). Data was also mean-centered and variables weighed to give them equal (unity) variance. Partial least square discriminant analysis (PLS-DA) was performed to cluster the individual gels according to similar protein

expression profiles. The correlation loading plots were used to identify the proteins which significantly contributed to class discrimination. The variable important plot (VIP) was used as a formal tool (Karp *et al.*, 2005b) based on the correlation loadings to identify the most important proteins describing the differences in protein expression profiles among the treatments. After selection of 100 proteins with PLS-DA and VIP procedure, these proteins were checked for downstream processing. Only successfully identified proteins were used to re-build a PLS-DA model. The details of the multivariate statistical procedure can be found in Chapter 3 (Pedreschi *et al.*, 2007). The final selected proteins were submitted to a confirmatory one-way ANOVA ($p < 0.05$) and treatment differences were assessed by a Tukey test ($p < 0.05$). Missing value imputation and log transformation was implemented in Matlab 7.0 (The Mathworks, Inc. 2007). PLS-DA was performed using The Unscrambler Version 9.0 (CAMO A/S, Trondheim, Norway) and the univariate statistical analysis using SPSS version 16.0 (Illinois, USA).

5.2.6 Protein identification

Proteins were first trypsin digested. The generated peptides were separated and identified through LC-ESI-MS/MS. The detailed approach described in Materials and Methods of Chapter 3 was used. Protein identification was carried out in collaboration with the Biomedical Research Institute, Hasselt University and School of Life Sciences, Transnationale Universiteit Limburg, Diepenbeek, Belgium.

5.3 Results

The 2-DE images for the different treatments can be observed in Figure 5.1. PCA analysis was first carried out to explore for a possible discrimination among the different treatments (Figure 5.2(a)) and it showed that the different conditions applied do have an influence on the protein levels of stored Conference pears. Since the applied treatments are known, PLS-DA analysis was carried out to enhance the discrimination among the different

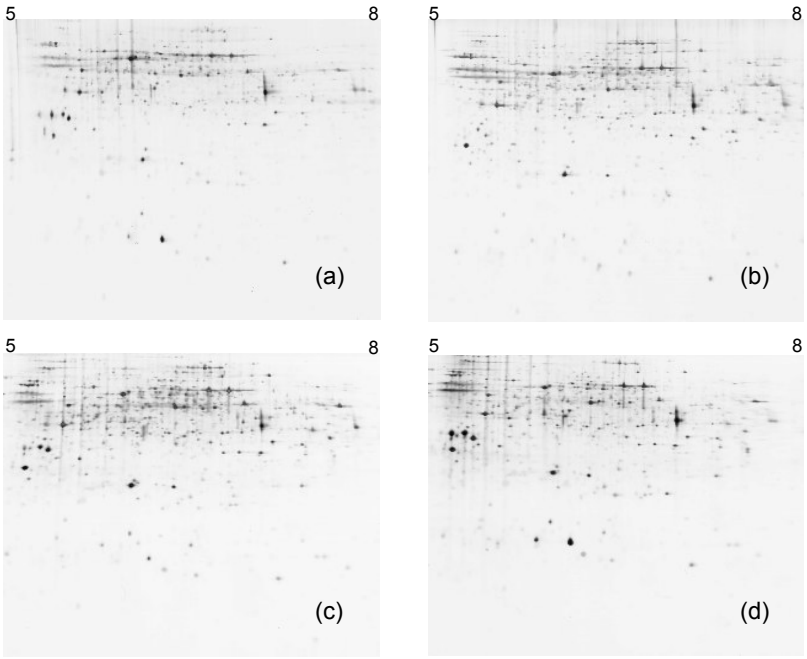

Figure 5.1: *2-DE images for the different CA conditions: (a) commercial condition, (b) browning inducing condition, (c) high CO_2 condition and (d) high O_2 condition. Separation of proteins was realized on a 24 cm strip, 5-8 pI.*

treatments (Figure 5.2(b)) and to correlate the different proteins with the applied treatments. The commercial CA condition was well separated from the high O_2 and high CO_2 CA conditions along LV1 and from the browning inducing condition along LV2. The total percentage explained variation corresponded to 17 % and 13 %, respectively. The oxygen concentration seems to be critical and might be favored also by an elevated carbon dioxide concentration.

In order to select the most important proteins involved in class discrimination, PLS-DA was rerun with only the 100 proteins selected through the VIP procedure. A better model was obtained in which 48 % and 24 % of the total X variance was explained by LV1 and LV2 respectively (data not shown). These 100 proteins were checked as being real spots and also for downstream processing (LC-ESI/MS-MS identification). Only 17 proteins remained and were successfully identified. With these proteins, a new PLS-DA model was obtained that explained 47 % and 22 % of the total

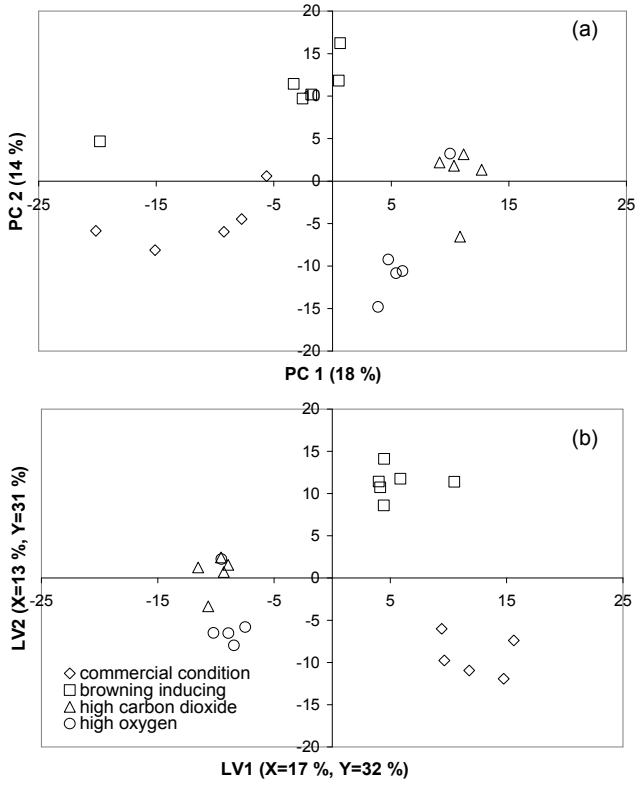

Figure 5.2: *PCA (a) and PLS-DA (b) score plots to assess for treatment differences considering all proteins.*

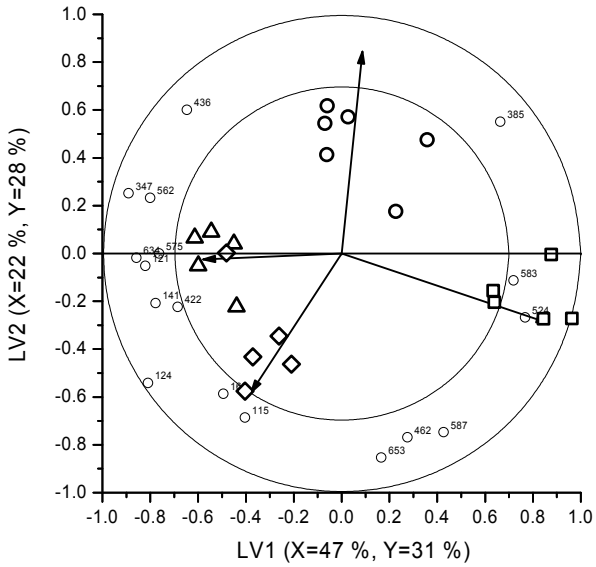

Figure 5.3: *Biplot obtained after filtering and keeping the successfully identified proteins. Seventeen proteins were kept and the model was re-built. Squares stand for the commercial condition, circles stand for the browning inducing condition, triangles stand for the high CO_2 condition and diamonds stand for the high O_2 condition. The percentage explained X and Y variance for the first two latent variables (LV1 and LV2) are shown on the axes.*

X variance and 31 % and 28 % of the Y variance with the first two latent variables (Figure 5.3). The commercial CA condition was clearly separated from the browning inducing condition and the high O_2 and high CO_2 conditions. One-way ANOVA confirmed that all seventeen selected proteins were significantly different between the treatments at a $p < 0.05$.

Energy metabolism related proteins were clearly important in distinguishing among the four treatments (Table 5.1). The glycolytic enzyme triosephosphate isomerase was up-regulated in all treatments as compared to the commercial condition. NAD dependent malate dehydrogenase (spot 583) was down-regulated for all the treatments as compared to the commercial treatment (Table 5.1). In addition, adenosine kinase was up-regulated in all treatments as compared to the commercial condition. Cobalamine independent methionine synthase isoforms were up-regulated for the high CO_2 and high O_2 conditions as compared to the commercial and browning inducing

conditions. However, the specific role of the oxygen and carbon dioxide concentration was not evident. The cell division protein 48 remained unchanged except for the high O_2 conditions, presumably as a direct response to the elevated oxygen concentration as compared to the other conditions. Defense related enzymes, such as the molecular chaperone HSP70 was mainly down-regulated under the browning inducing condition compared to the rest of the evaluated conditions. Selenium binding protein was also up-regulated in all conditions as compared to the commercial condition. 40S ribosomal SA protein seems to be affected by the carbon dioxide concentration. Thus, it was statistically similar for the browning inducing condition and high CO_2 condition compared to the remaining conditions characterized by a low carbon dioxide concentration. The polygalacturonase inhibiting protein was up regulated for all conditions except the optimal commercial condition. The major allergen Pyrc 1 was clearly down regulated in the conditions of low oxygen and high carbon dioxide (Table 5.1). Ethylene biosynthesis related proteins were clearly dependent on the oxygen and carbon dioxide concentrations. Thus, ACC oxidase was down regulated under very low oxygen and high carbon dioxide concentrations such is the case of the browning inducing condition.

Table 5.1: *Protein abundance expressed as fold change of a treatment compared to the commercial condition. The 17 selected proteins involved in treatment differences were identified by means of LC-ESI-MS/MS and cross species matching with sequences in databases. Different letters within a row stand for significant differences ($p < 0.05$) based on a Tukey test performed with the log transformed data. Results are expressed as fold change taking as reference (1) the commercial condition. Values greater than one indicate up-regulation while values smaller than one indicate down-regulation. Proteins have been group according to biological function. Significant non-redundant peptides[1] validated by Mascot/Sequest. The Mascot validation criteria used: expect value ≤ 0.05. The Sequest validation criteria used: cross correlation xC for 1^+, 2^+ and 3^+ charged ions ≥ 1.8, 2.5 and 3.5, respectively.*

Protein name	Spot number	Protein ID	Treatment				Peptides[1]
			Commercial	Browning inducing	High CO_2	High O_2	
I. Energy metabolism							
UDP glucose phosphorylase	385	Mdo.584	$(1.00)^{a}$	$(1.10)^{a}$	$(0.34)^{b}$	$(0.26)^{b}$	9
Triosephosphate isomerase	634	Mdo.184	$(1.00)^{a}$	$(1.60)^{b}$	$(2.40)^{c}$	$(2.14)^{b,c}$	5
NAD-dependent malate dehydrogenase	575	Mdo.20392	$(1.00)^{a}$	$(1.80)^{b}$	$(2.15)^{b}$	$(2.53)^{a}$	4
NAD-dependent malate dehydrogenase	583	Mdo.20392	$(1.00)^{a}$	$(0.26)^{b}$	$(0.004)^{c}$	$(0.36)^{b}$	2
Adenosine kinase	565	Mdo.19310	$(1.00)^{a}$	$(28.90)^{b}$	$(46.80)^{b}$	$(49.90)^{b}$	5
II. Defense related							
HSP70	160	Mdo.1243	$(1.00)^{a,b}$	$(0.68)^{a}$	$(2.24)^{b}$	$(2.34)^{b}$	9
Selenium binding protein	347	Mdo.3762	$(1.00)^{a}$	$(48.90)^{b}$	$(61.60)^{b}$	$(60.20)^{b}$	2
4-Hydroxyphenyl pyruvate dehydrogenase	422	Mdo.3417	$(1.00)^{a}$	$(1.17)^{a,b}$	$(2.24)^{c}$	$(1.83)^{b,c}$	2
Polygalacturonase inhibiting protein	436	Mdo.1066	$(1.00)^{a}$	$(6.88)^{c}$	$(5.46)^{b,c}$	$(3.07)^{b}$	3
Major allergen Pyrc1	653	Mdo.3014	$(1.00)^{a}$	$(0.15)^{b}$	$(0.42)^{a,b}$	$(1.00)^{a}$	5
III. Protein synthesis							
Cell division cycle protein 48	115	Mdo.253	$(1.00)^{a}$	$(0.75)^{a}$	$(1.45)^{a}$	$(3.10)^{b}$	7
Cobalamine-independent methionine synthase	121	Mdo.18819	$(1.00)^{a}$	$(1.70)^{a,b}$	$(3.89)^{b}$	$(2.95)^{b}$	10
Cobalamine-independent methionine synthase	124	Mdo.18819	$(1.00)^{a}$	$(1.00)^{a}$	$(56.20)^{b}$	$(60.00)^{b}$	3
Glycyl tRNA transferase	141	Mdo.9548	$(1.00)^{a}$	$(1.91)^{a,b}$	$(4.68)^{b}$	$(4.47)^{b}$	2
40S ribosomal protein SA	462	Mdo.5965	$(1.00)^{a}$	$(0.33)^{b}$	$(0.52)^{a,b}$	$(1.00)^{a}$	2
IV. Ethylene biosynthesis							
ACC oxidase	524	Mdo.3349	$(1.00)^{a}$	$(0.59)^{a}$	$(0.45)^{a}$	$(0.55)^{a}$	4
ACC oxidase	587	Mdo.3349	$(1.00)^{a}$	$(0.18)^{a}$	$(0.25)^{a,b}$	$(0.74)^{a,b}$	3

5.4 Discussion

Our results show that profound responses are imposed on primary metabolic pathways such as respiration during CA storage of Conference pears. Cells respond to the storage conditions by implementing survival strategies to adapt to the stress conditions. In this context, we could clearly show that under browning inducing conditions and high O_2 and high CO_2 conditions, there was an up-regulation of the key enzyme: triosephosphate isomerase as compared to the commercial condition. Since triosephosphate isomerase is at the crossroads of glycolysis, gluconeogenesis and the oxidative and reductive pentose phosphate pathways (Dennis *et al.*, 1998), it plays a critical role in metabolic flexibility towards anaplerotic pathways. These pathways provide molecules needed to carry on basic metabolism as well as defensive mechanisms against stresses. This is the case for NADPH which is not only needed for biosynthetic processes such as membrane repair damage but is also indispensable for the ascorbate glutathione antioxidant system. In a previous metabolomic study on core breakdown in 'Conference' pears (Pedreschi *et al.*, 2009a), some indication of pentose phosphate pathway activation was found as an alternative source for obtaining reducing equivalents. Our results indicate that respiratory pathways at least at the level of glycolysis and TCA cycle are clearly being affected by the conditions applied. Previous studies on pear cells, showed that exposure to for instance elevated carbon dioxide concentrations results in down regulation of glycolytic enzymes such as phosphofructokinase and pyrophosphate fructokinase with little change in other glycolytic enzymes (Kerbel *et al.*, 1990). The fact that adenosine kinase was also up regulated in all conditions as compared to the commercial condition might be related to the increased energy demand to surpass the stressful storage conditions (Dobrota, 2006).

Respiratory pathways provide carbon skeletons for aminoacid synthesis which are the building blocks of proteins. From the obtained proteomics data in this study, no clear effects about global protein synthesis alteration are observed. The activation of the two isoforms of cobalamine independent methionine synthase that were up-regulated for the high CO_2 and high O_2

conditions as compared to the commercial and browning inducing conditions seems to be related to the activation of methionine related pathways for polyamine production and ethylene biosynthesis. Methionine, as mentioned previously, can be further converted into polyamines. Polyamines, in turn are involved in maintaining ion balance, in chromatin protection, and in decreasing the generation of reactive oxygen species. An increase in the polyamine putrescine was observed on pears submitted to browning inducing conditions (Pedreschi *et al.*, 2009a) even though this brown tissue is expected to be the result of a collapsed antioxidant system and an over production of ROS. In addition to putrescine increase, several different defense mechanisms exist to scavenge ROS such as the ascorbate-glutathione cycle, tocopherols, polyphenols, among others. Thus, the higher increase of putrescine in brown tissue compared to sound tissue reveals that several defense mechanisms are simultaneously taking place and that putrescine alone is not enough to fight against the high ROS generation. In a previous study on banana meristems, increased levels were observed and positively correlated with the survival rate after cryopreservation (Ramon *et al.*, 2002). Increased levels of methionine synthase have also been reported in barley leaves submitted to salt stress (Narita *et al.*, 2004). Due to the storage stress, several defense related proteins were affected. For instance, the chaperone molecule HSP70 essential for maintenance and restoration of protein homeostasis during stress (Baniwal et al., 2004) was down-regulated as the oxygen concentration diminished. This might be due to the ATP depletion as low oxygen levels generally results in a decrease in the pool of chaperones as these are highly dependent on the levels of phosphorylation (Martinus *et al.*, 1995). The 40S ribosomal protein SA has been speculated to be involved in the formation of the translation initiation complex essential for protein synthesis. During yeast fission, this protein was shown to be essential for cell viability and its genetic depletion caused a complete inhibition of the 40S ribosomal subunit production (Perreault *et al.*, 2008). For Conference pears, the 40S ribosomal protein SA was down-regulated under elevated CO_2 conditions.

Finally, the major allergen Pyrc 1 was clearly down-regulated by a high carbon dioxide content and very low oxygen concentration. In a previous

study with different cultivars of apples, controlled atmosphere storage (3 °C, 2.5 % O_2, 1 % CO_2) was shown to reduce the allergenicity by 15 % compared with storage in air at 2 °C (Bolhaar *et al.*, 2005). At least in apples, a decrease of allergenicity has been related to an oxidative reaction. Formation of *o*-quinones formed by oxidation of polyphenols by the enzyme polyphenol oxidase (PPO) was suggested to modify the tertiary structure of allergens and participate in the cross linking of proteins, thus reducing allergenicity (Bolhaar *et al.*, 2005; Garcia *et al.*, 2007). Our results showed that controlled atmosphere storage conditions with elevated carbon dioxide concentrations reduce the presence of the allergen which is involved in defense mechanisms and shows sequence homology with the ones found in apple.

Figure 5.4 summarizes under browning inducing conditions a postulated series of events occurring during storage. Further studies should focus on studying the direct effect of gas conditions on protein level changes. Under browning inducing conditions, the respiration pathway is impaired involving up regulation of triosephosphate isomerase suggesting activation of pathways like the pentose phosphate to maximize the production of reducing molecules needed for membrane damage repair and antioxidant system. TCA enzymes were also regulated. The synthesis of enzymes involved in methionine related pathways needed for the synthesis of ethylene related proteins are reduced. As a consequence of the alteration of primary metabolism pathways, defensive mechanisms are induced as well.

5.5 Conclusions

The influence of storage conditions on the protein profiles was demonstrated. Sub-optimal storage conditions impaired respiration pathways and activated defensive mechanisms in order to surpass the stressful conditions. Proteins clearly down regulated by too low oxygen or too high carbon dioxide concentration are ACC oxidase and the major allergen Pyrc 1 which in turn is related to defensive mechanisms. It should be remarked that even though proteins are not statistically significant in terms of abundance change (an

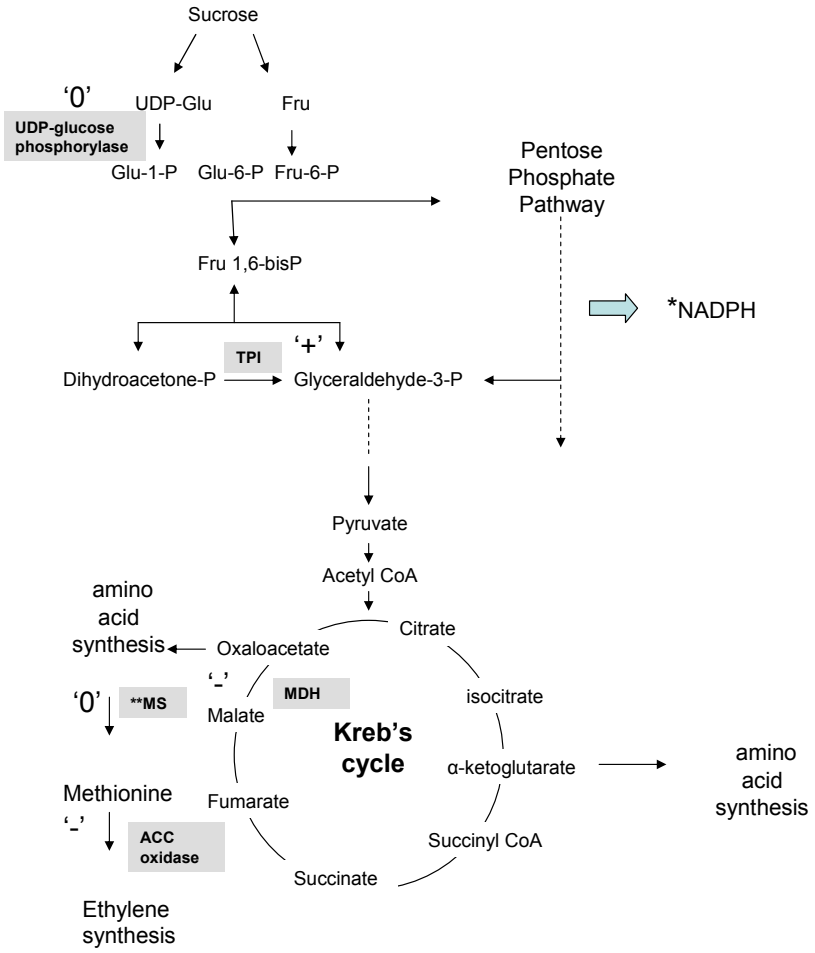

Figure 5.4: *A postulated overview of the different metabolic pathways and processes affected by browning inducing conditions (1 % O_2, 10 % CO_2, no-precooling and fruit from late harvest). *NADPH can be further utilized by the ascorbate-glutathione cycle. TPI: triosephosphate isomerase, MDH: malate dehydrogenase, MS: cobalamine independent methionine synthase. The '+', '-' and '0' synmbols stand for up-regulation , down-regulation or no change as compared to the commercial condition.*

implied univariate approach), small changes all added up might become important. These small changes are in fact accounted when using a multivariate statistical approach. In addition the presence of certain amount of protein might not be correlated to the activity of the protein. We believe, this proteomics approach is not only powerful in terms of providing an overview of what is happening at a fundamental level but in terms of selecting for example appropriate biomarkers (proteins) as to be able to detect metabolic disorders at an early stage thus reducing economical losses. Further studies should focus on isolating the effects of the different factors such as gas concentrations, harvest time, pre-cooling effect and other pre-harvest factors known to be involved in metabolic disturbances. In addition, more simplified experimental systems in which for instance the barrier to gas transport is reduced and the short term effect is evaluated should be used (See Chapter 6).

Chapter 6

Short term changes of protein profile in pear tissue during storage

6.1 Introduction

Pre-harvest and post-harvest factors (Franck *et al.*, 2007) influence the appearance of core breakdown disorder in 'Conference' pears. However, it is widely accepted that the CA composition triggers the onset of the disorder. The critical factor is the internal concentration of oxygen within the cells and not really the environmental concentration of oxygen. This internal concentration is affected by the crop resistance to oxygen diffusion, rate of utilization and the differential partial pressure between the crop exterior and interior (Kays, 1991). Kays (1991) classifies low oxygen stress in three categories: (i) severe stress when anaerobic conditions exist, significant losses are observed and can eventually end up in cell death, (ii) moderate stress when the oxygen availability is above that leading to anaerobiosis, (iii) mild oxygen stress that does not result in injury and can be used to extend shelf life and maintain quality of the products (Kays, 1991). Depending on the duration of the stress, the plant will be capable of recycling the metabolites

formed such as ethanol and pyruvate because these reactions are reversible when returning to aerobic conditions. Undesirable changes are manifested in the aroma and flavor profiles. The way the product will respond to carbon dioxide is dependent on the nature of the product, duration of the exposure, concentrations within the tissue and the internal oxygen concentration (Kays, 1991). High carbon dioxide levels favor the incidence of internal and external physiological disorders in different commodities.

The use of gel-based proteomics requires a huge time investment in protocol optimization and data processing. We have shown in the previous chapters the evolution followed in terms of understanding core breakdown by using proteomics means together with improving the approach followed to analyze the data. As we proceed, other challenges appear. The use of classical dyes used in the previous chapters, even though still valid if the number of replicates is high enough, have the problem of high technical variation. Differential in gel electrophoresis (DIGE) reduces the gel to gel variation allowing the multiplexing of samples and the introduction of an internal standard.

With the introduction of more sensitive MS equipment, spot overlap becomes evident in 2-DE protein separation (Campostrini *et al.*, 2005; Hunsucker & Duncan, 2006). Within a spot with several protein identifications, the one involved in the regulation could be incorrectly assigned. More frequently the information is discarded, as it is unclear which protein species has the highlighted fold change or even whether the fold change is a composite for different proteins each with differing relative expression levels. Different approaches to counteract the problem involve the use of narrow range IPG strips, sample fractionation methods, different sample preparation conditions and modification of conditions during 2-DE (Hunsucker & Duncan, 2006). The use of spectral counts can partially circumvent this problem. Spectral counting consists of counting the total number of spectra representing identified peptides for a certain protein (Liu *et al.*, 2004; Deutsch *et al.*, 2008) and has been shown to correlate closely with protein concentration (Liu *et al.*, 2004).

This chapter aims to understand the effect of the CA composition on the metabolism of 'Conference' pear, regardless of storage time and actual browning development. Thus, the current chapter focuses on protein expression changes in thin pear slices submitted to extreme gas compositions using 2-DE DIGE combined with a rigorous statistical analysis of the data preceding LC-ESI MS/MS identification of the relevant proteins. The challenge of spot overlapping described before will also be faced and tackled. The results of this chapter have been published in Pedreschi *et al.* (2009c)[5].

6.2 Materials and methods

6.2.1 Plant Material

Pears (*Pyrus communis* cv. Conference) were harvested in the orchard of the Centre for Fruit Culture in Rillaar (Belgium). All pears were stored under optimal controlled atmosphere (CA) conditions. Pears were picked at the commercial harvest date on 16/9/2006 as determined by the Flanders Centre of Postharvest Technology (Belgium). The fruit were submitted to pre-cooling in air at $-1\,^{\circ}C$ for three weeks before applying optimal controlled atmosphere conditions of 2.5 % O_2 and 0.7 % CO_2 following commercial protocols. Pears were stored for 8 months under these commercial storage conditions.

6.2.2 Sample preparation

Pears taken from the commercial storage conditions were cut perpendicularly to the stem-calyx axis at 5 cm from the calyx. Thin slices of tissue samples (1.5 mm tick and 2.5 cm diameter) were taken from the equatorial region excluding the skin and core. For each condition, 1.5 L respiration jars were filled with 150 g of pear tissue slices evenly distributed in layers. The 1.5 L jars were connected in series. CA conditions of 20 % O_2 and 80

[5]Pedreschi et al. (2009). Proteomics-gel based approach to study metabolic changes in pear tissue during storage. Journal of Agricultural and Food Chemistry, submitted.

% N_2 (air) and 10 % CO_2 and 90 % N_2 (anoxia) were generated to flush the jars at 10 L/h. Before entering the jars, the CA mixtures were humidified. Gas composition was monitored using a PBI Dansensor, Model Chekmate O_2 (Zr) CO_2-100 % (Denmark). The control consisted of pear slices coming directly from the commercial storage and they were immediately frozen in liquid nitrogen until further analysis. Jars were incubated at 1 °C for 5 days. In all cases four independent biological replicates composed of slices from ten sampled pears were obtained for each condition (control, air and anoxia).

6.2.3 Protein extraction and CyDye labelling

Proteins were extracted with the phenol extraction following by precipitation in 100 mM ammonium acetate in methanol as detailed in Chapter 3. The obtained protein pellets were stored at −80 °C until analysis. Total protein concentrations were determined using the Bio-Rad DC protein assay following the manufacturer's guidelines (Bio-Rad, UK). The protein pellets were rehydrated in DIGE buffer (7 M urea, 2 M thiourea, 4 % CHAPS, 5 mM magnesium acetate, 10 mM Tris pH 8.0). The pH was adjusted to 8.5 as required using 50 mM NaOH. Proteins were labeled using the fluorescent Cyanine dyes developed for DIGE (GE Healthcare, USA) following the manufacturer's guidelines. Thus, 50 μg of proteins were labeled with 400 pmol of amine reactive Cyanine dyes dissolved in fresh anhydrous dimethyl formamide. The two-dye approach recommended by Karp *et al.* (2007) was used. Four biological replicates per treatment (control, air, anoxia) were independently labeled with Cye 3. Cye 5 was used to label the internal standard composed of equal amounts of all samples. The labeling reaction was incubated in the dark for 30 min and quenched with 10 nmol lysine. An equal volume of 2x sample buffer (7 M urea, 2 M thiourea, 4 % CHAPS, 20 mg/mL DTT, and 2 % Pharmalites 3-10) was added to each of the labeled samples. Rehydration buffer (7 M urea, 2 M thiourea, 4 % CHAPS, 10 mg/mL and 1 % Pharmalites 3-10) was added to make up a final volume of 450 μL prior to IEF. Both the internal standard labeled with Cye 5 and the

sample labeled with Cye 3 were mixed and run together in the same strip.

6.2.4 2-DE

Linear IPG strips (24 cm long) of pH 4-7 (GE Healthcare, USA) were rehydrated with the CyeDye labeled samples for 10 h at 20 °C at 20 V using the IPGphor II apparatus (GE Healthcare, USA). The following steps of the IEF included: 1 h at 500 V, 1 h at 1000 V followed by 8.2 h at 8000 V. The last step consisted of 24000 Vh at 8000 V. After IEF completion, strips were equilibrated individually for 15 min in 10 mL equilibration buffer (8 M urea, 30 % glycerol, 1 % SDS, 100 mM Tris-HCl pH 6.8) containing 2 % w/v DTT and subsequently for 15 min in 10 mL equilibration buffer containing 2.5 % iodoacetamide. Second dimension separation was performed in an Ettan DALT Twelve system (GE Healthcare, USA) with lab cast 1.0 mm SDS polyacrylamide gels (12.5 %). Gels were run overnight at 1.5 W/gel.

6.2.5 Protein visualization

Labeled proteins were visualized in a Typhoon^{TM} 9410 imager (GE Healthcare, USA). Cye 3 images were scanned using a 532 nm laser and 580 nm band pass (BP) emission filter. Cye 5 images were scanned using a 633 nm laser and a 670 nm BP30 emission filter. Gels were scanned at a 100 μm resolution. The PMT was set to ensure maximum pixel intensity between 40000 and 60000 pixels. Gel analysis was performed using Progenesis SameSpots (Nonlinear Dynamics, UK), a 2-DE analysis software package.

6.2.6 Data analysis

A complete statistical analysis was carried out using both univariate and multivariate statistics on the log standardized abundance where the log standardized abundance is the Cye 3 sample spot volume divided by the Cye 5 standard sample spot volume after ratiometric normalization. By using both approaches independently, not only absolute changes in terms of

protein expression but also correlations and concerted changes in expression can be assessed.

6.2.6.1 Univariate statistics

One-way analysis of variance (ANOVA) was carried out at $p < 0.01$ and $p < 0.05$ respectively in order to assess for absolute protein changes among the different treatments. The false discovery rate (FDR) was assessed by calculating q-values using the p-values (Storey & Tibshirani, 2003) consequently both a p-value and a q-value are calculated for each spot. The q-value is a measure of significance in terms of FDR. Since the q-value approach relies on the use of the correct statistical test for the experimental design (Karp *et al.*, 2007), the two-dye DIGE scheme was used. By calculating the q-values, the user has control over the FDR as differing p-value thresholds can be chosen for differing levels of false call rates. The FDR estimates how many from the spots declared to be significant, are expected not to be significant at all. Differentially expressed proteins were manually checked as being proper spots before submitting them for protein identification. Pair-wise comparisons were carried out by using a Tukey test ($p < 0.05$) in SPSS version 15 (Chicago, Illinois, USA) only on those proteins declared to be significant by one-way ANOVA. A Tukey test was chosen as post-hoc test to find which treatment mean significantly differed from one another after the ANOVA model was significant. This test is very conservative and corrects for the multiple testing problem being more suitable for multiple comparisons than carrying out several independent t-tests.

6.2.6.2 Multivariate statistics

Data pre-processing steps included mean centering and standardizing the variance. Principal component analysis (PCA), an unsupervised technique was carried out as a first exploration of the data and to identify possible outlying gels through the 95 % Hotelling's T^2 limit (Johnson & Wichern, 1998). Partial least squares discriminant analysis (PLS-DA) analysis, a su-

pervised technique, was carried out to sharpen the discrimination among the treatments according to similar protein expression profiles as detailed in Pedreschi *et al.* (2008b). No data imputation was used because the image analysis software utilized generates complete datasets. The variable importance plot (VIP) was used as formal tool based on the correlation loadings to identify the most relevant proteins involved in class distinction. Further details can be found in Pedreschi *et al.* (2007, 2008b). The VIP procedure was first run to select the 120 most important proteins. Selected proteins were manually checked as being real spots. After checking for real spots, the VIP procedure was run again on the remaining proteins. Only proteins successfully identified by LC-ESI MS/MS identified which were independently selected from the univariate ($p < 0.05$ and $q < 0.1$) and multivariate statistical analysis were used to build the final PLS-DA model. Because of the small number of observations, cross-validation was applied to test the performance of the models. PCA and PLS-DA analysis were performed using Unscrambler 9.6 (CAMO A/S, Trondheim, Norway).

6.2.7 Protein identification

Mass spectrometry experiments were performed using LTQ linear ion trap instrument fitted with a nanospray ion source (ThermoFisher, Waltham, MA) as detailed in Coulthurst *et al.* (2008). Since the number of *Pyrus communis* sequences in the public databases is very limited, MS/MS data analysis and cross species ID were applied for protein identification. Data were submitted to the Mascot search algorithm (Matrix Science, London UK) and searched against the GenBank non-redundant Viridiplantae-specific protein sequence database, using a fixed modification of carbamidomethyl and a variable modification of oxidation (M). In a second identification round, a Mascot search was performed against the *Malus domestica* EST sequences from Unigene database of December 8^{th}, 2007. Proteins assigned on the basis of two or more peptides were considered as confidently identified. When more than one protein was assigned within a spot, spectral counting was used as a rough estimate of protein abundance (Liu *et al.*, 2004). Protein

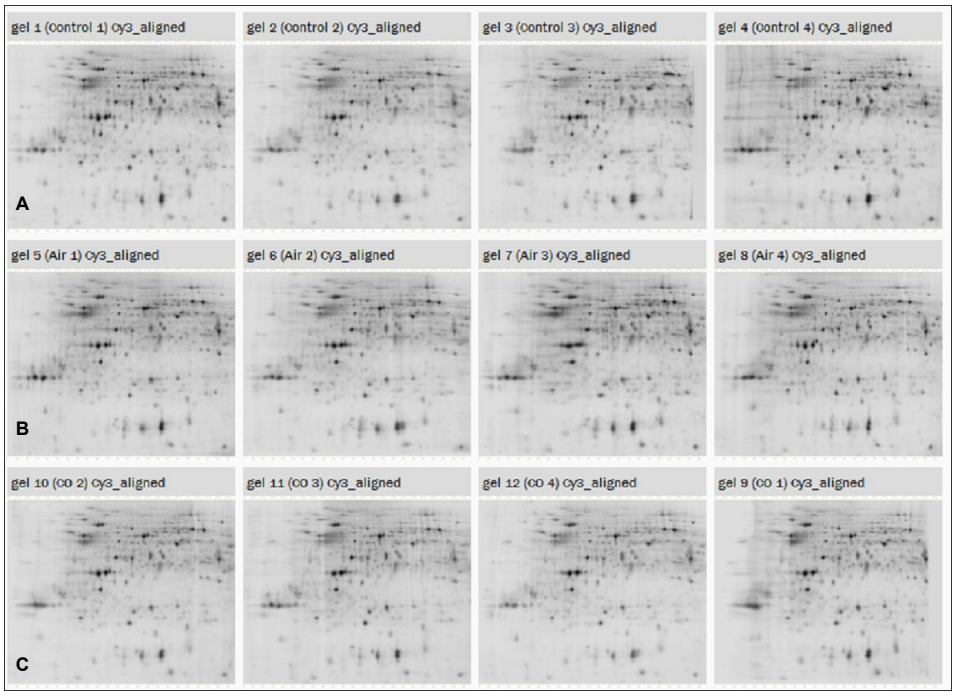

Figure 6.1: *2-DE DIGE maps for the different applied treatments. Row A: control (2.5 % O_2, 0.7 % CO_2), Row B: high oxygen or air (20 % O_2) and Row C: high carbon dioxide or anoxia (10 % CO_2). Proteins were separated on a 24 cm strip, 4-7 pI.*

identification was carried out in collaboration with the Cambridge Centre for Proteomics, University of Cambridge, United Kingdom and Prometa at Katholieke Universiteit Leuven.

6.3 Results

Even though extreme conditions were applied in terms of gas composition, the effect of the different treatments, did not trigger large changes in protein expression that could be visually seen (Figure 6.1) for the two conditions tested (air or anoxia) with respect to the control (commercial CA storage). Quantitatively it was found that the observed fold changes in protein expression levels were limited to a maximum of 4.

6.3.1 Spots selected through univariate statistics

The one-way ANOVA revealed 105 significant spots at a $p < 0.01$ value from which 56 were confirmed as real spots and suitable for identification. By applying the false discovery rate approach with a q-value threshold of 0.05, six out of the 56 were estimated to be false positives.

6.3.2 Spots selected through multivariate statistics

The PCA model generated with all variables included, revealed already a good discrimination among the different treatments (PC1 and PC2 were able to explain 22 % and 14 % of the total variance; data not shown). No-outlying gels were found. A PLS-DA model as a supervised technique sharpened the discrimination among the treatments and further analysis focuses on this multivariate technique. The PLS-DA model with all spot data revealed good discrimination among the different gas conditions being able to explain 93 % of the observed variation between the treatments based on the first two latent variables. In order to narrow down the number of proteins selected for further work, a new PLS-DA model was built based only on the 120 most important spots selected through the VIP procedure. This reduced model was still able to explain 86 % of the variation between the treatments with the first two latent variables.

6.3.3 Bringing both approaches together

From the 105 spots selected through one way ANOVA ($p < 0.01$), 75 were also selected through PLS-DA and VIP 120 procedure. Of the 120 spots chosen by the VIP procedure 112 spots had a $p < 0.05$ value. Spots that were identified as significant by either the univariate or the multivariate method were sent for MS identification after confirming they were real spots. Thus, 63 spots were submitted for identification out of which 43 spots were selected by both approaches and a new PLS-DA model based on these proteins was built (Figure 6.2A). The remaining spots were excluded from the analysis.

Focusing the multivariate technique on the spots that provide the discrimination allows a clear visualization as to which proteins species contribute to the separation in the multivariate space. The explained variance between the treatments was 93 % accounted for the two first latent variables when the 63 spots were included. The remaining variation was still very well explained by the remaining real 43 proteins obtaining a good discrimination between the treatments (Figure 6.2(B)).

6.3.4 LC-ESI-MS/MS identification of selected spots and correlation patterns

The statistically relevant spots selected through univariate and multivariate tools were analyzed by LC-ESI MS/MS. Fifty three out of 63 spots (84 %) yielded a confident match with pear or protein sequences from GenBank. The additional Mascot search against the apple EST database confirmed most of these identifications and increased the identification rate to 94 % (59 out of 63 proteins). The use of Malus ESTs is a valuable source data for the investigation of the poorly documented pear proteome. The presence of multiple proteins within a spot was clearly evident from the identifications (Figure 6.3). Thus, spectral counts were used as a rough estimate of protein abundance. Within a spot with multiple protein identifications, if the total number of spectral counts was significantly higher for a certain identification ($> 70\%$), it was considered as a singlet. A previous study showed that the top hit proteins are the most abundant on average contributing to 75 % of the spot intensities (Yang *et al.*, 2007). Thus, assigning the change in fluorescence or staining intensity to the most abundant or top hit protein is quite reasonable because the impact of the low abundance proteins appears limited. The PLS-DA model built with singlets is shown in Figure 6.2(B).

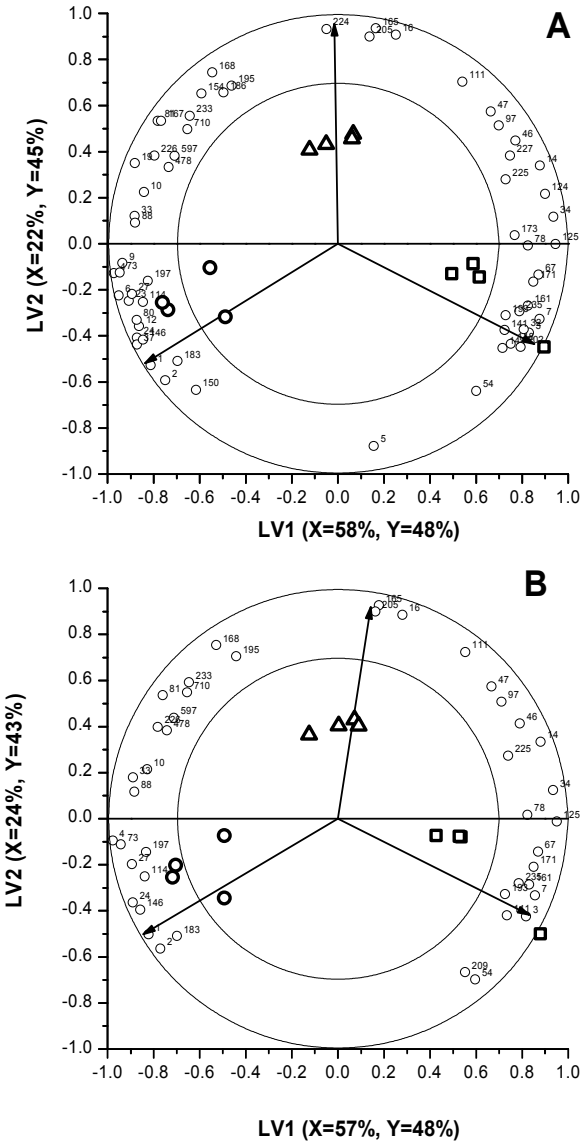

Figure 6.2: *PLS-DA biplot built with the (A) 63 proteins selected through univariate and multivariate statistics and sent for LC-ESI-MS/MS identification and (B) 43 identified proteins considered as singlets. Sample scores and loadings (proteins, small black dots) are superimposed. The Y-loadings are represented by the length of the arrows. The percentage explained variances are indicated on the axes. The analysis was based on the correlation matrix. Open small circles represent the different proteins. Squares represent control (2.5 % O_2, 0.7 % CO_2) conditions, circles represent high oxygen or air (20 % O_2) conditions and triangles represent high carbon dioxide or anoxic (10 % CO_2) conditions.*

6.4 Discussion

6.4.1 Statistical analysis

The high level of overlap of species identified as significant between the univariate and multivariate technique demonstrates the robustness of the technique and give confidence that highly significant changes are being identified. The unsupervised PCA technique revealed a good discrimination among the different treatments validating that the model was correct. The use of the supervised PLS-DA technique sharpened the discrimination among the treatments. The biplots (Figure 6.2) picture the marked discrimination among the treatments. The latent variable 1 (LV1) of the PLS-DA model is able to explain the variation between the control and oxygen while the latent variable 2 (LV2) explains the variation between the control and anoxia. Based on the marked discrimination of the treatments, it can be clearly observed that certain proteins are highly correlated to a specific treatment.

6.4.2 Spot overlap deserves special attention

With the introduction of more sensitive MS equipment, spot overlap has become more evident in 2-DE protein separation (Lu *et al.*, 2007; Xia *et al.*, 2007). This issue represents a serious problem in terms of relative quantification for further biological interpretation of the data. The presence of several proteins within a spot arises as in practice more than one protein can migrate to the same location on a 2-DE gel due to the large dynamic range of proteins within a cell. Additionally, unresolved proteins might be accompanied by contaminants. Different approaches to counteract the problem involve the use of narrow range IPG strips, sample fractionation methods, different sample preparation conditions and modification of conditions during 2-DE (Hunsucker & Duncan, 2006). As pear proteome is poorly characterized, the use of gel-free approaches is not a feasible alternative, since cross-species identification is the sole option for a poorly characterized genome. However, even when spot overlap exists, meaningful data can still

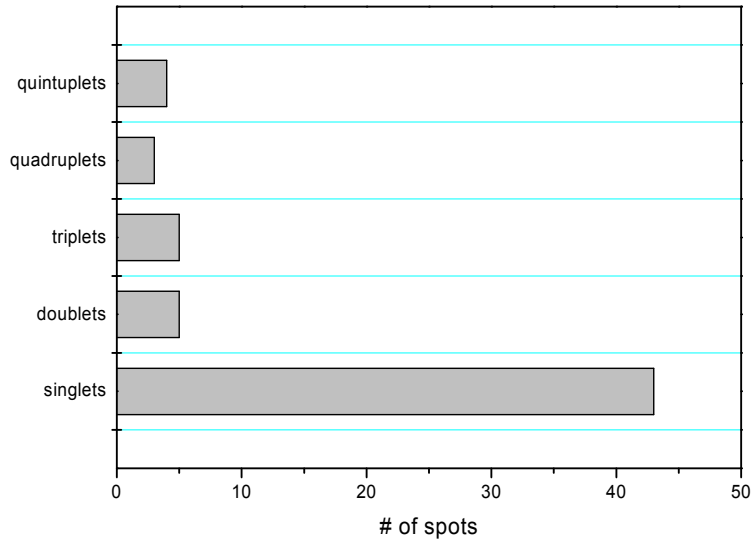

Figure 6.3: *Representation of the number of singlets, doublets, triplets, quadruplets and quintuplets found after LC-ESI-MS/MS identification and applying spectral counts as a rough indicative of protein abundance.*

be extracted. In this study, we used spectral counts to assign major proteins with mixed spectra. Spectral counting involves counting the total number of spectra representing identified peptides for a certain protein (Liu *et al.*, 2004; Deutsch *et al.*, 2008). Almost 80 % of the analyzed 63 proteins contained multiple proteins. By applying spectral counts as rough estimate of protein abundance, 43 of the 59 successfully identified proteins were identified as singlets (Figure 6.3). To be considered as singlet, the top hit ranked protein represented at least 70 % of the total spectral counts.

Spectral counts correlate excellent with protein concentration, in contrast to peptide counts that correlate poorly (Liu *et al.*, 2004). However, there are certain shortcomings that must be mentioned such as the fact that the approximation of abundance from the repeat peptide observations per protein totally ignores the size of the protein; large proteins contribute with more peptides than small ones, resulting in an overestimation if the data are not normalized (Lu *et al.*, 2007; Xia *et al.*, 2007). The data should also be normalized for the expected number of tryptic peptides (Nesvizhskii *et al.*,

2007). Additionally, depending on the instrumental setup for data acquisition, some peptides will be detected more easily than others and others will never be detected even if they are abundant in the sample (Colinge *et al.*, 2005; Deutsch *et al.*, 2008). Recently, Yang *et al.* (2007) has successfully applied an exponentially modified protein abundance index (emPAI) to determine the abundance of the individual proteins comprised within a spot containing multiple proteins. Although spectral counts correlate very well to protein abundance, they might not necessarily correlate very well to spot volume (e.g., minor protein components in terms of abundance might be more effectively labeled and thus account for most of the fluorescent signal).

6.4.3 Biological role of selected proteins

From the 43 identified singlets by spectral counts, a sub-selection of proteins for biological interpretation (Table 6.1) has been based on Tukey pair-wise comparisons ($p < 0.05$) following two main criteria: at least one of the treatments should differ from the control and there should be statistical differences between the treatments. The selected proteins have been classified based on their function and discussed accordingly. All comparisons are referred to control conditions.

6.4.3.1 The central metabolism

Within the glycolytic pathway, two enolase isoforms (24 and 33) were up-regulated in air as compared to the control. Previous studies on maize and other crops showed up-regulation of some isoforms of enolase under anoxia (Das & Uchimiya, 2002). Malate dehydrogenase (spot 46) was down-regulated in air conditions. Malic enzyme (spot 233) isoforms can be induced under stressful conditions to counteract the dependence on glycolysis and to produce extra energy by instead using the reserves of malic acid already present in the tissue.

It is expected that under anoxic conditions, glycolysis will be enhanced as a route to produce energy with a fermentative metabolism. Activation of

the pentose phosphate could be evidenced under anoxic conditions (transketolase, spot 125) and this finding provides further evidence to a previous metabolomics study in which there was pentose phosphate pathway activation under anoxic conditions (Pedreschi *et al.*, 2009a). The oxidative part of the pentose phosphate pathway is a major source of NADPH which is used for the synthesis of fatty acids and is important for the maintenance of the redox potential to protect against oxidative stress (Kruger & von Schaewen, 2003). In potato disc tubers under anoxia, enolase, aldolase, lactate dehydrogenase and alcohol dehydrogenase were up-regulated (Geigenberger *et al.*, 2000). However, we found no regulation of enolase under anoxia compared to the control but up-regulation of this enzyme under air conditions (Table 6.1). Previous works with whole pears and browning inducing conditions (Pedreschi *et al.*, 2007, 2008a) revealed that respiration pathways were at least partially involved in the appearance of core breakdown disorder in Conference pears. These results provide extra evidence of the response of respiration involved enzymes to the short term exposure to the currently tested gas conditions possibly leading to the appearance of core breakdown disorder. Acetyl CoA acetyltransferase (spot 27) certainly responds to air conditions being up-regulated compared to the control. This enzyme is involved in the mevalonate pathway that serves as a base for cell membrane maintenance. These results are consistent with the up-regulation of the pentose phosphate pathway given that mevalonate production is NADPH requiring. At a certain point, core breakdown results from membrane disruption. In summary, central metabolism pathways are subtly altered in the short term exposure of pear slices to the tested gas conditions.

Table 6.1: *Log standarized abundance (SA) ± stardard error (SE) for the proteins considered as singlets and chosen through one-way ANOVA and PLS-DA and VIP procedure. From the 63 proteins selected from independent univariate and multivariate statistical analysis, those proteins that were significantly different from the control and between the treatments by a pair-wise comparison Tukey test ($p < 0.05$) are displayed with an asterisk. Different superscript letters account for statistical differences ($p < 0.05$). The Mascot validation criteria used: expect value ≤ 0.05. The Sequest validation criteria used: cross correlation xC for 1^+, 2^+ and 3^+ charged ions ≥ 1.8, 2.5 and 3.5, respectively.*

Protein name	Spot ID	Protein ID	Treatment			Peptides[1]
			Control	Air	High CO_2	
I. Metabolism						
ATP synthase subunit d, mitochondrial	225	Mdo. 7184	$0.030^{a} \pm 0.008$	$-0.017^{b} \pm 0.013$	$0.018^{a,b} \pm 0.007$	7
Pyruvate dehydrogenase E1 β subunit	161	Mdo.3217	$0.027^{a} \pm 0.022$	$-0.048^{b} \pm 0.006$	$-0.030^{b} \pm 0.008$	7
ADP glucose synthase (*)	88	Mdo.1894	$-0.065^{a} \pm 0.021$	$0.048^{b} \pm 0.021$	$-0.000^{a,b} \pm 0.009$	12
Acetyl CoA acetyl transferase (*)	27	Mdo.9294	$-0.089^{a} \pm 0.033$	$0.118^{b} \pm 0.027$	$-0.032^{a} \pm 0.0222$	9
Phosphoglycerate dehydrogenase like	81	Mdo.10865	$-0.093^{a} \pm 0.010$	$0.012^{b} \pm 0.009$	$0.0267^{b} \pm 0.013$	11
Aconitase	226	Mdo.12067	$-0.022^{a} \pm 0.006$	$0.021^{b} \pm 0.008$	$0.015^{b} \pm 0.006$	17
Transketolase (*)	125	3559814	$0.074^{a} \pm 0.011$	$-0.020^{b} \pm 0.005$	$0.024^{c} \pm 0.011$	2
NADP-dependant malic enzyme (*)	233	Mdo.12341	$-0.048^{a} \pm 0.009$	$-0.020^{a,b} \pm 0.005$	$-0.014^{c} \pm 0.009$	15
NADP-dependant malic enzyme	710	Mdo.11996	$-0.066^{a} \pm 0.025$	$-0.018^{a} \pm 0.008$	$-0.016^{a} \pm 0.006$	13
Enolase (*)	24	Mdo.11920	$-0.072^{a} \pm 0.029$	$0.161^{b} \pm 0.016$	$-0.059^{a} \pm 0.023$	16
Phosphoglycerate kinase	168	Mdo.2160	$-0.053^{a} \pm 0.015$	$-0.011^{b} \pm 0.009$	$0.019^{b} \pm 0.005$	19
Phosphoglycerate kinase	195	Mdo.2160	$-0.048^{a} \pm 0.011$	$-0.014^{a,b} \pm 0.007$	$0.013^{b} \pm 0.013$	22
Malate dehydrogenase (*)	46	Mdo.15920	$0.048^{a} \pm 0.016$	$-0.127^{b} \pm 0.036$	$0.039^{a} \pm 0.018$	6
Enolase (*)	33	Mdo.11920	$-0.131^{a} \pm 0.042$	$0.065^{b} \pm 0.025$	$-0.022^{a} \pm 0.035$	5
II. Protein folding and stabilization						
Heat shock protein HSP 70	54	Mdo.1243	$0.066^{a} \pm 0.051$	$-0.037^{a,b} \pm 0.030$	$-0.082^{b} \pm 0.015$	33
Heat shock protein HSP 70	478	Mdo.1243	$-0.066^{a} \pm 0.036$	$0.018^{a} \pm 0.018$	$-0.003^{a} \pm 0.009$	40
Heat shock protein HSP 90	171	Mdo.423	$0.059^{a} \pm 0.013$	$-0.011^{b} \pm 0.011$	$0.011^{a,b} \pm 0.014$	16
Heat shock protein HSP 90	209	Mdo.423	$0.048^{a} \pm 0.016$	$0.014^{a,b} \pm 0.013$	$-0.006^{b} \pm 0.006$	10
III. Protein synthesis						
Translation initiation factor eIF-5A-2 (*)	16	77555893	$-0.035^{a} \pm 0.012$	$-0.118^{b} \pm 0.016$	$0.166^{c} \pm 0.018$	6
Nucleic acid binding protein	78	Mdo.6093	$0.101^{a} \pm 0.019$	$-0.024^{b} \pm 0.027$	$0.024^{a,b} \pm 0.008$	6
Translation elongation factor G	193	Mdo.6297	$0.054^{a} \pm 0.015$	$-0.009^{b} \pm 0.012$	$-0.007^{b} \pm 0.009$	14
CDC 48 like protein	597	Mdo.253	$-0.025^{a} \pm 0.030$	$0.039^{a} \pm 0.017$	$0.028^{a} \pm 0.005$	23

continued on next page

continued from previous page

Protein name	Spot ID	Protein ID	Treatment			Peptides[1]
			Control	Air	High CO_2	
Translation initiation factor eIF-4A (*)	14	Mdo.5235	$0.112^{a} \pm 0.020$	$-0.190^{b} \pm 0.052$	$0.036^{a} \pm 0.006$	19
Translation initiation factor eIF-4A (*)	146	Mdo.5235	$-0.050^{a} \pm 0.004$	$0.033^{b} \pm 0.007$	$-0.045^{a} \pm 0.010$	17
Translation initiation factor eIF-4A (*)	73	Mdo.5235	$-0.078^{a} \pm 0.013$	$0.049^{b} \pm 0.012$	$-0.031^{c} \pm 0.003$	26
IV. Transport						
Vacuolar ATPase subunit A	235	60592630	$0.020^{a} \pm 0.005$	$-0.010^{b} \pm 0.005$	$-0.006^{b} \pm 0.004$	24
V. Cellular communication						
14-3-3 like protein (*)	205	Mdo.5581	$-0.032^{a} \pm 0.007$	$-0.046^{a} \pm 0.013$	$0.010^{b} \pm 0.005$	12
VI. Stress related proteins						
Major allergen Mal d (*)	1	Mdo.13717	$-0.198^{a} \pm 0.053$	$0.330^{b} \pm 0.016$	$-0.202^{a} \pm 0.050$	8
Major allergen Mal d 1.03 (*)	4	Mdo.13717	$-0.204^{a} \pm 0.035$	$0.224^{b} \pm 0.023$	$-0.040^{c} \pm 0.047$	9
Major allergen Pyrc 1 (*)	183	3044216	$-0.001^{a} \pm 0.009$	$0.064^{b} \pm 0.013$	$-0.002^{a} \pm 0.015$	9
ACC oxidase (*)	47	4586409	$0.030^{a} \pm 0.021$	$-0.126^{b} \pm 0.024$	$0.044^{a} \pm 0.020$	10
Polygalacturonase inhibiting protein (*)	165	33087506	$-0.026^{a} \pm 0.014$	$-0.044^{a} \pm 0.006$	$0.030^{b} \pm 0.001$	15
Isoflavone reductase related protein	7	3243234	$0.316^{a} \pm 0.034$	$-0.039^{b} \pm 0.039$	$0.021^{b} \pm 0.033$	6
Superoxide dismutase	141	Mdo.1321	$0.105^{a} \pm 0.027$	$.022^{b} \pm 0.008$	$0.030^{a,b} \pm 0.018$	2
VII. Other proteins						
Endomembrane associated protein	3	Mdo.7202	$0.216^{a} \pm 0.089$	$-0.233^{b} \pm 0.104$	$-0.174^{b} \pm 0.047$	6
Hydrolase, hydrolyzing O-glycosyl (*)	34	Mdo.16650	$0.102^{a} \pm 0.008$	$-0.091^{b} \pm 0.030$	$0.019^{c} \pm 0.008$	2
Hydrolase, hydrolyzing O-glycosyl (*)	111	Mdo. 16650	$0.004^{a} \pm 0.017$	$-0.007^{b} \pm 0.016$	$0.033^{a} \pm 0.009$	2
Hydrolase, hydrolyzing O-glycosyl	10	Mdo.11995	$-0.143^{a} \pm 0.022$	$0.175^{b} \pm 0.040$	$0.118^{b} \pm 0.034$	3
Progesterone 5-beta reductase (*)	197	Mdo.6425	$-0.019^{a} \pm 0.009$	$0.042^{b} \pm 0.012$	$0.002^{a} \pm 0.007$	4
Clp C protease	67	Mdo.1094	$0.098^{a} \pm 0.007$	$-0.038^{b} \pm 0.022$	$-0.000^{b} \pm 0.019$	31
p-hydroxyphenyl pyruvate dioxygenase (*)	114	Mdo.3417	$-0.040^{a} \pm 0.013$	$0.058^{b} \pm 0.021$	$-0.016^{a} \pm 0.002$	2
Rubisco large subunit (*)	97	4098550	$0.003^{a} \pm 0.009$	$-0.102^{b} \pm 0.021$	$0.004^{a} \pm 0.022$	3

6.4.3.2 Allergens

The most pronounced changes in expression of proteins due to air or anoxic conditions were found with a series of allergen isoforms. They were consistently up-regulated in air conditions during the 5-day exposure of pear slices (Table 6.1) and highly correlated to most of the respiration involved enzymes (Figure 6.4). Previous studies conducted by our group on whole pears stored for long periods (6-8 months) under controlled atmosphere storage showed the same behavior (Pedreschi *et al.*, 2007, 2009a). A complete down regulation of these allergenic proteins was observed in browning inducing conditions of pear. These allergenic proteins show a high degree of sequence homology to the Mal d1 type allergens and Bet v 1, the major allergen of birch pollen (Karamloo *et al.*, 2001; Garcia *et al.*, 2007). People suffering the birch pollen allergy syndrome due to consumption of fruits and vegetables could benefit by consuming controlled atmosphere stored pears rather than air stored pears. A decrease in allergenicity has also been related to the action of the enzyme polyphenol oxidase (PPO) in apples (Garcia *et al.*, 2007). Both in the current and in previous studies on whole pears under browning inducing conditions, allergens were regulated (Pedreschi *et al.*, 2007). In brown pears (sub-optimally stored) the action of the enzyme PPO probably contributed to the total down regulation of these allergenic proteins compared to the sound tissue (sub-optimal conditions) which did not show a complete down regulation of these allergenic proteins. Thus, we suspect that total down regulation of these allergenic proteins during CA storage of pears might be correlated to the appearance of core breakdown. A partial down regulation is the consequence of the reduced oxygen concentration or high carbon dioxide concentration. Focus on the study of the role of these allergenic proteins and CA storage deserves further attention. Thus, elucidation of the sequence and behavior under different conditions of the different isoforms will be performed in future studies.

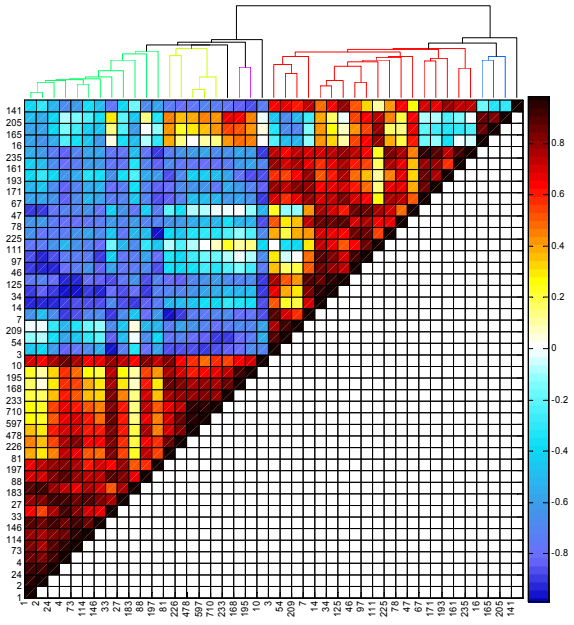

Figure 6.4: *Correlation map for the 43 proteins identified as singlets by means of LC-ESI MS/MS. Numbers represent the spot number of the proteins of Table 6.1. The color scale on the left goes from black '1' indicative of positive correlation to blue '-1' indicative of negative correlation.*

6.4.3.3 Protein synthesis related

Protein synthesis is an ATP dependent process. It has been previously reported in freshly cut potato tuber slices that as the oxygen concentration increases, protein synthesis increases as well (Geigenberger *et al.*, 2000). In this study, only two proteins involved in the synthesis of proteins (eukaryotic translation initiation factor eIF-4A, spots 146 and 73) were up-regulated under air conditions. Under anoxic conditions, the synthesis of anoxic enzymes is required. Thus, eIF-5A-2 (spot 16) was evidently up-regulated under anoxia (Table 6.1). Recently, it has been shown that eIF-5A-2 regulates programmed cell death caused by infection (Thompson, 2008). It might be possible after all that protein synthesis is mostly switched off except for certain anaerobic enzymes and that apoptosis is being induced. In a previous study with whole pears and long term exposure (Chapter 3), proteasomes believed to be involved in programmed cell death were up-regulated in tissue with visible symptoms of core breakdown (Pedreschi *et al.*, 2007).

6.4.3.4 Other stress related proteins

Contrary to previous studies on whole pears and long term exposure (Chapter 3), we found no differential expression of proteins involved the glutathione-ascorbate antioxidant system. ACC oxidase (spot 47) showed an opposite behavior compared to a previous study based on long term exposure of pears to different gas concentrations (Pedreschi *et al.*, 2007, 2009a). In this study, ACC oxidase was down-regulated in air. Polygalacturonase inhibiting protein (spot 165) was up-regulated under anoxic conditions. These results are in agreement to what we previously found in the long term exposure of whole pears to different gas concentrations (Chapter 3). Still, the reason for this behavior remains unknown but certainly deserves further attention specially because this enzyme is related to cell wall stress response and the final browning outcome is the result of membrane disruption (Franck *et al.*, 2007).

6.4.3.5 Proteins involved in regulatory processes

The 14-3-3 family protein (spot 205) was up-regulated under anoxia. This family protein has been reported to have roles in cell signaling, cell division, transcription and metabolism. In addition, it is also suspected that due to the existence of up to 13 genes in *Arabidopsis thaliana*, that particular isoforms have distinct biological functions (Roberts & Bruxelles, 2002). The specific response of this protein to gas concentrations deserves further research.

6.5 Conclusions

The current 2-DE DIGE approach confirmed previous studies on whole pears in terms of regulation of proteins involved in respiration, protein synthesis, ethylene responses and defense mechanisms in response to gas concentrations. In addition, anoxic conditions on pear slices revealed up-regulation

of a pentose phosphate pathway enzyme as an alternative route for production of reducing equivalents for further defense mechanisms and to skip ATP consuming steps. The role of allergenic proteins, as well as the polygalacturonase inhibiting protein deserves further attention. They responded clearly to the extreme gas concentrations applied and were previously found to be relevant in the appearance of the physiological core breakdown disorder after long term exposure of pears to sub-optimal storage conditions.

Chapter 7

General conclusions and future work

7.1 General conclusions

In order to study core breakdown in 'Conference' pears, it was first necessary to fine tune protein extraction, 2-DE and data analysis protocols. The highest protein yields were obtained when an extraction buffer of high pH containing high amounts of the reducing agent DTT was used. In order to draw strong conclusions from a gel-based proteomics study, it is highly recommended to begin with a good experimental design in which the biological variance is appropriately dealt with. Parametric statistical tests are more powerful, but in the univariate context the false positive rate must be controlled. Multivariate statistical analysis demands complete datasets, thus, missing values have to be dealt with appropriately. We showed that for our datasets, BPCA was superior in dealing with missing data. The use of univariate and multivariate analysis independently is a powerful approach and increases the confidence for a valid biological interpretation of the data.

Core breakdown is the consequence of a disturbance of the metabolism leading to the loss of homeostasis. Even though several pre and post harvest

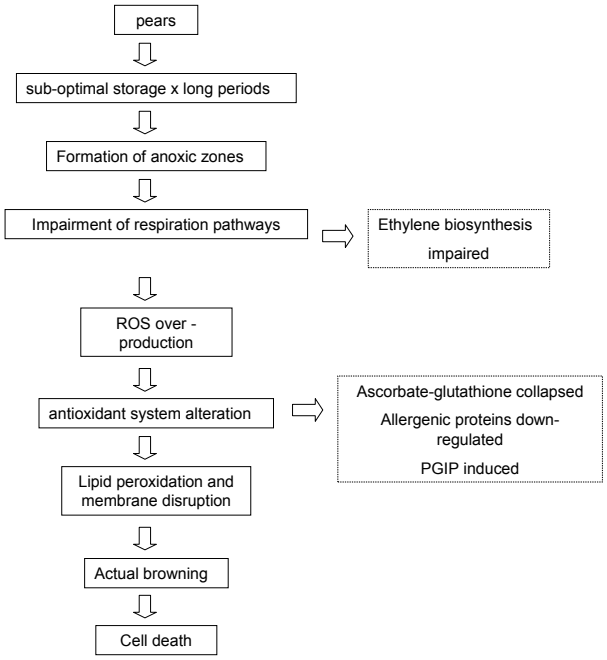

Figure 7.1: *Schematic overview of the series of events for browning development in pears.*

factors influence the incidence of the disorder, it is believed that extreme CA conditions trigger the disorder. A schematic overview of the series of events that are believed to lead to browning development is shown in Figure 7.1.

Pears stored for long periods under browning inducing conditions (low oxygen concentration and high carbon dioxide concentration) form anoxic zones due to gas gradients across the pear tissue (Lammertyn *et al.*, 2003b,a; Ho *et al.*, 2006c, 2008). Toward the center of the fruit, the oxygen concentration drops significantly. On the other hand, the carbon dioxide concentration increases toward the center of the fruit. As a consequence the metabolism responds to the spatial distribution of the atmosphere composition (aerobic closer to the peel; hypoxic and anoxic in the center) leading to differences in energy requirements. The effect of gas gradients across the tissue and

the corresponding different energy requirements could be evidenced by disturbance and/or impairment of respiration pathways to different degrees in brown, sound and healthy tissues respectively (Figure 7.2 and 7.3). The dependence of production of energy on glycolysis under stressful conditions (anoxia, brown tissue) was partially circumvented by the induction of malic enzyme isoforms which use the reserves of malic acid already present in the tissue (Figure 7.2). Methionine related pathways for polyamine and ethylene biosynthesis seemed to be activated (Figure 7.2 and 7.3).

Loss of homeostasis favors an over production of reactive oxygen species (ROS) that the pear antioxidant system can not cope with leading to a permanent state of oxidative stress (Figure 7.1). The main antioxidant system, glutathione-ascorbate, collapsed in brown tissue as evidenced by the total down regulation of several enzymes pertaining to this pathway (Figure 7.2). This antioxidant system is dependent on NADPH molecules provided by either the malic enzyme reaction and the pentose phosphate pathway (Figure 7.2 and 7.3). PR isoforms referred as allergens were repressed in brown tissue while PGIP was induced. The role of these enzymes in pathogenesis is well documented but not much is known about their role in anoxia.

Under a collapsed antioxidant system, lipid peroxidation was triggered leading to membrane disruption and furthermore to cell death (Figure 7.1). The mechanism of cell death in brown tissue seems to be programmed cell death. Different proteasome subunits were up-regulated and cell death was obvious from the lower total protein content of brown tissue (Figure 7.2).

The short term storage of pear slices under stressful conditions confirmed some changes also observed in the long term storage of whole pears. Some differences were also observed. Understanding the direct effect of the CA composition on the metabolism of pears on the short term independent of browning and gas gradient formation confirmed some previous findings. Subtle changes in respiration involved enzymes were observed. Major changes of

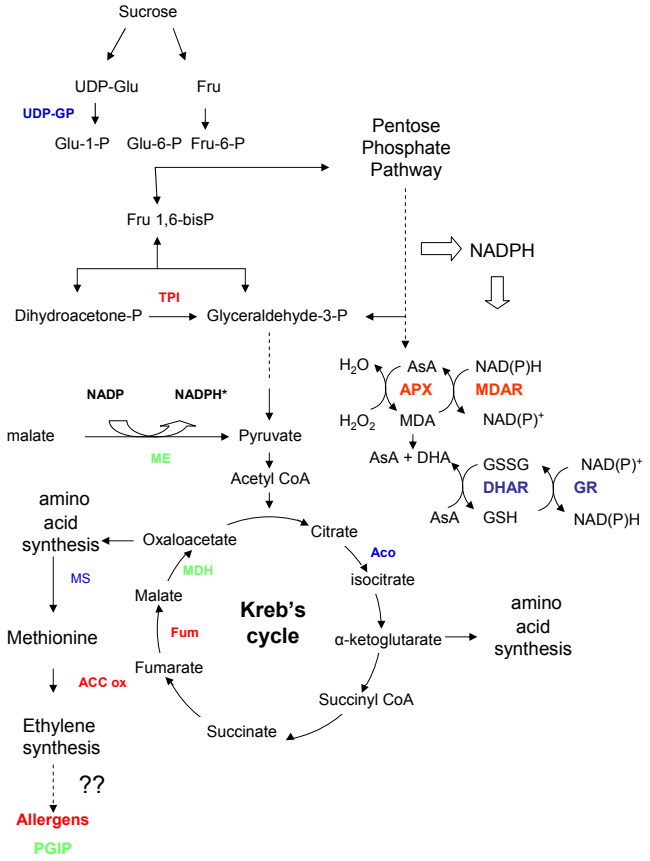

Figure 7.2: *Metabolic changes implicated in brown tissue under browning inducing conditions (1.0 % O_2, 10.0 % CO_2; no-pre cooling and fruit from late harvest). Black: no statistical change, red: down-regulation, blue: not analyzed, green: up-regulation. Up-down or no regulation is expressed relative to commercial conditions (2.5 % O_2, 0.7 % CO_2). UDP-GP: UDP-glucose phosphorylase, TPI: triosephosphate isomerase, ME: malic enzyme, APX: ascorbate peroxidase, MDAR: monodehydroascorbate reductase, DHAR: dehydroascorbate reductase, GR: glutathione reductase, MDH: malate dehydrogenase, Fum: fumarase, Aco: aconitase, MS: methionine synthase, ACC Ox: ACC oxidase, PGIP, polygalacturonase inhibiting protein, Eno: enolase, CoAT: acetyl CoA transferase, TK: transketolase.*

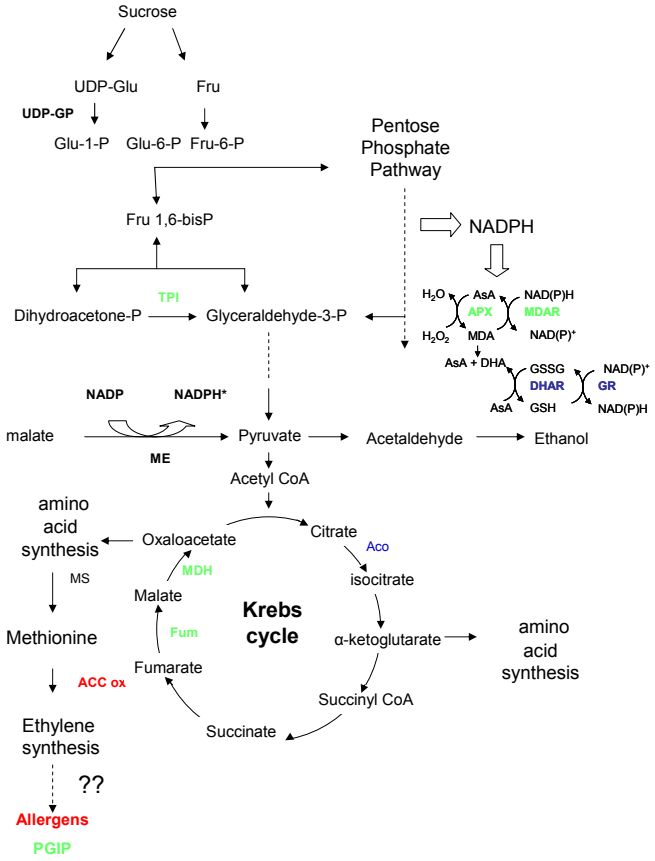

Figure 7.3: *Metabolic changes implicated in sound tissue under browning inducing conditions (1.0 % O_2, 10.0 % CO_2; no-pre cooling and fruit from late harvest). Black: no statistical change, red: down-regulation, blue: not analyzed, green: up-regulation. Up-down or no regulation is expressed relative to commercial conditions (2.5 % O_2, 0.7 % CO_2). UDP-GP: UDP-glucose phosphorylase, TPI: triosephosphate isomerase, ME: malic enzyme, APX: ascorbate peroxidase, MDAR: monodehydroascorbate reductase, DHAR: dehydroascorbate reductase, GR: glutathione reductase, MDH: malate dehydrogenase, Fum: fumarase, Aco: aconitase, MS: methionine synthase, ACC Ox: ACC oxidase, PGIP, polygalacturonase inhibiting protein, Eno: enolase, CoAT: acetyl CoA transferase, TK: transketolase.*

regulation were observed at the malic enzyme and pentose phosphate pathway (Figure 7.4). These results support our hypothesis about the need of activation of pathways providing reducing equivalents needed for membrane protection and defensive mechanisms (Figure 7.4).

Non significant changes were observed in the ascorbate-glutathione cycle but the allergenic and PGIP proteins involved in defense related mechanisms changes were observed. Browning development seems to start as soon as the ascorbate-glutathione antioxidant system collapses which is highly dependent on the redox status of the tissue. These allergenic proteins as well as PGIP are potential candidate markers to detect at an early stage and track core breakdown in 'Conference' pears.

7.2 Future work

In future work continuing from this thesis, the following aspects could be addressed:

- We have shown that respiration pathways play a key role in core breakdown development. One of the limitations of 2-DE, is the bias towards highly abundant proteins requiring prefractionation techniques in order to study the low abundant fraction. Many of the enzymes involved in respiration pathways might belong to this low abundant fraction. Thus organelle isolation is of key interest. The amount of mitochondria in pear tissue is very low. In order to increase mitochondria yields, protoplasts were isolated by enzymatic means (cellulose plus pectinase). This isolation process due to the temperature of extraction and long exposure time induced different cellular responses that would mask any treatment effect. Therefore, protoplast isolation as an intermediate step for mitochondria isolation does not seem to be a reasonable step.

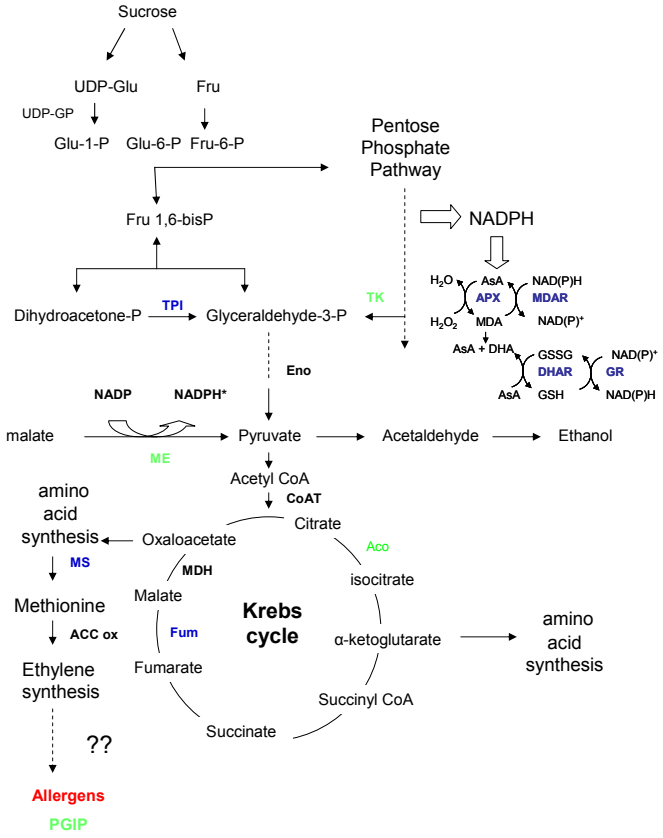

Figure 7.4: *Metabolic changes implicated in pear slices under anoxic conditions (0.0 % O_2, 10.0 % CO_2). Black: no statistical change, red: down-regulation, blue: not analyzed, green: up-regulation. Up-down or no regulation is expressed relative to commercial conditions (2.5 % O_2, 0.7 % CO_2). UDP-GP: UDP-glucose phosphorylase, TPI: triosephosphate isomerase, ME: malic enzyme, APX: ascorbate peroxidase, MDAR: monodehydroascorbate reductase, DHAR: dehydroascorbate reductase, GR: glutathione reductase, MDH: malate dehydrogenase, Fum: fumarase, Aco: aconitase, MS: methionine synthase, ACC Ox: ACC oxidase, PGIP, polygalacturonase inhibiting protein, Eno: enolase, CoAT: acetyl CoA transferase, TK: transketolase.*

- The isolation of mitochondria, so far has not been successfully carried out starting from either fruit tissue or leaves and using either mechanical disruption or enzymatic means. Isolation was carried out by means of differential centrifugation on a Percoll gradient. In the Percoll gradient, it was impossible to obtain the different layers indicating the different organelles even though several combinations of centrifugation force and time and Percoll percentages were tried. Thus, additional work is needed. This is an exhausting task because it demands trial and error work. The use of sucrose instead of Percoll gradients might be an option. To assess for purity of mitochondria, the activity of the marker enzyme cytochrome c oxidase was tracked in the different fractions. We were able to optimize this assay to assess for mitochondria purity.

- Several isoforms of the major allergen Pyrc 1, identified as Mal d1, were observed to respond differently to the treatments. The role of these proteins have been studied mainly in relation to pathogenesis. Since these proteins might have a role in defense responses due to gas concentrations, there is a need for further studies. The first step might be the sequence identification of these different isoforms. These proteins seem to respond clearly to the oxygen and carbon dioxide concentrations applied and might be interesting markers to track core breakdown. Multiple Response Monitoring (MRM) should be explored as a suitable MS approach to determine isoform abundances. (Fusaro *et al.*, 2009).

- The polygalacturonase inhibiting protein also popped up as candidate marker in the different experiments. Apparently it was always induced under anoxic conditions. The role of this protein has been studied in a pathogen context but not specifically under oxygen deprivation stress. There is need to direct efforts toward understanding the role of this enzyme in relation to oxygen deprivation.

- Core breakdown develops with time. There is need for proteomic studies that include time as a variable besides the effect of the gas compo-

sition. In addition, the pre-harvest factors play an important role in determining the incidence of core breakdown. Therefore, these factors should be included in future studies.

- Membrane disruption is a key event in core breakdown. Thus, future studies should focus on the membrane proteome which represents one third of the total proteome. Membrane proteins have roles in signaling, trafficking, transport, cell structure. As an example, the mitochondrial inner membrane is made up of protein complexes of the respiratory chain and proteins involved in triggering programmed cell death. These proteins are challenging to work with due to their hydrophobic nature, low abundance and heterogeneity. Membrane proteins are under-represented in 2-DE gels. Alternative proteomic platforms such as blue native PAGE, clear native PAGE, SDS/SDS PAGE and LC based approaches such as free flow electrophoresis and multidimensional protein identification technology should be further explored (Tan *et al.*, 2008). In addition, the sub-cellular localization of membrane proteins is essential for determining function. Localization of organelle proteins by isotope tagging (Sadowski *et al.*, 2008) should be further explored.

- Validation of the proteomics approach utilized through another low throughput technique for instance would be appropriate (e.g., measurement of enzymatic activity for some of the enzymes found to be differentially regulated).

- This thesis represents a first global attempt to understand a complex metabolic disorder. In addition, there is need to uplift this approach to a systems biology approach integrating data from different 'omics' platforms into an overall metabolic network model. Such model should allow for the integration of knowledge on the different mechanisms underlying homeostasis and help to identify the complex interplay of the various processes leading to the storage disorder. For instance, a better understanding of the metabolite synthesis and its regulation under specific conditions would be a great contribution.

- Instead of just measuring metabolite concentrations through a metabolomic approach, emphasis should be put toward sophisticated time lapse experiments measuring actual metabolite fluxes based on which a kinetics metabolic network model can be developed that can be coupled with our existing models in the area of gas transfer.

Bibliography

Abdi, N., Holford, P., & McGlasson, B. 2002. Application of two-dimensional gel electrophoresis to detect proteins associated with harvest maturity in stonefruit. *Postharvest Biology and Technology*, **26**(1), 1–13.

Alban, A., Olu, S., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., & Currie, I. 2003. A novel experimental design for comparative two-dimensional gel analyis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics*, **3**, 36–44.

Alferez, F., Sala, J. M., Sanchez-Ballesta, M. T., Mulas, M., Lafuente, M. T., & Zacarias, L. 2005. A comparative study of the postharvest performance of an ABA-deficient mutant of oranges I. Physiological and quality aspects. *Postharvest Biology And Technology*, **37**(3), 222–231.

Allison, P. 2000. Multiple imputation for missing data. *Sociological Methods and Research*, **28**, 301–309.

Alscher, R. G. 1989. Biosynthesis and antioxidant function of glutathione in plants. *Physiologia Plantarum*, **77**(3), 457–464.

Andersen, J.S., Svensson B, & Roepstorff, P. 1996. Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. *Nature Biotechnology*, **14**, 449–457.

Askari, H., Edqvist, J., Hajheidari, M., Kafi, M., & Salekdeh, G. H. 2006. Effects of salinity levels on proteome of *Suaeda aegyptiaca* leaves. *Proteomics*, **6**(8), 2542–2554.

Bartoli, C.G., Pastori, G.M., & Foyer, C.H. 2000. Ascorbate biosynthesis in mitochondria is linked to the electron transport chain between complexes III and IV. *Plant Physiology*, **123**, 335–343.

Bauchot, A.D., Reid, S.J., Ross, S.G., & Burmeister, D.M. 1999. Induction of apple scald by anaerobiosis has similar characteristics to naturally occurring superficial scald in Granny Smith apple fruit. *Postharvest Biology and Technology*, **16**, 9–14.

Blum, H., Beier, H., & Gross, H. J. 1987. Improved silver staining of plant-proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, **8**(2), 93–99.

Bolhaar, S. T. H. P., van de Weg, W. E., van Ree, R., Gonzalez-Mancebo, E., Zuidmeer, L., Bruijnzeel-Koomen, C. A. F. M., Fernandez-Rivas, M., Jansen, J., Hoffmann-Sommergruber, K. H., Knulst, A. C., & Gilissen, L. J. W. J. 2005. *In vivo* assessment with prick-to-prick testing and double-blind, placebo-controlled food challenge of allergenicity of apple cultivars. *Journal of Allergy and Clinical Immunology*, **116**(5), 1080–1086.

Bradford, M. M. 1976. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Analytical Biochemistry*, **72**(1-2), 248–254.

Bras, L.P., & Menezes, J.C. 2006. Dealing with gene expression missing data. *IEE Proceedings Systems Biology*, **153**, 105–119.

Brecht, J.K., Chau, K.V., Fonseca, S.C., Oliveira, F.A.R., Silva, F.M., Nunes, M.C.N., & Bender, R.J. 2003. Maintaining optimal atmosphere conditions for fruits and vegetables throughout the postharvest handling chain. *Postharvest Biology and Technology*, **27**, 87–101.

