

RESEARCH ARTICLE

Changes in cod muscle proteins during frozen storage revealed by proteome analysis and multivariate data analysis

Inger V. H. Kjærsgård¹, Mette R. Nørrelykke² and Flemming Jessen¹

¹ Department of Seafood Research, Danish Institute for Fisheries Research, Søltøfts Plads, Lyngby, Denmark

² Department of Perception and Functionality, Technological Institute, Kolding, Denmark

Multivariate data analysis has been combined with proteomics to enhance the recovery of information from 2-DE of cod muscle proteins during different storage conditions. Proteins were extracted according to 11 different storage conditions and samples were resolved by 2-DE. Data generated by 2-DE was subjected to principal component analysis (PCA) and discriminant partial least squares regression (DPLSR). Applying PCA to 2-DE data revealed the samples to form groups according to frozen storage time, whereas differences due to different storage temperatures or chilled storage in modified atmosphere packing did not lead to distinct changes in protein pattern. Applying DPLSR to the 2-DE data enabled the selection of protein spots critical for differentiation between 3 and 6 months frozen storage with 12 months frozen storage. Some of these protein spots have been identified by MS/MS, revealing myosin light chain 1, 2 and 3, triose-phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, aldolase A and two α -actin fragments, and a nuclease diphosphate kinase B fragment to change in concentration, during frozen storage. Application of proteomics, multivariate data analysis and MS/MS to analyse protein changes in cod muscle proteins during storage has revealed new knowledge on the issue and enables a better understanding of biochemical processes occurring.

Received: April 20, 2005

Revised: August 1, 2005

Accepted: August 5, 2005

Keywords:

Chemometrics / Cod sequences / Fish muscle / Frozen storage / Principal component analysis

1 Introduction

Proteomics offers a comprehensive approach to study biochemical systems by expanding the investigation from single proteins to a wide range of proteins at once [1]. Often prote-

omics encompasses analysis of a considerable number of gels each representing a large number of protein spots. Univariate data analysis can reveal changes for several single protein spots but multivariate data analysis of proteomic data enables identification of protein spots with interrelated changes. The changes in spot pattern can be correlated to conditions given by the experimental setup and thereby optimise the earnings of biological information given by proteome analysis ([2], and references within).

Fish meat deteriorates much faster than other muscle-based foods and storage conditions are critical. Therefore, onboard freezing immediately after catch can be an optimal choice to ensure high-quality products to the consumer. However, high frozen storage temperature (-20 to -10°C), fluctuating temperatures or prolonged frozen storage will lead to a quality loss. Due to a substantial content of poly-

Correspondence: Dr. Inger V. H. Kjærsgård, Department of Seafood Research, Danish Institute for Fisheries Research, Søltøfts Plads, Building 221, DK-2800 Lyngby, Denmark
E-mail: ivk@dfu.min.dk
Fax: +45-45-88-4774

Abbreviations: DPLSR, discriminant PLSR; MAP, modified atmosphere packed; MLC, myosin light chain; PCA, principal component analysis; PLSR, partial least squares regression; PC, principal component; r , correlation coefficient; RMSEP, root mean squared error of prediction

unsaturated fatty acids fish lipids are highly susceptible to oxidation [3] resulting in increased rancidity. Therefore, lipid oxidation contributes significantly to the deterioration in fatty fish, whereas in lean fish, as for example cod, parameters as protein aggregation/changes and increased water loss are mostly responsible for quality decline, which is often evaluated as increase in toughness and loss of juiciness by sensory evaluation [4, 5].

The stability and solubility of proteins from frozen fish have been studied for decades ([4] and references within) and a vast number of different fish species has been evaluated. However, it is not a simple task since protein stability and solubility is affected by a number of different factors such as species, freezing temperature, time span of frozen storage and physiological condition of the fish before freezing. Also, it has to be taken into account whether the frozen food in question comes from whole fish, fish fillets or fish mince [5]. Recent studies on extractability and thermal stability of proteins from different frozen fish species such as hake [6] and cod [7] have shown changes in the solubility of myosin heavy chain and α -actin to be responsible for some of the protein solubility alterations occurring during frozen storage.

The optimal method for identification of proteins in proteomics is mass spectroscopy since it only requires low amounts of protein and can be conducted directly on proteins excised from the 2-D polyacrylamide gels [8]. However, identification of fish proteins in proteomics is challenging due to the relatively low number of fish protein sequences and ESTs present in the public databases [9]. Therefore, it is essential to obtain protein sequences by either Edman degradation or MS/MS sequencing which enables identification by matching with homologue proteins from different species. This approach previously led to successful identification of hake proteins [10].

The strength of multivariate data analysis is that the patterns of variations between the samples are revealed and application of this approach on proteomic data is progressively becoming more common [11–14]. In multivariate data analysis, principal component analysis (PCA) is used to represent a matrix of observed data by a matrix of systematic variation and a matrix of noise [15]. PCA gives an overview of relationships or 'groupings' within multivariate data describing the observations as the sum of data structure and noise. Data represented in the p -dimensional space (p represents the number of variables) are reduced into a few principal components (PC1, PC2, PC3 *etc.*), where PC1 describes the biggest variation in the data set, PC2 the second biggest and so forth. The results of PCA analysis are shown graphically in a 'score plot' and a 'loadings plot'. The score plot shows if there are any groupings among the different samples, and the loadings plot shows which variables describe the variation between the samples. PCA of 2-DE data is typically conducted on a data matrix (**X**), with the spot intensities given in columns, related to the sample in a given row. PCA is based on the data in the **X** matrix and the calculations are done without relating to any knowledge on the sample con-

ditions. In partial least squares regression (PLSR), two data matrices **X** and **Y** are related by a linear multivariate model where the structure of matrix **X** and matrix **Y** are modelled [16]. The data are represented as described for PCA; however, the data for both the **X** and **Y** matrices are presented in the loadings plot enabling the prediction of how the different y variables are modelled by the x variables.

The aim of this study was to combine 2-DE with multivariate data analysis to characterise changes in cod muscle proteome during different storage conditions, and to use multivariate data analysis to explore how the pattern in protein spot composition correlates with the time course of the experiment. To achieve the full benefit from the study some of the cod protein spots of interest were identified with MS/MS sequencing.

2 Materials and methods

2.1 Materials

Unstained low molecular weight protein standard, Mark 12 (LMW) standard was purchased from Invitrogen, Life Technologies (Carlsbad, CA, USA). Pharmalyte 4–6.8, Pharmalyte 5–8, 87% glycerol and SDS were from Amersham Biosciences (Uppsala, Sweden). All other chemicals were reagent/analytical grade and were purchased from either Merck (Darmstadt, Germany) or Sigma (St. Louis, MO, USA).

2.2 Storage trials and experimental design

The cod (*Gadus morhua*) samples were selected from a large factorial design with two frozen storage temperatures (-20 and -30°C), four storage periods (3, 6, 9, 12 months) and five chill storage periods (0, 3, 7, 14, 21 days after frozen storage). All batches were initially stored at -30°C for 3 months, which was the time required for transportation of the cod from the catching ground in a freeze trawler to the institute. A detailed description of the experimental setup is given elsewhere [17] and an overview of the storage conditions analysed by 2-DE is given (Fig. 1). For all storage conditions protein extraction was done in duplicate on two different fillets.

2.3 Protein extraction and 2-DE

The detailed description of protein extraction, IEF and size separation on SDS-PAGE is described elsewhere [18]. In brief, proteins were extracted in 2% SDS and 100 mM DTT and separated according to charge with 1-D Immobiline DryStrips (Amersham Biosciences), 18 cm long and with a linear pH gradient of 3–10. For each sample 60 μg was applied to Immobiline DryStrips by cup loading. After separation of the proteins in the first dimension, for a total of 60 000 Vh, the proteins were separated according to size in

the second dimension on 12% w/v SDS-PAGE [19] as described [18]. The 2-D gels were silver-stained with AgNO_3 according to Hochstrasser *et al.* [20] with minor modifications as described [18].

All samples were run in three batches of eight to ten gels. To avoid confounding due to experimental setup the samples were shuffled around so duplicate samples were run in different batches.

The raw spot volumes are publicly available at www.difres.dk/product_uk/Public_data. The gel images can be available upon request.

2.4 Preparative 2-DE for MS/MS peptide sequencing

For MS/MS analysis 2-DE was conducted by applying 450 μg protein by overnight rehydration and the gels were stained with colloidal blue (Invitrogen). The other conditions for 1-DE and 2-DE were as described in Section 2.3.

2.5 Data analysis of 2-DE

Silver-stained 2-D gels were digitised using a 420 OE (PDI, BioRad, Hercules, CA, USA), flatbed scanner with a pixel size of 84.7×84.7 pixels. PDQuest 7.1.0 Discovery Series (BioRad). The 2-D analysis image software was used to locate protein spots and match spots among gels. Spot matching was accomplished by manually defining anchor spots followed by automated matching of the remaining spots.

The quantity of each spot was calculated during spot detection and Gaussian fitting measured as $\text{OD}/(0.1 \text{ cm}^2)$. To ensure a uniform quantification, a normalisation of the spots in each gel was conducted as follows: The $\text{OD}/(0.1 \text{ cm}^2)$ of each spot in a gel was divided by the $\text{OD}/(0.1 \text{ cm}^2)$ of all matched and unsaturated spots (peak intensity is within the linear range of the scanner) in the gel and a value as ppm was given. In the 2-D gels five spots were saturated (peak intensity outside the linear range of the scanner). In these cases the spot quantity was calculated by summing the intensities of the pixels inside the defined boundary and their quantities are comparable with regular, Gaussian-fitted spots.

2.6 Multivariate data analysis

Multivariate data analysis was conducted on 2-DE data generated by PDQuest using 'The Unscrambler Programme version 7.6' (CAMO ASA, Oslo, Norway). In the multivariate data analysis the gels represent the samples, spot numbers the variables and the intensity of a certain spot in a given gel the x -value. To eliminate any risk of misinterpretation the five saturated spots (1710, 3412, 3506, 3508 and 3509) in the data set were weighted down by a factor 0.001 in all multivariate calculations.

2.7 Data normalisation and mean centring

The 2-D data were mean centred and standardised before multivariate modelling. The variables in multivariate data analysis are generally mean centred thereby giving each variable a mean of zero. This facilitates the evaluation of the between sample variation of the variables, but not the absolute values of the variables.

Likewise the variables are often standardised by weighting by multiplication of each data point in the X matrix by the $(\text{SD} + B)^{-1}$. The most usual scaling factor is the inverse of the SD, *e.g.* B equals zero. However, in 2-DE data there is a risk of over-amplifying noise since small values are scaled up and large values are scaled down by multiplication by the inverse of the SD [21]. Since 2-D data are of a semi-quantitative nature this may introduce artefacts since especially weak spots are afflicted with some uncertainty [18, 22]. Conversely they may also be the most interesting protein spots in a data set, if they consistently appear or disappear according to the given experimental conditions. Finding a proper value for the constant B in the standardisation formula may help circumvent this problem and thereby attract attention to the protein spots giving most information in the given data set. Extracting the variable statistics, and plotting a scatter plot of the mean *versus* the SD can be helpful for finding a good estimate of this value. Different values for B were tested and a B value of 2500 was found to give the most consistent results where spots identified to change during storage by univariate data analysis were also highlighted to be important for the models.

2.8 PCA

A PCA analysis was conducted on all centred and normalised data, with B set to 2500 using all samples. The optimal number of PCs was found to be 2. The score plot of PC1 *versus* PC2 was examined for separation of clusters related to frozen storage time, freezing temperature or storage time in modified atmosphere packing (MAP) at 2°C .

Likewise the loading plot of PC1 and PC2 was examined to reveal which spots define the differences between the different groups.

2.9 PLSR and discriminant PLSR (DPLSR)

In order to clarify if any correlation of frozen storage time, freezing temperature or storage time at 2°C in MAP was hidden in the data set, PLSR modelling was applied. In DPLSR the Y matrix (*e.g.* the design matrix with quality indicators) is defined as a binary vector with the values 0 or 1 as design variables of the categories, and the X matrix as a set of other variables, in this case as the spot intensities [21]. In the initial phase of PLSR modelling, it was tested if the changes in protein pattern correlated the real storage conditions of either frozen storage time (days or months), freezing temperature ($^\circ\text{C}$) or storage time at 2°C in MAP (days); how-

ever, this did not lead to any conclusive PLSR models. This is probably due to the fact that alteration in protein pattern does not correlate in a linear manner to either month of storage, temperature in degree Celsius or days of storage. Therefore, different approaches were applied, for instance the natural logarithm of the values, the inverse values or inverse natural logarithm in the Y matrix, but none of these approaches led to any valuable PLSR models. However, since the PCA revealed the presence of storage-dependent groups, the data were analysed with DPLSR in order to optimise the separation between the different storage conditions and thereby determine which spots are significant for each storage condition.

A DPLSR model was calculated including all samples, e.g. values of 1 or 0 for frozen time (3, 6 or 9 months), freezing temperature (-30 or -20°C) or storage time at 2°C in MAP (0, 7 and 14 days). For the purpose of cross validation, the samples were divided into two different separate sets. Since the analogue samples were made up individually (different cod filets/different preparations/different 2-DE), they can be regarded as distinct observations and be used to validate the models by cross model validation [21]. This cross validation was used in all multivariate analyses. The design matrix contains qualitative indicators for the main effects, e.g. storage temperature (-20 or -30°C), frozen storage time (3, 6, 12 months) and chilled storage time (0, 7, 14 days).

First the three quality indicators were tested individually and DPLSR was conducted with respect to storage temperature, days on ice in MAP or frozen storage time. In DPLSR the correlation between the predicted and the measured gives an idea about which samples are well predicted and which are not. The correlation coefficient (r) is a unit-free measure of the goodness-of-fit between the two variables x and y within a certain set of N samples. If $r = 1$ there is a perfect fit with positive slope, if $r = -1$ there is a perfect fit with negative slope and if $r = 0$ there is no fit at all [21]. The root mean square error of prediction (RMSEP) is a measure of the average difference between predicted and measured response values at the prediction stage. Therefore, a low RMSEP is an indicator of a good model. The validated variance for the x or y variables express how well the model will perform with similar new data [15].

In the final DPLSR calculations based on 288 protein spots all design variables were included; however 'storage time at 2°C MAP' and 'frozen temperature' were weighted down by a factor of 0.001. Therefore, they do not influence the model; however, they can still be visualised in the correlation plot.

2.10 Peptide sequencing by nanoLC-ESI-MS/MS

Selected proteins were excised from 2-D gels stained with either colloidal blue or silver stain without glutardialdehyde. The proteins were in-gel digested with trypsin (Promega, Madison, WI, USA, modified sequencing grade) [23], without previous reduction and alkylation of the proteins. The

digestion mixtures were separated using a CapLC (Waters, Milford, MA, USA) and on-line sequenced by ESI-MS/MS positive mode using a Q-TOF2 mass spectrometer (Waters, Micromass, Manchester, UK). Peptides were trapped on a $75\ \mu\text{m} \times 10\ \text{mm}$ precolumn and separated on a $75\ \mu\text{m} \times 150\ \text{mm}$ analytical column, both packed with ReproSil-Pur C18-AQ $3\ \mu\text{m}$ (Coricon, Knivsta, Sweden) using an in-house column packing system (Proxeon, Odense, Denmark). Mobile phase A was 2% ACN, 0.1% formic acid and mobile phase B was 80% ACN, 0.1% formic acid. Peptides were loaded onto the trapping column with 0.1% formic acid with a flow rate of $5\ \mu\text{L}/\text{min}$ in 10 min, followed by a gradient elution through the analytical column. The gradient was 7% B in 0–10 min, 7–20% B in 10–12 min, 20–60% B in 12–40 min, 60–100% B in 40–50 min and the flow rate was $250\ \text{nL}/\text{min}$. The nanoLC column was connected to a silica PicoTip emitter, $10\ \mu\text{m}$ tip (New Objective, Woburn, MA, USA) in the mass spectrometer interface. The capillary voltage on the PicoTip was 2.3 kV, and the cone voltage was 45 V.

Only parent ions with two or three charges were chosen for MS/MS and the data were acquired in a data-dependent mode. The resulting MS/MS data were managed by Mass Lynx 4.0 software (Micromass).

2.11 MS data processing

The MS/MS dataset were processed using the ProteinLynx Global Server 2.1 software (Micromass) and used to search the NCBI nr protein database. Most peptides were identified in this way. Additional peptides were manually sequenced with Pepseq option in the MassLynx software. All amino acid sequences were submitted to the MS BLAST search algorithm [24] (<http://dove.embl-heidelberg.de/Blast2/msblast.html>) for obtaining a total score for identity. The default settings were used, but the alignments were done against the database sp_nrrdb. This database covers all nonredundant protein databases from Swiss-Prot, SwissProtNew, SptremblNew and Sptrembl.

3 Results

3.1 2-DE

Differences in the muscle proteome of cod (*G. morhua*) has been studied under 11 different storage conditions, thereby evaluating the storage effects due to frozen storage temperatures, frozen storage time and storage time after thawing at 2°C in MAP (Fig. 1).

The proteins in cod muscle extracts were separated by 2-DE, and silver staining revealed 1500–2000 spots as shown in a representative gel (Fig. 2). At protein sizes above the dashed square the resolution was not optimal and therefore the data analysis was concentrated on the partial proteome with the optimal resolution, corresponding to proteins with pI values

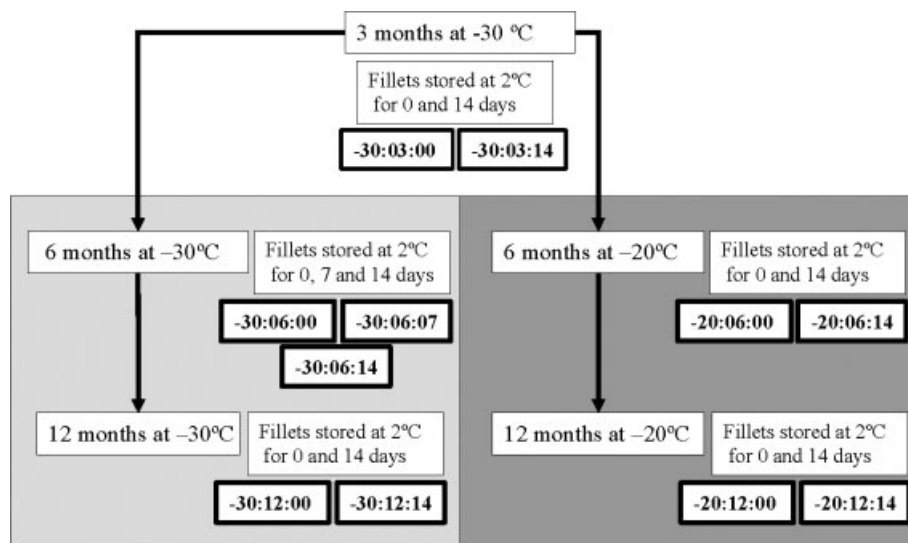


Figure 1. Flow chart of the different storage conditions for cod analysed by 2-DE reported here. Fish fillets were initially frozen at -30°C for 3 months. Hereafter the fish were kept at either -30 or -20°C for 3 or 9 more months. Numbers in the boxes with a thick frame show the sample names. Numbers correspond to the following values: frozen storage temperature: months of frozen storage: storage time on ice in MAP. For example -20:12:14 denote that the fillets, after the initial 3 months at -30°C , have been stored at -20°C for 9 months and for 14 days on ice in MAP.

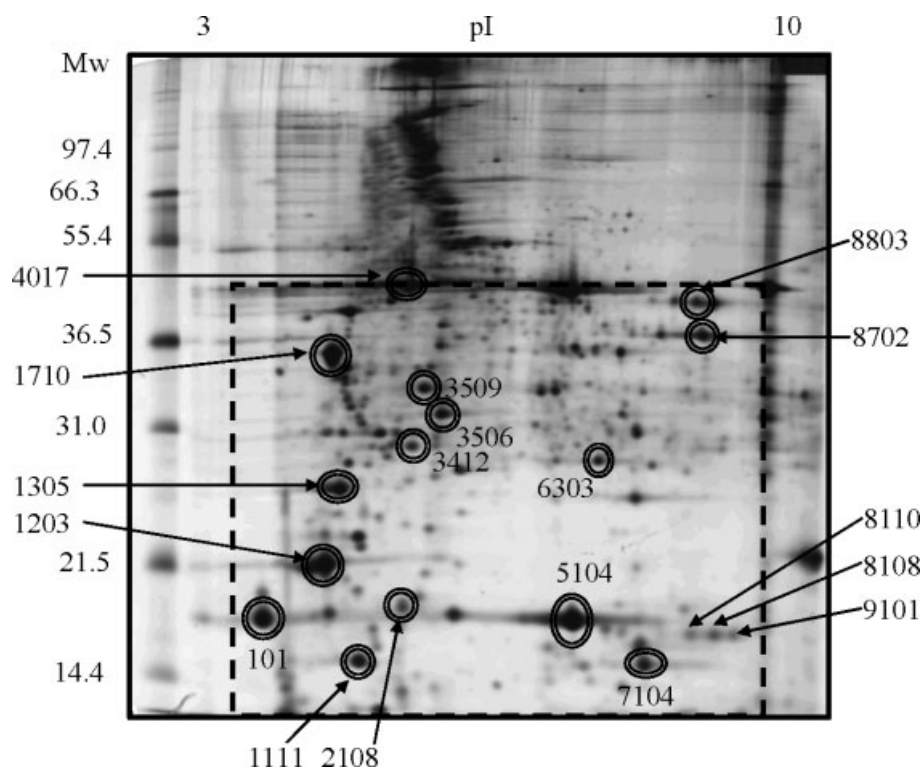


Figure 2. Representative 2-DE from cod muscle protein extracted after 6 months at -30°C and 7 days at 2°C in MAP. Extracted muscle proteins ($60\text{ }\mu\text{g}$) were first separated by linear pH 3–10 Immobiline DryStrips, followed by a size separation on 12% SDS-PAGE. Gels were silver-stained. The square marked by the dotted line specifies the area of the gels used for further proteome analysis. Spots marked by a circle are identified by MS/MS sequencing. Selected spots found to change in concentration according to storage time are labelled with their spot number.

of 3.5–9.0 and molecular weights of 13–40 kDa. A total of 22 samples (11 analogues) were separated by 2-DE, but three gels were omitted during final analysis due to cracks. The remaining 19 2-D gels represents proteins extracted under 11 different storage conditions (Fig. 1). Five representative partial 2-D gels representing some of the different storage conditions analysed are shown (Fig. 3). A total of 504 protein spots are found in the analysis set and 89 of these are represented in all gels. Visual examinations of

the 2-D gels reveal that six rather abundant protein spots disappear at the basic end after 12 months of frozen storage (Fig. 3).

3.2 Multivariate data analysis

In order to extract the optimal information from the data obtained from matching 2-D gels with PDQuest, multivariate analysis was conducted as earlier reported [25].

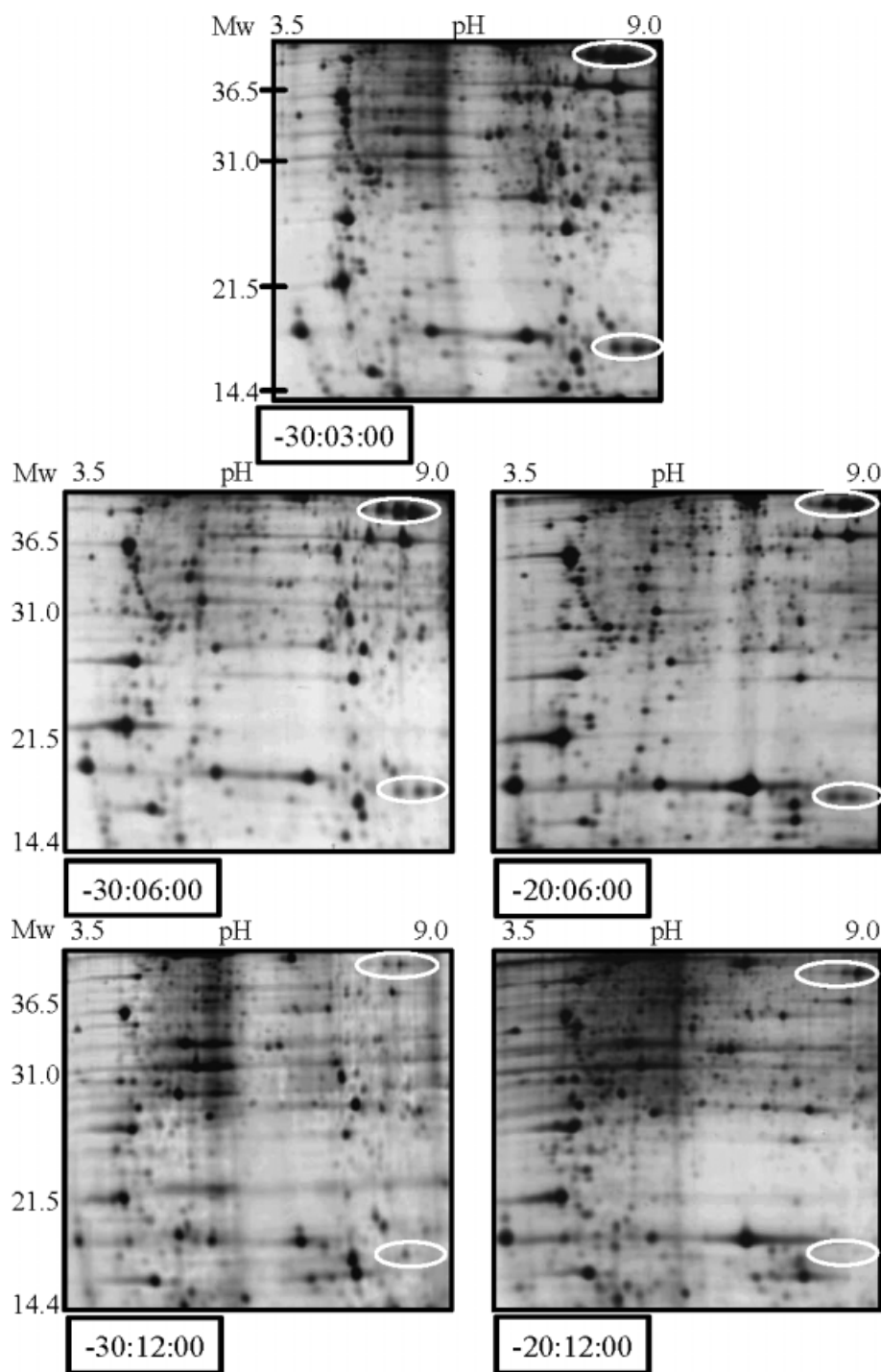


Figure 3. Representative 2-DE of cod muscle extracted at five different storage conditions. Extracted muscle proteins (60 μ g) were first separated by linear pH 3–10 Immobiline Dry-Strips, followed by a size separation on 12% SDS-PAGE. Gels were silver-stained. The partial proteome used for further analysis is shown for samples showing 2-DE of proteins extracted after storage at different conditions. Samples are named as described in Fig. 1. White eclipses are used to highlight the protein spots that disappear in intensity upon prolonged frozen storage.

