ELSEVIER

Carbohydrate Polymers 41 (2000) 163–174

Carbohydrate Polymers

www.elsevier.com/locate/carbpol

Starch molecular structure and phosphorylation investigated by a combined chromatographic and chemometric approach

A. Blennow[a,*], S.B. Engelsen[b], L. Munck[b], B.L. Møller[a]

[a]*The Royal Veterinary and Agricultural University, Plant Biochemistry Laboratory, Department of Plant Biology, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark*

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

Received 19 January 1999; received in revised form 27 April 1999; accepted 1 June 1999

Abstract

Structural features of starch were studied with special emphasis on the relationship between starch phosphorylation and starch chain length distribution comparing a chemometric approach with classic statistics. Starches prepared from 44 plant species were analysed with respect to the degree of phosphorylation and chain length distribution of the neutral unit chains, prepared by enzymic isoamylase debranching, using high performance anion exchange chromatography with pulsed amperometric detection (HPAEC/PAD). Chemometric algorithms such as Principle Component Analysis (PCA), Non-Negative Alternating Least Squares Regression (NNALSR) and Partial Least Squares regression (PLS) were used to analyse the systematic variation in the chromatograms and compared to Gauss decomposition. Detailed relations between chain length data and structural elements of the current cluster/bocklet model of native starch granules were revealed. By using PCA, both crystal polymorphs and botanical origin of the starch were accurately predicted. Using PLS, a strong correlation (0.93) was obtained between the chain length distribution and the degree of phosphorylation in the potato starch group. The use of chemometrics as an efficient tool to classify and predict starch functionality is documented.

Keywords: Starch; Chemometrics; Chromatography; Phosphorylation

1. Introduction

Starch is a polysaccharide of fundamental importance as a digestible polymer in food industry and as a constituent in various non-food products (Ellis, et al., 1998). Native starch consists of two different types of glucose polymers, amylose and amylopectin. The main glucosidic linkage between glucose units in amylose is in α-1,4 configuration, while in amylopectin additional α-1,6 glucosidic bonds make branch points in the macromolecule (Manners, 1989). Starch is stored as granules in the plant and the morphology and crystallinity of a starch granule are determined by the amylopectin fraction which constitutes about 70–80% of normal starch. Amylopectin is thought to be arranged in the granule as clusters of radially oriented chains organised in super helical and semi-crystalline blocks (Gallant, Bouchet & Baldwin, 1997). The proposed model has emerged mainly from chain length distribution analysis of debranched amylopectin (Hizukuri, 1986), electron microscopy (Oostergetel & van Bruggen, 1993), polarised light microscopy (French, 1972), electron diffraction microscopy and fibre X-ray crystallography (Imberty, Chanzy, Pérez, Buléon & