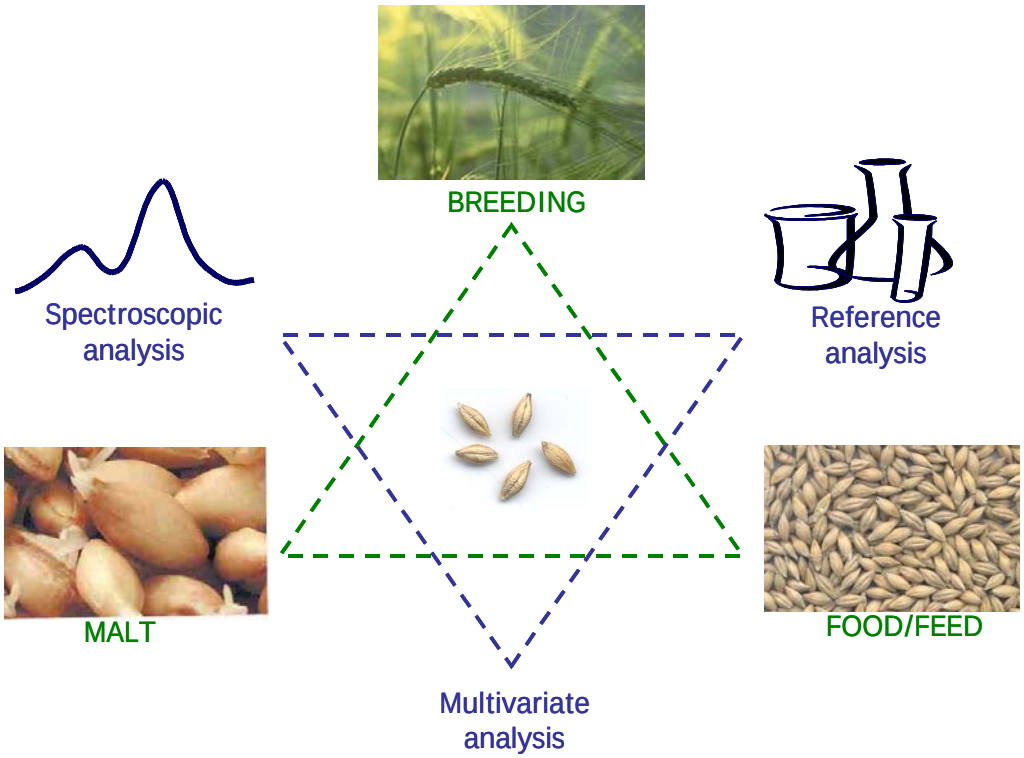
Screening analyses for quality criteria in barley

Predicting germinative and physical-chemical properties
by spectroscopy evaluated by multivariate data analysis



$$\mathbf{X} = TP' + E$$

Ph.D. thesis by

Birthe Pontoppidan Møller
M.Sc. Agricultural Science

Food Technology
Department of Food Science
The Royal Veterinary and Agricultural University
Rolighedsvej 30, 1958 Frederiksberg C
Denmark

THE ROYAL VETERINARY AND AGRICULTURAL UNIVERSITY

Screening analyses for quality criteria in barley

Predicting germinative and physical-chemical properties by spectroscopy evaluated by multivariate data analysis

Ph.D. thesis by
Birthe Pontoppidan Møller
M.Sc. Agricultural Science

Supervisor
Professor Lars Munck
Food Technology
Department of Food Science
The Royal Veterinary and Agricultural University
Rolighedsvej 30
DK-1958 Frederiksberg C

Screening analyses for quality criteria in barley - Predicting germinative and physical-chemical properties by spectroscopy evaluated by multivariate data analysis

ISBN 87-7611-042-7

Printed by Samfundslitteratur Grafik, Frederiksberg, Denmark

Preface

This thesis is based on a series of research projects conducted at The Royal Veterinary and Agricultural University (KVL).
I am greatly indebted to my enthusiastic and inspiring supervisor Professor Lars Munck without whom I would never have reached so far. I am also grateful to him for giving me the opportunity to do this work and present me to the way of multivariate thinking.
The staffs at Food Technology and at Plant Food Science are thanked for their scientific and social support. Jesper Pram Nielsen, Lars Nørgaard, Elisabeth Micklander, Lisbeth Hansen and Gilda Kischinovsky are warmly appreciated for their big help and support. Also the former employees Kirsten Wilms, Anna Marie Hansen and Karen Thomsen together with the student helpers Helene Fast-Seefeldt, Anne-Marie Jacobsen and Marie Kaas are acknowledged for their crucial assistance in everything from harvest of plants to chemical analyses.

The work with high lysine barley has been made in an inspiring cooperation with Susanne Jacobsen and Ib Søndergaard, BioCentrum, DTU to whom I am grateful. Liljana Nesic and Marianne Petersen are acknowledged for their assistance in making 2-DE and amino acid determinations.

Jens Bertelsen, Danny Klysner and the rest of the staff at Højbakkegård, KVL are acknowledged for their help performing field and green house experiments. Also Monica Frank, AgroLab and Skånska Lantmännen together with Fyens Stifts Patriotiske Selskab are appreciated for collecting samples and making field experiments.

The research presented in this thesis originates from different research projects. The Skånska Lantmännen Foundation with Lennart Wikström initiated the research on malting barley and are much appreciated. The participants in The Danish Cereal Network and in the two EU-projects “Qualigrain” and “Cascade” are highly acknowledged for inspiring discussions and meetings as well as financial support. Also Thomas Börjesson, ODAL as well as the Centre for New Sensor Systems for the Measurement of Food Quality, Erhvervsfremmestyrelsen are much appreciated for the required financial support during the period of my work.

Finally collegues, friends and family are warmly acknowledged for their big patience with me during this work.

Copenhagen, March 2004

Birthe Pontoppidan Møller

Summary

Barley screening analyses for quality parameters developed by the use of pattern recognition multivariate data analysis (chemometrics) of traditional data sets and instrumental analyses have shown the following results.

- Earlier developed heat-stress germination models for prediction of vigour potential (VP) in malting barley based on statistical distributional assumptions should be re-evaluated. A new class of heat resistant barley was found by Principal Component Analysis (PCA) (Papers I and II) where the germination decay curve with heat does not follow the classic statistical model.

- Germination percentage day 1 is a more convenient and sensitive estimate for germination speed or "vigour" than the germination index (GI). They are highly correlated (r=0.99). "Viability" - live seeds - can be expressed by germination percentage from day 3 or by using the Tetrazolium test to determine the number of living grains. By plotting "vigour" against "viability" in a two-way germinative classification plot, barley can be classified predicting two malting quality parameters (extract% and ß-glucan in wort mg/L) in seven groups with high precision (Paper III). This tool should be useful as a new early selection criterion for malt quality in industry.

- Barley seeds with "viability" > 92% was analysed by two different set of measurements for physical-chemical seed characteristics (Near Infrared Transmission spectroscopy (NIT) and ten physical-chemical parameters). It was found that "vigour" and extract and ß-glucan in wort can be predicted independently by the two methods by Partial Least Square Regression (PLSR) analysis (Papers III and IV). The physical-chemical seed structure analysed by these two instrumental methods is suggested to be the limiting factor for the accessability of substrate to the embryo for growth. This function is also related to malt modification. Spectroscopic calibrations to "vigour" should be further developed as a tool to predict "vigour" by using NIT for inspecting the malting barley *on-line* as a first selection of promising material already at the farmer/elevator level. The lines selected by NIT should be checked for "viability" by germination three days or by the Tetrazolium embryo staining test. A malting barley quality evaluation with the germinative classification plot (Paper III) should then be performed combining separate estimates for "vigour" and "viability".

- It is possible to detect and define previously unknown parameters with regard to the physical-chemical diversity of the barley endosperm, by using Near Infrared Reflection spectroscopy (NIR) evaluated by chemometrics (Papers V-VII). A new class of high ß-glucan/low starch mutants was found (Paper VI). The very high ß-glucan mutants have a potential use in dietary foods. The investigation establishes a basis for a further analysis by

genetics and biotechnology to define gene function and expression introducing the concept of the digitized (spectroscopic) phenome (Paper VII).

- Pattern recognition multivariate data analysis (PCA and PLSR) of datasets with relevance in plant breeding, biotechnology and in the malting and brewing industry have been useful in establishing new often surprising results (Papers I-VII). End users working with instruments and chemical and malting analyses should be able to make their own classifications and calibrations, visually to identify quality outliers in a graphic display by PCA. Such identification allows for selection of high quality malting barley for the industry and is shown to be a helpful tool in plant breeding to correct the gene background for a negative genetic covariate quality complex to a high-lysine mutant gene (Paper VII) which is based on a multigene family. Instrumental data evaluated by chemometrics should allow "data breeding" as an innovative interaction between plant breeding in the field on one hand and selection of barley lines by their spectroscopic data in a PCA plot on the other (Paper VII).

Resumé

Anvendelsen af multivariat data analyse (kemometri) på traditionelle og spektroskopiske datasæt til estimering af bygkvalitet har givet følgende resultater.

- Den tidligere udviklede spiringsmodel til forudsigelse af vitalitet ved brug af varmestress bør revurderes. En ny klasse af varmeresistente bygprøver blev fundet med Principal Component Analyse (PCA) af spiringskurverne som var opnået ved varmebehandling. Disse fulgte ikke den klassisk statistiske model med sigmoid kurveform (Artikel I og II).

- Spiringsprocenten dag 1 er et mere følsomt og praktisk estimat for vitalitet (spiringshastighed) end spiringsindeks (GI). Disse er stærkt korrelerede (r=0.99). Viabiliteten (antal levende kerner) kan udtrykkes ved at bestemme spiringsprocenten dag 3 eller anvende Tetrazolium testen (farvning af levende kerner). Ved at plotte vitalitet mod viabilitet i et spirings-klassifikations plot, kan bygprøver klassificeres i syv grupper i forhold til to maltkvalitetsparametre (ekstrakt% og ß-glukan i urt mg/L) (Artikel III). Denne metode vil kunne bruges i industrien som en tidlig selektion af malt kvalitet.

- Ved at analysere byg med viabilitet >92% ved analyse med to forskellige metoder til bestemmelse af fysisk-kemiske kernekarakteristika (NIT spektroskopi og ti fysisk-kemiske parametre) er det muligt, at forudsige vitaliteten (udtrykt som spiring dag 1) og maltkvalitet-kriterierne ekstrakt og ß-glukan i urt uafhængigt af hinanden ved brug af Partial Least Square Regression (PLSR) (Artikel III og IV). Den fysisk-kemiske kernestruktur, som udtrykkes ved de to ovennævnte metoder, foreslåes, at være den begrænsende faktor for tilgængeligheden af substratet for kimen under vækst. Denne funktion er også relateret til malt modifikation (enzymatisk nedbrydning af cellevægge). Spektroskopiske kalibreringer til vitalitet bør videreudvikles til en *on-line* NIT metode til forudsigelse af vitalitet som første selektion af maltbygkvalitet hos landmanden/kornmodtagelsen. De selekterede prøver skal efterfølgende kontrolleres vedrørende viabilitet ved en 3 dags spiring eller ved en Tetrazolium test. En evaluering med spiringsklassifikations-plottet (Artikel III) bør efterfølgende udføres ved at kombinere de to estimater for vitalitet og viabilitet.

- Det er muligt at finde og definere ukendte parametre med hensyn til fysisk-kemisk diversitet af byg ved brug af Nær Infrarød Reflektions spektroskopi (NIR) evalueret med kemometri (Artikel V-VII). En ny klasse mutanter med højt ß-glukanindhold og lavt stivelsesindhold er fundet med NIR spektroskopi. Denne kan danne grundlag for videre genetiske og bioteknologiske analyser for at definere genfunktionen i relation til ß-glukan/stivelsessyntesen i byg. Mutanterne med højt indhold af ß-glukan er af potentiel interesse i fremtidens kost som fiberkilde.

- Multivariat data analyse (PCA og PLSR) af datasæt med relevans for planteforædling, bioteknologi og malt- og bryggeriindustrien er anvendt til at danne grundlag for de nye resultater. Slutbrugerne som arbejder med spektroskopi og billedanalyse instrumenter samt kemi og maltanalyser bør kunne lave deres egne klassifikationer og kalibreringer samt identificere outliers. En sådan brug af multivariat analyse er nødvendig for at industrien skal kunne udføre en selektion efter høj maltkvalitet baseret på et multivariat grundlag, og ikke alene bruge NIT spektroskopi til at bestemme vand og protein med instrument fabrikantens datamodeller. NIT spektroskopi har vist sig, at være et meget anvendeligt værktøj til at korrigere for et negativt genetisk kovariat kvalitetskompleks for en højlysin bygmutant baseret på selektion for en ændret genbaggrund. Instrumentelle data evalueret med kemometri vil kunne bruges til "data forædling" som er en innovativ interaktion mellem på den ene side planteforædling i marken og på den anden side udvælgelse af byglinier udfra byggens NIT data i PCA og PLSR scoreplots.

List of publications included in the thesis

This thesis is based on the publications listed below and are included in full text. The papers will be referred to as Papers I to VII throughout the text.

Paper I:

Birthe Møller and Lars Munck (2002). Seed vigour in relation to heat sensitivity and heat resistance in barley evaluated by multivariate data analysis. *Journal Institute of Brewing* **108** (3), 286-293.

Paper II:

Birthe Møller, José Luis Molina-Cano and Lars Munck (2002). Variation in malting quality and heat resistance in the malting barley variety "Alexis". *Journal Institute of Brewing* **108** (3), 294-302.

Paper III:

Lars Munck and Birthe Møller (2004). A new germinative classification model of barley for prediction of malt quality amplified by a Near Infrared Transmission spectroscopy calibration for vigour "on-line" both implemented by multivariate data analysis. *Journal Institute of Brewing* **110** (1), 3-17.

Paper IV:

Birthe Møller (2004). Near Infrared Transmission spectra of barley of malting grade represent a physical-chemical fingerprint of the sample that is able to predict germinative vigour in a multivariate data evaluation model. *Journal Institute of Brewing* **110** (1), 18-33.

Paper V:

Lars Munck, Jesper Pram Nielsen, Birthe Møller, Susanne Jacobsen, Ib Søndergaard, Søren Balling Engelsen, Lars Nørgaard and Rasmus Bro (2001). Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. *Analytica Chimica Acta* **446**, 171-186.

Paper VI:

Lars Munck, Birthe Møller, Susanne Jacobsen and Ib Søndergaard (2004). Spectral multivariate indicators for mutant endosperm genes evaluated by chemometrics reveal a new mechanism for substituting starch with (1→3,1→4)-ß-glucan in barley. *Journal of Cereal Science*. Submitted.

Paper VII:

Susanne Jacobsen, Ib Søndergaard, Birthe Møller, Torben Desler and Lars Munck (2004). The barley endosperm as a data interface for expression of genes and gene combinations at different levels of biological organisation explored through pattern-recognition data evaluation. *Journal of Cereal Science*. Submitted.

Abbreviation list

AACC	American Association of Cereal Chemists
Al	Alexis, barley variety
A/P	Amide/Protein
Ar	Ariel, barley variety
BG	(1→3, 1→4)-ß-glucan
Bl	Blenheim, barley variety
Ca	Carula, barley variety
CZ	Czech Republic
D	Deutchland, Germany
der	Derivate
DK	Denmark
E	Espania, Spain
EBC	European Brewery Convention
Et	Etna, barley variety
g%n	Germination percentage day n
GC	Germinative Classification
GE	Germinative Energy
GH	Germination Homogenity
GI	Germination Index
HI	Hardness Index
ICC	International Association for Cereal Chemistry
i-PLS	Interval Partial Least Squares Regression
Ly	Lysimax, barley variety
Me	Meltan, barley variety
MGT	Mean Germination Time
MLR	Multiple Linear Regression
MSC	Multiplicative Signal Correction
N	Normal barley variety
NE	The Netherlands
NIR	Near Infrared Reflection
NIT	Near Infrared Transmission
P	Protein
PC	Principal Component
PCA	Principal Component Analysis
PLSR	Partial Least Squares Regression
RE	Relative Error
RMSECV	Root Mean Squared Error of Cross Validation
SD	Standard Deviation
SKCS	Single Kernel Characterization System
SU	Suomi, Finland
TKW	Thousand Kernel Weight
VP	Vigour Potential

Table of contents

1. Strategy and aim of the thesis

The approach of this thesis is to develop analytical screening methods combining the different research disciplines analytical chemistry, spectroscopy and multivariate data analysis and their applications in the fields of agronomy, plant breeding, biotechnology as well as in the malt, brewing, food and feed industries. It is obviously not possible to cover all these disciplines and applications in depth. However, the main objective of this thesis is to demonstrate in an interdisciplinary effort that specialists can gain new knowledge, time and money by exploring the great potential of spectroscopy and multivariate data analysis in the development of screening analyses useful in many of the application fields outlined above. The screening methods and their applications put forward in this thesis should be regarded as preliminary investigations to demonstrate the obvious possibilities for science and industry in using instrumental screening methods evaluated by multivariate data analysis.

Selecting high-quality barley

Plant breeding and industrial utilization of barley involve a selection made by man of the best plants and grain batches within a variable population. Selection is made possible by exploring and promoting variability in the most important quality criteria.

The main objective in the past has been to find means to increase grain yield that is dependent on many environmental and genetic influences. There are, however, several other requirements related to germination and malting characteristics of physiological and physical-chemical nature, which are central to barley breeding for malt, feed and human consumption. For a successful result in detecting, propagating and exploiting diversity, effective screening methods for the most important quality traits are essential. Therefore, a major effort is made in this thesis to develop such methods.

Quality requirements

The targets for breeding of malting barley are numerous, as the farmers, maltsters and brewers each have specific requirements for new varieties. Consumer demand for clean raw material has led to a shift in the production systems towards less use of pesticides, so the farmers need to grow plants that are resistant to fungi and can compete with weeds.

The maltsters have several specific quality requirements for the malting varieties[43]. For example, they want barley with easy removal of dormancy, improved vigour that is maintained during storage and a fast and uniform endosperm modification implying the establishment of a high hydrolytic enzymatic potential.

The brewers require varieties yielding a malt quality within the specifications (e.g. regarding extract yield, malt modification, enzymatic activity, soluble N, ß-glucan content), aiming at a beer with a stable foam, good taste and storage stability and low oxidation potential[43].

The breeders try to comply with the wishes of the industry, but can only breed for improved quality if simple and efficient screening methods are available[43,68] and if there is enough genetic diversity in the underlying breeding material.

Screening methods for quality are required not only in plant breeding and in the malt industry, but also for quality grading and payment at the commercial elevators in grain companies as well as in the feed and food industry.

The ultimate aim in industry today is to introduce quality control *on-line* or *at-line* thus abandoning the quality control laboratory to save time and money. This is now possible thanks to near infrared spectroscopy now used routinely in the cereal industry for protein and water analyses[13]. This thesis aims at extending the NIR and NIT spectroscopy application in analysing barley for quality.

The quality complex

The overall quality concept for e.g. malt production is a multi-facetted complex in which several analytical methods have to be employed. It would always be an advantage if fewer and faster methods could replace the traditional methods, while still using the latter for validation. The quality requirements that should be fulfilled can be divided into four groups: Grain composition, physiological criteria, process functionality and safety.

Grain composition consists of a wide range of chemical and physical quality criteria. The chemical properties include parameters such as protein, starch, fat and fibre content. The physical features focuse on seed hardness, form and cell wall thickness (ß-glucan content).

The physiological criteria involve seed viability (live seeds), germination speed and homogeneity as well as enzymatic activity.

The process functionality is related to how the raw material functions in malt processing and to the effects of differences in physiological and physical-chemical composition. It is evaluated in laboratory-scale experiments such as micro malting involving characters such as extract (%) and ß-glucan (mg/L) in wort which are important economic criteria in full-scale malt production.

A high ß-glucan content of barley is negative in malting[15,52,65] but positive in human nutrition[78]. It may block the filtration in the brewing process. The same trait leads to a high content of soluble fibres in human diets which has an important function for preventing

high cholesterol in blood and development of cancer[26,50,51,78]. The amino acid composition (essential amino acids - lysine) is an important nutritional factor in feed barley, which has initiated breeding programmes for high lysine barley[57].

Safety requirements such as the presence of insects or toxic substances e.g. from fungi, will not be discussed further in this thesis.

Many different laboratory analyses are thus needed to achieve a full quality characterisation of the quality of a cereal for a specific purpose. These analyses are both time-consuming and expensive and are inadequate to meet the demand for more rapid and cost-effective direct analyses in the cereal industry.

Organisation of the thesis

The thesis is divided into the following parts: Section 2 describes methods for determining grain quality, whereas Section 3 describes the multivariate data analysis which is essential in this thesis. Sections 4 and 5 are examples of areas where the use of multivariate data analysis has made it possible to describe and explore new hypotheses and connections. Section 4 describes applications within germination and Section 5 deals with the advantages of spectroscopic screening methods and multivariate data analysis in plant breeding and biotechnology. Section 6 concludes the claims of the thesis.

2. Experimental methods

2.1 Overview of the analytical methods used

Table 1 lists the 25 methods used in this thesis divided into groups (germinative, spectroscopic, morphological, physical, chemical analyses and micro malting) including references.

The properties of barley and malting quality can be described as manifest (apparent) and indirect parameters. The manifest parameters are related to kernel morphology (e.g. 1000 kernel weight) and the chemical properties of the raw barley such as protein and ß-glucan. The indirect parameters are properties that are only obtained by processing e.g. by germinating the grains to obtain a germination characterisation curve or by malting the samples for establishing values such as extract % and mg/L ß-glucan in wort.

It is the ultimate aim of non-destructive screening analyses combined with multivariate data analysis to predict the results of destructive chemical analyses manifest in the grain and the important indirect parameters that are obtained downstream in the production chain from the barley material.

Table 1. Overview of the methods used in this thesis.

Method	Analysis	Reference	Used in
Germinative	Germination energy (GE)	EBC Analytica 3.6.2, 1998[7]	Art II-IV
	Germination capacity (GC)	EBC Analytica 3.5.2, 1998[7]	Art I, III-IV
	Vigour potential (VP)	Aastrup et al, 1989[2]	Art I-II
Spectroscopy	Near infrared transmission	Infratec 1225 Food and Feed Analyzer, Foss Tecator, Höganäs, Sweden	Art III-IV
	Near infrared reflection	NIR Systems 6500, Foss, USA	Art V-VII
Morphological	GrainCheck	GrainCheckTM 310, Foss Tecator, Höganäs, Sweden	Art III-VI
Physical	Thousand kernel weight	EBC Analytica 3.3, 1987[6]	Art III-IV
	Hardness index	SKCS 4100, Perten North America, Reno, NV, USA	Art III-VI
Chemical	Dry matter	ICC 110/1[9]	Art I-VII
	Protein	EBC Analytica 3.2, 1987[6]	Art II-VII
	ß-glucan	EBC Analytica 3.11.2, 1987[6]	Art II-VII
	Starch	AACC 76-13[8]	Art V-VII
	Amylose	BeMiller, 1964[11]	Art VI
	Fat	AACC 30-25[8]	Art V-VII
	Soluble and Insoluble fibres	AACC 32-05[8]	Art V-VI
	Ash	ICC 104/1[9]	Art II
	Alfaamylase activity	ICC 108[9]	Art I-II
	Amide	Munck , 1972[55]	Art V-VII
	Amino acid	Barkholt and Jensen, 1989[10]	Art V-VII
	2D electroforese	Jacobsen et al, 2001[37]	Art V and VI
Micro maltings	Extract	EBC Analytica 4.4, 1987[6]	Art II, Art III-IV*
	Betaglucan in wort	EBC Analytica 3.11.2, 1987[6]	Art II, Art III-IV**
	Kolbach Index	EBC Analytica [6]	Art II
	Colour	EBC Analytica 4.7, 1987[6]	Art II
	Viscosity	EBC Analytica 8.2, 1987[6]	Art II
*Extract was determined with a refractometer			
**In-house colour binding method by which ß-glucan and congo red develop a colour complex which is measured with a spectrophotometer			
AACC = American Association of Cereal Chemists, EBC = European Brewery Convention, ICC = International Association for Cereal Chemistry			

2.2 Germinative measurements

A fast and even germination in malting barley is an important quality criterion because it allows the maltsters to be more flexible in the choice of process parameters to meet the customers requirements. In this thesis two germination methods are used. One determines germination energy (GE) by germinating the grains in petri dishes with water (EBC 3.6.2)[7] and one determines germination capacity (GC) using immersion in 0.75% H_2O_2-solution to overcome dormancy after harvest (EBC 3.6.1)[7]. GE is used in Papers II-IV and GC in Papers I, III-IV. From these methods the speed of germination is calculated obtaining the germination index (GI) (Paper I-IV) and germination homogeneity (GH) (Papers III and IV) after Riis and Bang-Olsen[70].

The vigour potential (VP) is determined according to Aastrup *et al.*[2] and is investigated in Papers I and II.

The methods are further described in Section 4.1.

2.3 Spectroscopy

Near infrared spectroscopy[85-86] combined with multivariate data analysis (Section 3) is an established economical, non-destructive technique widely used in the food and feed industries. The method is fast, and often more than one component can be determined at the same time. In barley, calibrations have been made, among others, for moisture[22], protein[21], starch[41,80], amino acids[84], hot water extract[54] and crude fibre[31].

Henry[34], Law[44] and Czuchajowska *et al.*[16] attempted to determine total ß-glucan content in barley by near infrared reflection spectroscopy (NIR), but were unable to obtain a satisfactory prediction (r=0.69). Delwiche *et al.*[18] have shown the possibilities of using NIR and multivariate data analysis to detect wheat-rye chromosome translocations with high precision by measuring flour from wheat kernels. Wang *et al.*[83] used non-destructive NIR measurements on single kernels to predict the number of dominant R-alleles influencing colour in wheat. NIR has also been used to identify amylopectin/amylose in mutants of wheat[19]. The non-destructive NIT measurements on whole kernels was utilised by Campbell *et al.*[14] to classify starch mutants in maize with PCA and PLS with different degrees of precision. In Papers V-VII specific spectral gene indicators are investigated by using NIR measurements on barley flour from normal barley and from different endosperm mutants.

NIT spectrometers are, as briefly described in Section 1 already used in the cereal industry for determining protein and moisture content[13,87]. The spectroscopic technique produces "global" results within the limits of the performance of the instruments. That is, in principle

it grasps a holistic fingerprint of the entire state of the physics and chemistry of the sample where protein and moisture content are just two of the many physical and chemical signatures that are latent in the spectra[58]. NIT spectrometers can thus be used as multimeters to predict several different analytes, if reliable calibrations are available.

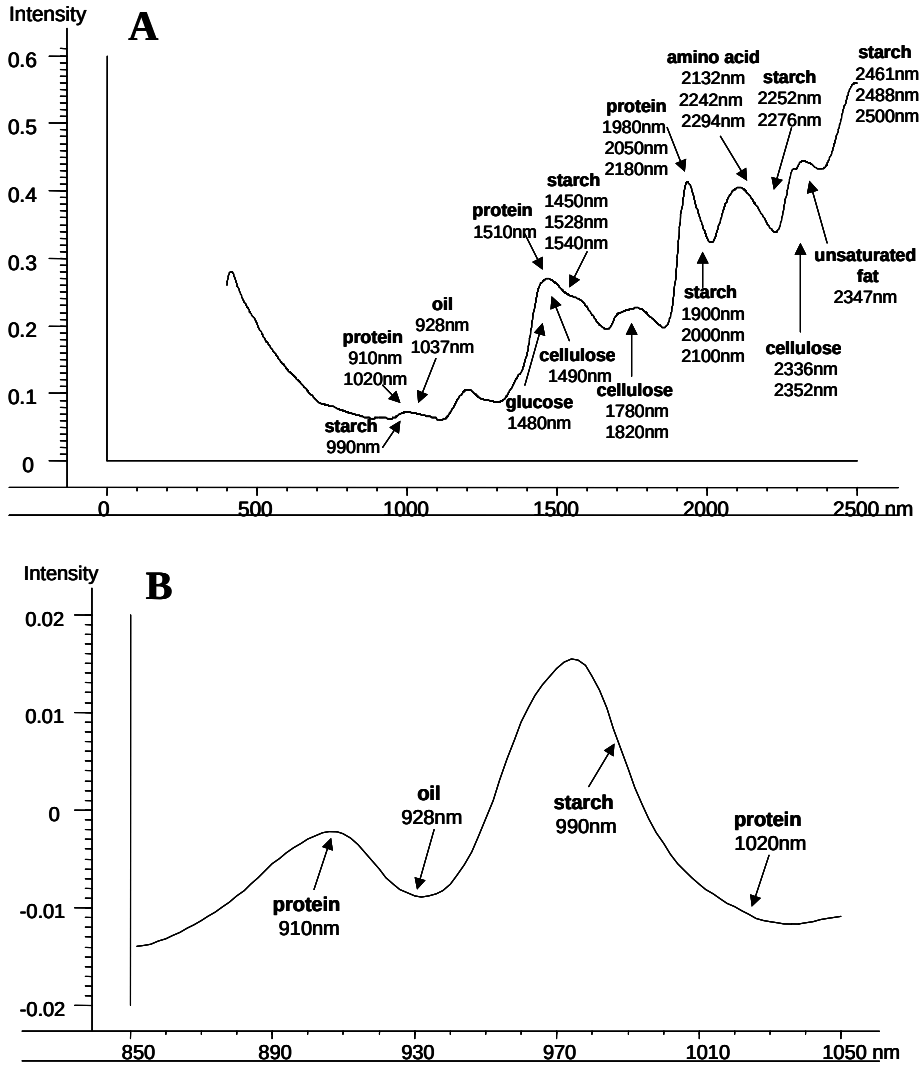


Figure 1. Example of spectrum of barley obtained by **A.** NIR and **B.** NIT.

Figure 1A shows an example of a multiplicative signal corrected (MSC) NIR spectrum of barley flour obtained at 400-2500nm. Wavelengths indicative of chemical bonds in amino acids, cellulose, glucose, oil, protein, starch and unsaturated fat are shown. In Figure 1B first derivates of a NIT spectrum obtained on whole barley kernels are displayed together with some of the chemical analytes (bonds) that are indicative of wavelengths in the area of

850-1050nm. As can be seen for especially the NIR spectrum, it is not only one single wavelength that is connected to for example protein. It is a complicated covariate function depending on different repetitive manifestations of infrared vibrations in the NIR area[64]. The spectral measurements have a very high degree of reproducibility. Therefore, even small differences in the spectral plot of samples, almost unrecognisable on visual inspection, may contain important reproducible information, as is demonstrated in Papers V-VII.

In order to detect and utilise these subtile changes in spectra, multivariate data analysis is essential as will be discussed in the following.

2.4 Morphological (imaging) measurements

Seed brightness as well as grain size and shape are associated to malting performance[30,43,61], making a seed imaging analysis instrument interesting to use to determine barley quality. Serious weather conditions during grain maturation promote mould growth which results in dark kernels. In this thesis the GrainCheck instrument (Foss Tecator, Sweden) has been used to obtain digital images of hundreds of grains. The images are used to estimate different kernel parameters such as kernel size, kernel shape and colour of every single grain. This procedure is automatic and fast; the results are available within a few minutes. In this thesis the average values of the kernel parameters are used as a fast manifest description of the kernel characteristics. GrainCheck has been used in Papers III and IV.

2.5 Physical measurements

A Single Kernel Characterization System (SKCS) 4100 instrument (Perten Instruments Inc., Reno, NV, USA) originate developed to determine the hardness of wheat has been used for determination of hardness in barley. Experiments have shown that hardness of malting barley was associated to malting performance[37,61] (Papers III and IV). As mentioned in Paper IV, the instrument used is built for wheat and therefore, analyses of barley can be problematic if the barley samples have not been cleaned and graded carefully. The barley grains are longer and narrower compared to wheat kernels that are short and round and can therefore be rejected of the instrument. The rejected number of kernels is correlated to thousand-kernel weight. They were included in the data analyses in Papers III-IV. The instrument company should adapt the software of the instrument to barley to reduce the numbers of rejected grains.

2.6 Chemical measurements

Several different reference methods have been used to determine quality in grains (Table 1). Before NIT was an integrated part of the agricultural industry, protein was determined by the Kjeldahl method. Today, other major components are determined such as starch, fat and fibre. Amino acid analyses and 2-dimensional electrophoresis are examples of more deeply penetrating complex analyses used in Papers V-VII to define the biochemical effect on the phenotype of specific genes and as a validation of the spectral gene indicators obtained.

3. Data analysis

The most simple form of data analysis is overviewing and comparing means of classes and evaluating analytes in simple two dimensional plots as is or in the form of ratios. This type of study is a useful way of getting to know the data set.

An aspect of data representation that is not strictly part of statistics, but definitely affects the results obtained, is pre-processing of the data. The natural units for the variables involved in a multivariate analysis problem are often different. For example, one variable in an analysis could have a range of 0.2 to 0.4 (e.g. amide) and another a range of 9 to 20 % (e.g. protein). A normalisation operation - autoscaling - is used to convert each raw variable in a data set to a standardised variable with zero mean and unit variance[40].

3.1 Pre-processing of data

In this thesis first derivate is used on NIT spectra (Papers III and IV) and multiplicative signal correction (MSC) is used on NIR spectra (Papers V-VII). The two methods are described briefly in the following.

The first or second derivatives are common transformations of continuous data-reducing noise and are often applied in spectroscopy. If there is a constantly increasing or decreasing basis line this can be eliminated by using the first or second derivate[12]. The first derivate is used for NIT data in Papers III and IV, reducing constantly changing baselines.

Spectroscopic measurements of flour often display light-scattering effects due to physical differences for example in particle size that influence the results. MSC[48] is a spectral transformation method that can compensate for two general undesired effects, amplification and offset, which can be removed from the raw spectral signals to prevent them from dominating over the chemical signals. By using MSC it may be possible to save one or more PLSR components in modelling the relevant **y**[27]. The MSC is based on the fact that light scatter's wavelength dependency on particle size is different from that of the chemically based light absorption. By using data from many wavelengths it is possible to distinguish between absorption and scatter[30]. MSC is used in Papers V-VII.

3.2 Multivariate classification by Principal Component Analysis

In the real world as in food production chains the underlying processes are indirectly observed from their manifest values. The manifest primary data gives a complex view of the ongoing processes, which often may be compressed into more simplified underlying latent functional factors because of the essential dependence (covariance) of primary data.

In 1901 the classic statistician Karl Pearson[66] published a paper "On lines and planes of closest fit to systems of points in space" and laid the foundation to multivariate analysis by pattern recognition (Principal Component Analysis - PCA). These methods were further developed by the social scientists and introduced into chemistry in the beginning of the 1970's and named chemometrics. The introduction of the computer revolutionised multivariate analysis by introducing an experimental mathematical approach being able to process and overview large data libraries by an instructive graphical data interface.

In contrast to classic statistics that looks upon all variables as manifests the data program in multivariate analysis produces the underlying principal components (PC's) by combining different amounts of the measured variables as latent variables according to their distribution in data space.

The aim of using the PCA logarithm is to let the data set "speak for itself" as a whole by unsupervised modelling interaction of the used variables by defining the underlying latent factors and use these in a graphic scatter plot to characterize the variable pattern of each sample. The PCA biplots, where different PC's are compared, contain information about the patterns of variables from each sample and about the association of each variable to the samples in a biplot. It is thus possible to address the patterns of variables from each individual sample in relation to other samples in a PCA biplot (see example in Section 3.2.1).

In PCA the data matrix (**X**) is decomposed into a score matrix (T), a loading matrix (P) and the residuals (E), resulting in **X**=TP' + E.

The principle of the PCA is that the algorithm treats the data by plotting the data pattern of samples in a n-dimensional space where n is the number of variables. A first PC axis will run through the data, describing most variation. The score values of the samples are the perpendicular distances to the PC axis. This is PC1. PC2 will be placed orthogonal to the first one through the centre of the data space, describing second largest variation, third largest variation with PC3 and so on. By this procedure the PCA algorithm finds the main directions in the data set.

By plotting different combinations of principal components as axes, patterns in data are easily explored (e.g. how the samples (scores) group and the variables (loadings) correlate, and if some samples are more influenced by a certain variable than others).

These plots are used as a basis for a later more precise definition of the relationship between samples and variables by involving prior knowledge and by supervised deductive methods[59]. A distinct example of this principle is given regarding analysis of germination curves in Section 4 (Paper III) where the functionality of the latent factors (principal components PC's) were revealed, so that PC1 was found to describe vigour (germination speed) and PC2 described the viability (germination amount). The PC's are aimed to be

interpreted *a posteori* computation. When such an interpretation is possible it is a sign that the data analysis has been successful in depth and been founded in *a priori* knowledge.

A few scientists associated with plant breeding[77] have for a long time argued for a multivariate approach to model the covariate characteristics unique for each genotype by using for example PCA. The main problem in their approach has been in obtaining data of high quality, which is difficult when focusing on complex phenotypic environmentally dependent traits like genotypic stability of yield under favourable and stressed conditions[51]. The high quality of spectral data from products such as barley makes them especially suited for multivariate pattern recognition evaluation. In malt and brewing research multivariate analyses have been used to compress the complex malt quality data set into few components[38-39,56,69].

PCA is used to explore data in Papers I-VII. An example is given in Section 3.2.1.

3.2.1 An example of PCA classification

Table 2 gives an overview of 54 barley samples analysed for protein, amide and ß-glucan. In classical statistics one would look at the variables one by one and test whether the variables are normally distributed or if there is a tendency toward a class separation. As seen in Figure 2A-B, protein (and amide, not shown) seem to follow approximately a normal distribution curve, whereas ß-glucan is deviating. However, obviously the number of samples is too low to allow a firm decision. It is impossible to classify this sample set unsupervised by distributional assumptions. As is seen in the following, this is possible by a pattern recognition multivariate approach.

If the variables are linked to each sample as chemical "spectra", with three "wavelengths" they will appear as seen in Figure 2C. Because the ranges of the variables differ from each other (protein 13.3-20.5, amide 0.26-0.60, ß-glucan 2.2-20.0), the data are standardised by dividing the sample data by the standard deviation. This gives all variables the same range of variation. Prior knowledge is now placed "on the shelf" letting the data speak for them selves in an unsupervised classification[58].

An example of a PCA classification analysis is shown in the biplot in Figure 2D where data from the chemical spectra from each sample (Figure 2C) are compressed to represent one point in a PCA score plot.

Table 2. Chemical composition of the 54 samples in Figure 2.

	Protein	Amide	A/P	ß-glucan		Protein	Amide	A/P	ß-glucan
1	18.2	0.35	12.0	5.3	28	16.6	0.47	17.7	5.8
2	18.2	0.33	11.3	5.3	29	17.2	0.46	16.7	5.5
3	16.6	0.29	10.9	3.6	30	15.5	0.44	17.7	5.3
4	17.8	0.27	9.5	2.2	31	15.6	0.42	16.8	5.2
5	17.0	0.26	9.5	2.3	32	15.8	0.41	16.2	7.2
6	17.6	0.32	11.4	2.5	33	15.1	0.45	18.6	6.0
7	16.6	0.31	11.7	3.6	34	15.2	0.39	16.0	6.0
8	17.8	0.36	12.6	5.0	35	15.1	0.39	16.1	5.3
9	17.1	0.31	11.3	7.2	36	16.3	0.42	16.1	6.0
10	15.9	0.36	14.1	6.3	37	16.4	0.46	17.5	5.7
11	19.2	0.38	12.4	8.5	38	17.2	0.47	17.1	5.4
12	16.6	0.38	14.3	8.6	39	16.9	0.47	17.4	9.0
13	16.9	0.41	15.2	19.6	40	16.3	0.44	16.9	5.3
14	18.4	0.49	16.6	19.8	41	14.6	0.39	16.7	6.7
15	15.7	0.37	14.7	20.0	42	15.9	0.41	16.1	6.9
16	18.9	0.49	16.2	13.7	43	13.3	0.36	17.0	5.3
17	18.1	0.45	15.6	11.9	44	14.5	0.39	16.8	5.6
18	16.5	0.41	15.6	12.3	45	20.5	0.60	18.3	6.7
19	17.3	0.41	14.8	13.5	46	16.3	0.43	16.5	5.1
20	17.3	0.43	15.5	13.5	47	16.8	0.44	16.4	4.4
21	16.1	0.39	15.1	13.5	48	17.8	0.48	16.8	6.0
22	15.4	0.43	17.4	5.6	49	16.7	0.46	17.2	6.1
23	16.0	0.44	17.2	6.0	50	16.5	0.44	16.7	6.3
24	16.1	0.44	17.1	5.1	51	17.3	0.48	17.3	5.0
25	14.6	0.38	16.2	6.8	52	17.5	0.49	17.5	6.6
26	14.6	0.39	16.6	5.3	53	17.7	0.48	17.0	6.2
27	15.6	0.41	16.4	6.0	54	18.2	0.48	16.5	6.0
1-3:*lys3a* , 4-5:*lys3m* , 6-7:*lys3b* , 8-9:*lys3c* , 10-12:*lys3a5g* , 13-15:*lys5f* , 16-21:*lys5g* , 22-54:normal									

In this very simplified example it is seen that the samples in the PCA plot separate into three clusters: Group 1 with samples located diagonally from the top left corner to the right bottom corner, Group 2 (encircled) with samples located in the top right corner and Group 3 (encircled) in the bottom left corner. The three loadings (analyte variables) are located to the right in the plot with ß-glucan on top close to Group 2, amide in the middle and protein at the bottom of this line. When looking in the laboratory journals we can now name the samples shown in Figure 2D. It is seen that the samples in Group 2 are very high in ß-glucan, whereas Group 3 has higher protein and lower amide and ß-glucan content than any of the samples in Group 2. The samples in Group 1 have values in between these two extremes. It is also seen that the samples in Group 1 are normal barley lines and *lys3a5g* double recessives, Group 2 are high ß-glucan *lys5f* and *lys5g* mutants, while Group 3 are *lys3a*, *lys3b*, *lys3c* and *lys3m* mutants with high lysine and low amide content.

The data was thus first explored in an unsupervised PCA classification and afterwards related to our prior knowledge.

Here it is possible to investigate and monitor each single sample and its behaviour compared to the others. In classical statistics one needs a large population and the results are the variance for the data set or groups within it. It is difficult if not impossible to follow a single sample.

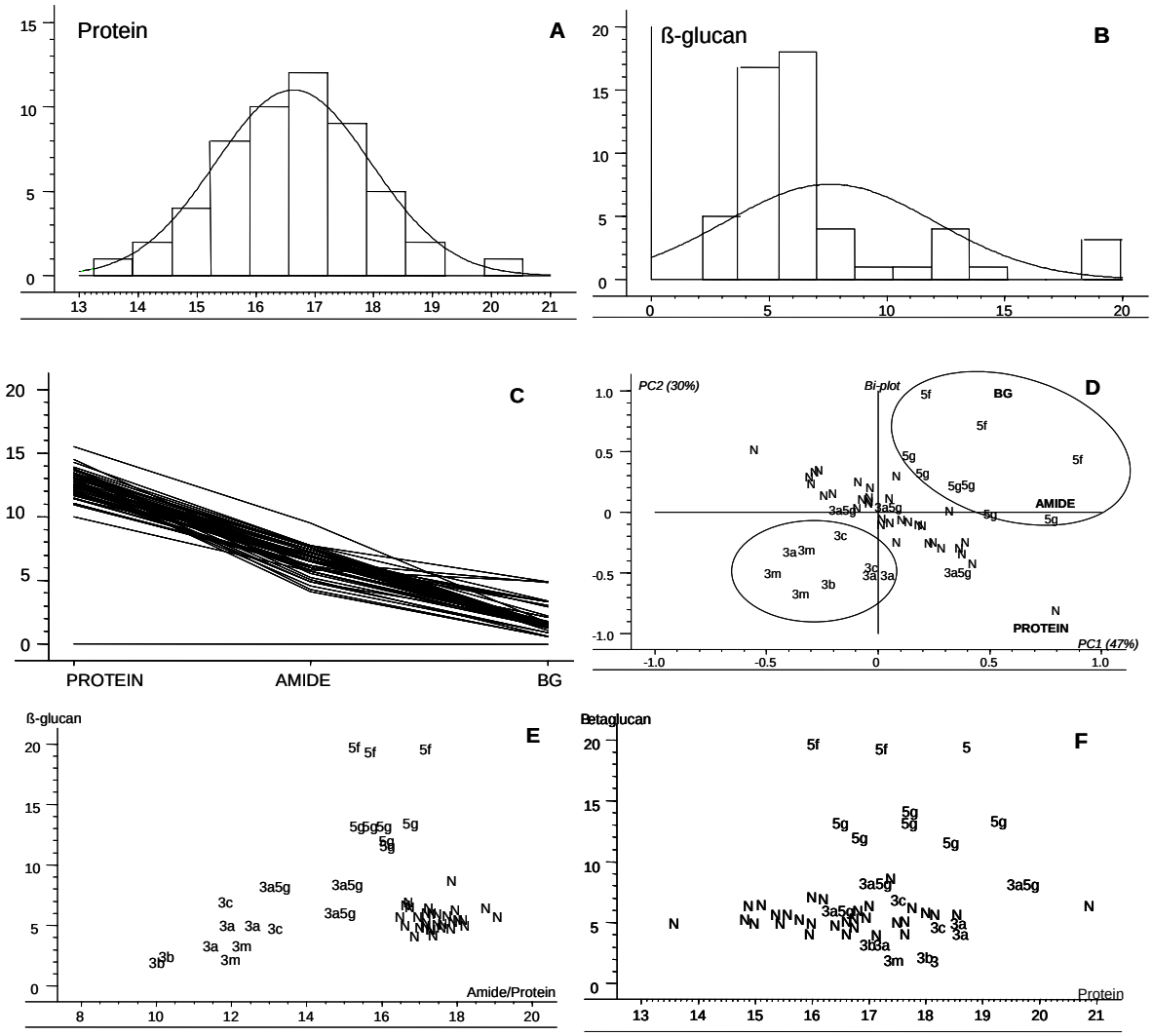


Figure 2A-B. Distribution of values of chemical analyses of 54 barley lines: protein% (A) and ß-glucan% (B) **C.** Patterns from the 54 samples in the form of "spectra" with three "wavelengths" for protein, amide and ß-glucan (normalised values). **D.** PCA biplot (scores and loadings) with principal components PC1 (abscissa) and PC2 (ordinate) evaluating each sample from the material in C featuring normal barley (N) and the genes *lys3* (alleles *a*,*b*,*c*,*m*) and *lys5* (alleles *f*,*g*). The biplot demonstrates that *lys3* is associated with high protein, normal barley with high amide and *lys5* with high ß-glucan. **E-F**. Two-dimensional simple abscissa-ordinate plots of A/P - ß-glucan (E) and Protein - ß-glucan (F).

From the PCA in Figure 2D it is clear that the amide, protein and ß-glucan content are reliable criteria for classifying the different classes of high-lysine, high ß-glucan and normal barley lines. However, by plotting the amide/protein index (A/P) against ß-glucan, as seen in Figure 2E, we obtain the same classification as in the PCA plot in Figure 2D. Here samples of *lys3* mutants are low in ß-glucan and A/P values, *lys5* mutants have high ß-glucan values and medium A/P. The double recessives (*lys3a5g*) are located in between, while the normal barley lines have high values of A/P and medium values of ß-glucan. The same simple classification strategy is less successful using the combinations protein-to-ß-glucan (Figure 2F) and protein-to-amide (not shown). It is thus concluded that in this data set simple ratio plots are just as effective as multivariate analysis with PCA. However, as will be demonstrated in the following, multivariate data analysis is essential in evaluating more complex data sets.

By using the unsupervised classification with PCA (or by ratio plotting) we may now classify the samples and test their significance of variance in a classical statistical analysis (Table 3).

Table 3. Level of significance for a Duncan test calculated for each chemical parameter of the chemical analyses as well as samples divided into significant groups

		Significant groups		
Parameter	**Level of significance**	**A**	**B**	**C**
Protein	5%	*lys3*, *lys5*, *lys3a5g*	*lys5*, normal	-
Amide	1%	*lys5*, *lys3a5g*, normal	*lys3*	-
A/P	1%	normal	*lys5*, *lys3a5g*	*lys3*
Betaglucan	1%	*lys5*	*lys3a5g*	*lys3*, normal

Using Duncan's multiple range tests for every single chemical parameter it is seen that A/P index and ß-glucan give the best significance in classification and that they can complement each other in differentiating between the four genotypes.

Science to date has focused on the genotype in genetics, plant breeding and biotechnology and largely left the phenotype to others[82]. Probability statistics is instrumental in studying gene linkage and mapping as well as in gene sequences. However, modern statistics based on maximum likelihood cannot handle large highly covariate data sets (like spectra) which are characteristic for individual biological objects (phenotypes)[58]. Classical statistics has also difficulty in handling information from individual phenotypes (e.g. a grain sample from a genetically homogenous barley variety) discretely, but has to compare them in populations for classification, putting a great demand on the number of individuals in the different classes.

A strength of multivariate data analysis is that every sample is individually defined as a more or less unique pattern, and this makes it possible to investigate unbalanced data sets as well. It is clear that a small number of samples within a group compared to many samples in another group makes it difficult to draw general conclusions about group differences. However, valid conclusions can be drawn on the individual level of all samples if they are analysed by the complete set of parameters.

When the 54 samples from Table 2 are measured with NIR (now 1050 data points instead of 3), they can also be classified in a PCA. The same pattern is seen as with the three chemical variables, namely the four different groups from before signifying the genetic diversity of the spectral physical-chemical parameters (Paper VI and Section 5). The NIR data material will be discussed in the following section.

3.3 Multivariate correlation by Partial Least Squares Regression

PLSR focuses on finding the relevant information in the pattern of the variables **X** for describing one or several *a priori* defined **y** characteristics and can be described as:

$\mathbf{y} = \mathbf{X}b + E$

y is the measured reference value or the value that should be predicted from the earlier made calibration, e.g. protein content.

X is the data matrix, e.g. NIR measurements.

b contains regression coefficients that are determined during the calibration step.

E is the residuals, e.g. model errors and noise.

One important feature of PLSR is the ability to model covariate data. This is in contrast to Multiple Linear Regression (MLR) which does not cope well with covariate data, but is the classical way of building a regression model using several X-variables. PLSR analysis is used in Papers III-VII. See example in Section 3.3.1.

3.3.1 An example of PLSR correlation

When the variables in Table 2 are increased from 3 chemical to 1050 variables with NIR spectra (400-2500nm) in Figure 3A the use of MLR is inapplicable because of a high covariance of the data set. The NIR measurement gives, as indicated in Figure 3A, a fast physical-chemical fingerprint of the sample. Therefore it is expected that there should be some information from the chemical analysis in the NIR spectra. This is tested using the supervised PLSR algorithm for predicting protein from NIR spectra (Figure 3B).

In the PLSR analysis using NIR for predicting protein (calibrating to protein information in Table 2) the algorithm selects latent components in the form of mean spectra – **X** representing different aspects for the prediction of protein (**y**). By carrying out data experiments a full cross-validation validation is performed[49]. The regression coefficient (r = 0.96), error RMSECV (0.4) and RE (5.1) are calculated as well.

The PLSR scores are found in a slightly different way than in the case of PCA, namely by taking the variation in **y** during the decomposition of **X** into account, and optimising the covariance between the scores of the two structures. In Figure 3C a PLSR score plot with **y** = protein and **X** = NIR is shown for the 54 samples. It is evident that a classification is possible with the same accuracy here as for the three chemical variables with PCA (Figure 2D) or by ratio plotting (Figure 2E).

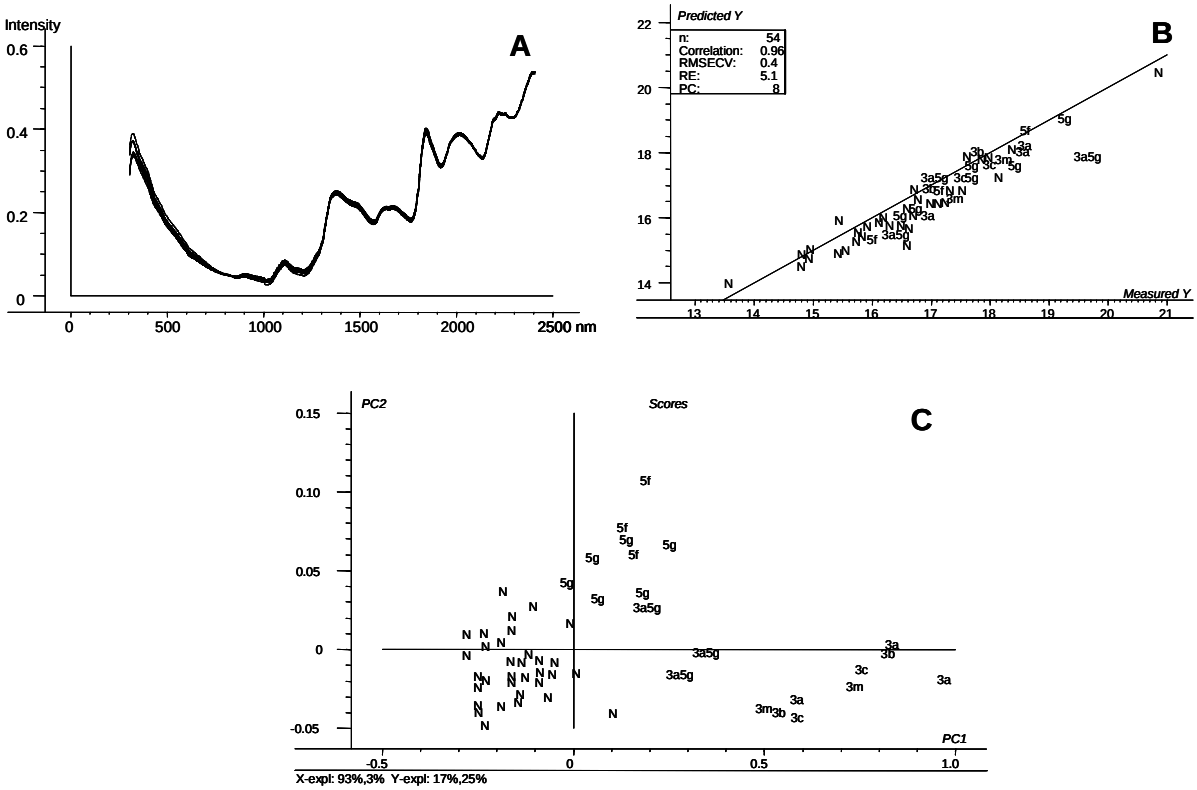


Figure 3. Example of PLSR prediction of protein from NIR spectra (MSC).
A. The NIR spectra (MSC). **B.** Measured versus predicted values. **C.** PLSR score plot.

In Table 4 the results from the NIR PLSR calibrations of the four different chemical parameters from Table 2 are shown using the spectrograph as a multimeter. It is seen that protein, amide and ß-glucan all can be predicted with a correlation coefficient higher than 0.96. However, to achieve these high correlations it is necessary to use 8-10 PC's, which

makes the models more complex. The RE values are quite large, but it is expected that RE would decrease if more samples were used.

Table 4. PLSR prediction of protein, amide, A/P and ß-glucan from NIR spectra

	r	RMSECV	PC	RE	n
Protein (P)	0.96	0.4	8	5.08	54
Amide (A)	0.97	0.01	10	2.94	54
A/P	0.88	1.1	1	11.62	54
ß-glucan	0.98	0.9	8	4.89	54

3.4 Validation

3.4.1 Statistical multivariate validation

As in all data analyses it is important to validate the results. In PLSR data analysis there are two possibilities of validation based on data experimentation: test set and cross-validation. The best way is to make the analyses on one data set, and then test the results on another representative data set taken randomly from the same data population - the test set. This is, however, not always possible if the sample set is small, e.g. because the reference analyses are expensive (like e.g. amino acid analyses)[27].

Another possibility is to validate by cross-validation. The principle is that the models are validated by taking one sample out and predicting this sample from the rest of the samples. Then include the sample again, exclude the next and so on. Full cross-validation has been used in all the papers in this thesis.

RMSECV and RE

The performance of the regression models is evaluated by its prediction error in terms of root mean squared error of cross-validation (RMSECV). The relative error (RE) in percentage is calculated as (RMSECV/($\mathbf{y}_{max}$-$\mathbf{y}_{min}$))*100 where $\mathbf{y}_{max}$ is the highest reference value and $\mathbf{y}_{min}$ the lowest reference value of the **y** parameter in question (Papers III and IV).

RMSECV is used in Papers III-V and RE is used in Papers III, IV and VII.

Outliers

PLSR prediction (e.g. for estimating protein by NIR) implies an inductive model where the next unknown sample is supposed to follow the model. This can be checked by the outlier

control in **X** in which the computer automatically controls whether or not the characteristic multivariate fingerprint from a sample belongs to the model (the calibration population). This is the scientific basis for the revival of induction on a large scale now made possible by new software[59]. This data technology is now used worldwide to screen for quality in, for example, agricultural and pharmaceutical products, but remains largely unused as an exploratory unsupervised tool in other areas of science such as genetics, plant breeding and biotechnology[58].

Outliers can depend on different circumstances. One type arises when wrong measurements create outliers, which should not be included in the model before checking the analysis. Another group of samples with extremely high or low content of the investigated constituent outside the calibration range should be checked further before the analysis. These outliers may improve the model when included by extending the range of values, thereby making the model more robust for future predictions. On the other hand, this type of outliers can also be too extreme, dominating the other patterns too much, as in the example of the extreme barley variety Lysimax in Paper IV.

In plant breeding the latter group of outliers can be very important, because the outlier can reveal itself as a potential mutant such as the six new high ß-glucan and low-starch mutants found in the example with PCA classification of NIR spectra in Paper VI.

3.4.2 Conceptual validation

The essence of chemometrics is to allow the statistical and conceptual evaluation to proceed stepwise with the support of a graphic display integrated by the experience of the same researcher instead of dividing the evaluation into statistical and conceptual tasks made by different persons[49,59]. The critical explorative analysis is facilitated by obtaining information from the same sample set by means of different methods (Papers III-VII) (e.g. by chemistry and spectroscopy) and then analysing them separately by unsupervised PCA, keeping prior knowledge "on the shelf"[58]. By letting data speak for themselves, such as in the search for suitable estimates for "vigour" and "viability" in Papers III and IV, new hypotheses can be created in a truly explorative investigation and tested by studying the chemical and spectroscopic data sets in correlation studies such as the PLSR[59] analyses demonstrated in Papers III-VII. The relation between samples and variables can be estimated in PCA biplots, as discussed in Papers I-VII. This procedure facilitates an overview at the same time, as details are revealed in depth in a sequential hypothesis-generating fashion.

Outliers in PLSR and PCA can be checked by visual inspection of score, correlation and influence plots (Paper III) and further analyses of the outlier sample to define their

character, as were especially used in detecting the role of the "viability" outliers in Papers III and IV.

Supervised classical statistics may lead to false results due to biased prior assumptions regarding, for example, the distributional (sigmoid) characteristics of heat treatment germination decay curves discussed in Papers I and II. An unsupervised PCA analysis of the germination decay curves demonstrated surprisingly the existence of heat resistant barley samples which were prevously neglected by other seed scientists.

4. A multivariate approach to germinative analyses in the barley industry and plant breeding

In the following, vigour and germination is discussed from the maltsters point of view where the aim is to characterise malt quality by the simplest possible method.

A malting company producing 375.000 t of malt a year will require about 500.000 t of barley. In Canada[68], the average size of a barley unload is about 75 t and an unload typically represents barley from a single permit or location. The yearly intake of barley would therefore represent about 6.667 permits. On the average for every sample accepted, two are rejected. A grain inspector could, therefore, look at about 20.000 submitted samples each year. The volume of work required to conduct germinations on even half this number, not to mention preload and unload samples, is prohibitive. Germination analyses are presently determined on composites of pooled samples, but values on all individual permits would be preferred. To accommodate these needs, a test that would provide a near-instant estimate of barley's germination potential and/or vigour is required[68].

4.1 Background to the development of germinative analyses

Traditionally, germinative analyses are only used as a prequalification criterion for malting barley (e.g. as germinative energy (GE) at 3 days germination as the deciding evaluation). In Papers III and IV the aim is to integrate information from the whole germination curve together with other criteria in the evaluation of malting barley.

Determining germination properties early in the season after harvesting is influenced by dormancy of the barley sample especially prevalent after wet growing seasons and in special varieties. To overcome dormancy the GC method can be used where H_2O_2 breaks the dormancy.

The determination of germination speed is supposed to be correlated to vigour, and was first systematically taken into account in malting when the brewing scientists Riis and Bang-Olsen[70-71] at Carlsberg defined the germination index (GI) as an expression of germination speed (3-5 days of germination). GI is considered to be the most effective method available to describe vigour and is investigated in Papers I-IV. GI is an indirect way of measuring the enzymatic activity important for modification[43]. In order to check the malting quality of barley, plant breeders and maltsters use micro malting to evaluate extract yield and about ten other malting quality factors. To comply with the demand concerning homogenous germination, Riis and Bang-Olsen[70] also developed an index for germination homogeneity (GH) that is investigated in Papers III and IV.

The Vita-scope and other germ-staining (Tetrazolium) techniques have been proposed, as has the use of hydrogen peroxide to detect viability (live seeds) in dormant seed samples[1,4,46]. However, all alternatives to the traditional methods suffer from one defiency or another, and there is as yet no technique that meets all the criteria that would suit a selector's needs. A simple-to-perform test is required that would reliably estimate the germination potential of 100-200 samples per day both early after harvest and after storage[68].

Seed lots with the same high germination test results in the laboratory may perform quite differently under stressful conditions. Under such conditions high vigour seed lots have greater capacity for seedling survival, as resistance to viability decreases under stress[23-25,32,68,74].

The assessment of seed vigour is based on the physical or physiological performance of a seed lot. These performance parameters include 1) changes in biochemical processes in the germ itself and in the endosperm, 2) the rate and uniformity of germination and seedling growth and 3) germination capability when exposed to stress conditions[79]. In every group of analyses there are different methods used today (Figure 4). In this thesis artificial ageing in heat (Papers I and II) and rate of germination determined as GI and by 24 hours germination (Papers I-IV) were the method used.

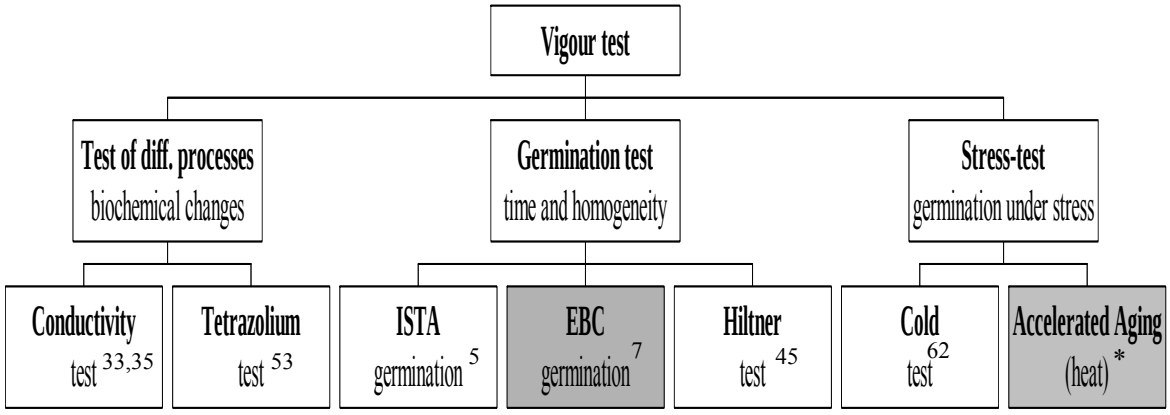


Figure 4. Methods to determine vigour, according to the literature.

It is well known that seed quality will deteriorate with time and that the rate of deterioration is dependent upon temperature and seed water content influenced by environmental conditions during storage and the length of storage period. This is the basis in artificially aged vigour models developed by Ellis and Roberts[23-25] based on seed research since 1911[32]. Ellis and Roberts[25] presume that ageing in a seed sample is normally distributed,

and that loss in germination is the last stage in a continuous process of different signs of ageing. They[23-25] claim that by using their germination model they can, by knowing the initial germination quality of the sample before and after a standard heat treatment, predict germination under all other storage conditions (temperature, moisture and time) after storage in n number of years.

Aastrup *et al*.[2] further developed the vigour model. They also presumed that the ageing of a seed lot was normally distributed (Figure 5A) and that the vigour potential (VP) could be defined by transformation of the germination percentage values into probits achieving a straight line (Figure 5B, Table 5). The intersection of the ordinate with the tangent of the line determined VP for the sample.

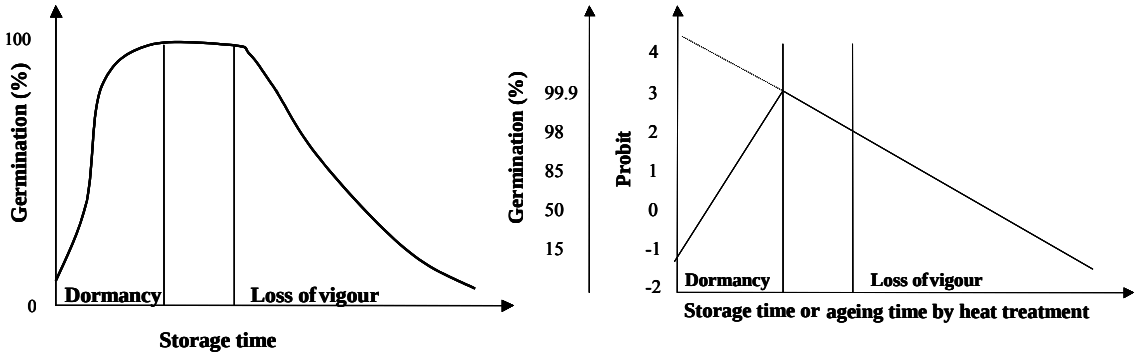


Figure 5A. Maximal germination as a function of storage time and **B.** Maximal germination as a function of storage time plotted on probability paper, according to Aastrup *et al*., 1989[2].

Table 5. Percentage of germination transformed to probit values (See Figure 5B)

Germination %	Probit
99.9	3
98	2
85	1
50	0
15	-1

Seed vigour is partly determined by genotype and partly by environment in a complex relation that is studied in Papers I-IV.

4.2 Germination heat stress experiments to estimate seed vigour

In order to investigate how vigour was influenced by genetics and environment more than 100 barley samples were collected in 1993 in cooperation with the Swedish grain company ”Skånska Lantmännen” (Paper I). Germination properties from one day of germination to eight days were determined in all the samples, and a sample collection was made

representing a large range of germination characteristics for GC and GE. Four of these samples were artificially aged and grown together with some of the naturally aged samples at three locations in 1994 and in subsequent years (Paper IV).

Using the methodology of Ellis and Roberts[25] and Aastrup *et al.*[2] to determine vigour presented as vigour potential (VP), it was seen that the method was not adequate for characterising all kinds of samples. When heat-treating both naturally and artificially aged samples for 4½ hours, 68˚C with water content at 12.0%, three different curve forms as well as intermediates were detected (Figure 6A). One sigmoid curve following the model (No. 11), one curve with initial (and following) low germination percentage (No. 03), and most surprisingly samples resisting the heat treatment with a germination of 90% or more after 4½ hours at 68˚C (No. 29) were found. Using the principle of Ellis and Roberts[25] and of Aastrup *et al.*[2], the germination curves were transformed to probits, so VP could be determined as the intersection of the curve tangent with the ordinate, according to Figure 5B. In Figure 6B it is seen that sample No. 11, VP=4.4 following the vigour model according to the Aastrup *et al.*[2] model has a higher VP compared to the heat-resistant sample resisting the stress No. 29, VP=3.1. According to the definition of vigour, a high-vigour seed sample should be able to resist heat stress on viability (survival), sample No. 29 should therefore possess a higher vigour than the heat-sensitive sample No. 11, which was not the case. The VP definition also implies that a high vigour seed should undergo fast and homogeneous germination.

Using the unsupervised algorithm PCA to reveal the pattern (Figure 6C) of the heat decay germination curves (Figure 6A) it is seen that the heat-sensitive samples (e.g. sample No. 11) are located at the upper half of the plot, whereas the samples with initial low germination percentage are located to the bottom left part of the plot. The heat-resistant samples (e.g. sample No. 29) appear in the bottom right part of the plot. Another discovery has occured. The “ID” notation (Figure 6C) refers to the GI values from untreated samples, indicating that the three sample groups roughly divide into samples with high GI values (Group I: Heat-sensitive samples), intermediate GI values (Group II: Heat-resistant samples) and low GI values (Group III: Samples with low initial germination percentage).

When average GI on untreated samples is calculated for the three groups there is a significant statistical difference between GI for Group I (heat sensitive) and the other groups, but not between Groups II and III (Paper I).

From this PCA evaluated experiment a hypothesis could be generated suggesting that heat-resistant germs “hide” in the grain, and by having a slow germination the grains may adapt better to the stressful environment than the heat-sensitive samples being killed because they are fast germinators. This hypothesis has to be further investigated before more generalised conclusions can be drawn.

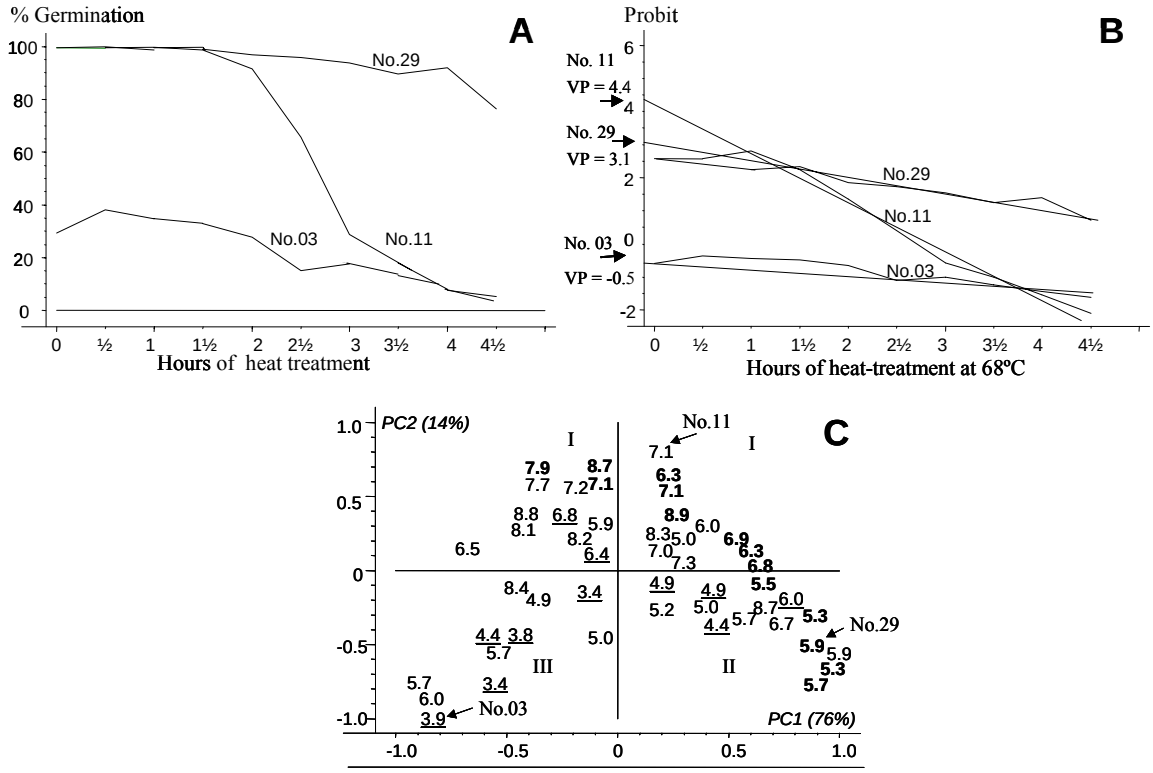


Figure 6. (Paper I) **A.** Germination decay curves with germination percentage according to time of heat treatment for three samples with different heat sensitivity (See comments in text). **B.** Germination curves from A transformed to probit. **C.** PCA score plot of germination decay curves. PC1:2. The numbers refer to the germination index (GI) of unheated samples for each sample. Underlined samples are artificially aged according to the Carlsberg model. Bold samples are harvested in 1994.

In ten sub-samples of two barley varieties (Ca108725 and Klages), Aastrup *et al.*[3] found high correlation coefficients (r=0.95 and r=0.92) between VP and GI. GI is calculated on the untreated sample (zero), before determining ageing by VP using 9 heat treatments. In our test in Paper I there was not seen such a significant correlation in the total material. When the material was divided into natural and artificially aged samples, there was no correlation for the naturally aged samples, however, a weak positive correlation (r=0.63) was found for the artificially aged samples. If the artificially aged samples were further divided according to variety, it seemed like the correlation between VP and initial GI is dependent of variety thereby explaining the results of Aastrup *et al.*[3] made entirely on artificially aged material.

Investigating earlier experiments from literature since 1911 with heat-treatment of primarily barley (Table 6) demonstrates that these were made with one to three varieties producing sub-samples by heat-treatment of a limited number of original barley field lots which all were artificially aged. In order to test the vigour model of Ellis and Roberts[25] and further developed by Aastrup *et al.*[2], 50 barley samples from 7 different varieties grown in two different years were used (Paper I). In contrast to earlier authors, both artificially and naturally aged samples were included. We even succeeded in finding naturally aged samples with a broader range in VP than the artificially aged samples.

Table 6. Historic overview of references and experiments of earlier vigour investigations using heat treatment

Author	Species	Varieties	Locations	Harvest year	Reference
Goodspeed	1, barley	-	-	-	Goodspeed, 1911[32]
Robertson	3, wheat	1, Marquis	-	3, 1922,1926 and 1931	Robertson, 1939[76]
	barley	1, Colsess			
	oat	1, Colerado			
Roberts	3, wheat	-	-	-	Roberts, 1960[72]
	barley				
	oat				
Roberts	1, wheat	1, Atle	-	-	Roberts, 1961[73]
Roberts and Abdalla	3, barley	1, Proctor	1, England a	-	Roberts and Abdalla, 1968[75]
	bean	1, Claudia	1, England b		
	pea	1, Meteor	1, England c		
Ellis and Roberts	1, barley	3, Proctor	-	1, 1972	Ellis and Roberts, 1980a and b[23-24]
		Golden promise			
		Julia			
Ellis and Roberts	4, barley	1, Proctor	-	2, 1963 and 1972	Ellis and Roberts, 1981[25]
	rice	6		-	
	maize	3		-	
	onion	1		-	
Aastrup et al	1, barley	2, Klages	2, USA	1, 1987	Aastrup et al, 1989[2]
		Ca108725*)	DK		
Møller and Munck	1, barley	7, Alexis	26	2, 1993 and 1994	Møller and Munck, 2002[Paper I]
		Ariel			*50 samples in total*
		Blenheim			
		Carula			
		Etna			
		Lysimax			
		Meltan			
Møller et al	1, barley	1, Alexis	17	1, 1994	Møller et al, 2002[Paper II]

*) Ca108725 = (Prisca *(Triumph *Emir))

To investigate in more detail if heat-resistance was influenced by environment a sample set of the variety Alexis grown at 17 different locations all over Europe were collected. Heat-sensitive as well as heat-resistant samples were also detected. In this experiment no systematic pattern was found between growing location (country) and heat resistance. As in the earlier investigation, no correlation between VP and GI on unheated samples was found in these naturally aged sample set either (Paper II).

In all 67 barley samples grown 1993-1994 have been analysed with the artificial ageing concept, of which 56 samples were naturally aged. Earlier investigations presented in the literature from 1911-1989 all studied artificially aged samples. It is apparent that the underlying assumption based on probability distributional assumptions of data has partly misled these investigators. With PCA it is possible to compare whole germination curves from different samples and assert (Papers I and II) that approximately 30% of barleys are extremely heat resistant. Earlier investigators should thus have found such samples, but have probably dismissed them as experimental errors.

It is concluded that the classical vigour loss model for heat treatment may be used as a worst-case prediction for germination (Paper II), but it does not address the variation found in practice, including the possible advantage of exploiting the naturally occurring heat resistance. It has not been possible to find a reproducible cause (genetic or environmental) for the occurrence of heat-resistant barley, but possibilities of a physiological mechanism is discussed in Paper II. The heat stress concept for an estimate of vigour should thus not be relevant for maltsters. Other stress factors such as development and survival under oxygen and water stress would be more relevant to study in malting barley.

4.3 Germinative classification

The germination curves (percentage of germinated grains for Days 1-8, Figure 7A) were used in an unsupervised PCA calculation for the 17 samples of the variety Alexis grown at different locations all over Europe (Figure 7B) previously discussed (Paper II). The PCA biplot shows that three samples grown in Spain (4E, 5E and 6E) are located to the left in the plot, whereas all the samples from Finland (SU) are located in bottom right corner. In the top right corner are samples from Germany (D), Czech Republic (CZ), The Netherlands (NE) and two of the samples grown in Denmark (DK). The squared “A” in the middle is the average of the 17 samples. Germination percentages from days 1 to 8 are shown in the plot as loadings. It is seen that Day 1 percentages are located in the top, Day 2 closer to the rest of the loadings, and Days 3-8 are located near each other in a group. This can also be seen from the germination curves in Figure 7A, while the curve form after 3-8 days is more leveled for the 17 samples.

To investigate in more detail why the samples are located as they are in the PCA plot in Figure 7B, the ID notation is changed in Figure 7C and refers to germination percentage after one day (bold) and after three days. The 3-day germination percentage is taken as the most conveniently measured representative for the close loading cluster 3-8. From this it is seen that there is a clear gradient in germination percentage after three days along the abscissa from left to right, and for germination percentage after one day along the ordinate from below to above. Now it is possible to ascertain the meaning of the “hidden” principal

components PC1 and PC2 in the plot. PC1 mainly describes the variation due to germination percentages Day 3. This axis can approximately be described to represent "viability" (quotation marks symbolise estimated viability). The germination percentage after three days does not increase very much to Day 8 (Figure 7A), and this factor can therefore give an estimate of living grains. PC2 mainly describes the variation due to germination percentage Day 1. Concluded from earlier investigations where GI and g%1 correlates well (Papers I and II), PC2 can be described as an expression of germination speed "vigour" (quotation marks symbolise estimated vigour).

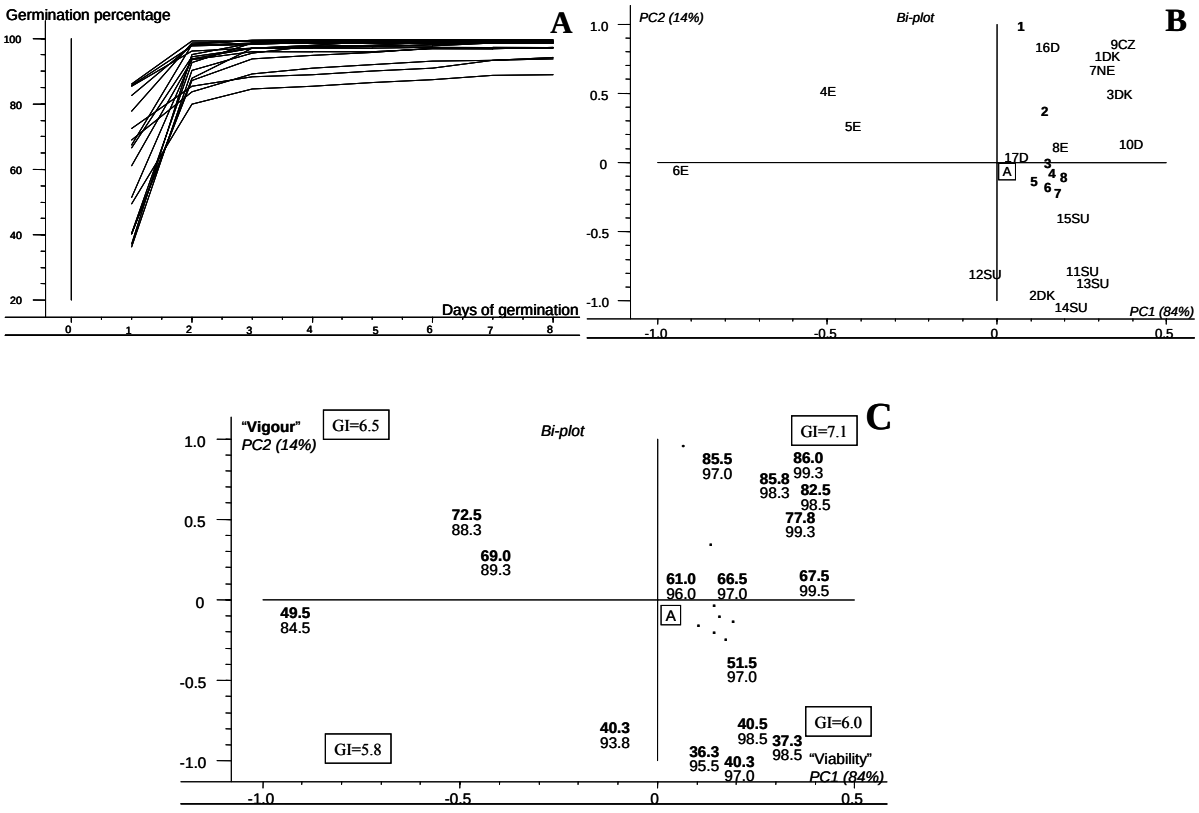


Figure 7. Multivariate evaluation of germination profiles g%1-g%8 for 17 untreated Alexis barley seed samples grown in EBC trials in Europe in 1994. **A.** Germination profiles for the 17 samples. **B.** PCA biplot (PC1:2) of the germination profiles for the 17 samples No. 1-17. Letters denote country symbols. Figures in bold are loadings g%1(**1**) –g%8 (**8**). **C.** Same PCA as B, but with identification of each sample position by figures for g%1 in bold and g%3 in normal font. GI = germination index for each quadrant. See text for discussion.

It is seen that it was possible to identify and name the latent components that describe most of the variation in the germinative data. In Figure 7C the mean germination index (GI) of the samples in the four quadrants are also displayed. These values support the idea that there is a vigour gradient along the ordinate. From this PCA example a hypothesis can be

generated that g%1 and g%3 could be used in a simple two-dimensional germinative abscissa-ordinate plot for malting quality classification. This was tested in a malting analysis of the 17 samples displayed in Paper III. It was further confirmed in studying another data set (Paper III) where 42 micro malted barley samples with a large variation in "vigour" (abscissa) and "viability" (ordinate) are plotted as seen in Figure 8A. The area with high germination percentage is enlarged in Figure 8B. Two quality levels of both "viability" (92 and 98%) and "vigour" (30 and 70%) are introduced in the plot. This leads to a division of the barley samples into 7 classes: 1.1, 1.2, 1.3, 2.1, 2.2, 2.3 and 3.0.

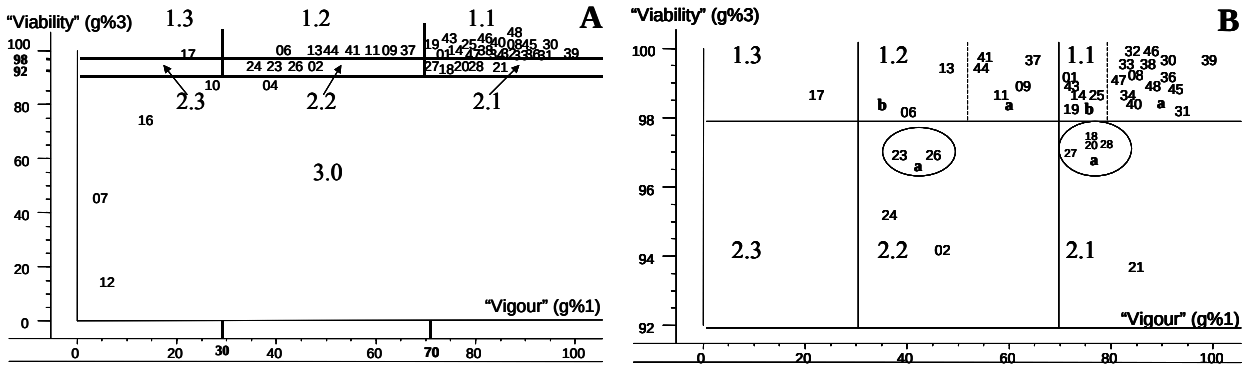


Figure 8. Germinative energy classification for a malting barley material (n=42). "Vigour" g%1 (abscissa) and "viability" g%3 (ordinate). **A.** Overview. **B.** Enlargement of classifcation plot with "viability" ≥ 92%. See text for discussion.

Average and standard deviation for barley and malt quality for these classes are shown in Table 7. It is clearly seen that the "vigour" component is complementary to the germinative energy component "viability" in differentiating the whole material with regard to extract % and to an even greater degree with regard to the critical quality criteria ß-glucan in wort, revealing the dependence of cytolytic activity in the malt on a swift and complete germination. The mean values of the malting barley classes reveal clear gradients in these important quality criteria. First, from the right to the left along the "vigour" axis from 79.0 to 37.3 g%1; (Class 1: extract 83.2-79.3%, ß-glucan in wort 180.4-266.5 mg/L; Class 2: extract 79.8-79.1%, ß-glucan in wort 199.2-297.4 mg/L) and second, from above to below along the "viability" g%3 axis from 99.2 to 95.9 germinative energy %; (Class 1.1 versus 2.1 extract 83.2-79.8%, ß-glucan in wort 180.4-199.2 mg/L and for Class 1.2 versus 2.2; extract 81.6-79.1% and ß-glucan in wort 202.0-297.4 mg/L). The feed barley Class 3 is clearly unsatisfactory for malting, as seen from the figures from the individual samples (Paper III) with mean figures of 70.1% for extract and 382.2 mg/L for ß-glucan in wort. GI has a high correlation of r=0.99 in this material with g%1.

It is concluded that the proposed two dimensional classification system with the barley material tested here is highly sensitive for predicting and discriminating the levels of extract (%) and ß-glucan in wort (mg/L) which are central parameters in the barley malt quality complex. It is suggested that further malting barley quality research should be directed to utilise vigour and viability information from germination curves for quality classification.

Table 7. Germination, barley and malt analyses for 41 reference barley samples classified in Figure 8. Mean and standard deviations of the different classes.

Class	n	"Vigour"	"Viability"	GI (g1-g3)	GH (g1-g3)	TKW	Extract	BGwort	BG	P
1:1	16	79.0 ± 8.6	99.2 ± 0.5	8.3 ± 0.6	56.3 ± 10.0	43.6 ± 2.1	83.2 ± 1.8	180.4 ± 39.7	4.1 ± 0.3	9.3 ± 0.5
1:2	7	50.3 ± 8.5	99.2 ± 0.7	6.6 ± 0.4	54.5 ± 7.6	41.1 ± 3.9	81.6 ± 3.0	202.0 ± 81.8	3.9 ± 0.3	10.0 ± 1.2
1:3	1	17.8	98.8	5.4	54.5	35.3	80.6	257.0	3.9	10.2
2:1	5	73.1 ± 5.3	96.6 ± 1.6	8.0 ± 0.4	54.8 ± 4.1	37.1 ± 1.8	79.8 ± 5.4	199.2 ± 52.2	4.0 ± 0.5	10.2 ± 0.5
2:2	4	37.3 ± 5.1	95.9 ± 1.4	6.1 ± 0.3	46.7 ± 1.1	40.7 ± 1.6	79.1 ± 1.9	297.4 ± 53.3	4.1 ± 0.2	12.7 ± 0.8
3:0	5	13.6 ± 14.6	63.0 ± 31.8	4.8 ± 0.9	42.6 ± 5.8	40.0 ± 1.9	70.1 ± 7.6	382.2 ± 94.4	3.8 ± 0.2	12.1 ± 1.5

There are two sources of inspiration for the suggestion to upgrade the germinative energy and capacity concepts with a supplementary classification system, as visualised in Figure 8. The first is the suggestion given by the PCA algorithm to divide the material into two separate, basically independent representations (PC's) of "vigour" and "viability". The second is the fact that maltsters and brewers, when considering GI as an expression of vigour, in practice always have to check this information against viability expressed as GE for 3 or 5 days, because a high GI can occur in spite of low GE, for example by inmix of dead kernels. It is thus concluded in Paper III that a univariate quality index like GI although composed of several quality parameters is not optimal as a quality estimate. A multivariate or at least a bivariate approach is needed for quality classification proper as demonstrated by the germinative classification plot featuring separate estimates for "vigour" and "viability".

4.4 Predicting the germinative parameters

It would be advantageous if it were possible to predict the indirect parameters such as germination properties from the set of ten physical-chemical manifest ones like hardness and morphological imaging barley data. The supervised chemometric algorithm PLSR can test this. The results according to germination properties are shown in Table 8 (Paper IV).

When using 42 samples in the prediction of g%1 GC, the first PLSR correlation (Table 8) is r=0.78 (one PC with the parameters TKW, HI, Volume, Width, Round, Intensity, BG and P ordered according to importance, RE=18.9).

Table 8. PLSR Jack-knife correlations between the 10 manifest parameters (TKW, HI, P, BG, width, length, area, volume, round, intensity) as (**X**) and hidden germination and malting variables (**y**). Samples with low viability GE (<92%) = underlined, medium viability GE (92-98%) = **bold**, high viability GE (>98%) = normal.

y	No.	Step*	r	RMSECV	RE	PC**	n	Total outlier samples removed	Significant variables***
g%1 GC	a01	0	0.78	16.33	18.9	1	42		TKW, HI, Volume, Width, Round, Intensity, BG, P
g%3 GC	a02	0	0.63	2.79	18.6	1	42		TKW, HI, P, Width, Round, Volume
GH GC	a03	0	0.77	8.17	17.1	1	42		HI, TKW, Round, Width, Volume, Intensity, Length, P
g%1 GE	a04	0	0.73	17.09	18.1	1	42		P, Round, Length, Width, Volume, Intensity
g%1 GE	a05	I	0.84	12.41	14.5	1	35	M07, A08, A09, A12, **B21**, A37, B41	P, Round, Length, Width, Volume
g%1 GE	a06	I	0.94	7.77	9.1	4	35	M07, A08, A09, A12, **B21**, A37, B41	P,Width, Round, Length, Volume
g%3 GE	a07	0	0.39	13.98	16.5	1	42		P, TKW
	a08	I	0.56	3.60	14.5	1	40	M07,A12	P, HI, TKW
	a09	II	0.73	1.74	14.5	2	39	M07,A12, M16	P, INT, TKW
GH GE	a10	0	0.70	6.17	14.9	2	42		Length, Round, P, Area

* Step of outlier selection from influence plot ** Minimum value of residual validation variance *** Variables ordered after degree of importance

The same PLSR method as for "vigour" g%1 (GC) is used to predict "viability" g%3 (GC) from the set of the ten manifest variables of which TKW, HI, P, Width, Round and Volume are significant with a correlation of r=0.63 (RE=18.6, Table 8). The sequence of significant variables according to their importance is almost identical between g%1 and g%3. However, g%1 gives much better predictions than g%3.

With respect to GE determined after three to six years of storage, the PLSR correlation to the set of the ten physical-chemical variables is r=0.73 (one PC, RE=18.1) with a changed sequence of significant variables compared to GC (P, Round, Length, Width, Volume, Intensity). The influence plots (Paper III) shows seven outliers (removed in two steps, Table 8), two of which have extremely low "viability" g%3 (M07 and A12) and B21 with reduced "viability" g%3. When the seven outliers are removed, the correlation improves to r=0.84 (one PC, RE=14.5) with r=0.94 for four PC's (RE=9.1). The pattern of importance is unchanged. There is a clear tendency for low "viability" outliers in the other correlation models given in Table 8. This is especially apparent in the "viability" g%3 GE prediction in Table 9 where the correlation coefficient is improved from r=0.39 (one PC, RE=16.5) to r=0.73 (two PC, RE=14.5) when removing three outliers which all have low "viability" (g%3). It is also concluded in both the GC and GE cases that the prediction of "viability" g%3 from the ten parameters has a significantly less correlation coefficient than that of "vigour" g%1.

We can thus conclude that "vigour" g%1 to a surprisingly great extent can be predicted by the set of the ten physical-chemical parameters. From this information a new hypothesis can be generated, as described in Paper III. It seems that there are two major functional factors that influence germination quality and malt modification. The two factors which have to be considered when detecting malt quality are 1) the physiological vigour of the embryo and 2) the physical-chemical structural characteristics of the endosperm. The

underlying physical-chemical factors, which are assumed to affect germination rate of the seedling in the second category, are suggested to be related to the availability of the substrate needed for germination. Seed morphology, hardness, cell wall thickness (β-glucan content), protein and starch (extract potential) are the factors included. These physical-chemical factors seem to play a much more pronounced role in influencing "vigour" than the physiological condition of the embryo in a malting barley material within a range of acceptable viability g%3 for malting.

From the above-mentioned investigations it was seen that ten physical-chemical manifest parameters were able to roughly predict germination properties of a sample. Using NIT spectroscopy a physical-chemical fingerprint is likewise obtained. The ten physical-chemical manifest parameters are here expanded to 100 variables. Therefore, it is expected that NIT will also be able to predict germination.

In Table 9 a relatively high correlation of r=0.86 for "vigour" g%1 GC (ten PC's) and r=0.76 for "viability" g%3 GC (four PC's) is seen using a first derivate of NIT spectra. As with the prediction of hidden variables using the ten manifest parameters in Table 8 there is a clear tendency that g%3 gives lower predictions with NIT than g%1 and that outliers have a low viability. The low vigour outliers in **y** in the NIT correlations can obviously not predict the physiological process of germination connected to the embryo in dry, not-germinated kernels. However, the hypothesis drawn from Table 8 is supported by demonstrating that "vigour" g%1 (GC) can be predicted by a separate set of physical-chemical variables (i.e. NIT).

The most important variable in the prediction of "vigour" (GC) in Table 8 is HI, followed by Volume, Width, TKW, Round and Protein. Most of these parameters are obtainable with NIT (Paper IV) with correlations of r=0.94 for HI, r=0.74 for Width, r=0.92 for TKW, r=0.77 for Round and r=0.97 for protein. The rather good prediction for "vigour" from NIT measurements is therefore expected to stem from the physical-chemical properties manifest in the grains, which are essential for access of nutrients to the embryo influencing germination speed (g%1), as discussed in Paper III.

The use of two independent instrumental screening analyses (GrainCheck, hardness and chemical analyses versus NIT) to measure the physical-chemical status of the samples supports the notion that germination and malting properties are heavily dependent on the physical and chemical properties of the barley samples which may be used for prediction of these indirect hidden parameters (Papers III-IV). We therefore conclude that these findings support a new way of looking at the relevance of germination parameters in brewing science which also opens up new possibilities for prediction of important variables by instrumental methods for direct *on-line*, *at-line* quality control in industry by NIT.

Table 9. NIT (1. der.) prediction of germination, malting data and chemical-physical data for samples of Alexis, Blenheim and Meltan. Samples with low viability GE (<92%) = underlined, medium viability GE (92-98%) = bold, GE >98% = normal.

y	No.	Step*	r	RMSECV	RE	PC**	n	Outliers***
g%1 (GC) 1993-99	b01	0	0.86	12.4	16.4	10	52	
g%3 (GC) 1993-99	b02	0	0.50	3.0	31.2	1	52	
	b03	I	0.76	1.5	15.8	4	49	M17,**B21**,**B28**
g%1 (GC) 1993-96	b04	0	0.92	10.4	12.1	10	42	
g%3 (GC) 1993-96	b05	0	0.59	2.9	19.3	1	42	
	b06	I	0.77	1.6	17.3	2	39	M17, **B21**, **B28**
GH (GC) 1993-96	b07	0	0.78	8.1	17.1	5	42	
g%1 (GE) 1993-96	b08	0	0.74	16.8	17.8	4	42	
	b09	I	0.77	14.9	15.7	4	41	~~A12~~
	b10	II	0.80	13.6	14.3	3	38	~~A12~~,~~M16~~,**A20,A27**
g%3 (GE) 1993-96	b11	0	0.31	14.4	17.0	1	42	
	b12	I	0.68	1.8	15.7	1	39	~~M07,A12,M16~~
	b13	II	0.80	0.9	3.4	3	37	~~B04,M07,A10,A12,M16~~
GH (GE) 1993-96	b14	0	0.59	6.9	24.4	4	42	
	b15	I	0.75	4.4	17.1	4	37	**B21**,A31, A37, A39, M44

* Step of outlier selection from influence plot *** Total outlier samples removed from correlation

** Minimum value of residual validation variance

With the strategy of focusing on the structural factor by PLSR and identifying the physiological (viability) nature of the outliers in **y** the surprising conclusion is reached that germination speed "vigour" in this investigation (Papers III and IV) has a much more pronounced structural component than physiological within the range of viability which is characteristic for malting barley. The g%3 variable also reflects seed structure to some degree but with a much lower correlation to the structural parameters than g%1. It is therefore concluded that the structural physical-chemical factor is the main determinator for vigour, defined as the early growth rate of the emerging plantlet in barley of malting grade (Paper III). As in Table 8 with the ten manifest parameters NIT predictions of GE g%1 and GE g%3 are improved by removing the outliers which in NIT spectroscopy all were found to be low in "viability" g%3.

These preliminary results can be interpreted as follows: Substrate availability for the germ is of importance for fast sprouting and is related to the function of how to "unlock" the complex physical and chemical structure of the food store – the endosperm. This function should also be identical with the aims of the maltster to obtain a fast malt modification (a low malt modification resistance) in dissolving cell walls and in enzyme spreading in the endosperm. Fast germination, i.e. high "vigour", should therefore be operative for the malsters as an indicator of an efficient malt modification representing the structural functional factor related to physics and chemistry. Thus, by securing a high "viability" the

structural functional factor becomes limiting in malting and brewing performance (Paper III).

The outliers with low "viability" (g%3) that have been found in the models in Tables 8 and 9, are deviates in **y** (g%3) and not in **X** (NIT or the ten physical-chemical parameters). When removal of outliers determined in **X** no improvements in correlation coefficients are found. The detected outliers in **X** do not show low "viability". This indicates that neither NIT nor the ten physical-chemical parameters can be used for predictions of "viability" in unknown samples. This is in accordance with the initial hypothesis that physical-chemical analyses should not be able to trace the physiological properties (low viability). A separate method for "viability" is thus needed as a supplement to "vigour" (g%1) to remove low "viability" outliers. The germination percentage after 3 days (or more correctly 8 days germination) or the Tetrazolium test could be used for such a purpose. By removing low "viability" outliers using the Tetrazolium test, the NIT analysis can be used to predict "vigour" from the high "viability" samples. This could be done within one-two hours.

Alternatively to NIT measurements the maltsters could germinate samples in 24 hours as well as determine the percentage of living grains with the fast tetrazolium test (approximately one hour analysis time), and from here obtain an abscissa-ordinate classification where samples divide according to malt quality. The g%1-g%3 classification with data derived from germination Tetrazolium tests and/or NIT calibrations should be able to be developed to a convenient tool in classification of barley for malt.

The advantages and possibilities using multivariate data analysis to explore different kinds of data predictions of indirect variables by fast instrumental methods (grain image analysis and spectroscopy) have thus been shown.

5. Multivariate analysis in plant breeding and biotechnology

The possibilities and advantages of using multivariate data analysis in plant breeding and biotechnology on regular data sets and in combination with instrumental methods are discussed in the following.

5.1 Field trials: A PCA adaptability biplot

To exemplify the possibilities using multivariate data analysis in plant breeding, a simplified example is used regarding gene-environment interaction in the study of germination and agronomic field characteristics. The 51 samples (Table 10) are of seven varieties and the variables determined are agronomical properties (number of plants, number of ears, yield (kg/plot) and grading) as well as germination characteristics (number of seeds germinated on Days 1, 2, 3 plus total germination after 3 days as well as GI). The unsupervised PCA is used to obtain an overview of one dataset, as in the example in Section 3. The agronomic and germination variables are analysed by making spectra of variables for each of the 51 samples, the principle of which is demonstrated for the chemical analyses in Figure 2C. The algorithm determines which samples are similar and different, making it possible to get a holistic representation of the total data set of this material.

Table 10. Mean and std.dev. for germination parameters and field data from samples of Alexis, Carula, Meltan grown in three locations: Scania, Zealand and Funen

	All samples				Zealand			
	Total	**Alexis**	**Carula**	**Meltan**	**Total**	**Alexis**	**Carula**	**Meltan**
n	51	14	9	10	16	4	3	3
g%1	38.9±24.8	51.2±21.7	36.5±11.7	36.8±27.8	36.1±12.5	40.1±2.3	32.2±3.0	17.0±5.4
g%3	96.6±2.7	97.4±1.2	99.0±0.7	93.8±2.8	97.7±1.5	96.8±1.6	99.3±0.7	96.8±0.6
GI	6.3±1.1	6.8±1.1	6.1±0.5	6.3±1.3	6.1±0.5	6.2±0.2	6.0±0.3	5.3±0.1
plant	224±44	226±33	193±55	232±35	238±40	240±18	227±72	244±16
ear	762±148	798±144	701±122	808±205	840±117	881±68	695±118	943±109
yield	7.55±1.06	7.59±1.19	7.56±0.86	7.95±1.10	7.30±0.63	7.72±0.17	6.61±0.66	7.94±0.21
grading>2.5	82.1±9.5	87.9±3.3	91.3±2.3	80.8±7.6	83.8±9.0	90.4±0.7	93.6±0.7	85.3±6.8
	Scania				**Funen**			
	Total	**Alexis**	**Carula**	**Meltan**	**Total**	**Alexis**	**Carula**	**Meltan**
n	18	3	3	3	17	7	3	4
g%1	15.3±12.8	23.8±5.5	26.6±4.9	14.3±0.8	66.4±12.6	69.4±12.8	50.6±6.3	68.6±7.0
g%3	95.7±3.4	97.3±1.1	98.5±0.7	91.0±2.4	96.5±2.4	97.8±0.9	99.2±0.5	93.8±1.7
GI	5.4±0.4	5.6±0.2	5.7±0.2	5.3±0.1	7.6±0.8	7.7±0.8	6.7±0.3	7.7±0.5
plant	232±58	265±31	145±38	262±34	203±14	201±14	207±15	200±19
ear	613±88	545±9	592±32	549±21	846±90	858±44	817±83	902±143
yield	6.97±1.20	5.66±0.16	8.07±0.52	6.61±0.58	8.40±0.62	8.34±0.69	8.01±0.44	8.95±0.52
grading>2.5	79.1±11.6	89.7±2.2	91.3±0.7	76.3±7.2	83.6±6.7	85.7±3.2	88.9±1.7	80.7±8.2

In Figure 9 the PCA biplot is shown and it is seen that the varieties divide into three groups according to location. To the left are samples grown in Scania, in the middle those grown on Zealand and to the right those grown on Funen. Carula (Ca) differs from the other varieties and Meltan (Me) is the other extreme. The Carula samples for all locations are located together in the middle/bottom of the plot, indicating less sensitivity for environmental conditions than the Meltan variety, which divides according to location (Figure 9).

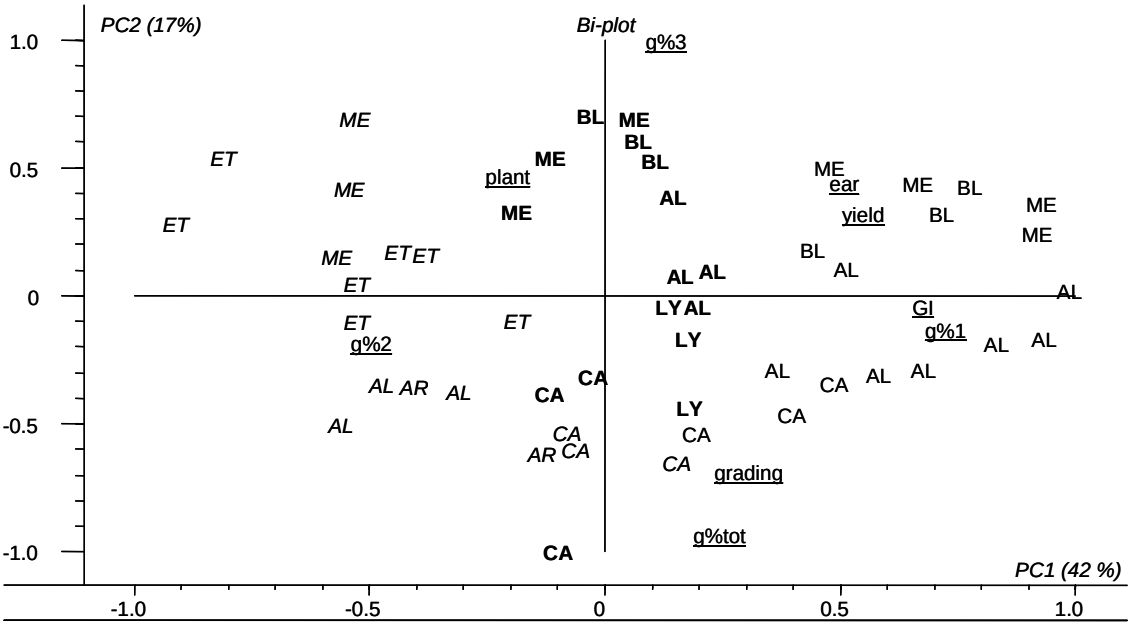


Figure 9. PCA biplot (PC1:2) of barley samples grown in three locations (*AL*=Scania, **AL**=Zealand, AL=Funen) in 1994. Varieties: AL=Alexis, AR=Ariel, BL=Blenheim, CA=Carula, ET=Etna, LY=Lysimax, ME=Meltan. Loadings (underlined): g%1-3=germination % day 1, 2, 3, plant=plants/m^2, ear= ears/m^2, yield=yield/parcel, grading=%grains > 2.5mm.

The variables are also shown in the PCA biplot in Figure 9, revealing the relationship between samples and varieties. It is seen that the variables ears, yield, g%1 and GI are located to the right, as are the samples grown on Funen. This implies that samples from Funen have higher numbers of ears, higher yield and a faster germination than those from the other two locations. Samples from Scania are placed to the left, as are the variables numbers of plants and g%2. Samples from Scania germinate more slowly than those from the other locations, which is seen in the plot where the samples are placed opposite g%1 and GI and near g%2. Furthermore, the plot shows that GI and g%1 are placed near each other, indicating a high correlation (r=0.99). Similarly, number of ears is placed near yield with a correlation of r=0.82. PC1 describes the growing location characterised by

germination speed (g%1 and GI), number of ears and yield, whereas PC2 characterises the varieties and are stretched by g%3 and total germination as well as grading properties.

In Table 10 an overview of mean and standard deviation of the variables is shown for the Alexis, Carula, Meltan and the average of the 51 barley samples divided according to growing location. The overview strengthens the conclusions from above that Carula is nearly resistant to environmental influence, whereas Meltan is most sensitive.

These results demonstrate the usefulness of PCA in enabling an unsupervised interactive and fast overview of a mixed data material as a basis for an in-depth discussion in order to evaluate the stability of the barley varieties. This is expected to be useful in plant breeding. Although this example only involves a few variables, the PCA plot facilitates a rapid overview of the individual nature of each sample, thus generating ideas (hypotheses) that can be checked by classical statistics and by pattern recognition multivariate correlation studies (e.g. PLSR).

5.2 The barley endosperm as a data interface for expression of specific genes

It is previously shown (Section 4) that NIT spectroscopy (850-1050 nm) on whole kernels measuring 200 wavelength, whereby 100 variables (sparing each second wavelength) are obtained, is a strong tool for classifying a sample, because the spectra give a valid fingerprint of the physics and chemistry of each sample[58] (Papers V-VII). In this example the wavelength area is increased using NIR spectroscopy (400-2500 nm) on milled barley flour, obtaining 2100 variables for every sample, which is reduced before calculation to 1050.

Figure 10A shows the MSC-corrected NIR spectra from 49 samples of barley grown in greenhouses, outdoor pots and in the field. The barley material consists of normal barley varieties (N) and two mutants (*lys5f* and *lys5g*) selected at Risø as high-lysine lines with the dye-binding method[20]. The mutants are in the same locus and are identical to Risø mutant 13 (background Bomi) and mutant 29 respectively (background Carlsberg II)[20]. In this material there are additional recombinants (doubled monoploids) from crosses between these mutants and normal barley carrying these genes[20].

By using the unsupervised algorithm PCA, patterns according to genetic parameters as well as for environment are seen (Figure 10B). Normal samples are located diagonally in the top left corner, whereas mutants and recombinants are spread diagonally in the bottom right corner. Furthermore, it is seen that samples grown in greenhouses are located diagonally in the bottom left corner, samples from field are located in the middle and samples grown in outdoor pots are placed diagonally in the top right corner.

In an unsupervised PCA, where one does not know the origin of the samples, mean spectra from the sample groups can now be formed and inspected using prior spectro-chemical knowledge[64] (Paper VII) in order to define analyses for chemical validation.

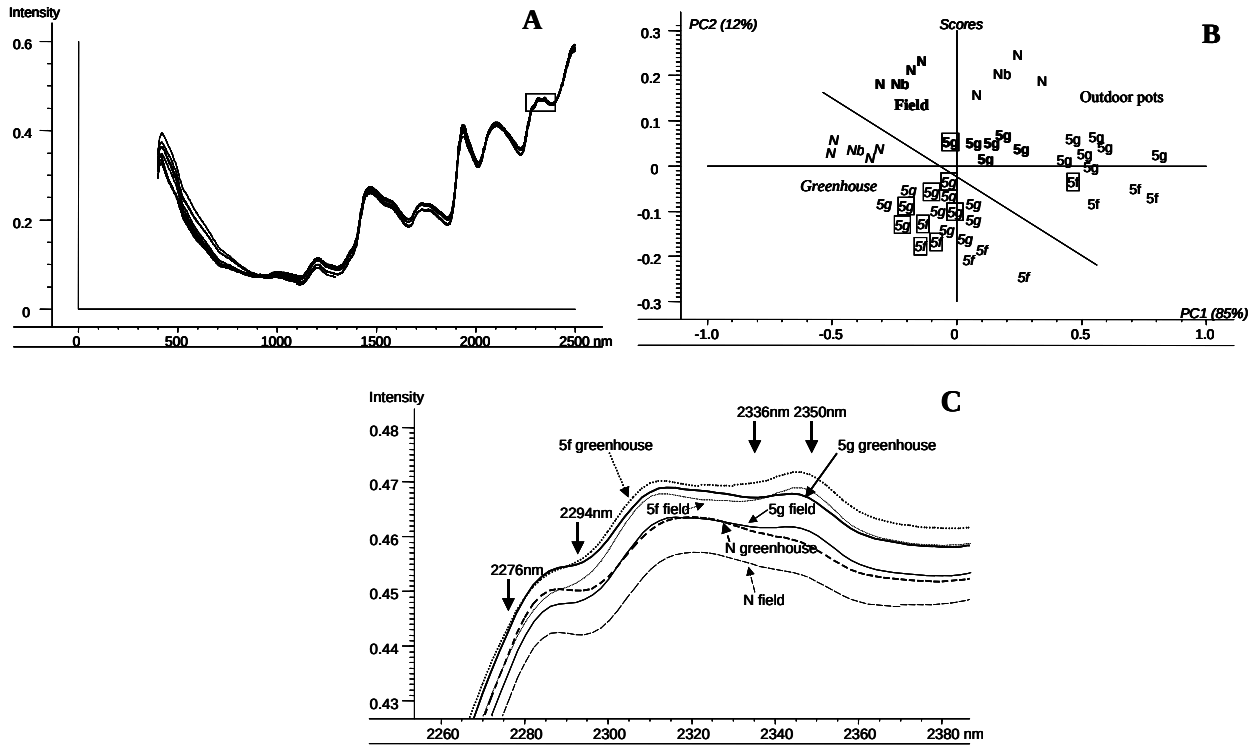


Figure 10A. Average NIR spectra (MSC) from normal barley grown in greenhouse (n=5) and field (n=8), *lys5f* mutants and crosses grown in greenhouse (n=6) and outdoor pots (n=4) and *lys5g* mutants and mutant recombinants crosses grown in greenhouse (n=13), outdoor pots (n=7) and field (n=6). **B.** PCA (PC1:2) score plot of NIR spectra (MSC) for 49 samples of normal barley (N, Nb=Bomi) and mutants and mutant recombinants crossings of *lys5f* (*5f*) and *lys5g* (*5g*) grown in three environments: *greenhouse*, **field** and in pots outdoors. Squared samples are original mutants **C.** Average spectra of normal barley (N) and mutants of *lys5f* (*5f*) and *lys5g* (*5g*) grown in greenhouse and field.

Because samples located near each other in a PCA plot indicate similar pattern of variables, here NIR wavelengths, mean spectra from the different groups can be characterised by visual inspection. In Figure 10C the area from 2260-2380 nm is enlarged for the average spectra of mutants of *lys5f*, *lys5g* and normal barley samples grown in greenhouses and in the field. It is easily recognised that the average curve form of the normal barley (N) is quite different from those of the mutants. However even between the mutants of *lys5f* and *lys5g* differences in their average curve form are seen. In comparing these curves with

Figure 1A, where the corresponding wavelength areas[64] of different chemical compounds and bonds are shown, it is seen that the mean spectra in Figure 10C are different from each other at the wavelength about 2276nm correlating to starch[64], at 2294nm describing the content of amino acids[64], at 2336nm which correlates with cellulose content[64] and at the area around 2350nm which is related to cellulose[64] (2352nm) and unsaturated fat[64] (2347nm). The spectra cannot be used as a direct inspection of content of, for example, amino acids without prior collection of spectra from many samples that can be correlated to the reference values, ending up with a model that can predict the reference value from new unknown sample spectra. However, the comparison of the mean spectra in Figure 10C with the spectra in Figure 1A can be used roughly to get ideas as to where the samples differ from each other. From here it is possible to make hypotheses about differences in chemical composition and to evaluate these by choosing proper reference analyses for final validation.

When comparing growing locations in Figure 10C it is seen that the curve form for each mutant genotype is approximately the same, but the offset base line indicates growing location where the greenhouse has a higher intensity than the field. As indicated from the PCA in Figure 10B, the *lys5f* and *lys5g* mutations are drastic enough to obtain full penetrance in the recombinants that are classified together.

The differentiation between the genetic and environmental effects have also been demonstrated in Papers V and VII where the small wavelength area between 2290 and 2360 nm provides unique spectroscopic signatures for the *lys3a* gene and the wild type, while environmental effects are registered as offsets from the baseline.

It can thus be concluded that by using NIR spectroscopy and the unsupervised PCA algorithm it is possible to detect environmental and genetic differences within a sample set, where the curve form is more related to genetics and the offsets from the baseline mainly describe environment.

Although it is possible to some extent to evaluate spectra by inspection through the human brain it is not possible to grasp the full spectral information by visual inspection alone. Global and local PLSR's (e.g. interval i-PLS,[63]) are needed as well as multivariate correlation studies with chemical analyses in PLSR for validation. An example of the use of i-PLS is given in Paper V.

To further investigate the possibilities of NIR and PCA the two-locus gene system from Paper VI is taken into account. Here it was seen that it was possible to classify samples into four clusters: normal barley (N), *lys3* mutants (*3a*, *b*, *c*, *m*), *lys5* mutants (*5f*, *g*) and double recessives (*3a5g*) by using spectroscopic data. The grouping was explained by chemical analyses which revealed *lys5* locus as extremely high ß-glucan-producing at an almost normal amino acid composition, while *lys3* produced a drastically changed amino acid

pattern (real high-lysine barley). If NIR spectra of seven samples of five new genotypes with unknown carbohydrate chemical composition (*lys4d*, 16, 449, w1 and w2) are included in the data set for testing the classification method, the new samples are included in the original clusters (Figure 11). Mutant 16, 449 and w1 were all included in the ß-glucan rich cluster around *lys5*. Sample w2 is included in the upper part of the normal barley cluster, whereas *lys4d* is included in the *lys3* cluster in the bottom right of the plot. To explain why the new samples are located as they are, the chemical reference analyses are then made and interpreted. From this it is seen that the samples located in the ß-glucan rich cluster of *lys5* samples actually have a high content of ß-glucan (mutant 16=15.2%, mutant 449=13.5% and w1=15.4%). The *lys4d* sample has a low A/P index and a low ß-glucan content, as is the case with *lys3a*, *b* and *m*. This is also confirmed by the amino acid composition, where it seems that *lys4d* is changed in the direction of *lys3a* compared to the less amino acid changed mutant 16 (Paper VI). Samples w1 and w2 are expected to be waxy mutants according to notation in our gene library, but the amylose content indicates that w1 has an almost normal amylose content of 20.3 %, but is very high in ß-glucan (15.4%) and low in starch. Thus, w1 is not a classical waxy high amylopectin/low amylose mutant with slightly increased ß-glucan content, but rather a low starch/high ß-glucan mutant as the mutants in the *lys5* cluster, as described in Paper VI. The other waxy mutant w2 located in the upper high ß-glucan part of the cluster of normal barleys is waxy with high amylopectin and a low amylose content (4.2%). It has a ß-glucan content on the high side (7.0%) compared to the normal barley samples (average 6.5%). It is noteworthy that the *lys3c* allele which contrary to the other three *lys3* alleles has normal ß-glucan content even though it is classified in the low ß-glucan/high lysine cluster in the right corner of the PCA plot in Figure 11. This points out a limitation of the method.

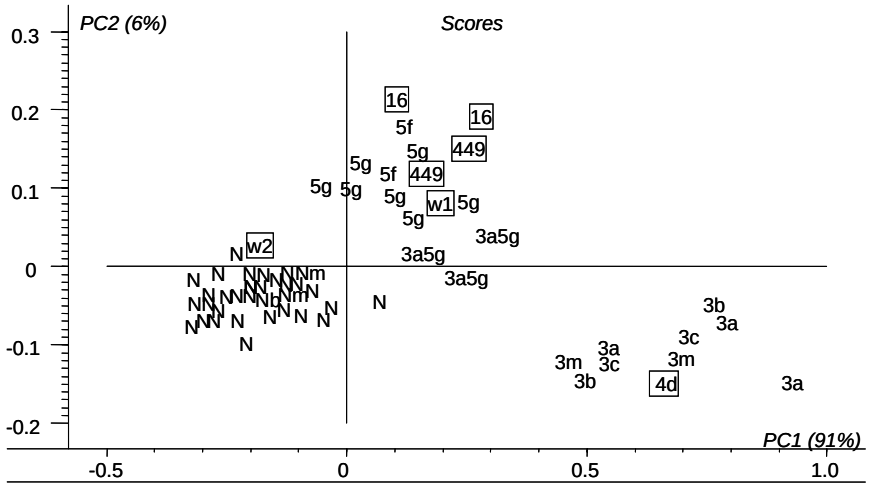


Figure 11. PCA (PC1:2) of NIR spectra (MSC) from the 54 barley samples grown in greenhouse and 7 samples of 5 different mutants (Paper VI).

The NIR spectrograph detects the genetic covariance (pleiotropy and linkage) where the investigated genes give a characteristic chemical-physical fingerprint of the seed phenotype (endosperm). Because each mutant and its gene background are approximately isogenic, it is possible to get a clear-cut spectral fingerprint of each mutant. The fingerprint is modified by crossing and selection, but the *lys5g* and *lys5f* mutants are drastic enough to allow a classification of the segregating homozygotic genotypes after crosses approximately in the same class as the original mutants (Paper VII). Also another mutant No. 95 was found to be included in the high ß-glucan, low starch class (Paper VI).

Thus, by evaluating a mutant gene bank by NIR spectroscopy and chemometrics a new class of six low starch, high ß-glucan genes were detected (Paper VI) where starch was seemingly exchanged for ß-glucan, an endosperm component of great interest for human nutrition[78]. New extreme high ß-glucan genotypes on a previously unobtainable level were detected with ß-glucan content up to 15-20% (*lys5f*). The sum of starch and ß-glucan was only slightly less in these genotypes compared to normal barleys. As discussed in Paper VI, further genetic research should be done to identify the number of loci involved (at least two; *lys5* and *lys16,* according to Doll[20]). The finding of this new class of endosperm low starch, high ß-glucan metabolic mutants generates a hypothesis pointing toward a close connection between starch and ß-glucan regulation that should stimulate further studies with these mutants using biotechnological tools to define the enzymes which should be monitored in the early stage of seed synthesis.

The above discussion has shown the strength of NIR spectroscopy and multivariate data analysis in classifying variation in different kind of samples and datasets as well as in predicting chemical properties from spectra. The latter quality will be further confirmed in the following example.

Amino acid composition, starch, ß-glucan, amide, dry matter, protein, fat, insoluble and soluble fibres were determined in 34 samples of *lys3a*, *lys5f*, *lys5g*, *lys3a5g* and normal barley described in Paper VII. In Table 11 correlations and predictions are compared for ß-glucan, starch and amide. By simple diallel correlation calculation between the three variables, it is seen that ß-glucan and starch indicate a negatively correlation. This is in accordance with the earlier mentioned theory (Paper VI) that *lys5f* and *lys5g* samples compensate for low starch content by increasing the content of ß-glucan.

If the gross chemical analyses (protein, fat, insoluble and soluble fibre, dry matter, ß-glucan, amide and starch) are used as **X** predicting ß-glucan, starch or amide as **y** (by leaving the predicted variable out), correlation coefficients from 0.78 (ß-glucan) to 0.96 (amide) are seen (Table 11).

Table 11. Diallel correlation and PLSR predictions of BG, starch and amide from the other gross chemical analyses, amino acid analyses and NIR spectra.

	X	y	r	RE	PC	n	Significant variables
Diallel correlation's	BG	Starch	-0.76				
	BG	Amide	0.47				
	Starch	Amide	-0.44				
PLSR predictions	Seven chemical analyses	BG	0.78	17.4	1	30	Starch, DM, Protein, Amide
Jack-knife		Starch	0.91	12.4	1	27	Protein, BG, Fat, Amide, DM
		Amide	0.96	1.4	5	29	-
	17 amino acids	BG	0.86	15.0	3	30	ser, tyr, val, met, ala, his, gly
		Starch	0.77	35.1	2	25	ile, met, leu, val, asp, lys, arg
		Amide	0.87	13.0	3	32	met, ile, thr, phe
PLSR predictions	NIR (1050 wavelengths)	BG	0.99	4.5	8	30	Significant variables not
		Starch	0.95	18.6	2	27	calculated
		Amide	0.94	13.3	5	32	

A few extreme samples have been removed as outliers but are not further described here. In the three PLSR predictions it is interesting to notice, that the model predicting amide has no significant variables according to the Jack-knife calculation, however, still ending up with a high correlation coefficient (r=0.96). This could be evaluated as the seven measured chemical properties all are equally needed as a part of a complex to predict amide, so that no specific variable is more important than the others.

If the 17 amino acids are used to predict ß-glucan, starch and amide, ß-glucan increase in correlation coefficient whereas starch and amide decrease, however still with correlation coefficients larger or equal to 0.77. The relative errors are quite large in these models. However, these predictions still indicate that the amino acid dataset also is part of a great chemical web within each individual barley phenotype (grain sample) which can be used as a data interface for classifying specific gene and environmental effects as earlier discussed (Paper VII).

From the above it seems that NIR spectroscopy would be an advantage to use because it gives a more complex physical-chemical fingerprint of every sample in a large wavelength area. Looking at the PLSR correlations in Table 11 high correlation coefficients are obtained (a few outliers are removed). The PLSR prediction for ß-glucan has a high number of PC's (8) indicating a complex model, however with no outliers (the two samples missing has not been analysed due to lack of material).

All predictions described here are calculated on a limited number of samples, and are therefore only indicative. However, these predictions show that the chemistry of a biological material is of a covariate character dependent on many chemical bonds. This shows the usefulness and strength of NIR spectroscopy in predictions of chemical composition as well as in the pattern recognition classification between different samples and genotypes.

5.3 Data breeding

To investigate if spectroscopy and chemometrics also can be used in plant breeding an example is given where *lys3* mutants are crossed with normal barleys to obtain lines with high lysine content and improved kernel quality as well as with increased starch content and improved yield. A library of original mutants, normal barley lines and yield-improved crosses[57] was evaluated by NIR.

A section of this data bank from material grown in the field in 1991 is demonstrated in a PCA (Figure 12A) which shows the pattern of the 14 NIR spectra 400-2500nm characteristic for 6 normal barleys, 2 mutants and 7 crosses in F5-10.

It is clearly seen that the samples divide into two clusters: one with normal barleys (Bomi, Minerva and Triumph) to the left, and the *lys3* mutants *lys3a* and *lys3m* (squared) and their high-lysine segregates to the right. Lysiba and Lysimax are commercial high-lysine varieties with verified improved yield (about 10-15% better than the original mutant). The lines 502, Lysiba and Lysimax are thus *a priori* defined as positive selections. They are in the NIR PCA situated closer to the normal samples than for example line 531, which is a negative selection confirmed by its location to the far right.

If an unsupervised PCA is performed on chemical data (protein, amide, A/P, starch, ß-glucan and rest = (100- (protein + starch +ß-glucan)), almost the same pattern (Figure 12B) as with NIR spectra (Figure 12A) is seen, indicating the physical-chemical basis of the NIR measurements.

The PCA of chemical data in Figure 12B shows that samples of Triumph, which are known to have a very high starch content, are located close to the starch variable, indicating that these samples are highly influenced by this variable. Opposite, still to the left, are Minerva and Bomi samples highly influenced by amide, A/P and ß-glucan. To the very right the mutants *lys3a* and *lys3m* are placed with the high-lysine segregates and together with the variables protein and rest. The samples of Lysiba and Lysimax (Group 1) are located closer to the normal varieties than the others, indicating a positive selection (with regard to starch) as in the PCA with the NIR spectra. Samples 502 and 556 (Group 2) are located close together, whereas samples 505, 531 and 538 (Group 3) are placed in between the mutants.

When comparing the chemical data of these four groups one can easily see (Table 12) that Group 1 with Lysiba and Lysimax has protein and starch content more closer to normal varieties. The amide content and A/P index is, however, more closely related to the mutants. This means that they still are likely to have a content of essential amino acids as high as the original high-lysine mutants combined with an increase in starch (mean) from 48.7 to 52.6 %. Group 2 is intermediate in starch content with reduced protein content,

while Group 3 has as high protein content as high as the original mutants with only a slight increase in starch.

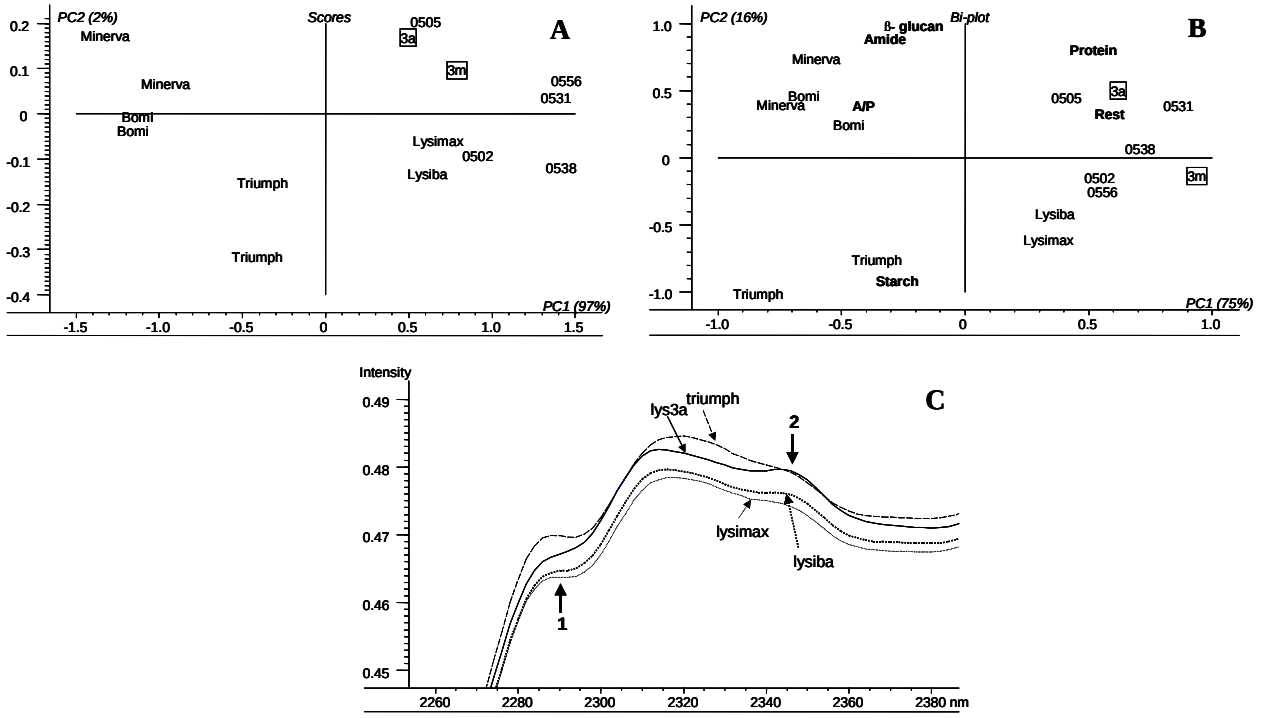


Figure 12A. PCA score plot of NIR (MSC) spectra from normal barley (Bomi, Minerva, Triumph), mutants (*lys3a*, *lys3m*) as well as high lysine recombinant lines (0502, 0505, 0531, 0538, 0556, Lysiba, Lysimax). **B.** PCA biplot of chemical data (protein, ß-glucan, amide, A/P, starch) at the same material.**C.** Comparison of the MSC NIR area 2260-2380nm for samples of Triumph, mutant *lys3a* and high lysine recombinant lines Lysiba and Lysimax). Numbers are referred to in text.

Table 12. Average and standard deviation of chemical data for the five groups.

	Normal (n=6)	Group 1	Group 2	Group 3	Group 4
Protein (P)	11.3±0.4	11.7±0.1	11.7±0.1	12.6±0.2	12.5±0.2
Amide (A)	0.28±0.03	0.21±0.007	0.21±0.007	0.22±0.02	0.23
A/P	15.5±0.9	11.0±0.3	10.9±0.4	10.7±0.8	11.4
Starch	54.6±2.5	52.6±0.5	50.0±0.1	49.4±1.5	48.7±0.2
ß-glucan	4.7±1.1	3.1±0.1	3.1±0.2	3.1±0.3	2.8±0.5
Rest (100-P+S+BG)	29.5±1.8	32.7±0.5	35.3±0.3	34.9±1.8	36.1±0.5

Group 1= Lysiba, Lysimax Group 2= 502, 556 Group 3= 505,531,538 Group 4= *lys3a*, *lys3m*

The wavelength area used in earlier investigations (2260-2380nm) is compared for spectra from samples of the normal barley Triumph, *lys3a* and from the positive selected lines Lysiba and Lysimax. From Figure 12C it is seen that the curve forms for Triumph and *lys3a* are very different, and that those of Lysiba and Lysimax are intermediate between

lys3a and normal barley. The improved high-lysine lines have adopted some characteristics from the normal curve form. In the wavelength area 2285 to 2295nm (arrow marked 1) the Triumph curve form is horizontal and *lys3a* is diagonal, while Lysiba and especially Lysimax are approaching the normal horizontal condition.

It is also seen that the bulb in *lys3a* at the arrow marked 2 (Figure 12C) is much reduced in Lysiba and especially Lysimax indicating a more normal barley state. This area correlates to unsaturated fat (2347nm) and cellulose (2352nm). The content of these analytes should be tested further to detect the reasons why the samples separate from each other. A lower content of unsaturated fat and cellulose (fibre) in Lysiba and Lysimax should be expected.

We will now further qualify the results of PCA clustering and visual inspection in a PLSR correlation study involving the 15 samples using the NIR spectra to predict the chemical composition. High PLSR predictions were obtained by PLSR models for six chemical parameters as **y** with NIR as **X** in Table 13.

Table 13. PLSR prediction of chemical data (**y**) by NIR spectra (**X**) for samples in Table 12.

y	r	PC	RMSECV	RE
Protein	0.90	3	0.3	11.8
Amide	0.98	3	0.01	0.1
A/P	0.99	3	0.4	6.0
Starch	0.79	1	1.8	15.6
ß-glucan	0.998	9	0.1	1.8
Rest	0.85	1	1.6	15.3

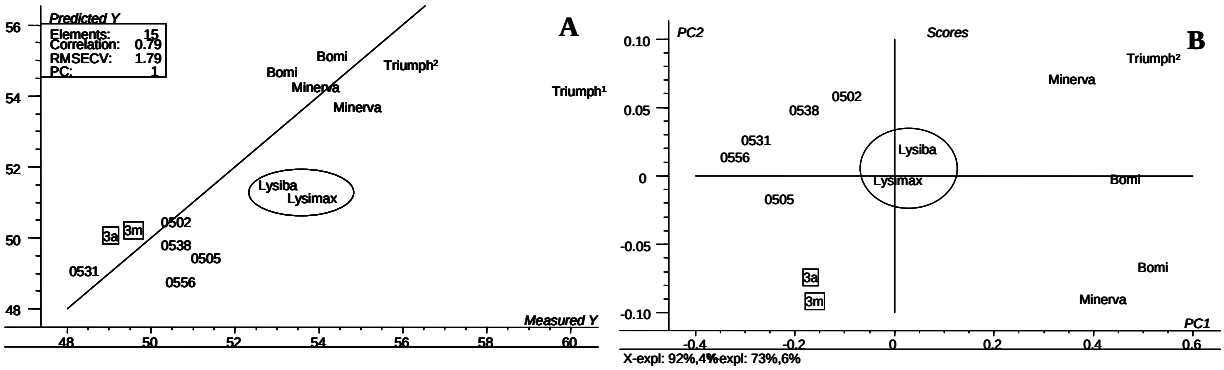


Figure 13A. PLSR prediction plot of starch (**y**) by NIR measurements (**X**) from the sample set in Table 13. **B.** PLSR score plot.

Focusing on the PLSR correlation plots predicting the important parameter for yield, starch from NIR (r=0.79) are seen in Table 13 and Figure 13A, respectively. The lower correlation of NIR data is due to the Triumph outlier marked 1. If the outlier is eliminated, a correlation coefficient of r=0.85 is obtained for prediction of starch with NIR. The score plots of the NIR data (one outlier removed, Figure 13B) reveal a classification where the high-starch

lys3 recombinants Lysiba and Lysimax have moved in the direction of the normal barleys, indicating an improvement in chemical composition. When breeding for a plump high-yielding condition, yet retaining the high-lysine condition, it is necessary to check the amide/protein ratio that is kept low in the *lys3a* recombinants (Table 12). The low number of samples in this experiment needs verification with an independent data set. This was obtained with essentially the same conclusion by another set of the same samples grown under greenhouse conditions, not presented here.

If the environmental offset of the spectra could be compensated for by a chemometric adjustment, a combined PCA of NIR spectra from field and greenhouse could be obtained. Such a development should be helpful in plant breeding.

This investigation indicates again that near infrared spectroscopy is a strong tool which can characterise even a limited number of samples in a breeding material.

It is seen that a large numbers of reproducible data points in spectroscopy increase the possibility of a reliable identification and classification by chemometrics of each single sample. The separate treatment of the evaluation of chemical and NIR data supports the conclusions.

The above shows that by obtaining whole NIR spectra and evaluating them with the unsupervised algorithm PCA the plant breeders have a strong tool for detecting and directing diversity. This chemometric selection procedure is called "data breeding"[60]. By measuring NIR on flour or NIT on whole kernel samples both large and relatively small physical-chemical differences between varieties or recombinant lines and between extreme mutants can be detected. In principle, it is like working in the field, where plant breeders use their eyes to detect agronomic and morphological differences and similarities in patterns between populations as well as single plants. By using spectroscopy evaluated by PCA score and loading plot they can now also use their eyes to observe patterns to detect chemical-physical differences between populations and single plants.

We have thus proved that it is possible by NIR to detect and breed for a modification of the negative parts (low starch, high protein, high fibre), parts of a pleiotropic complex such as the *lys3* gene while maintaining the high-quality (high lysine, low A/P index) condition. This is done by manipulating the whole genetic background of the mutant gene in cross breeding materials where NIR spectra evaluated by chemometrics in a PCA or PLSR score plot can be used as a selection criterion.

6. Conclusion

Why multivariate analysis should be adopted in science and industry as a complement to traditional statistics

It is obvious that PCA and PLSR should be routinely used as a convenient and reliable tool for the plant breeders and the cereal industry to identify the physical-chemical nature of optimal and deviating samples by analysing germination curves and evaluating NIR or NIT spectral or imaging seed data (GrainCheck), as demonstrated in Papers I-VII.

PLSR/multivariate analysis is regularly used in the software for NIT spectrometers as ”black boxes”, for industrial calibrations (e.g. for protein) without the possibility for the user to intervene and understand the nature of multivariate calibration. This is an example of the hidden nature of technology. In this case multivariate analysis is used for prediction of for example protein and water without exploiting its full potential by using the whole (spectral) data sets for classification of samples and for prediction of other analytes where the user may perform own calibrations.

The two dimensional germinative quality classification g%1-g%3 for malting barley presented in this thesis should be able to be used routinely by the industry in combination with the viability detection by the Tetrazolium method for a 24 hours test. We have found that “vigour” (g%1) has a strong physical-chemical background near connected to malt modification as demonstrated by the promising predictions by NIT. A NIT test should further be developed as a convenient screening method *at-line* or *on-line* in cooperation with the malting and brewing industry and with the instrument manufactures.

It would be helpful, if PCA software and instructions for use could be included in the data programmes of the instrument. Special short courses should be targeted to the end users of the instruments. Multivariate analysis is essential in evaluating in depth large multivariate covariate datasets where classical statistics cannot be used.

From above results it can be concluded that if the purpose is to discover new physically-chemically related endosperm genes in a mainly homozygotic background, there is no need for a complete proteome analysis. To characterise a pleiotropic covariate fingerprint of a gene involving a few factors taken from the phenotype (as in Figure 2A-F), the metabolome or the proteome should in most cases be sufficient, if the gene has a major expression effect. However, in such endeavours multivariate, chemometric pattern recognition data analysis is a necessary and central tool in handling the covariate information (Paper VI).

Section 5.3 illustrates that several minor covariate gene expressions of a physical-chemical nature from the genetic background can be manipulated by NIR detection and PCA without detailed knowledge of the genetic background.

This new approach (Papers V-VII) is able to record biodiversity as a characteristic covariate physical, chemical spectral fingerprint of the barley seed phenotype reflecting the genetic variation in chemical composition, including recording significant differences in proteome patterns and their consequences. Specific information on the biological organisational level (macro) from the "global" data set can be selected and modelled by calibration to the chemical organisational level (micro) for validation[58].

The effect of the high lysine gene *lys3a* on the proteome level is quite well established[57] (Paper V-VII), affecting albumines, globulines and prolamines. In addition pleiotropic effects occur which, for example, change starch, ß-glucan and fat content[57]. All this variation represents, as discussed by Munck[57], "a cage of covariance" connected to the specific gene that to some extent could be opened up by modern multivariate analysis of data from spectroscopy as a reproducible physical-chemical fingerprint (Paper VII).

It is thus possible to evaluate the physical-chemical diversity of barley gene banks and mutation materials on the phenotypic spectral seed level using an exploratory approach with a minimum of *a priori* hypotheses. First, spectral data are taken as a fingerprint of the "global" physics and chemistry of the sample and interpreted by PCA to define "normal" and "deviating" barley. Secondly, a *posteriori* measurement of the compositional and structural status of the indicated aberrant is defined by PLSR correlations to chemical analyses and validated to proteomic fingerprints. The chemical analyses could be suggested from the spectra by prior spectroscopic knowledge. Automated data selection through i-PLS of "hot" wavelengths from the spectra could be performed. These could have a high correlation to the tentatively selected chemical analyses. The exploratory strategy is thus hypothesis generating, ending with a specified advanced indication, which should be verified and expanded in further data breeding experiments with new plant materials and analytical methods to define its genetic and chemical implications[60].

The application of this technology at the canopy level to understand and control the realisation of genetic diversity in plant growth and production would be a difficult, but most rewarding challenge[58].

7. References

1. Aastrup, S. (1988). A review of quick, reliable and simple check methods for barley and malt based on the Carlsberg seed fixation system. *Journal of American Society of Brewing Chemists* **46**, 37-43.

2. Aastrup, S., Riis, P. and Hansen, J.R. (1989). High vigour - the basis for high malting barley quality. *Proceedings of the European Brewing Congress*, Zurich, IRL Press: Oxford, 171-178.

3. Aastrup, S., Riis, P. and Munck, L. (1990). Controlled removal of dormancy renders prolonged storage of sprouting resistant malting barley. The Fifth International Symposium on Pre-harvest Sprouting in Cereals, West view press, San Francisco, USA, 329-337.

4. American Society of Brewing Chemists (1976). Methods of Analysis. The Society: St. Paul, Minnesota, USA.

5. Anonymous (1985). International rules for seed testing. *Seed Science and Technology* **13**, 299-355.

6. Anonymous (1987). *Analytica-EBC*, Brauerei- und Getränke-Rundschau, Zurich, Switzerland.

7. Anonymous (1998), *Analytica-EBC*, Brauerei- und Getränke-Rundschau, Zurich, Switzerland.

8. Anonymous (1998). *American Association of Cereal Chemists Approved Methods*, 9th edition, AACC, St. Paul, Minnesota, USA.

9. Anonymous (1998). *ICC-standards*, Verlag Moritz Schäfer, Detmold, Germany.

10. Barkholt, V. and Jensen, A.L. (1989). Amino acid analysis: Determination of cysteine plus half-cysteine in proteins after hydrochloric acid hydrolysis with disulfide compound as additive. *Analytical Biochemistry* **177**, 318-322.

11. BeMiller, J.N. (1964). IodometricDetermination of Amylose - Amperometric Titration. Methods in Carbohydrate Chemistry IV, Academic Press, 165-168.

12. Bro, R. (1996). Multivariabel Kalibrering. DSR-tryk, Frederiksberg, Denmark.

13. Büchmann, N.B., Josefsson, H. and Cowe, I.A. (2001). Performance of European artificial neural network (ANN) calibrations for moisture and protein in cereals using the Danish near-infrared transmission (NIT) network. *Cereal Chemistry* **78**, 572-577.

14. Campbell, M.R., Sykes, J. and Glover, D.V. (2000). Classification of single and double mutant corn endosperm genotypes by near infrared transmittance spectroscopy. *Cereal Chemistry* **77**, 774-778.

15. Chandra, G.S., Proudlove, M.O. and Baxter, E.D. (1999). The structure of barley endosperl – An important determinant of malt modification. *Journal of the Science of Food and Agriculture* **79**, 37-46.

16. Czuchajowska, Z., Szczodrak, J. and Pomeranz, Y. (1992). Characterization and estimation of barley polysaccharides by near-infrared spectroscopy. I. Barleys, starches and β-D-glucans. *Cereal chemistry* **69**, 413-418.

17. Delouche, J.C. and Baskin, C.C. (1973). Accelerated ageing technique for predicting the relative storability of seed lots. *Seed Science and Technology* **1**, 427-452.

18. Delwiche, S.R., Graybosch, R.A. and Peterson, C.J. (1999). Identification of wheat lines possessing the 1AL.1RS or 1BL.1RS wheat-rye translocation by near-infrared reflectance spectroscopy. *Cereal Chemistry* **76**, 255-260.

19. Delwiche, S.R. and Graybosch, R.A. (2002). Identification of waxy wheat by near-infrared reflectance spectroscopy. J*ournal of Cereal Science* **35**, 29-38.
20. Doll, H. (1983) Barley seed proteins and possibilities for their improvement, in: W. Gottschalk, H.P. Muller (Eds.), Seed Proteins: Biochemistry, Genetics, Nutritive Value, Martinus Nijhoff, The Hague, 205-223.
21. Donhauser, E.G., Linsenmann, O. and Faltermeier, E. (1983). Evaluation of near infrared spectrophotometry in the malting and brewery laboratory. *Monatsschrift für Brauwelt* **36**,474-478.
22. Downey, G. (1985). Estimation of moisture in undried wheat and barley by near infrared reflectance. *Journal of Science in Food and Agriculture* **36**, 951-953.
23. Ellis, D.H. and Roberts, E.H. (1980a). Improved equation for the prediction of seed longevity. *Annals of Botany* **45**, 13-30.
24. Ellis, D.H. and Roberts, E.H. (1980b). The influence of temperature and moisture on seed viability period in barley (*Hordeum viaticum L.*). *Annals of Botany* **45**, 31-37.
25. Ellis, D.H. and Roberts, E.H. (1981). The quantification of ageing and survival in orthodox seeds. *Seed Science and Technology,* **9**, 373-409.
26. Englyst, H. N. and Hudson, G. J. (2000). Carbohydrates. I: *Human Nutrition and Dietetics*. 10th edition. Ed.: Garrow, J. S.; James, W. P. T. & Ralph, A. Churchill Livingstone, Edinburgh.
27. Esbensen, K. (2002). Multivariate Data Analysis in practise "Unscrambler", CAMO, Oslo, Norway.
28. Fiala, F. (1989). Report of seed vigour test committee 1986-1989. *Seed Science and Technology* **15**, 507-522.
29. Gebhardt, D.J., Rasmusson, D.C. and Fulcher, R.G. (1993). Kernel morphology and malting quality variation in lateral and central kernels of six-row barley. *American Society of Brewing Chemists* **51**, 145-148.
30. Geladi, P., MacDougall, D. and Martens, H. (1985) Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* **39**, 491-500.
31. Gerstenkorn, P. (1985). NIR-spectroscopy as a practical method for measurement of crude fiber in barley. *Getreide Mehl Brot* **39**, 167-170.
32. Goodspeed, T.H. (1911). The temperature coefficient of the duration of life of barley grains. *Botanical Gazette* **51**, 220-224.
33. Hampton, J.G. (1995). In: H.A. van de Venter (ed.) Seed vigour testing, *Proceedings from 24th Congress of the International Seed Testing Association*, Copenhagen, 10-28.
34. Henry, R.J. (1985). Near-Infrared Reflectance analysis of carbohydrates and its application to the determination of (1-3),(1-4)-ß-D-glucan in barley. *Carbohydrate Research* **141**, 13-19.
35. Hibbard, R.P. and Miller, E.V. (1928). Biochemical studies on seed viability. I. Measurements of conductance and reduction. *Plant physiology* **3**, 335-352.
36. Holmes, M.G. (1990). Relationships between pearling resistances of barleys and the extract potential of the subsequent malts. *Journal Institute of Brewing* **96**, 123-124.
37. Jacobsen, S., Nesic, L., Petersen, M. and Søndergaard, I. (2001). Electrophoresis **22**, 1242-1245.
38. Jacobsen, T. (1980). Et nyt begrep innenfor kjemisk analyse: Kjemometri, variasjoner I sporelemtinnholdet i malt belyst ved hjelp av faktoranalyse. *Brygmesteren* **6**, 149-158.

39. Jacobsen, T. (1986). Malting technology and the properties of malt: A cluster analysis study. *Journal Institute of Brewing* **92**, 177-180.

40. Jurs, P.C. (1986). Pattern recognition used to investigate multivariate data in analytical chemistry, *Science* **232**, 1219-1224.

41. Kim, H.O. and Williams, P.C. (1990). Determination of starch and energy in feed grains by near infrared reflectance spectroscopy. *Journal of Agriculture and Food Chemistry* **38**, 682-685.

42. van Laarhoven, H.P.M., Douma, A.C. and Angelino, S.A.G.F. (1997). Image analysis as practical tool in barley-malt-beer chains. EBC Congress, Maastricht, The Netherlands, 183-189.

43. Larsen, J. (1994). Targets for tailored breeding. In: Eur.Brew.Conv.Monogr., Symposium on malting technology, Andernach, 152-164.

44. Law, D.P. (1985). Selection for malting quality in barley breeding using near infrared reflectance. In Proceedings of the First International Symposium on Near Infrared Reflectance Spectroscopy, Royal Australian Chemical Institute, Cereal Chemistry Division, Melbourne, Australia.

45. Lindenberg, W. and Bulat, H. (1955). „Triebkraft", Ziegelgruswert und Tetrazoliumwert. *Saatgutwirtschaft* **7**, 315-319.

46. MacLeod, A.M. (1950) Determination of germination capacity of barley by means of tetrazolium salts. *Journal Institute of Brewing* **60**, 125-134.

47. Martens, H., Jensen, S.A. and Geladi, P. (1983). Nordic Symposium on Applied Statistics, Stokkand Forlag Publ., 208-234.

48. Martens, H. and Martens, M. (2000). Modified Jack-knife estimation of parameters uncertainty in bilinear modelling by partial least squares regression (PLSR), *Food quality and preference* **11**, 5-6.

49. Martens, H. and Næs, T. (1989). Multivariate Calibration, Wiley, New York.

50. McCleary, B.V. and Prosky, L., (2001). Advanced dietary fibre technology, MGB Books Ltd, Bodmin, Cornwall.

51. McIntosch, G.H., Whyte, J., McArthur, R. and Nestel, P.L., (1991). Barley and wheat foods: Influence on plasma cholesterol concentration in hypercholesterolemic men. *American Journal of Clinical Nutrition* **53**, 1205-1209.

52. Molina-Cano, J.L., Ramo, T., Ellis, R.P., Swanston, J.S., Bain, H., Uribe-Echeverria, T. and Perez-Vendrell, A.M. (1995). Effect of grain composition on water uptake by malting barley: A genetic and environmental study. *Journal Institute of Brewing* **101**, 79-83.

53. Moore, R.P. and Goodsell, S.F. (1965). Tetrazolium test for predicting cold test performance of seed corn. *Journal of Agronomy* **57**, 489-491.

54. Morgan, A.G. and Gothard, P.G. (1979). Rapid prediction of malt hot water extract by near infrared reflectance spectroscopy studies on barley. *Journal Institute of Brewing* **85**, 339-343.

55. Munck, L., (1972). Improvement of nutritional value in cereals, *Hereditas* **72**, 1-128.

56. Munck, L. (1991). Quality criteria in the production chain from malting barley to beer – some views on the logistic problems with a few examples (dormancy, vitatility and β-glucans). *Ferment* **4**, 235-241.

57. Munck, L. (1992). The case of high-lysine barley breeding, in: P.R. Shewry (Ed.), Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, CAB International, Wallingford. U.K, 573-602.

58. Munck, L. (2003). A new holistic exploratory approach bridging the genotype and phenotype, in: R. Von Bothmer, T. van Hintum, H. Knüpffer, K. Sato (Ed.), Diversity in barley (*Hordeum vulgare*), Elsevier Science B.V., Amsterdam, The Netherlands, Chapter 11, 227-245.

59. Munck, L., Nørgaard, L., Engelsen, S.B., Bro, R. and Andersson, C. (1998). Chemometrics in food science – a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics and intelligent laboratory systems* **44**, 31-60.

60. Munck, L., Møller, B. and Pram Nielsen, J. (2000). From Plant Breeding to DataBreeding - How New Multivariate Screening Methods Could Bridge the Information Gap between the Genotype and the Phenotype. *Proceedings from VIIth IBGS Congres*, Adelaide, Australia.

61. Nielsen, J. P. (2003). Evaluation of malting barley quality using explorative data analysis. II The use of kernel hardness and image analysis as screening methods. *Journal of Cereal Science,* **38**, 173-180.

62. Nijenstein, J.H. (1995). Soil cold test – European perspective, In: H.A. van de Venter (ed.) Seed vigour testing, *Proceedings from 24th Congress of the International Seed Testing Association*, Copenhagen, 34-52.

63. Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L. and Engelsen, S.B. (2000). Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy* **54**, 413-419.

64. Osborne, B.G., Fearn, T., Hindle, P.H. (1993). Practical NIR spectroscopy with applications in Food and Beverage analysis. Longman, Harlow, UK.

65. Palmer, G.H. (2000). Malt performance is more related to inhomogeneity of protein and β-glucan breakdown than to standard malt analyses. *Journal Institute of Brewing* **106**, 189-192.

66. Pearson, K. (1901) On lines and planes of closest fit to systems of points in shape. *Philosophical Magazine* **2**, 559-572.

67. Perry, D.A. (1981). Handbook of vigour test methods, International Seed Testing Association, Zürich, Switzerland.

68. Pitz, W.J. (1990). An analysis of Malting Research. *Journal of American Society of Brewing Chemists* **48**, 33-44.

69. von Reiner, L. (1981). Die Faktorenanalyse in der brautechnologischen Forschung. *Brauwissenschaft* **24**, 403-412.

70. Riis, P. and Bang-Olsen, K. (1991). Breeding high vigour barley – a short cut to fast maltings, *Proceedings of the European Brewery Convention Congress*, Lisbon, IRL Press: Oxford, 101 –108.

71. Riis, P., Meiling, E. and Peetz, J. (1995). Determination of germinative percentage and germination index – collaborative trial and ruggedness testing, *Journal of the Institute of Brewing* **101**, 171 –175.

72. Roberts, E.H. (1960). The viability of cereal seed in relation to temperature and moisture. *Annals of Botany* **24**, 12-31.

73. Roberts, E.H. (1961). Viability of cereal seed for brief and extended periods. *Annals of Botany* **25**, 373-380.

74. Roberts, E.H. (1973). Predicting the storage life of seeds. Seed science and technology 1, 499-514.

75. Roberts, E.H. and Abdallah, F.H. (1968). The influence of temperature, moisture and oxygen on period of seed viability in barley, broad bean and peas. *Annals of Botany* **32**, 97-117.

76. Robertson, D.W. (1939). Effect of relative humidity on viability, moisture content and respiration of wheats, oats and barley in storage. *Journal of Agricultural Research* **59**, 281-291.

77. Romagosa, I. and Fox, P.N. (1993). Genotype x environment interaction and adaptation in Plant Breeding (eds.) Hayward, M.D., Bosemark, N.O. and Romagosa, I., Chapman and Hall, London, UK. 372-421.

78. Salmeron, J., Manson, J.F., Stampfer, M.J., Colditz, G.A., Wing, A.L. and Willet, W.C. (1997). Dietary fiber, glycemic load and risk of non-insulin dependent diabetes mellitus in woman, *Journal of the American Medical Association* **277**, 472-477.

79. Spears, J.F. (1995). An introduction to seed vigour testing, *In*: H.A. van de Venter (ed.) Seed vigour testing, *Proceedings from 24th Congress of the International Seed Testing Association*, Copenhagen, 1-9.

80. Szczodrak, J., Czuchajowska, Z. and Pomeranz, Y. (1992). Characterization and estimation of barley polysaccharides by near-infrared spectroscopy. II. Estimation of total β-D.glucans. *Cereal Chemistry* **69**, 419-423.

81. Tekrony , D.M. (1995). Accelerated ageing, In: H.A. van de Venter (ed.) Seed vigour testing, *Proceedings from 24th Congress of the International Seed Testing Association*, Copenhagen, 53-72.

82. Waddington, C.H. (1969). in A. Koestler, J.R. Smythies (Eds.), Beyond Reductionism. The Alpbach Symposium, Radius Book/Hutchinsson, London, 357.

83. Wang, D., Dowell, F.E., Lacey, R.E., (1999). Predicting the number of dominant R-alleles in single wheat kernels using visible and near-infrared reflectance spectra. *Cereal Chemistry* **76**, 6-8.

84. Williams, P.C., Preston, K.R., Norris, K.H. and Starkey, P.M. (1984). Determination of amino acids in wheat and barley by near-infrared reflectance spectroscopy. *Journal of Food Science* **49**,17-22.

85. Williams, P. and Norris, K. (1987). Near Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, Minnesota, USA.

86. Williams, P.C. and Sobering, D.C. (1993). Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *Journal of Near Infrared Spectroscopy* **1**, 25-32.

87. Wilson, P.R. (1993). Analysis of whole cereal grains by near infrared transmission. *Aspects of Applied Biology, Cereal Quality III* **36**, 477-480.

Seed Vigour In Relation To Heat Sensitivity And Heat Resistance In Barley Evaluated By Multivariate Data Analysis

Birthe Møller[1,2] and Lars Munck[1]

ABSTRACT

J. Inst. Brew. 108(3), 286–293, 2002

The vigour loss model based on normal distribution after artificially ageing by heat treatment of barley seeds developed by Ellis and Roberts and further developed at Carlsberg by Aastrup et al. introducing vigour potential (VP), does not adequately describe seed vigour for all barley samples. In a preliminary investigation we have identified heat-resistant barley samples. In this investigation we found untreated barley samples from the field where heat treatment as high as 68°C for 4 h at 12% water content only decreases germination from 99.0% to 93.8% compared with 94.8% to 0.0% for some of the heat-sensitive barleys following the above mentioned model. The correlation between germination velocity measured by the germination index (GI) of untreated samples and VP is not consistent when comparing different barley material. It is concluded that the classic vigour loss model for heat treatment may be used as a worst case prediction for germination, but it does not address the variation found in practice, including the possible advantage of exploiting the naturally occurring heat resistance.

Key words: Barley, heat resistance, heat treatment, principal component analysis, vigour.

INTRODUCTION

The main requirement for malting barley is a complete, even and preferably rapid germination; that is, the barley should possess a high vigour. Riis and Bang-Olsen[19] found that slow-germinating low-vigour barley which had an end germination percentage comparable to fast-germinating high-vigour barley demanded up to 42% longer malting time.

Moreover, vigour is important in the grain industry that produces seed for farmers for sowing, because low-vigour seeds will give a reduced grain yield, especially under stressed weather conditions[8]. A reduced vigour can affect the plant emergence in the field, so that the crop is not successfully established. In extreme cases, it is not possible to compensate for this effect by using more seed. Furthermore, plants from low-vigour seeds do not achieve as high yield per plant as plants from high-vigour seeds[8].

All these conditions may influence the production costs and the quality of the end product – barley for malting.

The circumstances that influence the ageing process have been investigated by several authors in many different experiments where methods to determine loss of vigour have been developed. These methods can roughly be divided into three categories: 1) tests monitoring biochemical changes, 2) germination tests where germination time and homogeneity are investigated, and 3) stress tests where seeds are exposed to a stressful environment, either prior to imbibition or during germination. Vigour tests included in this category are accelerated ageing by heat treatment, cool germination, controlled deterioration and the Hiltner test using Ziegel gravel[23].

With regard to heat treatment as a stress source, many theories have been proposed to exploit heat treatment experimentally in order to accelerate the ageing process.

Survival of barley with constant moisture content was studied after varying periods of heat treatment at 50-100°C in closed containers. Life duration was defined as a chemical reaction according to the equation of Arrhenius[9]. The temperature coefficient Q10 was determined as the factor by which the death rate of seed increases when temperature increases 10°C.

Groves[14] studied the connection between life duration after heat treatment at different moisture contents and the temperature coefficient Q10 for ageing, but did not find a clear relationship.

Robertson et al.[22] demonstrated that the decrease in germination percentage due to heating was dependent on water content (relative humidity). Kernel age and storage condition before testing also influenced artificial ageing. They suggested that if one knew how much humidity the grain could tolerate at a certain time and temperature, it would be possible to use this information as a guide in the choice of storage conditions. Robertson et al. were the first to propose the possibility of using temperature, humidity and germination percentage in a model to determine vigour[22].

Robertson's theory expressed in words influenced the creation of the mathematical models of Roberts and Ellis.

Roberts[21] assumed that ageing in a seed sample adhered to a normal distribution model, and saw the ageing process as a sigmoid curve. He described the relationship between the half-vital-period (the time it takes to kill 50% of the

[1] The Royal Veterinary and Agricultural University, Dairy and Food Science, Food Technology, Rolighedsvej 30, Dk-1958 Frederiksberg C

[2] Corresponding author: E-mail: bm@kvl.dk

Publication no. G-2002-0826-044

kernels) and the temperature and water content, and found a linear correlation[21].

A better correlation was found between predicted and measured values of loss of vigour, when water content (%) was transformed to logarithmic values[9]. The earlier equations were written together resulting in Ellis and Roberts' vigour equation[11].

$$V = K_i - p/10^{(K_e - C_w * \log m - C_h * t - C_q)}$$

V = probit % germination, K_i = probit % germination at beginning of storage, p = storage time in days, K_e = 9.983, C_w = 5.896, C_h = 0,04, C_q = 0.000428, m = moisture content, t = temperature

The values of the constants were determined from "survival curves" for a sample of the barley variety Proctor stored in 52 different environments from –20 to 90°C with moisture contents between 5 and 25%[10].

Based on the theoretical and empirical vigour model of Ellis and Roberts[11], Aastrup et al.[1] at Carlsberg defined the vigour potential (VP) in probits (probability scale) , also precluding that germination follows a sigmoid curve based on the assumption of normal distribution of ageing with storage time[1]. The life story of an ideal barley sample can thus be described as a decrease in dormancy followed by complete germination ending with the ageing process, which will result in loss of vigour. When the percentage of germination is transformed in a probability scale to probit (e.g. VP 3 = 99.9%, VP 0 = 50%, VP –1 = 15%), a straight line will occur, and VP can be determined as the intersection with the y-axis[1].

A barley lot will often contain a few grains that are dead due to mechanical damage and not because of loss of vigour. For this reason Aastrup et al.[1] introduced Pn_v as the part of the normally distributed barley population, eliminating the effect of grain damage in the equation.

Two barley varieties, one original sample of each, were used by Aastrup et al.[1] to test the model, where 5 subsamples of each sample/variety were artificially aged for varying periods of time at 60°C at a water content of 12%.

It was assumed that accelerated ageing through heat treatment was indicative of decrease of vigour during long-term storage, which occurs in practice. There have been few publications regarding vigour potential since 1989. Apparently, the VP model is little used in practice.

Instead, the germination index (GI) determined on unheated seeds defined by Riis and Bang-Olsen[19] was adopted as a measure of vigour, and seems to be more widely used.

In the following we will investigate vigour as a function of heat treatment on a wide range of barley varieties from different harvest conditions. The material will be studied with and without accelerated ageing by heat treatment expressed as germination percentage, vigour potential and germination index using multivariate data analysis for evaluation of the whole germination curves.

MATERIALS AND METHODS

Plant material

All samples were stored at 4°C before analysing.

A. Barley samples from 1993 with natural differences in vigour: 21 samples of Alexis, Ariel, Blenheim, Etna and Meltan varieties were used. The samples were grown in different farm locations in Southern Sweden collected and received from Skånska Lantmännen, Malmö after harvest 1993.

B. Barley samples from harvest 1993 with artificial differences in vigour due to heat storage: Alexis, Carula, Etna and Lysimax varieties were used. The 4 samples, which displayed initial germination percentage of 97.3 to 98.5, were harvested in 1993 and artificially aged by high temperature storage at 58°C (See Heat Treatment) yielding 4 samples of Alexis (0, 10, 30, 40 h), 3 samples of Carula (0, 10, 30 h), 5 samples of Etna (0, 10, 30, 40, 50 h) and 3 samples of Lysimax (0, 10, 30 h).

C. Barley from harvest 1994 with natural differences in vigour: Samples from A and B were grown in 1994 in Southern Sweden and in Denmark on the islands of Zealand and Funen. 14 samples of Alexis, Blenheim, Lysimax and Meltan varieties were chosen to determine VP.

Weather conditions 1993 and 1994

Seed vigour is heavily dependent on weather conditions and research on vigour under field conditions is severely hampered by the fact that only one to two years out of ten show drastic effects on germination.

In 1993 the spring and early summer in Southern Scandinavia were characterised by rain deficit. From sowing in March to July it only rained 50% of the average amount. July was colder than normal and had 16% fewer hours of sun than normal. However, July had 50% more rain than usual, most of it in the last two weeks of the month. In August the precipitation was normal, although most of the rain fell in the two first weeks of the month[6]. Thus, the humidity conditions made the harvest in 1993 quite difficult, and most of the grain was harvested with more than 15% moisture content, causing in some cases severe losses in germination capacity.

The spring (March) of 1994 was wet with 100% more rain than the normal amount. Therefore, the sowing of the experiments was delayed until late April. In June there was 20% more rain than usual. July was dry and hot, and August had normal precipitation, but most of it fell in late August after the barley was harvested[7]. This resulted in dry and good quality seed.

It is concluded that 1993 and 1994 were contrasting years with regard to the effect of weather conditions on harvest conditions and germination of the produced seeds.

Germination analysis

Two different germination analyses were made: (1) to determine total germination percentage and germination index (GI) after harvest using a H_2O_2 solution to remove dormancy, and (2) to determine the percentage of germinated kernels after heat treatment, after which the vigour potential (VP) was calculated. Petri dishes and water were used for this analysis (see below). In both analyses the samples were placed in a dark Refritherm incubator at 20°C. Four replicates were made for every sample and the standard deviation between the replicates was less than 5%.

1. Germination Index (GI)[19]: 4 × 100 kernels were steeped and germinated in a beaker with 50 mL 0.75% H_2O_2 solution added for three days. Percentage of germinated kernels (n) was calculated and removed after 24, 48 and 72 h. Every day the H_2O_2 solution was changed and

new was added to each sample. This test was carried out on samples immediately after harvest. Germination Index (GI) was calculated according to the following equation[19]:

$$GI = \frac{10 \times (n24 + n48 + n72)}{n24 + (n48 \times 2) + (n72 \times 3)}$$

n = % germinated kernels after 24, 48 and 72 h of germination

2. Vigour Potential (VP)[1]: 4 × 100 kernels were germinated in 90 mm petri dishes with two layers of filter paper (Whatman No. 1) and 5 mL H_2O for 8 days at 20°C. Germinated kernels were counted and removed every third day. This was carried out at least two months after harvest; so most dormancy was broken.

Heat treatment

Heat treatments were carried out with two different purposes:

1. To produce artificially aged samples by heat storage for comparison with naturally aged samples: The 4 samples, Alexis, Carula, Etna and Lysimax, all contained 12% water. Ageing (loss of vigour) was achieved by heat treatment of barley lots in watertight plastic bags in a water bath at 58°C for 0-50 h. Subsamples were removed every 10 h. These samples were analysed in order to determine VP and included in the field experiment in 1994.

2. To determine vigour potential (VP) according to Aastrup et al.[1]: Vigour potential was determined by heat-treating 10 subsamples of every barley lot at 12% water content contained in welded plastic aluminium bags in a water bath at 68°C for 0-4½ h, where subsamples were removed every 30 min. After this treatment every subsample was germinated according to germination method 2. The results from 8 days of germination were plotted, yielding a germination curve dependent on heat treatment from which the vigour potential was calculated using the Carlsberg vigour model[1].

If the moisture content in a sample was too high, the sample was spread out in a thin layer at room temperature until the moisture content was below 12%. The samples were corrected to 12.0% water content by the addition of the missing volume of water. The sample was then shaken and left for 24 h in a closed plastic box, after which the moisture content was checked.

Chemical analysis

Moisture content was determined according to ICC 110/1[3].

Activity of α-amylase was determined according to ICC 108[3].

Data analysis

Principal Component Analysis (PCA) was performed according to Martens and Næs[15] using the software "Unscrambler" from CAMO ASA, Norway. The aim of the PCA algorithm is to determine the latent factors or principal components (PCs) in the data set, which describe most variation. Based on vector algebra the algorithm calculates and compresses the data material (whole germination curves after heat treatment) into scores for principal components, which are plotted on a score plot. The position of a sample in the score plot expresses the pattern of the corresponding germination curve, so those samples with similar scores reflect the same pattern. The variables involved (germination percentage (g) after heat treatment in e.g. 4½ h (g4½) and untreated (g0), expressed as loadings, can be plotted together with the scores in a biplot. A score for a germination curve placed near a loading indicates a high influence by the variable.

RESULTS

In Table IA the differences and ranges in vigour potential (VP), germination percentage and germination index (GI) of all the barley varieties naturally and artificially aged are shown. These results show that the germination index (GI) varies from 4.9 to 8.9 for the naturally aged samples and from 3.4 to 8.7 for the artificially aged samples. Three samples of naturally aged Alexis had a final germination percentage after 3 days (method 1) less than 90%, whereas 5 of the artificially aged samples by heat storage (Carula 30 h; Etna 30, 40, 50 h; Lysimax 30 h) had less than 90% germination. In the naturally aged samples the level of VP varies from 0.6 to 4.4, whereas in the arti-

TABLE IA. Min-Max values for vigour potential (VP), germination index (GI) and total germination percentage for the 50 samples. Min-Max values of α-amylase activity for 28 samples of the total.

Variety	Samples	VP	GI	Germ_tot	α-Amylase
Naturally aged 1993–1994					
Alexis	12	0.6–4.4	4.9–8.9	49.8–98.8	0.2–2.0 (n = 12)
Ariel	2	0.8–1.7	8.3–8.4	95.8–97.0	—
Blenheim	7	1.5–2.9	5.7–8.8	93.5–97.0	0.1–1.2 (n = 5)
Etna	2	0.9–2.1	5.0	95.8–97.5	—
Lysimax	2	2.6–2.9	6.8–6.9	98.8–99.5	0.1 (n = 2)
Meltan	10	1.2–3.1	5.2–7.3	97.0–99.3	0.1–0.7 (n = 8)
Artificially aged					
Alexis – untreated	1	1.0	8.2	98.5	0.4 (n = 1)
heat-treated 10-40 h	3	0.3–1.4	4.4–6.8	95.3–98.3	—
Carula – untreated	1	1.2	8.1	97.3	—
heat-treated 10-30 h	2	0.5–1.0	3.9–6.4	25.0–95.3	—
Etna – untreated	1	1.5	5.0	97.3	—
heat-treated 10-50 h	4	0.1–1.5	3.4–5.0	15.3–94.8	—
Lysimax – untreated	1	2.3	8.7	97.8	—
heat-treated 10-30 h	2	0.4–2.2	3.8–6.0	32.5–91.8	—

TABLE IB. Average value and variation for vigour potential (VP), germination index (GI) and % total germination in artificially aged samples, naturally aged samples in total and for the two separate years.

	n	VP	GI	Germ.%
Artificially aged	11	0.9 ± 0.8	4.8 ± 1.2	73.3 ± 16.2
Naturally aged	35	2.0 ± 0.8	6.7 ± 1.2	94.2 ± 4.8
Naturally aged 1993	21	1.8 ± 0.8	6.6 ± 1.3	93.0 ± 5.9
Naturally aged 1994	14	2.4 ± 0.4	6.7 ± 1.2	96.4 ± 1.3

ficially aged samples the level of VP varies from –0.5 to 2.2, which is markedly lower. Alexis is an exception here, because the given heat treatments for 10 and 30 h apparently have higher VP (VP = 1.4 and 1.3) than the untreated sample (VP = 1.0). The reason for this will be discussed later.

A comparison of the average values of the artificially and naturally aged samples in Table IB shows that the naturally aged samples have higher values of VP, GI and total germination compared to the artificially aged samples produced by heat storage. It is furthermore seen that the naturally aged samples grown and harvested in 1994 have higher values of VP, GI and total germination compared to the naturally aged samples harvested in 1993.

The extreme germination curves (Fig. 1A) obtained by heat treatment at 68°C with 12% water content from ½-4½ h, were selected from the PCA classification plots (Fig. 2A). One group of samples followed the vigour model of Ellis and Roberts[11] with a strong decrease in germination after 1-2 h of heat treatment, resulting in reverse S-shaped curves as represented by No. 11 Alexis (Fig. 1A). No. 29 Meltan (Fig. 1A), displaying horizontal linear curves where germination exceeded 90% even after 4 h of heat treatment represents another group of apparently heat-resistant samples. The third group also showing horizontal curves, but on a low germination level already in the untreated sample, is represented by the heat-stored sample No. 03 Carula.

In untreated barley samples from the field we found that germination after heat-treatment at 68°C for 4½ h in the extreme samples decreased from 99.0% to 93.8% in the heat-resistant samples compared with 94.8% to 0.0% for some of the heat-sensitive barleys following the above mentioned model.

The Carlsberg model[1] was used to calculate the vigour potential (VP) in a probit scale. As an example, the germination heat treatment curves from Fig. 1A are shown in Fig. 1B where the probit scale is used instead of percentage of germination. The germination curve plotted on probability paper follows approximately a straight line – the tangent of the curve, which is extrapolated to the y-axis where the intersection gives VP.

One would think that sample No. 29 should have the highest VP, because this sample was more resistant to the stress factor of heat compared with the other two samples. This is not the case with the Carlsberg model. As one can see in Fig. 1B, the heat-resistant sample No. 29 has a VP of 3.1, whereas the heat-sensitive sample No. 11 (following the VP model) has a VP of 4.4. The probit curves for the two samples have the same intersection at the Y-axis, but the intersection of the tangent for the two curves which defines VP makes a difference between the two samples according to the slope. The third sample (No. 03) has the lowest VP, as expected.

Principal component analysis (PCA) is used to represent the different patterns of the germination curves for the total number of barley samples (n = 50) (Fig. 2A). The PC 1 and PC 2 explain 76% and 14% of the variation respectively. The notations refer to variety, and behind every score point in the PCA is a corresponding germination curve, as shown in Fig. 1A. Samples lying close to each other in the plot have similar scores and patterns with re-

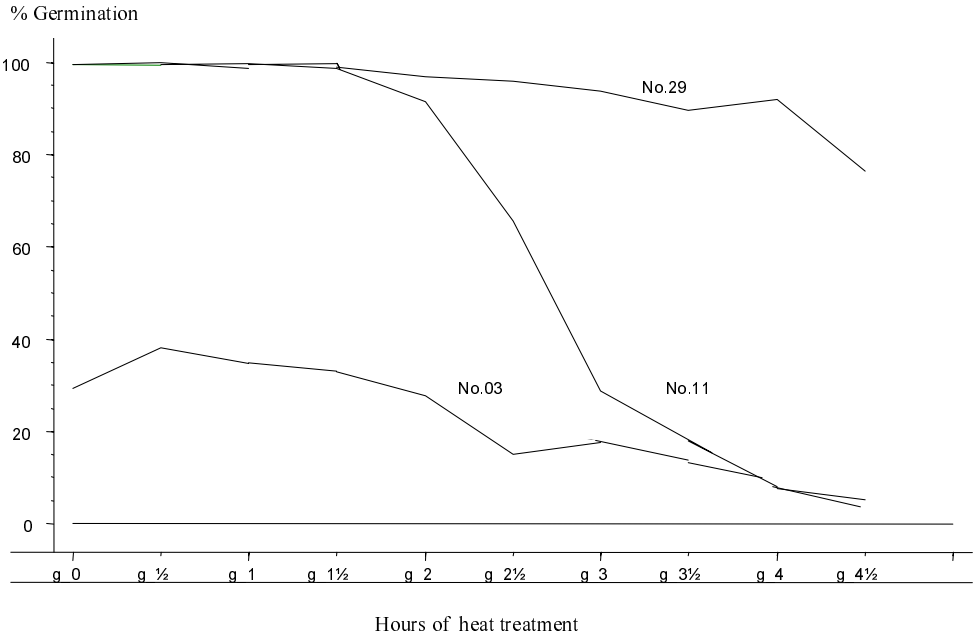


FIG. 1A. Germination curves for samples No. 03 (Carula, artificially aged 30 h), No. 11 (Alexis, naturally aged) and No. 29 (Meltan, naturally aged). Total germination percentage after heat-treatment at 68°C in 0-4½ h, where sub-samples are removed every ½ h.

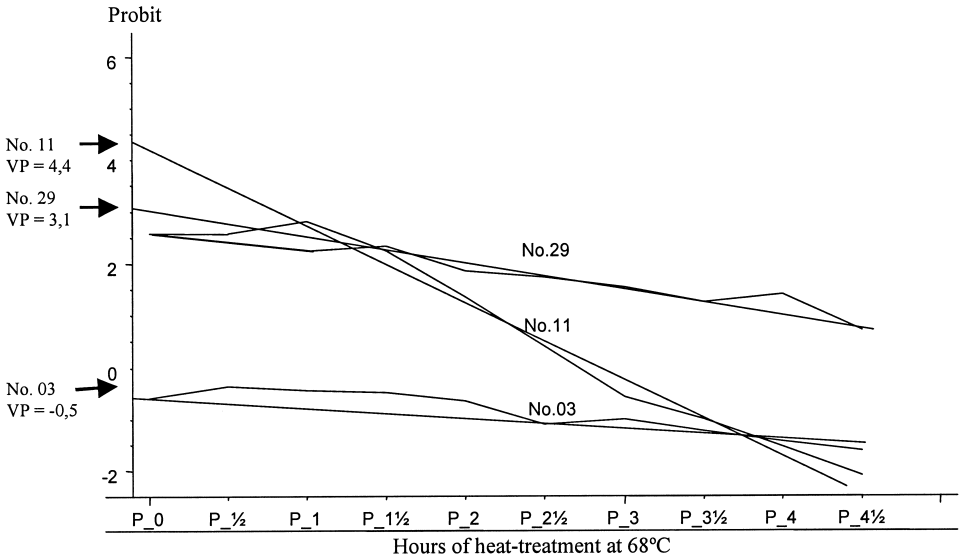


FIG. 1B. Germination curves for samples No. 03, 11 and 29. Total germination percentage in probit after heat treatment at 68°C in 0-4½ h where sub-samples are removed every ½ h. The points where the tangents meet the y-axis (arrows) is defined as the initial vigour potential (VP).

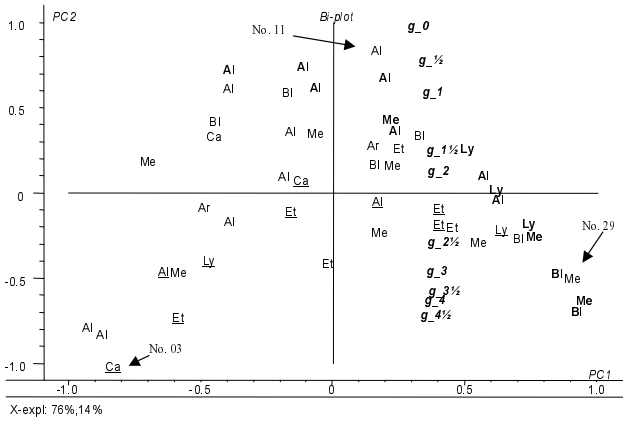


FIG. 2A. Biplot (PC1:PC2) made on germination curves for the heat treated samples harvested in 1993 and 1994. Al = Alexis, Ar = Ariel, Bl = Blenheim, Ca = Carula, Et = Etna, Ly = Lysimax, Me = Meltan. Underlined samples are artificially aged with heat storage before the heat treatment procedure according to the Carlsberg model. Bold samples are harvested in 1994. All others are from 1993. Loadings: e.g. g0 = total % germination in untreated sample and g½ = heat treated ½ h at 68°C.

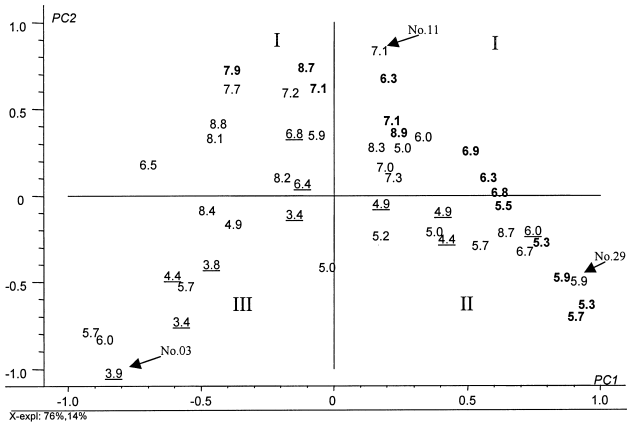


FIG. 2B. Plot with scores (PC1:PC2) made on germination curves for the 50 heat treated samples harvested in 1993 and 1994. The numbers refer to the germination index (GI) for every sample. Underlined samples are artificially aged before heat treatment according to the Carlsberg model. Bold samples are harvested in 1994.

gard to the germination curve. As can be seen in the biplot, there is no systematic division between varieties or between naturally and artificially aged (underlined) samples (Fig. 2A). The corresponding loadings for the variables regarding germination after heat treatment (g0 – g4½) in the biplot in Fig. 2A reveal a systematic trend in the position of the variables ranging from germination percentage at 8 days for the unheated sample g0 in the top right corner to the corresponding values for heat-treatment at 4½ h below. The long heat-treatment variables appear in the same corner as the heat-resistant samples like No. 29 (Fig. 2A).

In Fig. 2B the same plot (loadings not displayed) is shown as in Fig. 2A, although the notation refers to germination index (GI) for the samples instead of variety. It appears as if the samples in the upper half (Group I) have higher GI values than the samples in the lower half. Here there are two groups: one to the bottom left containing samples with low GI (Group III) and one to the bottom right containing samples with intermediate GI values (Group II). The three examples of germination curves No. 11, No. 29 and No. 3 illustrated in Fig. 1A are representatives for each of the three groups I, II and III, respectively. The samples with reverse S-formed curves following the vigour model are placed in the upper half of the score plot (Group I) with high values of GI. The samples with low initial germination percentages (Group III) are placed bottom left (lowest GI values), while the heat-resistant samples with high germination percentage in the untreated sample (as well as in the initially heat-stored samples) are

TABLE II. Average and 5% confidence interval for GI, germination percentage (8 days) of samples, untreated and heat treated at 68°C (12.0% H_2O) in 4½ h calculated for Group I, II and III. Different letters refer to significant differences (5% level) between groups. Activity of α-amylase for 28 samples divided in the three groups are shown as well.

	n	VP		G% untreat.		G% 4½h		α-Amylase	
Group I	24	7.2 ± 0.4	A	95.6 ± 2.0	A	17.9 ± 7.9	A	0.4 ± 0.3 (n = 18)	A
Group II	15	5.7 ± 0.6	B	94.1 ± 3.5	A	53.6 ± 12.4	B	0.2 ± 0.1 (n = 6)	A
Group III	11	5.0 ± 1.0	B	56.8 ± 11.2	B	12.3 ± 16.3	A	0.7 ± 0.9 (n = 4)	A

placed bottom right (Group II), where the GI values are intermediate between the other two groups (Fig. 2B).

When average GI for the three groups is calculated, there is a clear statistical difference between the groups (Table II). The confidence interval (5% level) shows that there is significant difference between GI for Group I and the other groups, but there is no significant difference between GI for Groups II and III. Comparing the average germination percentage with the untreated samples for the three groups, it is seen that there is no significant difference between Group I (heat sensitive) and Group II (heat resistant). Group III however is significantly different from the two groups due to the low initial germination percentage of the samples belonging to this group. This picture changes when looking at the average germination percentage for the groups after 4½ h of heat treatment. Now there is a clear significant difference between Group I and Group II. The samples in Group III are defined as heat resistant because the decrease in germination percentage from untreated to 4½ h of heat treatment are low. However when the germination percentage is initially low, the germination percentage after heat treatment will also be low. Therefore Group III is not significantly different with respect to GI from Group I (Table II). It is concluded that group III is irrelevant for the malting industry due to its low initial germination percentage.

DISCUSSION

The samples grown and harvested in 1994 have higher average values of VP, GI and total germination percentage than the samples from 1993 (Table IB). This is as expected, because the weather conditions during harvest were almost optimal in 1994, but difficult in 1993. With regard to the heat resistance and heat sensitivity, it is not possible to see a systematic difference between the two years. In Figures 2A and B it is seen that samples from 1994 (in bold) are placed from the left top to the right bottom corner, showing both heat-sensitive and heat-resistant samples together with the samples from 1993.

From the example in Fig. 1B it is obvious that the model of vigour potential is not adequate to describe all barley samples. In some cases when samples show heat resistance, the VP will be lower for the sample resisting stress than for the heat-sensitive sample, which reacts to stress. Another example is the artificially aged samples of Alexis, which have higher vigour potentials compared to the untreated sample of Alexis (Table IA). This is due to the fact that the tangents for the two heat-stored samples have a larger slope than the unheated sample, resulting in a higher VP value compared to the untreated raw sample.

When comparing VP with germination index (GI), which is another potential indicator of vigour, the heat-sensitive samples in this experiment have higher GI than the heat-resistant samples. Aastrup et al.[2] found a high correlation (r = 0.95 and r = 0.92) between VP and GI for two barley samples (varieties Ca108725 and Klages) divided in 10 subsamples, which were artificially aged by heat storage.

Table III shows the correlation coefficient (r) between VP and GI from our experiment. There is no significant correlation for the total material (r = 0.33). When the samples are divided into naturally and heat stored samples, there is no correlation for the naturally aged samples (r = –0.09), but there is a weak positive correlation between VP and GI for the heat stored samples (r = 0.63). When the heat-stored samples are divided into varieties and correlation coefficients are calculated, Alexis has a low correlation value (r = 0.42), but r is higher for Lysimax, Etna and Carula separately and combined (Lysimax, r = 0.86; Etna, r = 0.86; Carula, r = 0.95 and combined r = 0.65). It seems that the correlation between VP and GI in heat stored samples is dependent on variety and that samples that are artificially aged by heat storage have a relatively high correlation coefficient, confirming the results of Aastrup et al.[2]. This result deviates from the naturally aged samples, which display a low correlation coefficient between VP and GI.

GI is determined after germination for three days whereas VP is determined after 8 days of germination. Germination for 8 days is not interesting for the malting industry, but 8 days was chosen to obtain the extreme criterium for vitality and to compare with earlier experiments[2].

TABLE III. Correlation between vigour potential (VP) and germination index (GI).

Correlation VP-GI	n	r
All samples	50	0.33
Group I	24	–0.07
Group II	15	0.4
Group III	11	0.33
Naturally aged	35 + 4	–0.09
Alexis	12 + 1	0.02
Ariel	2	—
Blenheim	7	–0.78
Etna	2 + 1	0
Lysimax	2 + 1	–0.84
Meltan	10	–0.18
Naturally aged 1993	21 + 4	0.01
Naturally aged 1994	14	–0.63
Artificially aged + untr. control	11 + 4	0.58
Artificially aged	11	0.63
Alexis	4	0.42
Carula	3	0.95
Etna	5	0.86
Lysimax	3	0.86
C+E+L	11	0.65

From our experiments it is demonstrated that the vigour model developed by Ellis and Roberts[11] using heat treatment and further developed with vigour potential at Carlsberg[1] is not adequate to classify and describe all barley samples. In addition to the sigmoid curve postulated by the statistical model[1,11] there are at least two other extreme reaction models (Fig. 1A), both comparatively insensitive to heat treatment, but exhibiting high and low germination levels. There is a continuous variation between the three models, which is best detected and visualised by multivariate data analysis (PCA).

A review of the published literature[1,9-13,20-22] shows that earlier experiments with heat treatment of barley were made with very few varieties producing sub-samples by heat treatment of a very limited number of original barley field lots which all were artificially aged. In order to test the vigour models of Ellis and Roberts[11] and Aastrup et al.[1] in our experiment we tested 50 barley samples from 7 different varieties grown in two different years and, in contrast to the other authors, included both artificially and naturally aged samples. Furthermore, we have succeeded in finding naturally aged samples with a broader range in VP than the artificially aged samples.

The results of the current study confirmed those of Riis and Bang-Olsen[19] who performed experiments with heat-treated samples (artificially aged) for which GI was determined. The heat treatment was carried out on four different samples of different varieties: Alexis, Ariel, Triumph and Prisma. The samples reacted differently to the heat treatment, where Ariel in this experiment was most influenced by the heat treatment as opposed to Alexis that in this case was most resistant to the heat. All the samples had lower GI when they were heat-treated[19].

Pre-germination has been shown by several authors to cause increased sensitivity with regard to storage[4,5,18].

In order to elucidate if the observed heat sensitivity and resistance in our experiment is related to pregermination or de novo produced enzymes, the activity of α-amylase was analysed on 28 of the 50 samples.

As can be seen from Table IA the α-amylase activity is generally low in our material. There is no indication of increased α-amylase activity with respect to either the heat sensitive or the heat resistant samples (See Table II).

The analyses show that the content of α-amylase differ from 0.1-2.0 units α-amylase. No correlations between content of α-amylase and e.g. vigour or heat sensitivity is found. As an example the content of α-amylase for the heat-sensitive sample "Alexis no. 11" is compared with the heat resistent "Meltan no. 29" (Fig. 1A and 1B). The sensitive sample has an activity of 0.3 units α-amylase whereas the heat-resistent is 0.2 units α-amylase.

CONCLUSIONS

We have found heat-sensitive samples following the vigour model[1,11] as well as heat-resistant samples[16] with more than 90% germination after heat treatment at 68°C for 4½ h, which do not follow the model. Heat-sensitive samples in our material have a tendency towards a faster germination and therefore a higher GI than heat-resistant samples.

We conclude that the above mentioned model for vigour as related to heat treatment may only be used as a worst case scenario for prediction of survival.

The genetic and environmental background for heat resistance of germination found in several barley samples in this investigation will be discussed in a subsequent publication[17].

ACKNOWLEDGEMENTS

The Seed Company "Skånska Lantmännen" through the SL Foundation and the Danish Cereal Network are greatly thanked for financing this study. We are very grateful to Monica Frank and the rest of the staff at Agrolab, Eslöv for collecting and analysing the barley samples from the 1993 harvest according to germination capacity. We are also grateful to Gilda Kischinovsky for her comments to the article.

REFERENCES

1. Aastrup, S., Riis P. and Hansen, J.R., High vigour – the basis for high malting barley quality. Proceedings of the European Brewing Convention Congress, Zurich, IRL Press: Oxford, 1989, 171-178.
2. Aastrup, S., Riis, P. and Munck, L., Controlled removal of dormancy renders prolonged storage of sprouting resistent malting barley. Fifth International Symposium on Pre-Harvest Sprouting in Cereals, Norway, Westview Press: Oxford, 1989, 329-337.
3. Anonymous, Standard methods of the International Association for Cereal Chemistry (ICC) ICC-standards, Verlag Moritz Schäfer: Detmold, 1998.
4. Bason, M.L., Ronalds, J.A. and Wrigley, C.W., Prediction of safe storage life for sound and weather-damaged malting barley. *Cereal Foods World* 1993, **38**, 361-363.
5. Carn, J.D., Alpha amylase indicates problems with malting of stored barley. *Food Technology in Australia,* 1982, **34**, 82-83.
6. DMI (Danish Meteorological Institute), *Weather in Denmark,* 1993.
7. DMI (Danish Meteorological Institute), *Weather in Denmark,* 1994.
8. Ellis, R.H., Seed and seedling vigour in relation to crop growth and yield. *Plant Growth Regulation,* 1992, **11**, 249-255.
9. Ellis, R.H. and Roberts, E.H., Improved equation for the prediction of seed longevity. *Annals of Botany,* 1980, **45**, 13-30.
10. Ellis, R.H. and Roberts, E.H., The influence of temperature and moisture on seed viability period in barley (*Hordeum distichum* L.). *Annals of Botany,* 1980, **45**, 31-37.
11. Ellis, R.H. and Roberts, E.H., The quantification of ageing and survival in orthodox seeds. *Seed Science and Technology,* 1981, **9**, 377-409.
12. Ellis, R.H. and Roberts, E.H., An investigation into the possible effects of ripeness and repeated threshing on barley seed longevity under six different storage environments. *Annals of Botany,* 1981, **48**, 93-96.
13. Goodspeed, T.H., The temperature coefficient of the duration of life of barley grains. *Botanical Gazette,* 1911, **51**, 220-224.
14. Groves, J.F., Temperature and life duration of seeds. *Botanical Gazette,* 1917, **58(3)**, 169-189.
15. Martens, H. and Næs, T., Multivariate Calibration, Wiley: New York, 1993, 35-72 and 97-100.
16. Møller, B. and Munck, L., Barley seed vitality in relation to heat-susceptibility and -resistance. The VII International Barley Genetics Symposium, University Extension Press: University of Saskatchewan, Saskatoon, Canada, 1996, 646-648.
17. Møller, B., Molina-Cano, J.L. and Munck, L., Variation in malting quality and heat resistance in the malting barley variety "Alexis". *Journal of the Institute of Brewing,* 2002, **108(3)**, 294-302.
18. Pitz, W.J., Rapid and objective methods for estimation of pre-

germination and viability in barley. *Journal of the American Society of Brewing Chemists,* 1991, **49(3)**, 119-127.
19. Riis, P. and Bang-Olsen, K., Breeding high vigour barley – a short cut to fast maltings. Proceedings of the European Brewing Convention Congress, Lisbon, IRL Press: Oxford, 1991, 101-108.
20. Roberts, E.H. and Abdallah, F.H., The influence of temperature, moisture and oxygen on period of seed viability in barley, broad bean and peas. *Annals of Botany,* 1968, **32**, 97-117.
21. Roberts, E.H., The viability of cereal seed in relation to temperature and moisture. *Annals of Botany,* 1960, **24**, 12-31.
22. Robertson, D.W., Lute, A.M. and Gardner, R., Effect of relative humidity on viability, moisture content and respiration of wheat, oats and barley in storage. *Journal of Agricultural Research,* 1939, **59**, 281-291.
23. Spears, J.F. An introduction to seed vigour testing. In: Seed Vigour Testing, Proceedings from 24th Congress of the International Seed Testing Association, H.A. van de Venter Ed., Copenhagen, 1995, 1-9.

(Manuscript accepted for publication April 2002)

Variation in Malting Quality and Heat Resistance in the Malting Barley Variety "Alexis"

Birthe Møller[1], Jose Luis Molina-Cano[2] and Lars Munck[1,3]

ABSTRACT

J. Inst. Brew. 108(3), 294–302, 2002

In earlier studies concerning vigour, where subsamples are heat-treated before germination there was found heat-sensitive as well as heat-resistant barley samples. The vigour model developed by Ellis and Roberts and further developed at Carlsberg, could only describe the heat-sensitive barleys. Seventeen samples of the "Alexis" variety grown widely in Europe were collected from the EBC trials in 1994 in order to see if heat resistance in barley was influenced by different growing conditions. We found both heat-sensitive samples following the vigour model as well as pronounced heat-resistant samples, but these were not divided according to growing conditions. The germination curves dependent on heat treatment and germination time were evaluated by Principal Component Analysis (PCA). Heat-resistant barley samples could be differentiated from heat-sensitive samples already after ½ h of heat treatment at 68°C (12% moisture) and after 3 days of germination. The barley samples were analysed with regard to malting quality. The PCA evaluation of the data divided the samples according to growing location, mainly due to differences in protein and β-glucan. However, the malting analyses could not describe the differences in heat resistance and sensitivity of the barleys. The biochemical background of the heat resistance found is discussed on the basis of literature. Our findings should give an experimental basis for exploiting a biochemical principle for heat resistance, which is formed during grain filling and consumed during storage and germination.

Key words: Heat resistance, malting barley quality, principal component analysis, vigour.

INTRODUCTION

Seed vigour is among the most important characters in malting barley, but also one of the most difficult to measure. Different experiments and vigour models have been developed, for example, including heat stress and vigour loss curves after different periods of heat treatment. Ellis and Roberts' vigour model[9] was based on the assumption that ageing in a seed sample adheres to a normal distribution model, where the ageing process follows a sigmoid curve. Based on the model of Ellis and Roberts[9], Aastrup et al.[2] further developed the vigour model defining the Vigour Potential (VP) which corrects for dead kernels. In earlier studies[14] barley samples have been identified where heat treatment as high as 68°C for 4½ h only decreases germination from 99.0% to 93.8% compared with 94.8% to 0.0% for some of the heat-sensitive barleys which follow the above-mentioned model.

In the following we will investigate if it is possible to find both heat-resistant and heat-sensitive samples within the same variety. Furthermore, we will investigate if heat resistance is related to location of growing and to malting quality. We have chosen the variety "Alexis" as an example, because it has been a well-accepted malting barley in many countries in Europe.

MATERIALS AND METHODS

Seventeen barley samples of the variety "Alexis" were collected from locations in Finland (Su), Germany (D), Spain (E), Denmark (Dk), the Czech Republic (Cz) and the Netherlands (Nl) in 1994.

Ten of the samples were material from EBC-trials grown in each of the four regions: North, West, Central and South Europe. Two of the ten samples were from the same location, Jokoinen in Finland (11Su and 12Su).

The samples were collected and analysed for malting quality at Centre UdL-IRTA. In the summer of 1995 KVL received the samples, and analysed them in the autumn for germination properties. The samples were stored at 4°C.

Growing conditions

The trials in Spain were sown in January and harvested in June, whereas the rest of the trials were sown from late March to early May and harvested from late July to early September. The samples in Spain had the longest growing season (162 days), whereas the other samples had a growing season between 87 and 119 days.

Temperature in the month when the plants headed was highest in Germany and Finland. With regard to precipitation, Irlbach in Germany (16D) and Pälkäne (15Su) in Finland had the most rain (75-85 mm), whereas the four locations in Spain (4E, 5E, 6E and 8E) and Jokoinen (11Su and 12Su) in Finland had less than 10 mm of rain during that month.

Thus, the collected set of samples from the same variety show great differences in growing conditions.

[1] The Royal Veterinary and Agricultural University, Dairy and Food Science, Food Technology, Rolighedsvej 30, Dk-1958 Frederiksberg C
[2] Centre UdL-IRTA, Lleida, Spain. E-mail: JoseLuis.Molina@irta.es
[3] Corresponding author: E-mail: lmu@kvl.dk

Publication no. G-2002-0826-045

Analyses determined at the Royal Veterinary and Agricultural University (KVL), Denmark

Moisture content was determined according to ICC 110/1 and ash content (A) according to ICC 104/1[6].

A. Germination. 4 × 100 kernels were germinated in petri dishes with two layers of filter paper and 4 mL of water for 8 days at 20°C in a dark Refritherm incubator. The percentage of germinated kernels was determined every day, and after 8 days the total germination percentage was determined. Four replicates were made for every sample, standard deviation between the replicates was less than 5%.

The Germination Index (GI) was calculated on the untreated samples according to Riis and Bang-Olsen[16] from the germination percentages after 24, 48 and 72 h of germination.

B. Heat treatment. To determine the vigour potential each sample was divided into 10 subsamples. If the samples had a water content higher or lower than 12% they were exposed to the relative humidity and temperature of the laboratory and moistened to 12.0% H_2O according to Møller and Munck[15]. The subsamples (water content = 12%) were heat treated in sealed aluminium bags at 68°C in a water bath from 0 to 4½ h, where sub-samples were taken every half hour. After these treatments 4 × 100 kernels from every sub-sample were germinated and a value of vigour potential (VP) could be calculated from the germination curve using the Carlsberg vigour model[2].

Due to lack of material, the two-hour heat treatment for samples 1-6 (Denmark and Spain) and 13-14 (Finland) and additional three-hour heat treatment for sample 16 (Germany) was left out. Thus in these samples the number of heat-treatments were reduced from 10 to 8. Furthermore, the two-, three- and four-hour treatments were left out for sample 17D.

The missing values in the heat-treatment germination curves were interpolated.

Barley and malt quality analyses determined at the University of Lleida, Spain

Protein (P) and β-glucan (BG) were determined in raw barley[5]. The micromaltings were performed at the University of Lleida, Spain. Insoluble β-glucan (BIS), Colour (C), Extract (Ext), Total protein in malt (MTP), Soluble protein in malt (MSP), Kolbach (K) and Viscosity (V) were determined after malting according to EBC[5] and McCleary and Glennie-Holmes[11].

Multivariate data analysis

Principal Component Analysis (PCA) was performed according to Martens and Næs[10]. The aim of the PCA algorithm is to calculate the latent factors, which describe the variation in the data set as Principal Components (PC's). Based on vector and algebra the algorithm calculates and compresses the data (e.g. whole germination curves after heat treatment) into scores for principal components which are plotted in a scoreplot. Samples with similar scores have similar patterns of their germination curves. The variables involved (e.g. total germination percent after heat treatment from 0 (g%0) to 4½ h (g%4½)) expressed as loadings can be plotted together with the scores in a biplot. A score position for a germination curve placed near a loading indicates a high influence by that variable.

RESULTS

Vigour estimated by the effect of heat stress on germination

The 8-day germination curves for heat treatment of 0-4½ h for extreme and intermediate samples of Alexis in Fig. 1 (selected from Fig. 2) show a great variation involv-

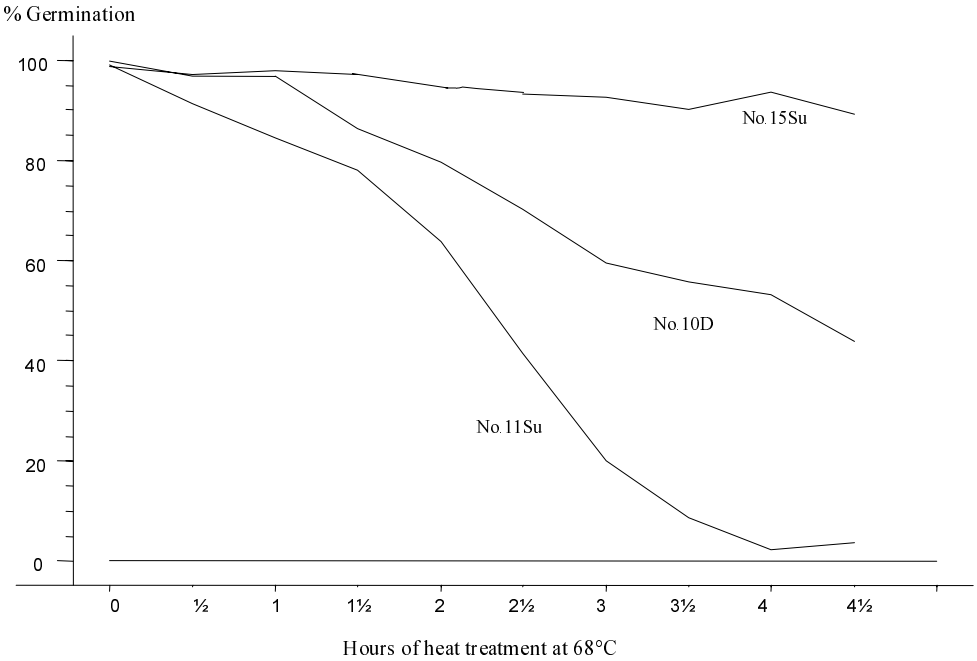


FIG. 1. Germination curves for three extreme Alexis samples (marked with squares in FIG. 2) determined as total germination percentage after heat treatment at 68°C for 0-4½ h followed by eight days of germination.

ing both heat-resistant and heat-sensitive samples as well as intermediates. The curves change from reversed S-shaped to almost horizontal curves. There are samples (Fig. 1) which are heat-sensitive and follow the statistical reversed s-shaped curve form for the vigour-model developed by Ellis and Roberts[9] and Aastrup et al.[2] such as 11Su, as well as heat-resistant samples according to Møller and Munck[15] such as 15Su which have a total germination up to 98% after 4½ h of heat treatment at 68°C. In between the extreme germination curves there are intermediates such as sample 10D.

Using the Principal Component Analysis (PCA) algorithm to represent data from the patterns of the germination curves, a biplot with scores and loadings is shown in Fig. 2. In the biplot the scores of the samples depict the form of the germination curves. Samples with similar germination curves appear near each other in the score plot. The position of the extreme samples (squared in Fig. 2) 11Su, 10D and 15Su discussed earlier represent the different categories of heat sensitive, intermediate and resistant. The variables (0-4½ h of heat treatment) represented as loadings (encircled in Fig. 2) illustrate how these variables influence the sample scores. Variables close to samples show a high degree of influence. In the biplot in Fig. 2 the heat-sensitive samples 2Dk, 4E, 5E, 6E and 11Su (underlined in Fig. 2) are placed to the left, and the heat-resistant samples such as 3Dk, 7Nl, 8E, 9Cz, 13Su, 14Su, 15Su, 16D and 17D are placed to the right. In between are the intermediate samples 1Dk, 10D and 12Su. These samples are more closely connected with the heat-resistant samples than the heat-sensitive. The position of the loadings in Fig. 2 shows that the variables g%1½ (germination percentage after 1½ h of heat-treatment) to g%4½ (germination percentage after 4½ h of heat-treatment) are placed in the lower right corner near the heat-resistant samples, indicating high influence (high germination values) of the heat-treated samples.

The heat treatment germination curves in Fig. 1 are based on the final germination percentage after 8 days of germination displayed as a PCA plot in Fig. 2. The development in germination (1, 2, 3 and 6 days) from the entire sample material is presented in PCA score plots in Figs. 3A-D. In the beginning, after one day of germination, all the samples in the score plot are placed on a vertical line to the left, except sample 16D, which is an outlier placed to the right (Fig. 3A). Sample 16D has a very fast germination with at least 85% germinated kernels after 24 h in the samples heat-treated up to 1 h. The other barley samples do not differ at 24 h of germination according to heat sensitivity.

After two days of germination (Fig. 3B) there is a tendency towards a differentiation, so that the heat-sensitive samples are placed in the bottom left corner and the heat-resistant samples (except 16D) are placed in the top right corner.

After three days of germination (Fig. 3C) a complete separation into heat-sensitive samples and heat-resistant samples has taken place. The intermediate samples such as 1Dk, 10D and 12Su are placed near the heat-resistant samples, indicating that at this stage they have a similar germination profile after 3 days of germination. After 6 days germination (Fig. 3D) the pattern is almost the same as for 8 days of germination (Fig. 2), where the earlier fast germinating outlier sample 16D has now been included in the heat-resistant group. Heat-sensitive samples are placed to the left, heat-resistant samples to the right, and samples

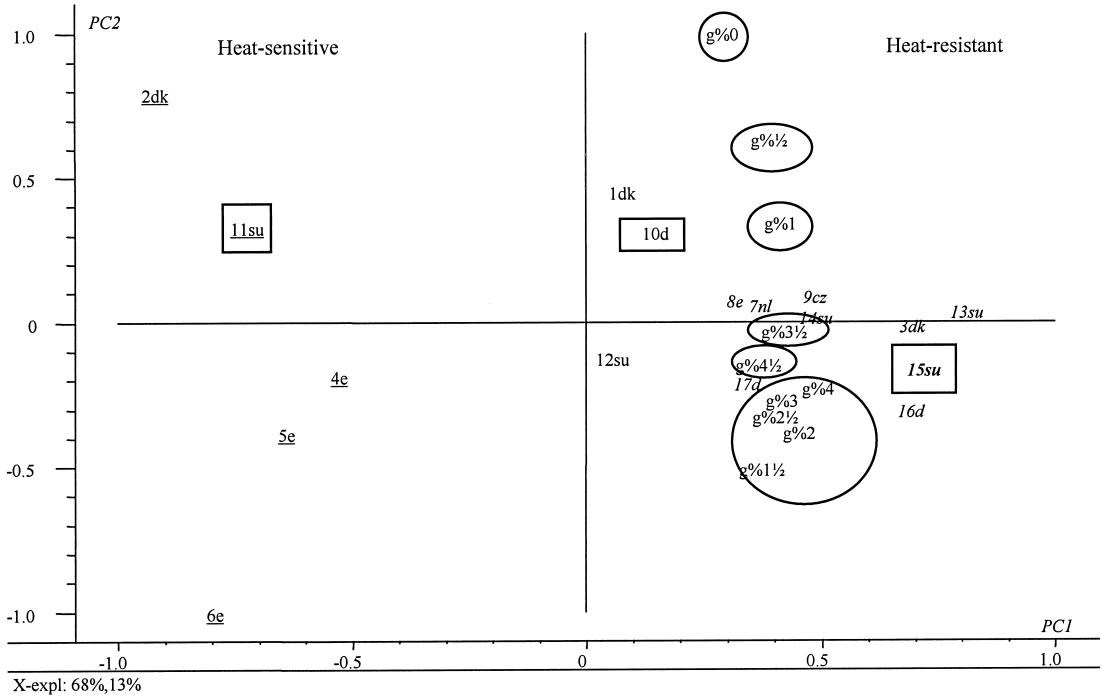


FIG. 2. Biplot for the heat treatment curves of the 17 samples. CZ = Czech Republic, D = Germany, DK = Denmark, E = Spain, NL = Netherlands, SU = Finland. Variables are percentage germinated kernels (8 days germination) after heat treatment at 68°C for 0-4½ h. Underlined samples = heat sensitive, *italic samples* = heat resistant. Variables (loadings) are marked with a circle. G%½ = germination percentage after ½ h heat treatment. Heat treatment curves for the squared samples are displayed in FIG. 1.

1Dk, 10D and 12Su are intermediate to the two groups, but are closest to the heat-resistant samples.

Another way to look at this extensive data set is to separately address the development of germination from 1 to 8 days for the untreated material and for the different heat-treatments. An example is shown in Figs. 4A-B for untreated samples (Fig. 4A) and samples heat-treated for ½ h (Fig. 4B). In the score plot for the untreated samples (Fig. 4A) there is not a clear division between the heat-sensitive and heat-resistant samples, although three of the five heat-sensitive samples are placed to the left in the plot, separated from the rest of the samples. When looking at the curves for development of germination from 1 to 8 days in samples heat treated for ½ h in a PCA plot for pattern recognition, it is already possible to see a clear differentiation between heat-sensitive and heat-resistant samples (Fig. 4B). Loadings in the biplot are marked with a circle and are placed in a vertical line from right top corner to right bottom corner. The variable d1 (percentage of germinated kernels after 1 day of germination) is placed closest to sample 16D, because 16D is the sample with the highest percentage of germinated kernels after one day of germination. The heat-resistant samples have a higher germination percentage than the heat-sensitive samples after heat treatment, and it is therefore natural that these variables are placed close to the heat-resistant samples, as can be seen in Fig. 4B.

It is not possible with classical statistics to treat the pattern of germination curves as a whole as is done in PCA. Instead differences at different levels of heat treatment

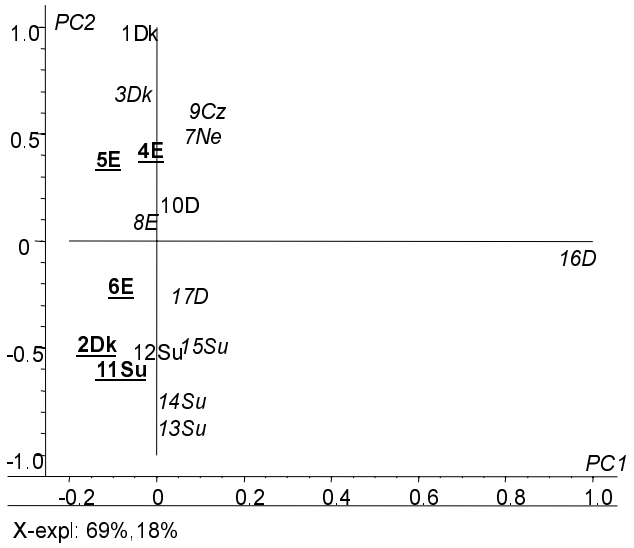


FIG. 3A. Score plot for germination heat treatment curves after 1 day of germination. Underlined = heat sensitive, *italics* = heat resistant samples.

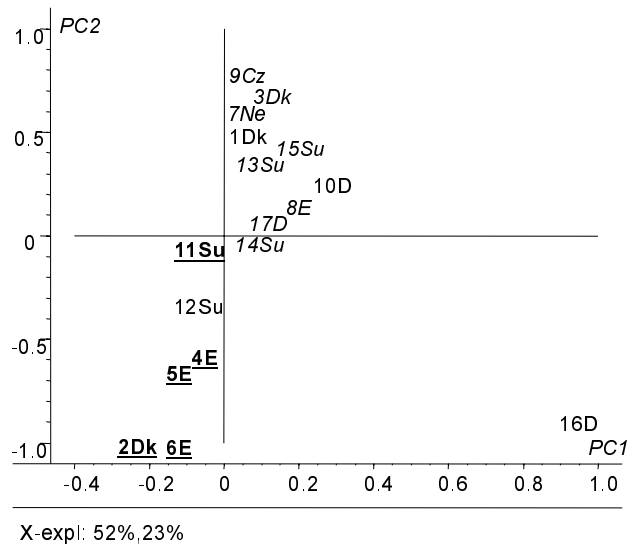


FIG. 3B. Score plot for germination heat treatment curves after 2 days of germination. Underlined = heat sensitive, *italics* = heat resistant samples.

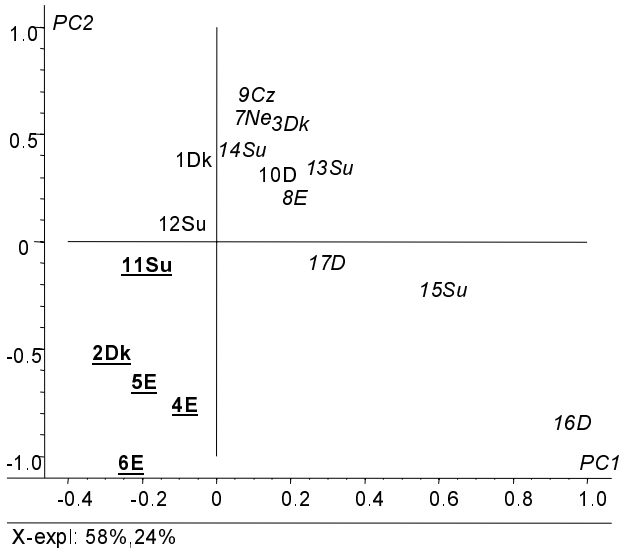


FIG. 3C. Score plot for germination heat treatment curves after 3 days of germination. Underlined = heat sensitive, *italics* = heat resistant samples.

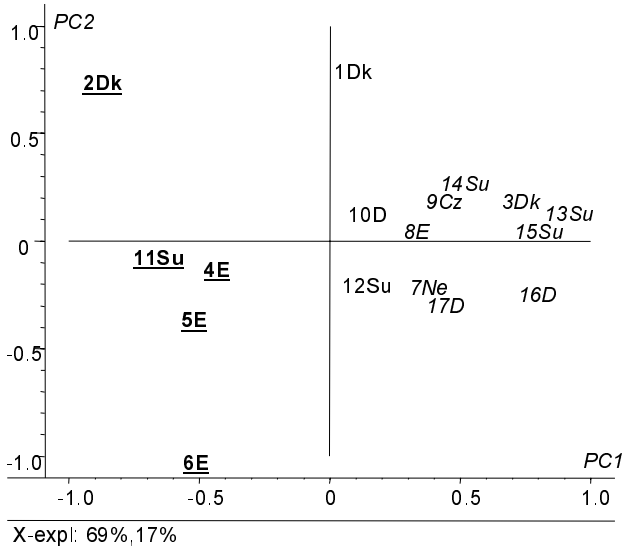


FIG. 3D. Score plot for germination heat treatment curves after 6 days of germination. Underlined = heat sensitive, *italics* = heat resistant samples.

and germination time, including a control, were compared (5% confidence interval) for the three groups selected from the PCA plot: heat sensitive, heat resistant and intermediate after 3 days of germination for the untreated and ½ h heat treated samples. From Table I it is seen, that there is no significant difference between the untreated samples after three days of germination. However comparing the 3 day germination percentage for the groups after ½ h heat-treatment, there is a significant difference (mean 82.1% and 96.2%) between the heat sensitive and heat resistant groups, where it is possible to differentiate between extreme heat sensitive and extreme heat resistant samples.

The difference in germination between the heat resistant and heat sensitive groups increases at 3½ h heat-treatment to 53 absolute percentages (mean 79.0% and 26.0% respectively). This tendency is conserved at 4½ h heat-treatment (mean 67.7% and 23.7% respectively). However, at 4½ h of heat-treatment this difference is not statistically significant at a 5% level, because of an increased variation induced by heat-treatment.

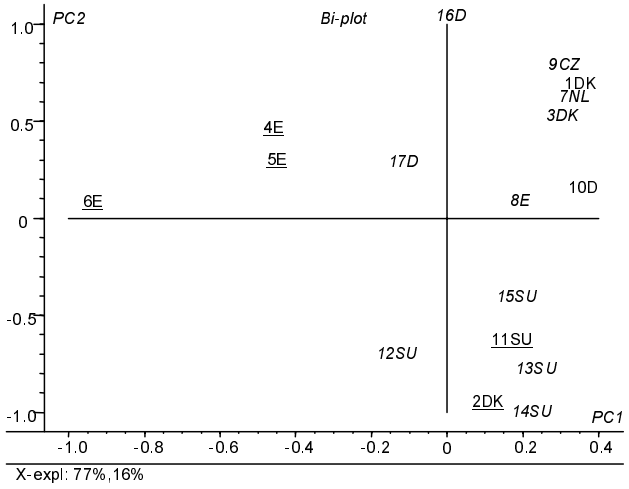


FIG. 4A. Biplot (loadings not displayed) for germination development (1-8 days) for untreated samples. Underlined = heat sensitive, *italics* = heat resistant samples.

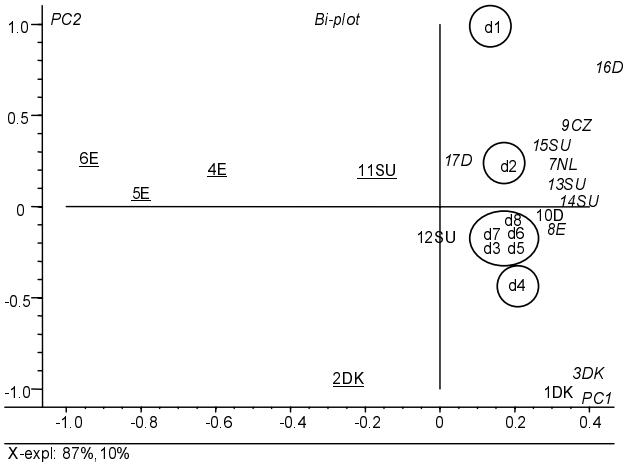


FIG. 4B. Biplot for germination development (1-8 days) for samples heat treated for ½ h. Underlined = heat sensitive, *italics* = heat resistant samples. Variables (loadings) d1-d8 are germination percentage on days 1-8.

Barley and malt quality analyses

The variation of the barley and malt quality variables for the samples grown in different countries is documented in Table II and visualised in the PCA biplot in Fig. 5. As can be seen from the figure, the samples divide according to national location following principal component 1 (PC1). Samples grown in Spain are placed to the left in the biplot; samples from Finland are placed in the middle, and the samples from Denmark, Germany, the Czech Republic and the Netherlands are to the right. In the biplot the loadings for the variables concerning protein (P, MTP, MSP) and β-glucan (BG, BIS) are placed to the left together with the Spanish samples, indicating high levels. Three samples from Finland (11Su, 12Su, 14Su) are placed to the left in the plot along with the Spanish samples, although closer to the middle. Sample 13Su from Finland is placed near the variable viscosity (V), indicating a high influence on the sample. When comparing 13Su with the other samples, it is found that 13Su has the highest value of viscosity (V = 1.72). Protein in barley (P) and total protein in malt (MTP) are placed at the same position in the plot, indicating a high correlation (r = 0.98). Soluble protein in malt (MSP) and colour (C) are placed in the bottom left corner near samples 4E and 5E. This is as expected, because samples 4E and 5E are those with the highest values for these variables (MSP = 7.40 and 6.89, C = 9.90 and 8.50). Samples 6E and 8E are placed in the top left corner near beta-glucan (BG) and water insoluble β-glucan (BIS), which is evident when sample 6E has the highest value of BG and BIS compared to the other samples (BIS = 3.17, BG = 4.95). Sample 8E, which is placed near BG, has a high value of BG as well (BG = 4.43).

There seem to be small differences between the rest of the samples, except sample 16D, which is placed in the bottom right corner near the variable K. This is as expected, since 16D has a very high value of Kolbach Index (K = 54.23) compared to the other samples, which is consistent with its rapid germination development.

TABLE I. Average and 5% confidence interval for germination percentage after three days of germination for untreated and ½ h heat treated samples, and for germination after eight days for samples heat treated 3½ and 4½ h. All calculated for heat sensitive, heat resistant and intermediate samples. Different letters refer to significant differences (5% level) between groups.

		3 Days of germination		8 Days of germination	
	No. of samples	Untreated	½ hour	3½ hour	4½ hour
Heat sensitive	5	91.2 ± 3.6 A	82.1 ± 4.3 A	26.0 ± 9.3 A	23.7 ± 13.2 A
Heat resistant	9	97.1 ± 1.9 A	96.2 ± 2.5 B	79.0 ± 6.4 B	67.7 ± 9.9 A
Intermediate	3	97.2 ± 2.7 A	95.1 ± 4.8 AB	66.3 ± 19.5 AB	45.0 ± 40.2 A

TABLE II. Data for the 17 samples of Alexis grown on different locations in Europe in 1994. Co = country, P = protein, BG = total beta-glucan, BIS = water insoluble beta-glucan, C = colour, Ext. = extract, MTP = total protein (malt), MSP = soluble protein (malt), K = Kolbach index, V = extract viscosity, A = ash, Vp = vigour potential, Gtot = total germination percentage after 72 h, GI = germination index.

No.	Co	P	BG	BIS	C	Ext	MTP	MSP	K	V	A*	VP*	G_tot*	GI*
1	Dk	9.97	3.40	2.15	2.20	78.31	11.38	4.32	37.99	1.59	1.82	1.5	99.25	8.5
2	Dk	8.84	4.16	2.85	2.20	81.60	9.52	3.73	39.18	1.70	2.35	4.8	98.50	5.7
3	Dk	8.88	3.86	2.91	2.00	81.99	10.10	3.98	39.41	1.82	1.80	1.7	99.25	8.5
4	E	14.10	4.26	3.06	9.90	79.41	15.24	7.40	48.55	1.62	2.00	1.1	94.25	6.4
5	E	13.91	4.24	2.97	8.50	80.84	14.73	6.89	46.74	1.62	1.98	1.1	93.75	6.6
6	E	13.45	4.95	3.17	4.00	78.78	15.06	5.50	36.55	1.79	1.73	2.3	89.00	5.9
7	Ne	9.16	3.65	2.38	2.20	81.67	9.71	4.23	43.51	1.68	2.12	3.3	99.00	8.4
8	E	12.58	4.43	2.99	3.20	78.52	14.41	5.49	38.07	1.60	1.79	3.3	98.50	7.1
9	Cz	9.25	3.56	2.17	3.00	79.93	10.39	4.09	39.32	1.66	2.14	2.6	99.25	8.8
10	D	9.05	3.70	2.54	2.50	81.51	9.85	4.65	47.22	1.74	1.94	1.8	99.75	7.5
11	Su	12.47	4.05	2.08	2.50	78.41	14.60	5.73	39.26	1.56	1.79	2.2	99.00	6.0
12	Su	13.30	4.35	2.45	2.50	78.45	13.64	5.70	41.79	1.68	1.92	2.6	97.00	5.6
13	Su	10.41	4.12	2.45	2.00	79.14	11.80	4.35	36.90	1.72	1.89	2.8	100.00	6.0
14	Su	11.42	4.10	2.90	2.50	79.30	12.45	5.21	41.88	1.63	1.89	2.3	99.00	5.7
15	Su	9.42	3.94	2.67	4.00	81.19	10.39	5.00	48.11	1.65	1.85	2.0	98.75	6.7
16	D	10.32	3.60	1.15	7.00	83.06	11.71	6.35	54.23	1.55	2.27	2.0	97.25	8.8
17	D	8.91	3.41	1.95	2.50	81.82	9.99	4.66	46.63	1.58	1.90	2.0	97.25	6.9

*Analyses made at the Royal Veterinary and Agricultural University, Copenhagen, Denmark.
All other analyses made at University of Lleida, Lleida, Spain.

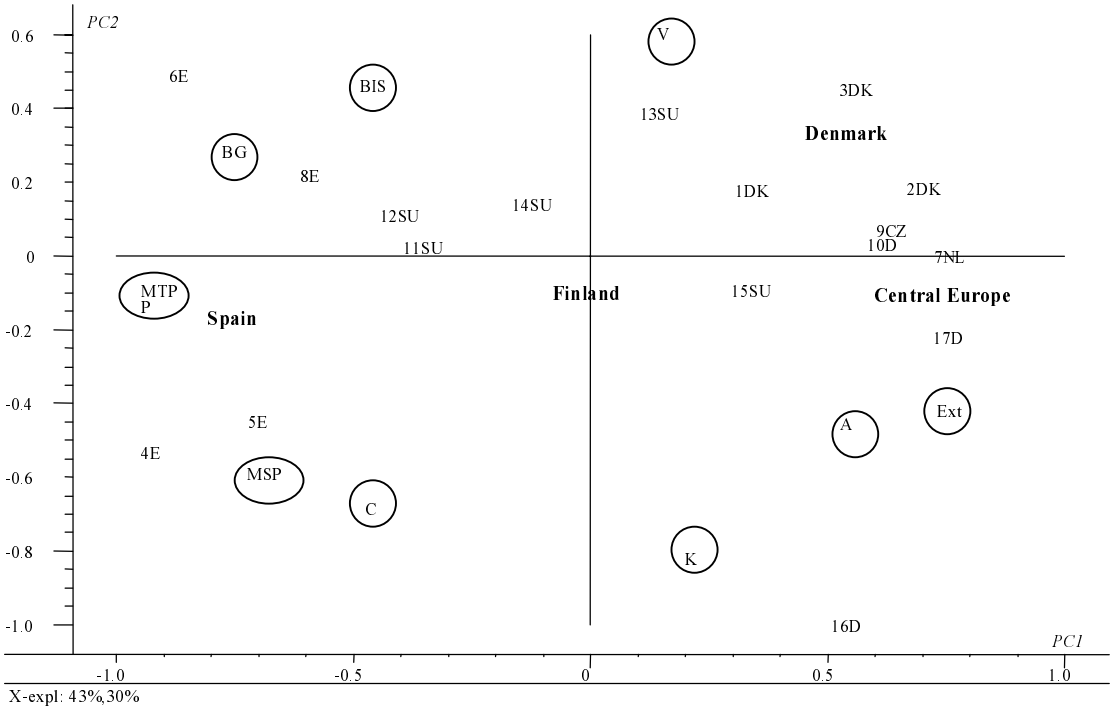


FIG. 5. Biplot for the chemical data from the 17 Alexis samples. CZ = Czech Republic, D = Germany, DK = Denmark, E = Spain, NL = Netherlands, SU = Finland. Variables are data as shown in TABLE 1. P = protein, BG = total β-glucan, BIS = water insoluble β-glucan, C = colour, Ext. = extract, MTP = total protein (malt), MSP = soluble protein (malt), K = Kolbach index, V = viscosity, A = ash.

DISCUSSION

The practical significance of the heat stress seed vigour model – Model not adequate to describe vigour

In Fig. 1 it is clearly seen that the vigour model developed by Ellis and Roberts[9] and further developed at Carlsberg[2] is not adequate to describe vigour, confirming the earlier published results by Møller and Munck[14,15]. There are heat-sensitive samples following the vigour model as well as heat-resistant samples that do not follow their model.

The sample with the highest value of vigour potential (VP) according to Aastrup et al.[2] is 2Dk (VP = 4.8), which is the most heat-sensitive of the 17 samples. This should not be the case, as the heat-resistant samples can resist stress factors (heat) better than sample 2Dk, and therefore are supposed to have a higher vigour potential. The average values for VP seen in Table III indicate that the aver-

age value for the heat-resistant samples is 2.42 compared to VP = 2.30 for the heat-sensitive samples. The intermediate samples (1Dk, 10D and 12Su) have an average VP = 1.97, which is the lowest average value of the three groups, although these intermediate samples can resist the stress in the form of heat better than the heat-sensitive samples.

We have earlier studied 50 barley samples from seven different malting and feeding varieties, including naturally as well as artificially aged samples, where we have also registered heat-sensitive as well as severe heat-resistant samples[15]. In the present study we have chosen to work with only one variety "Alexis" and with naturally aged samples from very different growing conditions. The results demonstrate that it is possible to find heat-sensitive as well as heat-resistant samples within the same variety.

Using Germination Index to describe vigour

Germination Index (GI) is another possible measurement for vigour[16]. A comparison of the average values of GI for the three groups reveals their distribution is different for VP. The average value for GI for the heat-resistant group is 7.38, the group with the intermediate samples has an average value of 7.19, while the heat-sensitive samples have the lowest average value of the three groups ($GI_{average}$ = 6.10). The differences between the groups with respect to GI and VP are not significant.

Aastrup et al.[3] found high correlation coefficients (r = 0.95 and r = 0.92) between VP and GI in 10 subsamples of two barley varieties (Ca108725 and Klages) which were artificially aged. Our material with naturally aged samples shows no correlation between VP and GI (r = –0.20). In an earlier experiment[15] we found that the heat-sensitive samples had slightly higher values of GI compared to heat-resistant samples. The difference was significant. This is not the case in this set of samples. However in the present experiment GI was determined in Petri dishes with water in contrast to test tubes with 0.75% H_2O_2 solution for the earlier experiment by Møller and Munck[15].

Future methods to evaluate vigour

It is concluded from the above discussion that the model of accelerated ageing by short- time heat treatment to determine vigour introduced by Ellis and Roberts[9] is obviously not valid for heat resistant and intermediate barley samples which naturally can be found in abundance. The vigour model of Ellis and Roberts[9] and further developed by Aastrup et al.[2] could, however, be used as a last resort, rejecting the samples with the lowest vigour. By using the heat stress vigour model it should be possible according to its advocates to calculate the vigour loss after long-term storage at a lower temperature. One would think that heat treatment at 50-68°C does not affect the same biochemical and physiological processes that in practice will lead to a change in vigour under field and storage conditions.

Alternative methods worth investigating could include other types of stress, e.g. the Hiltner test, where a layer of Ziegel gravel stresses germination[17].

In the malting house the aim is to create an optimal environment for the germinating barley, and it is seldom that the barley is stressed by lack of water or extreme temperatures. Instead, the most common stress for barley in the malting house is lack of oxygen. Therefore, a better research hypothesis to support the malting industry might be to induce stress with different rates of oxygen instead of different periods of heat treatment. However, such a method would be difficult to standardise. Therefore, the best candidate method to estimate vigour would be to use the speed of germination, calculating the germination index determined at optimal conditions with filter paper in Petri dishes[16]. This method is the easiest to standardise and it is now commonly used all over the world.

The causes of heat sensitivity and resistance

The large difference between heat-resistant and heat-sensitive barley observed in this study in the variety "Alexis" and in other varieties[14,15] is an interesting phenomenon as

TABLE III. Chemical and germination analysis for determination of malting quality. Average value and standard deviation according to location and heat sensitivity. Abbreviations: see TABLE II.

Country	n	P	BG	BIS	C	Ext	MTP	MSP
Dk	3	9.23 ± 0.64	3.81 ± 0.38	2.64 ± 0.42	2.13 ± 0.12	80.63 ± 2.02	10.33 ± 0.95	4.01 ± 0.30
E	4	13.51 ± 0.68	4.47 ± 0.33	3.05 ± 0.99	6.40 ± 3.30	79.39 ± 1.04	14.86 ± 0.37	6.32 ± 0.98
D	3	9.43 ± 0.78	3.57 ± 0.15	1.88 ± 0.70	4.00 ± 2.60	82.13 ± 0.82	10.52 ± 1.04	5.22 ± 0.98
Central Eur.	2	9.34 ± 0.56	3.58 ± 0.11	2.04 ± 0.54	3.44 ± 2.01	81.60 ± 1.12	10.33 ± 0.81	4.80 ± 0.90
Su	5	11.40 ± 1.55	4.11 ± 0.15	2.51 ± 0.30	2.70 ± 0.76	79.30 ± 1.13	12.58 ± 1.63	5.20 ± 0.57
Heat-sensitive	5	12.55 ± 2.17	4.33 ± 0.36	2.83 ± 0.43	5.42 ± 3.56	79.81 ± 1.36	13.83 ± 2.42	5.85 ± 1.42
Heat-resistant	9	10.04 ± 1.27	3.85 ± 0.33	2.40 ± 0.59	3.16 ± 1.58	80.74 ± 1.56	11.22 ± 1.53	4.82 ± 0.78
Intermediate	3	10.77 ± 2.24	3.82 ± 0.49	2.38 ± 0.20	2.40 ± 0.17	79.42 ± 1.81	11.62 ± 1.91	4.89 ± 0.72

Country	n	K	V	A	Gtot	GI	VP
Dk	3	38.86 ± 0.76	1.70 ± 0.12	1.99 ± 0.31	99.00 ± 0.43	7.44 ± 1.56	2.7 ± 1.9
E	4	42.48 ± 6.04	1.66 ± 0.09	1.88 ± 0.14	93.88 ± 3.89	6.49 ± 0.49	1.9 ± 1.1
D	3	49.36 ± 4.23	1.62 ± 0.10	2.04 ± 0.20	98.08 ± 1.44	7.70 ± 0.99	1.9 ± 0.1
Central Eur.	2	46.18 ± 5.48	1.64 ± 0.08	2.07 ± 0.15	98.50 ± 1.17	8.06 ± 0.87	2.3 ± 0.6
Su	5	41.59 ± 4.19	1.65 ± 0.06	1.87 ± 0.05	98.75 ± 1.09	5.99 ± 0.43	2.4 ± 0.3
Heat-sensitive	5	42.06 ± 5.26	1.66 ± 0.09	1.97 ± 0.24	94.90 ± 4.07	6.10 ± 0.39	2.30 ± 1.51
Heat-resistant	9	43.12 ± 5.64	1.65 ± 0.08	1.96 ± 0.17	98.69 ± 0.92	7.38 ± 1.19	2.40 ± 0.61
Intermediate	3	42.33 ± 4.64	1.67 ± 0.08	1.89 ± 0.06	98.67 ± 1.46	7.19 ± 1.47	1.97 ± 0.57

such and needs an explanation. If heat resistance could be controlled in cereal stocks, heat treatment, for example against fungi or insects, might be feasible as an alternative to chemicals. This phenomenon studied at 12% water content is apparently not to any higher degree associated to the genotype. Future studies should further elucidate the importance of different levels of water content other than 12% in heat treatment to differentiate between sensitive and resistant types.

Effects due to location and climate

With regard to the positive effect of climate on germination, it is seen (Table II) that three samples from Spain (No. 4E-6E) show low germination values (94.00-89.00%) compared to all other samples (100.00-97.25%). It could not be excluded that the low germination percentage of the Spanish samples were due to pregermination. Due to lack of material α-amylase activity was not determined in order to test pregermination. We have in earlier experiments confirmed that pregermination does not relate to heat-sensitivity or heat-resistance[15]. In this material extreme heat-resistant and heat-sensitive samples were found in three of the six countries (DK, E and SU) with widely different climates. It is thus concluded that the heat-sensitive heat-resistant mechanisms is not likely to be related to climatic conditions.

The PCA score plot for the germination curves (Fig. 2) on heat treated samples displays no specific pattern for growing location. In samples from Spain (E), Finland (Su) and Denmark (Dk) heat-sensitive as well as heat-resistant samples are found. In the figures describing the germination development (Figs. 3A-D and 4A-B) there is no systematic difference due to growing location either.

In relation to differences in malting quality due to climatic conditions, it is observed that precipitation in the three-week period after heading could influence the content of β-glucan in grain. Aastrup[1] used six barley genotypes with large inherited differences in β-glucan contents. Duplicate sets of plants were grown in plant growth chambers under conditions giving excellent vegetative growth for eight weeks. One set of plants was rained upon; the other set received the same amount of water to the roots. At full ripeness the total β-glucan content for the rain-treated plants was significantly lower than in grains from the control plants. In a study of barley grown in field on different locations in Denmark Aastrup[1] found that the extract viscosity was negatively correlated to the precipitation and to relative humidity. These field results confirm that extract viscosity and thus the soluble β-glucan content are affected during later stages of ripening by both precipitation and relative humidity.

A correlation of precipitation from the period of supposed heading for the samples in this experiment with the total content of β-glucan at harvest reveals a trend ($r = -0.67$). The exact week for heading was not registered.

Differences in malt quality

In contrast to heat stress germination data, there is a clear location influence on the barley and malt quality analyses (Table III and Fig. 5). This differentiation is mainly due to the effect of climate on protein and β-glucan content. This is in accordance with Molina-Cano et al.[12] who studied differences in water uptake in two varieties at two contrasting locations (Spain and Scotland) in two consecutive years. Molina-Cano et al.[12] found that two groups of grain components are relevant to water uptake: reserve proteins and β-glucans. The horizontal axis in the PCA separated Spain from Scotland, and expressed the joint effect of protein and β-glucans. The vertical axis in the PCA expressed the negative effects of water insoluble β-glucans, total proteins and the B/C hordein ratio on both final water uptake and malt extract yield, and separated the samples grown in Spain in 1991 from those grown in Spain in 1992.

There are no significant differences in the chemical data between the heat-sensitive samples and the heat-resistant samples. In Table III it is seen that the group of heat-sensitive samples has higher average values for protein (P), total malt protein (MTP) and soluble malt protein (MSP) than the heat-resistant group. The intermediate samples have average values in between. The differences are, however, not significant.

The biochemistry of heat sensitivity and resistance in relation to water content and heat treatment

Accelerated ageing by heat treatment at 12% moisture content as reported in this paper has been used by several investigators[2,8,9]. Prolonged storage of wheat and barley seeds at 12% moisture content at moderate temperature (35°C) for 28 days led to a decrease in protein synthesis after imbibition in water for 20 h at 20°C[8]. Germination deteriorated from 96% to 92%. No remarkable changes were found in the synthesis in the embryo of the most "normal" germination proteins. Increasing moisture content during heat treatment in wheat to 14.5% drastically decreased germination to 5% and the protein synthesis after imbibition of a range of specific proteins[7]. Imposing a 4 h stress period at 40°C during water imbibition produced heat shock proteins (HSP's) in heat-treated (12% water at 35°C) barley and wheat as well as in controls.

The results were confirmed using the 2,3,5 triphenyltetrazoliumchlorid assay on embryos. The authors hypothesized a thermotolerant principle. The HSP's from 60-14 kD after heat shock declined from 1.5 to 12 h of imbibition, whereas the high molecular HSP's at about 68 and 94 kD were increasing. Abernethy et al.[4] discuss the possibility that HSP's could be synthesised as a normal part of seed development and retained during seed desiccation and early germination during which time they undergo degradation. Temperature stress during maturation could be the critical factor in releasing heat resistance.

Our results strongly support the build-up of a heat resistance principle during maturation of barley, although we have not been able to trace the conditioning window nor the influence needed during maturation. The fact that extreme heat-resistant barley samples can be found in great numbers in practice makes further biochemical studies of the underlying heat resistant principle possible. In such a study the possibility of inducing heat resistance by gene-environment interaction, preharvest and/or post-harvest conditioning should be investigated. In parallel, the possible positive practical implications of seed heat resistance for the industry should be studied.

CONCLUSION

From these results it is clear that heat-resistance is not dependent on variety, as we have succeeded in finding both heat-sensitive and heat-resistant samples within the same variety. The barley variety Alexis was grown at different locations with extreme differences in climate. It has not been possible to see a pattern in heat resistance according to location. During kernel development, different genes are activated at different times[13]. Environmental conditions during a critical but short period of development might shift the phenotype of the seed between heat-resistance and heat-sensitivity. The role of the heat shock proteins as well as other biochemical protection systems should be investigated in this context.

REFERENCES

1. Aastrup, S., The effect of rain on betaglucan content in barley grains. *Carlsberg Research Communication,* 1979, **44**, 381-393.
2. Aastrup, S., Riis, P. and Hansen, J.R., High vigour – the basis for high malting barley quality. Proceedings of the European Brewing Convention Congress, Zurich, IRL Press: Oxford, 1989, 171-178.
3. Aastrup, S., Riis, P. and Munck, L., Controlled removal of dormancy renders prolonged storage of sprouting resistant malting barley. Fifth International Symposium on Pre-Harvest Sprouting in Cereals, Norway , Westview Press: Oxford, 1989, 329-337.
4. Abernethy, R.H., Thiel, D.S., Petersen, N.S. and Helm, K., Thermotolerance is developmentally dependent in germinating wheat seed. *Plant Physiology,* 1988, **89**, 566-576.
5. Anonymous, European Brewery Convention: Analytica-EBC, 4th edition, Bruhin AG Druckerei: Freienbach, Switzerland, 1987.
6. Anonymous, Standard Methods of the International Association for Cereal Chemistry (ICC) ICC-standards, Verlag Moritz Schäfer: Detmold, 1998.
7. Dell'Aquila, A., Wheat seed ageing and embryo protein degradation. *Seed Science Research,* 1994, **4**, 293-298.
8. Dell'Aquila, A., Corona, M.G. and Di Turi, M., Heat-shock proteins in monitoring aging and heat-induced tolerance in germinating wheat and barley embryos. *Seed Science Research,* 1998, **8**, 91-98.
9. Ellis, R.H. and Roberts, E.H., The quantification of ageing and survival in orthodox seeds. *Seed Science and Technology,* 1981, **9**, 377-409.
10. Martens, H. and Næs, T., Multivariate Calibration, Wiley: New York, 1993, 35-72, and 97-100.
11. McCleary, B.V. and Glennie-Holmes, M., Enzymic quantification of (1-3),(1-4)-*B*-D-glucan in barley and malt. *Journal of the Institute of Brewing,* 1985, **91**, 285- 295.
12. Molina-Cano, J.L., Ramo, T., Ellis, R.P., Swanston, J.S., Bain, H., Uribe-Echeverria, T. and Perez-Vendrell, A.M., Effect of grain composition on water uptake by malting barley: a genetic and environmental study. *Journal of the Institute of Brewing,* 1995, **101 (2)**, 79-83.
13. Munck, L., Improvement of nutritional value in cereals. *Hereditas,* Berlingska Boktryckeriet: Lund, Sweden, 1972, **72**, 48-53.
14. Møller, B. and Munck, L., Barley seed vitality in relation to heat-susceptibility and resistance. The VII International Barley Genetics Symposium, University Extension Press, University of Saskatchewan, Saskatoon, Canada, 1996, 646-648.
15. Møller, B. and Munck, L., Seed vigour in relation to heat sensitivity and heat resistance in barley evaluated by multivariate data analysis. *Journal of the Institute of Brewing,* 2002, **108 (3)**, 286-293.
16. Riis, P. and Bang-Olsen, K., Breeding high vigour barley – a short cut to fast maltings. Proceedings of the European Brewing Convention Congress, Lisbon, IRL Press: Oxford, 1991, 101-108.
17. Spears, J.F., An introduction to seed vigour testing. In: Seed Vigour Testing, Proceedings from 24th Congress of the International Seed Testing Association, H.A. van de Venter Ed., Copenhagen, 1995, 1-9.

(Manuscript accepted for publication April 2002)

A New Germinative Classification Model of Barley for Prediction of Malt Quality Amplified by a Near Infrared Transmission Spectroscopy Calibration for Vigour "On Line" Both Implemented by Multivariate Data Analysis

Lars Munck[1] and Birthe Møller

ABSTRACT

J. Inst. Brew. 110(1), 3–17, 2004

A new germinative two-dimensional classification plot fully compatible to the current EBC analyses (EBC methods 3.5–3.7) is proposed for malting barley based on separate estimates for "vigour" (24 h germination) as abscissa with limits at 70% and 30% and for "viability" (72 h germination) as ordinate with limits at 98% and 92%. Early detection of germination by image analysis was improved by utilising the auto fluorescence of the root cap. The seven hierarchical germinative classes visualise the quality differences in a consistent way, ordering classes according to falling extract % and increasing wort β-glucan (mg/L).

It was surprising to discover that significant barley Near Infrared Transmission (NIT) spectroscopy based Partial Least Squares Regression prediction models for "vigour" and "viability" were obtained after removing the PLSR outliers. The majority of these were found to be low in vigour.

It was concluded after experimental validation that the physical-chemical structure of the seed, reflected by the correlation of the barley NIT spectral fingerprints to germination speed, is connected to the availability of substrate for germ growth. This is another aspect of the speed of malt modification.

An automated combination instrument for measuring physical-chemical and seed germination parameters is suggested for quality control and to establish an *on-line* NIT calibration network for integrated germinative and malting quality classification.

Key words: Germinative classification, malting barley, seed vigour, seed viability, malt quality, near infrared transmission spectroscopy.

The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.

[1] Corresponding author. E-mail: lmu@kvl.dk, bm@kvl.dk
Web page: www.models.kvl.dk

Publication no. G-2004-0317-204

INTRODUCTION

The need for more informative germinative malting barley analyses

Optimal germination performance is without a doubt the most important quality criterion for malting barley[5,11,14,16,40,41]. The industry and trade are dependent on reproducible and representative laboratory analyses for germinative capacity (GC%) and energy[37] (GE%) and for germination speed (GI-germination index[38,39]) as expressed in the methods 3.5–3.7 in EBC Analytica[3].

It is surprising that even today germination data do not seem fully integrated with malting data in barley quality evaluation, but are rather used as univariate pre-qualification criteria with respect to live (viable) seeds at 3–5 days germination with GC% and GE% methods. In the following we will, inspired by the results from multivariate pattern recognition data analysis (also called chemometrics[17,18,44]) of germination profiles, define the optimal practical criteria for vigour and viability. We will find that these criteria should basically represent two different dimensions in malting barley analysis, analogous to those of acceleration and mass in physics. This paper focuses on how to utilise germination information more effectively for prediction of malt quality by a new two way dimension germinative classification. It also aims at prediction of germinative data by calibration to automated instrumental analytical methods calibrated by chemometric models to speed up these very slow germination analyses.

MATERIALS AND METHODS

Two barley materials (I–II) are utilised, the results of which are published elsewhere[22,24] but are here recalculated and presented in a completely new form to support our case. The materials were:

I. The 17 samples of the malting barley Alexis[24] grown in different places in Europe in 1994 were collected and analysed for malting quality at Centre UdL-IRTA, Spain[24] by Dr. José Luis Molina-Cano. The same collection was also heat treated at our institute[24] to study artificial ageing[6,7] as a measure for vigour and to calculate the vigour potential (VP) according to Aastrup *et al.*[1].

II. A barley material of 42 samples[22] of the varieties Alexis, Blenheim and Meltan grown in Southern Scandinavia, collected in 1993–1996, was analysed for GE, seed physical-chemical analyses and malt quality after cold storage (7°C, 13.5% water) in 1999. Møller[22] publishes a detailed study on this material in this issue. The barley material[22] was collected under a number of years with different growing conditions in order to obtain a representative material spanning a wide variation in quality parameters.

Methods

Germination analyses

a. Germinative energy (GE) % was determined using the BRF method (EBC method 3.6.2)[3]. This method was used on 4 × 100 grains. GI and GH were calculated for all methods according to Riis and Bang Olsen[38] and Analytica-EBC (EBC method 3.7)[3].

b. Heat-treatment to determine vigour potential. This was determined for Material I by Møller *et al.*[24] according to Aastrup *et al.*[1].

c. A new method for detection of early germination by inspection under fluorescent light. In screening for vigour, reproducible early detection of chitting is fundamental. We have found that inspection under fluorescent light (366 nm) is useful, because the newly developed root caps display a characteristic blue fluorescence (Fig. 1), making them stand out against the seed background. The fluorescence is tentatively assigned to ferulic acid, which binds to the cell walls[29].

The image from each single kernel is degraded in transversal lines for an area in the middle and for both ends of the grain, and the average intensity in these areas is determined. The intensities are compared, and the image analysis concludes whether the grain is germinated or not, based on the average intensities.

We have preliminary tested two germination methods for determination of g%1; germination with hydrogen peroxide (GC[3]) and the BRF-method[3] (GE). The correlation for manually counting with and without using the light table (visual inspection) is quite good (H_2O_2: r = 0.99; BRF: r = 0.98). Comparing manual inspection with digital image analysis the GE BRF-method[3] (r = 0.99) is as precise as for visual inspection according to the illumination method by the light table, whereas for germination using H_2O_2, the correlation is much lower (r = 0.89). It appears that the hydrogen peroxide bleaches the root head, so the separation with image analysis between the emerging root cap and the rest of the grain is not as good as when using BRF (GE) germination.

It can be concluded that it is possible to use fluorescent light to detect the early chitted grains by image analysis in a GE screening. The image analysis should be optimised to detect the sprouts after 24 hours automatically. The experiments are preliminary. It should be possible to develop the method further to a prototype, e.g. combined in an apparatus, which also determines single seed form properties, light reflectance intensity and hardness as suggested by the work of Møller[22].

d. Physical, chemical and malt quality analyses. Thousand Kernel Weight (TKW), seed form parameters, light reflection intensity, NIT-spectroscopy and hardness were analysed according to Møller[22].

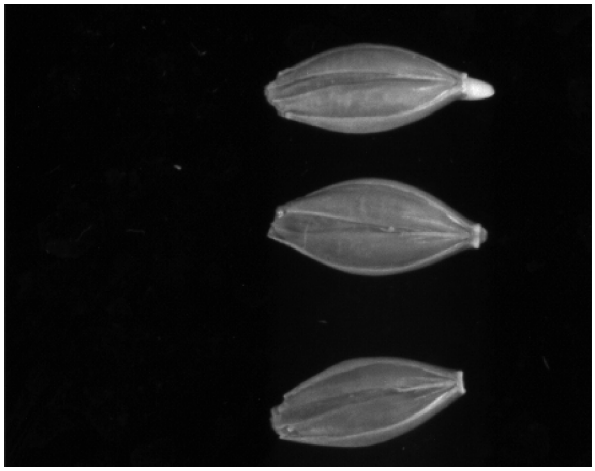


Fig. 1. Image of three kernels illuminated by fluorescence light after 24 hours of germination. The grain above is germinated; the two grains below are not germinated.

Extract and wort colour were determined on Material I at Centre UdL-IRTA[24]. The (1→3, 1→4)-β-glucan in barley and wort was analysed by the calcofluor method[2,30].

e. Multivariate data analysis. Principal Component Analysis (PCA) and Partial Least Squares regression (PLSR) were performed as described by Martens and Næs[17] and Martens and Martens[18] using the "Unscrambler" software from CAMO A/S, Trondheim, Norway. The principal components indicated as PC's in the PLSR analyses with the "Unscrambler" software are mathematically not identical with the PC's denoted in the PCA analysis. The importance of the **X** variables is evaluated using Jackknife validation proposed by Martens and Martens[18].

Abbreviations

BG	(1→3, 1→4)-β-glucan in barley
BGwort	(1→3, 1→4)-β-glucan in wort
C	Wort colour
EBC	European Brewery Convention
EXT	Extract
g%1–8	Germination percentage day 1–8
GC	Germination Capacity
GE	Germination Energy
GH	Germination Homogeneity
GI	Germination Index
HI	Hardness Index
NIT	Near Infrared Transmission
P	Protein
PC	Principal Component see comment in Materials and Methods **e**
PCA	Principal Component Analysis
PLSR	Partial Least Squares Regression
RE	Relative Error in percentage
RMSECV	Root Mean Square Error of Cross-Validation
ROUND	Kernel roundness
STEEP	% Water uptake after 24 hours of steep
TKW	Thousands Kernel Weight
VP	Vigour Potential
WIDTH	Width of kernels

RESULTS

Vigour and viability identified by multivariate data analysis as independent principal components from a set of barley germination data[24] (Material I)

In contrast to probability statistics used on germination data[6,7] that is based on distributional assumptions, it is possible by multivariate data analysis to classify the form of each individual barley germination curve (Fig. 2A) with a minimum of *a priori* assumptions. Initially we are thus leaving Material I (Table I) in the hands of a chemometric model – Principal Component Analysis (PCA) in order to characterise the relationship between samples (1–17) and the germination parameters (g%1–g%8) letting data speak for themselves.

The data (Table I) from 17 germination profiles from samples of the Alexis variety (1–8 days germination, Material I) are visualised in Fig. 2A. The PCA in Fig. 2B is a combined score/loading plot (biplot) featuring the different germination observation days 1–8 (g%1–g%8) as variables (loadings) marked in bold (**1–8**). Neighbouring sample score points have similar germination profiles. The position of the loadings can be interpreted so that sample scores characterising the whole germination profile for e.g. sample 1DK, 3DK, 7NE, 9CZ and 16D located near to day 1 (**1**) are fast germinating while samples 2DK, 11SU, 13SU and 14SU situated below in the opposite direction along the ordinate are slow germinators. It is obvious that the signs **1** (g%1) and **2** (g%2) above are distributed along the ordinate together with a cluster of **3–8** (g%3–g%8) further below with its members situated close to each other.

One of the foremost tasks of chemometric data analysis besides data reduction is to support or even extend the scientific language in an evaluation dialogue between the graphic data interface and prior knowledge. This is done in continuing the evaluation process of the PCA plot, with an interpretation plot to ascertain the character of the PC's. We may now mark each sample point in Fig. 2B with its values for g%1 (in bold) and g%3 in Fig. 2C. Three-day germination percentage (g%3) is taken as the most conveniently measured representative for the close loading cluster g%3–g%8. A glance at the thus labelled PCA bi-plot convinces us that there is a clear gradient in g%3 along the abscissa from left to right and for g%1 along the ordinate from below to above. It is now possible to interpret the meaning of the graphic representations PC1 and PC2 (principal components) of the PCA plot as "viability" (abscissa; g%3; germination "mass") and "vigour" (ordinate; g%1; germination "acceleration") respectively and to mark the axes accordingly (Fig. 2C). The orthogonal character of the PC's inherent in the PCA algorithm used above depicts their basic independence in variance set by the algorithm. The mean GI's of the samples in the four quadrants displayed in Fig. 2C also supports the idea that there is a vigour gradient along the ordinate. In this material there is a reasonable correlation (r = 0.92) between g%1 and GI (as compared to r = 0.73 for g%2, and r = 0.46 for g%3) thus supporting the notion that germination day 1 (g%1) can be used as a convenient estimate for germination velocity.

A simple two-dimensional plot with "vigour" as abscissa and "viability" as ordinate for the classification of malting barley in germination energy and capacity classes (Material I)

In interpreting the principal components of the previous PCA (Fig. 2B and Fig. 2C) of the EBC Alexis barley material grown in widely different locations in Europe, we generated a hypothesis that the information from the 1–8 day germination curves could be simplified into two parameters one for vigour and one for viability. This leads us to the simple two-dimensional plot in Fig. 3A, which represents a reduction of the original data material of germination profiles in Table I. Here we approximate vigour as g%1 (abscissa) and viability as g%3 (ordinate). In order to avoid confusion we put the designates for vigour and viability in Fig. 3A and in the following figures and text in quotation marks as "vigour" and "viability" to distinguish their character as estimates. In order to visualise effectively the range of measurement, g%1 is chosen as

Table I. Germination profile values 1–8 days (GE conditions EBC 3.6.2) for the 17 Alexis samples in Fig. 2. Germination index (GI) and germination homogeneity (GH) according to EBC 3.7. The letters in the sample notation denotes country. Vigour potential (VP) is calculated after heat treatment according to Aastrup *et al.*[1]. Background data from the paper of Møller *et al.,* 2002[24] previously unpublished. (Material I).

	1	2	3	4	7	8	GI	GH	VP
1DK	85.8	97.8	98.3	98.8	99.3	99.3	8.8	64.6	1.5
2DK	36.3	90.3	95.5	97.8	98.5	98.5	6.0	42.6	4.8
3DK	77.8	98.5	99.3	99.3	99.3	99.3	8.2	56.5	1.7
4E	72.5	85.5	88.3	89.0	93.3	94.3	8.3	52.2	1.1
5E	69.0	83.8	89.3	91.0	93.3	93.8	7.8	42.7	1.1
6E	49.5	80.0	84.5	85.5	88.8	89.0	6.8	40.4	2.3
7NL	82.5	98.0	98.5	98.5	99.0	99.0	8.6	61.3	3.3
8E	66.5	94.5	97.0	97.3	98.5	98.5	7.5	47.5	3.3
9CZ	86.0	99.3	99.3	99.3	99.3	99.3	8.8	66.0	2.6
10D	67.5	98.3	99.5	99.8	99.8	99.8	7.5	50.2	1.8
11SU	40.5	92.8	98.5	98.5	98.8	99.0	6.1	41.3	2.2
12SU	40.3	87.3	93.8	—	97.0	97.0	6.1	39.2	2.6
13SU	37.3	95.0	98.5	—	99.0	99.0	6.0	45.6	2.8
14SU	40.3	87.8	97.0	—	99.3	99.3	5.9	36.1	2.3
15SU	51.5	93.8	97.0	—	98.8	98.8	6.6	43.7	2.0
16D	85.5	96.3	97.0	97.0	97.3	97.3	8.9	64.5	2.0
17D	61.0	93.5	96.0	96.0	96.8	97.3	7.2	46.1	1.8

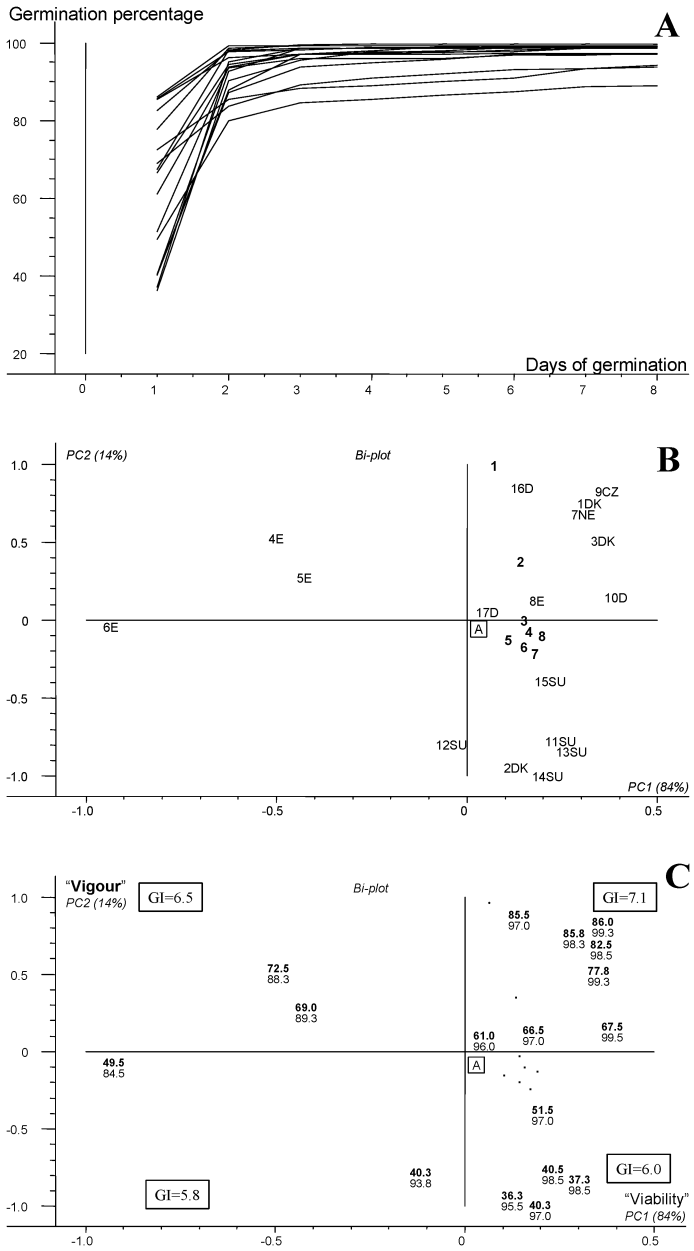


Fig. 2. Multivariate evaluation of germination profiles g%1–g%8 (see Table I[24]) for 17 untreated, non-dormant Alexis barley samples grown in EBC trials in Europe in 1994. **A.** Germination profiles of the 17 samples. **B.** Principal component analysis (PC1 (abscissa); PC2 (ordinate)). Biplot of the germination profiles for the 17 samples no 1–17. Letters denote country symbols. Figures in bold are loadings (variables) g%1 (**1**)–g%8 (**8**). **C.** Same PCA as B but with identification of each sample position by figures for g%1 in **bold** (estimate for "vigour") and g%3 normal text (estimate for "viability"). GI = Germination index 3 days mean for each quadrant. See text for discussion.

the abscissa (*x*-axis) because it has a wider range compared to the ordinate (*y*-axis) g%3, keeping in mind that the human vision can more precisely evaluate horizontal patterns compared to vertical. We will thus focus on "vigour".

As first guidelines for the germinative classification based on GE[3] analysis conditions, we tentatively set the demarcation lines for "viability" at 95% and for "vigour" 70%, as outlined in Fig. 3A. It is clearly seen that the simple classification plot (Fig. 3A) differentiates the samples equally well as the PCA-model with the same material in Fig. 2B. In order to evaluate the classification we display the malt analyses of the quartile classes in Fig. 3A in Table II. There is a clear tendency in reduction of extract and increase in barley β-glucan, protein and wort colour from classes 1:1 and 1:2 to 2:1 and 2:2. Class 1:1

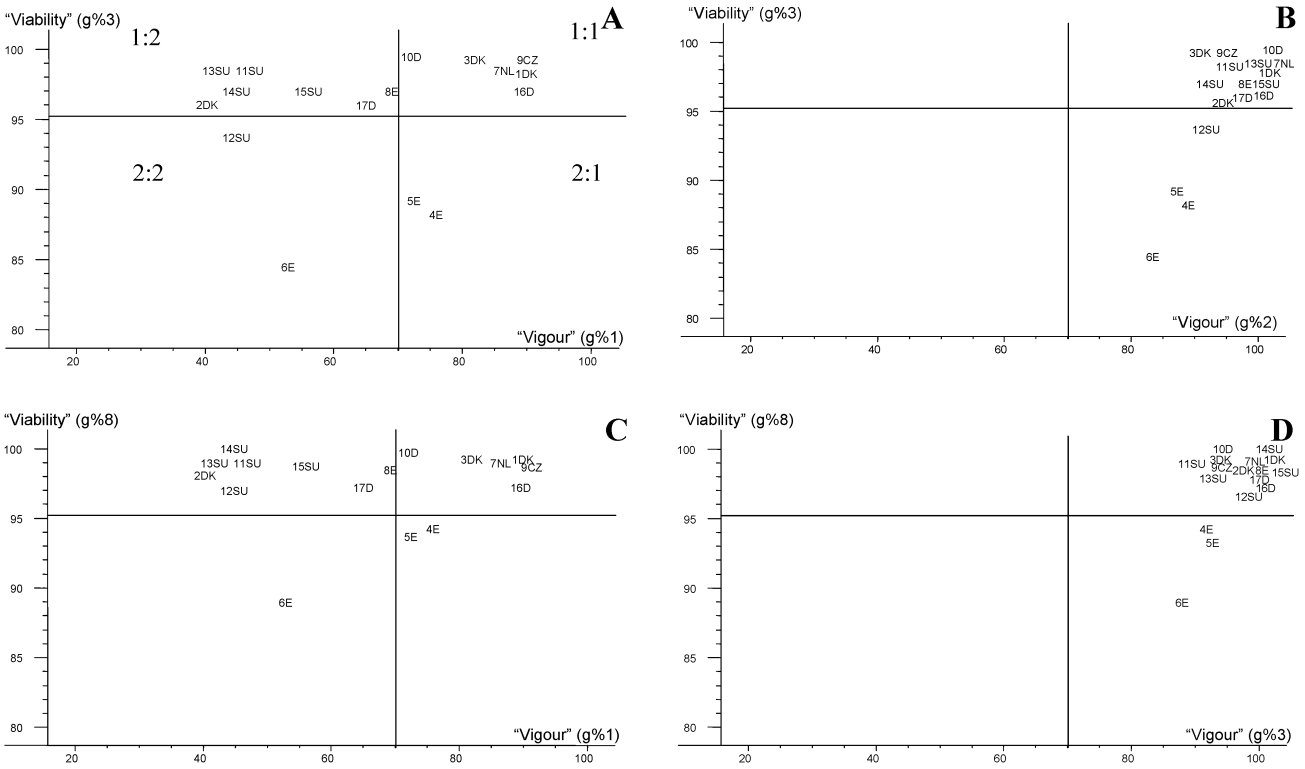


Fig. 3. Germinative classification plots of the Alexis samples[24] from Fig. 1 (Material I). **A.** "Vigour"(abscissa) defined as g%1, "viability" (ordinate) as g%3. For barley and malt analyses see Table II. **B.** "Vigour" g%2 (abscissa); "viability"(ordinate) g%3. **C.** "Vigour" g%1 (abscissa); "viability" g%8 (ordinate). **D.** "Vigour" g%3 (abscissa); "viability" g%8 (ordinate). For discussion see text.

Table II. Evaluation of the classes in the germinative classification in Fig. 3A (Material I, $n = 17$)[4].

	n	"Vigour" (g%1)	"Viability" (g%3)	GI	GH	TKW	Ext	BGbarley	P	C
1:1	6	80.8 ± 7.2	98.6 ± 0.9	8.4 ± 0.5	60.5 ± 6.1	45.3 ± 5.6	81.1 ± 1.7	3.6 ± 0.2	9.4 ± 0.6	3.2 ± 1.9
1:2	7	46.0 ± 12.2	97.1 ± 1.1	6.3 ± 0.6	43.3 ± 3.8	43.0 ± 8.1	80.0 ± 1.5	4.0 ± 0.3	10.6 ± 1.6	2.7 ± 0.7
2:1	2	70.8 ± 2.5	88.8 ± 0.7	6.5 ± 0.2	47.5 ± 6.8	—	80.1 ± 1.0	4.3 ± 0.0	14.0 ± 0.1	9.2 ± 1.0
2:2	2	44.9 ± 6.5	89.1 ± 6.5	5.8 ± 0.2	42.4 ± 0.6	42.4 ± 0.6	78.6 ± 0.2	4.7 ± 0.4	13.8 ± 0.1	3.3 ± 1.1

has the highest level of "vigour", "viability", GI as well as germination homogeneity (GH). We will now comment on how different assignments of "vigour" and "viability" influences the resolution of the classification of these samples.

We first judge g%2 as a candidate for "vigour" in Fig. 3B. We note from the PCA biplot in Fig. 2B that its loadings (**2**) is positioned between the sign for g%1 (**1**) earlier assigned for "vigour" and the tight cluster (**3–8**) of g%3–g%8 indicative of "viability". The g%2 trait has thus confounding information for both "vigour" and "viability". Using g%2 for "vigour" and g%3 for "viability" in a classification in Fig. 3B thus greatly narrows the span of "vigour" compared to the assignment of g%1 in Fig. 3A.

Using the conservative estimate for "viability" g%8 in combination with g%1 for "vigour" (Fig. 3C) does not substantially change the picture from our earlier comparison of "vigour" g%1 and "viability" g%3 in Fig. 3A, other than being less discriminative for the outlier 12SU. By applying g%8 for "viability" the samples move on the average 2% upwards along the viability axis (ordinate) compared to g%3 (Fig. 3A). Thus g%3 is for analytical time reasons the most practical estimate for "viability".

The question is now to what extent the g%3 trait carries information regarding vigour when we plot it as an estimate for "vigour" against the conservative viability estimate g%8? Not very much as seen by the tight cluster of the whole material in Fig. 3D compared with the great differentiation power of our comparison g%1 versus g%3 in Fig. 3A. The extreme outlier samples 4E, 5E and 6E are however still classified as outliers below the 95% demarcation line for "viability".

We thus conclude that the EBC methods[3] featuring 3 and 5 days of germination are overwhelmingly indicative for viability. We claim that in a classification plot defined as "viability" g%3 (which almost all in the malting industry seem to agree on), a differentiating estimate for "vigour" such as one-day germination as proposed in the diagram (Fig. 3A) would be a useful tool for the industry. In the following we will further test this option while discussing possibilities of improving the precision of the method.

A proposal for a two-dimensional germinative classification for vigour and viability demonstrated on a barley material with highly diverse germination and malting performance (Material II)

In order to support a proposal for a classification system using g%1 and g%3 as estimates for vigour and viability respectively we now select a part of a material of barley of varying composition and malting quality (Material II), described in detail by Møller[22]. The 42 barley samples were harvested in 1993–96, cold stored at 7°C and analysed in 1999. The classification plot (GE conditions) appears as an overview in Fig. 4A and is enlarged with regard to the malting barley classes in Fig. 4B. Here we have introduced two levels of both "viability" g%3 (92% and 98%) and "vigour" g%1 (70% and 30%). We thus arrive at a tentative classification system of six malting barley classes 1.1, 1.2, 1.3 and 2.1, 2.2, 2.3 as well as a feed barley class 3.0 (Fig. 4A).

Table III shows mean values and standard deviations for barley and malt quality for these classes. In order to trace possible gradients within the classes, they are divided into sub clusters **a** and **b**, as displayed in Fig. 4B and in Table III. Values for extreme samples are also given. Now we use the germination, chemical and malting analyses from Table III to evaluate the classification of the barley samples in Fig. 4A and Fig. 4B.

The material has a large variation in "vigour", "viability", GI and GH due to variety and year of production. It is clearly seen that the "vigour" component is complementary to the germinative energy component "viability" in differentiating the whole material with regard to extract % and to an even greater degree with regard to BGwort, revealing the dependence of cytolytic activity in the malt on a swift and complete germination. The mean values of the malting barley classes reveal clear gradients in these important quality criteria. First, from the right to the left along the "vigour" axis from 79.0 to 17.8 g%1; (class 1: extract 83.2–80.6%, BGwort 180.4–257.0; class 2: extract 79.8–79.1%, BGwort 199.2–297.4) and second, from above to below along the "viability" g%3 axis from 99.2 to 95.9 g%3; (class 1.1 versus 2.1 extract 83.2–79.8%, BGwort 180.4–199.2 and for class 1.2 versus 2.2; extract 81.6–79.1% and BGwort 202.0–297.4).

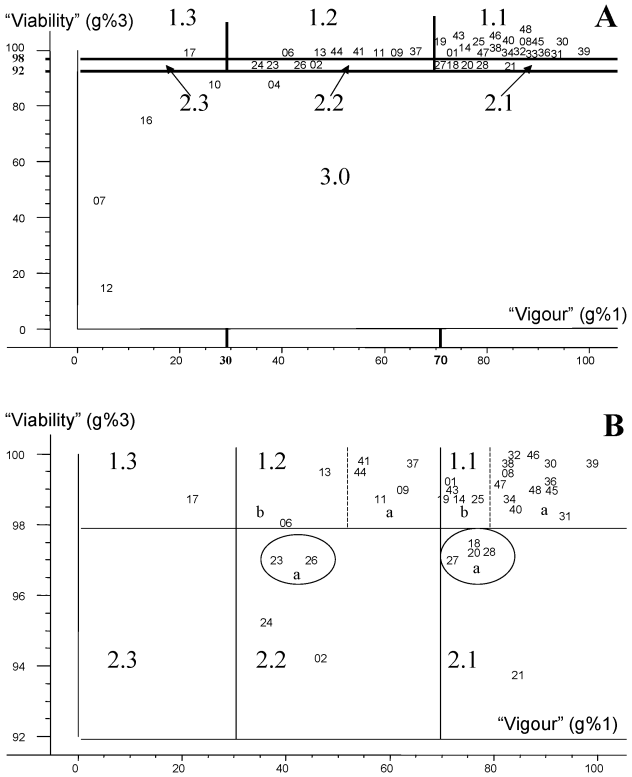


Fig. 4. Germinative classification for the malting barley reference material (n = 42) (Material II). "Vigour" g%1 (abscissa); "viability" g%3 (ordinate). For discussion of the limits and classes see text. For barley and malt analyses see Table III. **A.** Overview. **B.** Enlargement of the malting grade classes in the plot with "viability" ≥ 92%.

Table III. Germination and barley and malt analyses for the 42 reference barley samples classified in Fig. 4. Mean and standard deviations of the different classes including values from extreme samples (Material II).

Class	n	"Vigour" (g%1)	"Viability" (g%3)	GI (g1–g3)	GH (g1–g3)	TKW	Extract (%)	BGwort (mg/l)	BG (%)	P (%)
1:1	**16**	**79.0 ± 8.6**	**99.2 ± 0.5**	**8.3 ± 0.6**	**56.3 ± 10.0**	**43.6 ± 2.1**	**83.2 ± 1.8**	**180.4 ± 39.7**	**4.1 ± 0.3**	**9.3 ± 0.5**
a	11	83.7 ± 5.5	99.4 ± 0.6	8.6 ± 0.5	59.3 ± 9.3	44.3 ± 1.8	83.9 ± 1.1	171.4 ± 37.1	4.0 ± 0.3	9.2 ± 0.4
b	5	68.8 ± 3.2	98.9 ± 0.2	7.6 ± 0.2	49.7 ± 8.6	42.1 ± 1.9	81.6 ± 2.1	200.2 ± 42.0	4.3 ± 0.3	9.5 ± 0.6
1:2	**7**	**50.3 ± 8.5**	**99.2 ± 0.7**	**6.6 ± 0.4**	**54.5 ± 7.6**	**41.1 ± 3.9**	**81.6 ± 3.0**	**202.0 ± 81.8**	**3.9 ± 0.3**	**10.0 ± 1.2**
a	5	54.7 ± 4.5	99.4 ± 0.5	6.8 ± 0.3	57.3 ± 7.2	42.3 ± 3.9	82.5 ± 2.9	176.2 ± 37.2	3.9 ± 0.3	9.4 ± 0.7
b	2	39.5 ± 5.3	98.8 ± 1.1	6.2 ± 0.3	47.7 ± 1.9	38.2 ± 2.1	79.3 ± 2.6	266.6 ± 151.6	3.7 ± 0.0	11.4 ± 1.4
1:3	**1**	**17.8**	**98.8**	**5.4**	**54.5**	**35.3**	**80.6**	**257.0**	**3.9**	**10.2**
2:1	**5**	**73.1 ± 5.3**	**96.6 ± 1.6**	**8.0 ± 0.4**	**54.8 ± 4.1**	**37.1 ± 1.8**	**79.8 ± 5.4**	**199.2 ± 52.2**	**4.0 ± 0.5**	**10.2 ± 0.5**
a	4	71.3 ± 3.8	97.3 ± 0.2	7.8 ± 0.2	53.1 ± 2.4	37.3 ± 2.0	79.1 ± 5.9	209.7 ± 53.9	4.2 ± 0.3	10.0 ± 0.5
no21	1	80.5	93.8	8.6	61.2	36.6	82.9	157.2	3.2	10.6
2:2	**4**	**37.3 ± 5.1**	**95.9 ± 1.4**	**6.1 ± 0.3**	**46.7 ± 1.1**	**40.7 ± 1.6**	**79.1 ± 1.9**	**297.4 ± 53.3**	**4.1 ± 0.2**	**12.7 ± 0.8**
a	2	37.4 ± 4.8	97.0 ± 0.0	6.1 ± 0.2	47.1 ± 1.7	40.0 ± 0.5	80.1 ± 2.6	261.1 ± 55.7	4.0 ± 0.1	12.1 ± 0.3
no2	1	42.5	94.3	6.4	45.9	43.0	78.7	341.7	4.3	13.7
no24	1	32.0	95.3	5.9	46.4	39.9	77.5	326.0	4.0	12.9
3:0	**5**	**13.6 ± 14.6**	**63.0 ± 31.8**	**4.8 ± 0.9**	**42.6 ± 5.8**	**40.0 ± 1.9**	**70.1 ± 7.6**	**382.2 ± 94.4**	**3.8 ± 0.2**	**12.1 ± 1.5**
no4	1	34.3	88.0	6.0	41.7	42.6	78.9	330.5	4.2	14.5
no10	1	22.5	90.0	5.3	39.5	37.6	76.3	309.1	3.8	11.4
no16	1	9.3	75.3	4.6	37.3	39.8	61.2	321.3	3.6	11.0
no7	1	0.3	46.5	3.7	52.3	40.8	64.1	532.8	3.7	12.3
no12	1	1.5	15.3	4.6	42.2	39.2	70.0	417.3	3.9	11.1

Table IVA. PLSR predictions of Extract and BGwort from germination parameters g%1, g%2 and g%3 for Material II (n = 42).

	r	RMSECV	PC*	RE	n	Sign. variables
Extract	0.85	2.79	3	11.43	42	non
BGwort	0.80	50.92	1	12.42	42	g%1, g%2, g%3

*Minimum value of residual validation variance

Table IVB. NIT (1. der.) PLSR prediction of germination data for the same sample set as in Table IVA. Samples with low viability g%3 GE conditions (<92%) = underlined, medium viability (92–98%) = **bold**, viability >98% = normal text.

	Step**	r	RMSECV	PC*	RE	n	Removed outliers in each step
g%1 GE	0	0.74	16.81	4	17.8	42	
	I	0.77	14.94	4	15.7	41	A12
	II	0.80	13.55	3	14.3	38	M16, **A20**, **A27**
g%3 GE	0	0.31	14.42	1	17.0	42	
	I	0.68	1.88	1	15.7	39	M07, A12, M16
	II	0.80	0.89	3	3.4	37	B04, A10

*At minimum value of residual validation variance
**Step of outlier selection from an influence plot

As seen from Table III, there are also clear gradients within the classes in extract and BGwort. An exception is subclass 1.2a containing five samples which have slightly higher extract and lower wort β-glucan values than the preceding higher subclass 1:1b. These differences are, however, not significant. Sample 21, which is a negative outlier in class 2.1, has in spite of its low 93.8% "viability" a much better malting quality than its position would indicate. This is partly due to its low β-glucan content in the barley (3.2% compared to the mean of 4.2% of subclass 2a (n = 4)) leading to low BGwort (157.2) and to its high "vigour" (80.5%) leading to high extract (82.5%). Note that the samples 02 and 24 from the inferior subclass 2.2 with higher "viability" (94.2 and 95.3%) than sample 21, but with lower "vigour" (42.5 and 32.0%), are much lower compared to sample 21 in extract (78.7 and 77.5%) and much higher in BGwort (341.7 and 326.0). The "vigour" component thus brings an essential differentiating element into the germinative classification plot.

The feed barley class 3.0 is clearly unsatisfactory for malting, as seen from the figures from the individual samples with mean figures of 70.1% for extract and 382.2 for BGwort. GH (Table III) has, as in the example discussed previously (Table II), a tendency to be positively correlated to "vigour" (g%1). GI has a high correlation of r = 0.99 in this material with g%1. It is concluded that the proposed germinative classification system with the barley material tested here is highly sensitive for predicting and discriminating the levels of extract (%) and BGwort (mg/L), which are the central parameters in the barley malt quality complex.

Defining the theoretical basis for the vigour and viability concepts by multivariate analysis explaining why vigour can be predicted by NIT spectroscopy (Material II)

Multivariate correlation between the germination profile and malt quality variables. We will now briefly explore the multivariate connections between the variables in our previous classification data set (Table III, Fig. 4A and Fig. 4B) with the aim to generate new hypotheses inspired by finding structures in data, interpreted by experience in brewing science. Partial Least Squares Regression (PLSR) analysis[12,17] is a two-way matrix (**X**, **y**) correlation model based on a decomposition principle comparable to the one block (**X**) method PCA. It is here possible to judge the significance of the variables and their contribution in the total correlation[18].

In understanding how the germinative classification system works we first asked ourselves how the germination profile (g%1, g%2, g%3) as **X** is related to each of the parameters extract% and BGwort (mg/L) as **y**. Table IVA shows a significant positive PLSR correlation between extract% and the whole germination profile of r = 0.85, however, none of the individual germination profile parameters are significant. An analogous PLSR correlation with BGwort (mg/L) displays a significant negative correlation coefficient of r = 0.80 where a high BGwort (mg/L) implies low germination values. Here all three germination parameters give significant individual contributions. It is concluded as expected from the evaluation of the germinative classification plot in Fig. 4 and Table III that the germination profile seen as a multivariate whole is highly informative regarding important malt quality parameters.

Obviously there are two major functional factors, which have to be taken into consideration when breaking down the endosperm to extract and BGwort:

1. The physiological factor of germ viability, i.e. plant hormone and enzyme dissemination into the endosperm for enzyme induction through the aleurone tissue necessary for malt modification.
2. The structural factor, i.e. the physical-chemical endosperm structure and composition of importance for resistance to malt modification and for the remaining β-glucan in wort.

Predicting germination variables from NIT spectroscopy profiles by PLSR. We did not expect that our two physical-chemical sets of analyses would be able directly to sense physiological-biochemical changes related to viability in the germ constituting less than 5% of the entire seed weight. At an early stage in our investigation

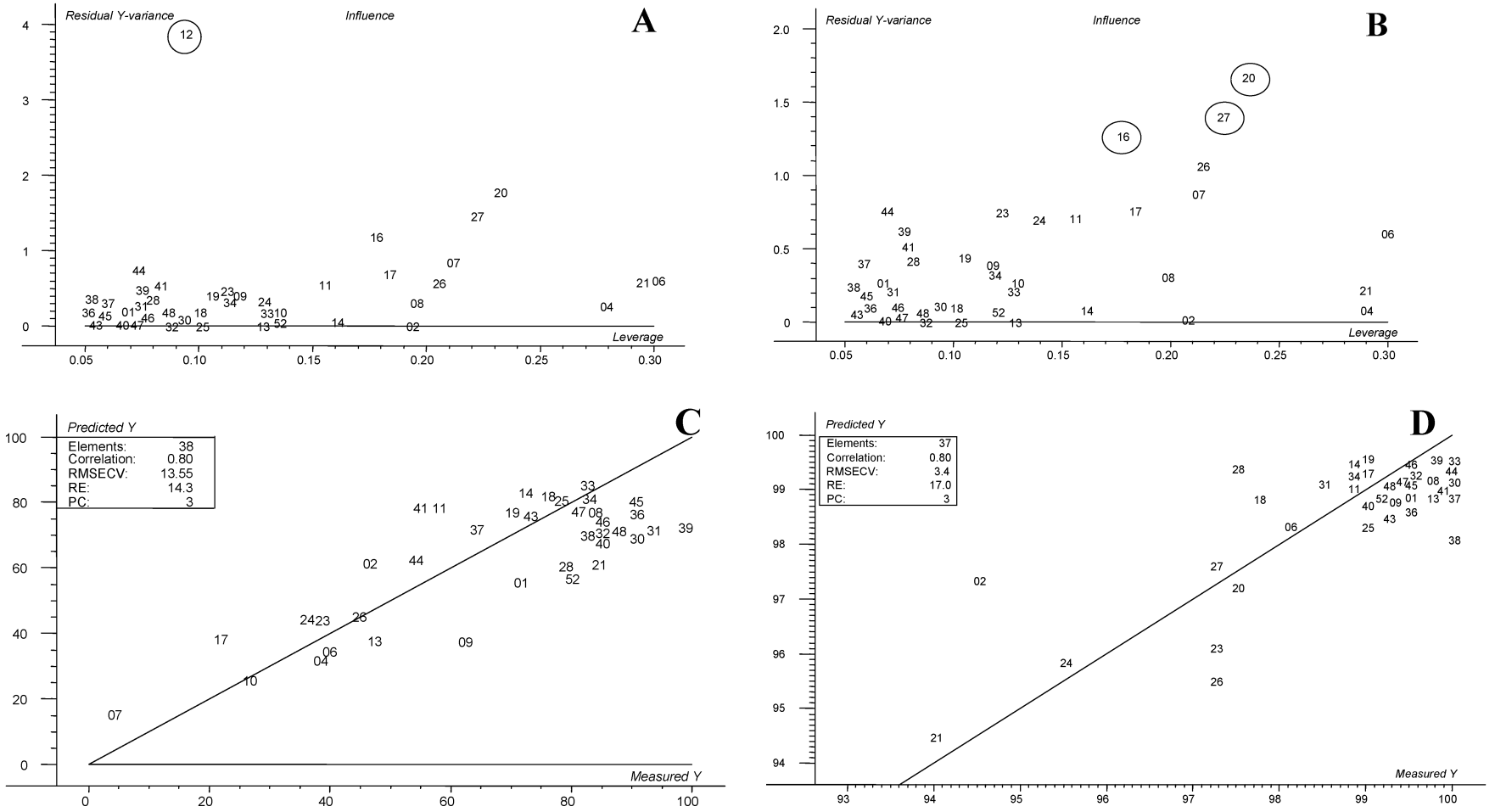


Fig. 5. NIT spectroscopy predictions for "vigour" g%1 and "viability" g%3 of the barley material (n = 42) classified in Fig. 4. Letters and numbers denote samples. Outliers are encircled and removed in later step. See discussion in text. **A.** Influence plot (step 0) for outliers in prediction of g%1 from NIT. One outlier will be removed. **B.** Influence plot (step I) for outliers in predic tion of g%1 from NIT. Three outliers will be removed. **C.** g%1 PLSR correlation plot, step II (n = 38). Three PC's for g%1. Four outliers in all are removed. See discussion in text. **D.** g%3 PLSR correlation plot, step II (n = 37). Seven outliers in all are removed. See discussion in text.

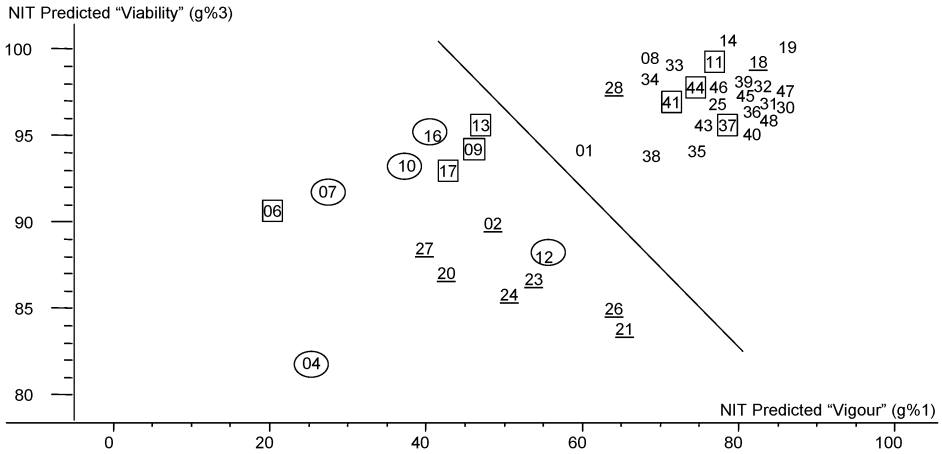


Fig. 6. Germinative classification plot g%1 "vigour" (abscissa) and g%3 "viability" (ordinate) values predicted by PLSR correlations with NIT. See discussion in text.

with material II[22] we were surprised that the non-destructive NIT analysis on whole kernels, which should reflect the physics and chemistry of the seed (functional factor 2), could make reasonable PLSR predictions of "vigour" g%1 – a character which we first thought belonged to category 1 (the physiological factor) defined above. Thus we could get reasonable predictions of "vigour" of r = 0.74 (four PC's, RE = 17.8) with all samples (n = 42) as seen in Table IVB.

In multivariate analysis it is important to select and define the nature of outliers. This is done by consulting a PLSR influence plot[17,18] of the g%1 PLSR correlation to NIT data above as shown in Fig. 5A. Here sample 12 is clearly identified as an outlier. Removing this sample the correlation coefficient is improved to r = 0.77 (Table IVB, four PC's, RE = 15.7) giving a new influence plot (Fig. 5B) where three new outliers were found 16, 20 and 27. A recalculation (Table IVB) improves the correlation to r = 0.80 (Table IVB, three PC's, RE = 14.3) for the 38 samples as seen in the correlation plot in Fig. 5C.

A similar stepwise procedure of outlier identification by influence plots is presented for g%3 "vigour" PLSR correlation to NIT data in Table IVB. Here outlier removal leads to a high increase in the correlation coefficient from r = 0.31 (one PC, RE = 17.0) to r = 0.80 (three PC's, RE = 14.3) in two steps. The last correlation is presented as a plot in Fig. 5D. Five outliers 07, 12, 16, 04 and 10 were identified. When consulting the germinative classification plot in Fig. 4A; 07, 12, 16, 04 and 10 are classified as class 3.0 feed barley while the outlier samples 20 and 27 from Fig. 5C belong to class 2:1 medium grade malting barley.

It is concluded that samples with low "viability" are outliers in the PLSR NIT correlations for prediction of "vigour" and even more marked in those for "viability".

It should be pointed out that the outliers with low "viability" g%3 identified by the PLSR influence plots above are deviates in **y** (g%3) and not in **X** (NIT). We have checked that there are no improvements in the correlation coefficient by removal of **X** outliers. These do not show low "viability". This implies that "viability" cannot be predicted in unknown samples by NIT spectroscopy. This is in accordance with our initial hypothesis that physical-chemical analyses like NIT should not be able directly to trace physiological factors represented by low germ viability (e.g. dead and slowly geminating kernels). If information for "viability" had been carried in NIT data – this would have been possible. A separate method for "viability" such as g%3 (GE) or theoretically more correct g%8 or by the tetrazolium staining test for live germs (GC^3) is thus needed as a supplement to "vigour" g%1 to obtain a reliable germinative classification plot. If one wants to evaluate the theoretical aspects of germ "viability" comparing different kernels under identical substrate conditions, one has probably to excise the germ and resort to embryo culture *in vitro*.

In Fig. 6 we surprisingly see that we can, to some extent, repeat our germination classification plot from Fig. 4A by using the original NIT PLSR predictions mentioned above (no outliers removed) with "vigour" g%1 and "viability" g%3 for each of all 42 samples. The demarcation line between the two classes – high quality above to the right and low quality below to the left – was tentatively drawn to fit in between the main clusters in the plot without any *a priori* judgement, regarding individual sample data. The NIT germinative classification plot thus obtained (Fig. 6) is successful in making a complete separation of the extreme classes 3.0 (feed quality, encircled) and 1.1 (highest malt quality, normal figures). The low malting barley classes 2.1–2.3 are divided with seven samples (underlined) to the left in the plot and two outliers in the high germination quality cluster to the right. The low vigour part of the premium class 1, – classes 1.2 and 1.3 (squared) are divided with four samples in each of the two clusters. If the dividing line between good and low malting quality is drawn between classes 1.3 and 2.1, six samples out of 42 were thus wrongly classified. This preliminary coarse NIT classification with a limited material is a promising indication and demonstrates that further investigations to develop a germinative screening method based on NIT with a much larger calibration material should be profitable.

Checking the results from NIT prediction of germinative parameters by a separate set of data. In our

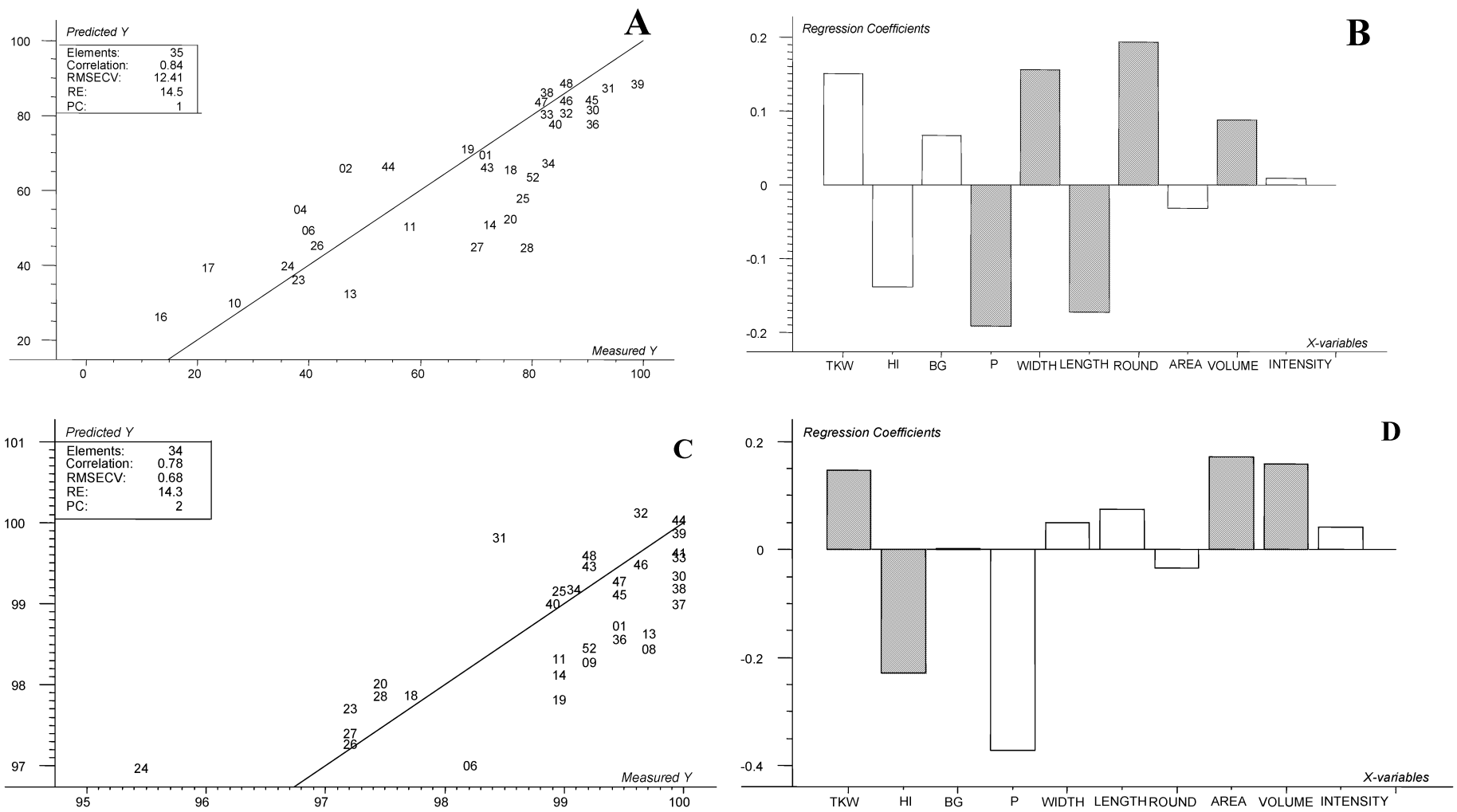


Fig. 7. PLSR correlations and Jacknife regression coefficient plots for the barley material in Fig. 4 and Table III. X = 10 physical-chemical barley analyses (see abbreviations, text and figures). **A.** PLSR correlation "vigour" g%1(GE conditions) = y. **B.** Regression coefficient plot for 7A. Significant variables shadowed. **C.** PLSR correlation "viability" g%3 (GE conditions) = y. **D.** Regression coefficient plot for 7C. Significant variables shadowed.

multivariate approach to confirm the hypothesis of a major structural impact on "vigour" founded on our NIT PLSR predictions discussed above, we made a separate analysis on the same barley samples using a separate set of ten physical-chemical analyses[22] (Fig. 7).

In Fig. 7A–D we have included the ten physical-chemical criteria manifest in the intact ungerminated seed as a representation for structure which is correlated in a PLSR analysis with "vigour" g%1 and "viability" g%3 (GE conditions) with Jack-knife validation. The ten analyses are two chemical analyses (BGbarley and protein analysed on bulk) and eight physical analyses: hardness, seed weight and six automatic seed analyses from the automatic imaging instrument: width, length, round, area, volume and light reflectance intensity. The destructive seed hardness index (HI) is analysed on a separate instrument. In order to eliminate range and numerical differences in parameters, which highly influence the correlation coefficients, we regularly employ scaling as a pre-treatment to data. We obtain, as with NIT spectroscopy, a significant PLSR correlation (r = 0.73, one PC, RE = 18.1) with g%1 ("vigour") as **y** and the ten physical-chemical parameters as **X**. When inspecting the influence plots in two cycles we note two "viability" g%3 outlier samples 07 and 12, which are below 92% (Fig. 7A) and one outlier 21 in the medium viability category below 98%. Four outliers; samples 08, 09, 37 and 41 cannot be explained by deviating "viability" properties. In Fig. 7A we have displayed the correlation plots of the barley material when we have eliminated the seven outliers. The correlation coefficient has increased to r = 0.84 (one PC, RE = 14.5) with five significant variables; P, round, length, width and volume (Fig. 7B). Using four PC's in this outlier-corrected material (influence plot not presented here) we obtain a correlation coefficient of r = 0.94 and an error of RE of 9.1.

The correlation of "viability" g%3 to the physical-chemical parameters is much lower than for "vigour" in the whole material (r = 0.39 versus r = 0.73, one PC) as with the NIT analysis. When inspecting the influence plot (not presented here) we identify eight outliers, six having low and one medium low "viability" and the last outlier with medium "viability" and high "vigour" (samples 02, 04, 07, 10, 12, 16, 17 and 21). The correlation plot without the eight outliers is presented in Fig. 7C. It has a distinctly lower correlation coefficient of r = 0.66 (RE = 22.7) compared to r = 0.84 (RE = 14.5) for "vigour" in Fig. 7A. The four significant variables (Fig. 7D) are HI, area, volume and TKW.

Connecting information to explain the physical-chemical basis of germination. Møller[22] explained the ability of NIT to predict "vigour" g%1 by the high individual correlation predictions of NIT to each of the majority of the ten physical-chemical parameters in this barley material. Extract% and BGwort could also be predicted by NIT and by the ten physical-chemical parameters[22]. Most of the outliers in the latter correlations were also found to be samples with low viability. The tight connection between the malt parameters and the germination profile is demonstrated in Table IVA.

It is clear that the multivariate correlation's demonstrated above to a great extent explain the predictive qualities of the "vigour" g%1–"viability" g%3 germinative classification plots, how they relate to critical quality criteria and why "vigour" g%1 together with germinative energy "viability" g%3, should be integrated in malting barley evaluation, as an essential complement to the barley and malt analyses.

With our strategy of focusing on the structural factor by PLSR correlation's to NIT and to the ten physical-chemical variables, identifying the physiological ("viability") nature of the outliers, we reach the surprising conclusion that germination speed "vigour", g%1 in this investigation, has a much more pronounced structural component than physiological within the range of viability which is characteristic for malting barley. The g%3 characteristic also reflects to some degree seed structure, but with a much lower correlation to the structural parameters than g%1 and with a larger number of low viability outliers. It seems that the NIT correlations are more robust than those of the set of ten physical-chemical parameters indicating that NIT data should represent the most complete physical-chemical fingerprint of the two screening methods.

We thus arrive at the important conclusion that the structural physical-chemical factor is the main determinate for vigour, defined as the early growth rate of the emerging plantlet in barley of malting grade. We interpret these preliminary results as follows. The substrate availability for the germ is of importance for fast sprouting and is related to the function of how to unlock the complex physical and chemical structure of the food store – the endosperm. This function should also be identical with the aims of the maltster to obtain a fast malt modification in dissolving cell walls and in enzyme spreading in the endosperm. Fast germination, i.e. high "vigour" g%1, should therefore be operative for the maltsters as an indicator of an efficient malt modification representing the structural factor related to physics and chemistry and to extract and BGwort performance respectively.

Thus the structural physical-chemical factor becomes limiting for seed vigour and malting and brewing performance of the barley samples that fulfil the classic germination energy qualification limit of 92%.

DISCUSSION

Classification systems for malting barley quality as tools in trade, industry, plant breeding and research

We have two sources of inspiration for our suggestion to upgrade the germinative energy (EBC 3.6)[3] and capacity (EBC 3.5)[3] concepts with a supplementary classification system as visualised in Fig. 4A and Fig. 4B. The first is the suggestion given by the PCA algorithm (Fig. 2B) to arrange the material (Table I) into two separate basically independent representations of vigour and viability (Fig. 2C). The second is the fact that maltsters and brewing scientists[11,19] in practice when considering GI, as an expression for vigour must always check this information against viability expressed as GE% for 3 or 5 days. Fig. 4A and Fig. 4B represent an informative visualisation of this reflective process, which is dramatically simple.

We have found a high correlation between GI and g%1 in the two barley materials presented here. GI could well be used instead of g%1 as an estimate for vigour. We suggest, however, that g%1 should be preferred. First, because it is much more responsive than GI with regard to the vigour trait. Secondly, because day 1 germination information from either the germinative capacity (EBC 3.5.2)[22] or energy (EBC 3.6) methods[3] could be combined with the fast tetrazolium (EBC 3.5.1) screening method for viability in a germinative classification plot for a diagnosis within 24 hours according to the layout in Fig. 4. Further studies should be made to check the reproducibility of the g%1 estimation for "vigour" in comparison with GI (EBC 3.7) testing the fluorescence tool in non–hydrogen peroxide germination methods for improvement of automated counting at an early germination stage.

The "vigour" and "viability" classification test may also be used in micro and pilot maltings and in full-scale trials as an essential description of germinative "acceleration" and "mass" inherent in the germination quality trait of fundamental importance in industrial malting barley utilisation and economical evaluation. The promising use of NIT screening for an indicative germinative classification (Fig. 6) *on-line* alone or in combination with the Tetrazolium test for "viability" will be discussed in the Conclusions.

The necessary mathematical tools for evaluation of complex covariate data sets

Pattern recognition multivariate analysis is indispensable when evaluating large horizontal data sets with many strongly correlated (covariate) variables such as germination profiles and NIT spectra. If such data has high quality, the fingerprint of variables will in a unique way characterise each sample. Such a characterisation is not possible in classic statistics[17,18,26]. This property makes it also possible for chemometrics to handle incomplete and unbalanced data sets, which are more difficult to handle with classic methods[17,18].

The graphic interface of PCA and PLSR score plots facilitates an interactive overview of the complex data set combined with an in depth scrutiny of the detailed connections between samples and variables as demonstrated in our examples. The chemometric analysis is hypothesis generating letting the data set at first speak for itself with a minimum of *a priori* hypotheses. Validation by prior knowledge and further chemical analyses are dynamically included after the first evaluation. Statistical validation of errors in PLSR is made by data experimentation within the data set by cross validation or by comparing a calibration set with a test set sampled from the same population. The conclusions are supported by evaluating in parallel two separate materials e.g. Material I and II and by comparing separate sets of analyses e.g. the PLSR correlation of g%1 "vigour" to NIT-spectroscopy on one hand and to the set of ten seed physical-chemical analyses on the other as demonstrated above.

Seed scientists such as Ellis and Roberts[6,7,10] used classic probability calculus and curve fitting when evaluating germination profiles and seed deaths with storage time by accelerated ageing by heat. A seed sample, which could resist heat, was supposed to have a high vigour[4]. Because of the hard assumptions of their statistical models they could only use one germination variable (e.g. g%8) at a time to study resistance in the decay of viability as a marker for vigour.

In malting research Favier[8] and Woods *et al.*[43] used curve fitting to model the germination rate of dormant seeds during storage and Aastrup *et al.*[1] utilised Ellis and Roberts[6,7] heat stress probit model for calculating vigour potential (VP). These scientists analysed a limited number of samples and did not test their models by a separate test set. The VP's calculated by probit analysis could not predict germination vigour in Material I, Table I (VP to GI = –0.37)[24]. The accelerating ageing theory using heat stress is thus far from predicting the vigour of the brewing malt either as GI or as g%1[22]. In malting technology other form of stresses i.e. from oxygen depletion and water should be more adequate to study.

We do not claim a global calibration for "vigour" g%1 by NIT-spectroscopy but rather that our preliminary results are indicative for the feasibility of a new direction in future malting barley research. This is demonstrated by our validated hypothesis on the importance of seed structure for vigour, first presented in this publication. We can also conclude that in order to achieve our results we have to address the sensitivity (range) as well as the reproducibility of the analyses. Thus, BGwort (Table III) is a much more suitable parameter than wort viscosity and BGbarley for spanning the functional factor of modification resistance[2,30] just as g%1 is more sensitive than GI as a marker for "vigour".

Compressing malting barley quality data into information – quality index versus hierarchical classification

It is in this context important to realise that there is nothing such as a combined index for vigour[4] and viability[4] because both give unique and complementary information. We have found that their information is best expressed as classes in two-dimensional graphic representations. The seed agronomist Heydecker[10] thus solemnly concludes (p. 225): "Attempts to express germination speed and percentage in one combined index are well meant but confusing". Hampton and Coolbear in 1990[9] agree (p. 225) "It seems clear to us . . . that it is most unlikely that any one aspect of behaviour, whether germinative, physiological or biochemical will be a universally reliable index of all aspects of seed vigour".

However, in brewing science the idea of a malting barley quality index has for a long time been in focus, as discussed and presented by Monnez *et al.*[25] and Molina-Cano[21] in 1987, the latter publication resulting in an EBC index[21]. Most brewing scientists throughout the years[15,21,25,31], in their efforts to classify or to construct a malting quality index for barley, seem to have forgotten to include standardised parameters under GE or GC conditions with regard to viability and vigour. Our results and the above-cited literature indicate that germination parameters for "viability" (g%3) and especially for "vigour" (g%1) are informative enough to be included in malt quality data.

Molina-Cano[21] constructed the Q malting barley index ranging from 1 (feed barley) to 9 (highest malting quality) in three steps. First, five critical characteristics – extract yield, Kolbach index, apparent final attenuation, viscosity and diastatic power – were selected. The reference value was defined as the overall mean for all varieties, locations and years. In the second step an index of quality was defined for each characteristic ranging from 1–9 based on the reference value and the position of the actual sample value in a normal distribution curve. In the third step an overall index was obtained by a weighted linear combination of the indices of quality for each characteristic where the coefficients were based on the judgement of an expert committee. The Q index was tested as a tool for malting barley quality classification for trials in the years 1982–1985 under variable conditions[21]. Our brewing research group[31] has recently suggested a refinement of the EBC-method[21]. An acceptance/rejection profile (membership curve) with scores between 0–1 is worked out by experts and related to the range of values for each parameter. Instead of a constant coefficient for each characteristic, fuzzy logic was used to calculate an overall quality index (OQI) value from the memberships curves. This method was able to adequately rank a limited material of malt analyses from 50 spring and winter barley samples. There was surprisingly a reasonable PLSR correlation between the OQI value and NIT spectroscopy[31], which confirms the conclusions regarding NIT predictions of malt quality parameters in this barley material and by Møller[22].

In evaluating malting barley quality, the importance of different parameters indicative for the practical use of the barley cultivars will change in different years[15,16]. A univariate quality index is not likely to function optimally in such a dynamic reality. Monnez *et al.*[25] (Fig. 9, p. 484) therefore used multivariate pattern recognition analysis to obtain two hierarchical indices, one for the suitability of malting and one for the suitability of brewing for use at two and three levels respectively combining expert validation with barley and malt analyses. The indices are displayed in an abscissa-ordinate plot, divided into classes resembling our germinative energy classification plot for "vigour" and "viability"(Fig. 4A and B) but with a much more complex primary data set.

A quality ranking must be able to recognise the pattern of analysis characteristics of each individual barley sample in a larger calibration context of feed and malting barleys. This is only possible by a multivariate chemometric approach. This could be done, if a large jointly used database could be set up for the benefit of brewers, maltsters and plant breeders as a source of artificial intelligence to visualise the position of each new barley sample in a PCA for classification. Van Lonkhuijsen *et al.*[15] have demonstrated the feasibility of such a strategy in a limited scale where it was possible to reduce a considerable number of analyses without losing information.

In a PCA with 186 commercial malt samples Munck[27] was able to reduce the 11 quality analyses to three "functional combination factors", obtained by identifying the nature of the first three principal components. These factors (PC's) were:

1. "Chemistry" (extract plus a range of enzyme influenced analyses)
2. "Physics" (malt hardness, cell wall thickness (β-glucans), resistance to malt modification[2])
3. "Protein"

In another trial from our research group with a smaller material of 50 spring and winter barley samples only the "physics" functional factor prevailed in a clear cut[35] manner.

The "functional factor" scores (PC's) or function specific indices could be used in classification plots analogous to those for germination classification in Fig. 4A and Fig. 4B in a hierarchical way in different combinations, such as 1, 2, 2:3 and 1:3, as suggested by Monnez *et al.*[25]. They could alternatively be used in a three dimensional classification. We are convinced that if sensitive and reproducible analyses are developed and selected and a broad calibration material is obtained, two to three composite factors covering malting barley functionality should suffice. Such a strategy would be more adaptive to different environmental conditions and much more informative than a simple univariate malting barley quality score and would not increase the analytic workload compared to the present situations. Instead the number of analyses should be able to decrease considerably. The germinative parameters "vigour" and "viability" have their obvious place in such a malting barley quality classification system.

CONCLUSIONS

The role of germinative analyses in developing international malting barley quality control systems based on NIT spectroscopy and image analysis

The simple germinative classification for malting barley presented here could be directly tested and used in practice today with marginal additional costs. Our findings that the major component of the "vigour" parameter is related to the structure of the endosperm makes instrumental analyses such as NIT spectroscopy and image analysis attractive. These parameters are important for the accessibility of the endosperm food store to the embryo and for the speed of malt modification. The non-destructive NIT spectroscopy methods calibrated to protein% and water% are used routinely today for on-line grading of barley by seed elevators and maltsters. The positive experience with global calibrations using NIT with regard to water and protein[20] demonstrates that it is possible to calibrate the spectrometers via the Internet. Reproducibility of the reference analyses is often a greater problem than the predictive results obtained from the spectra.

An extended, semi-intelligent, updated database in service for the malting barley industry and trade analyses is a gigantic, far-sighted task which requires international cooperation on a large scale between transnational communities such as EU, industrial branch collaborations like EBC, instrument companies and a network of industries and universities. Experience from the ring tests in the EBC analytical committee tells us that it is the sensitivity of the reference analyses and the reproducibility between labora-

tories, including the necessary germination methods[37] and micro maltings, that are the weak points. The issue now is how the great potential of NIT spectroscopy[33,36] may be expanded as indicated here to all the physical and chemical parameters in barley, which constitute the physiological[28] and technological basis[5] for germination and malting. In order to overcome the exorbitant costs to obtain global calibrations with NIT and to improve precision in establishing the desired data library we suggest that the analyses in the calibration step should also be automated, preferably as a compounded instrument, combining several analyses.

Maltsters[5] are interested in fast chitting, but slow development of roots and acrospires, while focusing on the speed of malt modification and enzyme development within the endosperm. An instrument for this purpose should be able to analyse 300 seeds in a few minutes with dry single seed analyses starting with seed form/weight parameters and colour and moisture analyses and finishing with a destructive hardness test[36]. The instrument should also be able to work with wet seeds automatically to determine germinated seeds at g%1 and g%3 using fluorescence for early identification. Thus, varieties with an early chitting and slow development of roots and acrospires could be selected for the best compromise for obtaining a high yield of malt and a fast modification. The processes inside the barley endosperm should be speeded up by proper counter selection where excessive development of external organs of the seed should be retarded. Such an instrument could be used as stand-alone in the seed laboratories of the seed dealers, maltsters and plant breeders e.g. to study water sensitivity[13] as well as a calibration instrument for NIT spectroscopy. Tentatively the instrument could even include single seed NIT measurements[34] for control of barley and malt seed homogeneity. By combining information regarding vigour from NIT spectroscopy with that of the Tetrazolium test (EBC 3.5.1)[3] for viability, a germinative classification could be performed within two hours, if a high-quality "global" calibration for g%1 to NIT can be established. The preliminary results obtained with Material II points out that an indicative instant germinative classification could be possible by NIT prediction of "vigour" g%1 and "viability" g%3 alone (Fig. 6). This finding needs further confirmation and explanation.

We have indicated earlier[26,32] that it should be possible to select improved malting barley by its NIT spectrum as a total physical-chemical spectral fingerprint of the sample positioned on a PCA plot with an ideal barley spectrum as reference. Translated to the task envisaged here, it should be possible by calibration of NIT to "vigour" g%1, checking lines with "viability" g%3 below 92% to breed for the whole physical-chemical complex of endosperm availability as a substrate for the embryo. This characteristic should be identical to the ease of malt modification. By means of a set of barley and malt standards spanning the whole quality range documented and distributed by a few acknowledged inter-calibrated laboratories all users could, within the limits of the method, preliminarily evaluate their own barley by their NIT spectra in a PCA. It is in this context essential to disseminate the rather simple use of basic chemometrics[17,18,44] to the whole production chain from plant breeding to malting and brewing. Such knowledge is now in the high days of information technology hiding in the "black box" software of the NIT instruments currently used. Here it does not contribute to the kind of multivariate thinking which today is necessary for successful complex problem solution. All participants in the barley to beer production chain should be able to make a PCA on their own data and know how to interpret the biplot. Chemometrics[44] in brewing science and technology has a potentially more direct profitable impact compared to, for example, biotechnology in solving complex problems.

A dynamic international malting barley quality reference data and sample library calibrated to measurements to standardised instruments by multivariate analysis would be able to speed up and rationalise quality grading in industry, plant breeding and trade to reach a new standard of rationalisation and precision. The two-dimensional "vigour" g%1/"viability" g%3 classification plot suggested here could be used directly in practice for convenient quality grading under germinative energy and capacity conditions and should be included in the software of a future malting barley grading network based on NIT spectroscopy.

It can be concluded that problems associated with agrobiological based technology, dependent on natural variation, only can be solved by a moderate but longsighted enduring investment in research in order to build up a representative data base for reliable calibrations and predictions. This implies an acknowledgement of the unique covariate properties that characterise each biological individual[26,42] such as a homozygotic barley line or variety. These properties can only be understood by an open ended interactive exploratory experimental strategy implemented by multivariate pattern recognition data analysis[44].

ACKNOWLEDGEMENTS

The Seed Company "Skånska Lantmännen" Malmö, Sweden through Lennart Wikström at the SL Foundation, the Danish Cereal Network (B5) and the EU-Qualigrain FAIR CT 97-3648 project are greatly thanked for initiating and further financing our long-sighted research 1994–2002. We appreciate the work of Kees Van Haaren, Pajbjerg Plant Breeding Station who has performed the micro malting and malt analysis. Dr. José Luis Molina-Cano, Centre Udl-IRTA, Spain is gratefully acknowledged for letting us use his barley material in dataset I previously published[24]. Inspiring discussions regarding fluorescence in germination detection with Professor Gary Fulcher, Minnesota State University, St. Paul, U.S.A is greatly acknowledged. Gilda Kischinovsky and Lars Nørgaard are greatly thanked for their comments to the paper.

REFERENCES

1. Aastrup, S., Riis, P. and Hansen, J.R, High vigour – the basis for high malting barley quality. Proceedings of the European Brewing Convention Congress, Zurich, IRL Press: Oxford, 1989, 171–178.
2. Aastrup, S., Malt quality control. *Ferment,* 1995, **8**(**3**), 187–195.
3. Anonymous, European Brewery Convention: Analytica-EBC. Fachverlag Hans Carl:Nürnberg, 1998.
4. Anonymous. Seed vigour testing handbook. Contribution no. 32. The handbook of seed testing. The Association of Official Seed Analysts, USA, 1983, p. 93.
5. Briggs, D.E., Some aspects of barley grain physiology. In: Malting and Brewing Science, Briggs, D.E., Hough, J.S., Ste-

vens, R. and Young, T.W. (Eds.), Chapman and Hall: London Vol. 1, 1995, pp. 39–57.
6. Ellis, D.H. and Roberts, E.H., The influence of temperature and moisture on seed viability period in barley (*Hordeum viaticum* L.). *Annals of Botany,* 1980, **45**, 31–37.
7. Ellis, D.H. and Roberts, E.H., The quantification of ageing and survival in orthodox seeds. *Seed Science and Technology,* 1981, **9**, 373–409.
8. Favier, J.F., A model for germination during dormancy loss in *Hordeum vulgare. Annals of Botany,* 1995, **76**, 631–638.
9. Hampton, J.G. and Coolbear, P. Potential versus actual seed performance – can vigour testing provide an answer? *Seed Science and Technology,* 1990, **18**, 215–228.
10. Heydecker, W., Vigour. In: Viability in seeds. E.H. Roberts, Ed., Chapman and Hall: London, 1972, 209–251.
11. Kleber, W., Kamm, G., Matsuyama, H. and Tsumura, Y., Grain size and germination energy of malting barley. *Brauwelt,* 1969, **109**(**17**), 262–263.
12. Krzanowski, W.R., Principles of Multivariate Analysis – A user's perspective. *Oxford Statistical Science Series*, 1988, **23**, 586.
13. Krauss, G. and Djalali, M.A., Untersuchungen über Keimenergie und Wasserempfindlichkeit der Braugerste (Investigations into germinating energy and water sensitivity of brewing barley). *Brauwelt,* 1969, **22**(**9**), 248–253.
14. Larsen, J., Targets for tailored breeding. In: Symposium on Malting Technology, Andernach, European Brewing Convention. Monograph, Fachverlag Hans Carl: Nürnberg, 1994, 152–164.
15. van Lonkhuijsen, H.J., Douma, A.C. and Angelino, S.A.G.F., Evalution of a malting barley quality assessment system, *Journal of the American Society of Brewing Chemists,* 1998, **56**(**1**), 7–11.
16. Lu, M.Q., O'Brien, L. and Stuart, I.M., Barley malting quality and yield interrelationships and the effect on yield distribution of selection for malting barley quality in early generations, *Australian Journal of Agricultural Research,* 2000, **51**, 247–258.
17. Martens, H. and Næs, T., Multivariate Calibration, Wiley: New York, 1989.
18. Martens, H. and Martens, M., Modified Jack-knife estimation of parameters uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Quality and Preference,* 2000, **11**, 5–16.
19. McCafferty, K.A. and Bryce, J.H., Improvements in the characterisation and quality of vigour in malting barley. *Seed Science and Technology*, 1999, **27**, 633–644.
20. Meurens, M. and He Yan, S., Applications of vibrational spectroscopy in brewing. In: Handbook of Vibrational Spectroscopy. J.M. Chalmers, and P.R. Griffiths, Eds., John Wiley & Sons: Chichester, UK, 2002, **5**, pp. 3663–3671.
21. Molina-Cano, J.L., The EBC and malt committee index for the evaluation of malting quality in barley and its use in breeding. *Plant Breeding*, 1987, **98**, 249–256.
22. Møller, B., Near infrared transmission spectra of barley of malting grade represent a physical-chemical fingerprint of the sample which is able to predict germinative vigour in a multivariate data evaluation model. *Journal Institute of Brewing,* 2004, **110**(**1**), 18–33.
23. Møller, B. and Munck, L., Seed vigour in relation to heat resistance in barley evaluated by multivariate data analysis. *Journal of the Institute of Brewing*, 2002, **108**(**3**), 286–293.
24. Møller, B., Molina-Cano, J.L. and Munck, L., Variation in malting quality and heat resistance in the malting barley variety "Alexis". *Journal of the Institute of Brewing,* 2002, **108**(**3**), 294–302.
25. Monnez, J.M., An approach to the estimation of brewing quality in barley and malt. *Journal of the Institute of Brewing,* 1987, **93**, 477–486.
26. Munck, L., Detecting diversity – a new holistic, exploratory approach bridging phenotype and genotype. In: Diversity in Barley (*Hordeum vulgare*), R. Von Bothmer, T. Van Hinton and H. Knupffer Eds., Elsevier Science: Amsterdam, 2003, chapter 11, pp. 227–245.
27. Munck, L., Quality criteria in the production chain from malting barley to beer. *Ferment,* 1991, **4**, 235–241.
28. Munck, L., Aspects on the regulation of dormancy and vitality in barley seeds and the possible relation between vitality and antimicrobial proteins. In: European Brewing Convention monograph XV, Symposium on Plant Biotechnology, Helsinki, Fachverlag Hans Carl: Nürnberg, 1989, pp. 199–129.
29. Munck, L., Fluorescence analysis of foods, L. Munck, Ed., Longmans Scientific: London, 1989.
30. Munck, L., Jørgensen, K.G., Ruud-Hansen, J. and Hansen, K.T., The EBC methods for determination of high molecular weight β-glucan in barley, malt, wort and beer. *Monatsschrift für Brauwissenschaft*, 1989, **4**, 162–166.
31. Nielsen, J.P., Bro, R., Larsen, J. and Munck, L., Application of fuzzy logic and near infrared spectroscopy for malt quality evaluation. *Journal of the Institute of Brewing,* 2002, **108**(**4**), 444–451.
32. Nielsen, J.P and Munck, L., Exploring malting barley data using chemometrics. In: 8th Int. Barley Gen. Symp., Adelaide, Australia, S. Logue, Ed., Department of Plant Science, Adelaide University, Adelaide, South Australia, 2000, pp. 258–261.
33. Nielsen, J.P. and Munck, L., Prediction of malt quality on whole grain and ground malt using near infrared spectroscopy and chemometrics. In: Near Infrared Spectroscopy: Proc. of the 9th Int. Conf. Davies, A.M.C. and Giangiacomo, R. Eds., NIR Publications: Chichester, UK, 2000, 713.
34. Nielsen, J.P., Pedersen, D.K. and Munck, L., Development of non-destructive screening methods for single seed characterisation of wheat. *Journal of Cereal Chemistry,* 2003, **80**(**3**), 274–280.
35. Nielsen, J.P. and Munck, L., Evaluation of malting barley quality using explorative data analysis. I. Extraction of information from micro-malting data of spring and winter barley. *Journal of Cereal Science,* 2003, **38**, 173–180.
36. Nielsen, J.P., Evaluation of malting barley quality using explorative data analysis. II The use of kernel hardness and image analysis as screening methods. *Journal of Cereal Science,* 2003, **38**, 247–255.
37. Recheneder, E. and Narziss, L., Comparison of methods for determining germinative energy. *Brauwelt International,* 1989, **4**, 338, 345–346, 348–349.
38. Riis, P. and Bang-Olsen, K., Breeding high vigour barley – a short cut to fast maltings, Proceedings of the European Brewery Convention Congress, Lisbon, IRL Press: Oxford, 1991, 101–108.
39. Riis, P., Meiling, E. and Peetz, J., Determination of germinative percentage and germination index – collaborative trial and ruggedness testing, *Journal of the Institute of Brewing,* 1995, **101**(**3**), 171–175.
40. Swanston, J.S. and Taylor, K., The effects of different steeping regimes on water uptake, germination rate, milling energy and hot water extract. *Journal of the Institute of Brewing,* 1990, **96**(**1**), 3–6.
41. Ulonska, E. and Baunder, B., Investigation of the value of water uptake and germination for the estimation of malting quality in Barley. In: Proc. 3d International Barley Genetics Symposium, Verlag Karl Thiemig: Munich, Germany, H. Gaul, Ed., 1975, pp. 579–593.
42. Watkins, S.M., Hammock, B.D., Newman, J.W. and German, J.B., Individual metabolism should guide agriculture towards foods for improved health and nutrition. *American Journal of Clinical Nutrition,* 2001, **74**, 283–286.
43. Woods, J.L. Favier, J.F. and Briggs, D.E., Predicting the germinative energy of dormant malting barley during storage. *Journal of the Institute of Brewing,* 1994, **100**, 257–269
44. www.models.kvl.dk. Chemometric and Spectroscopy group, Food Technology, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Denmark, website for information on chemometrics and spectroscopy.

(Manuscript accepted for publication December 2003)

Near Infrared Transmission Spectra of Barley of Malting Grade Represent a Physical-Chemical Fingerprint of the Sample That Is Able to Predict Germinative Vigour in a Multivariate Data Evaluation Model

Birthe Møller

ABSTRACT

J. Inst. Brew. 110(1), 18–33, 2004

The physiological and physical-chemical basis of barley germination was studied. Vigour was defined as germination percentage after 24 h and viability as that of 72 h. The barley samples were analysed under germination capacity and energy conditions after harvest and after long time cold storage at 7°C three–six years. These parameters were each correlated by Partial Least Squares Regression (PLSR) to two separate multivariate data sets: a set of ten physical-chemical parameters and to Near Infrared Transmission (NIT) spectra (850–1050 nm). Surprisingly high correlation coefficients for each of these two data sets were obtained especially with vigour, extract (%) and β-glucan in wort (mg/L) when outliers with viability below 92% were removed. Hard, slowly germinating seeds were more resistant to decay in vigour and viability storage than soft seeds. This change could be predicted by PLSR correlations to the two physical-chemical multivariate methods. Vigour was a more sensitive indicator for the ability to store than viability. The steep criterion was also found to have a physical-chemical basis. The results indicate that NIT calibrations can be used to predict vigour in malting grade barley.

Key words: Malt quality, Near Infrared Transmission spectroscopy, physical-chemical properties, seed imaging analysis, seed viability, seed vigour.

INTRODUCTION

The malting barley quality complex[5,6,14,20,21,28–30] consists of a wide range of physical and chemical criteria that are manifest already in the intact barley grain as well as indirect parameters, which are first developed during malting and brewing. To overview such an elaborate quality complex, with a minimum of hard assumptions constitutes a great challenge to the brewing chemist. However, thanks to multivariate pattern recognition data analysis[14–16,20,21] also called chemometrics, it is now possible to identify the more and less unique analytical pattern of each individual barley sample for classification and correlation and to trace the influence of important variables as visualised on a graphic display. Now an extended sequential data dissection and component synthesis and identification is possible. Then new hypotheses can be generated and verified to prior knowledge and to further cycles of data analyses and experiments.

Multivariate data analyses have been rather sporadically used for analysing data in brewing research[10,11,14,19–21,28,29,31].

Instrumental screening methods for quality *at-line* and *on-line* have been exploited in the production chain of the brewing industry throughout the last 15 years. Examples are Near Infrared Transmission (NIT) spectroscopy[17] and seed imaging and hardness[29] determination instruments, which will be further utilised in this experimental multivariate approach to elucidate the relations between germinative properties and physical-chemical properties, in order to predict malt quality.

Our research group[25–27,38] has focused on finding a simple and relatively fast criterion for seed vigour. We demonstrated[25], that the barley samples could be classified according to malt quality in a plot with estimated "vigour" (g%1) as abscissa and "viability" (g%3) as ordinate.

This paper is an extended in depth analysis of the barley material harvested in 1993–1999 used by Munck and Møller[25] for germinative classification here also including the storage and steep aspects.

Vigour and viability of seeds are the most important quality criteria in malting barley[6,25,32,33,35]. Because the germination parameters are highly influenced by weather conditions, it is difficult to get a reliable estimation of malting barley quality by testing varieties for one year in one environment only. In Northern Europe there will typically be one to two extreme years with unacceptable vigour and viability during a 10-year period.

It is therefore important to make seed collections in years yielding low vigour seeds, because each such year is unique and cannot be simulated by artificial treatments. The chemical-physical composition[2,18,19,37] can be influenced by temperature, precipitation, fertilisers[1,7] and

The Royal Veterinary and Agricultural University, Department of Dairy and Food Science, Food Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.

Corresponding author. E-mail: bm@kvl.dk

Publication no. G-2004-0317-205

Table I. Sample plan with sample identification and analysis of "vigour" g%1 GC conditions shortly after harvest (germination capacity method). A1, A2, A3 (Alexis), B1 (Blenheim), L1 (Lysimax), M1, M2 (Meltan) were delivered by farmers (0th generation) in 1993 and propagated in 1994–1999. Samples grown on Zealand (z), Funen (f) and in Scania (s). Sample identification example: A8[57] = Alexis grown in 1998 with sample number 57.

	1993	1994	1995	1996	1997	1998	1999
A1	77,0 (f) [1]	37,3 (z) [8]	85,8 (z) [18]	83,1 (z) [30]	45,5 (z) [50]	72,5 (z) [56]	-
		19,0 (s) [9]	35,5 (s) [19]				
		74,0 (f) [10]	46,3 (f) [20]				
A2	68,3 (s) [2]	42,8 (z) [11]	-	97,6 (z) [31]	-	-	-
		85,5 (f) [12]					
A3	2,8 (s) [3]	22,5 (s) [13]	-	98,4 (z) [32]	-	-	-
B1	81,0 (s) [4]	33,3 (z) [14]	37,8 (f) [21]	93,8 (z) [33]	-	-	-
L1	85,8 (z) [5]	55,3 (z) [15]	72,8 (z) [22]	-	-	-	-
M1	32,3 (s) [6]	58,5 (f) [16]	77,8 (s) [23]	85,3 (z) [34]	-	-	-
M2	12,0 (s) [7]	15,0 (s) [17]	75,3 (s) [24]	81,9 (z) [35]	-	-	-
		A1z	80,5 (z) [25]	86,6 (z) [36]	52,5 (z) [51]	79,3 (z) [57]	-
			52,5 (s) [26]				
			40,3 (f) [27]				
		A1s	-	94,6 (z) [37]	-	-	-
		A2z	-	95,2 (z) [38]	88,1 (z) [52]	74,5 (z) [58]	-
		A2f	-	95,4 (z) [39]	-	-	-
		A3s	-	97,3 (z) [40]	-	-	-
		B1z	61,5 (f) [28]	89,3 (z) [41]	-	-	-
		L1z	72,5 (z) [29]	96,9 (z) [42]	75,5 (z) [53]	75,0 (z) [59]	-
		M1f	-	84,9 (z) [43]	-	-	-
		M2s	-	79,0 (z) [44]	-	-	-
			A1zz	86,3 (z) [45]	-	-	-
			A1zf	96,3 (z) [46]	-	-	-
			A1oz	90,0 (z) [47]	-	-	-
			A1of	90,6 (z) [48]	-	-	-
			L1oz	94,0 (z) [49]	-	-	-
				A1ozz	50,3 (z) [54]	-	-
				A2zoz	76,5 (z) [55]	83,0 (z) [60]	-
				L1ozz	-	77,3 (z) [61]	-
						A2zoooz	85,3 (z) [62]
						L1zoooz	80,8 (z) [63]
						L1ozzoz	90,5 (z) [64]

0 growing generation; 1st growing generation; 2nd growing generation; 3rd growing generation; 4th growing generation

weathering including microorganisms[12,24,36], which in malting may influence steeping conditions[12,13,35,36], water sensitivity[6] and dormancy[6]. The climate also affects levels of pre-germination[7].

In order to obtain a sample set, which is suitable for over viewing and comparing the different quality control methods and for demonstrating the advantage of multivariate data analysis, we have grown the same batch of seeds of different samples in a range of years with widely different climatic conditions (Table I). In order to preserve vigour the material was stored dry under refrigerated conditions.

MATERIALS AND METHODS

Materials

A rare extreme season with a difficult and wet harvest occurred in 1993 in Denmark and Southern Sweden producing a great variation in seed vigour. Table I gives an overview of the origin of the samples and their "vigour" (defined as g%1) measured by EBC 3.5.2[3] (GC method) shortly after harvest. The samples are marked with numbers 1–64. Seeds from the seven original batches from 1993 listed to the left in Table I were planted in 1994–1998 as the 0 growing generation of seeds. The harvested seeds from 1994, called first growing generation, were planted in 1995–98. Harvested samples from 1995, called second growing generation were planted in 1996 and so on until the year 1999. The sample numbers are used to facilitate identification of specific samples in the figures. All three locations were good soils (JB 5-6) representing a geographical area of approximately 200 km^2 in Southern Scandinavia. In all, 64 samples were collected, 63 of which were subjected to analysis of malt quality (See Table I for sample identification and Tables IIA–B, IIIA–B, IVA–B and VA–B for results). Seed size was very low in 1998 and 1999 presumably due to a high level of weeds in the field those years, a condition that did not affect germination properties. All samples were stored at 7°C with 13.5 ± 0.8% moisture. There was no remaining dormancy in the samples in 1999 when the samples were micromalted.

The mean 3-day (72 h) germination energy (g%3) of the material fell from 97.1% to 94.7% during the storage period of six years from 1993 to1999.

Alexis and Blenheim are categorised as malting varieties. The reason for including the varieties for feed, Meltan and Lysimax, is that it is necessary to include extreme barley varieties in order to expand the range of important parameters for an improved multivariate evaluation of the efficiency of the instrumental screening methods. Lysimax has an extreme mutant gene (*lys3a*), which causes a deviating amino acid composition high in lysine[22,23]. This variety is included for theoretical reasons because of its high level of hydrophilic proteins, fast germination characteristics and low β-glucan content.

Analytical methods

A. Germination analyses (Tables I–III)

Two different germination analyses were carried out with 100 seeds in four replicates. In both methods percentages of germinated kernels (*n*) were calculated and the germinated kernels were removed after 24, 48 and 72 h.

(1) Determination of total germinative capacity (GC) in percentage (Tables I and IIA–B). Less than two months after harvest (1993–1999) seeds ($n = 64$) were immersed in a 0.75% H_2O_2 solution to remove dormancy, according to EBC standard method 3.5.2[3]. Every day the H_2O_2 solution was changed and new added to each sample.

(2) Determination of the germinative energy (GE) percentage (EBC 3.6.2)[3] (Table IIIA–B) was performed for samples grown 1993–1996 in the year 1999 after storage three–six years at 7°C (47 samples in total). Kernels were germinated in 90 mm petri dishes with two layers of filter paper (Whatman No. 1) and 4 mL H_2O.

In both analyses (GC and GE) the samples were placed in a dark Refritherm incubator at 20°C. The standard deviation between the replicates was less than 5%.

Germination Index (GI) and Germination Homogeneity (GH) were determined for both GC and GE analyses[32,33] according to EBC 3.7[3].

(3) Estimation of vigour and viability (Tables I–III). Vigour was defined as 24 h germination percentage or g%1(here denoted as "vigour") in the application of both EBC methods 3.5.2[3] (GC) and 3.6.2[3] (GE) as suggested by Munck and Møller[25]. Viability is defined as 72 h germination percentage or g%3 (here denoted as "viability")[25].

B. Physical kernel parameters (Tables IVA–B and VIA)

These analyses were made in 1999 together with the chemical analyses, NIT analyses and the pilot malting.

Thousand-kernel weight: 1000 kernels of every sample were weighed in three replicates.

Seed imaging measurements: A GrainCheck™ 310 instrument (Foss Tecator, Höganas, Sweden) was used to determine the kernel size and light reflectance parameters and their distribution: width, length, roundness, area, volume and total reflected light intensity. The instrument uses digital image analysis to determine the parameters for every single kernel measured in bulk and to calculate the average and standard deviation values for the sample which consists of 300 kernels. In total 60 samples were measured.

Hardness of barley seeds: Kernel hardness (Hardness Index – HI) was determined by the Perten SKCS 4100 (Single Kernel Characterisation System, Perten North America, Reno, NV, USA). The hardness index refers to the American wheat classification system defined by the United States Department of Agriculture, Technical Service Division of the Grain Inspection, Packers and Stockyard Administration. The Perten SKCS 4100 determines individual kernel weight, moisture content, diameter and crushing force profiles described as hardness index[29]. HI was determined as an average of the 300 kernels. Samples with HI < 33 are characterised as soft, 33 < HI < 46 as semi-soft, 46 < HI < 59 as semi-hard, and samples with HI > 59 as hard. The instrument is optimised for wheat and accepts only kernels with a width of more than 2.2 mm and a round wheat-like seed form. Analysing barley there are a great number of rejections (a mean of 464 to obtain 300 kernels). Number of rejected kernels is therefore considered in the investigation. It can be concluded that even if HI is dependent on rejected kernels it gives valuable, specific information. The instrument should be optimised for barley. HI was determined on 58 samples.

C. Chemical analyses (Tables IIA–B and VA–B)

The chemical composition was measured after harvest (1993–1999) according to the standard methods: dry matter (ICC 110/1)[4], protein content (EBC methods 3.2)[3] and content of β-glucan (EBC method 3.11.2)[3]. The α-amylase activity in barley was determined according to ICC 108[4] to study pre-germination.

D. Steeping properties (Table IIA–B)

Steeping characteristics were determined in a separate experiment by weighing 5.2 grams of whole grains before and after steeping in 15 mL 2.5% H_2O_2 for 24 h. The weight enhancement (%) was calculated. This procedure was developed in 2002 in order to facilitate NMR measurements of water uptake (not reported here), allowing the seeds to germinate in immersed state for a long time without change of fluid while still obtaining enough oxygen for germination. In total 62 samples were measured.

E. Malt analysis (Table VA–B)

Micromalting and malt analyses were performed on 63 samples in 1999 with the cold stored grains on a pilot plant system developed and built by the Pajbjerg Plant Breeding Station (The Pajbjerg Foundation, DK-8300 Odder, Denmark). The standardised steeping programme was 8 h with water, 16 h air-break, 9 h with water. Steeping and germination were performed at 16°C and 100% RH. Germination was performed in 87 h. The kilning programme was 16 h at 45°C, 2 h at 65°C and 6 h at 85°C. A sample of Alexis was used as a standard. Extract was determined with a refractometer and β-glucan in wort was determined using an in-house colour-binding method by which β-glucan and Congo red develop a colour complex, which is measured with a spectrophotometer.

F. Near infrared transmission (NIT) spectroscopy (Table VIB)

In 1999 spectra were obtained for intact whole kernels stored 1993–1999 ($n = 62$) when measuring bulk samples of 60.0 g using an Infratec 1225 Food and Feed Analyzer (Foss Tecator, Höganäs, Sweden). The spectrophotometer measures spectra in the wavelength area of 850–1050 nm, collecting data every second nm yielding 100 data points for each of the 62 samples measured.

G. Multivariate data analysis

Data analysis was performed according to Martens and Næs[15] using the software "Unscrambler" version 7.6 SR-1 from CAMO A/S, Trondheim, Norway for Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR)[16] where Jack-knife validation calculates the important variables[17]. Different combinations of principal components (PC's) are tested, and the combination show-

ing the most interesting results are shown (primarily PC1: PC2). The principal components indicated as PC's in PLSR analysis with the 0"Unscrambler" software are mathematically not identical with the PC's denoted in PCA analysis. Data processing was performed by scaling the physical-chemical and physiological analyses (1/std.dev.). The NIT spectra were transformed to the first derivate. The performance of the regression models is evaluated by its prediction error in terms of root mean square error of cross-validation (RMSECV). The relative error (RE) in percentage is calculated[28] as

$$(\text{RMSECV}/(\mathbf{y}_{max} - \mathbf{y}_{min})) * 100$$

where $\mathbf{y}_{max}$ is the highest reference value and $\mathbf{y}_{min}$ the lowest reference value of the **y** parameter in question.

Abbreviations

A	Alexis
AREA	Area of kernel (mm^2)
B	Blenheim
BG	(1-3,1-4)-β-glucan in barley
BGwort	(1-3,1-4)-β-glucan in wort (mg/L)
C	Wort colour
EBC	European Brewery Convention
EXTRACT	Extract yield (%)
g%1–3	Germination percentage day 1–3
GC	Germination Capacity Method EBC 3.5.2[4]
GE	Germination Energy Method EBC 3.6.2[4]
GH	Germination Homogeneity[32,33]
GI	Germination Index
HI	Hardness Index
INTENSITY	Total intensity
L	Lysimax
LENGTH	Kernel length (mm)
M	Meltan
NIR	Near Infrared Reflection
NIT	Near Infrared Transmission
P	Protein
PC	Principal Component in PCA and PLSR, see Materials and Methods **G**
PCA	Principal Component Analysis
RE	Relative Error in percentage
PLSR	Partial Least Squares Regression
RMSECV	Root Mean Square Error of Cross-Validation
ROUND	Kernel roundness
STEEP	% Water uptake after 24 h of steep
TKW	Thousand Kernel Weight
VOLUME	Volume of kernel (mm^3)
VP	Vigour Potential
WIDTH	Kernel width (mm)

RESULTS AND DISCUSSION

An overview of the experimental material with means, ranges and standard deviations (Tables I–V)

The germination profiles under GC conditions for years and varieties are displayed in Fig. 1A and Fig. 1B respectively. Mean "vigour" (Table IIA) for the years 1993 and 1994 (51.3 and 44.3%) are especially low in contrast to the best malt quality years 1996 and 1999 (90.8 and 85.5%). However, when estimating "viability" as g%3 the mean results in 1995 (93.6%) are even lower than in 1993 (94.9%) and in 1994 (96.3%) compared to 99.1% in 1996. These differences are reflected in the germination profiles in Fig. 1A and may be explained by the available meteorological data from Zealand.

As discussed in the following, the year 1995 gave rise to exceptionally hard seeds (HI) (Table IVA) with the lowest steep value (Table IIA). Precipitation was exceptionally low during July and August (51 mm) of 1995 compared to the wet years 1993 (181 mm) and 1994 (169 mm). The corresponding values are for 1996 (87 mm), 1997 (108 mm), 1998 (168 mm) and for 1999 (143 mm). The years 1993 and 1995 were regarded as difficult barley quality years for the malting industry, but for different reasons: too much precipitation in 1993 and too little in 1995. The other years were acceptable to excellent (1996).

There was a pregermination tendency in 1999 (three samples only) (Table IIA) including the malting variety Alexis, which, however, did not seem to influence germination ("viability"g%3 = 98.9%).

With respect to germination properties (GC conditions) of the four varieties (Fig. 1B, Table IIB), Lysimax is the fastest germinator ("vigour" g%1 = 79.7%). It is the only variety which displays field germination in addition to the one sample of Alexis harvested in 1999 (α-amylase in samples L9[63]: 41.6 units, A9[62]: 53.7 units).

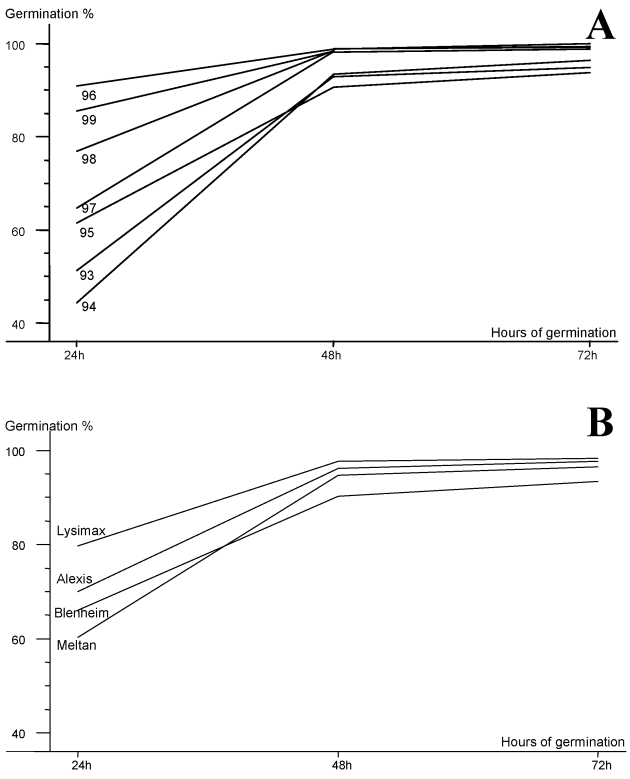


Fig. 1A. Average germination percentage at 24, 48 and 72 hours according to the germination capacity (GC) method after harvest according to harvest year (data Table IIA). **B.** Average germination percentage at 24, 48 and 72 hours according to the germination capacity (GC) method after harvest according to variety (data from Table IIB).

Table IIA. Average and std. dev. of germination capacity, α-amylase (barley) and steeping parameters divided according to harvest year.

	1993	1994	1995	1996	1997	1998	1999	Total
n	7	10	12	20	6	6	3	64
g%1 (GC)	51.3 ± 34.8	44.3 ± 23.7	61.5 ± 18.2	90.8 ± 5.9	64.7 ± 17.5	76.9 ± 3.8	85.5 ± 4.9	77.5 ± 15.8
"Vigour"	2.8–85.8	15.0–85.5	35.5–85.8	81.9–98.4	72.5–83.0	80.8–90.5	80.8–90.5	2.8–98.4
g%2 (GC)	92.8 ± 7.8	93.3 ± 4.5	90.6 ± 5.7	98.7 ± 0.5	98.3 ± 1.2	98.4 ± 0.8	98.3 ± 1.0	96.8 ± 2.8
	75.8–98.8	85.3–98.6	78.3–96.3	97.4–99.4	96.8–99.9	97.5–99.6	97.6–99.5	75.8–99.9
g%3 (GC)	94.9 ± 6.6	96.3 ± 3.5	93.6 ± 4.4	99.1 ± 0.3	99.5 ± 0.4	98.9 ± 1.0	98.9 ± 0.6	97.9 ± 2.0
"Viability"	80.5–99.3	88.3–99.3	84.5–98.5	98.5–99.6	99.3–100.0	97.5–99.8	98.3–99.5	80.5–100.0
GI (GC)	7.1 ± 1.7	6.6 ± 1.2	7.4 ± 1.0	9.3 ± 0.5	7.5 ± 1.0	8.1 ± 0.3	8.8 ± 0.4	8.4 ± 0.9
	4.9–8.8	5.4–8.9	6.1–8.8	8.3–9.9	6.4–9.0	7.8–8.5	8.4–9.2	4.9–9.9
GH (GC)	58.8 ± 6.4	50.3 ± 7.9	49.9 ± 7.9	71.9 ± 9.0	52.7 ± 8.5	56.3 ± 2.7	63.8 ± 6.6	59.34 ± 11.9
	50.2–69.6	42.7–64.0	38.2–62.4	59.0–85.9	44.4–66.5	53.1–59.1	59.2–71.3	38.2–85.9
α-Amylase	0.43 ± 0.53	0.18 ± 0.08	0.11 ± 0.03	0.12 ± 0.04	0.2 ± 0.2	1.0 ± 1.2	47.7 ± 8.6	4.1 ± 13.2
		(*n* = 9)		(*n* = 19)		(*n* = 3)	(*n* = 2)	(*n* = 58)
	0.1–1.5	0.1–0.3	0.1–0.2	0.1–0.2	0.1–0.6	0.1–3.1	41.6–53.7	0.1–53.7
Steep	38.1 ± 1.6	36.9 ± 2.7	35.6 ± 2.0	37.3 ± 2.3	37.8 ± 2.1	42.6 ± 1.6	41.5 ± 1.9	38.0 ± 2.4
	(*n* = 5)							(*n* = 62)
	36.4–40.3	31.4–40.3	32.6–38.7	33.2–43.2	35.5–41.1	40.7–44.8	39.3–42.9	31.4–44.8

Table IIB. Average and std. dev. of germination capacity, α-amylase (barley) and steeping parameters divided according to variety.

	Alexis	Blenheim	Lysimax	Meltan
n	37	6	11	10
g%1 "Vigour" (GC)	70.0 ± 25.5	66.1 ± 26.2	79.7 ± 11.8	60.2 ± 29.3
g%2 (GC)	96.1 ± 4.8	90.2 ± 7.6	97.6 ± 2.0	94.7 ± 4.9
g%3 "Viability" (GC)	97.5 ± 3.7	93.4 ± 5.4	98.7 ± 0.8	96.5 ± 3.8
GI (GC)	8.0 ± 1.4	7.8 ± 1.6	8.4 ± 0.8	7.4 ± 1.4
GH (GC)	60.1 ± 13.0	54.9 ± 14.7	61.7 ± 10.8	56.5 ± 6.5
α-Amylase	0.2 ± 0.1 (*n* = 33)	0.3 ± 0.5	10.1 ± 20.0 (*n* = 9)	0.1 ± 0.1
Steep	37.3 ± 2.7 (*n* = 36)	37.3 ± 1.3	41.0 ± 3.0 (*n* = 10)	36.4 ± 2.1

The feed barley Meltan (mean estimated "vigour" g%1 = 60.2%) is the slowest in germination compared to the malting barley varieties Alexis (70.0%) and Blenheim (66.1%). Blenheim, however, shows the lowest mean estimated "viability" with a mean value of 93.4% compared to 96.5% for Meltan, 97.5% for Alexis and 98.7% for Lysimax.

The germination results (GE conditions) after cold storage in Tables IIIA–B as well as the physical-chemical and malt parameters given in Tables IVA–B and VA–B are elaborated upon in the discussion together with the multivariate evaluation.

Evaluating physical-chemical and germinative malting parameters of the barley material by PCA

In this study the PCA biplot in Fig. 2A summarises the 19 variables of which 11 are classified as "manifest" on the barley raw material level and eight as "indirect" characteristics, only to be attained through germination and malting. The PCA biplot presents a convenient overview over the relations between the 63 barley and malt sample (for identification see Table I) and the 19 analytical parameters (Tables IIA–B, IVA–B and VA–B) where the sample symbols marked according to variety, year and number lying near to each other represent similar quality profiles. Variable symbols lying near to each other are positively correlated

It is seen that the variables describing the abscissa – PC1 (Fig. 2A, 34% of the variance) are the physical and germinative parameters HI, Rej (rejected grains during HI determination), Round, Width, Volume, TKW and the GC germination parameters g%1, g%3, GI and GH as well as Extract while the ordinate PC2 (25% of the variance) is mainly described by Steep, Intensity, Length, BGwort and Area. The sign Extract is adjacent to the germination parameters g%1, g%3, GI and GH in a cluster. They are thus positively correlated.

The light reflection intensity value registered with the RGB camera in the GrainCheck instrument gives a low value with darker weathered seeds as for the rainy harvest year 1993 (marked 3), while the dry year in 1995 (5) gives

Table IIIA. Average and std. dev. of germination energy and moisture parameters after 3–6 years of storage according to harvest year.

	1993	1994	1995	1996
n	5	10	12	20
g%1 "Vigour" (GE)	36.1 ± 24.1	43.9 ± 29.7	58.4 ± 18.3	78.3 ± 12.7
g%2 (GE)	76.6 ± 35.2	82.8 ± 29.2	95.7 ± 2.2	98.2 ± 2.0
g%3 "Viability" (GE)	85.2 ± 22.1	87.4 ± 26.5	97.1 ± 1.4	99.3 ± 0.6
GI (GE)	5.9 ± 1.4	6.5 ± 1.5	7.2 ± 1.0	8.3 ± 0.9
GH (GE)	47.8 ± 4.7	49.6 ± 8.3	51.7 ± 4.7	59.4 ± 10.5
Moisture, min-max	13.3–14.5	13.3–15.4	12.9–14.5	11.7–13.8

Table IIIB. Average and std. dev. of germination energy and moisture parameters after 3–6 years of storage according to variety.

	A93–96	B93–96	L93–96	M93–96
n	26	6	5	10
g%1 "Vigour" (GE)	67.1 ± 22.0	64.5 ± 18.3	70.3 ± 24.9	36.2 ± 25.9
g%2 (GE)	93.9 ± 17.4	94.8 ± 6.4	97.7 ± 1.4	80.8 ± 28.5
g%3 "Viability" (GE)	95.2 ± 16.4	96.2 ± 4.6	98.4 ± 0.9	89.8 ± 18.0
GI (GE)	7.7 ± 1.2	7.5 ± 1.0	8.0 ± 1.5	6.0 ± 1.4
GH (GE)	55.4 ± 8.7	52.9 ± 7.0	60.4 ± 12.3	47.2 ± 7.6
Moisture, min-max	12.4–15.3	11.7–14.1	12.9–13.7	12.7–15.4

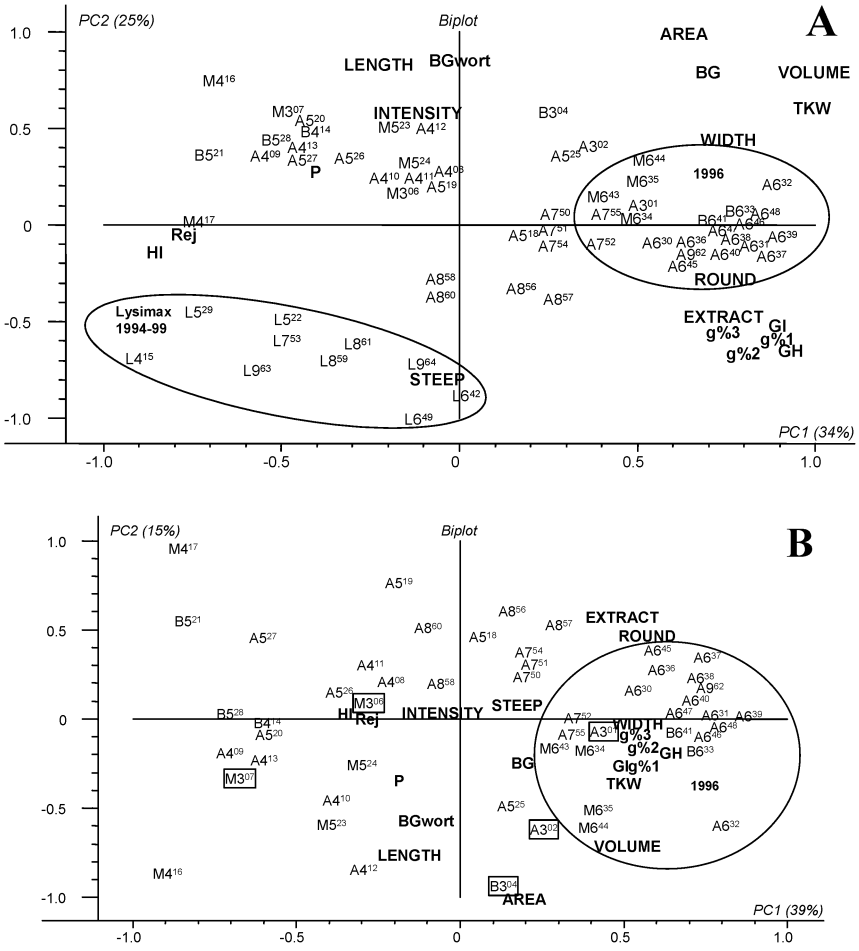


Fig. 2A. PCA biplot (PC1:2) overviewing all samples and physical-chemical malting parameters. For sample identification see Table I. Outliers $A3^{03}$ and $L3^{05}$ are removed. Germination method is GC^4. **B.** PCA biplot (PC1:2) for all samples without the extreme variety Lysimax. Outlier A^{03} is removed. Germination method is GC^4. See abbreviations in the Materials and Methods section.

Table IVA. Average and std. dev. of physical and morphological parameters according to harvest year.

	1993	1994	1995	1996	1997	1998	1999	Total
n	5	10	12	20	6	6	3	62
TKW	41.8 ± 1.5	38.1 ± 2.6	38.0 ± 2.6	43.8 ± 3.1	40.3 ± 3.5	31.9 ± 2.7	33.6 ± 7.8	39.9 ± 5.3
Round	0.29 ± 0.08	0.25 ± 0.02	0.28 ± 0.01 (*n* = 11)	0.30 ± 0.02 (*n* = 19)	0.30 ± 0.02	0.29 ± 0.02	0.28 ± 0.03	0.29 ± 0.03 (*n* = 60)
Length	8.80 ± 0.16	9.08 ± 0.34	8.76 ± 0.27 (*n* = 11)	8.56 ± 0.22 (*n* = 19)	8.60 ± 0.11	8.58 ± 0.18	8.58 ± 0.03	8.71 ± 0.29 (*n* = 60)
Width	3.65 ± 0.08	3.44 ± 0.08	3.51 ± 0.09 (*n* = 11)	3.60 ± 0.10 (*n* = 19)	3.62 ± 0.12	3.58 ± 0.10	3.48 ± 0.19	3.56 ± 0.12 (*n* = 60)
Area	22.51 ± 0.73	22.16 ± 0.89	21.87 ± 0.82 (*n* = 11)	22.19 ± 0.80 (*n* = 19)	21.86 ± 0.53	21.34 ± 0.52	21.32 ± 0.91	21.99 ± 0.81 (*n* = 60)
Volume	51.57 ± 2.68	48.17 ± 2.58	48.46 ± 2.54 (*n* = 11)	50.91 ± 2.78 (*n* = 19)	49.55 ± 2.42	47.47 ± 1.94	46.97 ± 4.33	49.38 ± 2.94 (*n* = 60)
Intensity	65.43 ± 1.33	74.83 ± 5.36	75.67 ± 2.00 (*n* = 11)	69.25 ± 1.91 (*n* = 19)	65.37 ± 2.10	65.46 ± 2.29	68.51 ± 2.40	70.23 ± 4.85 (*n* = 60)
HI	56.7 ± 8.9 (*n* = 4)	57.8 ± 9.3	65.8 ± 9.7 (*n* = 11)	41.7 ± 4.9	60.5 ± 11.4	58.1 ± 9.4 (*n* = 4)	45.5 ± 17.6	49.5 ± 11.9 (*n* = 58)
Rej. kernels HI	405 ± 102 (*n* = 4)	575 ± 171	567 ± 138 (*n* = 11)	356 ± 131	356 ± 131	536 ± 58 (*n* = 4)	448 ± 199	464 ± 163 (*n* = 58)

Table IVB. Average and std. dev. of physical and morphological parameters according to variety.

	Alexis	Blenheim	Lysimax	Meltan
n	36	6	10	10
TKW	40.8 ± 3.8	41.4 ± 3.5	32.2 ± 3.3	41.6 ± 3.4
Round	0.29 ± 0.02 (*n* = 34)	0.29 ± 0.01	0.27 ± 0.01	0.27 ± 0.02
Length	8.70 ± 0.31 (*n* = 34)	8.81 ± 0.20	8.51 ± 0.16	8.89 ± 0.28
Width	3.61 ± 0.08 (*n* = 34)	3.58 ± 0.09	3.37 ± 0.07	3.53 ± 0.08
Area	22.18 ± 0.50 (*n* = 34)	22.52 ± 0.56	20.64 ± 0.27	22.40 ± 0.77
Volume	50.38 ± 1.65 (*n* = 34)	51.16 ± 2.58	44.25 ± 1.02	50.07 ± 2.51
Intensity	70.92 ± 4.90 (*n* = 34)	71.61 ± 5.55	67.45 ± 5.09 (*n* = 10)	69.84 ± 3.38
HI	48.3 ± 9.7 (*n* = 33)	56.9 ± 13.2	68.6 ± 14.5 (*n* = 9)	54.2 ± 8.3
Rej. kernel HI	418 ± 130 (*n* = 33)	518 ± 245	584 ± 186 (*n* = 9)	454 ± 127

a light-coloured seed with high reflectance value (Table IVA). The Intensity sign on the PCA plot in Fig. 2A is consequently located near samples grown in 1995 (5).

Extract (Fig. 2A) is associated to Round kernels, while BGwort is related to Length, as in the study of Nielsen[29]. BGwort is negatively correlated to the variable Steep because it is located in the opposite direction along PC2 – the ordinate. Looking for patterns between samples from harvest years the samples of 1996 (6), which have by far the best malt quality, are all located very close to each other to the right. They are more closely connected to Extract and germination parameters compared to samples from the other harvest years, which are widely distributed, indicating a smaller variance in malting quality between the 1996 samples.

The PCA plot in Fig. 2A is influenced by the extreme variety Lysimax (marked L), of which nearly all samples are located in the bottom left corner near the Steep variable.

The Lysimax is an outlier barley variety because of its high steep percentage, low β-glucan content and fast germinations in spite of a hard kernel. It has been included to test the sensitivity of the instrumental methods.

When the samples of Lysimax are excluded (Fig. 2B), the associations between Extract and Round as well as

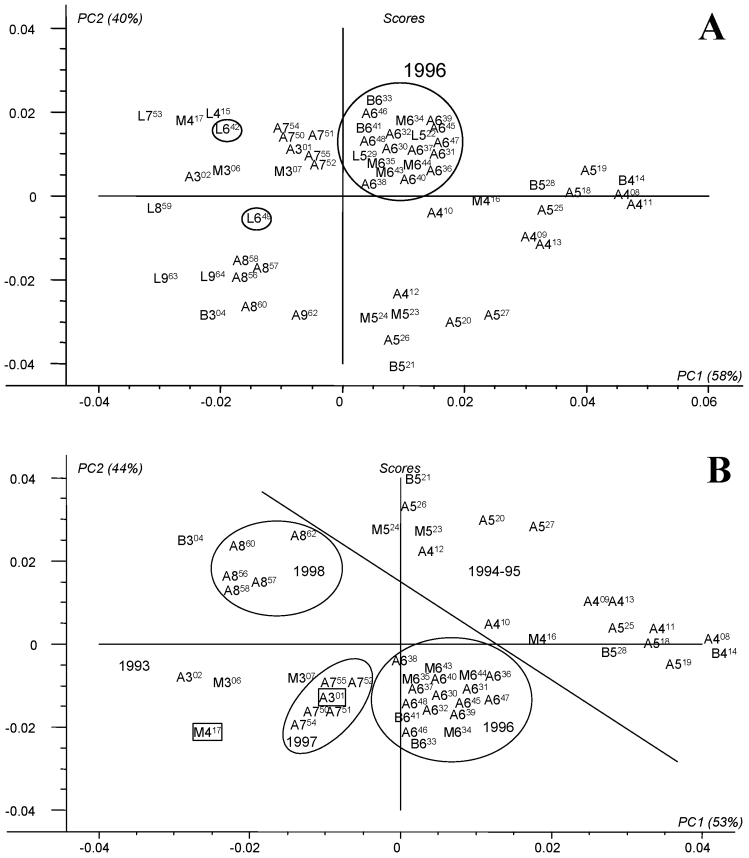


Fig. 3A. PCA score plot (PC1:2) for NIT-spectra (1.der.) for all samples. $A3^{03}$ and $L3^{05}$ not measured due to lack of material. **B.** PCA score plot (PC1:2) for NIT-spectra (1.der.) excluding samples of Lysimax.

BGwort and Length from Fig. 2A are confirmed. Now the samples do not divide according to variety as much as to harvest year. As in Fig. 2A, the samples harvested in 1996 (6) (encircled) are located together to the right of the PCA plot highly influenced of Extract and germination properties. Samples from 1997 (7) and 1998 (8) are placed in the upper half in the middle of the plot, whereas samples harvested in 1994 (4) and 1995 (5) are located to the left, indicating low malting quality. The original samples from the difficult harvest year in 1993 (3) are marked with a square in Fig. 2B. These samples were originally selected to describe a large variation in germination characteristics and are consequently located all over the plot.

We can thus conclude that there are tendencies towards a PCA classification according to harvest year in this material with 19 parameters.

The barley material in this investigation is far from a complete design with equal number of samples and varieties each year, which makes it difficult to evaluate with classic statistics based on means and variances of classes which each, should consist of several samples. It is important to note that this is not a problem in multivariate analysis based on pattern recognition where each sample is individually defined as a more and less unique pattern[8] in this case by 19 variables which all are complete. It is, however, clear that a small number of samples as for 1999 (n = 3) with only two varieties makes it difficult to draw general conclusions about that year. The conclusions, which can be drawn on the individual level for these three samples, however, still hold. The multivariate advantage is further discussed by Munck and Møller[23,25].

Classification by PCA of NIT spectra from barley samples

NIT spectra[17] facilitate in principle a global non-destructive physical-chemical fingerprint[23] of the barley samples, where samples are measured in the wavelength area 850–1050 nm obtaining every second measurement, resulting in 100 data points for every sample. This can be used for classification in PCA models and for PLSR calibrations[16,25] where the spectra can be used to predict physical-chemical values (e.g. protein).

NIT spectra were measured on the barley material (Table I) on 62 samples (two samples missing due to lack of material). A PCA plot reflecting the pattern of the whole spectra is displayed in Fig. 3A. Samples with similar spectral patterns are situated adjacent to each other in the plot. The samples divide into different groups, where the encircled samples to the right are harvested in 1996 (6). Two samples of Lysimax (L) (encircled) from 1996 (6) are located to the left outside the 1996 group, but closer to the other Lysimax samples. There is a tendency that the samples located in bottom right quadrant were

Table VA. Average and std. dev. of chemical and malt parameters according to harvest year.

	1993	1994	1995	1996	1997	1998	1999	Total
n	6	10	12	20	6	6	3	63
Protein	12.6 ± 1.6 (n = 5)	10.3 ± 0.8	10.5 ± 1.2	9.1 ± 0.4	10.8 ± 0.5	11.2 ± 0.4	10.8 ± 1.3	10.0 ± 1.1 (n = 62)
(barley)	10.2–14.5	9.0–11.4	9.2–12.9	8.6–9.8	10.2–11.6	10.7–11.6	9.3–11.8	8.6–14.5
BG	4.0 ± 0.3	3.7 ± 0.5	3.9 ± 0.6	3.9 ± 0.4	3.8 ± 0.4	3.4 ± 0.2	3.3 ± 0.5	3.8 ± 0.4
(barley)	3.6–4.3	2.5–4.5	2.5–4.5	2.9–4.5	3.2–4.2	3.2–3.6	2.9–3.8	2.5–4.5
Extract (%)	71.9 ± 12.9	77.5 ± 7.0	80.1 ± 3.7	82.5 ± 5.2	81.5 ± 2.3	82.4 ± 0.8	81.6 ± 2.1	80.7 ± 5.0
	49.2–83.0	61.2–84.0	70.3–84.1	81.1–85.6	78.2–84.0	81.7–83.7	80.1–83.9	49.2–85.6
BGw (mg/ml)	360.3 ± 121.3	231.1 ± 102.4	204.0 ± 67.5	173.1 ± 52.9	174.5 ± 28.7	174.4 ± 28.1	66.9 ± 7.1	181.2 ± 73.3
	163.2–532.8	61.3–321.3	111.9–328.0	66.3–244.9	153.6–229.3	141.5–210.0	58.7–71.2	58.7–532.8

Table VB. Average and std. dev. of chemical and malt parameters according to variety.

	Alexis	Blenheim	Lysimax	Meltan
n	37	6	10	10
Protein	10.1 ± 1.1 (n = 36)	10.5 ± 2.1	10.4 ± 1.2	10.8 ± 1.6
BG	4.0 ± 0.3	4.1 ± 0.5	3.0 ± 0.3	3.8 ± 0.1
Extract	81.1 ± 6.5	81.3 ± 1.9	81.2 ± 1.2	74.7 ± 8.8
BGwort	195.7 ± 74.2	233.6 ± 69.7	104.7 ± 44.3	302.9 ± 95.8

harvested in 1994 (4) and 1995 (5), most of the samples found in the top left quadrant were harvested in 1997 (7) and samples harvested in 1998 (8) and 1999 (9) are seen in the bottom left quadrant.

The PCA pattern obtained from the NIT analyses in Fig. 3A is related more to harvest year than to the pattern in the PCA plot of physical and chemical parameters in Fig. 2A, although all the samples of Alexis (A), Blenheim (B) and Meltan (M) harvested in 1996 (6) are located close together on both plots. There is a stronger tendency for a clear classification of Lysimax (L) in the analytical PCA plot in Fig. 2A than in that of the NIT in Fig. 3A.

When the extreme Lysimax samples are excluded from Fig. 3A, a much clearer division according to harvest years is seen (Fig. 3B). This is comparable to the change in the PCA classification plots between Fig. 2A and Fig. 2B with and without Lysimax respectively.

All samples harvested in 1998 (8) are located in the top left quadrant (Fig. 3B), all 1997 (7) samples in the bottom left quadrant, and the samples harvested in 1996 (6) are found in the bottom right quadrant. In the top right quadrant the samples from 1994 (4) and 1995 (5) are located close together. One exception is M4[17] (low "vigour" and normal "viability"), which is located in the opposite direction of the 1994 samples (4) as an outlier. Samples from 1993 (3) are located "all over" to the left side of the PCA plot (Fig. 3B), indicating larger variation in physical-chemical structure compared to e.g. 1997 (7) samples because the 1993 samples (3) were selected to obtain large variation.

In conclusion the resemblance of the sample classifications with the two different analytical methods in Fig. 2A–B and Fig. 3A–B indicates a firm physical-chemical basis for the NIT spectral analysis, which will be further supported by the high prediction values in the following discussion.

Prediction of barley germination and malt quality parameters from two different physical-chemical data sets by PLSR

It is of great interest for the plant breeding, malt- and brewing industry to be able to predict indirect parameters (e.g. germination parameters, extract, BGwort as **y**) from parameters manifest in the barley raw material e.g. by the fast non-destructive NIT spectroscopy measurements now used *on-line* by the grain industry to predict water and protein[17]. The precision of such measurements should be indicative enough to be able to classify barley samples *on-line* into two classes: High and low malting quality where only the first is accepted for further analytical scrutiny *at-line*[25]. Here two separate screening methods for manifest physical-chemical properties are compared: the set of ten parameters (six seed imaging measurements plus TKW, hardness, protein, BGbarley) and the NIT dataset in making such predictions by PLSR. This is done for two reasons. First to increase the validity of the predictions in a limited material and second to better understand how the NIT technology works. A great number of PLSR predictions are shown in Table VIA (a01–a20) using the set of ten variables and in Table VB (b01–b28) where NIT measurements are used (**X**). Outlier samples have been detected and removed stepwise by consulting the influence plots of the PLSR software as described by Munck and Møller in the adjacent paper in this issue[25]. It is a general tendency that the prediction of indirect parameters, after storage in both materials, produces outliers with a "viability" (g%3, GE) below 92%(Tables VIA–B). This observation is valid for 16 cases. Exceptions are a05, a06, b15 and b25. There are no such outliers in the g%1 and g%3 predictions (a01–a03, b01–b07) under GC conditions because of the low number of samples with viability below 92% before storage.

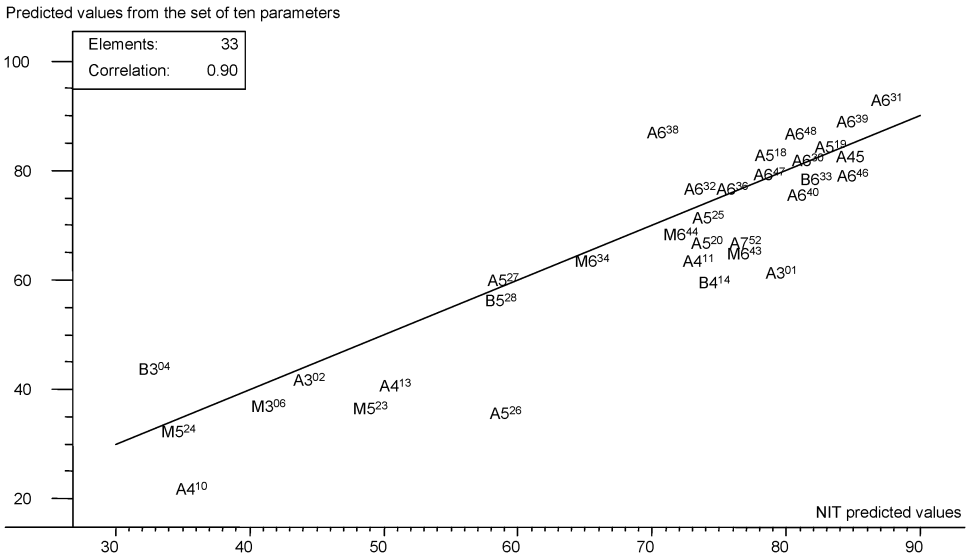


Fig. 4. NIT prediction of "vigour" g%1 GE correlated to the same prediction by the set of ten manifest physical-chemical variables

In order to obtain models, which are more relevant in practice, we have excluded the extreme variety Lysimax in the predictions except for a18–a20 (Steep, Extract and BGwort) from the ten physical-chemical parameters.