In the PCA analysis described here, each 2-D gel (*e.g.* sample) represents a storage condition and each spot number an x variable, where the different protein spot volumes reflect the nature of the sample. In the PCA score plot (Fig. 4A) the samples are segmented in two clusters, one representing samples that have been frozen for 12 months and one representing those that have been frozen for either 3

or 6 months. Also, it can be seen (Fig. 4A) that samples from the different gel electrophoresis batches are evenly distributed throughout the score plot, thereby confirming that there is no confounding between the results and the experimental setup. In the PCA analysis 19% of the variation in the samples is explained along PC1 and PC2. In the PCA loadings plot (Fig. 4B) the x variables, *e.g.* the spot numbers, are

y-value	X-value	Y-value resulting range	No. of DPLSR factors	Correlation coefficient	RMSEP ^{a)}	% valid. Y var. expl. ^{b)}
Frozen storage time, months	All spots (504)	3–12	2	0.97	0.16	51
	Selected spots ^{c)} (288)	3–12	2	0.97	0.13	59
Frozen storage temperature, °C	All spots (504)	−20 to −30	− ^{d)}	−0.29	0.57	− ^{d)}
Storage in MAP at 2°C, days	All spots (504)	0–14	2	0.40	0.45	23

d) No correlation between the protein spots and the γ value was established; therefore, no validation value is reported.

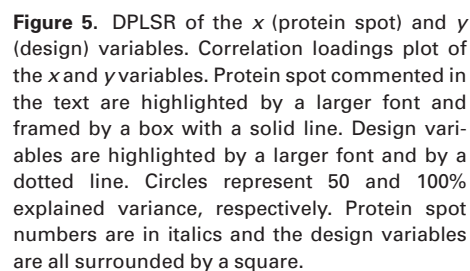
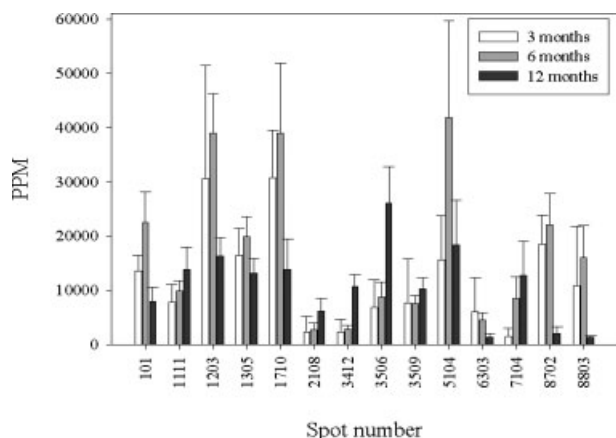


Table 2. Identification of cod protein spots 101, 1111, 1203, 2108, 1305, 1710, 3412, 3506, 3509, 4017, 5104, 6303, 7104, 8702, 8803, from cod by means of LC-MS/MS and cross-species matching with sequences in the databases

Spot no.	Description (species)	Accession no.	Total score	Peptides matched	Sequence coverage	MW _e ; pI _e	MW _{th} ; pI _{th}
101	MLC 3 (<i>Theragra chalcogramma</i>)	Q9DEA9_THECH	641	7	64.3% ^{a)}	17.6; 3.6	17.2; 4.3
1111	Fast skeletal muscle α -actin (<i>G. morhua</i>)	Q78AY8_GADMO	243	4	12.7%	14.5; 4.9	41.9; 5.3
1203	MLC 2 (<i>T. chalcogramma</i>)	Q9DEA8_THECH	612	6	45.6% ^{a)}	21.5; 3.9	19.1; 4.7
1305	MLC 1 (<i>T. chalcogramma</i>)	Q9IB19_THECH	677	7	43.8% ^{a)}	25.1; 4.3	21.2; 4.7
1710	Skeletal muscle tropomyosin1–2 (<i>Fugu rubripes</i>)	Q8JIM7_FUGRU	807	9	32.7% ^{a)}	36.4; 3.6	32.5; 4.7
2108	Fast skeletal myosin heavy chain (<i>G. morhua</i>)	Q8JIV5_GADMO	190	2	11.1%	18.3; 5.3	30.3; 8.7
3412	Fast skeletal muscle α -actin (<i>G. morhua</i>)	Q78AY8_GADMO	573	5	18.8%	30.5; 5.4	41.9; 5.3
3506	Fast skeletal muscle α -actin (<i>G. morhua</i>)	Q78AY8_GADMO	502	5	19.6%	32.5; 5.7	41.9; 5.3
3509	Fast skeletal muscle α -actin (<i>G. morhua</i>)	Q78AY8_GADMO	821	8	29.2%	34.2; 5.5	41.9; 5.3
4017	Fast skeletal muscle α -actin (<i>G. morhua</i>)	Q78AY8_GADMO	1217	14	43.8%	41.5; 5.4	41.9; 5.3
5104	Nuclease diphosphate kinase B (<i>Gillichthys mirabilis</i>)	Q9DFL9_GILMI	218	4	27.5% ^{a)}	17.1; 7.1	17.0; 7.45
6303	Triose-phosphate isomerase B (<i>Xiphophorus maculatus</i>)	Q7T315_XIPMA	584	8	39.7% ^{a)}	28.7; 7.4	26.5; 8.1
7104	Nuclease diphosphate kinase B (<i>G. mirabilis</i>)	Q9DFL9_GILMI	86	1	8.1% ^{a)}	14.1; 8.0	17.0; 7.45
8702	Glyceraldehyde-3-phosphate dehydrogenase (<i>G. morhua</i>)	Q8AWX8_GADMO	945	8	39.6% ^{a)}	37.9; 8.7	36.1; 8.2
8803	Aldolase A (<i>Brachydanio rerio</i>)	Q8JH72_BRARE	705	9	30.2% ^{a)}	40.0; 8.6	39.7; 8.4

MW_e and MW_{th} are experimental and theoretical molecular weights, respectively. MW is given in kDa. Accession number in Swiss-Prot and TrEMBL databases.

a) Indicates cross-species identification.