Brock, N., Shaffer, J., Blakesley, R., Lotz, M., & Tseng, G. 2008. Which missing value imputation method to use in expression profiles: a comparative study and two selection schemes. *BMC Bioinformatics*, **9**, 12.

Burmeister, D. M., & Dilley, D. R. 1995. A Scald-Like Controlled-Atmosphere Storage Disorder Of Empire Apples - A Chilling Injury-Induced By Co2. *Postharvest Biology And Technology*, **6**(1-2), 1–7.

Burmeister, D.M., & Harman, J.E. 1998. Effect of fruit maturity on the success of controlled atmosphere storage of Fantasia nectarines. *Acta Horticulturae*, **464**, 636–369.

Burssens, S., Himanen, K., van de Cotte, B., Beeckman, T., Van Montagu, M., Inz, D., & Verbruggen, N. 2000. Expression of cell cycle regulatory genes and morphological alterations in response to salt stress in *Arabidopsis thaliana*. *Planta*, **211**, 632–640.

Campostrini, N., Areces, L., Rappsilber, J., Pietrogrance, C., Dondi, F., Pastorino, F., Ponzoni, M., & Righetti, P.G. 2005. Spot overlapping in two-dimensional maps: a serious problem ignored for too long. *Proteomics*, **5**, 2385–2395.

Capitani, G., Hohenester, E., Feng, L., Storici, P., Kirsch, J. F., & Jansonius, J. N. 1999. Structure of 1-aminocyclopropane-1-carboxylate synthase, a key enzyme in the biosynthesis of the plant hormone ethylene. *Journal of Molecular Biology*, **294**(3), 745–756.

Carpentier, S., Coemans, B., Podevin, N., Laukens, K., Witters, E., Matsumura, H., Terauchi, R., Swenne., & Panis, B. 2008a. Functional genomics in a non-model crop: transcriptomics or proteomics? *Physiologia Plantarum*, **133**, 117–130.

Carpentier, S., Panis, B., Vertommen, A., Swennen, R., Seargeant, K., Renaut, J., Laukens, K., Witters, E., Samyn, B., & Devreese, B. 2008b. Proteome analysis of non-model plants: a challenging but powerful approach. *Mass Spectrometry Reviews*, **27**, 354–377.

Carpentier, S. C., Witters, E., Laukens, K., Deckers, P., Swennen, R., & Panis, B. 2005. Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics*, **5**(10), 2497–2507.

Carpentier, S. C., Witters, E., Laukens, K., Van Onckelen, H., Swennen, R., & Panis, B. 2007. Banana (Musa spp.) as a model to study the meristem proteome: Acclimation to osmotic stress. *Proteomics*, **7**(1), 92–105.

Casado-Vela, J., Selles, S., & Martinez, R. B. 2005. Proteomic approach to blossom-end rot in tomato fruits (*Lycopersicon esculentum* M.): Antioxidant enzymes and the pentose phosphate pathway. *Proteomics*, **5**(10), 2488–2496.

Casado-Vela, J., Sells, S., & Martnez, R.B. 2006. Proteomic analysis of tobacco mosaic virus-infected tomato (*Lycopersicon esculentum* M.) fruits and detection of viral coat protein. *Proteomics*, **6**, S196–S206.

Casati, P., Drincovich, M. F., Edwards, G. E., & Andreo, C. S. 1999. Malate metabolism by NADP malic enzyme in plant defense. *Photosynthesis Research*, **61**(2), 99–105.

Chang, J., Van Remmen H., Ward, W., Regnier, F., Richardson, A., & Cornell, J. 2004. Processing of data generated by two dimensional gel electrophoresis for statistical analyis: missing data, normalization and statistics. *Journal of Proteome Research*, **3**, 1210–1218.

Chevalier, F., Rofidal, V., Vanova, P., Bergoin, A., & Rossignol, M. 2004. Proteomic capacity of recent fluorescent dyes for protein staining. *Phytochemistry*, **65**, 1499–1506.

Chich, J.F., David, O., Villers, F., Schaeffer, B., Lutomski, D., & Huet, S. 2007. Statistics for proteomics: experimental design and 2-DE differential analysis. *Journal of Chromatography B*, **849**, 261–272.

Cho, S.M., Shin, S.H., Kin, K.S., Kim, Y.C., Eun, M.Y., & Cho, B.H. 2004. Enhanced expression of a gene encoding a nucleoside diphos-

phate kinase 1 (OsNDPK1) in rice plants upon infection with bacterial pathogens. *Molecules and Cells*, **18**, 390–395.

Colinge, J., Chiappe, D., Lagache, S., & Moniatte, M. 2005. Differential proteomics via probabilistic peptide identification scores. *Analytical Chemistry*, **77**, 596–606.

Cordwell, S.J. 1999. Microbial genomes and missing enzymes: redifining biochemical pathways. *Archives of Microbiology*, **172**, 269–279.

Coulthurst, S.J., Lilley, K.S., Hedley, P.E., Liu, H., Toth, I.K., & Salmond, G. 2008. DsbA plays a critical and multifaceted role in the production of secreted virulence factors by the phytopathogen *Erwinia carotovora* subsp. atroseptica. *Journal of Biological Chemistry*, **283**, 23739–23753.

Cui, T., Nakamura, K., Ma, L., Li, J., & Nayahara, H. 2005. Analyses of arbutin and chlorogenic acid, the major phenolic constituents in oriental pear. *Journal of Agricultural and Food Chemistry*, **53**, 3882–3887.

Danvind, J. 2002. PLS prediction as a tool for modelling wood properties. *Holz als Roh-und Werstoff*, **60**, 130–140.

Das, A., & Uchimiya, H. 2002. Oxygen stress and adaptation of a semi-aquatic plant: rice (*Oryza sativa*). *Journal of Plant Research*, **115**, 315–320.

Davey, M., Van Montagu, M., Inze D, Sanmartin, M., Kanellis, A. Smirnoff, N. Benzie I., Strain, J. Favell, D., & Fletcher, J. 2000. Plant L-ascorbic acid: chemistry, function, metabolism, bioavailability and effects of processing. *Journal of the Science of Food and Agriculture*, **80**, 825–860.

Davletova, S., Rizhsky, L., Liang, H. J., Zhong, S. Q., Oliver, D. J., Coutu, J., Shulaev, V., Schlauch, K., & Mittler, R. 2005. Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of Arabidopsis. *Plant Cell*, **17**(1), 268–281.

Dawson, S. P., Arnold, J. E., Mayer, N. J., Reynolds, S. E., Billet, M. A., Gordon, C., Colleaux, L., Kloetzel, P. M., Tanaka, K., & Mayer, R. J. 1995. Developmental changes of the 26-S proteasome in abdominal intersegmental muscles of Manduca-Sexta during programmed cell death. *Journal of Biological Chemistry*, **270**(4), 1850–1858.

de Castro, E., Barrett, D.M., Jobling, J., & Mitcham, E.J. 2008. Biochemical factors associated with a CO_2-induced flesh browning disorder of Pink Lady apples. *Postharvest Biology and Technology*, **48**, 182–191.

Delumeau, O., Morere-Le Paven, M. C., Montrichard, F., & Laval-Martin, D. L. 2000. Effects of short-term NaCl stress on calmodulin transcript levels and calmodulin-dependent NAD kinase activity in two species of tomato. *Plant Cell and Environment*, **23**(3), 329–336.

Dennis, D., Huang, Y., & Negm, F. 1998. *Plant Metabolism.* Adison Wesley Longman, London, UK. Chap. Glycolysis, the pentose phosphate pathway and anaerobic respiration, pages 105–123.

Deutsch, E.W., Lam, H., & Aebersold, R. 2008. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiological Genomics*, **33**, 18–25.

Dobrota, C. 2006. Energy dependant plant stress acclimation. *Reviews in Environmental Science and Biotechnology*, **5**, 1569–1705.

Dumont, D., Noben, J. P., Raus, J., Stinissen, P., & Robben, J. 2004. Proteomic analysis of cerebrospinal fluid from multiple sclerosis patients. *Proteomics*, **4**(7), 2117–2124.

Edwards, G.E., & Andreo, C.S. 1992. NADP-malic enzyme from plants. *Photosynthesis Research*, **31**, 1845–1857.

Eisen, M.B., Spellman, P.T., Brown, P.O., & Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academic of Sciences of the United States of America*, **95**, 14863–14869.

Espin, J., Garcia-Ruiz, P., Tudela, J., & Garcia-Canovas, I. 1998. Study of stereospecificity in pear and strawberry polyphenol oxidases. *Journal of Agricultural and Food Chemistry*, **46**, 2469–2473.

Faurobert, M., Mihr, C., Bertin, N., Pawlowski, T., Negroni, L., Sommerer, N., & Causse, M. 2007. Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiology*, **143**, 1327–1346.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., & Whitehouse, C.M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, **246**, 64–71.

Fernandez, A., Ogawa, J., Penaud, S., Boudebbouze, S., Ehrlich, D., van de Guchte, M., & Maguin, E. 2008. Re-routing of pyruvate metabolism during acid adaptation in *Lactobacillus bulgaricus*. *Proteomics*, **8**, 3154–3163.

Fernie, A.R., Carrari, F., & Sweetlove, L. J. 2004. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current Opinion in Plant Biology*, **7**, 254–261.

Fodoroff, N. 2006. Redox regulatory mechanisms in cellular stress responses. *Annals of Botany*, **98**, 289–300.

Franck, C., Baetens, M., Lammertyn, J., Scheerlinck, N., Davey, M. W., & Nikolaï, B. M. 2003. Ascorbic acid mapping to study core breakdown development in Conference pears. *Postharvest Biology and Technology*, **30**(2), 133–142.

Franck, C., Lammertyn, J., Ho, Q.T., Verboven, P., Verlinden, B. E., & Nicolaï, B. A. 2007. Review: Browning disorders in pear fruit. *Postharvest Biology and Technology*, **43**, 1–13.

Fridman, E., & Pichersky, E. 2005. Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Current Opinion in Plant Biology*, **8**, 242–248.

Fuji, K., Kondo, T., Yokoo, H., Yamada, T., Matsuno, Y., Iwatsuki, K., & Hirohashi, S. 2005. Protein expression pattern distinguishes different lymphoid neoplasms. *Proteomics*, **5**, 4274–4286.

Fusaro, V., Mani, D., Mesirov, J., & Carr, S. 2009. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature Biotechnology*, **27**, 190–198.

Garcia, A., Wichers, J. H., & Wichers, H. J. 2007. Decrease of the IgE-binding by Mal d 1, the major apple allergen, by means of polyphenol oxidase and peroxidase treatments. *Food Chemistry*, **103**(1), 94–100.

Gaspar, T., Franck, T., Bisbis, B., Kevers, C., Jouve, L., Hausman, J.F., & Dommes, J. 2002. Concepts in plant stress physiology. Application to plant tissue cultures. *Plant Growth Regulation*, **37**, 263–285.

Geigenberger, P. 2003. Response of plant metabolism to too little oxygen. *Current Opinions in Plant Biology*, **6**, 247–256.

Geigenberger, P., Fernie, A.R., Gibon, Y., & Christ, M. 2000. Metabolic activity decreases as an adaptive response to low internal oxygen in growing potato tubers. *Biological Chemistry*, **381**, 723–740.

Giraud, M., Westercamp, P., Coureau, J.F., Chapon, J.F., & Berrie, A. 2001. *Recognizing postharvest diseases of apple and pear.* Centre Technique Interprofessionel des Fruits et Legumes.

Giribaldi, M., Perugini, I., Sauvage, F.X., & Schubert, A. 2007. Analysis of protein changes during grape berry ripening by 2-DE and MALDI-TOF. *Proteomics*, **7**, 3154–3170.

Gomez-Galindo, F., & Sjiholm, I. 2007. Plant stres physiology: opportunities and challenges for the food industry. *Critical Reviews in Food Science and Nutrition*, **46**, 749–763.

Gonzalez-Candelas, L., Sanchez-Torres, P., Alamar, S., Estables, B., Ballester, A., Sanchez-Ballesta, M., Luch, Y., Gosalbes, M., Zacarias, L., Marcos, J., Lafuente, M., Forment, J.,

& Granell, A. 2005. Genomic approaches to postharvest biotic and abiotic stresses of citrus fruit. *Acta Horticulturae*, **682**, 225–232.

Görg, A., Obermaier, C., Boguth, C., Csordas, A., Diaz, J., & Madjar, J. 1997. Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis*, **18**, 328–337.

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

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

Gorny, J. R., & Kader, A. A. 1996a. Controlled-atmosphere suppression of ACC synthase and ACC oxidase in Golden Delicious apples during long-term cold storage. *Journal of the American Society for Horticultural Science*, **121**(4), 751–755.

Gorny, J.R., & Kader, A.A. 1996b. Regulation of ethylene biosynthesis in climateric apple fruit by elevated CO_2 and reduced O_2 atmospheres. *Postharvest Biology and Technology*, **9**, 311–323.

Gottlob-McHugh, S.G., Sangwan, R.S., Blakeley, S.D., Vanlerberghe, G., Ko, K., & Turpin, D.H. 1992. Normal growth of transgenic tobacco plants in the absence of cytosolic pyruvate kinase. *Plant Physiology*, **100**, 820–825.

Granell, A., Pons, C., Marti, C., Forment, J., Royo, C., Gradziel, R., Peace, C., Ogundiwin, E., & Crisoto, C. 2007. Genomic approaches innovative tools to improve quality of fresh cut produce. *Acta Horticulturae*, **746**, 203–211.

Granier, C., Inze, D., & Tardieu, F. 2000. Spatial distribution of cell division rate can be deduced from that of p34(cdc2) kinase activity in maize leaves grown at contrasting temperatures and soil water conditions. *Plant Physiology*, **124**(3), 1393–1402.

Grinberg, Y., James, P.E., & Swartz, H.M. 1998. *Oxygen transpot to tissue*. Plenum Press, New York. Chap. Are there significant gradients of pO_2 in cells?, pages 415–423.

Grisson, C.B., Canellas, P.F., & Wedding, R.T. 1983. Allosteric regulation of the NAD malic enzyme from cauliflower: activation by fumarate and coenzyme A. *Archives of Biochemistry and Biophysics*, **220**, 133–144.

Grove, H., Hollung, K., Uhlen, A. K., Martens, H., & Faergestad, E. M. 2006. Challenges related to analysis of protein spot volumes from two-dimensional gel electrophoresis as revealed by replicate gels. *Journal of Proteome Research*, **5**(12), 3399–3410.

Grung, B., & Manne, R. 1998. Missing values in principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **42**(1-2), 125–139.

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

Hajheidari, M., Abdollahian-Noghabi, M., Askari, H., Heidari, M., Sadeghian, S. Y., Ober, E. S., & Salekdeh, G. H. 2005. Proteome analysis of sugar beet leaves under drought stress. *Proteomics*, **5**(4), 950–960.

Hall, R.D., Brouwer, I.D., & Fitzgerald, M.A. 2008. Plant metabolomics and its potential application for human nutrition. *Physiologia Plantarum*, **132**, 162–175.

Haruta, M., Murata, M., Kadokura, H., & Homma, S. 1999. Inmunological and molecular comparison of polyphenol oxidase in *Rosaceae* fruit trees. *Phytochemistry*, **50**(6), 1021–1025.

Hjern, K., Alm, R., Canbck, B., Matthiesen, R., Trajkovski, K., Bjrk, L., Roepstorff, P., & Emanuelsson, C. 2006. Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant. *Proteomics*, **6**, 1574–1587.

Ho, E., Hayen, A., & Wilkins, M.R. 2006a. Characterisation of organellar proteomes: A guide to subcellular proteomic fractionation and analysis. *Proteomics*, **6**, 5746–5757.

Ho, Q., Verlinden, B. E., Verboven, P., Vandewalle, S., & Nicolaï, B. A. 2008. A continuum model for metabolic gas exchange in pear fruit. *PLoS Computational Biology*, **4**, e10000023.

Ho, Q. T., Verlinden, B. E., Verboven, P., & Nicolaï, B. A. 2006b. Gas diffusion properties at different positions in the pear. *Postharvest Biology and Technology*, **41**(2), 113–120.

Ho, Q.T. 2008. *Multiscale modelling of gas transport in fruit during controlled atmosphere storage*. Ph.D. thesis, Katholieke Universiteit Leuven.

Ho, Q.T., Verlinden, B.E., Verboven, P., Vandewalle, S., & Nicolaï, B. A. 2006c. A permeation-diffusion-reaction model of gas transport in cellular tissue of plant materials. *Journal of Experimental Botany*, **57**, 4215–4224.

Hollywood, K., Brison, D., & Goodacre, R. 2006. Metabolomics: Current technologies and future trends. *Proteomics*, **6**, 4716–4723.

Holtorf, H., Guitton, M.C., & Reski, R. 2002. Plant functional genomics. *Naturwissenschaften*, **89**, 235–249.

Honore, B., Ostergaard, M., & Vorum, H. 2004. Functional genomics studied by proteomics. *BioEssays*, **26**, 901–915.

Horgan, G.W. 2007. Sample size and replication in 2D gel electrophoresis studies. *Journal of Proteome Research*, **6**, 2884–2887.

Hoving, S., Gerrits, B., Voshol, H., Muller, D., Roberts, R.C, & Van Oostrum, J. 2002. Preparative two dimensional electrophoresis at alkaline pH using narrow range immobilized pH gradients. *Proteomics*, **2**, 127–134.