Germination parameters

Surprisingly high correlation coefficients are obtained for both "vigour" (g%1) and "viability" (g%3) before and after outlier removal. When comparing the datasets in Table VA and Table VB it is seen for both GC and GE conditions that g%1 (a01 and b01 for GC, a04–a06 and b08–b10 for GE) is better predicted r = 0.73–0.94 with higher correlation coefficients than for g%3 r = 0.39–0.80 (a02 and b02–b03 for GC and a07–a09 and b11–b13 for GE).

A high frequency of low viability among outliers (*GE-conditions*) *was identified after long-term storage with regard to the* g%1 (b10) *and especially with the* g%3 (a08–a09, b12–b13) *predictions. This is in accordance with our previous assumption*[25] *that the physiological condition of the germ should not be able to be predicted by the two physical-chemical screening methods. Removing the low viability outliers improves both types of correlations indicating a firm physical-chemical basis especially for "vigour"* g%1 *of seed samples with reasonable viability* (<92%).

With the set of ten physical-chemical parameters in Table VIA it is possible by Jack-knife validation to register the important variables in each correlation. These variables are ordered in sequence after falling importance in Table VIA. It is seen that two different patterns of these variables arise:

GC: TKW, HI, Volume, Width, Round, Intensity, BG, P (a01–a02),
GE: P, Round, Length, Width, Volume, Intensity (a04–a09).

GE g%3 (a07–a09) has fewer outliers than GE g%1(a04–a06). Germination homogeneity (GH) can be predicted by the set of ten variables (r = 0.70–0.77; a03, a10) with a similar pattern of important variables for GC as given above. It is clear that these differences in importance of the ten different variables as **X** for GC and GE predictions as **y** rests with the germination parameters because the set of ten variables has only been measured once in 1999 together with the malting analyses. We therefore presume that the six seed imaging parameters as well as TKW, protein, β-glucan and Hardness (HI) were not affected by storage at 7°C 13.5% water in three–six years.

This seems reasonable for at least the eight first parameters but should be checked in a future experiment.

To evaluate if the predicted values of "viability" (g%1) by the best models from the two sets of measurements correspond, a diagram is shown where the predicted values for each sample from NIT are plotted as abscissa against those from the ten variables as ordinate. *As is seen from Fig. 4 the correlation coefficient between the two prediction methods is* r = 0.90. *This result strongly supports our hypothesis*[25] *that "vigour"* (g%1) *in malting grade barley can be predicted from the physical-chemical measurements, because the predicted values obtained from two independent measurements correlates.*

Predicting the effect of storage on vigour and viability

It is again surprising to note that both the effect of three–six years storage at 7°C on "vigour" Δg%1 (GCg%1 – GEg%1) and "viability" Δg%3 (GCg%3 – GEg%3) have a profound physical-chemical basis. This is demonstrated by the significant predictions for both the set of ten parameters in Table VIA (a11; r = 0.76 respectively a12; r = 0.71) as well as for NIT VIB (b16; r = 0.80 respectively b17 r = 0.89). It is interesting to note that the parameter hardness (HI) is characterised as an important variable in the predictions a11 and a12.

It should be emphasised that the GC (peroxide treatment method EBC 3.5.2[3]) and the GE comparison (BRF method EBC 3.6.1[3]) is relative and that the respective

Table VIA. PLSR correlations with jack-knife validation between the ten physical-chemical parameters (TKW, HI, P, BG, width, length, area, volume, round, intensity) as **X** and indirect germination and malting variables as **y**. Material 1993–1996 without Lysimax samples (n = 42) in the top of the table. Samples with low viability GE (<92%) = underlined, medium viability GE (92–98%) = **bold**, high viability GE (>98%) = normal.

y	No.	Step*	r	RMSECV	RE	PC**	n	Total outlier samples removed	Significant variables***
g%1 GC	a01	0	0.78	16.33	18.9	1	42		TKW, HI, Volume, Width, Round, Intensity, BG, P
g%3 GC	a02	0	0.63	2.79	18.6	1	42		TKW, HI, P, Width, Round, Volume
GH GC	a03	0	0.77	8.17	17.1	1	42		HI, TKW, Round, Width, Volume, Intensity, Length, P
g%1 GE	a04	0	0.73	17.09	18.1	1	42		P, Round, Length, Width, Volume, Intensity
g%1 GE	a05	I	0.84	12.41	14.5	1	35	M07, A08, A09, A12, **B21**, A37, B41	P, Round, Length, Width, Volume
g%1 GE	a06	I	0.94	7.77	9.1	4	35	M07, A08, A09, A12, **B21**, A37, B41	P, Width, Round, Length, Volume
g%3 GE	a07	0	0.39	13.98	16.5	1	42		P, TKW
	a08	I	0.56	3.60	14.5	1	40	M07, A12	P, HI, TKW
	a09	II	0.73	1.74	14.5	2	39	M07, A12, M16	P, INT, TKW
GH GE	a10	0	0.70	6.17	14.9	2	42		Length, Round, P, Area
Δg%1	a11	I	0.76	14.59	16.3	1	36	A03, A10, A12, M16, **M23**, **M24**	Intensity, TKW, Width, Volume, HI, Round
Δg%3	a12	II	0.71	2.22	14.6	2	37	A03, M07, A10, A12, M16	Volume, Area, TKW, HI
Steep	a13	0	0.60	1.57	11.7	3	42		Intensity
Extract	a14	0	0.62	3.79	15.5	1	52	A03	Round, Length, P, HI, TKW, Area
Extract	a15	I	0.81	1.33	14.3	2	48	A03, M07, A12, M16, **A20**	P, Length, Round, HI, Area, Width
BGwort	a16	0	0.61	65.30	14.1	1	52	A03	P, Width, Intensity
BGwort	a17	I	0.78	39.90	11.4	4	50	A03, M07, A12	P, Width, Intensity
Predictions including the Lysimax samples									
Steep	a18	0	0.82	1.65	12.3	4	62		Intensity, TKW, HI
Extract	a19	0	0.82	1.24	13.3	3	58	A03, M07, A12, M16, **A20**	P, Length, Round, HI, Area, Intensity
BGwort	a20	0	0.69	65.57	23.4	2	63		P, Length, Area, BG, Round, Volume

*Step of outlier selection from influence plot
**Minimum value of residual validation variance
***Variables ordered after degree of importance

g%1 and g%3 values are not fully comparable. However, we have in an unpublished experiment compared GC and GE measurements side by side during another storage experiment with normal and heat damaged barley where dormancy was completely removed after four months. After four–six months of storage there were no significant differences in "viability" (g%3) between the GC and GE conditions while it seemed that the peroxide condition in GC increased "vigour" (g%1) compared to GE, but only for the untreated seeds. There was thus no negative effect on "vigour" (g%1) by the peroxide. The differences (Δ) between the GC and GE conditions given above should thus make sense and merit a more detailed study of the effect of storage on individual samples.

This is done in Fig. 5A and Fig. 5B for "vigour" (g%1) and for "viability" (g%3) respectively where the GC values at harvest are plotted as the abscissa against the GE values as the ordinate. A line drawn on these plots marks unchanged parameters. In Fig. 5A, all 1996 samples (6) are situated in the circle to the right marked 1996 together with a few samples from other years.

It is seen that samples of Blenheim (B) and Alexis (A) from 1994 (4) and 1995 (5) actually increase in "vigour" (Fig. 5A) when stored to 1999, whereas especially samples of Meltan (M) irrespectively of production year are unchanged or decrease during the same period. The cause of the recorded increase in vigour after storage could tentatively be related to a protective storage effect by a high HI of these samples from 1994 (4) and 1995 (5) (Table IVA). Remarkably, the samples from 1996 (6) are all placed below the line, indicating a significant loss in "vigour" after storage from this premium harvest, which also could be seen in Table IIIA. These samples are all very soft (Table IVA).

It is concluded from the results in this experiment that hard seeds store better than soft.

In contrast to our findings regarding "vigour" (g%1), estimated "viability" (g%3) (Fig. 5B) indicates less effect of harvest year on "viability" due to cold storage. There are, however, large differences due to storage of individual samples. The samples of Meltan seem to be more sensitive for decrease in "viability" after storage ($M3^{07}$, $M4^{16}$, $M5^{24}$, $M5^{23}$, $M6^{44}$), whereas a range of samples of Blenheim (B), Alexis (A) and Lysimax (L) from l994 (4) and 1995 (5) actually increase moderately as was registered with the "vigour" comparison. Cold storage of malting barley on a large scale is considered safe[8]. Our finding suggests that storage-sensitive barley batches should be considered before long-term cold storage.

The difference in germination behaviour is expected to be associated with moisture content. When moisture content data are used as ID in Fig. 5B (not displayed here), there is a weak trend showing a few samples where viability is affected by storage ($M3^{07}$, $A4^{12}$, $M4^{16}$, $A4^{10}$, $B3^{04}$) with a higher moisture content in the moderate range of 13.3–15.4% and a decrease in "vigour" after storage. The majority of samples were stored with water content of 11.7–14.0% (13.5 ± 0.8%). It is unlikely that these relatively small increases in water content can fully explain the drastic loss of "viability" of these samples.

The samples that have the largest reduction in "vigour" and "viability" have the lowest malting quality: $A4^{12}$ (Ex-

Table VIB. NIT (1.der.) prediction of germination, malting data and chemical-physical data for samples (1993–1999) of Alexis, Blenheim and Meltan (Lysimax excluded) (n = 52). GE was not available at 97–99 samples. Samples with low viability GE (<92%) = underlined, medium viability GE (92–98%) = **bold**, high viability GE (>98%) = normal.

y	No.	Step*	r	RMSECV	RE	PC**	n	Outliers***
g%1 (GC) 1993–99	b01	0	0.86	12.43	16.4	10	52	
g%3 (GC) 1993–99	b02	0	0.50	2.96	31.2	1	52	
	b03	I	0.76	1.50	15.8	4	49	M17, **B21**, **B28**
g%1 (GC) 1993–96	b04	0	0.92	10.41	12.1	10	42	
g%3 (GC) 1993–96	b05	0	0.59	2.89	19.3	1	42	
	b06	I	0.77	1.56	17.3	2	39	M17, **B21**, **B28**
GH (GC) 1993–96	b07	0	0.78	8.14	17.1	5	42	
g%1 (GE) 1993–96	b08	0	0.74	16.81	17.8	4	42	
	b09	I	0.77	14.94	15.7	4	41	A12
	b10	II	0.80	13.55	14.3	3	38	A12, M16, **A20, A27**
g%3 (GE) 1993–96	b11	0	0.31	14.42	17.0	1	42	
	b12	I	0.68	1.88	15.7	1	39	M07, A12, M16
	b13	II	0.80	0.89	3.4	3	37	B04, M07, A10, A12, M16
GH (GE) 1993–96	b14	0	0.59	6.94	24.4	4	42	
	b15	I	0.75	4.40	17.1	4	37	**B21**, A31, A37, A39, M44
Δg%1	b16	0	0.80	16.64	13.1	8	42	
Δg%3	b17	I	0.89	2.08	7.8	9	40	M07, A12
Steep	b18	0	0.61	1.91	14.2	1	52	
Extract	b19	0	0.73	3.33	13.6	5	52	
BGwort	b20	0	0.77	52.06	12.7	9	52	
TKW	b21	0	0.92	1.43	9.9	10	52	
HI	b22	0	0.94	3.33	7.4	10	49	
P	b23	0	0.97	0.31	5.2	10	52	
Width	b24	0	0.63	0.07	18.4	3	50	
	b25	I	0.74	0.05	13.2	3	49	M17
Length	b26	0	0.77	0.19	16.5	7	50	
Round	b27	0	0.77	0.01	12.5	5	50	
Intensity	b28	0	0.96	1.24	7.4	2	50	

*Step of outlier selection from influence plot
**Minimum value of residual validation variance
***Total outlier samples removed from correlation BGbarley, Area and Volume below r = 0.60

tract: 70.03%, BGwort: 417.33 mg/L), M3[07] (Extract: 64.11%, BGwort: 532.83 mg/L) and M4[16] (Extract: 61.23%, BGwort: 321.27 mg/L).

Cold storage can be seen as a stress treatment, which is likely to affect vigour as does accelerated ageing by short-term heat treatment[26,27,34], however, to a much lesser extent. Estimated "vigour" (g%1) has a much greater response than GI (Tables IIA–B and IIIA–B). For example, the relative decrease in mean "vigour" by cool storage at six years for the barley samples in 1993 was 30% (from 51.3 to 36.1%) compared to 16% for GI (from 7.1 to 5.9).

It is concluded by comparing Fig. 5*A with Fig.* 5*B that the "vigour"* (g%1) *criterion is a more sensitive indicator for effects due to storage than the "viability"* (g%3) *criterion.*

There are marked relative germinative improvements comparing initial GC with GE after storage regarding "vigour" (g%1) *and "viability"* (g%3) *for many of the samples grown in 1994* (*4*) *and 1995* (*5*) *featuring hard seeds with relatively low steep figures. These effects should also be significant in a long-time GE to GE storage comparison.*

The storage sensitivity of the variety Meltan (M) could not be explained by the physical-chemical criteria alone. A slow negative microbial effect on the ability to germinate at these conditions cannot be excluded where a harder more closed structure could have a guarding effect on the germ. Anti-microbial proteins[24] have been discovered in the endosperm e.g. (1→3)-β-glucanase and chitinase that are able to dissolve (1→3)-β-glucan and chitin, which constitutes the cell wall of the fungal hyphae. There are differences between barley varieties in this respect e.g. Lysimax used in this experiment is overproducing the above-mentioned antifungal proteins in large amounts[24].

Investigating the nature of the steep criterion

In this experiment it has only been possible to obtain a low correlation coefficient of r = 0.60–0.61 to predict Steep from the NIT measurements (Table VIB, b18) as well as from the ten physical-chemical parameters (Table VIA, a13). When the samples of Lysimax are included in the model with the ten physical-chemical variables, the correlation coefficient is increased to r = 0.82 (Table VIA, a18). This is not surprising because the samples of Lysimax have a higher average steep compared to the other varieties (Table IIB), so the number of samples with high values will be more balanced to the rest of the samples in this model. However the total range of Steep in the samples (31.4–44.8 in Table IIA) is the same with or without Lysimax, because one sample of Alexis deviates from the rest having a Steep value of 44.8%.

As for Ulonska and Bauner[36], significant correlations were not found between Steep and germination speed g%1 for either GE or GC analyses in this experiment. The significant variables in the steep correlations with the set of the ten parameters (a13 and a18) were Intensity, TKW,

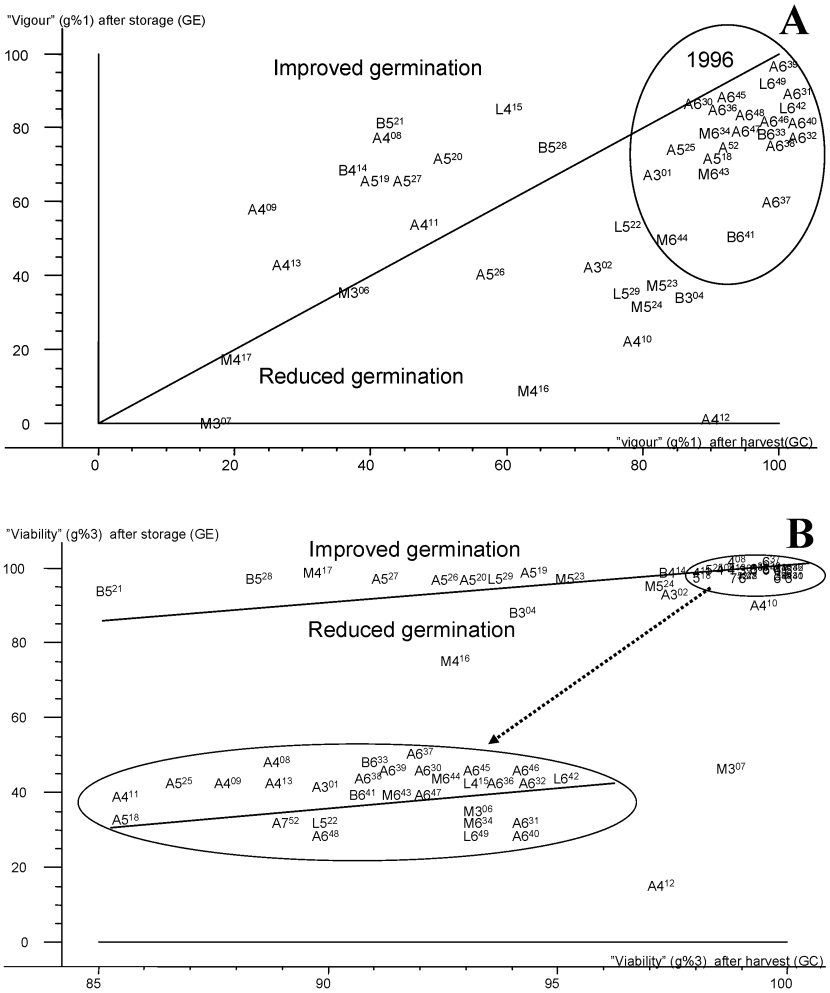


Fig. 5A. "Vigour" g%1 GC at harvest (abscissa) related to "vigour" g%1 GE after storage (ordinate) for samples grown in 1993–1996. **B.** "Viability" g%3 GC at harvest (abscissa) related to "viability" g%3 GE after storage (ordinate) for samples grown in 1993–1996. Abbreviations as mentioned in the Materials and Methods section.

and hardness (HI). The last two were expected[6,9] while the involvement of the intensity (light reflection) parameter could be indicative of a connection to weathering (indicator for dark seeds) related to softening of seeds.

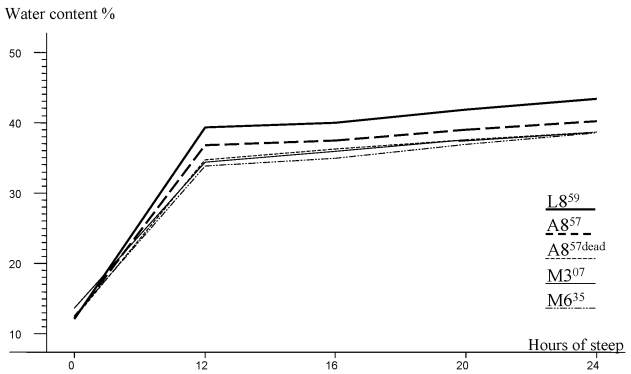


Fig. 6. Water uptake during 24 hours of steep for five samples of Lysimax, Alexis and Meltan. Germination values (g%1, g%3) for GC method: L8^{59} (75.0, 97.5) A8^{57} (79.3, 99.8), A8^{57dead} (0, 0), M3^{07} (12.0, 98.0), M6^{35} (82.0, 99.0).

Optimal steeping conditions are essential if barley's full malting potential is to be realised[6]. Water uptake rates vary between the year of harvest, and between different varieties of barley as seen in Tables IIA–B. The steep character has been found to be influenced by kernel size[6], endosperm structure and composition[18,35] including the amount of mealy or steely grains[6,9]. Ulonska and Bauner[36] concluded that barley varieties require specific combinations of steeping and germination time to reach optimal values of modification and malt quality. These conditions have not been possible to optimize in the present experiment.

In order to make a focused study on the steep parameter and its dependence on the physical-chemical structure of the seed in relation to viability we have therefore taken out five samples from Table I for a separate standardised experiment. The "vigour" (g%1) and "viability" (g%3) values for the five samples are given in the text of Fig. 6.

A sub-sample of Alexis A8^{57} (Table I) was heated to 100°C at low water content (12%) to produce a dead sample for comparison with the untreated one. In Fig. 6 the increase in water uptake (%) from 0 to 24 h of steep

for the five samples is shown involving immersion in dilute peroxide solution (as for steep values in Table IIA–B). It is seen that all samples have a similar water uptake curve form although reaching different steep levels. Lysimax ($L8^{59}$) has the fastest water uptake (ending at 43%) compared to the other samples, confirming the 24 h steeping results given in an overview for all samples in Table IIB. This effect is likely to be partly due to a higher amount of hydrophilic proteins in this mutant, but is also caused by smaller thousand kernel weight (TKW 28.6 g for $L8^{59}$) compared to the other varieties that show decreasing steep percentage with increasing TKW ($A8^{57}$: 33.6 g, $M3^{07}$: 40.8 g and $M6^{35}$: 44.9 g).

The living Alexis sample ($A8^{57}$) had the second largest water uptake and reached water content after 24 h of steep at 40%. This sample had a slightly higher water uptake already after 12 h of steep compared to the same heat-treated sample. This was as expected. It is interesting, however, that the dead sample $A8^{57}$ actually had a surprisingly high water uptake ending at 38% at 24 h, which is comparable with the living Meltan samples.

With respect to the germination properties of the samples (Fig. 6) $L8^{59}$, $A8^{57}$ and $M6^{35}$, all have a "vigour" (g%1) larger than 75% (Table II), whereas $M3^{07}$ has a g%1 at 12% and the dead $A8^{57}$ sample has a g%1 at 0%. The large difference in vigour between the Meltan samples is not influencing their steep profiles, which are almost identical (the small difference could be due to different TKW).

In the present material it is noteworthy (Table IVA) that the dry harvest year 1995 that produced very hard kernels (HI = 65.8) compared to the favourable year 1996 (HI = 41.7) also had the lowest mean steep value (35.6%) compared to 37.3% in 1996 (Table IIA). Likewise 1995 had lower germination homogeneity (GH = 49.9) than 1996 (GH = 71.9). The evidence given above indicates that differences between varieties in the physical-chemical properties of the seed endosperm are important for determining the steep (water uptake) which in the start is a passive process rather independent of "vigour" and "viability" as indicated in the special experiment in Fig. 6. However, the variation in the steep parameter kept at constant steeping time in the main material in this investigation is rather low if one removes the extreme non malting variety Lysimax, which is an exceptional outlier combining hard seeds with high steep. It is concluded that the multivariate aspects of the steep parameter should be studied with a barley material with variable hardness where the steep level was optimized for each sample under more realistic experimental conditions compared to the above experiment.

Malting parameters

Using the ten physical-chemical parameters to predict Extract and BGwort it is seen, that by removing up to five outliers (most with a low "viability") it is possible to obtain a correlation coefficient of $r = 0.81$ for Extract (two PC's) and $r = 0.78$ for BGwort (four PC's) (Table VIA, a14–a17). These are slightly better predictions than using NIT measurements ($r = 0.73$ and $r = 0.77$ respectively, b19–b20), however, with no outliers removed (Table VIB). The number of PC's (five–nine) were much higher for the NIT correlation indicating a higher level of complexity. The significant parameters in Table VIA for Extract were P, Length, Round, HI and Area confirming the relationships between these parameters which was discussed in the PCA classification in Fig. 2A–B demonstrating the usefulness of the multivariate stepwise approach in visualising data to get indications for providing predictions.

Short-circuiting the two data sets by PLSR predictions of each of the set of ten variables to NIT spectra

As indicated by comparing the classifications of the PCA's of the separate datasets in Fig. 2A–B and Fig. 3A–B there are resemblances, which are further, strengthened in the PLSR predictions in Tables VIA–B. The final proof that NIT spectroscopy represents a non-destructive fingerprint of the whole barley seed sample is demonstrated by the high NIT predictions of each of the set of the ten physical-chemical parameters in Table VIB. Especially Protein ($r = 0.97$), Intensity ($r = 0.96$), HI ($r = 0.94$) and TKW ($r = 0.92$, 10 PC's indicating high complexity) show high correlation coefficients (while the seed imaging parameters Round, Length, Width and Intensity vary from $r = 0.74$ to $r = 0.96$ with a lower number of PC's (two–seven). It is interesting to note that the secondary parameters Area and Volume calculated from the three other primary parameters (Length, Width, Round) by the software in the GrainCheck instrument are not significantly correlated to NIT data. The β-glucan parameter cannot be predicted by NIT spectroscopy or by the set of ten parameters in this investigation when Lysimax is excluded because of the low variation of β-glucan in this limited material (Table VB). It can therefore be concluded that a large NIT calibration sample set has to be collected which shows a broad diversity in every parameter of importance, if the *on-line* NIT technology should be able to be extended in practice from protein and water to predict other parameters such as g%1, BGbarley, HI, Extract and BGwort.

CONCLUSIONS

Multivariate data analysis classification with the algorithm PCA is a strong tool for obtaining a preliminary overview of associations between different variables and individual samples (Fig. 2A–B and Fig. 3A–B). For further verification and focusing, PLSR multivariate prediction of specific parameters is essential (Table VI). The results from the two independent screening analyses (The set of ten variables and NIT) to estimate the physical-chemical status of the germinative and malting parameters support our earlier conclusions[25] regarding the usefulness of germinative classification plots with g% as abscissa and g%3 as ordinate for malt quality prediction. Thus the indirect parameters such as germination and malting properties are heavily dependent on the manifest multivariate physical-chemical properties in barley which may be used for prediction. This is only possible by multivariate pattern recognition analysis.

The surprisingly high correlation coefficients obtained using NIT can obviously be explained by high prediction values by NIT of each of the set of the ten physical-chemical variables discussed above.

It is concluded that germination characteristics and malt quality are influenced by two major functional factors – the physiological viability of the germ and the physical-chemical structure of the endosperm. The first factor is assessed by removing the low vigour outliers in the correlations by the PLSR influence plot[16,25]. The second factor is related to "vigour" (g%1) by limiting the required substrate to the germ for germination and growth of the plantlet in live seeds as suggested by our research group[25].

It should be emphasised that the prediction models in Table VIA–B only consider the physical-chemical endosperm properties underlying "vigour" (g%1) and "viability" (g%3) and cannot predict the physiological state of the germ in a test set of barleys with unknown germination properties. Thus "vigour" (g%1) cannot be used alone in order to judge germinative quality but is dependent on g%3 or the tetrazolium test for viability as outlined in the two-dimensional germinative classification plot[25]. A single score value combining the g%1, g%2 and g%3 parameters such as in GI is not sufficient[25]. The results from this investigation explain in more detail the advantage of the multivariate basis of the germinative classification. It is clear that the physical-chemical structure in malting barley of normal malt quality is the fundamental and dominating factor for predicting germination speed "vigour" (g%1) by NIT. Therefore NIT spectroscopy is feasible for a preliminary screening *on-line* for potential vigour of samples which should be checked for viability.

The vigour criterion is a more sensitive indicator for storage effects than viability. The same conclusion applies to modelling critical criteria such as Extract and BGwort where vigour plays a central role.

This study suggests that the physical-chemical seed structure such as hardness/softness also plays a major role in regulating the water uptake (steep) and in protecting the germ during long time storage. These results should be further verified by an experimental multivariate data analytical approach.

The "viability" (g%3) trait besides information on the physiological dimension also carries a substantial physical-chemical component as demonstrated by the significant correlation coefficients for NIT in Table VIB, however, with lower values than those for "vigour" (g%1) and with a greater number of outliers. Besides the tetrazolium test, g%8 should be selected in future basic research as a more clean parameter for viability better reflecting the physiological status of the germ than g%3.

It is expected that a barley data bank[25] covering a large variation in physical-chemical composition, with a wide range of varieties grown in different years and climatic regions, will be able to reduce the relative high errors of the statistically significant prediction models obtained in this investigation made with a limited number of samples.

The establishment of an extensive source of semi-artificial intelligence by a range of representative calibration models is the prerequisite for employing non-destructive NIT spectroscopy as a first *on-line* selection criterion[25] for "vigour" and malting quality of barley, where the best deliveries with regard to "vigour" (g%1) can be checked for viability by the tetrazolium test *at-line* as a second selection within two hours.

ACKNOWLEDGEMENTS

This paper is a part of a Ph.D. thesis in Food Technology at the Royal Danish Veterinary and Agricultural University supervised by Professor Lars Munck to whom I am grateful for enthusiastic and inspiring discussions.

The Swedish Seed Company "Skånska Lantmännen" through Lennart Wikström at the SL-Foundation and the Danish Cereal Network are greatly thanked for initiating and further financing this study. For initiating and financing the steep investigation, Thomas Börjesson, SL-Odal are greatly thanked. I am very grateful to Monica Frank and the rest of the staff at Agrolab, Eslöv for collecting and analysing the barley samples from the 1993 harvest according to germination capacity. "Fyens Stifts Patriotiske Selskab", "Skånska Lantmännen" and the staff at "Højbakkegård" are greatly thanked for helping with field experiments. Also Kirsten Wilms, Anna-Marie Hansen and Lisbeth Hansen are greatly appreciated for their crucial assistance in analysing the samples. A great thanks goes to Kees Van Haaren, Pajbjerg Plant Breeding Station for performing the micro malting and malt analysis. I am also grateful to Ellen Jensen, Carlsberg Research Center for helping with references and to Gilda Kischinovsky for her comments on the article.

REFERENCES

1. Agu, R.C. and Palmer, G.H., The effect of nitrogen level on the performance of malting barley varieties during germination, *Journal Institute of Brewing,* 2001, **107**(**2**), 93–98.
2. Andersson, A., Elfverson, C., Andersson, R., Regner, S. and Åman, P, Chemical and physical characteristics of different barley samples, *Journal of the Science of Food and Agriculture,* 1999, **79**, 979–986.
3. Anonymous, European Brewery Convention: Analytica-EBC, Fachverlag Hans Carl: Nürnberg, 1998.
4. Anonymous, Standard methods of the International Association for Cereal Chemistry (ICC) ICC-standards, 1998, Verlag Moritz Schäfer: Detmold.
5. Bamforth, C.W. and Barclay, A.H.P., Malting technology and the uses of malt in barley. In: Chemistry and Technology. A.W. MacGregor and R.S. Bhatty, Eds., American Association of Cereal Chemists: St. Paul, MN, USA, 1993, pp. 297–354.
6. Briggs, D.E., Some aspects of barley grain physiology. In: Malting and Brewing Science, Briggs, D.E., Hough, J.S., Stevens, R. and Young, T.W., Eds., Chapman and Hall: London, Vol. 1, 1995, pp. 39–57.
7. Broadbent, R.E. and Palmer, G.H., Relationship between β-amylase activity, steeliness, mealiness, nitrogen content and the nitrogen fractions of the barley grain, *Journal Institute of Brewing,* 2001, **107**(**6**), 349–354.
8. Brunner, H., Cold preservation of malting barley, *Brauwelt Int.,* 1995, **13**(**5**), 420–424.
9. Chandra, G. S., Proudlove, M. O. and Baxter, D. E., The structure of barley endosperm – An important determinant of malt modification, *Journal of the Science of Food and Agriculture,* 1999, **79**, 37–46.
10. Jacobsen, T., New aspects of chemical analysis: Chemometrics – variations in trace elements in malt as shown by factor analysis, *Brygmesteren,* 1980, **6**, 149–158.
11. Jacobsen, T., Malting technology and properties of malt: cluster analysis study, *Journal Institute of Brewing,* 1986, **92**, 177–180.
12. Kelly, L. and Briggs, D.E., Relationships between the duration of steeping, grain microbes, grain maturity and the response of de-embryonated grains to gibberellic acid, *Journal Institute of Brewing,* 1993, **99**, 57–61.
13. Kitamura, Y., Yamada, K. and Yumoto, T., The initial absorption of water and the manifestation of physiological activities by barley kernels, *Monatsschrift für Brauwissenschaft,* 1990, **6**, 216–220.

14. Lonkhuijsen, H.J., van Douma, A.C. and Angelino, S.A.G.F., Evaluation of a malting barley quality assessment system, *J. Am. Soc. Brew. Chem.,* 1998, **56**(**1**), 7–11.
15. Martens, H. and Næs, T., Multivariate Calibration, Wiley: New York, 1989.
16. Martens, H. and Martens, M., Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR), *Food Quality and Preference,* 2000, **11**, 5–16.
17. Meurens, M. and Yan, S. H., Applications of vibrational spectroscopy in brewing. In: Handbook of vibrational spectroscopy. J. M. Chalmers and P. R. Griffiths, Eds., John Wiley & Sons: Chichester, UK, 2002, **5**, pp. 3663–3671.
18. Molina-Cano, J.L., Francesch, M., Perez-Vendrell, A.M., Ramo, T., Voltas, J. and Brufau, J., Genetic and environmental variation in malting and feed quality of barley, *Journal of Cereal Science,* 1995, **25**, 37–47.
19. Molina-Cano, J.L., Ramo, T., Ellis, R.P., Swanston, J.S., Bain, H., Uribe-Echeverria, T. and Perez-Vendrell, A.M., Effect of grain composition on water uptake by malting barley: a genetic and environmental study, *Journal Institute of Brewing,* 1995, **101**, 79–83.
20. Monnez, J.M., An approach to the estimation of brewing quality in barley and malt. *Journal Institute of Brewing,* 1987, **93**, 477–486.
21. Munck, L., Quality criteria in the production chain from malting barley to beer – some views on the logistic problems with a few examples (dormancy, vitality and β-glucans), *Ferment,* 1991, **4**(**4**), 235–241.
22. Munck, L., The case of high lysine barley breeding. In: Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, P.E. Shewry, Ed., CAB International: Great Britain, 1992, pp. 573–602.
23. Munck, L., A new holistic exploratory approach bridging phenotype and genotype. In: Diversity in Barley (*Hordeum vulgare*), R. von Bothmer, Th.J.L. van Hintum, H. Knupffer and K. Sato, Eds., Elsevier Science: Amsterdam, 2003, Chapter 11, 227–245.
24. Munck, L. Aspects on the regulation of dormancy and vitality in barley seeds and the possible relation between vitality and antimicrobial proteins. In: European Brewery Convention Monograph XV, Symposium on Plant Biotechnology, Helsinki, Fachverlag Hans Carl: Nürnberg, 1989, pp. 199–129.
25. Munck, L and Møller, B. A new germinative classification model of barley for prediction of malt quality amplified by a near infrared transmission spectroscopy calibration for vigour "on-line" both implemented by multivariate data analysis, 2004, *Journal Institute of Brewing,* **110**(**1**), 2–17.
26. Møller, B. and Munck, L., Seed vigour in relation to heat sensitivity and heat resistance in barley evaluated by multivariate data analysis, *Journal Institute of Brewing,* 2002, **108**(**3**), 286–293.
27. Møller, B., Molina-Cano, J.L. and Munck, L., Variation in malting quality and heat resistance in the malting barley variety "Alexis", *Journal Institute of Brewing,* 2002, **108**(**3**), 294–302.
28. Nielsen, J.P. and Munck, L., Evaluation of malting barley quality using explorative data analysis. I. Extraction of information from micro-malting data of spring and winter barley. *Journal of Cereal Science,* 2003, **38**, 173–180.
29. Nielsen, J.P., Evaluation of malting barley quality using explorative data analysis. II The use of kernel hardness and image analysis as screening methods, *Journal of Cereal Science,* 2003, **38**, 247–255.
30. Nischwitz, R., Cole, N.W. and MacLeod, L., Malting for brewhouse performance, *Journal Institute of Brewing,* 1999, **105**(**4**), 219–227.
31. Reiner, L. Die Faktorenanalyse in der Brautechnologische Forschung, *Brauwissenschaft,* 1971, **24**, 403–412.
32. Riis, P. and Bang-Olsen, K., Breeding high vigour barley – a short cut to fast maltings, Proceedings of the European Brewery Convention Congress, Lisbon, IRL Press: Oxford, 1991, pp. 101–108.
33. Riis, P., Meiling, E. and Peetz, J., Determination of germinative percentage and germination index – collaborative trial and ruggedness testing, *Journal of the Institute of Brewing,* 1995, **101**(**3**), 171–175.
34. Spears, J.F., An introduction to seed vigour testing. In: Seed vigour testing, Proceedings from 24th Congress of the International Seed Testing Association, H.A. van de Venter, Ed., International Seed Testing Association, Copenhagen, 1995, pp. 1–9.
35. Swanston, J.S. and Taylor, K., The effects of different steeping regimes on water uptake, germination rate, milling energy and hot water extract. *Journal of the Institute of Brewing,* 1990, **96**(**1**), 3–6.
36. Ulonska, E. and Baumer, M., Investigation of the value of water uptake and germination for the estimation of malting quality in barley, Proceedings of the Third International Barley Genetics Symposium, Garching, Verlag Karl Thiemig: Münich, Germany, 1975, pp. 579–593.
37. Zhang, G., Chen, J., Wang, J. and Ding, S., Cultivar and environmental effects on (1-3,1-4)-β-D-glucan and protein content in malting barley, *Journal of Cereal Science,* 2001, **34**, 295–301.
38. www.models.kvl.dk Chemometric and Spectroscopy group, Food Technology, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Denmark, website for information on chemometrics and spectroscopy.

(Manuscript accepted for publication December 2003)

ELSEVIER

Analytica Chimica Acta 446 (2001) 171–186

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics

L. Munck [a,*], J. Pram Nielsen [a], B. Møller [a], S. Jacobsen [b,1], I. Søndergaard [b,1], S.B. Engelsen [a], L. Nørgaard [a], R. Bro [a]

[a] *Chemometrics Group, Food Technology, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark*

[b] *Section of Biochemistry and Nutrition, The Technical University of Denmark, Søltofts Plads, Building 224, DK-2800 Kgs. Lyngby, Denmark*

Received 7 November 2000; received in revised form 22 March 2001; accepted 4 April 2001

Abstract

Evaluating gene effects on proteomes and the resulting indirect pleiotropic effects through the cell machinery on the chemical phenotype constitutes a formidable challenge to the analytical chemist. This paper demonstrates that near-infrared (NIR) spectroscopy and chemometrics on the level of the barley seed phenotype is able to differentiate between genetic and environmental effects in a PCA model involving normal barley lines and the gene regulator *lys3a* in different genetic backgrounds. The gene drastically changes the proteome quantitatively and qualitatively, as displayed in two-dimensional electrophoresis, resulting in a radically changed amino acid and chemical composition. A synergy interval partial least squares regression model (si-PLSR) is tested to select combinations of spectral segments which have a high correlation to defined chemical components indicative of the *lys3a* gene, such as direct effects of the changed proteome, for example, the amide content, or indirect effects due to changes in carbohydrate and fat composition. It is concluded that the redundancy of biological information on the DNA sequence level is also represented at the phenotypic level in the dataset read by the NIR spectroscopic sensor from the chemical physical fingerprint. The PLS algorithm chooses spectral intervals which combine both direct and indirect proteome effects. This explains the robustness of NIR spectral predictions by PLSR for a wide range of chemical components. The new option of using spectroscopy, analytical chemistry and chemometrics in modeling the genetically based covariance of physical/chemical fingerprints of the intact phenotype in plant breeding and biotechnology is discussed.

Keywords: Proteome; Near-infrared spectroscopy; Chemometrics; Analytical chemistry; Biotechnology; Barley; Plant breeding

1. Introduction

Spectroscopy has become fundamental in chemistry. Its discovery dates back to 1800 when the British astronomer William Herschel reported on the existence of "the invisible thermometrical spectrum" to the Royal Society [1]. However, spectroscopy first gained momentum when Abney and Festing in 1881 [2] first measured spectra of organic compounds. Since then, a number of highly informative spectroscopic techniques have been developed. One of the more recent developments is near-infrared (NIR) spectroscopy which has invaded analytical chemistry by

* Corresponding author. Tel.: +45-35-28-33-58; fax: +45-35-28-32-45; URL: http://models.kvl.dk.
E-mail address: lmu@kvl.dk (L. Munck).
[1] URL: www.be.dtu.dk.

0003-2670/01/$ – see front matter
PII: S0003-2670(01)01056-X

non-destructively predicting chemical composition, even in complex biological samples from agriculture [3]. These methods are now routinely used with high precision locally and internationally for controlling the composition of agricultural raw materials for feed and food such as cereals, milk, and meat.

It all started in the 1950s at the USDA laboratory in Beltsville, Maryland, USA, when Karl Norris and his group constructed a moisture metre [4] for wheat by NIR spectroscopy. The problem of interference in the moisture measurements from other chemical constituents was solved by data pretreatment through spectral derivatization and classical statistical wavelength selection by regression analysis made possible by the computer followed by multiple linear regression (MLR) to calibrate a few selected wavelengths to water measurements.

At the 7th World Cereal and Bread Congress in 1982, Martens and Jensen [5] introduced chemometric algorithms in utilizing information from whole NIR spectra. This was in the form of the partial least squares regression (PLSR) which later became fundamental in the development of software for NIR equipment dedicated for specific functions made by the instrumental industry.

In plant breeding, rapid screening methods for chemical composition and identification of specific genes are essential tools in classical breeding as well as in gene biotechnology. Recently, NIR spectroscopy has been shown to be able to detect the phenotypic effects of wheat–rye chromosomal translocations [6], and the chemical mechanism behind this classification was discussed.

In the 1960s, research in cereals was focused on obtaining genes for improved amino acid composition for nutritional purposes, especially with regard to the first limiting amino acid lysine [7]. The senior author of this paper was involved in developing a dye-binding method with acilane orange as an expression for the sum of the basic amino acids — lysine, arginine, and histidine — which was used as a ratio to protein (N × 6.25) to select the first *high-lysine* barley gene *lys1* from the world barley collection [8]. A more drastic ethylenimine-induced mutant M-1508 (gene *lys3a*) from the barley variety Bomi was isolated in 1973 by the Risø laboratory group [9] in Denmark employing the dye-binding method. The regulatory status of the Mendelian high-lysine gene *lys3a* was finally established in 1996 [10]. In a recent study, we have demonstrated that NIR spectroscopy is able to differentiate between five different high-lysine mutant genotypes [11] with characteristically different amino acid patterns.

In an autopollinated crop such as barley, spectroscopic screening is greatly facilitated analytically, because each line derived after six to nine generations of self-pollination can be considered homozygotic and thus genetically homogeneous. A mutation or a transfer of a specific gene to such a line is thus expressed in a genetically reproducible, controlled isogenic background and will show up in the spectroscopic physical/chemical fingerprint, if its chemical implications directly or indirectly are large enough. In development of NIR analytical methods, the applications have always been ahead of theory. We therefore aim at exploring the chemical basis of how NIR spectroscopy works in differentiating the *lys3a* phenotype from normal barley in two different environments. The genetically based diversity is first detected on the phenotypic level of biological organization by non-invasive spectroscopy and afterwards calibrated to destructive chemical analytical methods. This dialogue between data from the biological and chemical levels of organization is made possible by chemometric software and the computer.

2. Materials and methods

Total 125 different varieties of normal barley (O) and *lys3a* (X) lines based on crosses with these varieties were bred at the Carlsberg Research Laboratory, Valby, Copenhagen from 1973 to 1990, as represented in the figures and tables. They were grown in the field and/or in the greenhouse (V) together with the original *lys3a* mutant from Risø (M-1508) and its isogenic motherline Bomi. Seeds grown in the greenhouse tended to have a lighter color and higher protein content compared to those grown in the field. The whole seed barley samples were measured with a near-infrared transmission (NIT) instrument, Infratec (Foss Tecator AB, Høganäs, Sweden). The samples were milled in a hammermill (sieve 0.5 mm) and the whole flour was measured by a NIR instrument (Foss-NIR-Systems 6500, USA). The samples were analyzed for moisture, Kjeldahl protein (N × 6.25)

(Foss Tecator, Kjeltec) and for amides by alkali volatile nitrogen by adding 50 ml of 36% NOH alkali to 3 g of flour in the Kjeldahl destillation unit.

Twenty-one of these samples presented in figures and tables were also analyzed for starch (AACC 76-13), for fat (Foss Tecator, Soxtech), for β-glucan (Foss Tecator β-glucan analyzer system, Carlsberg), for soluble and insoluble fiber (Foss Tecator, Fibertech), and for amino acids after hydrolysis [12]. The 21 barley samples were extracted in order to obtain buffer-soluble (albumins, globulins) and ethanol-soluble (hordeins) proteins which were separated by a two-dimensional gel electrophoresis [13]. Principal component analysis (PCA) and PLSR analyses were performed by the "Unscrambler" software version 7.5 (Camo A/S Trondheim, Norway) with full cross-validation. The spectroscopic data were reduced by 50% by selecting information from every second wavelength with full cross-validation.

Interval-PLS (i-PLS) employed in the selection of wavelength areas [14] was performed on the 1050 NIR spectral data points divided into 30 equal intervals numbered 1–30, stating correlation coefficients (r) and error (RMSECV: root mean square error of cross-validation). An extension of i-PLS called synergy i-PLS (si-PLS) was employed to find the interval combinations of all possible combinations of intervals which give the highest correlation coefficients and the lowest errors.

3. Results from observations and experiments

3.1. Exploratory classification of separate spectral and chemical datasets by PCA

We will first investigate the non-destructive observation of batches of whole barley seeds by NIT spectroscopy and its ability to differentiate between *lys3a* and normal phenotypes. In Fig. 1A, second derivative NIT spectra between 860 and 1035 nm from 51 barley samples grown in the field are shown. We will now simulate the discovery of the *lys3a* gene, as it could have taken place with NIT spectroscopy instead of the dye-binding method [6–8]. In Fig. 1B, a PCA of 51 barley NIT spectra displays a normal barley population (O) and a *lys3a* mutant outlier (X). When this mutant is crossed with different normal barley genotypes, the segregants form two clusters (Fig. 1C) representing normal barley (O) and *lys3a* mutant (X) recombinants.

It is obvious that the non-destructive NIT spectroscopy on whole seeds is an attractive method for selection in plant breeding and that it is able to differentiate between the two extreme genotypes with a sufficient degree of precision. However, the short spectral range of 175 nm of NIT is mainly due to the third and fourth and partly the second overtones in the lower range of the near-infrared spectrum as limited by the silicon sensor. Further development of NIT spectroscopy with new sensors will reveal how far this technology can be expanded upwards at higher wavelengths to obtain less crude and more detailed spectra as with near-infrared reflection (compare Fig. 1A with D). Expanding upwards in the near-infrared spectrum, will give more specific chemical information, including from the combinatory region from 1900 nm and upwards [15]. Therefore, in order to explore these possibilities, we have chosen in the following to concentrate on NIR spectroscopy with photomultipliers measuring 1050 data points at every second wavelength from 400 to 2500 nm (Fig. 1D). This however introduces the drawback of having to mill the barley seed samples.

In Fig. 2 in a PCA plot with 125 NIR spectra constituting the whole barley material, we can identify four clusters where PC 1 differentiates between the genotypes normal (O) and *lys3a* (X) barley, while PC 2 differentiates between the barley grown in the field (O and X) and in the greenhouse (OV and XV). With few exceptions, the genetic differentiation is excellent. It would probably have been even better, if samples of the mutant had been compared with several samples of a normal barley with the same isogenic background.

Total 21 of these samples, 15 normal and 6 *lys3a* lines, were subjected to a detailed chemical analysis, including two two-dimensional gel electrophoresis analyses with a buffer and an ethanol extract of each sample in order to study the water- and salt-soluble albumins and globulins as well as the ethanol-soluble storage proteins, the prolamins (hordeins). By visual inspection, it was clearly possible to classify the two electrophoresis patterns of each of the pure samples of the *lys3a* genotype as different from those of the normal barleys. Two representative sets of electrophoresis, each for water/salt-soluble and ethanol-soluble proteins for the original 1508 gene a *lys3a* and the

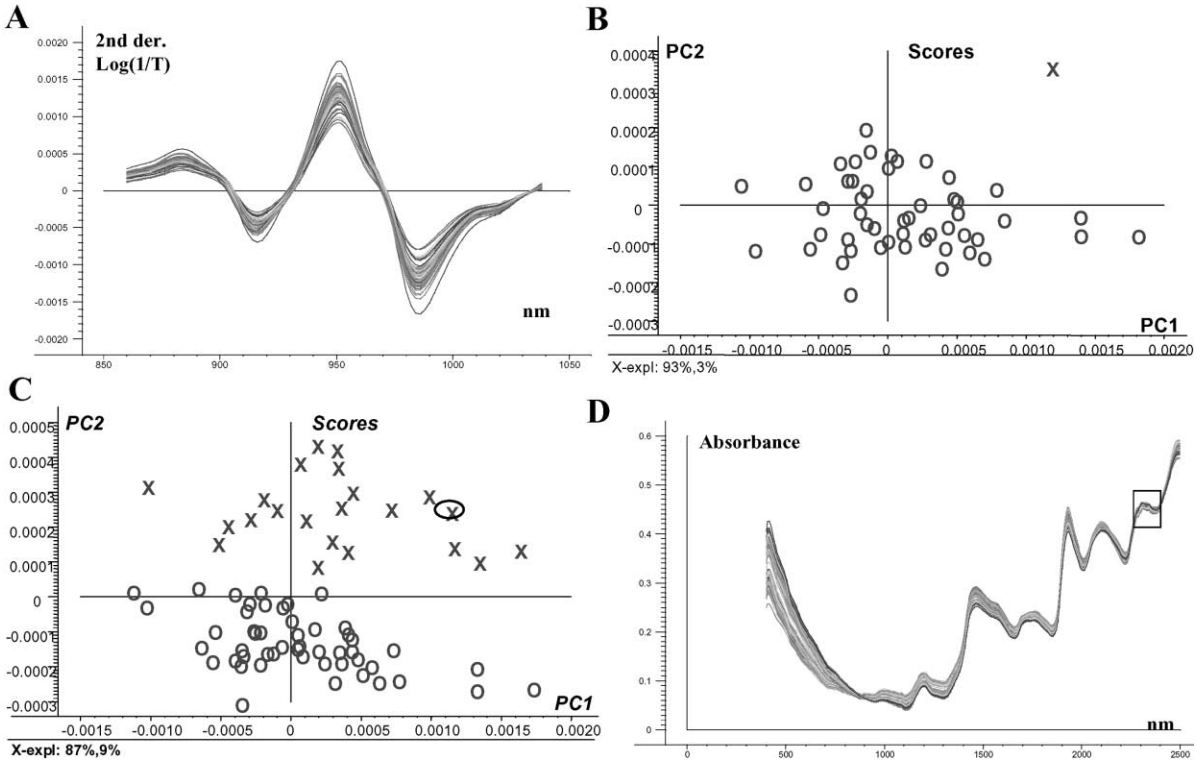


Fig. 1. (A) NIT spectra (850–1040 nm) measured non-destructively on whole seed samples of normal and *lys3a* barleys grown in the field presented as PCA plots in (B–D); (B) PCA of whole seed NIT spectra detecting a *lys3a* (mutant 1508) outlier (X) among normal barleys (O); (C) PCA of whole seed NIT spectra showing a segregating population for the *lys3a* gene (X) where the original 1508 mutant (encircled) has been crossed with normal barley (O); (D) 125 NIR spectra (400–2500 nm) from whole flour of milled seeds of normal and *lys3a* barleys shown in a PCA plot in Fig. 2. The squared area 2270–2360 nm (approximately identical with the i-PLS interval number 28) visually selected for difference between normal and *lys3a* spectra is presented enlarged in Fig. 6.

isogenic barley line Bomi are displayed in Fig. 3. The different proteomes give a good, general overview of the very high chemical complexity on the protein level which fuels the phenotypical variation discussed above.

The hordein patterns from a normal barley variety such as Bomi compared with those of the high-lysine mutant *lys3a* in Bomi shows that the normal line expresses a number of proteins that the mutant does not express (Fig. 3a and b). Contrary to this finding, the albumin/globulin fraction shows that the mutant line expresses a number of proteins that Bomi does not express (Fig. 3c and d). In the albumin/globulin fraction, the picture is rather complicated. A general conclusion about whether these differences are of a quantitative or a qualitative nature merits further studies during which the electrophoresis separation is optimized.

The changes in the proteome of the barley endosperm due to the *lys3a* gene are reflected in a drastic change in the amino acid composition of the total protein (see Table 1). The basic amino acids, including lysine, are increased together with aspargine, alanine, threonine, and valine, while glutamic acid, proline, amide nitrogen to protein nitrogen (A/P index) (see Table 2) and phenylalanine are markedly decreased. The change in the amino acid pattern of the 21 samples is expressed in the PCA biplot with scores and loadings in Fig. 4. There is a clear differentiation between *lys3a* (X) and normal barley samples (O). However, two of the samples, X and XV, are more intermediate, emerging from the same 508 line of the

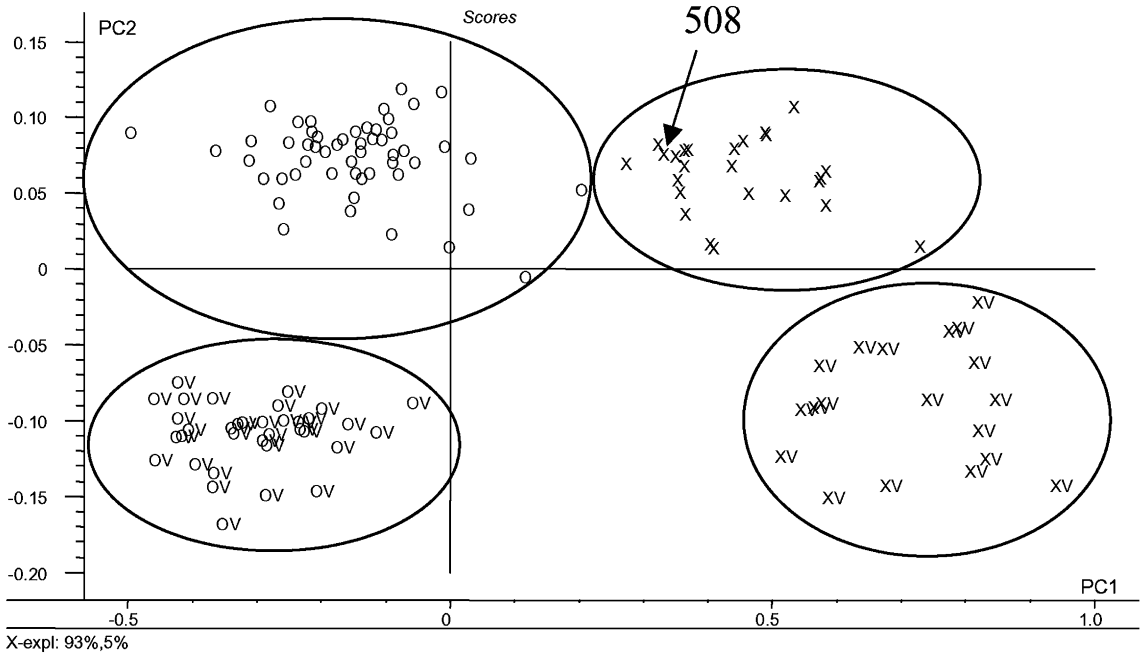


Fig. 2. PCA on NIR spectra from Fig. 1D from normal (O) and *lys3a* (OX) barleys grown in the field and in the greenhouse, OV and XV, respectively.

Carlsberg breeding program. The loadings from the different amino acids appear in three groups. On the right side of the PCA plot near the *lys3a* barleys (X and XV), there is a cluster of amino acids such as lysine and aspartic acid which are increased in this genotype. Situated above this are the amino acids with minor changes between the two genotypes, while to the left near to the normal (O and OV) samples are the amino acids like glutamic acid, proline, and A/P which are high in the normal genotypes. It is seen that the normal barleys grown in the greenhouse (OV) are near this amino acid cluster, because they have higher amounts of the amino acids typical for storage proteins like glutamic acid and proline compared to

Table 1
Selected barleys for amino acid determination (g per 100 g protein) from the material in Fig. 2

	Normal barley ($n = 15$)	Barley 1508 ($n = 4$)
Asp	4.78 ± 0.50	8.11 ± 0.57
Arg	4.51 ± 0.50	6.97 ± 0.42
Lys	3.10 ± 0.39	4.84 ± 0.28
Gly	2.93 ± 0.33	4.30 ± 0.32
His	v2.12 ± 0.21	2.76 ± 0.14
Tyr	2.81 ± 0.26	2.93 ± 0.22
Ala	3.07 ± 0.36	4.23 ± 0.28
Ser	3.17 ± 0.28	3.49 ± 0.33
Thr	2.63 ± 0.27	3.28 ± 0.30
Val	4.32 ± 0.45	4.92 ± 0.37
Met	1.43 ± 0.15	1.53 ± 0.14
Cys	1.56 ± 0.28	1.60 ± 0.41
Leu	5.98 ± 0.59	5.85 ± 0.49
Ile	3.12 ± 0.28	2.99 ± 0.19
Glu	21.62 ± 2.51	14.60 ± 1.69
Pro	9.63 ± 1.34	5.76 ± 0.64
Phe	4.51 ± 0.50	3.67 ± 0.22

Table 2
Chemical composition (% DM) of barleys for the same material selected for amino acid composition in Table 1

	Normal barley ($n = 15$)	Barley 1508 ($n = 4$)
Protein (N × 6.25)	12.76 ± 2.38	14.28 ± 1.84
Amide-N	0.32 ± 0.08	0.25 ± 0.05
Amide-N/N (A/P)	15.68 ± 1.08	11.53 ± 1.98
Beta-glucan	4.76 ± 0.77	3.88 ± 1.19
Fat	1.88 ± 0.18	3.22 ± 0.53
Starch	54.83 ± 4.18	50.95 ± 3.03
Insoluble fiber	10.91 ± 1.64	16.47 ± 0.91
Soluble fiber	2.87 ± 0.67	2.30 ± 0.69

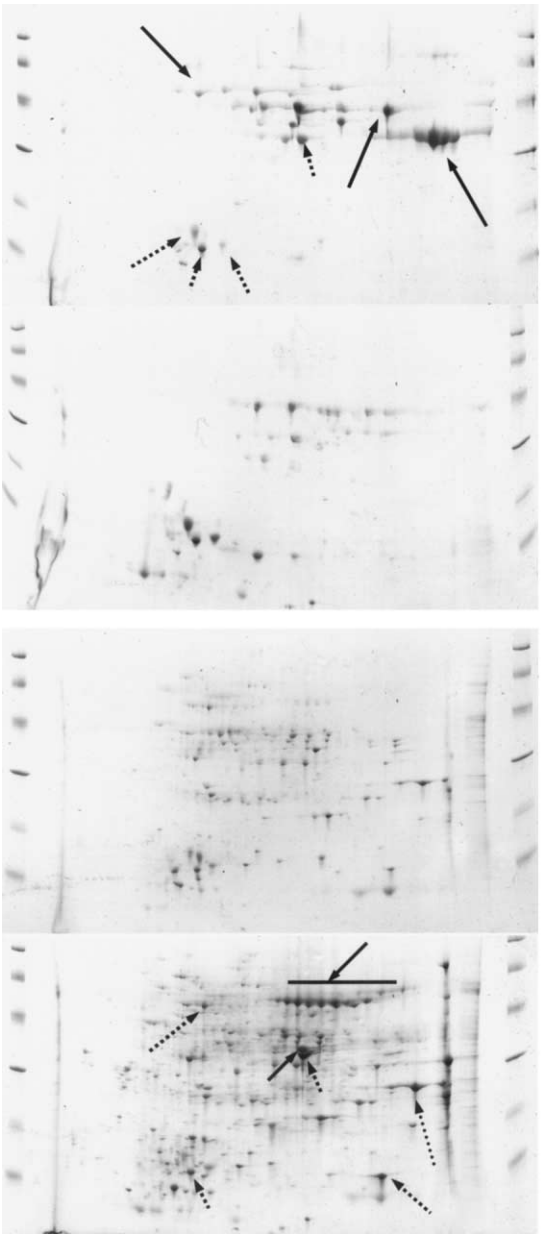


Fig. 3. Two-dimensional electrophoretic gels [13] of hordeins and albumins/globulins from the barley variety Bomi and its *lys3a* mutant. The gels were run using an immobilized pH gradient from 3 to 10 in the first dimension: (a) hordeins from Bomi; (b) hordeins from its *lys3a* mutant; (c) albumins/globulins from Bomi; (d) the albumins/globulins from its *lys3a* mutant. Some protein spots can be seen in the gels both from Bomi and the mutant, a subset of these common spots being indicated with dashed arrows for orientation. The full arrows indicate a subset of proteins that are only present in either Bomi or *lys3a*.

those grown in the field (O). This is caused by the higher protein content of the samples grown in the greenhouse (V) due to intensive nitrogen fertilization which especially increases the alcohol-soluble storage proteins, the hordeins.

Simply inherited Mendelian regulating genes like *lys3a* have complicated, indirect effects on the phenotype when the changed protein pattern influences the total endosperm cell machinery. This is reflected in a change in the total chemical composition with increases in *lys3a* barley for protein, fat and insoluble fiber and decreases in amide-N, beta-glucan, starch, and soluble fiber (Table 2 and Fig. 5). The PCA biplot of the total chemical composition in Fig. 5 parallels that of amino acid composition in Fig. 4, elucidating the different patterns in chemical composition due to genotype and growth environment. Also here, two of the *lys3a* lines, X508 and XV508, are intermediate. A close inspection of the seeds facilitated by the fact that *lys3a* seeds have a large embryo [7] reveals that these lines are not pure, but contain 39% (X508) and 60% (XV508) normal barley seeds on weight basis. This impurity is to some extent reflected in the position of the X508 sample adjacent to the normals (O) in the PCA plot of the whole material in Fig. 2; the spectrum of the XV508 sample was not included.

3.2. Establishing causal relationships between spectral, genetic, and chemical information by PLSR and wavelength selection

In spectroscopic evaluation, it is important from the onset of the investigation to carefully inspect the individual spectra. Inspired by the spectral variation, a trained NIR spectroscopist is able to select several wavelengths which may contribute to the chemical validation of the problem. As an example of visual selection, we will discuss a small spectral area displaying an interesting, fine structure between 2270 and 2360 nm marked by a square in Fig. 1D and enlarged about 20 times in Fig. 6A and B. In Fig. 6A, we can compare the spectrum of the original M-1508 mutant with that of its isogenic motherline Bomi, displaying two entirely different patterns. The Bomi spectrum compared to that of *lys3a* shows a more marked shoulder from about 2283 to 2295 nm and a maximum (instead of a decrease) at about 2320 nm, while the *lys3a* spectrum has a dual peak at about 2315 and 2345 nm.

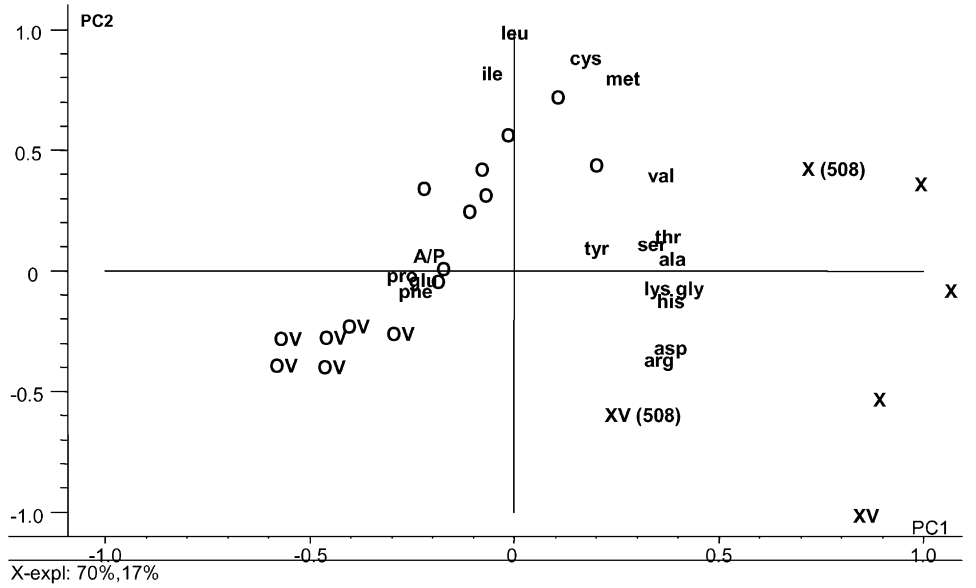


Fig. 4. A PCA biplot with scores and loadings of 21 amino acid analyses of normal (O) and *lys3a* (X) barley samples (see Table 1). V denotes barleys grown in the field; 508 is a putative *lys3a* line which is an outlier due to a contamination with normal barley seeds. The amino acid symbols denote loadings. The A/P symbol denotes loadings for the amide-N to total N index.

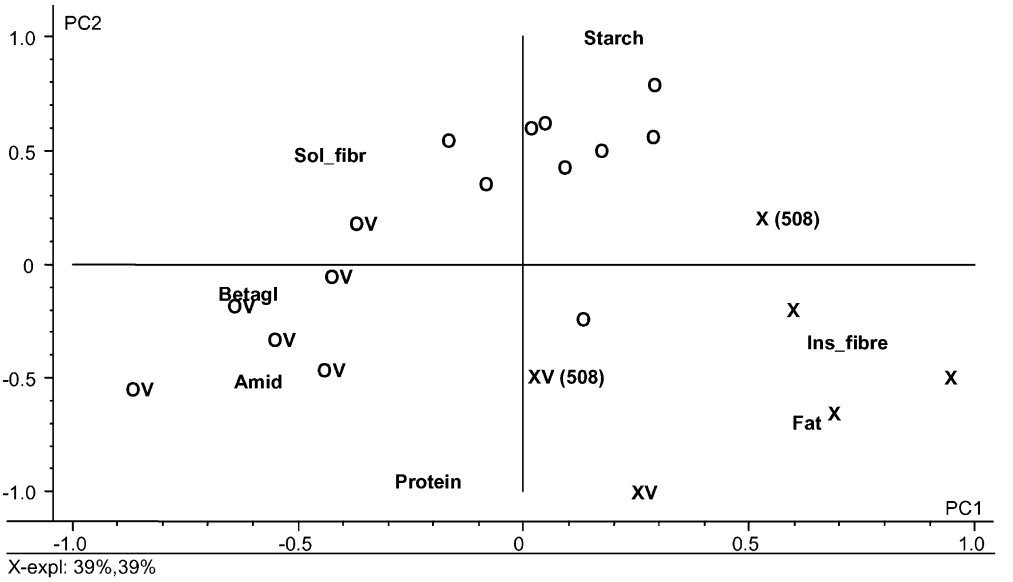


Fig. 5. A PCA biplot with scores and loadings of the chemical composition of the 21 normal (O) and *lys3a* (X) barleys from Table 2. V denotes barleys grown in the field; 508 is a *lys3a* outlier described in Fig. 4.

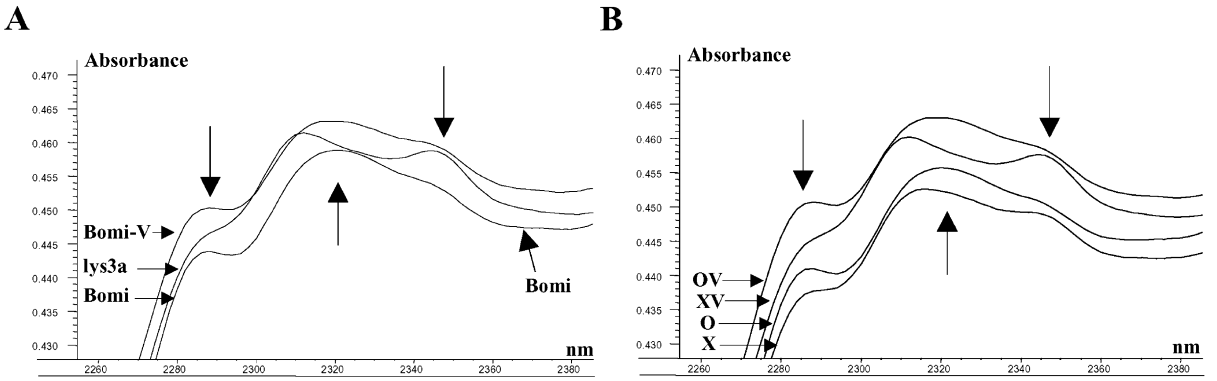


Fig. 6. Enlargement of part of the NIR spectrum 2290–2360 nm in Fig. 1D marked with a square. The arrows denote differences between the spectra which enable a classification discussed in the text — Fig. 1A: spectra of *lys3a* and its isogenic motherline Bomi grown in the field together with that of Bomi grown in the greenhouse (V); Fig. 1B: mean NIR spectra of normal barley (O, $n = 52$; OV, $n = 30$) and *lys3a* barley (X, $n = 25$; XV, $n = 18$).

Both these spectra represent samples grown in the field.

In comparison, Bomi-V (Fig. 6A) grown in the greenhouse (V) demonstrates essentially the same spectral form as Bomi grown in the field, but at higher absorbance reflecting the higher protein content. The conclusions from Fig. 6A regarding the isogenic lines and locations are confirmed for the barley varieties (O, OV) and *lys3a* crosses (X, XV) in Fig. 6B, displaying the mean spectra of the four classes: O ($n = 52$), OV ($n = 30$), X ($n = 25$), and XV ($n = 18$). Returning to Fig. 1D for comparison, an impressive reproducibility of the NIR spectral measurements is demonstrated. We may conclude that by visual inspection of the region 2270–2360 nm, it is possible to correctly classify spectra from normal and *lys3a* barley. The two contaminated deviating *lys3a* lines marked 508 discussed above show spectral characteristics intermediate between *lys3a* and the wildtype. Upon consulting the spectral table for chemical assignment [15], it appears that at 2294 nm there is an amino acid determinant (N–H and C=O) at the normal barley plateau of 2283–2295 nm. The normal barley peak at 2320 nm does not seem to coincide with any nitrogen bond information, but rather with CH information at 2310 nm (CH_2) and 2323 nm (CH_2). At 2336 nm, there is information on cellulose. The *lys3a* peak at 2345 nm is close to the $HC{=}CHCH_2$ indication at 2347 nm for unsaturated fat. It can thus be concluded that the small area of 90 nm between 2270 and 2360 nm, apparently unique for the *lys3a* genotype, is characterized not only by differences to normal barley in amino acids (protein), but also in carbohydrates (cellulose) and (unsaturated) fat, as confirmed in the chemical analyses (Table 2 and Fig. 5). These differences are to be related to the part of the proteome regulated by the *lys3a* gene which directs the machinery of the endosperm and germ tissue cells during development.

We will now supplement wavelength characterization by the naked eye with automatic selection using chemometric algorithms for data reduction into latent factors (principal components), such as with PLSR [5] from spectral intervals created by i-PLSR [12]. In order to indicate which wavelengths are dependant on genotype (normal and *lys3a*) and location (field and greenhouse) in the total material from Fig. 2, a discriminant PLSR was made with a 2×2 factor setup (wildtype = 1, *lys3a* = 0, field = 1, greenhouse = 0). The regression coefficients related to spectral wavelengths are presented in Fig. 7. The genotype component has a much higher profile than that of the location with positive and negative peaks at 495, 510, 1040, 1375, 1505, 1650, 1890, 1900, and 2400 nm. There were however large unique areas in the location loading which explained the satisfactory PCA classification in Fig. 2.

In further defining the chemistry behind the NIR spectroscopy of the barley material, we will now use PLSR to calibrate spectroscopic information on the level of the seed phenotype with two basic chemical

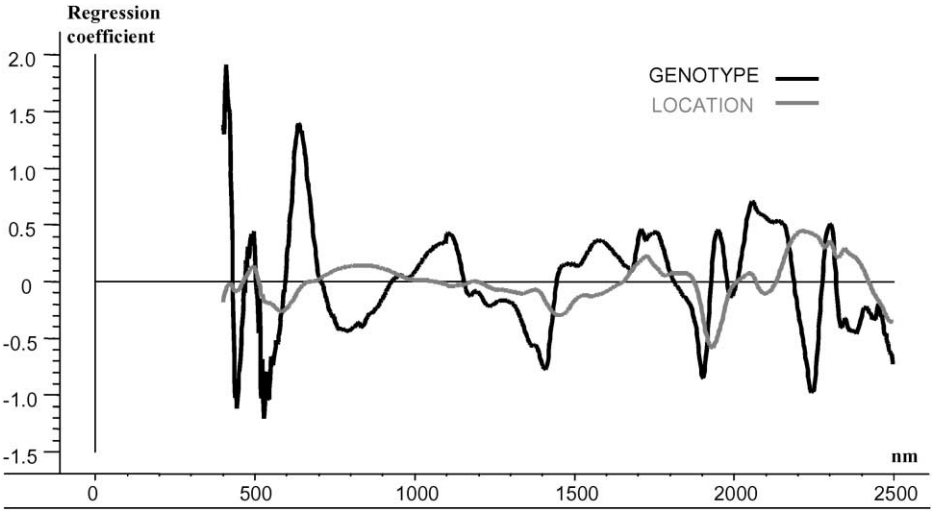


Fig. 7. Regression coefficient from a discriminant PLSR involving NIR spectra, x ($n = 125$) and genotype (normal and *lys3a* barley) and location (grown in greenhouse and in the field) y, from the data in Fig. 2.

analyses, Kjeldahl protein equal to N × 6.25 which represents univariate expression of a range of nitrogen bonds, and the more specific alkali volatile nitrogen analysis which mainly represents the amide-N group and small amounts of ammonium salts available in the barley seeds. In Fig. 8, high correlation PLSR models (six to seven PLS components) with full cross-validation are calculated based on whole spectra for prediction of protein (Fig. 8A), amide-N (Fig. 8B), and the ratio between amide-N and total N (Fig. 8C). It is seen that the barleys grown in the greenhouse marked V tend to have the highest amount of protein (Fig. 8A) and amide-N (Fig. 8B) content and that the *lys3a* genotype marked X generally has a lower content of amide-N compared to normal barley (O). This tendency is further reinforced in the PLSR model for prediction of the amide-N/total N ratio where there is a clear-cut clustering which separates the *lys3a* (X) genotypes from the normal (O). The amide-N/total N ratio available as spectroscopic information is thus one of the many spectral methods effective for screening for the *lys3a* genotype as an alternative to the dye-binding method [7].