**Figure 6.** Spot intensity of selected protein spots, shown according to frozen storage time, e.g. 3, 6 or 12 months. Error bars correspond to the 95% confidence interval. Spots were selected from the loading plot.

species (Table 2). Except for the fragments the discrepancy between the theoretical and experimental MW is within the window of $\pm 0.20\%$ [28]. The discrepancy between the theoretical and experimental pI values is not always within the ± 0.25 U limit [28]; however, pI values are less conserved between species than MW values.

For all listed identifications the total score of MS BLAST correspond to significant hits [24] and homologies are found with analogue proteins from several fish species, however listed here (Table 2) are those giving the highest hits.

3.4 Changes in protein concentration of selected proteins

The concentration of full-length myofibrillar proteins MLC1, MLC2, MLC3 (MLC–myosin light chain), e.g. spots 1305, 1203, 101, respectively, and tropomyosin (spot 1710) (Fig. 6) generally declines from 3 to 12 months of frozen storage

($p < 0.05$). Tropomyosin represents a saturated spot for the 3 and 6 months frozen storage and therefore the calculations of the concentrations might be suboptimal; it is however clear that there is a distinct decrease in spot volume for this protein from 3 or 6 months frozen storage to 12 months of frozen storage. Other myofibrillar identifications represent fragments namely a myosin heavy chain fragment (spot 2108) and four α -actin fragments (spot 1111, 3412, 3506 and 3509), whereof two (spot 3412 and 3506) increase about five-fold from 3 months of frozen storage to 12 months ($p < 0.05$). The peptides obtained for spot 1111 are all located in the N-terminal end (peptides identified between positions 3 and 63) and for spots 3412, 3506 and 3509 all peptides are located in the C-terminal end (peptides identified from positions 199 to 374, 228 to 374 and 98 to 374, respectively).

The other identified proteins changing in concentration during frozen storage are sarcoplasmatic proteins involved in glycolysis. Triose-phosphate isomerase (spot 6303), glyceraldehyde-3-phosphate dehydrogenase (spot 8702) and aldolase A (spot 8803) decrease five-, nine- and eight-fold, respectively ($p < 0.05$), from 3 to 12 months of frozen storage. Spot 5104 is full-length nucleic diphosphate kinase B, whereas spot 7104 is identified as a fragment of this enzyme. The changes in spot 5104 are significant ($p < 0.05$) from 3 to 6 months of frozen storage and for the fragment (spot 7104) there is a significant increase from 3 to 6 months and from 3 to 12 months of frozen storage.

4 Discussion

The storage conditions of fish are known to have a substantial effect on eating quality. The proteomic data studied here represent variations according to three different storage conditions, *e.g.* frozen storage temperature, frozen storage time and storage at 2°C in MAP after defrosting of cod fillets. To achieve optimal information from the proteome analysis described, multivariate data analysis was applied to search for interrelated changes in the protein pattern characterising the samples in the experimental setup. A visual inspection of the PCA score plot (Fig. 4A) reveals the variation in protein pattern primarily to be related to frozen storage time as the samples are clustered according to this parameter. The loadings plot (Fig. 4B) gives an indication about which protein spots have relatively high or low intensities according to the clusters appearing in the score plot. For the validated variation in the spot intensities (Fig. 4C) 19% is explained along PC1 and PC2. Hereof 17% is described along PC1 where the samples form two groups, *e.g.* 12 months frozen storage *versus* 3 or 6 months frozen storage. Along PC2 the grouping cannot be explained according to differences in storage conditions, or experimental setup, but may be due to individual differences of the investigated fish, which are not described, for instance sex or age. Alternatively, the semiquantitative nature of proteomics may weaken the interpretation of the results, where especially the staining technique can be

afflicted with uncertainties, in respect to reproducibility, minor day-to-day variations and a limited dynamic range of the protein stain [29]. Since the samples were mixed during 2-DE (Fig. 4A) the major problems due to day-to-day variations are omitted and the influence of proteins stained to saturation is absent since these are weighted down in the multivariate calculations. It is possible that more explicit results could have been obtained by for instance fluorescent staining methods, however the multivariate data analyses show that despite the described uncertainties a systematic variation can be subscribed to frozen storage time.

To get the optimal information from the data set DPLSR models with one design variable was studied at the time (Table 1). This did reveal a correlation of the protein pattern with frozen storage time but not with frozen storage temperature or storage time at 2°C in MAP.

Therefore, the obtained results reveal that, within the molecular weight and *pI* values of the proteins studied here, the biggest alterations in cod fillet protein composition can be subscribed to the frozen storage time (Figs. 3–6) and that distinct changes appear between 6 and 12 months of frozen storage. The samples analysed were selected to screen the three different storage parameters and it is therefore possible that a different approach could have given more explicit details of changes for every single storage parameter. However, by applying the more explorative approach we now know that the frozen storage time has a pronounced effect on the protein pattern in contrast to the effect of two different freezing temperatures (−30 or −20°C) or storage time in MAP after defrosting. This resulted in a large number of protein spots (288) correlating with the frozen storage time (Fig. 5) whereof some have been identified by MS/MS sequencing (Table 2).

Identification of cod proteins required cross-species analysis based on peptide sequencing, since distant evolutionary relationships typically result in orthologous proteins with less than 70% sequence identity unsuitable for cross-species identification by PMF [30]. The success of cross-species identification is improved by a modified version of the standard BLAST algorithm [24]. This programme is advantageous for identification of proteins from MS/MS sequences, since it is targeted at matching several relatively short peptide sequences in the query with a homologue protein sequence from the database entry. The algorithm is based on an ungapped BLAST search identifying all 'high-scoring pairs' regions of high local sequence similarity between individual peptides in the query and a protein sequence from the database entry.

Generally frozen storage inhibits microbial spoilage, but the water-holding capability diminishes and the muscle texture becomes harder and more flaky partly due to modifications leading to decreasing protein solubility [5, 31]. Alterations in protein solubility and formation of aggregates have previously been studied by gel electrophoresis under different storage conditions and for different fish species [6, 7, 32–34]. In these studies the fish proteins were sequentially extracted by first measuring the amount of protein soluble in NaCl,

then SDS and finally SDS with β -mercaptoethanol. A general tendency is that the solubility of proteins in NaCl decreases during prolonged frozen storage, thereby indicating a shift in protein solubility due to increased protein aggregation. That addition of the reducing agent β -mercaptoethanol increases the protein solubility highlights the importance of disulphide covalent bonds in aggregation. In the work described here, cod muscle proteins were extracted in 2% SDS and 100 mM DTT and the overall protein solubility did not change throughout the time course of the experiment. This indicates that the buffer used for protein extraction masks the overall differences in protein solubility; therefore, the general differences observed in 2-DE are due to cleavages in the protein chains and not due to different solubility of the proteins according to storage time. However, for some distinct proteins the solubility may change during storage, thereby giving the impression that they disappear even though they solely might have reduced solubility in SDS and DTT.