Hunsucker, S.W., & Duncan, M.W. 2006. Is protein overlap in two-dimensional gels a serious practical problem? *Proteomics*, **6**, 1374–1375.

Hunt, L., Lerner, F., & Ziegler, M. 2004. NAD new roles in signalling and gene regulation in plants. *New Phytology*, **163**, 31–44.

Hunt, S., Thomas, M., Sebastian, L., Pedersen, S., Harcourt, R., Sloane, A., & Wilkins, M. 2005. Optimal replication and the importance of experimental design for gel-based quantitative proteomics. *Journal of Proteome Research*, **4**, 809–819.

Jacobs, D.I., van der Heijden, R., & Verpoorte, R. 2000. Proteomics in plant biotechnology and secondary metabolism research. *Phytochemical Analysis*, **11**, 277–287.

Jenner, H. L., Winning, B. M., Millar, A. H., Tomlinson, K. L., Leaver, C. J., & Hill, S. A. 2001. NAD malic enzyme and the control of carbohydrate metabolism in potato tubers. *Plant Physiology*, **126**(3), 1139–1149.

Jessen, F., Lametsch, R., Bendixen, E., Kjaersgard, I. V. H., & Jorgensen, B. M. 2002. Extracting information from two-dimensional electrophoresis gels by partial least squares regression. *Proteomics*, **2**(1), 32–35.

Johnson, A.R., & Wichern, W.D. 1998. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, New Jersey.

Jorsten, R., Wang, H., Welsh, W., & Ouyang, M. 2005. DNA microarray data imputation and significance analysis of differential expression. *Bioinformatics*, **21**, 4155–4161.

Jung, K., Gannoun, A., Sitek, B., Meyer, H., Stuhler, K., & Urfer, W. 2005. Analysis of dynamic protein expression data. *REVSTAT-Statistical Journal*, **3**, 99–111.

Jung, K., Gannoun, A., Sitek, B., Apostolov, O., Schramm, A., Meyer, H., Stuhler, K., & Urfer, W. 2006. Statistical evaluation of methods for the analysis of dynamic protein expression data from a tumor study. *REVSTAT-Statistical Journal*, **4**, 67–80.

Kadam, S.S., Dhumal, A., & Shinde, N.N. 1995. *Handbook of fruit science and technology.* Marcel Dekker Inc, New York. Chap. Pear, pages 183–202.

Kader, A. 2002. Postharvest biology and technology of horticultural crops. *University of California, Division of Agriculture and Natural Resources*, **3311**, 39–47.

Kader, A.A. 1986. Biochemical and physiological basis for effects of controlled and modified atmospheres on fruits and vegetables. *Food Technology*, **40**, 99–100.

Karamloo, F., Scheurer, S., Wangorsch, A., May, S., Haustein, D., & Vieths, S. 2001. Pyr c 1, the major allergen from pear (*Pyrus communis*), is a new member of the Bet v 1 allergen family. *Journal of Chromatography B*, **756**, 281–293.

Karas, M., & Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10000 Daltons. *Analytical Chemistry*, **60**, 2299–2301.

Karp, N., Spencer, M., Lindsay, H., O'Dell, K., & Lilley, K. 2005a. Impact of replicate types on proteomic expression analysis. *Journal of Proteome Research*, **4**, 1867–1871.

Karp, N., McCormick, P.S., Russell, M.R., & Lilley, K.S. 2007. Experimental and statistical considerations to avoid false conclusions in proteomic studies using differential in gel electrophoresis. *Molecular and Cellular Proteomics*, **6**, 1354–1364.

Karp, N. A., & Lilley, K. S. 2005. Maximising sensitivity for detecting changes in protein expression: Experimental design using minimal CyDyes. *Proteomics*, **5**(12), 3105–3115.

Karp, N. A., Griffin, J. L., & Lilley, K. S. 2005b. Application of partial least squares discriminant analysis to two-dimensional difference gel studies in expression proteomics. *Proteomics*, **5**(1), 81–90.

Karp, N. A., Feret, R., Rubtsov, D. V., & Lilley, K. S. 2008. Comparison of DIGE and post-stained gel electrophoresis with both traditional and SameSpots analysis for quantitative proteomics. *Proteomics*, **8**(5), 948–960.

Kays, S.J. 1991. *Postharvest physiology of perishable plant products.* Chapman and Hall, London, UK. Chap. Stress in harvested products, pages 335–393.

Ke, D., Mateos, M., Siriphanich, J., Li, C., & Kader, A. 1993. Dioxide action on metabolism of organic and amino acids in crisphead lettuce. *Postharvest Biology and Technology*, **3**, 235–247.

Kerbel, E., Kader, A., & R., Romani. 1990. Respiratory and glycolytic response of suspension-cultured Passe Crassane pear fruit cells to elevated CO_2 concentrations. *Journal of the American Society of Horticultural Sciences*, **115**, 111–114.

Kim, M., Ahn, J. W., Jin, U. H., Choi, D., Paek, K. H., & Pai, H. S. 2003. Activation of the programmed cell death pathway by inhibition of proteasome function in plants. *Journal of Biological Chemistry*, **278**(21), 19406–19415.

Kjaersgard, I. V. H., Norrelykke, M. R., & Jessen, F. 2006. Changes in cod muscle proteins during frozen storage revealed by proteome analysis and multivariate data analysis. *Proteomics*, **6**(5), 1606–1618.

Knee, M. 1993. *Biochemistry of fruit ripening.* Chapman and Hall, London, UK. Chap. Pome fruit.

Krogh, M., Fernandez, C., Teilum, M., Bengtsson, S., & James, P. 2007. A probabilistic treatment of the missing spot problem in 2D gel electrophoresis experiments. *Journal of Proteome Research*, **6**(8), 3335–3343.

Kruger, N.J., & von Schaewen, A. 2003. The oxidative pentose phosphate pathway: structure and organization. *Current Opinion in Plant Biology*, **6**, 236–246.

Kumar Yadav, S., Singla-Pareek, S., Reddy, M.K., & Sopory, S.K. 2005. Transgenic tobacco plants overexpressing glyoxalase enzymes resist an increase in methylglyoxal and maintain higher reduced glutathione levels under salinity stress. *FEBS Letters*, **579**, 6265–6271.

Kuo, S.J., & Parkin, L. 1989. Chilling injury in cucumbers associated with lipid peroxidation as measured by ethane evolution. *Journal of Food Science*, **54**, 1488–1491.

Kussmann, M., Raymond, F., & Affolter, M. 2006. OMICS-driven biomarker discovery in nutrition and health. *Journal of Biotechnology*, **124**, 758–787.

Kwade, Z., Swiatek, A., Azmi, A., Goossens, A., Inze, D., Van Onckelen, H., & Roef, L. 2005. Identification of four adenosine kinase isoforms in tobacco BY-2 cells and their putative role in the cell cycle-regulated cytokinin metabolism. *Journal of Biological Chemistry*, **280**(17), 17512–17519.

Laborde, B., Moine-Ledoux, V., Richard, T., Saucier, C., Dubourdieu, D., & Monti, J. P. 2006. PVPP polyphenol complexes: A molecular approach. *Journal of Agricultural and Food Chemistry*, **54**(12), 4383–4389.

Lammertyn, J., Aerts, M., Verlinden, B. E., Schotsmans, W., & Nicolaï, B. M. 2000. Logistic regression analysis of factors influencing core breakdown in Conference pears. *Postharvest Biology and Technology*, **20**(1), 25–37.

Lammertyn, J., Scheerlinck, N., Jancsok, P., Verlinden, B. E., & Nicolaï, B. M. 2003a. A respiration-diffusion model for Conference pears I: model development and validation. *Postharvest Biology and Technology*, **30**(1), 29–42.

Lammertyn, J., Scheerlinck, N., Jancsok, P., Verlinden, B. E., & Nicolaï, B. M. 2003b. A respiration-diffusion model for Conference pears II. Simulations and relation to core breakdown. *Postharvest Biology and Technology*, **30**(1), 43–55.

Lane, C.S. 2005. Review Mass spectrometry-based proteomics in the life sciences. *Cell Molecular and Life Sciences*, **62**, 848–869.

Larrigaudiere, C., Lentheric, I., Pinto, E., & Vendrell, M. 2001. Short-term effects of air and controlled atmosphere storage on antioxidant metabolism in conference pears. *Journal of Plant Physiology*, **158**(8), 1015–1022.

Lee, J., Kim, S., Lee, J.Y., Yoo, E., Cheoul, M., Cho, M., Kim, B., & Bahk, Y. 2006. A differentially expressed proteomic analysis in placental tissues in relation to pungency during the pepper fruit development. *Proteomics*, **6**, 5248–5259.

Levin, A., Lurie, S., Zutkhi, Y., & Ben-Arie, R. 1995. Physiological effects of controlled atmosphere storage on Fiesta Red nectarines. *Acta Horticulturae*, **379**, 121–127.

Little, R.J., & Rubin, D.B. 1987. *Statistical analysis with missing data.* John Wiley and Sons, New York, USA.

Liu, H., Sadygov, R.G., & Yates, J.R. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry*, **76**, 4193–4201.

Loomis, W. D., & Battaile, J. 1966. Plant phenolic compounds and isolation of plant enzymes. *Phytochemistry*, **5**(3), 423–438.

Lougheed, E. C. 1987. Interactions of oxygen, carbon-dioxide, temperature, and ethylene that may induce injuries in vegetables. *Hortscience*, **22**(5), 791–794.

Lu, P., Vogel, C., Wang, R., Yao, X., & Marcotte, E. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology*, **25**, 117–124.

Lurie, S. 1992. Controlled atmosphere storage to decrease physiological disorders in nectarines. *International Journal of Food Science and Technology*, **27**, 507–514.

Lurie, S., & Crisosto, C. H. 2005. Chilling injury in peach and nectarine. *Postharvest Biology And Technology*, **37**(3), 195–208.

Majoul, T., Bancel, E., Triboi, E., Ben Hamida, J., & 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), 175–183.

Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A., & Hol, W. G. J. 1994. Crystal-structure of recombinant human triosephosphate isomerase at 2.8 angstrom resolution - triosephosphate isomerase-related human genetic-disorders and comparison with the trypanosomal enzyme. *Protein Science*, **3**(5), 810–821.

Marengo, E., Robotti, E., Cecconi, D., Hamdan, M., Scarpa, A., & Righetti, P. G. 2004. Identification of the regulatory proteins in human pancreatic cancers treated with trichostatin A by 2D-PAGE maps and multivariate statistical analysis. *Analytical and Bioanalytical Chemistry*, **379**(7-8), 992–1003.

Marengo, E., Robotti, E., Bobba, M., Liparota, M. C., Rustichelli, C., Zamoo, A., Chilosi, M., & Righetti, P. G. 2006. Multivariate statistical tools applied to the characterization of the proteomic profiles of two human lymphoma cell lines by two-dimensional gel electrophoresis. *Electrophoresis*, **27**(2), 484–494.

Martin, R.L., & Brancia, F.L. 2003. Analysis of high mass peptides using a novel matrix assisted laser desorption/ionization quadrupole ion trap time of flight mass spectrometer. *Rapid Communications in Mass Spectrometry*, **17**, 1358–1365.

Martinus, R. D., Ryan, M. T., Naylor, D. J., Herd, S. M., Hoogenrad, N. J., & Hoj, P. B. 1995. Role of chaperones in the biogenesis and maintenance of the mitochondrion. *FASEB Journal*, **9**(5), 371–378.

Mathesius, U., Imin, N., Chen, H.C., Djordjevic, M.A., J.J., Weinmann, Natera, S.H., Morris, A.C., Kerim, T., Paul, S., Menzel, C., Weiller, G.R., & Rolfe, B.G. 2002. Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. *Proteomics*, **2**, 1288–1303.

Mattheis, J.P., Buchanan, D.A., & Fellman, J.K. 1997. Volatile constituents of Bing sweet cherry fruit following controlled atmosphere storage. *Journal of Agricultural and Food Chemistry*, **45**, 212–216.

Meleth, S., Deshane, J., & Kim, H. 2005. The case of well conducted experiments to validate statistical protocols for 2D gels: different preprocessing=different lists of significant proteins. *BMC Biotechnology*, **5**, 1–15.

Millar, A.H., Mittova, V., Kiddle, G., Heazlewood, J. L., Bartoli, C.G., Theodorou, M.E., & Foyer, C.H. 2003. Control of ascorbate synthesis by respiration and its implications for stress responses. *Plant Physioloy*, **133**, 443–447.

Miller, I., Crawford, J., & Gianazza, E. 2006. Protein stains for proteomic applications: Which, when, why? *Proteomics*, **6**(20), 5385–5408.

Mittler, R. 2006. Abiotic stress, the field environment and stress combination. *Trends in Plant Science*, **11**(1), 15–19.

Mohammed, M., & Wickham, L.D. 1997. Ocurrence of chilling injury in golden apples (Spondias dulcis Sonn.). *Journal of Food Quality*, **20**, 91–104.

Moran, J. F., Becana, M., Iturbeormaetxe, I., Frechilla, S., Klucas, R. V., & Apariciotejo, P. 1994. Drought induces oxidative stress in pea plants. *Planta*, **194**(3), 346–352.

Morzel, M., Chambon, C., Hamelin, M., Lhoutellier, S., Sayd, T., & Monin, G. 2004. Proteome changes during pork meat ageing following use of two different pre-slaughter handling procedures. *Meat Science*, **67**, 689–696.

Narita, Y., Taguchi, H., Nakamura, T., Ueda, A., Shi, W. M., & Takabe, T. 2004. Characterization of the salt-inducible methionine synthase from barley leaves. *Plant Science*, **167**(5), 1009–1016.

Nelson, P. R. C., Taylor, P. A., & MacGregor, J. F. 1996. Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemometrics and Intelligent Laboratory Systems*, **35**(1), 45–65.

Nesvizhskii, A., Vitek, O., & Aebersold, R. 2007. Analysis and validation of proteomic data generated by tandem mass spectromety. *Nature Methods*, **4**, 787–797.

Neuhoff, V., Arold, N., Taube, D., & 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), 255–262.

Norden, B., Broberg, P., Lindberg, C., & Plymoth, A. 2005. Analysis and understanding of high-dimensionality data by means of multivariate data analysis. *Chemistry and Biodiversity*, **2**(11), 1487–1494.

Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., & Ishii, S. 2003. A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, **19**(16), 2088–2096.

O'Farrell, P.Z., Goodman, H.M., & O'Farrell, P.H. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell*, **12**, 1133–1142.

Olsson, I., Larsson, K., Palmgren, R., & Bjellqvist, B. 2002. Organic disulfides as a means to generate streak free two dimensional maps with narrow range basic immobilized pH gradient strips as first product dimension. *Proteomics*, **2**, 1630–1632.

Patton, W.F. 2002. Detection technologies in proteome analysis. *Journal of Chromatography B*, **771**, 3–31.

Pedreschi, R., Vanstreels, E., Carpentier, S., Hertog, M., Lammertyn, J., Robben, J., Noben, J. P., Swennen, R., Vanderleyden, J., & Nicolaï, B. M. 2007. Proteomic analysis of core breakdown disorder in Conference pears (*Pyrus communis* L.). *Proteomics*, **7**(12), 2083–2099.

Pedreschi, R., Hertog, M., Robben, J., Noben, J.P., & Nicolaï, B. A. 2008a. Physiological implications of controlled atmosphere storage on Conference pear (*Pyrus communis* L.): a proteomics approach. *Postharvest Biology and Technology*, **50**, 110–116.

Pedreschi, R., Hertog, M. L. A. T. M., Carpentier, S. C., Lammertyn, J., Robben, J., Noben, J. P., Panis, B., Swennen, R., & Nicolaï, B. M. 2008b. Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. *Proteomics*, **8**(7), 1371–1383.

Pedreschi, R., Franck, C., Lammertyn, J., Erban, J., Kopka, J., Hertog, M., & Nicolaï, B. A. 2009a. Metabolic profiling of Conference pears under low oxygen stress. *Postharvest Biology and Technology*, **51**, 123–130.

Pedreschi, R., Hertog, M., Lilley, K. S., & Nicolaï, B. A. 2009b. Proteomics for the food industry: opportunities and challenges. *Critical Reviews in Food Science and Nutrition*, accepted.

Pedreschi, R., Hertog, M., Robben, J., Lilley, K. S., Karp, N., Baggerman, G., Vanderleyden, J., & Nicolaï, B. A. 2009c. Proteomics-gel based approach to study metabolic changes in pear tissue during storage. *Journal of Agricultural and Food Chemistry*, submitted.

Perreault, A., Bellemer, C., & Bachand, F. 2008. Nuclear export competence of pre-40S subunits in fission yeast requires the ribosomal protein RSP2. *Nucleic Acids Research*, **36**, 6132–6142.

Pierpoint, W. 2004. *Methods in Molecular Biology.* Humana Press, Totowa. Pages 65–74.

Plaxton, W.C. 1996. The organization and regulation of plant glycolysis. *Annual Reviews in Plant Physiology and Molecular Biology*, **47**, 185–214.

Pons, C., Royo, C., Forment, J., Gadea, J., Lluch, Y., Granell, A., Kanellis, A.K., Zacarias, L., & Lafuente, M. 2005. A customized citrus cDNA microarray and its use in postharvest. *Acta Horticulturae*, **682**, 225–232.