In order to further dissect and explain the spectral information, the spectra were divided into 30 intervals of 70 nm each, giving 35 data points after 50% reduction. These were calibrated to protein and amide-N in fully cross-validated i-PLS [14] models. An example with the distribution of RMSECV along the spectral intervals is shown for amide-N in Fig. 8D for the whole barley material ($n = 125$) where the spectral interval number 23 (1940–2008) shows the lowest error. The covariance is generally high and the correlation coefficients with, for example, amide-N content vary typically between 0.99 (highest) and 0.80 (lowest) for five principal components for the 30 intervals.

In order to study synergy between the different spectral intervals, a si-PLS model was developed for computational time reasons limited here to two combined segments. Fig. 8E shows the two selected spectral intervals for amide-N, numbers 23 and 26, and the PLSR model is displayed in Fig. 8F.

For comparison of interval selection by si-PLSR, three data materials were constructed: normal barley ($n = 82$), *lys3a* barley ($n = 43$), and the total barley material normal + *lys3a* ($n = 125$). They were analyzed for si-PLS calibrated to protein and amide-N with full cross-validation. In Table 3, the intervals that were selected with maximum correlation coefficients are presented and the number of PLS components noted. The correlation coefficients and errors for the adjacent principal components were also selected and compared to a full-spectrum PLSR model. It is seen that the information is widely confounded and

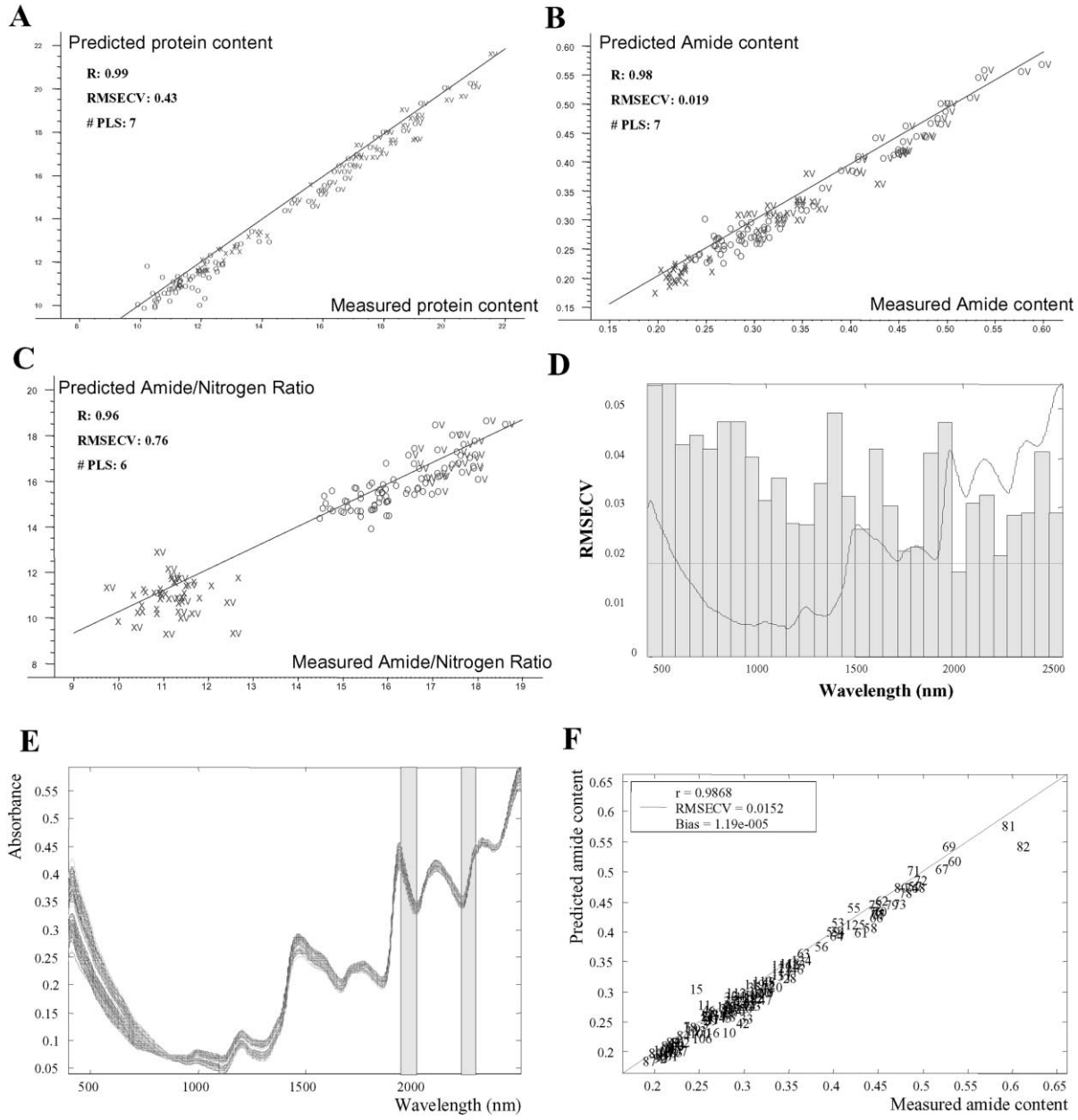


Fig. 8. PLSR regressions of the NIR material in Fig. 2 ($n = 125$) representing prediction of protein (N × 6.25) in (A), of amide in (B) and of the amide-N to total nitrogen ratio in (C). The RMSECVs for the prediction of amide content from NIR spectra of the 30 wavelength intervals chosen for the i-PLSR are shown in (D) and the two optimized wavelength intervals for amide prediction in (E), whereas (F) denotes the PLSR calculation for these two intervals.

Table 3
Selection of NIR spectral segments from different barley materials by si-PLSR calibrated to protein N and amide-N[a]

Calibration	Material											
	Normal (O) + *lys3a* (X) ($n = 125$)				*lys3a* (X) ($n = 43$)				Normal (O) ($n = 82$)			
	PLS components	Intervals	RMSECV	r	PLS components	Intervals	RMSECV	r	PLS components	Intervals	RMSECV	r
Protein N × 6.25	6	16–18	0.42	0.991	5	15–26	0.25	0.997	5	25–27	0.46	0.988
	7	20–27	0.39	0.992	6	20–27	0.25	0.997	6	24–27	0.43	0.989
	8	16–18	0.40	0.992	7	20–27	0.26	0.997	7	15–26	0.44	0.988
	7	FS[b]	0.43	0.990	7	FS	0.32	0.995	7	FS	0.46	0.988
Amide-N	7	12–23	0.015	0.987	6	25–26	0.0087	0.989	6	20–26	0.0161	0.986
	8	23–27	0.015	0.987	7	24–26	0.0076	0.992	7	20–26	0.0158	0.986
	9	15–23	0.014	0.988	8	23–27	0.0077	0.992	8	15–26	0.0162	0.986
	7	FS	0.019	0.980	7	FS	0.0196	0.944	7	FS	0.0167	0.985

[a] Interval synergi-PLSR selection for the barley NIR spectra; 30 spectral intervals, two synergetic components.
[b] Full spectrum.

Table 4
Chemical characterization [15] of wavelength intervals selected by si-PLSR in Table 3

Interval number	Selected in correlation with[a]	Wavelength interval (nm)	Chemistry [15]				
12	A	1170–1238				CH	
15	P, A	1380–1448	NH		ROH	CH	
16	P	1450–1518	NH, protein		ROH, starch cellulose		
18	P	1590–1658				CH	
20	P, A	1730–1798			Cellulose	CH	SH
23	P, A	1940–2008	NH	Amide	R–OH, starch		
24	P, A	2010–2078	NH, protein	Amide, C–O			
25	P, A	2080–2148		Amide	Starch, C=C (fat)		
26	P, A	2150–2218		Amide	Fat, CHO		
27	P, A	2220–2288	Amino acid		Starch	CH	
28[b]		2290–2358	Amino acid		Cellulose, fat	CH	

[a] P: protein; A: amide.
[b] Visually selected.

there is only a small improvement in the local models compared to the full-spectral models, except for amide-N in the *lys3a* material. In order to determine if the relatively small improvement in correlation coefficients and RMSECVs is significant, the models should be evaluated with an independent test set, if a close ranking is desired. Because of the small differences in correlation coefficients and errors, we have decided to further discuss all the intervals having the highest correlation coefficients and the lowest errors selected by si-PLS in Table 3 as a whole in Table 4. Two intervals (numbers 16 and 18) are selected for protein (P) only and one is selected for amide-N (A). Seven intervals are chosen by both types of correlations marked AP. In Table 4, the spectral regions for the intervals are defined and the chemical interpretation given from the literature [15]. We may now also compare to the previously discussed visually selected spectral segment 2270–2360 nm which was indicative for the *lys3a* genotype and which best coincides with the interval number 28 (2290–2358 nm). It is noteworthy that this segment is only in the middle of the ranges of correlations and is not prioritized by the si-PLS selection, although the correlation coefficients are only about 3% lower than those of the optimal segments. Interval number 28 contains mixed information from amino acids (protein) as well as from non-nitrogen components such as cellulose, fat, and C–H bonds. The mixed information is also prevalent for most other selected segments, such as numbers 15, 16, 23, 25, 26, and 27. Only segment 24 seems to be a clear-cut N indicator, while segment 20 indicates carbohydrates and SH groups, and numbers 12 and 18 are C–H indicators.

It can thus be concluded that the NIR spectrum contains repetitive confounded chemical information throughout the spectrum which gives a high degree of redundancy and which in combination with the high precision and repeatability of the measurement explains the versatility and robustness of NIR full-spectrum chemical PLSR predictions brought out by spectroscopic “multimeters” in practice. The consequences of utilizing the multivariate chemical analytical advantage in plant breeding and biotechnology are discussed in Section 4.

4. Discussion

At present, focus in biotechnology tends to be changing from genome sequencing to the concept of the proteome to describe the complement of proteins expressed by a cell tissue, for instance an endosperm. This is called the “the post-genome revolution” in an article in the 16 December 1999 issue of the journal Nature. The complexity of this challenge is lucidly illustrated in Fig. 3a–d comparing the two-dimensional electrophoresis protein patterns from buffer and ethanol extracts from the original Risø 1508 mutant *lys3a* and its isogenic motherline Bomi. They represent about 60% of the total seed proteins with the endosperm as the dominating tissue.

The differences due to just one gene are so great that they cause confusion in comparison of the electropherograms. However, when superimposing the two electrophoresis sets from the albumins/globulins and the prolamins (hordeins), respectively, a pattern of common spots marked with dashed arrows in Fig. 3 proves that the reproducibility is reasonable. The 2×19 electropherograms were made for the 15 normal barleys and the four pure mutant lines analyzed for chemical composition in Tables 1 and 2. The differences between the categories, the wildtype and *lys3a*, could be easily discerned in a blind test by visual inspection where a set of anonymized electropherograms were presented in random order. In the literature, the effect of the *lys3a* gene on the synthesis of about 15 proteins has been compared to other relatively well-studied mutants [7,16]. It constitutes only a fraction of the differences demonstrated in Fig. 3a–d. The *lys3a* gene is situated in chromosome 7 [17] and regulates [10] a range of structural genes in other chromosomes, for example in chromosome 5, the loci coding for the hordeins B (Hor 2), C (Hor 1), and D (Hor 3) produced in the endosperm [16]. The first two of those constituting a major part of the hordeins are drastically reduced in *lys3a*, contributing to a total reduction of the hordeins of 85% [18]. At the same time, however, the D-hordeins increase in *lys3a* [18]. The overall reduction of the hordeins was confirmed in the electropherograms when comparing the protein patterns of the alcohol soluble proteins in Fig. 3 (a: Bomi; b: *lys3a*). On the other hand, the protein pattern of the buffer-soluble albumins/globulins in Bomi (Fig. 3c) are in general quantitatively increased in *lys3a* (Fig. 3d) with great qualitative and quantitative differences. Thus, it seems as if the retardation through the *lys3a* gene of the synthesis of the hydrophobic ethanol-soluble hordeins results in a range of hydrophilic buffer-soluble fragments.

With regard to buffer-soluble proteins expressed in the triploid tissue of the endosperm and aleuron in the barley seed, large changes have been confirmed due to the gene *lys3a*. For example, there are decreases in beta-amylase and protein Z [19], while others are increased, e.g. the potential antifungal proteins [20], an amylase/subtilisine inhibitor, a chitinase, and a ribosome-inactivating protein [21,22].

In plant breeding and in biotechnology, there is a great need for screening methods in order to identify genes and gene effects. The genome and proteome concepts introduce a multivariate challenge, where multivariate screening methods such as spectroscopy and data analysis such as chemometrics are central. Recently, Delwiche et al. have demonstrated [6] that it is possible to detect certain wheat–rye chromosomal translocations by NIR spectroscopy. We have described [11] that in a mixed genetic background, it is possible by NIR to distinguish normal barley not only from the drastic high-lysine mutant Risø 1508, but also from the lesser high-lysine mutants Risø 13, 16, 29, and 95 [23] which have not yet been studied from a protein chemistry point of view.

In a classical approach, the biochemist and the biotechnologist tend to focus on specific genes, mechanisms and proteins, while forgetting about the side effects which are considered vital by the geneticist and the plant breeder and are collected in the concept of pleiotropic gene effects. Thus, the Risø mutant 56, in contrast to the regulating gene *lys3a* [24] is a mutation (deletion) in a structural gene Hor 2 in chromosome 5 coding for the B-hordeins, as documented by the absence of the RNA messenger [24]. There are compensatory increases in the C- and the gamma-hordeins [24]. Other pleiotropic effects due to the gene mutant 56, for example on the carbohydrate composition, have not been described but are most likely to occur, because mutants affecting hordein synthesis usually have decreased starch synthesis [7]. If so, the changes in the proteome as well as other changes in other chemical components derived from here could be detected in mutant 56 by NIR spectroscopy, preferably in an isogenic comparison.

Thus, NIR spectroscopy enables a physicochemical fingerprint of the phenotype on the level of phenotypical biological organization which can be compared and analyzed by chemometrics in a PCA, and validated to chemical analyses and knowledge by PLSR. It is thus possible, by defining what is normal barley, to identify and investigate outliers with unknown chemical composition and afterwards define their genetic and chemical status [11]. The material should be grown on the same site, although we have shown in our *lys3a* example that it is possible to separately model the environmental and genetic effects (Figs. 2 and 7). It is clear that a multivariate dataset, collected either by a range of univariate chemical analyses (Tables 1 and 2, Figs. 4 and 5) or more easily by a

spectroscopic method (Fig. 2), could facilitate differentiation between genetic and environmental effects compared to a classical approach. The key to utilizing this option is the ability to explore and to model covariance in the datasets by using chemometric methods.

As early as 1930, Bishop [25], in his nitrogen-regulation principle of the Osborne [26] protein fractions of barley, demonstrated a case of covariance implying that as a part of the total protein content, the hordein proteins increased and the albumins and globulins declined when the protein level was increased, for example with nitrogen fertilization. This implied a decline in the total protein content of lysine and essential amino acids and an increase in the content of amides, glutamic acid, and proline due to the composition of these proteins. This mechanism was almost considered a natural law [27] until the discovery of the first high-lysine mutants. Fig. 9A displays the high precision of the negative regression lysine g per 16 gN with total protein content ($N \times 6.25$) for the 15 normal barleys analyzed for amino acids. The introduction of the *lys3a* gene outliers (Fig. 9A) completely changes this picture. It has been confirmed [7] that both the *lys3a* and the *lys1* genes straighten out this correlation, so that now the lysine content of protein with these genes is independent of total protein content. This change is also reflected in comparing the two specific total amino acid patterns as a function of the protein content of the seed, each unique for normal barley and for the high-lysine mutant *lys3a* (Table 1) which is elaborated as a whole in the PCA in Fig. 4.

While the genetic data in our investigation are well defined and hard, the direct and indirect effects of this regulating gene on the endosperm proteome and the phenotype are extremely diverse and multivariate, thus requiring soft mathematical modeling. It is thus in practice impossible to study these effects as a whole without suitable multivariate analytical screening methods like spectroscopy and without chemometric evaluation. We might conclude that gene-dependent, specific, multivariate covariate correlation patterns like those between amino acids as a function of protein content in barley seeds are just as deterministic and environmentally independent a trait as blue and brown eyes in humans.

We will now discuss how the detection of the *lys3a* genotype may work on the NIR spectroscopic level. As seen in Fig. 9B, lysine can be reasonably predicted by full-spectrum NIR (RMSECV O.24 at five PLSR components). A straight 30-interval i-PLS selects interval 28 (RMSECV O.20 at seven PLSR components) which was earlier selected (Fig. 6 and Table 4) as a unique area for visual *lys3a* differentiation from normal barley. A si-PLSR with two synergy elements selects at five PLS components interval 27 together with 28 (RMSECV 0.15 at five PLSR components). However, at seven PLS components, si-PLSR combines the spectral elements 17 and 26 with a minimal error of RMSECV 0.13, approximately half of that for the whole spectral model. These results on lysine, based on 18 spectra, should be confirmed in studies with a larger independent material. However, the purpose of the exercise here is not to differentiate one or two "hot" areas in the NIR spectra from a close ranking list in defining the *lys3a* genotype, but rather to look at several areas with low prediction errors in order to explain how full-spectrum NIR works.

In Fig. 9C, it is seen that lysine mol% is highly negatively correlated to amide-N to N ratio ($r = -0.97$), which points to the possibility that the spectroscopic signature of the amide bond in this material ($n = 125$) could be a good indicator for a low content of lysine. Of the spectral elements, numbers 17, 26, 27, and 28, previously selected by si-PLS correlated with lysine, number 26 (2150–2218 nm) is indicative of amide [15] (Table 4) as well as (unsaturated) fat and the aldehyde group. Area 17 (1520–1588 nm) has information about R–NH_2, starch and the peptide bond, area 27 (2220–2288 nm) about amino acids, cellulose and (unsaturated) fat, while area 28 (2290–2358 nm) includes fiber in addition to the information kept in area 27. In classifying the *lys3a* gene and the wildtype (Figs. 1 and 2) and in predicting chemical constituents such as lysine (Fig. 9B) and protein and amide-N (Fig. 8A–C), the models chosen by the PLSR algorithm do not only rely on direct protein information but also on other different combinations, exploiting the pleiotropic covariate effects of the gene (Table 2). These are due to the influence of the specific proteome on the parts of the cell machinery which are important for the starch, fiber, and fat synthesis (Tables 3 and 4).

It is concluded that the redundancy of biological information on the genotypic DNA sequence level is also represented at the phenotypic level in the dataset read by the NIT/NIR spectroscopic sensor from the chemical/physical fingerprint containing specific

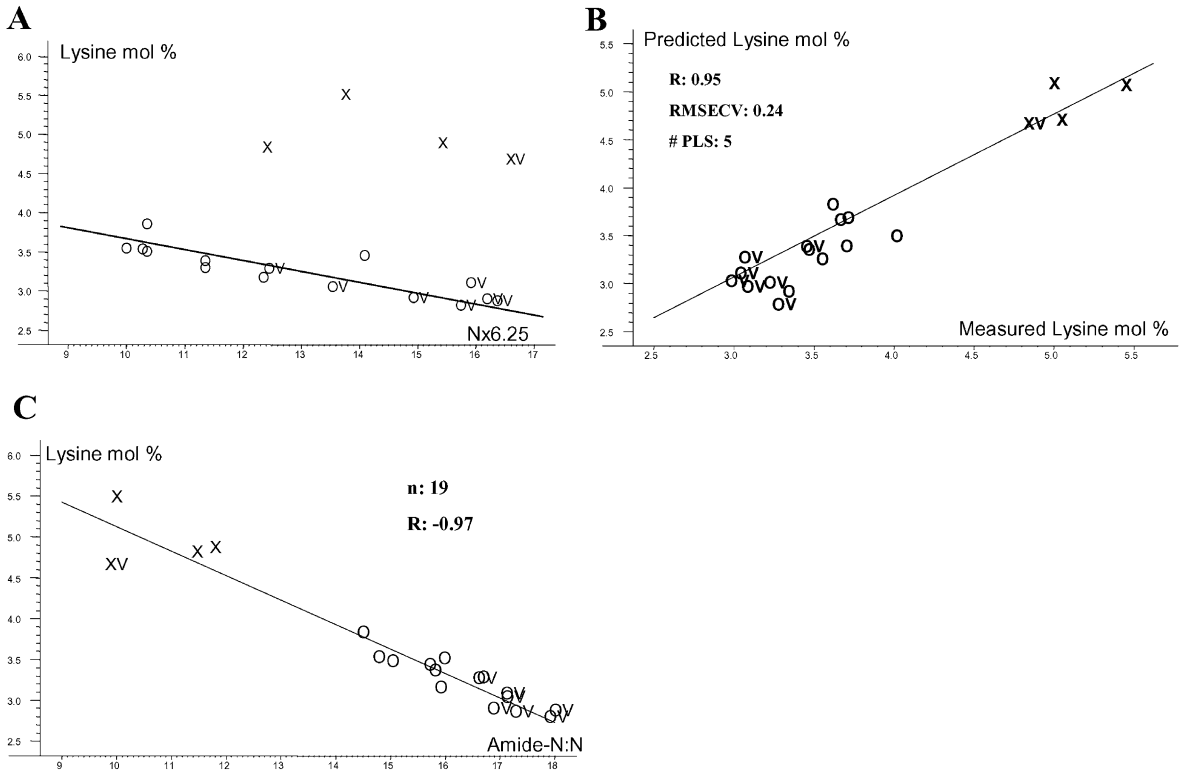


Fig. 9. (A) Data from Table 1 ($n = 19$): the regression line for normal barley (O) for lysine mol% to protein (N × 6.25), X denotes the *lys3a* outliers; (B) PLSR NIR prediction of lysine mol%; one outlier removed ($n = 18$); (C) correlation between lysine mol% and the ratio amide-N to total N.

genetic information which can be encoded by chemometrics. NIR spectroscopy may in the future find an extended use in selecting transformants not only coding for genetically engineered proteins [28], but also selecting for other genes by exploiting the pleiotropic effects as markers instead of using antibiotic-resistant genes. This could be done with great sensitivity and precision in an isogenic background with material grown under the same environmental conditions.

In barley, the *lys3a* gene exerts a negative pleiotropic effect on the starch content. Conventional breeding by changing the gene background has improved starch quantity without loss in protein quality [29]. By using NIR spectroscopy evaluated by chemometric methods, pleiotropic effects can be quantified as a whole and explained by chemical validation. For example, NIR spectroscopy combined with chemical validation makes it possible to define a pleiotropic covariate complex in high-lysine barley breeding for the *lys3a* gene. Thus, in a high-lysine barley breeding program, a chemometric model of selection in cross-breeding populations by NIR could be defined which neutralizes the negative parts of the pleiotropic complex by restructuring the multigene background. Classical, exploratory plant breeding could thus be upgraded to a high-tech analytical status and the cooperation with normative biotechnology improved.

Our example is a special case of proteome dynamics cast as a covariate chemical imprint in the desiccated seed endosperm tissue. The spectroscopic and chemometric screening concept presented here should have great advantages in isolating mutants and gene transformants by revealing covariate gene indicators in studying the dynamics of growth in cultures of cells and microorganisms.

The above-cited post-genome revolution by the proteome claimed by the journal Nature in December 1999 will not constitute the end point of biological

science. Already now, we are envisaging the possibility of utilizing the high precision and multivariate advantage of spectroscopy to separately model the covariate expressions of genetic and environmental variation by chemometrics. This enables us to overview and non-destructively, separately model the genetic and environmental variation at the highest level of biological organization the chemistry of the intact phenotype, the chemotype, within the limits of our sensors and our chemometric evaluation methods.

Acknowledgements

The research was kindly financed by the Danish Cereal Network, Department of Food and Agriculture. The authors thank Anna Marie Hansen, Peter Henriksen, and Kirsten Wilms for the laboratory work and Gilda Kischinovsky for assistance with the manuscript. The barley seed material emerges from a 40-year research cooperation with colleagues at the Swedish Seed Association, Svaløf, Sweden, and with The Risø National Laboratory and The Carlsberg Research Laboratory in Denmark. Preben Holm at the Danish Institute of Agricultural Sciences, Flakkebjerg, and Gary Fulcher from the University of Minnesota, St. Paul, USA, have contributed inspiring discussions.

References

[1] W. Herschel, Phil. Trans. R. Soc. (Lond.) 284 (1800) 284–292.
[2] W.W. Abney, R.E. Festing, Phil. Trans. R. Soc. (Lond.) 31 (1881) 416.
[3] P.C. Williams, K.H. Norris, Near-Infrared Technology in Agriculture and Food Industries, American Association of Cereal Chemists, St. Paul, MN, USA, 1987.
[4] K.H. Norris, Agric. Eng. (St. Joseph, MI, USA) 45 (1964) 370.
[5] H. Martens, S.Aa. Jensen, in: J. Holas, J. Kratochvil (Eds.), Proceedings of the 7th World Cereal and Bread Congress, Prague, June 1982, Elsevier, Amsterdam, 1983, p. 607.
[6] S.R. Delwiche, R.A. Greybosch, C.J. Peterson, Cereal Chem. 76 (1999) 255.
[7] L. Munck, in: P.R. Shewry (Ed.), Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, CAB International, Wallingford, UK, 1992, p. 573.
[8] L. Munck, K.E. Karlsson, A. Hagberg, B.O. Eggum, Science 168 (1970) 985.
[9] J. Ingversen, B. Køie, H. Doll, Experientia 29 (1973) 1151.
[10] M. Blom Sørensen, M. Muller, J. Skeritt, D. Simpson, Mol. Gen. Genet. 250 (1996) 750.
[11] L. Munck, in: R. Von Bothmer, et al. (Eds.), Diversity in Barley, Elsevier, Amsterdam, 2001, Chapter XII, manuscript accepted, in preparation.
[12] V. Barkholt, A.L. Jensen, Anal. Biochem. 177 (1989) 318.
[13] S. Jacobsen, L. Nesic, M. Petersen, I. Søndergaard, Electrophoresis 22 (2001) 1242–1245.
[14] L. Nørgaard, A. Saudland, J. Wagner, J. Pram Nielsen, L. Munck, S.B. Engelsen, Appl. Spectrosc. 54 (2000) 413.
[15] B.G. Osborne, T. Faern, P.H. Hindle, Practical NIR Spectroscopy, 2nd Edition, Longman, Harlow, UK, 1993.
[16] M. Kreis, P.R. Shewry, in: P.R. Shewry (Ed.), Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, CAB International, Wallingford, UK, 1992, p. 319.
[17] K.E. Karlsson, Barley Genet. Newslett. 7 (1977) 40.
[18] M. Kreis, P.R. Shewry, B.G. Forde, B.G. Miflin, Biochem. Gen. 22 (1984) 231.
[19] J. Hejgaard, S. Boisen, Hereditas 93 (1980) 311.
[20] R. Leah, H. Tommerup, I. Svendsen, J. Mundy, J. Biol. Chem. 266 (3) (1990) 1564.
[21] J. Mundy, J. Hejgaard, A. Hansen, L. Hallgren, K.G. Jørgensen, L. Munck, Plant Physiol. 81 (1986) 630.
[22] R. Leah, Barley Seed Antimicrobial Proteins, Dissertation, Department of Microbiology, University of Copenhagen, 1991.
[23] H. Doll, in: W. Gottschalk, H.P. Muller (Eds.), Seed Proteins: Biochemistry, Genetics, Nutritive Value, Martin Nijhoff, The Hague, 1985, p. 205.
[24] M. Kreis, P.R. Shewry, B.G. Forde, S. Rahman, B.J. Miflin, Cell 34 (1983) 161.
[25] L.R. Bishop, J. Inst. Brew. 36 (1930) 352.
[26] T.B. Osborne, The Vegetable Proteins, Longmans & Green, London, 1920.
[27] W. Postel, Zuchter 26 (1956) 211.
[28] M. Torrent, I. Alvarez, M.I. Geli, I. Dalcol, D. Ludevid, Plant Mol. Biol. 34 (1997) 139.
[29] K. Bang-Olsen, B. Stilling, L. Munck, Barley Genetics V, Proceedings of the Fifth International Barley Genetics Symposium, University of Okayama, Japan, 1987, p. 865.

Paper VI

Spectral multivariate indicators for mutant endosperm genes evaluated by chemometrics reveal a new mechanism for substituting starch with (1→3,1→4)-β-glucan in barley

Lars Munck, Birthe Møller, Susanne Jacobsen and Ib Søndergaard

---

VI

ABSTRACT

Near Infrared Reflection (NIR) spectroscopy was tested as a screening method to classify varieties and high lysine mutants from a barley collection by Principal Component Analysis (PCA). Inspecting the form of the mean spectra of the samples within each cluster from the PCA score plot, gene-specific patterns were identified in the area 2270-2360nm. The characteristic spectral signatures representing the *lys5* locus (Risø mutants 13 and 29) were found to be associated with large changes in percentage of starch and (1→3, 1→4)-β-glucan (BG). These alleles compensated a low level of starch (down to 30%) by a high level of BG (up to 15-20%) thus maintaining a constant production of carbohydrates around 50-58% within the range of normal barley.

The spectral screening tool was tested by an independent dataset with six mutants with unknown carbohydrate composition. Spectral data from four of these were classified within the high BG *lys5* cluster in a PCA. Their high BG and low starch content was verified. It is concluded that genetic diversity such as from gene regulated carbohydrate and storage protein pathways in the endosperm tissue can be discovered unsupervised from the phenotype by chemometric classification of a spectral library, representing the digitised phenome from a barley gene bank.

INTRODUCTION

The barley endosperm is a well conserved imprint of the physical-chemical dynamics of an approximately 35 days developmental process after anthesis which is regulated by specific genes according to a precise timetable set by the genotype partly independent of environment. Because all advanced barley lines due to selfpollination are almost homozygotic, mutants have near isogenic backgrounds for precise references. The desiccated seed/endosperm system is ideal to be explored by near infrared spectroscopy, avoiding interfering water peaks. The phenome (Watkins *et al*., 2001) is here regarded as an interface expressed as patterns of chemical bonds for the expression of specific genes (Munck, 2003). These are indirectly observed by spectroscopy as chemical-physical fingerprints developed on a graphic computer display by chemometric pattern recognition data models (Martens and Næs, 1989) such as Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR).

Near Infrared Transmission (NIT) measurements of spectra evaluated by PCA were used by Campbell *et al*. (2000) in maize to classify endosperm mutants. We have (Munck *et al*., 2001) demonstrated in Near Infrared Refection (NIR) spectroscopy PCA score plots, how unsupervised NIR information from barley flour can differentiate between normal and high lysine *lys3a* barley (Risø mutant 1508, Doll, 1983) as two different clusters. The genetic

differences were identified as specific signatures in NIR spectra especially at 2270-2380nm. Environmental differences are mainly expressed as offsets from the baseline. They are represented as systematic shifts between environments of gene specific clusters in the PCA score plot still maintaining the relative distances between the genotype clusters.

We will in this paper further demonstrate, how the spectral screening tools originally developed in food science (Osborne *et al*., 1993) can be used as a link between data from the genotype and phenotype, in order to detect hereto-unknown ways of gene expression in barley endosperm mutants (Munck, 2003).

MATERIAL AND METHODS

The barley mutant genes (Doll, 1983) investigated here consist of:

1. Four alleles in the *lys3* locus in chromosome 5 (new nomenclature, see Muravenko *et al*., 1991) with alleles Risø mutants *a* (1508), *b* (mutant 18) and *c* (mutant 19) in Bomi background and the Carlsberg mutant 1460 (Munck, 1992) in Minerva here called *lys3m*.
2. Two *lys5* alleles in chromosome 6 Risø mutants *5f* in Bomi (mutant 13) and *5g* in Carlsberg II (mutant 29) as well as the double recessive *lys3a5g*.

As an independent test set, four less studied mutants selected as high lysine are included as well as two waxy lines. They are: the Risø mutants *lys4d* (mutant 8, in chromosome 1) and mutant 16 (in chromosome 7) both induced in Bomi (Doll, 1983), the Italian mutants 95 and 449 induced in Perga by Di Fonzo and Stanca (1977) and two putative waxy (amylopectin) lines 1201 and 841878 of unknown origin previously imported to the Carlsberg collection maintained at KVL called w1 and w2.

The mutants and their background varieties and segregating crosses with normal barley as well as a range of normal barley varieties were grown under different conditions (field, outdoor pots, greenhouse) in different years.

The material was stored in closed containers in the refrigerator after they were equilibrated to the temperature and moisture of our laboratory. Thus two materials were obtained:

Material 1: 54 normal and original mutant lines (*lys3*, *lys5*, *lys3a5g*) grown in greenhouse in year 1998-2000.

Material 2: Nine lines from the test set of six mutants defined above, mainly grown in greenhouse.

The NIR analysis (on milled flour from ripe seeds) is described by Munck *et al*. (2001) together with the chemical analyses of protein, amino acids, amide, nitrogen, and starch. A determination of apparent amylose in starch from the two waxy lines was made with an iodine spectroscopic method on non-defatted isolated starch (BeMiller, 1964). Two methods of BG analysis were employed. The fluorimetric BG analysis with Calcofluor

(Munck *et al*., 1989) was used routinely and was checked with an enzymatic method specific for (1→3,1→4)-β-glucan (Anonymous, 1998).

Chemometric pattern recognition analysis of spectral data was performed unsupervised by Principal Component Analysis (PCA) for classification and supervised as predictions of chemical parameters by calibration through Partial Least Squares Regression (PLSR) analysis. For this the "Unscrambler" software (Camo A/S, Trondheim, Norway) was used according to Martens and Næs (1989).

RESULTS INCLUDING DISCUSSION OF THE ANALYTICAL METHODS

Exploring spectroscopic indicators for endosperm genes by NIR 400-2500nm (Material 1)

We start our spectral investigation unsupervised by measuring NIR spectra from whole milled flour of 54 barley lines grown in one environment (Material 1, greenhouse) which are presented in multiplicative signal corrected (MSC) form in Fig. 1A. On the basis of this data material, a PCA score plot between the principal components PC1 (*abscissa*) and PC2 (*ordinate*) is performed automatically in the computer and displayed in Fig. 1B. Four clusters are shown in the PCA plot.

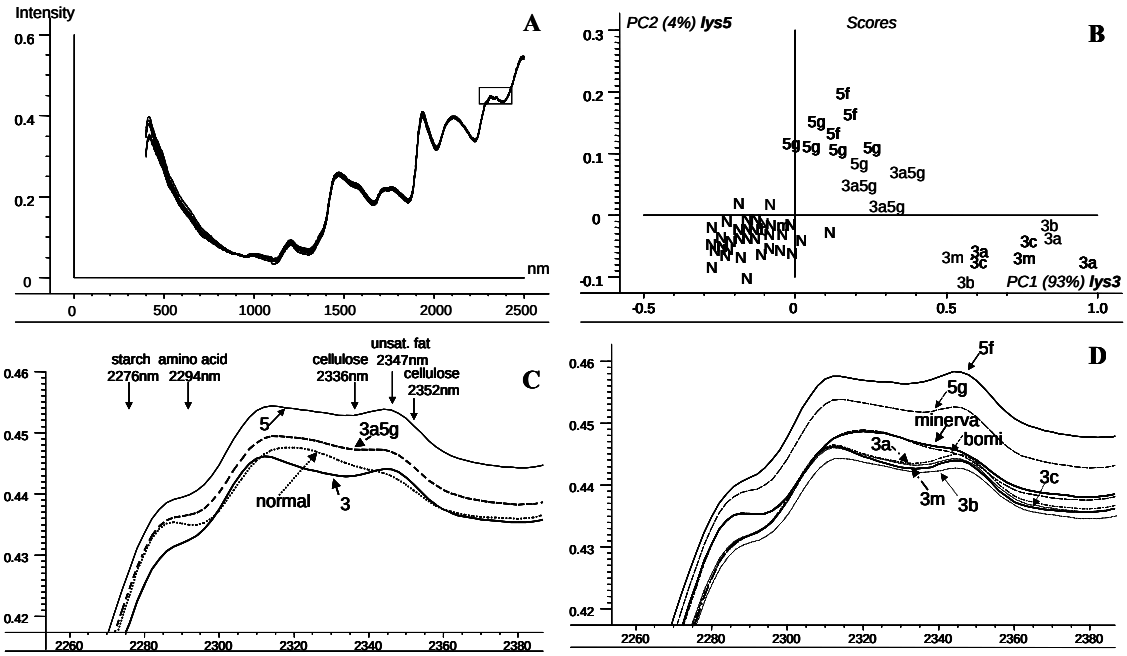


Figure 1A. NIR (MSC) spectra 400-2500nm of the 54 barley lines grown in greenhouse (Material 1). **B.** PCA scoreplot (PC1:2) of whole NIR spectra in A. Nb=Bomi, Nm=Minerva, for mutant identification see text. **C.** Mean spectra from enlarged area 2260-2380nm (marked with a square in A) of the four genotype clusters (normal, *lys3*, *lys5* and *lys3a5g*) revealed from the PCA in B. **D.** Comparing spectra (2260-2380nm) from *lys5f*, *lys5g*, *lys3a*, *lys3b*, *lys3c*, *lys3m* and the motherlines Bomi and Minerva (*lys3m*).

First, now we consult our identification list and find that the clusters reflect four different genotypes - normal barley N, *lys3* (four alleles *a*, *b*, *c* and *m*) along the PC1 axis and *lys5* (two alleles *f* and *g*) spanning the PC2 axis and with the double recessive *3a5g* in between.

The PCA score plot in Fig. 1B facilitates a reduction of the 54 spectra to four mean spectra each representing the clusters of normal (N), *lys3, lys5* and *lys3a5g* barley. When inspecting these four mean spectra, we can identify an interesting small area in the NIR spectra indicated with a square in Fig. 1A between 2270 and 2380nm and displayed enlarged in Fig. 1C-D.
The mean *lys3* and *lys5* spectral signatures in Fig. 1C are distinctly different from each other and from that of normal barley (mean), while *lys3a5g* (mean) is intermediate between those of *lys3* and *lys5*. In Fig. 1D the same conclusion can be drawn from spectra of individual samples of the four *lys3* alleles *lys3a, lys3b, lys3c, lys3m* and from the two alleles in *lys5*, *lys5f* and *lys5g*. The similar spectral responses for the samples of all the *lys3* alleles in Bomi background *lys3a*, *lys3b*, *lys3c* show the same response as the fourth *lys3* allele mutant *lys3m* in Minerva as demonstrated in Fig. 1D. The spectra from *lys5f* and *lys5g* have similar form, however, *lys5f* has a more extreme bulb at 2350nm, which confirms the more extreme position of *lys5f*, compared to *lys5g* in the PCA in Fig. 1B. The spectral differences between the barley reference varieties Bomi and Minerva and to most other normal lines are small.

It is thus concluded from the spectroscopic investigation that NIR spectroscopy evaluated by PCA is surprisingly effective in differentiating this genetic material grown in greenhouse. For best genetic separation, the material should be grown in the same environment (Munck *et al*., 2001). The spectroscopic signatures indicative for different gene loci and normal barley are clear cut and reproducible. The method is able to differentiate between alleles in the same locus. Thus, *lys5f* seems to be a more drastic mutant than *lys5g*. Differences in genetic background within the normal barley category (Bomi and Minerva) and within mutant alleles are less important than the effects of the mutants themselves (Munck *et al*., 2001, Jacobsen *et al*., 2004).

Explaining the results from the classification by NIR spectroscopy and PCA of the lys3/lys5 genes from Material 1 by chemical analyses
The NIR spectrum contains repetitive confounded physical-chemical information through out the NIR spectrum as primary, secondary, tertiary…. vibration overtones and combination bands from 2500-713nm emerging from the fundamental vibrations in the Infrared (IR) region 2500^{-5} x 10^{4}nm.

The NIR detection of the *lys*3 alleles (Fig. 1B) was not surprising. The *lys3a* genotype (Munck *et al*., 2001) is characterised by a low amide/protein N ratio (A/P) of 11.4 compared to its mother variety Bomi (A/P=16.3) due to the low content of hordeins rich in amides. It was then shown that the *lys3a* gene is likely to be detected in NIR because it mediates low amide content. Information for the amide bond is according to Osborne *et al*. (1993) distributed at 20 wavelengths in the NIR area 1430-2180nm. We also confirmed (Munck *et al*., 2001) a high correlation between lysine and amide content (r=-0.97). As discussed below amide detection is only a part of the definition of the *lys3a* phenotype by NIR spectroscopy.

The two *lys5* alleles *lys5f* (Risø mutant 13) and *lys5g* (Risø mutant 29*)* were selected by the dye-binding method (Doll, 1983) indicative for increase in basic amino acids such as lysine. The lysine content in these mutants was only slightly increased by about 10%. Later Greber *et al*. (2000) suggested that these mutants rather should be looked upon as gene lesions in the starch synthesis because they were considerably reduced in starch (50-75% compared to normal barley near isogenic controls).
However, if we compare the sum of starch and BG = (BG+S) content the picture changes (Table IA). *lys5g* now even seems to exceed normal barley (58.0 versus 53.9%), and even for the extreme mutant *lys5f* the low starch content (29.8%) is compensated with a high BG content (19.8%) to give a total sum as high as 49.4%.

It was thus surprising to note (Table IA) that the chemical analyses reveal that the *lys5* cluster in the spectral PCA in Fig. 1B additionally to a low level of starch is characterised by very high BG levels fully or partially compensating for the decrease in starch.

In this investigation, we were lucky to include the BG analysis because we have developed it as a routine analysis in the brewing industry (Munck *et al*., 1989). The extreme gene *lys5f* is as high as 19.8% compared to 13.3% for *lys5g* compared with normal barley's at 6.5%. Although a greenhouse environment regularly produces a higher BG and protein content than in the field (compare Tables IA with B) the effect of the *lys5* genes on BG is spectacular.

The allele *lys3m* (induced in Minerva) originally selected as a low BG mutant at Carlsberg (Munck, 1992) has an extremely low A/P index of 9.5 compared to 17.5 in Minerva. It is interesting to note that the mutant allele *lys3c* in Bomi differs from the other alleles in displaying a normal barley BG value of 6.1%. The other *lys3* alleles are all low in BG (approx 2.5%). The double recessive *lys3a5g* has A/P index and BG content intermediate between *lys3a* and *lys5g* barley verifying its intermediate position in the NIR classification by PCA between the *lys3* and *lys5* classes (Fig. 1B).

Table IA. Chemical properties of the 54 samples (Material 1) and the six original mutants (Material 2) grown in greenhouse

Greenhouse	n	BG	Starch (S)	BG + S	Protein	Amide	A/P	Fat	Lys	Glu
lys3a	3	4.73±0.98	40.4±1.0	45.2±0.1	17.7±0.9	0.32±0.03	11.41±0.56	3.51[d]	4.94[d]	14.86[d]
lys3b	2	3.05±0.78	-	-	17.1±0.7	0.32±0.01	11.51±0.22	-	-	-
lys3c	2	6.10±1.56	-	-	17.5±0.5	0.34±0.01	11.96±0.93	-	-	-
lys3m	2	2.25±0.01	39.3±1.3	41.6±1.3	17.4±0.5	0.27±0.01	9.50±0.06	-	-	-
lys5f	3	19.80±0.20	29.8±0.6[b]	54.8±6.0[b]	17.0±1.4	0.42±0.06	15.52±0.99	3.69[d]	3.32[d]	27.59[d]
lys5g	6	13.26±0.56	44.7[d]	58.2[d]	17.4±1.0	0.43±0.04	15.46±0.48	2.30±0.25[b]	3.76[d]	20.09[d]
3a5g	3	7.8±1.3	34.7±10.5[b]	43.2±10.6[b]	17.2±1.7	0.37±0.02	13.6±1.1	-	4.0[d]	20.1[d]
Normal incl. B	33	6.45±2.67[a]	47.8±1.0[c]	54.3±1.8[c]	16.2±1.3	0.44±0.01	19.95±0.62	1.94±0.16	3.05±0.15	24.13±0.83
Bomi (B)	1	6.80	48.8	55.6	14.6	0.38	16.24	1.74	3.27	22.90
lys4d	1	4.0	41.1	45.1	17.5	0.37	13.21	-	4.04	19.46
16	2	16.6±1.9	29.9[d]	45.1[d]	17.1±1.5	0.45±0.06	16.26±0.92	-	3.37[d]	22.50[d]
449	2	13.5±0	26.5[d]	40.0[d]	20.7±2.3	0.05±0.06	15.14±0.02	-	-	-
w1[e]	1	15.4	27.3	42.7	16.5	0.40	15.14	-	-	-
w2[e]	1	7.0	49.0	56.0	17.4	0.47	16.93	-	-	-
Total	63	10.6±4.6	39.9±8.2	48.0±6.8	16.0±2.7	0.42±0.06	15.55±2.17	2.36±0.68	3.45±0.63	22.66±3.44

[a] n=30, [b] n=2, [c] n=9, [d] n=1, [e]content of amylose: w1=20.3% and w2=4.2% of starch

Table IB. Chemical properties of 18 samples from Material 1 and 7samples from Material 2 grown in field

Field	n	BG	Starch (S)	BG + S	Protein	Amide	A/P	Fat	Lys	Glu
lys3a	1	3.1	48.5	51.6	12.7	0.23	11.36	2.63	-	-
lys3m	1	2.4	48.8	51.2	12.3	-	-	-	-	-
lys5f	1	16.5	-	-	-	-	-	3.77	3.8	19.9
lys5g	2	8.9±1.0	-	-	11.8±0.1	0.26[d]	13.7[d]	-	-	-
3a5g	1	-	-	-	15.5	0.28	11.3	-	4.8	15.6
Normal incl. B	13	4.5±0.8	55.1±2.2	49.7±.5	11.3±1.1	0.28±0.04[e]	15.4±0.6[e]	1.90±0.21	-	-
Bomi (B)	1	4.9	53.6	58.5	11.5	0.29	16.4	1.91	3.5	21.8
lys4d	1	4.1	-	-	12.9	0.29	14.0	-	4.2	19.1
16	1	12.0	-	-	13.8	0.32	14.5	-	-	-
95[f]	2	12.2	29.6[d]	41.8[d]	15.1±0.5	0.35±0.03	14.2±0.7	-	-	-
449	1	12.4	-	-	14.6	0.32	13.7	-	-	-
w1	1	15.6	-	-	-	-	-	-	-	-
w2	1	5.7	-	-	13.0	0.33	15.9	-	-	-
Total	27	6.9±4.2	49.9±8.0[g]	54.6±6.5[g]	12.3±1.7	0.29±0.04	15.1±1.1	2.16±0.64	3.7±0.5	21.0±2.7

[f]=one sample grown in outdoor pots [g] n=11

Simple chemical classification of the mutant genes

As is discussed by Jacobsen *et al*. (2004) there are many chemical ways to detect barley mutants because they give a range of specific complex physical-chemical imprints in the phenotype only detectable as a whole by multivariate pattern recognition analysis. It is surprising to note that very simple chemical plots like BG (*abscissa)* and A/P index (*ordinate)* in Fig. 2A and even starch (*abscissa)* and BG (*ordinate*) in Fig. 2B suffice for a successful gene classification as compared to the PCA of corresponding NIR data in Fig.1B. However, such a chemical classification must be supervised because it assumes *a priori* chemical knowledge.

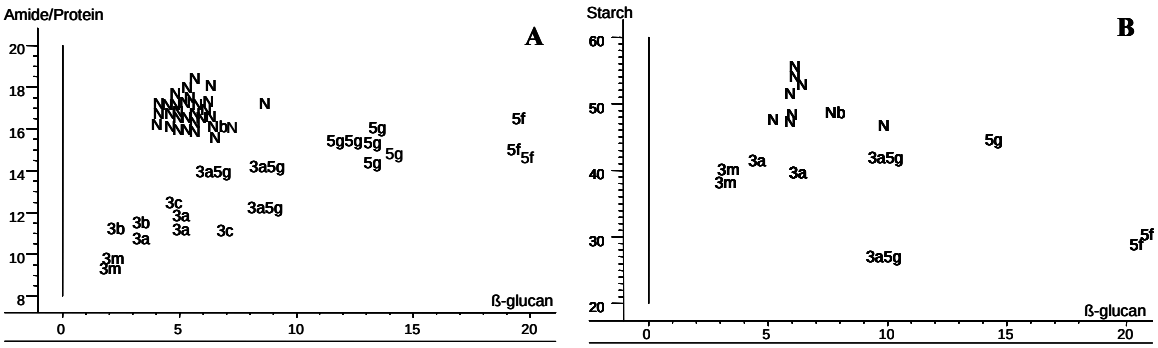


Figure 2A. Amide/Protein ratio (ordinate) plotted against ß-glucan. **B.** Starch (ordinate) plotted against ß-glucan.

The advantage of gene detection by NIR spectroscopy

The NIR approach is useful because the screening can be performed unsupervised, picking up a broad physical-chemical fingerprint of the phenome also including unexpected effects such as BG. The spectra can be interpreted by PCA representing the total effects of genetic covariance (pleiotropy and linkage) of the mutant gene preferably compared against a near isogenic background with barley material grown in the same environment.

This involves not only the detection of chemical bonds of obvious interest (here from amide, starch and BG). The indicative wavelengths for chemical bonds can be found in the spectroscopic literature (Osborne *et al*., 1993) giving a hint of which chemical analyses that should be performed for validation of the supervised NIR classification (See example in Fig. 1C). It also engages unexpected correlations such as with water content for which NIR spectroscopy is very sensitive. Thus it is an interesting observation that the high BG content on the expense of starch in *lys5* seems to result in a higher content of dry matter (and lower content of water) by approximately 1.5% (Table IA-B). This is presumably due to that more molecular water is bound in crystalline starch in the amyloplasts compared to water bound to BG in the endosperm cell wall. Thus the specific effect of water associated to the *lys5* gene is also included in the spectral classification together with a broad range of other side effects from the mutant gene.

These pleiotropic effects are automatically summed up in the gene specific spectral fingerprint by a PCA which can be chemically defined *a posteori* after measurement.

Testing the chemical discriminative power of the spectral NIR-PCA model introducing a separate test-set of six mutants (Material 2)

Nine samples from the six new genotypes in Material 2 were analysed for NIR spectroscopy and included in the full spectral PCA model in Fig. 1B presented in Fig. 3A. Seven of the barley samples were grown in greenhouse. The two samples of mutant 95 were grown in the field and outdoor in pots.

Mutant 16, mutant 449 and w1 (noted as a waxy line) were all included in the BG rich cluster around *lys5* with the two mutant 95 samples above to the right.

w2 (also considered as waxy) was included in the upper part of the normal cluster closer to the *lys5* area, while the *lys4d* barley sample resides in the very high lysine *lys3* area below to the right.

We may now compare spectra of the known and unknown barley samples in the interesting area 2270-2380nm. In Fig. 3B it can be seen that mutants 95 and 449 resembles *lys5f*. The two mutant 95 samples grown in the field (finely pointed line) and outdoors in pots (coarsely pointed line) both have similar form but are shifted above baseline due to the environmental difference. Mutant 16 in Bomi has a spectral form in the 2270-2380nm area which resembles *lys5g* as seen in Fig. 3C. As expected from the classification in Fig. 3A the spectrum of *lys4d* (also mutant in Bomi) has a similar form as *lys3a* (Fig. 3C) indicating major change in amino acid composition. w1 in the *lys5* cluster has a spectral form resembling *lys5f* while w2 is near to normal barley (Bomi) (Fig. 3D).

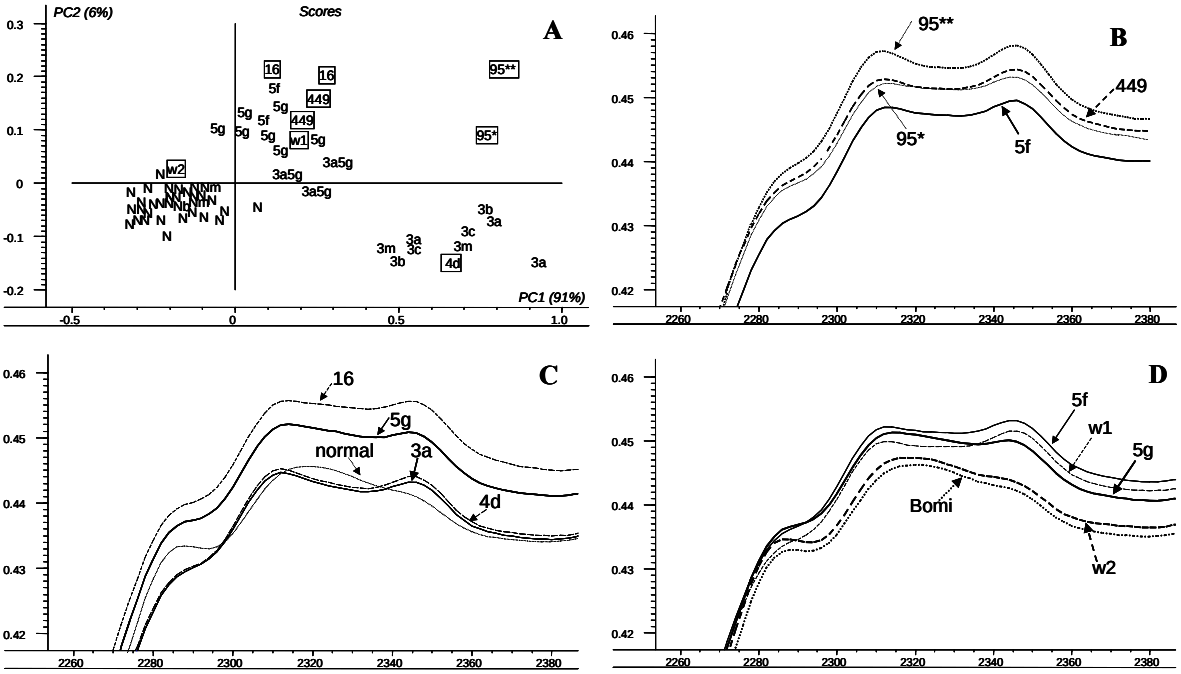


Figure 3A. PCA (PC1:2) of NIR (MSC) spectra from the 54 barley samples (Material 1) grown in greenhouse and nine samples of six different mutants in the test set (Material 2) discussed in the text. **B.** Comparing spectra (2260-2380nm) from the mutants grown in greenhouse *lys5f*, mutant 449 and mutant 95 grown in field* and in outdoor pots**. **C.** Comparing spectra (2260-2380nm) from the mutant *lys5g*, mutant 16 and normal barley. **D.** Comparing spectra (2260-2380nm) from the mutant w1, w2, *lys5f*, *lys5g* and bomi.

Chemical validation of the NIR-results from the test set

The result of the chemometric classification analysis of the spectral information in Figure 3A is validated in Table IA displaying the crude chemical composition of the six mutants, five of them grown in greenhouse except mutant 95, which have grown outdoors in the field and in pots (Table IB). All the new mutants positioned in the *lys5* cluster (or near to it like mutant 95) have strongly increased BG, namely mutant 16 (15.2%), mutant 449 (13.5%), and w1 (15.4%). The starch content of the new mutants is relatively more reduced than in *lys5g* and *lys5f*, having a starch plus BG level % d.m. clearly below the normal lines as well as with regard to *lys5g* and *lys5f* (Table IA). Also mutant 95 grown outdoors in the field and in pots is on the high BG side above the baseline in the PCA in Fig. 3A. It is high in BG (12.2 and 14.2%). It is clear that this mutant has a shifted position in the PCA plot because of environmental differences. As shown by its spectral form being near to *lys5f* (Fig. 3B) it should belong to the *lys5* cluster in the PCA score plot in Fig. 3B if grown in greenhouse.

lys4d, which was classified in the *lys3* cluster with changed amino acid pattern and low BG content, is confirmed to have a low A/P index and BG as is the case with *lys3a*, *lys3b* and *lys3m* (Table I). This is further verified with regard to the amino acid composition (lysine and glutamine) relative to Bomi in Table IA, which is significantly changed to the direction of *lys3a* in *lys4d*.

In evaluating the two proposed waxy mutants (Table IA and B) it is seen that w1 has a practically normal amylose content of 20.3% but is very high in BG (15.4%) and low in starch (42.7%). It can be concluded that w1 is not a classic waxy high amylopectin low amylose mutant with slightly increased BG, but rather a mutant in the new category of low starch/high BG mutants. w2 which has a NIR spectral form (Fig. 3D) closer to normal barley (Bomi) is waxy. It has a low amylose content (4.2%) and BG content on the high side (7.0%) compared to normal barley in Table I (mean 6.5%).

DISCUSSION: PERSPECTIVES ON (1→3,1→4)-β-GLUCAN REGULATING MUTANTS IN BARLEY

Recognising the genetic covariate spectral imprint of the mutant gene in the phenome

With the spectroscopic tools employed to cereal seeds, it is thus possible to detect unknown endosperm genes and mutants, which are deviating from normal lines by a PCA classification (Munck, 2003; Munck *et al*., 2001). In such a study it is difficult to differentiate between the two different sources of covariance pleitropy (biochemical gene applications) and linkage (association of adjecent genes on the chromosome). We therefore name the combined effect of pleitropy and linkage as “genetic covariance”.

This technology allows a truly exploratory strategy, with a minimum of *a priori* hypothesises, where the chemical effects of the gene are determined after selection with PCA with the spectra as preliminary guidelines (Osborne *et al*., 1993) for generating new hypotheses in a dialogue with *a priori* knowledge. In Table II we have connected the spectral NIR data from Figures 1 and 3 to the chemical data in Tables IA and B by a Partial Least Squares Regression (PLSR) analysis. Here relatively high correlations with BG, starch, protein and amide content are shown further explaining why NIR works for classifying barley mutants with PCA as demonstrated in Figures 1B and 3A. This example exploits the great capability of the NIR technology as a multimeter (provided reliable calibration), which is now, used world wide for quality control in agriculture and food products (Osborne *et al*., 1993). It is no doubt that such screening methods do have a promising more general future regarding overview and selection in plant breeding and in biotechnological research also involving cereal science and technology.

Table II. PLSR prediction of chemical properties from NIR (MSC) spectra

X	y	r	PC	RMSECV	RE	n	Outliers (n)
NIR	BG	0.87	2	2.32	13.0	60	
		0.92	2	1.78	10.0	59	1 (N)
	Starch	0.88	2	3.81	13.0	21	
	BG+S	0.75	1	4.40	17.3	20	
		0.78	1	4.27	16.8	19	1 (3a)
	Protein	0.93	5	0.55	6.1	63	
	Amide	0.86	4	0.03	8.8	63	
		0.93	5	0.02	7.4	60	3 (3m,N,95)
	A/P	0.89	3	0.97	10.6	63	
		0.95	5	0.61	7.9	60	3 (3m,3m,95)
	Fat	0.69	1	0.48	24.6	15	
		0.97	5	0.05	6.8	12	3 (3a, 5f, 5f)
	DM	0.80	4	0.55	14.0	63	
		0.87	4	0.43	12.4	62	1 (N)

Putative effect of a high BG content on earlier barley mutant screening methods

The high lysine Risø mutants were selected by DBC (Doll, 1983) using acilane orange that binds to basic amino acids (lysine, histidine and arginine) in an acid slurry of barley flour (Munck, 1972). After a reaction time, the sample was centrifuged and the supernatant measured in a short path spectrophotometer. It is remarkable that the *lys5* mutants have been able to be identified by the dye binding method when they only show an increase in basic amino acids by 6% (*lys5f* mutant 13) and 9% (*lys5g* mutant 29). Thus, a positive interference by the high BG content on the DBC measurements of these mutants could be hypothesised.

Likewise, we have found in our laboratory that it seems to be more difficult to define the waxy starch condition when testing whole barley flour high in BG in the classic potassium

iodine test. Thus, unknown plant breeders wrongly classified the w1 line as a waxy line with reduced amylose when it was included in our collection brought together from different sources at Carlsberg. As discussed above, w1 has a very high BG content but is normal in apparent amylose (Table I) (based on purified non defatted starch) although slightly at the low side (20.3%) compared to the normal range in apparent amylose of 21.3-25.8 (Tester *et al*., 1993).

Generating a hypothesis on carbohydrate regulation in the barley endosperm based on the six BG compensated starch mutants identified by NIR spectroscopy and verified by chemical analyses

Greber *et al*. (2000) suggested that the high lysine genes of barley could relate to carbohydrate synthesis and be used to understand starch synthesis. They tested 13 mutants where *lys3a, lys5f, lys5g* and Risø mutant 16 are of concern in this paper. *lys5f* (Risø mutant 13) showed an increased accumulation of ADP-glucose (ADPG) indicative for a reduction of starch synthesis (Greber *et al*., 2000). "A" starch granules were misshapen. However, the activities of starch related enzymes in this mutant are similar to or greater than those in wild-type barley at 14 days after pollination. There were no obvious differences in plastid membrane proteins. The other mutants listed above were near to normal in ADPG content. Practically all Risø *lys* mutants showed reduced seed size.

High to moderate levels of BG and high content of free sugars and even phytoglycogenes are in many mutant alleles associated with the amylopectin waxy gene (wax) in barley (Newman and Newman, 1992, Fujita *et al*., 1999). There are only slight reductions in starch level and seed size. It is interesting to note that the *lys5g* and mutant 16 here found to be high in BG has approximately normal levels of amylose (Tester *et al*., 1993). The sum of starch and BG in these mutants, as shown in Tables IA and B, however, approaches normal values in percentage of dry matter with *lys5g* as the best performer. Although mutant 16 seemed to be reduced by ten relative percentages in apparent amylose (Tester *et al*., 1993), it has a normal content of total amylose (from defatted starch) (Tester *et al*., 1993). In addition, the high lysine amino-acid mutants *lys3a* and *lys4d* in our study, displaying a reduction in BG are unchanged in amylose (Tester *et al*., 1993).

The BG content in barley reported here by the six BG compensated starch reduced mutants is extremely high compared to the review given e.g. by MacGregor and Fincher (1993) finding a range of 2.8-10.7% d.m. Newman and Newman (1992) mention a waxy, hulless, six rowed barley variety Azul grown under irrigation conditions in Arizona to have a BG content of 10-11%. There are amylose free waxy (wax) genes as well as alleles which contain amylose such as the w2 line (4.2% appearent amylose) reported in this investigation (BG 7.0%, see Table IA). The high amylose amo1 barley genes such as in Glacier AC38

(appearent amylose 40.6%) also have a moderately increased content of BG (7.9%) compared to a normal variety (BG 4.7%, appearent amylose 33.1%) as reported by Fujita *et al*. (1999). Swanston *et al*. (1995) demonstrated that the double recessive between a waxy (non amylose free) and the amo1 genes had 9.4% in BG compared to 3.6% of the controls. This was confirmed by Fujita *et al*., 1999 with a waxy amylose free allele combined with the gene amo1 in a double recessive that reached the high BG level of 12.4%, about 2.6 times higher than the control line. In this paper the six BG compensated starch mutants have a range in BG of 8.9-16.5% as grown in the field and 13.3-19.8% as grown in greenhouse compared to a mean of 4.5% and 6.5% respectively for a set of control varieties (Tables IA and B). This change in BG amounts from 2.0 to 3.7 times.

The BG values given here are analysed by the Calcofluor binding method based on fluorescence published by Munck *et al*. (1989). This method measures after an acid hydrolysis high molecular BG approximately above 10-20 kD with isolated barley BG as a standard. We have preliminarily checked this method with the current specific enzymatic (1→3, 1→4)-β-glucan kit (Anonymous, 1998) on the market and found a linear relation with a minimal offset up to at least 15% BG between the two methods. The extreme high levels of BG reported here for the six barley mutants are thus confirmed.

Real high lysine mutants seem to have reduced BG content. These are as analysed (Table IA) here *lys3a, lys3b* and *lys3m* in chromosome 5 and *lys4d* in chromosome 1. Note that the allele in the *lys3* locus *lys3c* has a normal BG content. This fact is likely to suggest that the high lysine and BG reducing traits here are controlled by adjacent genes and that the mutations involve chromosome segments of different lengths around the *lys3* locus in chromosome 5. There is a positive significant correlation between the A/P index and the BG content of *lys3* genotypes of r=0.83, which indicates a position effect of the different alleles. The reduced BG content should thus not be pleiotropic to the *lys3a*, *lys3b* and *lys3m* alleles but rather depend on a very tight linkage (which is very difficult to break by recombination) to an adjacent gene, which retards BG synthesis in three of the four *lys3* genotypes. It is in this context interesting to note that the real high lysine *lys4d* mutant gene (in another locus and chromosome) is also associated with reduced BG level compared to its mother variety Bomi (Tables IA and B).

BG is synthesised late in the kernel development and is dependent on environmental factors such as heat and precipitation. It is in normal barley inherited by a simple genetic additive system (Powell *et al*., 1985). When classifying barley spectra according to genotype with PCA, we need to utilise a comparable material grown in the same environment (e.g. greenhouse). In Tables IA and B, we have compared the BG content of the genotypes

presented here when grown in the field and in the greenhouse. Although the BG content is higher under greenhouse conditions, the differences between most mutants and normal barley are consistent. For example, the extreme mutant *lys5f* has a BG content in the greenhouse condition 19.8% compared to 16.5% in the field, while its mother variety Bomi was 6.8% and 5.2% respectively.

We are thus able to propose a hypothesis based on a new class of six high BG compensated starch mutants the nature of which was first identified with NIR spectroscopy as a screening method. These genes regulate the BG synthesis to be closely coupled and appearently competitive to starch synthesis in the developing endosperm. The sugar precursors available are shifted in their destiny accordingly, the cause of which may be found by further investigating the proteomes of the developing endosperm of the six genotypes by classic biochemical methods. Three of the six BG compensating starch mutants are confirmed to have a normal starch amylose to amylopectin composition. At least two gene loci *lys*5 (alleles *lys5f* and *lys5g*) in chromosome 6 and mutant 16 in chromosome 7 are involved.

CONCLUSION

In this investigation we have classified by spectroscopy ten of the classic "high lysine" barley mutants of the 20-30 available. We have found that half of those (*lys5f, lys5g,* mutant 16 in Bomi, and mutants 95 and 449 in Perga) combine low starch with excessive BG synthesis completely or partly compensating for the decrease in starch formation. As shown by Tester *et al*. (1993) *lys5g* (mutant 29) and mutant 16 were non waxy having a normal amylose content. This is also the case with the high BG low starch line 1201 w1 mutant identified in this investigation, previously wrongly classified as waxy. Thus three of the six high BG compensating starch mutants described in this paper has proved to have normal amylose content. The mechanism of high BG content in barley discussed before has in literature mainly been associated to either low (waxy) or high amylose (amo 1) genes. To our knowledge, data combining normal amylose barley with BG levels approaching 20% (*lys5f*, Table IA) have not been published before.

BG is of major importance in cereals for food because of its positive properties in human nutrition as a dietary fibre (McCleary and Brosky, 2001) for regulating the flow in the digestive system, lowering the cholesterol value in the blood and reducing the risk for colon cancer (McIntosch *et al.,* 1991). BG is also of importance for stimulating the immune system. Because the human body is hardly able to utilise BG as an energy resource, the energy value of a cereal product high in BG is correspondingly reduced implying a decrease in blood glucose as expressed in the glycemic index (GI) of great importance in

diabetes prevention and treatment (Salmeron *et al*., 1997). The very high BG level of the six barley BG mutants presented here makes them well suited as candidates for ingredients in functional foods low in energy and with a high level of soluble and insoluble fibre. However, these mutants challenge the food technologist to utilise this resource in development of food products of high sensory and nutritional value. They also give new options to the biotechnologist and the plant breeder in further improving the nutritional and technological value of high BG cereals e.g. by gene transfer of BG compensated starch reduced genes from barley to wheat to improve baking properties.

The introduction of spectroscopy and chemometrics to reveal gene expression patterns as discussed here and by Munck *et al*. (2001) and Munck (2003) on the level of the phenome. Watkins *et al*. (2001) has, however, further more general applications in biological research. Chemometric pattern recognition statistical methods such as PCA and PLSR models are now starting to be used more frequently in molecular biology to connect between different levels of biological organisation (Fiehn, 2002). Unsupervised screening methods such as NIR spectroscopy evaluated by chemometrics should be effective in revealing new metabolic mechanisms which later could be defined by classic biochemical analysis as demonstrated in this paper.

In a following article (Jacobsen *et al*., 2004) we will use the barley mutants from our barley gene collection as a model to expand on how the digitised spectral phenome and its genetic covariate imprints can be further realised and chemically understood by chemometrics and validation to classic biochemical analyses.

ACKNOWLEDGEMENTS

Centre for Advanced Food Research is gratefully acknowledged. Anne Blicher is thanked for performing amino acid analyses. In addition, Lisbeth Hansen, Kirsten Wilms and Anna-Marie Hansen are greatly appreciated for their crucial assistance analysing the samples for ß-glucan, protein and amide and making the spectroscopic measurements. Lars Nørgaard and Rasmus Bro at our chemometric group at KVL have kindly supported this investigation by valuable advice.

REFERENCES

Anonymous, 1998. ICC-standards, Verlag Moritz Schäfer, Detmold, Germany.

BeMiller, J.N., 1964. Iodometric Determination of Amylose - Amperometric Titration.
Methods in Carbohydrate Chemistry IV, Academic Press, 165.

Campbell, M.R., Sykes, J., Glover, D.V., 2000. Classification of single and double-mutant corn endosperm genotypes by near-infrared transmittance spectroscopy. *Cereal Chemistry* **77**, 774-778.

Di Fonzo, N., Stanca, A.M., 1977. EMS derived barley mutants with increased lysine content. *Genetica Agraria* **31**, 401-409.

Doll, H., 1983. Barley Seed Proteins and possibilities for their improvement, in: Gottchalk, W., Muller, H.P. (Eds), Seed Proteins, Martinus Nijhoff/Dr. W.Junk Publishers, The Hague, 205-223.

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

Fujita, M., Domon, E., Doi,Y., 1999. Grain and starch characteristics for the double recessive lines for amylose-free and high amylose gene in barley. *Breeding Science* **49**, 217-219.

Greber, B., Waite, D., Fahy, B., Hylton, C., Parker, M., Laurie, D., Smith, A.M., Denyer, K., 2000. Use of barley mutants to understand starch synthesis. *Proceedings 8th International Barley Genetics Symposium*. Adelaide October 2000. Volume 1, 196-198.

Jacobsen., S, Søndergaard, I., Møller, B., Desler, T., Munck, L., 2004. The barley endosperm as a data interface for the expression of specific genes at different levels of biological organisation explored through pattern-recognition data evaluation. *J.Cereal Science* submitted.

MacGregor, A.W., Fincher, G.B., 1993. in: MacGregor, A.W., Bhatty, R.S. (eds.), Barley - Chemistry and Technology, American Association of Cereal Chemists St. Paul, 247.

McCleary, B.V., Prosky, L., 2001. Advanced Dietary Fibre Technology, MPG Books Ltd, Bodmin, Cornwall.

McIntosch, G.H., Whyte, J., McArthur, R., Nestel, P.J., 1991. Barley and wheat foods: Influence on plasma cholesterol concentration in hypercholesterolemic men. *American Journal of Clinical Nutrition* **53**, 1205-1209.

Martens, H.and Næs, T., 1989. Multivariate Calibration, John Wiley, Chichester.

Munck, L., 1972. Improvement of nutritional value in cereals. *Hereditas* **72**, 1-128.

Munck, L., 1992. The case of high lysine barley breeeding, in: Shewry, P.R. (ed.), Barley: Genetics, Molecular Biology and Biotechnology, CAB International
Wallingford, U.K. 573-602.

Munck, L., Karlsson, K.E., Hagberg, A., Eggum, B.O., 1970. Gene for improved nutritional value in barley seed protein. *Science* **168**, 985-987.

Munck, L., 2003. Detecting diversity - a new holistic, exploratory approach bridging the geneotype and phenotype, in: Diversity in Barley *(Hordeum vulgare*), von Bothmer, R., van Hintum, T.J.L., Knupffer, H., and Sato, K. (eds.). Elsevier Science B.V., Amsterdam, The Netherlands, Chapter 11, 227-245.

Munck, L., Jørgensen, K.G., Ruud-Hansen, J., Hansen, K.T., 1989. The EBC methods for determination of high molecular weight ß-glucan in barley, malt, wort and beer. *Monatsschrift für Brauwissenschaft*, **4**, 162-166.

Munck, L., Nielsen, J. P., Møller, B., Jacobsen, S., Søndergaard, I., Engelsen, S.B., Nørgaard, L., Bro, R., 2001. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. *Analytica Chimica Acta* **446**, 171-186.

Muravenko, O.V., Badaeu, N.S., Badaeva, E.D., Kost, M.V., Pukhalskiy, V.A., Zelenin, A.V., 1991. Standardization of chromosome analysis of barley, in Munck, L. (ed), Barley Genetics VI, Munksgaard International Publication, Copenhagen, Vol. I, 293-296.

Newman, C.W., Newman, R.K., 1992. Nutritional aspects of barley seed structure and composition, in: Shewry, P.R. (ed.), Barley: Genetics, Molecular Biology and Biotechnology, CAB International,Wallingford, U.K. 351-368.

Osborne, B.G., Fearn, T., Hindle, P.H., 1993. Practical NIR spectroscopy with applications in Food and Beverage analysis, Longman Scientific and Technical, Harlow, U.K.

Powell, W., Calgari, P.D.S., Jinks, J.L., 1985. Genetical investigation into ß-glucan content in barley. *Theor. App. Gen.* **71**, 461-466.

Salmeron, J., Manson, J.F., Stampfer, M.J., Colditz, G.A., Wing, A.L., Willett, W.C., 1997. Dietary fiber, glycemic load and risk of non-insulin dependent diabetes mellitus in woman, *Journal of the American Medical Association* **277**, 472-477.