The causes of textural changes in meat and fish is of major interest and there has been an immense debate on the degradation pattern of specific proteins [35–43]. Even though a decreased solubility of α -actin has been reported [6, 7], actin degradation has generally not been observed [36, 37, 39] except in one study on pacific whiting (*Merluccius productus*) [38]. However, recent 2-DE studies have revealed fragmentation of pig α -actin during postmortem storage [44–46]. It is plausible that the detailed picture of the protein composition emerging from this technique gives a more comprehensive picture than the otherwise applied SDS-PAGE approach, thereby providing new information on the changes appearing in fish and meat after slaughter.

Within the time span of the experiment described here the full-length myofibrillar proteins MLC1 (spot 1305), MLC2 (spot 1203), MLC3 (spot 101), and tropomyosin (spot 1710) decline in concentration. The two α -actin fragments (spots 3412 and 3506) increase in concentration thereby implying an α -actin breakdown concurrent with prolonged frozen storage of cod. In both cod and pig α -actin is 377 aa long and the sequence identity between the pig and cod protein is 98%. It is remarkable that an α -actin fragment appearing in pig has a size of 32 kDa and a *pI* value of 5.3, which are basically the same as the values reported here for one of the cod fragments (spot 3506) with a size of 32.5 kDa and a *pI* value of 5.7. The high sequence identity between the two different species and the common characteristics of fragment size imply that the same type of protein degradation might occur in the two different species.

MLCs bind to the globular region of myosin heavy chain. Together the MLCs and myosin heavy chain are the main proteins forming the thick filaments. Besides an ATP- and two MLC-binding regions myosin heavy chain contains an actin-binding region, which is the major protein of the thin filaments. The results described here reveal a simultaneous decrease in concentration of MLC1, MLC2, MLC3, tropomyosin and increase in α -actin fragments. These contractile proteins are abundant in the muscle, and therefore degrada-

tion, changes in interactions and bindings of these proteins may signify an important point of the correlation between frozen storage time and textural damage. During the storage experiment two α -actin fragments (spots 1111 and 3509) and a myosin heavy chain fragment do not change significantly in concentration. These fragments might represent intermediate fragments generated from full-length proteins where the intermediate fragments are cleaved further into smaller fragments not identified here.

During frozen storage triose-phosphate isomerase, aldolase A and glyceraldehyde-3-dehydrogenase decrease about five-, eight- and nine-fold, respectively, from 3 to 12 months of frozen storage indicating that the enzymes disappear from the cod fillets. All three enzymes are involved in the glycolytic pathway and structurally triose-phosphate isomerase binds to other glycolytic enzymes already bound to the thin filaments in the muscle sarcomer, such as aldolase A and glyceraldehyde-3-dehydrogenase [47]. The breakdown of nuclease diphosphate kinase B is verified by the increase in concentration of spot 7104. Nuclease diphosphate kinases are involved in multiple cellular functions and are required for synthesis of nucleoside triphosphates (NTP) other than ATP by transferring the terminal phosphates from ATP. The enzymes are located in the nucleolus but might also be colocalised with actin [48].

The increase of α -actin fragments during frozen storage verifies that a degradation of the thin filaments occurs. Since triose-phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase and aldolase A are located in this region of muscle sarcomers, alterations in their binding pattern is expected concurrently with α -actin degradation. This degradation is expected to result in a softer structure of the thin filaments leading to an increased exposure of the enzymes present, in respect to both proteolysis and mechanical stress. This might explain the simultaneous degradation of triose-phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase and aldolase A.

The biochemical processes in fish muscle are expected to be different before and after rigor, *e.g.* oxygen depletion, during initial freezing and during prolonged frozen storage. In a previous study on postmortem changes during ice storage it is found that the protein spot corresponding to triose-phosphate isomerase increases about three-fold in concentration within the first 18 h of storage [18], which is the time point when rigor begins. This increase in concentration of triose-phosphate isomerase during ice storage from just slaughtered until rigor occurs indicates [18] a shift in metabolism in the postmortem muscle probably due to ATP depletion. Likewise triose-phosphate isomerase increases about five-fold in pig during postmortem storage [45]. In the study on protein changes during ice storage of cod, MLC2 and MLC3 have been found to increase about three times within the first 18 h of ice storage [18]. Interestingly a similar study on postmortem storage of pig for 72 h shows the concentration of MLC2 to increase about five times during postmortem storage [45, 46]. It is remarkable that some of the proteins changing in concentra-

tion are the same during ice storage, frozen storage and in species as different as cod and pig; this implies that similar biochemical mechanisms are involved during all processes.

The choice of proteins for MS/MS sequencing was based on a combination of protein spots with a sufficient size for sequencing and protein spots that appeared to change during frozen storage. The protein spots changing in concentration represents an average of different storage temperatures and days of storage in MAP and it is therefore likely that a more distinct pattern could evolve if the analysis was applied on more samples from each condition or for instance at different temperatures. Future work would therefore include a more comprehensive analysis of the protein spots and extended focus on the multivariate data analysis eventually searching for a connection between the alterations of protein pattern with different sensory aspects of the fish during different storage conditions.

The results demonstrate protein changes induced due to frozen storage time to have major influence on the protein pattern compared to those induced due to different freezing temperatures or storage time in MAP. Furthermore, combining proteomics and multivariate data analysis, to the samples described here, has drawn the attention to remarkably related proteins and protein fragments in the otherwise immense amount of proteomic data. It is notable how the identified proteins fall into groups where the proteins are connected. Three of the sarcoplasmic proteins identified this way are located in the same subcellular structure in the muscle fibre and all three disappear with a similar pattern from the cod muscle. For the fourth sarcoplasmic proteins the full-length proteins disappear coherently with an increase of concentration of the fragment. In respect to the myofibrillar proteins all three MLCs and tropomyosin are degraded in a similar pattern and at the same time two α -actin fragments increase in concentration. One of the α -actin fragments is also found in pig [45] thereby indicating that the same type of protein degradation appears in both fish and pig meat. The results achieved here provide an important basis for future studies on changes in protein composition/protein degradation during storage of cod and fish in general.

We acknowledge economic support from the Danish Ministry of Agriculture and Fisheries grant 93S-945-A99-00058 and grant FSK03-DFU-11. We are grateful to Karin Reimers for excellent technical assistance.