Ramon, M., Geuns, J., Swennen, R., & Panis, B. 2002. Polyamines and fatty acids in sucrose precultured banana meristems and correlation with survival rate after cryopreservation. *CryoLetters*, **23**, 345–352.

Righetti, P.G., & Bossi, A. 1997. Isoelectric focusing in immobilized pH gradients: an update. *Journal of Chromatography B*, **699**, 77–89.

Roberts, M., & Bruxelles, G. 2002. Plant 14-3-3 protein families: evidence for isoform specific mutations. *Biochemical Society Transactions*, **30**, 373–378.

Rocco, M., D'Ambrosio, C., Arena, S., Faurobert, M., Scaloni, A., & Marra, M. 2006. Proteomic analysis of tomato fruits from two ecotypes during ripening. *Proteomics*, **6**(13), 3781–3791.

Roe, M., & Griffin, T. 2006. Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics*, **6**, 4678–4687.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamsom, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., & Pappin, D. 2004. Multiplexed protein quantification in *Saccharomices cerevisae* using amine reactive isobaric tagging reagents. *Molecular and Cell Proteomics*, **3**, 1154–1169.

Sadowski, P., Groen, A., Dupree, P., & Lilley, K. 2008. Sub-cellular localization of membrane proteins. *Proteomics*, **8**, 3991–4011.

Sala, J. M., Sanchez-Ballesta, M. T., Alferez, F., Mulas, M., Zacarias, L., & Lafuente, M. T. 2005. A comparative study of the postharvest performance of an ABA-deficient mutant of oranges II. Antioxidant enzymatic system and phenylalanine ammonia-lyase in non-chilling and chilling peel disorders of citrus fruit. *Postharvest Biology And Technology*, **37**(3), 232–240.

Salekdeh, H., Siopongco, J., Wade, L., Ghareyazie, B., & Benett, J. 2002. Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics*, **2**, 1131–1145.

Salveit, M.E. 2003. Is it possible to find an optimal controlled atmosphere? *Postharvest Biology and Technology*, **27**, 3–13.

Saquet, A. A., Streif, J., & Bangerth, F. 2003. Energy metabolism and membrane lipid alterations in relation to brown heart development in Conference pears during delayed controlled atmosphere storage. *Postharvest Biology and Technology*, **30**(2), 123–132.

Saravanan, R. S., & Rose, J. K. C. 2004. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics*, **4**(9), 2522–2532.

Savenstrand, H., & Strid, A. 2004. A *Pisum sativum* glyoxysomal malate dehydrogenase induced by cadmium exposure. *DNA Sequence*, **15**(3), 206–208.

Shapiro, S., & Wilk, M. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.

Shen, S. H., Jing, Y. X., & Kuang, T. Y. 2003. Proteomics approach to identify wound-response related proteins from rice leaf sheath. *Proteomics*, **3**(4), 527–535.

Shevchenko, A., Wilm, M., Vorm, O., & Mann, M. 1996. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Analytical Chemistry*, **68**(5), 850–858.

Shevchenko, A., Sunyaev, S., Loboda, A., Shevehenko, A., Bork, P., Ens, W., & Standing, K.G. 2001. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. *Analytical Chemistry*, **73**, 1917–1926.

Shulaev, V., Cortes, D., Miller, G., & Mittler, R. 2008. Metabolomics for plant stress response. *Physiologia Plantarum*, **132**, 199–208.

Siegel, S.C. 1988. *Non parametric statistics for behavioral sciences.* McGraw Hill Book Company, Singapore.

Sommer, A., Fischer, P., Krause, K., Boettcher, K., Brophy, P.M., Walter, R.D., & Liebau, E. 2001. A stress-responsive glyoxalase I from the parasitic nematode *Onchocerca volvulus*. *Biochemical Journal*, **353**, 445–452.

Steen, H., & Mann, M. 2004. The ABCs (and XYZs) of peptide sequencing. *Nature Reviews Molecular and Cell Biology*, **5**, 699–711.

Storey, J.D., & Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academic of Sciences of the United States of America*, **100**, 9440–9445.

Sullivan, J., Shirasu, K., & Deng, X. 2003. The diverse roles of ubiquitin and the 26S proteasome in the life of plants. *Nature Reviews Genetics*, **4**, 948–958.

Sweetlove, L. J., Heazlewood, J. L., Herald, V., Holtzapffel, R., Day, D. A., Leaver, C. J., & Millar, A. H. 2002. The impact of oxidative stress on Arabidopsis mitochondria. *Plant Journal*, **32**(6), 891–904.

Tan, S., Tan, H.T., & Chung, M. 2008. Membrane proteins and membrane proteomics. *Proteomics*, **8**, 3924–3932.

Tecsi, L. I., Smith, A. M., Maule, A. J., & Leegood, R. C. 1996. A spatial analysis of physiological changes associated with infection of cotyledons of marrow plants with cucumber mosaic virus. *Plant Physiology*, **111**(4), 975–985.

Theodorou, M.E., & Plaxton, W.C. 1995. *Environment and plant metabolism: Flexibility and acclimation*. BIOS Scientific Publishers, London, UK. Chap. Adaptations of plant respiratory metabolism to nutritional phosphate deprivation, pages 70–109.

Thompson, J.E. 2008. eIF5A is involved in pathogen-induced cell death and development of disease symptoms in *Arabidopsis thaliana*. *Plant Physiology*, **148**, 479–489.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., & Davison, M. 2001. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, **1**(3), 377–396.

Tonutti, P., Beghelo, M., & Bonghi, C. 2008. Use of genomic tools to study peach fruit postharvest physiology. *Acta Horticulturae*, **804**, 77–84.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics*, **17**(6), 520–525.

Tsunezuka, H., Fujiwara, M., Kawasaki, T., & Shimamoto, K. 2005. Proteome analysis of programmed cell death and defense signaling using

the rice lesion mimic mutant cdr2. *Molecular Plant Microbe Interactions*, **18**(1), 52–59.

Tuomainen, M. H., Nunan, N., Lehesranta, S. J., Tervahauta, A. I., Hassinen, V. H., Schat, H., Koistinen, K. M., Auriola, S., McNicol, J., & Karenlampi, S. O. 2006. Multivariate analysis of protein profiles of metal hyperaccumulator Thlaspi caerulescens accessions. *Proteomics*, **6**(12), 3696–3706.

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

Urfer, W., Grzegorczyk, M., & Jung, K. 2006. Statistics for proteomics: a review of tools for analyzing experimental data. *Proteomics Practical Applications*, **1**, 48–55.

Van Breusegem, F., Vranova, E., Dat, J. F., & Inze, D. 2001. The role of active oxygen species in plant signal transduction. *Plant Science*, **161**(3), 405–414.

Veltman, R. H., Kho, R. M., van Schaik, A. C. R., Sanders, M. G., & Oosterhaven, J. 2000. Ascorbic acid and tissue browning in pears (*Pyrus communis* L. cvs Rocha and Conference) under controlled atmosphere conditions. *Postharvest Biology and Technology*, **19**(2), 129–137.

Veltman, R. H., Lentheric, I., Van der Plas, L. H. W., & Peppelenbos, H. W. 2003. Internal browning in pear fruit (*Pyrus communis* L. cv Conference) may be a result of a limited availability of energy and antioxidants. *Postharvest Biology and Technology*, **28**(2), 295–302.

Verboven, P., Kerckhofs, G., Mebatsion, H.K., Ho, Q.T., Temst, K., Wevers, M., Cloetens, P., & Nicolaï, B.M. 2008. 3-D exchange pathways in pome fruit characterised by synchrotron X-ray computed tomography. *Plant Physiology*, **147**, 518–527.

Verlinden, B.E., de Jager, A., Lammertyn, J., Schotsmans, W., & Nicolaï, B.M. 2002. Effect of harvest and delaying controlled at-

mosphere storage conditions on core breakdown incidence in Conference pears. *Biosystems Engineering*, **83**, 339–347.

Vikram, A., Prithiviraj, B., & Hamzehzarghani, H. 2004. Volatile metabolite profiling to discriminate diseases of McIntosh apple inoculated with fungal pathogens. *Journal of the Food Science and Agriculture*, **84**, 1333–1340.

Vincent, D., Wheatley, M. D., & Cramer, G. R. 2006. Optimization of protein extraction and solubilization for mature grape berry clusters. *Electrophoresis*, **27**(9), 1853–1865.

Vranova, E., Inze, D., & Van Breusegem, F. 2002. Signal transduction during oxidative stress. *Journal of Experimental Botany*, **53**(372), 1227–1236.

Wang, W., Scali, M., Vignani, R., Spadafora, A., Sensi, E., Mazzuca, S., & Cresti, M. 2003. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis*, **24**(14), 2369–2375.

Washburn, M.P., Wolters, D., & Yates, J.R. 2001. Large scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, **19**, 242–247.

Wayne, F. 2005. Polygalacturonase-inhibiting protein activity in cantaloupe fruit as a function of fruit maturation and tissue origin. *Journal of Plant Pathology*, **111**, 67–76.

Wills, R., McGlasson, B., Graham, D., & Joyce, D. 1998. *Postharvest, an introduction to the physiology and handling of fruit, vegetables and ornamentals.* CAB Interntional, Wallingford Oxon.

Wilm, M., & Mann, M. 1994. Electrospray and taylor cone theory, doles beam of macromolecules at a last. *International Journal of Mass Spectrometry*, **136**, 137–180.

Wold, S., Esbensen, K., & Geladi, P. 1987. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**(1-3), 37–52.

Wollnick, H. 1993. Time of flight mass analyzers. *Mass Spectrometry Reviews*, **12**, 89–114.

Wood, J., White, I. R., & Cutler, P. 2004. A likelihood-based approach to defining statistical significance in proteomic analysis where missing data cannot be disregarded. *Signal Processing*, **84**(10), 1777–1788.

Xia, Q., Hendrickson, E.L., Wang, T., Lamont, R.J., Leigh, J.A., & Hackett, M. 2007. Review: Protein abundance ratios for global studies of prokaryotes. *Proteomics*, **7**, 2904–2919.

Xuan, H., Streif, J., Pfeffer, H., Dannel, F., Romheld, V., & Bangerth, F. 2001. Effect of pre-harvest boron application on the incidence of CA-storage related disorders in Conference pears. *Journal of Horticultural Sciences and Biotechnology*, **76**, 133–137.

Yan, S. P., Tang, Z. C., Su, W., & Sun, W. N. 2005. Proteomic analysis of salt stress-responsive proteins in rice root. *Proteomics*, **5**(1), 235–244.

Yang, Y., Thannhauser, T.W., Li, L., & Zhang, S. 2007. Development of an integrated approach for evaluation of 2-D gel image analysis: impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow. *Electrophoresis*, **28**, 2080–2094.

Yang, Z., Floyd, D., Loeber, G., & Tong, L. 2000. Structure of a closed form of human malic enzyme and implications for catalytic mechanism. *Nature Structural Biology*, **7**, 251–257.

Yergey, A.L., Coorsen, J.R., Backlund, P.S., Blank, P.S., Humphrey, G.A., Zimmerberg, J., Campbell, J.M, & Vestal, M.L. 2002. De novo sequencing of peptides using MALDI/TOF-TOF. *Journal of the American Society of Mass Spectrometry*, **13**, 784–791.

Yoon, D., Lee, E., & Park, T. 2007. Robust imputation method for missing values in microarray data. *BMC Bioinformatics*, **8**, S6.

Zerbini, P.E., Rizzolo, A., Brambilla, A., & Grassi, M. 2002. Loss of ascorbic acid during storage of Conference pears in relation to the appearance of brown heart. *Journal of the Science of Food and Agriculture*, **82**, 1–7.

Zhou, H.W., Lurie, S., Lers, A., Khatchitski, A., Sonego, L., & Ben Arie, R. 2000. Delayed storage and controlled atmosphere storage of nectarines: two strategies to prevent woolliness. *Postharvest Biology and Technology*, **18**, 133–141.

Zorb, C., Schmitt, S., Neeb, A., Karl, S., Linder, M., & Schubert, S. 2004. The biochemical reaction of maize (*Zea mays* L.) to salt stress is characterized by a mitigation of symptoms and not by a specific adaptation. *Plant Science*, **167**(1), 91–100.

List of publications

International journal publications

Campos, D., Betalleluz, I., Tauquino, R., Chirinos, R., Pedreschi, R. 2009. Nutritional and functional characterization of Andean Chicuru (*Stangea Rhizanta* Sinom henrici). *Food Chemistry*, 112, 63−70.

Chirinos, R., Betalleluz, I., Huaman, A., Arbizu, C., Pedreschi, R., Campos, D. 2009. HPLC-PAD characterization of phenolic compounds from Andean oca (*Oxalis tuberosa* Mol.) tubers and their contribution to the antioxidant capacity. *Food Chemistry*, 113, 1423−1251.

Pedreschi, F., Travisany, X., Reyes, C., Troncoso, E., Pedreschi, R. 2009. Kinetics of reducing sugar extraction during blanching of potato slices. *Journal of Food Engineering*, 91, 443−447.

Pedreschi, R., Franck, C., Lammertyn, J., Erban, A., Kopka, J., Hertog, M., Nicolaï, B.M. 2009. Metabolic profiling of Conference pears under low oxygen stress. *Postharvest Biology and Technology*, 51, 123−130.

Pedreschi, R., Hertog, M., Robben, J., Karp, N., Lilley, K., Baggerman, G., Vanderleyden, J., Nicolaï, B. M. 2009. Proteomics-gel based approach to study metabolic changes in pear tissue during storage *Journal of Agricultural and Food Chemistry*, submitted.

Pedreschi, R., Hertog, M., Lilley, K., Nicolaï, B. M. 2009. Proteomics for the food industry: opportunities and challenges. *Critical Reviews in Food Science and Nutrition*, accepted.

Campos, D., Chirinos, R., Pedreschi, R. 2008. Extraction, recovery and purification of natural antioxidants from Andean crops. *Electronic Journal of Environmental, Agricultural and Food Chemistry*, 7, 3150−3154.

Chirinos, R., Campos, D., Costa, N., Arbizu, C., Pedreschi, R., Larondelle, Y. 2008. Phenolic profiles of Andean mashua (*Tropaeolum tuberosum* Ruiz and Pavon) tubers: identification by HPLC-DAD and evaluation of their antioxidant activity. *Food Chemistry*, 106, 1285−1298.

Chirinos, R., Warnier, M., Campos, D., Pedreschi, R., Rees, J., Larondelle, Y. 2008. Antioxidant properties of mashua (*Tropaeolum tuberosum* Ruiz and Pavon) phenolic extracts against oxidative damage using biological in vitro assays. *Food Chemistry*, 117, 98−105.

Pedreschi, R., Hertog, M., Carpentier, S., Robben, J., Noben, J.P., Lammertyn, Panis, B., Swennen, R., Nicolaï, B. 2008. Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. *Proteomics*, 8, 1371−1383.

Pedreschi, R., Hertog, M., Robben, J., Noben, J.P., Nicolaï, B. M. 2008. Physiological implications of controlled atmosphere storage of Conference pears (*Pyrus communis* L.): a proteomic approach. *Postharvest Biology and Technology*, 50, 110−116.

Chirinos, R., Rogez, H., Campos, D., Pedreschi, R., Larondelle, Y. 2007. Antioxidant phenolic compounds in mashua (*Tropaeolum tuberosum* Ruiz and Pavon) tubers: study of extraction conditions. *Separation and Purification Technology*, 55, 217−225.

Pedreschi, F., Leon, J., Mery, D., Moyano, P., Pedreschi, R., Kaack, K., Granby, K. 2007. Color development and acrylamide content of pre-dried potato chips. *Journal of Food Engineering*, 79, 786−793.

Pedreschi, F., Moyano, P., Santis, N., Pedreschi, R. 2007. Physical properties of pre-treated potato chips. *Journal of Food Engineering*, 79, 1474−1482.

Pedreschi, R., Cisneros-Zevallos, L. 2007. Phenolic profiles of Andean purple corn (*Zea mays* L.). *Food Chemistry*, 100, 956−963.

Pedreschi, R., Vanstreels, E., Carpentier, S., Robben, J., Noben, J.P., Hertog, M., Lammertyn, J., Vanderleyden, J., Swennen, R., Nicolaï, B. M. 2007. Proteomic analysis of core breakdown disorder in Conference pears (*Pyrus communis* L). *Proteomics*, 7, 2083−2099.

Chirinos, R., Campos, D., Betalleluz, I., Giusti, M., Schwartz, S., Tian, Q., Pedreschi, R., Larondelle, Y. 2006. HPLC-DAD/HPLC-MS profiling of anthocyanins from Andean mashua tuber (*Tropaeolum tuberosum* Ruiz and Pavon). *Journal of Agricultural and Food Chemistry*, 54, 7989−7097.

Pedreschi, R., Cisneros-Zevallos, L. 2006. Antioxidant and antimutagenic activities of phenolic fractions from a purple corn extract. *Journal of Agricultural and Food Chemistry*, 54, 4557−4567.

Pedreschi, R., Campos, D., Noratto, G., Chirinos, R., Cisneros-Zevallos, L. 2003. Andean yacon root (*Smallanthus sonchifolius* Poepp. Endl) fructooligosaccharides as a potential novel source of prebiotics. *Journal of Agricultural and Food Chemistry*, 51, 5278−5284.