Swanston, J.S., Ellis, R.P., Starkt, J.R., 1995. Effects on grain and malting quality of genes altering barley starch composition. *J.Cereal.Science* **22**, 265-273.

Tester, R.F., Morrison, W.R., Schulman, A.H., 1993. Swelling and gelatinization of cereal starches. V. Risø mutants of Bomi and Carlsberg II barley cultivars. *Journal of Cereal Science* **17**, 1-9.

Watkins, M.S., Hammock, B.D., Newman, J.W., German, J.B., 2001. Individual metabolism should guide agriculture towards foods for improved health and nutrition.
American Journal of Nutrition **74**, 283-286.

Paper VII

The barley endosperm as a data interface for expression of genes and gene combinations at different levels of biological organisation explored through pattern-recognition data evaluation

Susanne Jacobsen, Ib Søndergaard, Birthe Møller, Torben Desler and Lars Munck

---

VII

ABSTRACT

The ripe barley endosperm is considered as a data interface for the expression of mutants and gene complexes. It is based on a stable physical chemical structure which can be read by Near Infrared Reflection spectroscopy (NIR), interpreted by pattern recognition data analysis (chemometrics) and validated to biochemical and genetic data. The chemical background for the recent identification by NIR and Principal Component Analysis (PCA) of two (1→3,1→4)-β-glucan (BG) compensated low starch mutants (*lys5f* , *lys5g)* earlier selected as moderately high lysine is studied as a model together with the extreme high lysine mutant *lys3a* in near isogenic backgrounds and in crosses. 25 chemical analyses gave a gene classification in a PCA comparable to that of NIR. Plots between BG and ratios e.g. amide/protein were also effective. Six proteins were followed during seed synthesis giving specific gene expression patterns by PCA. NIR by PCA is effective in selecting for a changed gene background to the *lys3a* mutant, improving starch and reducing fibre. *lys5g* shows minor changes in the proteome pattern compared to its higher expression allele *lys5f* and the *lys3a* mutant. It is concluded that a gene-specific physical-chemical pattern from the endosperm phenome is almost as stable as its initial state - the DNA sequence of the corresponding gene.

INTRODUCTION

In this paper we aim at further investigating the chemical background of the barley endosperm as a data interface in order to explain the physical (spectral) gene specific patterns obtained by Munck *et al*., 2004 for a new class of (1→3,1→4)-β-glucan (BG) compensated low starch mutants. The mutants are compared to the mother lines and other normal barley's. Two BG-starch mutant alleles one extreme (BG:16.5-19.8%) - *lys5f* and one less extreme (BG: 8.9-13.3%) - *lys5g* and one very high lysine mutant (148% increase) - *lys3a* are selected. Additionally the double recessive recombinant *lys3a5*g is included. We will use proteome analysis (2-DE) including amino acid seqencing of “hot” spots and a range of other chemical analyses including amino acids in order to define differences between the genotypes. Chemometric pattern recognition analysis through PCA and PLSR is performed on data with a set of 25 analytes to study the specific effects of mutant genes in different gene backgrounds. These results are compared to the PCA classification of the same material by Near Infrared Reflection (NIR) spectroscopy of the same samples, in order to confirm the feasibility of a total spectral physical-chemical fingerprint to represent the phenome (Watkins *et al*., 2001) of the endosperm tissue in digitised form (Munck *et al*., 2004).

MATERIAL AND METHODS

Two barley materials were used:

Material 1. 34 barley samples earlier analysed by NIR spectroscopy (400-2500nm) by Munck *et al*. (2004) were analysed. The material was grown outdoors in the field and in pots and indoors in greenhouse at KVL as indicated in the Figures. The mutant genotypes including original mutants and mutant recombinants from crosses originate from Risø National Laboratory, Denmark (Doll, 1983). They were originally selected as the moderate high lysine mutants - *lys5f* (mutant 13 in Bomi) and *lys5g* (mutant 29 in Carlsberg II) and the very high lysine mutant *lys3a* (Risø mutant 1508 in Bomi). The double recessive recombinant *lys3a5g* was also included. The mutants and their mother varieties were compared to a set of normal barley cultivars (n=16).

Material 2. Consists of the original *lys3a, lys5g, lys3a5g* genotypes and the control varieties Bomi and Carlsberg II grown at the farm of the Carlsberg Research Laboratory analysed by Desler (1987).

Material 3. Consists of 84 samples in total. Two *lys3a* mutants and 26 *lys3a* segregants, three *lys5g* and ten *lys5g* segregants as well as 39 normal varieties including two samples of the *lys3a* mother line Bomi and two samples of *lys5g* motherline Carlsberg II. The samples have been grown outdoors in 1991 and 2000.

Material 4. Consists of three high lysine *lys3a* semi-commercial varieties Piggy, Lysimax and Lysiba selected from a material from Carlsberg Plant breeding (Munck, 1992) improved by cross breeding for better yield and grain quality resulting in improved starch content. Three normal controls are also included. The material is grown in the field.

2-DE: The supernatant of the alcohol and salt extracts were subjected to 2-DE in duplicate for each sample and genotype (Material 1) comparing both greenhouse and field conditions according to Jacobsen *et al*. (2001). The differences recorded were confirmed from samples in both environments. Here (Fig. 1 and 2) only the results from the greenhouse samples are reported. Briefly, the extract (30 µl) was mixed with rehydration solution (350 µl) containing 8M urea, 2% CHAPS, 0.3% DTT, 0.5% IPG buffer, pH 3-10 L, and a trace of orange G. The complete solution (400 µl) was loaded on an Immobiline DryStrip (18 cm length), pH 3-10 L (Amersham Biosciences), and run on a IPG-pher electrophoresis unit. After IEF, the strips were equilibrated for the second dimension, casted 12.5% on ready-made SDS-gels (ExelGel SDS gradient 12-14% Amersham Biosciences) according to the manufacturer. The casted gels were run on an Ettan Dalt six (Amersham Biosciences) and ready made gels were run on a Multiphor II unit. After SDS-PAGE the proteins were fixed and stained with Coomassie Brilliant Blue G-250.

Immunochemical analyses of six water and buffer soluble barley proteins during grain filling were made by Desler (1987) using radial immunodiffusion on agar slabs stained by

Commasie Blue according to Jonassen (1980). Endosperms were harvested at different times after anthesis, freeze-dried and milled in an agate pistle mill and extracted by buffer in a whirly mixer as described by Jonassen (1982). The immuno assay was kindly obtained at the Carlsberg Research Laboratory, Valby, Denmark from Dr. John Mundy now at the University of Copenhagen.

Amino acid analysis: Amino acid hydrolysis of samples was performed in evacuated and sealed tubes at 100^{o}C for 24h (Barkholt and Jensen, 1989).
N-Terminal amino acid sequencing of individual protein spots electroblotted onto PVDF-membranes from 2-DE electrophoresis was carried out by Edman degradation on a Procise 494 protein sequencer (Applied Systems, Foster City, CA, USA) according to Radova *et al*. (2001).
Gross chemical analyses and NIR spectroscopy were made according to Munck *et al*. (2004).

PCA and PLSR were transferred from the social and economic statistics to chemistry as chemometrics around 1970 where the application of these algorithms were further developed by cross validation and outlier control (Martens and Næs, 1989) as seen in the "Unscrambler" software version 7.5 (Camo A/S Trondheim, Norway) which was used here. Chemical data was normalised and the NIR spectra were multiplicative signal corrected (MSC). All results were crossvalidated.

RESULTS AND DISCUSSION

2-DE proteome analysis of seeds from mutants and reference barleys comparing the alcohol soluble (hordein) fraction (Material 1) with the salt-soluble. Samples are selected from Material 1

In genetics and biotechnology genes are classified as regulators communicating with gene promoters in different parts of the genome and as structural genes coding for specific proteins e.g. enzymes of importance for carbohydrate synthesis.

The 2-DE analysis show in general that the proteome pattern is specific for each genotype where the environmental influence is much less specific.
As expected the very high lysine *lys3a* genotype (Risø 1508) which is a regulator (Blom Sørensen *et al*., 1996*)* is extreme with regard to reducing a range of alcohol soluble proteins (Fig. 1E) while increasing several salt soluble (Fig. 2E). The *lys3a5g* double recessive is dominated by the *lys3a* influence.

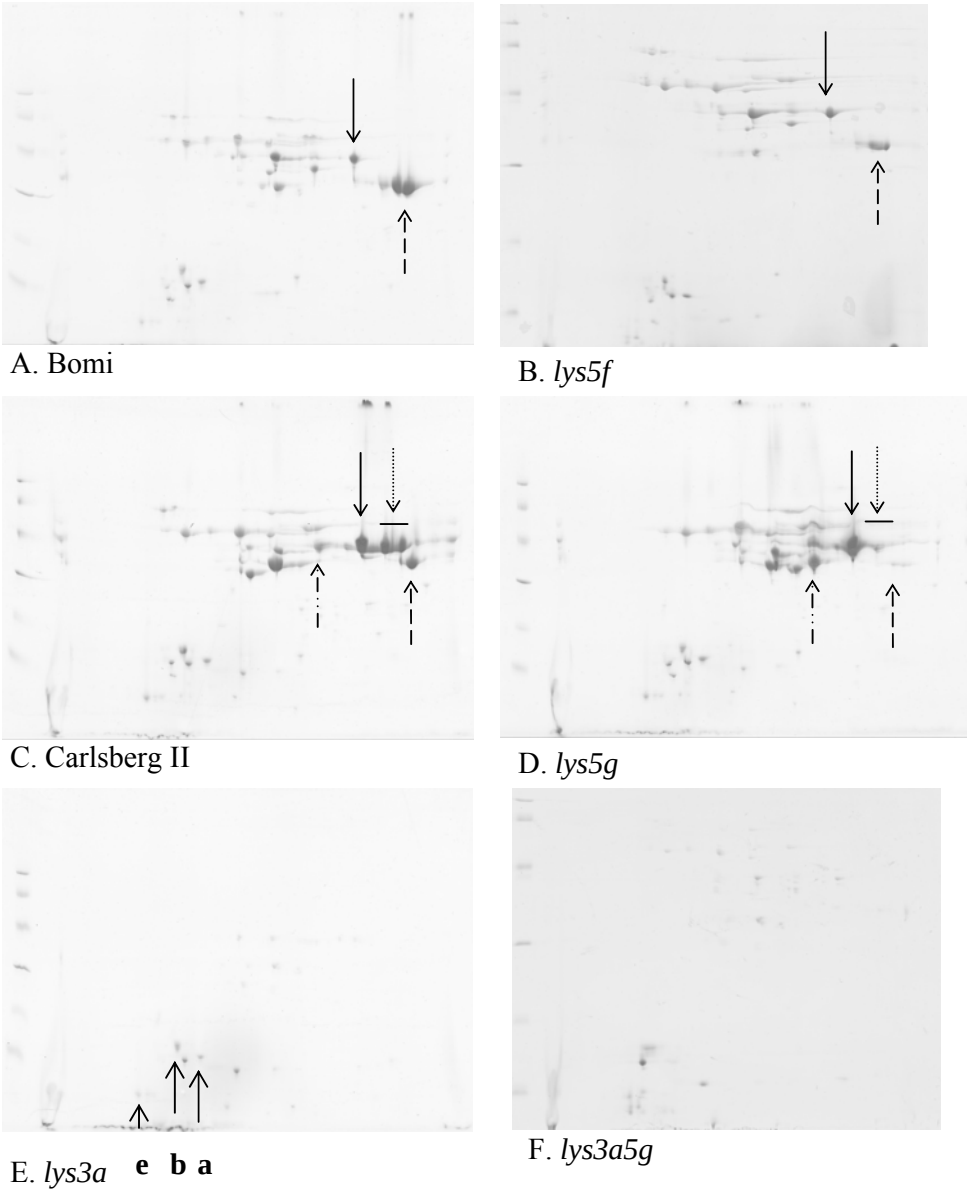
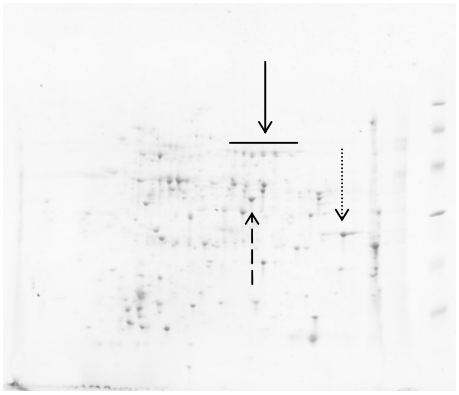


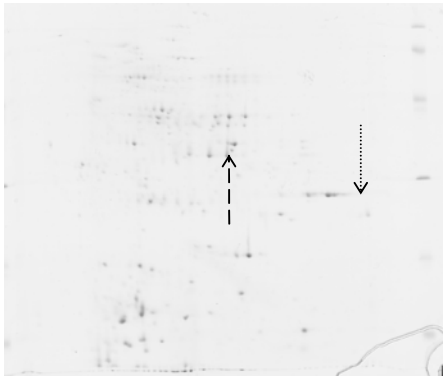
Fig. 1. 2-DE of the alcohol soluble protein fraction (hordeins). All barley samples were grown in greenhouse. First dimension electrophoresis all gels: Immobilized pH gradient 3-10. Second dimension electrophoresis, gels A, C, D and E: Ready-made gels 12-14% run on the Multiphore equipment. Gels B and F: Cast gels 12.5% run on the Ettan Dalt six equipment.

The molecular weight markers to the left are from top 94, 67, 43, 30, 20.1, 14.4 kDa. The B-hordein can be seen in the range 35-55 kDa, C-hordeins 55-85 kDa and D-hordeins 105 KDa.

– – – – – – →	1	See results and discussion
·····→	2	See results and discussion
– · – · – · – →	3	See results and discussion
——→	4	No sequence obtained. Presumeably N-terminal blocked
	a	alfa-amylase/trypsin inhibitor Cma
	b	alfa-amylase/trypsin inhibitor Cmd
	e	subtilisin-chymotrypsin inhibitor Cl-1B

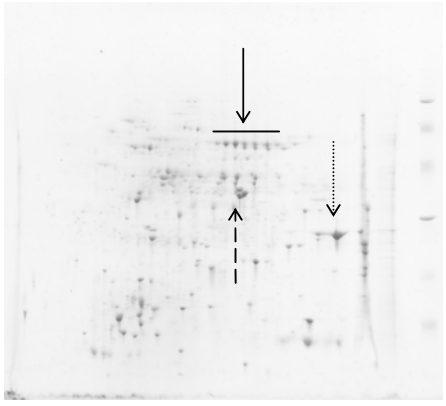


A. Bomi

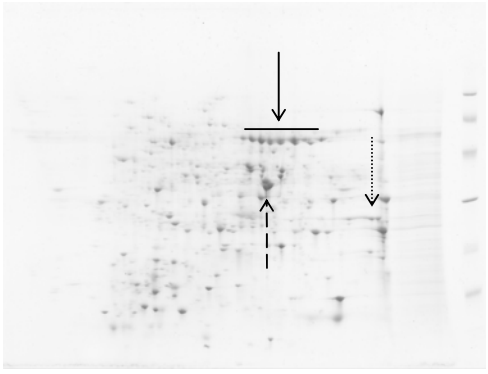


B. *lys5f*

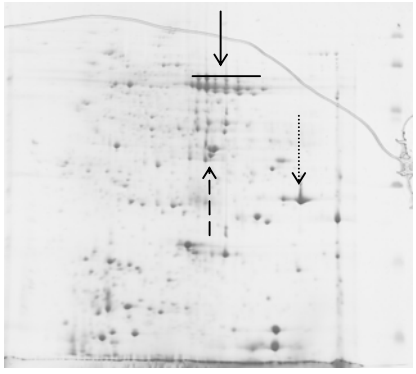
C. Carlsberg II



D. *lys5g*



E. *lys3a*



F. *lys3a5g*

Fig. 2. 2-DE of salt soluble protein fraction (albumins). All barley samples were grown in greenhouse. First dimension electrophoresis all gels: Immobilized pH gradient 3-10. Second dimension electrophoresis, gels A, C, D and E: Ready-made gels 12-14% run on the Multiphore equipment. Gels B and F: Cast gels 12.5% run on the Ettan Dalt six equipment.

- - - - - - →	c	Peroxidase BP1
·····→	d	Chitinase C
——→	f	N-terminal blocked

The *lys5f* mutant is almost as drastic as *lys3a* but shows an opposite reaction. It has a near unchanged alcohol soluble protein pattern (Fig. 1B) compared to its mother variety Bomi (Fig.1A) but is greatly reduced in salt soluble proteins (Fig. 2B versus 2A). The rows of six proteins (full line arrow), which are available in Bomi (Fig. 2A), Carlsberg II (Fig. 2C) and *lys5g* (Fig. 2D) in comparative amounts, are greatly reduced in *lys5f* (Fig. 2B) while they are much increased in *lys3a* (Fig. 2E) and *lys3a5g* (Fig. 2F). In general, the two latter genotypes show an increased level of most salt soluble proteins. The protein pattern of the salt soluble fraction in the mutant *lys5f* (Fig. 2B) is thus over all weak with seemingly missing spots.

The surprisingly large differences between the normal varieties Bomi and Carlsberg II are demonstrated in detail in the protein patterns of the hordeine fraction in Fig. 1A and Fig. 1C. The proteins at the mark 2 and 3 arrows points out protein spots in seemingly new positions in Carlsberg II compared to Bomi. At the mark 1 arrow a protein spot to the left is missing in Carlsberg II compared to the pattern in Bomi.
The differences between Carlsberg II (Fig. 1C, Fig. 2C) and its mutant *lys5g* (Fig 1D, Fig 2D) are much less than those between Carlsberg II and Bomi especially with regard to the hordeins. However the *lys5g* mutant has increased production compared to its mother line of the proteins marked with arrows in both the protein 2-DE analyses.

The *lys5g* gene having minor effects on the proteome patterns might thus be a candidate for a specific structural gene involved in starch and BG synthesis. However, as we will see in the following the *lys5g* gene has major genetic covariate effects outside the realm of carbohydrate synthesis. Like its allele *lys5f* and *lys3a* it mediates an increased fat content in the seed. Greber *et al*. have shown that *lys5f* shows an increased accumulation of ADP glucose indicative for a reduction of starch synthesis while *lys5g* is uneffected. It could thus be tentatively suggested that the major lesion of the *lys5f* mutant affects a chromosome segment with several more genes compared to *lys5g*. In studying gene specific spectral and chemical patterns in the barley endosperm it is not possible to differentiate between the indirect biochemical consequences of the gene (in genetics called pleiotropy) and linkage with adjacent genes on the chromosome. We will therefore in the following use the term "genetic covariance" for the combined effect of pleiotropy and linkage.

Selecting proteins spots changed by the mutants for confirmation of function by protein sequencing
The drastic hordein effect on the *lys3a* gene is confirmed in Fig. 1E and Fig. 1F. The remaining spots below in the lower left corner are low molecular weight proteins (10-15 kDa), which are also extracted with the salt solution as displayed in the same gel area in Fig. 2A-F. As these proteins are well separated from the other proteins in the gel and are

not N-terminally blocked, it is easy to N-terminal protein sequence these proteins. They also show quantitative variation between different genotypes. The spots marked a, b and e in Fig. 1E were sequenced (Table I) and will be further commented in the discussion.
Salt soluble protein spots (arrows c, d and f in Fig. 2) were analysed by N-terminal protein sequence analyses. However, results showed that the proteins in the area marked with arrow f were N-terminally blocked, because no sequence was seen even though the high amounts of these proteins characterising *lys3a* (Fig. 2E) and *lys3a5g* (Fig. 2F) were applied to the N-terminal protein sequencer.

Especially in *lys5g* (Fig. 2D), *lys3a* (Fig. 2E) and *lys3a5g* (Fig. 2F) the peroxidase (arrow marked c) is expressed in high amounts. The amounts of peroxidase (c) in normal barley Bomi (Fig. 2A) and Carlsberg II are low. In mutant *lys5f* (Fig. 2B) it is comparable to Bomi. Bomi and Carlsberg II do have comparable salt soluble protein fractions both when analysing the protein pattern and the individual amounts in protein spots.

The chitinase (Table I) spot marked d in Fig. 2, is especially high in *lys5g* (Fig. 2D), *lys3a* (Fig. 2E) and *lys3a5g* (Fig. 2F) compared to the other mutants and normal barley.

Table I. Sequences of protein from blots from 2-DE gels (See positions marked in the 2-DE gels in Fig. 1A and Fig. 1B.

*	**Protein**	**Sequence**	**Reference**
a	alfa-Amylase/trypsin inhibitor CMa	TGQYCYAGMGLPSNP	Barber et al, 1986
b	alfa-Amylase/trypsin inhibitor CMd	AAAATDCSPGVAFPT	Shewry et al, 1984
c	Peroxidase BP1	AEPPVAPGLSF	Johansson et al, 1992
d	Chitinase C	SVSSIVSRAQF	Leah et al, 1987
e	Subtilisin-chymotrypsin inhibitor CI-1B	MEGSVPKYPE	Williamson et al, 1988

We have in total sampled six characteristic spots which are increased in the mutant genotypes *lys3a*, *lys5g* and *lys3a5g* marked a-f marked in the 2-DE gels in Fig. 1 and 2 and sequenced them (Table I). It is interesting to note that they are all in the antimicrobial category. The protein fractions were extracted from dry desiccated kernels. At this stage the isolation of the metabolic proteins e.g. responsible for the BG/starch effects of the mutant genes *lys5f* and *lys5g* from the partly degraded endosperm should not be expected. Instead storage proteins and antimicrobial proteins of importance for securing the food store for the embryo should prevail. It is clear that in studying the biochemical effects of mutants, proteome or metabolome fingerprints from a series of developing endosperms should be preferred. We will now present such a preliminary study.

Immunological analysis of six antimicrobial proteins in developing seeds of the lys3a, lys5f, lys5g, lys3a5g mutant genotypes and the reference lines Bomi and Carlsberg II (Material 2)

In consulting our previous research we find an unpublished experiment (Desler, 1987) using immuno chemical methods to follow the seed synthesis of antimicrobial proteins. It involves the *lys3a, lys5g, lys3a5g,* Bomi and Carlsberg II genotypes and the chymotrypsin inhibitors CI-1 and CI-2, the protein synthesis inhibitor (PSI), the barley α-amylase subtilisin inhibitor BASI, the chitinase C and protein Q (Mundy *et al*., 1986; Hejgaard and Bjørn, 1985). Desler's Protein C is identical with Chitinase C in Table I. The characteristic synthesis curves for the six proteins and five genotypes are seen in Fig. 3A. In all cases, the double recessive has the most rapid synthesis of all six proteins. The mother varieties for the mutants Bomi and Carlsberg II have less synthesis at the different sampling times compared to the mutants in most cases. *lys3a* is manifested as a slow starter in comparison with *lys5g* of CI-1, CI-2, BASI, Q and C proteins, but ends up with larger amounts of synthesis of these proteins at harvest.

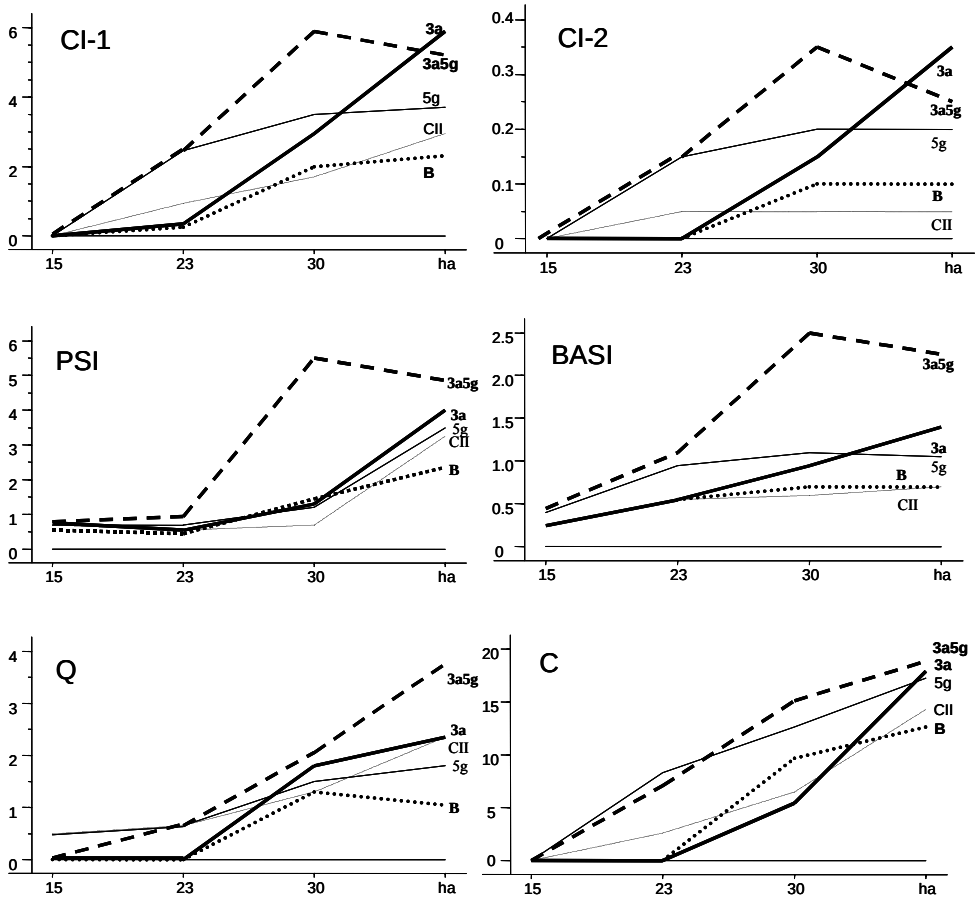


Fig. 3A. Development of the anti-microbial proteins CI-1, CI-2, BASI, PSI, Q and C (mg/g flour) at 15, 23, 30 days after anthesis and after harvest (ha). Radial immunodiffusion by Dessler (1987). X-axis: days after anthesis, ha=harvest. Y-axis: Quantification of protein for B=bomi, CII=Carlsberg II, 3a=*lys3a*, 5g=*lys5g*, *3a5g*= double recessive *lys3a5g*. For discussion see text.

By making a global PCA on the protein synthesis curves in Fig. 3A we arrive at a convenient overview in Fig. 3B modelling the response pattern of the six proteins for each genotype and sampling time. The trajectories reveal that *lys5g* and *lys3a5g* are the fastest starters followed by Carlsberg II, Bomi and *lys3a*. From day 23 after anthesis, *lys3a* accelerates considerably when its protein synthesis leads to very high final levels at harvest. This is also the case for *lys3a5g* indicated by the extreme positions of these samples to the right in the plot. Thus the qualities of the two genes *lys5g* and *lys3a* of fast protein synthesis in the start and in the end of the kernel formation period respectively are combined in the double recessive *lys*3*a5g*. We can thus conclude that each mutant gene performs a characteristic chemical fingerprint in the "endosperm display" which is available in the 2-DE patterns in Fig. 1 and 2, as well as in the PCA of the set of six arbitrarily chosen proteins (Fig. 3A and B). We will now further investigate the gene and environment discriminating ability of a new set of chemical analyses by chemometrics and compare it with that of NIR-spectroscopy.

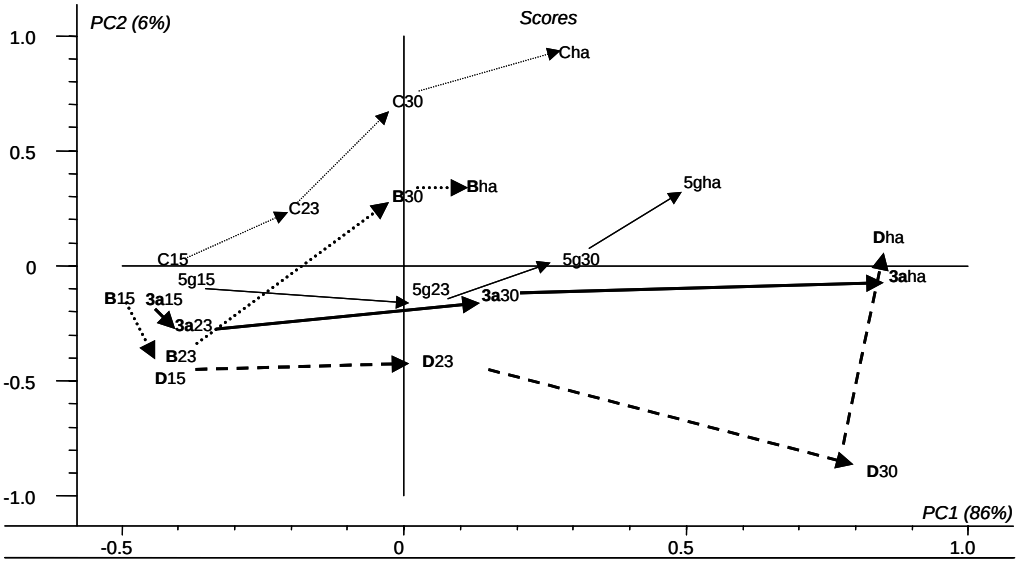


Fig. 3B. Compression PCA regarding the six anti-microbial proteins CI-1, CI-2, BASI, PSI, Q and C (mg/g flour) for each sample at four developmental stages (15, 23, 30 days after anthesis and after harvest (ha)). Five genotypes: B=bomi, C=Carlsberg II, 3a=*lys3a*, 5g=*lys5g*, D=double recessive *lys3a5g*.

Establishing a chemical data set of barley mutants reference lines, mutant segregants and normal varieties (Table II, material 1)

The data set involves 27 chemical variables of 25 analyses from 34 samples grown in three environments involving eight gross chemical analyses and 17 amino acids. The high BG low starch status of *lys5g* and even more *lys5f* is clearly seen in Table II. The BG compensating effect (BG +S) in this material is about 90% when comparing to the Bomi and Carlsberg II controls (Table II). There is a clear tendency of a higher content of dry matter (DM) in the BG rich low starch mutants. This is presumably due to more crystalline

bound water in starch which should be less than in the cell wall bound to the amorph BG. The *lys3a* and *lys3a5g* genotypes have as expected high lysine and low glutamic acid content. They also have a low content of amide nitrogen and amide to protein (A/P) index. *lys3a* is low in BG while the *lys3a5g* double recessive is intermediate. The small difference in hordeins between *lys5f* and *lys5g* on one side and their control varieties Bomi and Carlsberg II respectively on the other as seen in the 2-DE results are reflected in the small changes in the amino acid patterns from the wildtype to the mutant.

It is surprising to note that a high fat content in the seed is not only asociated with the *lys3a* gene (Munck 1992) but also with the *lys5* gene with both *f* and *g* alleles. The fat content is doubled in *lys5f*. There is a tendency for an increase in insoluble fibre (e.g. lignocellulose) in mutants with decreased starch and seed size. Soluble fibre% follows the BG content.

Table II. Relative values (normal = 100) for chemical properties of *lys3a*, *lys5g*, *lys5f*, *3a5g* and normal varieties (Material 1)

		Normal = 100	**Bomi***		**CAll***		***lys3a***		***lys5g***		***lys5f***		***3a5g***	
	n	16	2	Rel	1	Rel	4	Rel	6	Rel	5	Rel	3	Rel
1	Starch	55,1	52.3±4.9	95	56.8	103	51.0±0.9	93	46.2±4.0[f]	84	30.0±0.8[g]	54	28.3±1.6[g]	51
2	BG	4,9	6.0±1.1	122	4.8	98	3.2±0.6	65	10.5±2.3	214	17.8±2.4[h]	363	8.6±0.1[g]	176
3	S+BG	59,8	56.6±1.4	95	61.6	103	51.8±4.5	87	53.4±5.4[f]	89	50.5±1.7[g]	84	47.4±16.5[g]	79
4	DM	90,4	90.5±0.3	100	90.0	100	90.1±0.3	100	91.5±0.5[f]	101	91.8±0.6[g]	102	91.0±0.5[g]	101
5	Amide	0,3	0.34±0.06	103	0.25	76	0.25±0.04	76	0.34±0.07	103	0.38±0.03[e]	115	0.35±0.06	106
6	A/P	16,0	16.3±0.1	102	15.7	98	10.9±0.8	68	14.7±0.6	92	14.7±0.6[e]	92	12.7±0.0	79
7	Fat	1,9	1.8±0.1	95	1.6	84	3.4±0.6	179	2.9±0.8	153	3.7±0.1[g]	195	-	-
8	Ins Fibre	12,9	14.0[d]	109	14.2	110	17.1±0.6[e]	133	16.1±0.2[g]	125	-	-	-	-
9	Sol Fibre	3,6	3.7[d]	103	3.7	103	1.9±0.8[e]	53	6.6±1.7[g]	183	-	-	-	-
10	Protein	12,6	12.8±2.5	102	9.98	79	14.4±2.0	114	14.2±2.5	113	16.0±0.7[e]	127	17.1±1.9	136
11	LYS	3,3	3.4±0.2	103	3.44	104	5.1±0.3	155	3.7±0.2	112	3.6±0.3	109	4.7±0.7	142
12	HIS	2,1	1.93±0.1	92	2.02	96	2.7±0.1	129	2.1±0.1	100	2.1±0.1	100	2.4±0.3	114
13	ARG	4,0	3.9±0.1	98	4.13	103	6.0±0.5	150	4.1±0.2	103	4.2±0.2	105	5.5±0.7	138
14	ASP	5,7	5.4±0.0	95	5.64	99	9.3±0.5	163	6.5±0.4	114	6.3±0.4	111	9.3±2.1	163
15	GLU	22,7	22.4±0.8	99	21.49	95	14.7±0.8	65	20.7±1.1	91	22.9±3.0	101	16.9±2.8	74
16	PRO	13,3	13.1±0.0	98	11.36	85	8.0±0.7	60	12.8±0.8	96	12.8±1.2	96	9.6±2.1	72
17	THR	3,6	3.5±0.0	97	3.88	108	4.3±0.3	119	3.6±0.1	100	3.6±0.2	100	4.0±0.3	111
18	SER	5,0	5.1±0.1	102	5.57	111	5.3±0.2	106	5.0±0.3	100	5.5±0.2	110	5.2±0.2	104
19	GLY	7,1	7.0±0.1	99	7.63	107	10.1±0.2	142	7.6±0.3	107	7.0±0.6	99	9.1±0.9	128
20	ALA	6,0	6.2±0.1	103	6.30	105	8.0±0.2	133	6.6±0.3	110	6.4±0.6	107	7.5±0.7	125
21	TPCYS	2,1	2.7±0.5	129	2.49	119	2.2±0.2	105	2.3±0.2	110	2.1±0.4	100	2.2±0.2	105
22	VAL	6,0	6.2±0.1	103	6.61	110	6.7±0.1	112	6.1±0.4	102	5.4±0.3	90	6.2±0.4	103
23	MET	1,5	1.6±0.1	107	1.55	103	1.6±0.1	107	1.5±0.1	100	1.4±0.1	93	1.5±0.0	100
24	ILE	3,8	3.8±0.0	100	3.68	97	3.5±0.1	92	3.8±0.3	100	3.4±0.1	89	3.3±0.3	87
25	LEU	7,2	7.4±0.2	103	7.55	105	6.9±0.2	96	7.2±0.2	100	7.0±0.5	97	6.8±0.3	94
26	TYR	2,4	2.4±0.0	100	2.48	103	2.5±0.1	104	2.3±0.1	96	2.3±0.1	96	2.4±0.0	100
27	PHE	4,2	4.3±0.0	102	4.19	100	3.4±0.1	81	4.1±0.2	98	4.3±0.2	102	3.6±0.4	86

*Included in the normal category

[a] n=14, [b] n=12, [c] n=13, [d] n=1, [e] n=3, [f] n=5, [g] n=2, [h] n=4

Classification of genetic and environmental effects in the chemical data set by PCA (Material 1)

We may now look on the data in Table II after normalisation as 34 sample spectra with 27 "wavelengths" or variables. It is now possible to overview and work out the detailed relations in this small but complex data set to study the pattern (covariation) of these

"wavelengths" in a PCA. In the score plot in Fig. 4A adjacent samples have similar chemical spectra while in the loading plot the position of adjacent variables indicates a high positive correlation. Both graphs in Fig. 4A and 4B can be evaluated as transposed over each other. For a convenient orientation the position of the variables and samples respectively are marked with squares in the two figures. The *lys3a* genotype (Fig.4A) is thus positioned near to the amino acid cluster including lysine (LYS in Fig. 4B) indicating that *lys3a* is rich in this amino acid.

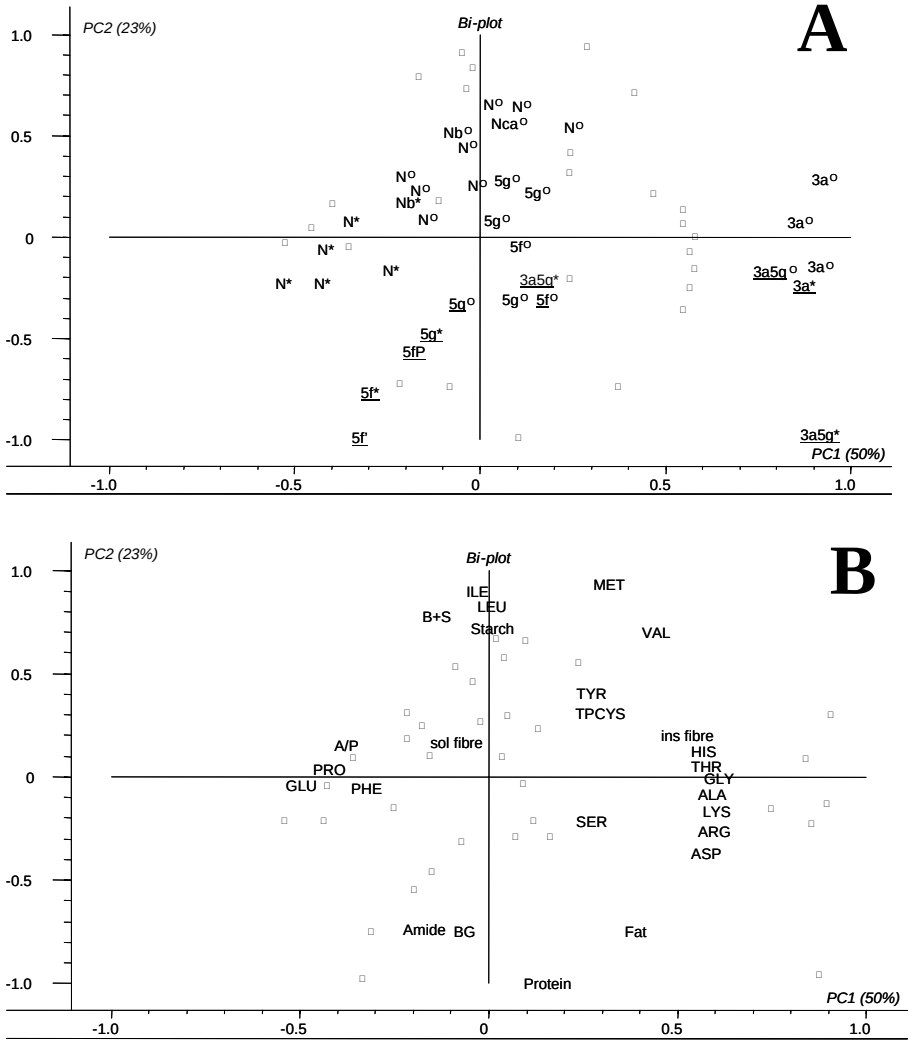


Figure 4. PCA biplot (PC1:2). Effects on classification of genotype and environment. Data from amino acid and chemical measurements on *lys3a*, *lys5g*, *lys3a5g* and normal (N) barley (Nb=bomi, Nca=Carlsberg II) (Material 1, Table II). * = indoors, O = outdoors, P = pots outdoors. **A.** Scores (samples) identified, □ = loadings see Figure 4B. Underlined samples are the original mutants **B.** Loadings (variables) identified, □ = scores see Figure 4A.

We now see a clear segregation in the patterns of the normal (N) and the mutant genotypes with *lys3a* far away to the right and *lys5g* more near to the normal barley's. The extreme BG rich mutant *lys5f* is positioned further away below to the right. The original mutants and their mother varieties Bomi (Nb) and Carlsberg II (Nca) are underlined. Their positions in relation to the other mutant samples not underlined represents mutant recombinants in crosses with other normal barley's. The distribution of the samples indicate that these genes have a greater impact on the chemical composition than their genetical backgrounds. There is a clear systematical environmental effect on the PCA clustering where, samples grown indoors (*) and outdoors in pots (^{P}) are pushed downwards to the PC1 axis compared to samples grown in the field (^{O}). This is partly due to an increased level of protein in the two first environments. It is now possible to define the characters of the principal components (PC´s) spanning the PCA loading plot (Fig. 4B). Thus PC1 is defined by two negatively correlated clusters with GLU, PRO, PHE and the A/P index to the left and a cluster with HIS, THR, GLY, ALA, LYS, ARG AND ASP to the right. PC2 is similarly defined by two negatively correlated clusters consisting of starch, BG+S, ILE, LEU in the upper part of the plot and Amide, BG and Protein in the lower part.

Comparing the chemical and the NIR data set by their PCA´s (Material 1)

We may now compare the chemical PCA fingerprint in Fig. 4A with that of the NIR spectra of the same samples in Fig. 5A and find that they essentially give the same separation between the genotypes. The NIR plot is, however, better differentiated
than the chemical plot pointing out that NIR spectroscopy represents the more complete physical-chemical fingerprint. There is a similar systematic disposition of the samples due to the environment in the NIR PCA plot (Fig. 5A) compared to the chemical PCA plot (Fig. 4A). As seen in the spectral window of 2270-2380nm, which is the spectral area which is most easy to interpret visually (Munck *et al*., 2001), the four different genotypes (Fig. 5B) are characterised by their spectral form while the environmental effect primarily consists of an offset from the baseline for samples grown indoors(*). Such samples tend to have a higher content of protein and BG content and lower starch than samples grown in the field (^{O}).

Classification by simple chemical ratios in two-dimensional plots

In defining the first barley high lysine gene *lys1* (Hiproly) earlier investigated by the senior author of this article (Munck *et al*., 1970; Munck, 1972), it was observed that ratios between specific amino acids were useful because they were largely independent of the environment. It was also found that a very high number of such ratios between many different chemical constituents were able to model *lys1* and other high lysine genes such as *lys3a* (Munck, 1972; Munck *et al*., 2001). The fundamental rationality of ratios e.g. A/P index for *lys3* genotypes (Table II) is thus a simple bivariate precursor to multivariate data

analysis. As shown in Fig. 5C and 5D simple ratios in an *abscissa-ordinate* plot could suffice for a successful genetic classification such as the amide to protein (A/P) index and ASP to GLU as *ordinate´s* and BG as *abscissa*. Thus the simple efficiency in classifying the genotypes by a few (six) anti microbial proteins presented above is verified on the basic biological organisational level of chemical bonds.

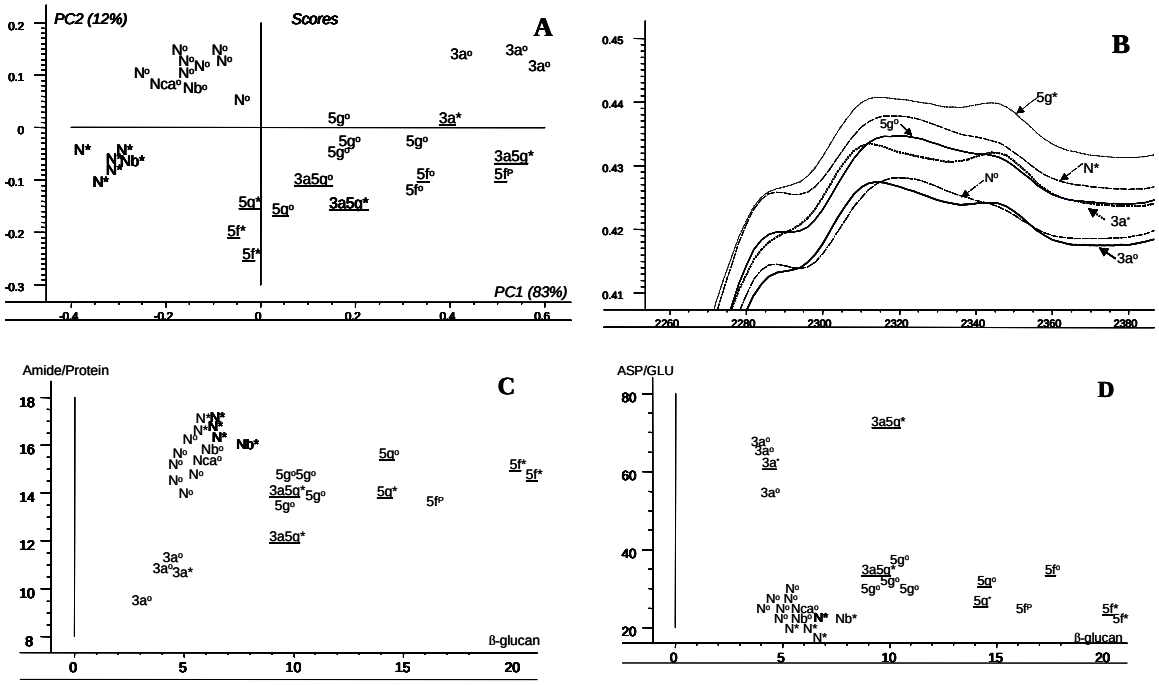


Figure 5A. PCA NIR (MSC) for 34 samples (Material 1), * = indoors, O = outdoors, P = pots outdoors. **B.** Enlarged spectral area (2260-2380nm) for *lys3a*, *lys5g* and average of spectra from normal barley's grown outdoors and indoors. **C.** Amide/Protein ratio (ordinate) plotted against (1→3,1→4)-ß-glucan. **D.** Asp/Glu ratio versus (1→3,1→4)-ß-glucan. Underlined samples are mutants.

Even though each gene has so many multivariate expression possibilities in the great web of epigenesis, gene detection as such need not to be complicated with our present data software and instrumentation. It is characteristic that the same endosperm mutants in barley and maize can be found directly or indirectly by a wide range of completely different screening strategies. They can be classified as high lysine (Doll, 1983; Munck, 1992), as carbohydrate (Greber *et al*., 2000) or as morphological mutants as for example shrivelled (Ullrich and Eslick, 1978) or opaque seeds (as in *lys3* seeds) (Munck, 1992) dependent on the focus of the scientist. All the nine spontaneous endosperm mutants selected for shrunken seeds analysed by Ullrich and Eslick (1978) displayed a higher lysine content of the protein compared to normal lines.

One of the best examples of how genetic covariance (pleiotropy and linkage) is expressed on the level of the phenotype is the fact that the Carlsberg mutant 1460 in Minerva (Munck,

1992; Munck *et al*., 2004) which was originally selected at Carlsberg by Aastrup to have a decreased BG content (by decreased acid extract viscosity) was found to be a *lys3* allele (*lys3m*) with a characteristic change in amino acid composition. However the allele *lys3c* has normal level of BG (Munck *et al*., 2004) putatively indicating different strongly linked genes for the amino acid change and decrease in BG.

The reason for the above redundancy in the exploration of endosperm mutants has not been discussed in literature as far as we now. It is, however, the major theme in this investigation. It is obvious that the answer resides in the tools of multivariate pattern recognition data analysis (chemometrics).

The effect of different gene backgrounds for the spectral expression of lys genes (Material 3)

We will now in more detail study the important issue of how the expressions of the *lys5g* and *lys3a* genes in the endosperm mediated by NIR spectroscopy are influenced by different genetic backgrounds. In Fig. 6A we are first visually inspecting local (2270-2380nm) mean NIR spectra from the original mutants *lys3a* (**3**) and *lys5g* (**5**) and their segregants (**x**) with normal barley. Their mean spectra are compared with those of the reference varieties Bomi and Carlsberg II and 39 other normal varieties. It is concluded that it seems that the spectral imprints of these two genes are rather well conserved. This impression is confirmed by a PCA evaluation of the corresponding global (400-2500nm) mean NIR spectra in Fig. 6B featuring three distinct clusters. We will in the discussion further give an example how the modification of the expression of a gene by a changed gene background can be followed as a whole physical-chemical fingerprint by NIR spectroscopy.

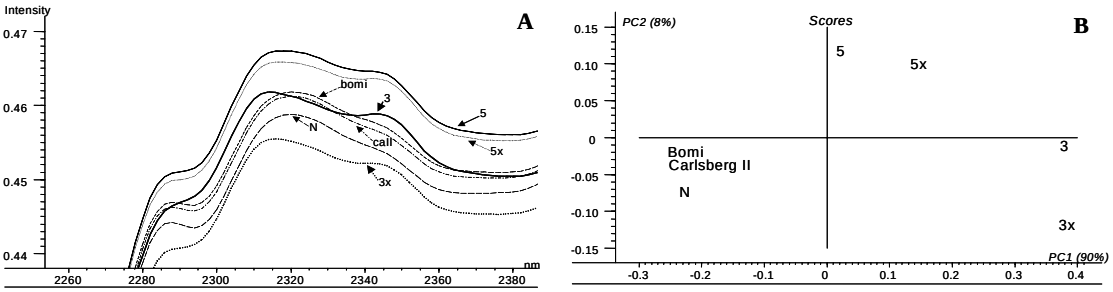


Figure 6. Enlarged area of the wavelength area 2260-2380nm. A and B comparing the effects of original mutants. **A.** Average NIR spectra (MSC) of Material 3. N = normal barley (n=39), bomi (n=2), caII (n=2), 3 = *lys3a* (n=2), 3x = *lys3a* seggregants (n=26), 5 = *lys5g* (n=3), 5x = *lys5g* seggregants (n=10). **B.** PCA on whole average spectra (400-2500nm) from Fig. 6A.

Breeding for an improved gene background to a major mutant gene by spectroscopy and chemometrics

Thus it was possible by cross breeding and selection for a plump seed with high yield to improve starch content in the *lys3a* genotype (Munck, 1992) almost to match some normal genotypes in seed quality and starch content, while keeping the favourable composition of essential amino acids. This is reflected in the NIR spectra as demonstrated in Fig. 7A and B (Material 4). First the area 2270-2380nm of a spectrum from the original *lys3a* mutant (starch 48.5%) is compared in Fig. 7A to spectra from the improved *lys3a* genotypes Piggy, Lysiba and Lysimax with starch contents of 49.9, 52.2 and 52.9% respectively and to those of the normal varieties Minerva (53.0%) and Triumph (59.2%). Low starch content together with high fibre (cellulose), and high fat content constitutes a negative pleiotropic covariate complex the improvement of which can be studied by NIR as a whole in Fig. 7A where wavelengths indicative for these components are marked above in the plot.

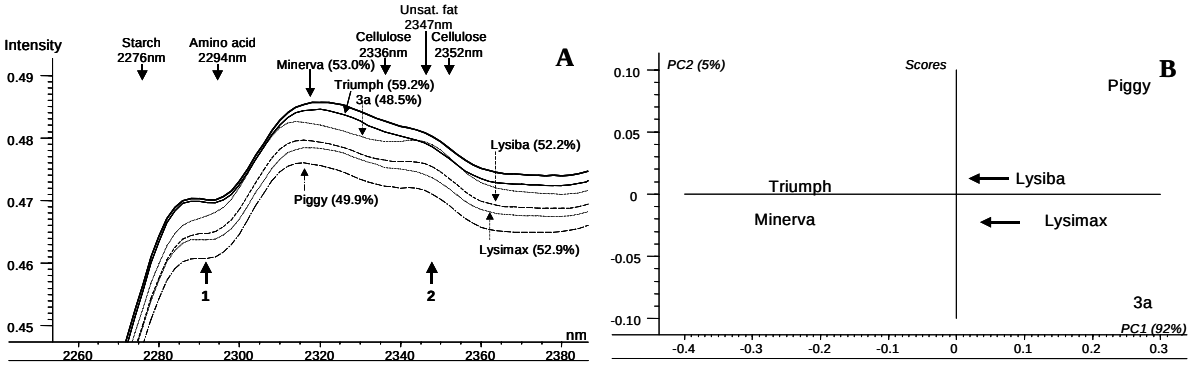


Figure 7. A. Mutants, normal and segregants with different starch content. *lys3a* and segregants 531 and lysimax, *lys3m*, Minerva and Triumph. Grown in field. **B.** PCA on whole average spectra (400-2500nm) from Fig. 7A (Material 4).

It is clearly seen that the improved *lys3a* genotypes (especially Lysimax) show normalisation of spectra approaching the normal genotypes Minerva and Triumph at arrow 1 (increased flattening of the plateau at 2285-2295nm) and arrow 2 (decrease in the bulb at 2340-2350nm). Visual inspection can only give a faint idea of how the covariate nature of complex spectra can be characterised by multivariate data analysis. In order to check the visual impression a PCA analysis is made on the corresponding whole spectra 400-2500nm in Fig. 7B. This "global" PCA evaluation clearly demonstrate that the semi commercial varieties Lysiba and Lysimax are approaching the normal controls in the direction of the arrows. The change in amino acid quality is intact as checked by the A/P index which is highly correlated to lysine (r=-0.94) (A/P Lysimax=10.8, Lysiba=11.2, *lys3a*=11.4, Minerva=15.8).

The dynamics of the influence of the segregating genome to modify the phenotypic expression of the pleiotropic complex of a mutant gene (Nielsen and Munck, 2000) can thus as shown here be evaluated by NIR spectroscopy and chemometrics. Positive segregants can be selected by "data breeding" (Munck *et al*., 2000) from a PCA of NIR data by comparing to the position of a high quality control varieties without elaborate chemical analyses. The effect can be checked after selection by chemical analysis.

It is evident that most endosperm mutants are more or less deleterious for the physiological balance of seed synthesis where every normal high yielding barley line has found a subtle balance point, which the mutant disturbs. As shown here it is, however, possible to adopt a new gene background to the mutant gene finding a new point of balance.

Short-circuiting the chemical, amino acid and NIR data sets by bi and multivariate correlations (Table III, Material I)

We will now further qualify the endosperm as a phenomic data interface carrying characteristic physical chemical imprints from the genotype as proved in Fig. 4B and Fig. 5A.

In order to further penetrate the finely structured web of interactions which we have detected in Material 1 we will as an example study how the three most important variables starch, BG and amide (as **y**) can be predicted by PLSR in Table III featuring three different data sets - seven chemical analyses, 17 amino acids and 1050 NIR wavelengths (as **X**). By Jack-knife validation it is possible to identify the variables which significantly stand out as the most important contributors in the correlations. The PLSR algorithm composes new parameters - Principal Components (PC's 1, 2, 3….) - by combining different amounts of the **X** variables. The PC´s are sequentially orthogonal (freely variable) to each other in the multivariate data space. Therefore the PC´s fulfills the requirements for a classic Multivariate Regression (MLR) analysis which completes the calculation of the PLSR correlation coefficient (r). A high number of significant PC´s indicate a complex model. In Table III we have compared the PLSR models of the three data sets with the simple diallel correlation's between the starch, BG and amide variables. There is a tendency that the numerical values of the correlation coefficients (r) predicting the three (**y**) variables improve with the complexity of the data sets (**X**) the diallel (r=0.44-0.76 **X**=1), the chemical (r=0.78-0.96, **X**=7), the amino acid (r=0.77-0.87, **X**=17) and the NIR (r=0.94-0.99, **X**=1050).

Table III. Diallel correlation and PLSR predictions of BG, starch and amide from the other gross chemical analyses, amino acid analyses and NIR spectra in Material 1, see Table II.

	X	y	r	RE	PC	n	Significant variables
Diallel correlation's	BG	Starch	-0.76				
	BG	Amide	0.47				
	Starch	Amide	-0.44				
PLSR predictions Jack-knife	Seven chemical analyses	BG	0.78	17.4	1	30	Starch, DM, Protein, Amide
		Starch	0.91	12.4	1	27	Protein, BG, Fat, Amide, DM
		Amide	0.96	1.4	5	29	-
	17 amino acids	BG	0.86	15.0	3	30	ser, tyr, val, met, ala, his, gly
		Starch	0.77	35.1	2	25	ile, met, leu, val, asp, lys, arg
		Amide	0.87	13.0	3	32	met, ile, thr, phe
PLSR predictions	NIR (1050 wavelengths)	BG	0.99	4.5	8	30	Significant variables not calculated
		Starch	0.95	18.6	2	27	
		Amide	0.94	13.3	5	32	
		Lys	0.94	8.6	3	34	
		Glu	0.91	9.7	3	34	

In studying the principle of interaction in biology transferred as covariance to the data set many (high quality) variables often give more reliable and simpler solutions (although empirical) than fewer variables. NIR spectroscopy gives obviously the most complete chemical fingerprint. Starch gives a less complex model in NIR prediction with less number of PC´s (2) compared to amide (5) and BG (8). NIR is also able to predict amino acids such as lysine and glutamic acid partly because their correlation with the amide bond which has several detection sites in the NIR spectra (Munck *et al*., 2001). In judging the number of significant Jack-knife variables in Table III it is seen that it is possible to obtain a high level of prediction without any significant variables as seen with the correlation amide (**y**) to the set of seven chemical variables (r=0.96). This is a sign of that all seven variables in **X** are equally tightly dependent on each other so that no one can be singled out as more important than the other. Still the PLSR prediction of amide from the set of seven chemical variables as a whole is highly significant (r=0.96).

We might conclude that the gene-specific, multivariate covariate classification patterns demonstrated in Fig. 4A and B (chemical data set) and Fig. 5A (NIR data set) and by the gene specific spectral signatures in Fig. 5B, 6A and 7A are just as real as their initial state - the DNA sequence of the corresponding gene. These gene specific patterns constitutes "a cage of of covariance" a black hole of information redundancy (Munck, 1992, p. 594). However, as shown in this paper it is possible to lock up "the cage" to find some of the underlying causes of genetic covariance from e.g. NIR spectra using chemical and genetic data for validation. The strength with the barley endosperm model is that such a study is possible in a near isogenic background which facilitates the study of the expression of specific genes as genetic covariance one by one. As of now this aspect on epigenesis is almost a virgin land in science.

It is quite clear that there is a firm scientific basis for a reversed genetic engineering strategy where the scientist is first exploring e.g. barley endosperms with a minimum of elaborate hypotheses by measuring with non invasive NIR spectroscopy in order to get indications of specific gene patterns through a PCA on spectra. Afterwards these indications can be followed up by genetic and biochemical analyses and finally by sequencing the DNA of the gene.

Perspectives on the multivariate approach to the phenome as a data interface

The strategy of chemometrics is to combine both internal data operations such as cross validation by external data e.g. by comparing several separate data sets from the same samples as demonstrated here for a verification of the physical-chemical nature of NIR data.

The chemometric methods are more robust for implementing incomplete data sets (such as the one in Table II) compared to the classic analysis of variance. It is not possible to study the important principle of covariance (fingerprints) on the level of individual samples (phenotypes) by classic statistics because it presumes free variation between the variables e.g. in Multiple Linear Regression (MLR) analysis (Martens and Næs, 1989, Munck, 2003). As seen in a historical perspective the pleiotropic component of the principle of genetic covariance has thus been neglected because of the limitations in classical statistics associated with genetics, where focus is on the gene linkage component by studying the variance of a few genes instead of pattern recognition of the multivariate physics and chemistry of a tissue from a biological individual such as the barley endosperm.
The examples of PLSR prediction given here are indicative because the limited material gives releatively high errors (RE´s). They give, however, because of the validation by separate chemical and physical (NIR) data sets a realistic insight in the basic multivariate, covariate character of the chemistry of a biological material which is useful in understanding how NIR spectroscopy works in genotype classification.

In the future new chemometric software should be developed that can handle the environmental more complex offset effects of spectra which can not be dealt with by MSC in order to focus on the genetic response available at the endosperm data interface. The present tedious data evaluation techniques for 2-DE methods for proteome analysis should be improved by automatic fingerprinting through new hardware and software paving the way for chemometric evaluation, facilitating a dialogue with data from the other levels of biological organisation. One approach to multivariate data analysis of 2-DE protein patterns has been done recently by using a massive data reduction before using chemometric methods (Schultz *et al*., 2004). This greatly reduces the calculation time.

CONCLUSION

The classic developmental geneticist C.H. Waddington defined already in 1969 the phenotype and epigenetic data spaces as interfaces for gene expression. He expressed the need of a classification data software which could handle complex data information on the basis of the biological individual which he was not aware of excisted at that time (Waddington, 1969).

Such a study is today possible combining spectroscopic instruments with chemometrics as with the PCA and PLSR data programs as demonstrated here. The intricate web of covariate connections revealed by chemometric analysis of spectral and chemical data is a necessary holistic complimentary aspect to the reductionistic classic biochemical and biotechnological methods in understanding epigenesis. It can be expressed as genetic covariance covering the combined effects of gene pleiotropy and linkage reflecting cellular and chromosomal structural and chemical relationships.

The proteome analysis of the dry seeds (mainly endosperm tissue) gives an imprint of the preceding dynamics of endosperm synthesis. In this investigation focusing on the chemical background to the gene classification of ripe seeds by NIR, we are not likely to be able to identify crucial synthesising enzymes like ADP-glucose (which is reduced in *lys5f* - mutant 13, Greber *et al*., 2000). Instead we find a range of storage proteins (Shewry, 1993; Shewry and Morell, 2001) and proteins (Leah *et al*., 1987; Mundy *et al*., 1986) of importance for microbial resistance which both are important for the endosperm as a food storage organ for the embryo but secondary to the carbohydrate synthesis. However, as we have found here, the final levels of these secondary proteins in the seed are effective in giving an indirect covariate imprint emerging from the primary mutant lesion in the DNA. This imprint is also available at the metabolome level as changes in starch and BG and at the total phenomic data interface level as read by NIR spectroscopy and chemometrics. Thus NIR micro-spectroscopy and PCA should be able to be used to indicate gene transfer in cell cultures as proposed by Munck *et al*. (2001).

The advantage of using these techniques to study gene expression in near isogenic backgrounds is obvious but as indicated here NIR spectroscopy is also effective in handeling a multivariate polygenic trait as a whole by selection through a PCA interface as for the improvement of the *lys3a* pleiotropic complex previously discussed.
The concept of the barley endosperm as a data interface representing the phenome for the expression of specific genes should be able to be transferred to other tissues and species.

Discovering genes as multivariate signatures in the phenome (the phenotypic data interface) is now used as a technology e.g. to differentiate fungal species (Frisvad *et al*., 1998) and to

diagnose cancer genes (Sukuta and Bruch, 1999). Here these methods are used to investigate the chemical background of the spectral detection of a new class of carbohydrate mutants (Munck *et al*., 2004) in barley empirically. This is done without addressing in depth a much-needed supportive extended theory on epigenesis
e.g. to clarify the pleiotropic effects on the phenome of typical regulating and structural genes in an isogenic background.

The current gene and protein sequencing analyses have been supplemented with new computerized instrumental techniques (Fiehn, 2002; Thiellement *et al*., 1999; Templin *et al*., 2002), which are harvesting data from the epigenetic data space (Waddington, 1969) e.g. the proteome (Vienne and Zivy, 1999) and metabolome by destructive analyses such as micro array technology, 2-DE, HPLC-MS, GC-MS, LC-MALDI-TOF-MS (Gottlieb *et al*., 2002) as well as with immunological methods. Such large collections of automatised instrumental data from destructive methods as well as from the intact phenotype (the phenome) by potential non-invasive observation through spectroscopy are open for exploratory discovery. They contain more information than the present scientists and science as a whole are able to hypothesise. It is therefore rational to introduce a data mining strategy to these very large data libraries. Such databanks can be explored by the introducing already excisting pattern recognition software to find surprising gene functions like BG compensated starch mutants which could be further investigated with traditional biotechnological methods to find the underlying gene sequence for the altered gene(s).

One of the foremost post-genomic challenges is the application of spectroscopy to obtain a digitised databank representing collections of phenomes, approaching the data quality of the genomic data libraries. A phenomic data bank can only be evaluated by modelling the genetic and environmental influence on the covariate physical-chemical interactions in the cell or tissue. Here each biological individual is the target as the unit of primary interest for the expression of genes and life. For this purpose pattern recognition data analyses such as those in chemometrics are indispensible.

ACKNOWLEDGEMENTS

The work of Torben Desler (1987) is greatly acknowledged. The experiments in his M.Sc. thesis from 1987 were performed at Carlsberg Research Laboratory, Valby, Denmark with L. Munck as supervisor. The supervisor at the Department of General Microbiology, The University of Copenhagen, Denmark, was associate professor Ole Matson. The investigation was supported by Dr. John Mundy, Carlsberg Research Laboratory, now at The University of Copenhagen, Denmark, who kindly supplied the antibodies. Ljiljana Nesic and Anne Blicher is greatly thanked for performing the biochemical analyses at

DTU. Lisbeth Hansen, Kirsten Wilms and Anna-Marie Hansen are greatly appreciated for their assistance in further chemical and spectroscopic measurements at KVL. This investigation had been impossible without the cooperation with enthusiastic colleagues at the plant breeding institute at Risø and Carlsberg, Denmark and at Svaløf, Sweden. 1963-91 in building up the the mutant barley collection used here.

REFERENCES

Barber, D., Sanchez-Monge, R., Mendez, E., Lazaro, A., Garcia-Olmedo, F., Salcedo, G., 1986. New α-amylase and trypsin inhibitors among CM-proteins of barley (*Hordeum vulgare*) *Biochim. Biophys. Acta* **869**, 115-118.

Barkholt, V., Jensen, A.L., 1989. Amino acid analysis: Determination of cysteine plus half-cystine in proteins after hydrochloric acid hydrolysis with a disulfide compound as additive. *Anal. Biochem.* **177**, 318-322.

Blom Sørensen, Müller, M., Skeritt, J., Simpson, D., 1996. Hordeine promoter methylation and transcriptional activity in wild-type and mutant endosperm. *Mol. Gen. Genet.* **250**, 750-760.

Desler, T., 1987. M.Sc. Thesis, Immunokemisk karakterisering af syntesen af specifikke proteiner under ontogenesen af kerner fra højlysin mutanter af byg (*in Danish*), University of Cophenhagen, Department of Plant Physiology and Carlsberg Research Laboratory, 91.

Doll, H., 1983. Barley Seed Proteins and possibilities for their improvement, in: Gottchalk, W., Muller, H.P. (Eds), Seed Proteins, Martinus Nijhoff/Dr. W.Junk Publishers, The Hague, 205-223.

Fiehn, O., 2002. Metabolomics - the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155-171.

Frisvad, J.C., Filtenborg, O., Trane, U., 1998. The use of secondary metabolites in fungal taxonomy, in: Frisvad, J.C., Bridge, P.D., Arora, D.K. (eds.), Chemical taxonomy. Marcel Dekker, New York, 289-319.

Gottlieb, D.M., Schultz, J., Petersen, M., Nesic, L., Jacobsen, S., Søndergaard, I., 2002.

Determination of wheat quality by mass spectrometry and multivariate analysis. *Rapid Comm. in Mass Spectrometry* **16**, 2034-2039.

Greber, B., Waite, D., Fahy, B., Hylton, C., Parker, M., Laurie, D., Smith, A.M., Denyer, K., 2000. Use of barley mutants to understand starch synthesis. *Proceedings 8th International Barley Genetics Symposium*. Adelaide, October 2000. Volume 1, 196-198.

Hejgard, J., Bjørn, S.E., 1985. Four major basic proteins of barley. Purification and partial characterisation. *Physiol. Plant.* **64**, 301-307.

Jacobsen, S., Nesic, L., Petersen, M., Søndergaard, I., 2001. Classification of wheat varieties: Use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and an artificial neural network. *Electrophoresis* **22**, 1242.

Johansson, A., Rasmussen, S.K., Harthill, J.E., Welinder, K.G., 1992. cDNA, amino acid and carbohydrate sequence of barley seed-specific peroxidase BP1. *Plant Mol. Biol.* **18**, 1151-1161.

Jonassen, I., 1980. Characteristics of Hiproly Barley I. Isolation and characterisation of two water soluble high lysine proteins. *Carlsb. Res. Comm.* **45**, 47-58.

Jonassen, I., 1982. Segregation of high and low chymomtrypsin inhibitor CI-2 content in crosses between hiproly and normal barley genotypes. *Carlsb. Res. Comm.* **47**, 305-312.

Laemli, U.K., 1970. Cleavage of structure proteins during the assembly of the head of bacteriophage T4. *Nature* **277**, 680.

Leah, R., Mikkelsen, J.D., Mundy, J., Svendsen, I., 1987. Identification of a 28.000 Dalton endochitinase in barley endosperm. *Carlsberg Reseach Communication* **52**, 31-37.

Martens, H. and Næs, T.1989. Multivariate Calibration. John Wiley, Chichester.

Munck, L., 1972. Improvement of nutritional value in cereals. *Hereditas* **72**, 1-128.
Munck, L., 1992. The case of high lysine barley breeeding, in: Shewry, P.R. (ed.), Barley: Genetics, Molecular Biology and Biotechnology, CAB International
Wallingford, U.K, 573- 602.
Munck, L., 2003. Detecting diversity - a new holistic, exploratory approach bridging the geneotype and phenotype, in: Diversity in Barley *(Hordeum vulgare*), von Bothmer, R., van Hintum, T.J.L., Knupffer, H. and Sato, K. (eds.). Elsevier Science B.V., Amsterdam, The Netherlands, Chapter 11, 227-245.
Munck, L., Karlsson, K.E., Hagberg, A., Eggum, B.O., 1970. Gene for improved nutritional value in barley seed protein. *Science* **168**, 985-987.
Munck, L., Nielsen, J., Møller, B. (2000). From Plant Breeding to Data Breeding - How New Multivariate Screening Methods Could Bridge the Information Gap Between the Genotype and the Phenotype. *Proceedings 8th International Barley Genetics Symposium*. Adelaide, October 2000. Volume 1, 179-182.
Munck, L., Nielsen, J. P., Møller, B., Jacobsen, S., Søndergaard, I., Engelsen, S.B., Nørgaard, L., Bro, R., 2001. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. *Anal. Chim. Acta* **446**, 171-186. John Wiley, Chichester.
Munck, L., Møller, B., Jacobsen, S., Søndergaard, I., 2004 Spectral multivariate indicators for mutant endosperm genes evaluated by chemometrics reveal a new mechanism for substituting starch with (1→3,1→4)-β-glucan in barley. *J. Cereal Science* submitted.
Mundy, J., Hejgaard, J., Hansen, A., Hallgren, L., Jørgensen, K.G., Munck, L., 1986. Differential synthesis in vitro of barley aleurone and starchy endosperm proteins. *Plant Physiology* 81, 630-636.
Nielsen, J.P and Munck, L., Exploring Malting Barley Data Using Chemometrics. *Proceedings 8th International Barley Genetics Symposium.* Adelaide, October 2000. Volume 2, 258-261.
Radova, A., Sebela, M., Galuszka, P., Frebort, I., Jacobsen, S., Faulhammer, H.G., Pec, P. 2001. Barley polyamine oxidase: Characterisation and analysis of the cofactor and the n-terminal amino acid sequence. *Phytochem. Anal*. **12**, 166-173.
Schultz, J., Gottlieb, D.M., Petersen, M., Nesic, L., Jacobsen, S., Søndergaard, I., 2004. Explorative data analysis of two-dimensional electrophoresis gels. *Electrophoresis*, **25**, 502-511.
Shewry, P.R., Lafiandra, D., Salcedo, G., Aragoncillo, C., Garcia-Olmedo, F., Lew, E.J.-L., Dietler, M.D., Kasarda, D.D., 1984, N-terminal amino acid sequences of chloroform-methanol-soluble proteins and albumins from the endosperm of wheat, barley and related species. *FEBS Letter* **175**, 359-363.
Shewry, P.R., 1993. Barley seed proteins, in: MacGregor, A.W., Bhatty, R.S. (eds.) Barley: Chemistry, technology, American Association of Cereal Chemists, St. Paul, USA, 131-197.
Shewry, P.R., Morell, M., 2001. Manipulating cereal endosperm structures, development and composition to improve end-use properties. Advances in Botanical research incorporating advances in Plant Patology, Academic Press, London, U.K. 34, 165-236.
Sukuta, S., Bruch, R., 1999. Factor analysis of cancer Fourier transform infrared evanescent wave fiberoptical (FTIR-FEW) spectra. *Lasers in surgery and medicine* **24**, 382-388.
Templin, F.T., Stoll, D., Schrenk, M., Traub, P.C., Vöringer, C.F., Joos, T.O., 2002. Protein microarray technology. *Trends in Biotechnology* **20**, 160-166.
Thiellement, H., Bahrman, N., Damerval, C., Plominon, C., Rossignol, M., Santoni, M., de Vienne, D., Zivy, M., 1999. Proteomics and two-dimensional electrophoresis – proteomics for genetic and physiological studies in plants. *Electrophoresis* **20**, 2013.
Ullrich, S.E., Eslick, R.F., 1978. Lysine and protein characterization of spontanouse shrunken endosperm mutants of barley. *Crop Science* **18**, 809-812.
Vienne, D., Zivy, M., 1999. Proteomics for genetic and physiological studies in plants. *Electrophoresis* **20**, 2013-2026.

Waddington, C.H., 1969. The theory of evolution today, in: Koestler, A., Smythies, J.R. (eds.) Beyond Reductionism. The Alpbach Symposium, Radius Book/Hutchinsson,
London, U.K., 357-395.
Watkins, M.S., Hammock, B.D., Newman, J.W., German, J.B., 2001. Individual metabolism should guide agriculture towards foods for improved health and nutrition.
Am. J. of Nutr. **74**, 283-286.
Williamson, M.S., Forde, J., Kreis, M., 1988. Molecular cloning of two isoinhibitor forms of chymotrypsin inhibitor 1 (CI-1) from barley endosperm and their expression in normal and mutant barleys. *Plant Mol. Biol*. **10**, 521.