5 References

- [1] Blackstock, W. P., Weir, M. P., *Trends Biotechnol.* 1999, 17, 121–127.
- [2] Dowsey, A. W., Dunn, M. J., Yang, G. Z., *Proteomics* 2003, 3, 1567–1596.
- [3] Frankel, E. N., in: Frankel, E. N. (Ed.), *Lipid Oxidation*, Bell & Bain, Glasgow 1998, pp. 187–225.
- [4] Brown, W. D., in: Bechtel, P. J. (Ed.), *Muscle as Food*, Academic Press, San Diego, CA (USA), 1986, pp. 405–451.
- [5] Mackie, I. M., *Food Rev. Int.* 1993, 9, 575–610.
- [6] Careche, M., del Mazo, M. L., Fernandez-Martin, F., *J. Sci. Food Agric.* 2002, 82, 1791–1799.
- [7] Tejada, M., Careche, M., Torrejon, P., delMazo, M. L. *et al.*, *J. Agric. Food Chem.* 1996, 44, 3308–3314.
- [8] Jensen, O. N., Larsen, M. R., Roepstorff, P., *Proteins Suppl.* 1998, 2, 74–89.
- [9] Pineiro, C., Barros-Velazquez, J., Vazquez, J., Figueras, A., Gallardo, J. M., *J. Proteome Res.* 2003, 2, 127–135.
- [10] Pineiro, C., Vazquez, J., Marina, A. I., Barros-Velazquez, J., Gallardo, J. M., *Electrophoresis* 2001, 22, 1545–1552.
- [11] Taylor, J., Giometti, C. S., *Electrophoresis* 1992, 13, 162–168.
- [12] Jacobsen, S., Nesic, L., Petersen, M., Søndergaard, I., *Electrophoresis* 2001, 22, 1242–1245.
- [13] Klenø, T. G., Leonardsen, L. R., Kjeldal, H. O., Laursen, S. M. *et al.*, *Proteomics* 2004, 4, 868–880.
- [14] Verhoeckx, K. C. M., Bijlsma, S., de Groene, E. M., Witkamp, R. F. *et al.*, *Proteomics* 2004, 4, 1014–1028.
- [15] Esbensen, K. H., in: Esbensen, K. H. (Ed.), *Multivariate Data Analysis-In Practice*, Camo Asa/Camo process As 2000.
- [16] Wold, S., Sjöström, M., Eriksson, L., *Chemom. Intell. Lab. Syst.* 2001, 58, 109–130.
- [17] Bøknæs, N., Jensen, K. N., Guldager, H. S., østerberg, C. *et al.*, *Lebensm.-Wiss. u. -Technol.* 2002, 35, 436–443.
- [18] Kjærsgård, I. V. H., Jessen, F., *J. Agric. Food Chem.* 2003, 51, 3985–3991.
- [19] Görg, A., Obermaier, C., Boguth, G., Harder, A. *et al.*, *Electrophoresis* 2000, 21, 1037–1053.
- [20] Hochstrasser, D. F., Harrington, M. G., Hochstrasser, A. C., Miller, M. J., Merrill, C. R., *Anal. Biochem.* 1988, 173, 424–435.
- [21] Martens, H., Martens, M., in: Martens, H., Martens, M. (Eds.), *Multivariate Analysis of Quality-An Introduction*, John Wiley & Sons, Chichester 2000.
- [22] Voss, T., Haberl, P., *Electrophoresis* 2000, 21, 3345–3350.
- [23] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, 68, 850–858.
- [24] Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A. *et al.*, *Anal. Chem.* 2001, 73, 1917–1926.
- [25] Jessen, F., Lametsch, R., Bendixen, E., Kjærsgård, I. V. H., Jørgensen, B. M., *Proteomics* 2002, 2, 32–35.
- [26] Westad, F., Martens, H., *J. Near Infrared Spectrosc.* 2000, 8, 117–124.
- [27] Martens, H., Martens, M., *Food Qual. Prefer.* 2000, 11, 5–16.
- [28] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K. *et al.*, *Bio/Technology* 1996, 14, 61–65.
- [29] Garfin, D. E., *Trac-Trends Anal. Chem.* 2003, 22, 263–272.
- [30] Lester, P. J., Hubbard, S. J., *Proteomics* 2002, 2, 1392–1405.
- [31] Sotelo, C. G., Pineiro, C., Perez-Martin, R. I., *Z. Lebensm. Unters. Forsch.* 1995, 200, 14–23.
- [32] Tejada, M., Huidobro, A., Mohamed, G. F., *J. Sci. Food Agric.* 2003, 83, 113–122.
- [33] Badii, F., Howell, N. K., *J. Sci. Food Agric.* 2002, 82, 87–97.
- [34] Saeed, S., Howell, N. K., *J. Sci. Food Agric.* 2002, 82, 579–586.
- [35] Seki, N., Watanabe, T., *J. Biochem.* 1984, 95, 1161–1167.

- [36] Seki, N., Tsuchiya, H., *Nippon Suisan Gakkaishi* 1991, 57, 927–933.
- [37] Benjakul, S., Seymour, T. A., Morrissey, M. T., An, H. J., *J. Food Sci.* 1997, 62, 729–733.
- [38] Lin, T. M., Park, J. W., *J. Food Sci.* 1996, 61, 536–539.
- [39] Taylor, R. G., Geesink, G. H., Thompson, V. F., Koohmaraie, M., Goll, D. E., *J. Anim. Sci.* 1995, 73, 1351–1367.
- [40] Geesink, G. H., Morton, J. D., Kent, M. P., Bickerstaffe, R., *J. Food Sci.* 2000, 65, 1318–1324.
- [41] Verrez-Bagnis, V., in: Olafsdóttir, G., Luten, P., Dalgaard, P., Careche, M., Verrez-Bagnis, V., Martinsdóttir, E., Heia, K. (Eds.), *Proceedings of the Final Meeting of the Concerted Action "Evaluation of Fish Freshness"*, International Institute of Refrigeration, Nantes, France 1997, pp. 229–237.
- [42] Verrez-Bagnis, V., Noel, J., Sautereau, C., Fleurence, J., *J. Food Sci.* 1999, 64, 240–242.
- [43] Verrez-Bagnis, V., Ladrat, C., Noelle, J., Fleurence, J., *J. Sci. Food Agric.* 2002, 82, 1256–1262.
- [44] Lametsch, R., Roepstorff, P., Bendixen, E., *J. Agric. Food Chem.* 2002, 50, 5508–5512.
- [45] Lametsch, R., Karlsson, A., Rosenvold, K., Andersen, H. J. *et al.*, *J. Agric. Food Chem.* 2003, 51, 6992–6997.
- [46] Morzel, M., Chambon, C., Hamelin, M., Sante-Lhoutellier, V. *et al.*, *Meat Sci.* 2004, 67, 689–696.
- [47] Stephan, P., Clarke, F., Morton, D., *Biochim. Biophys. Acta* 1986, 873, 127–135.
- [48] Bosnar, M. H., de Gunzburg, J., Bago, R., Brecevic, L. *et al.*, *Exp. Cell Res.* 2004, 298, 275–284.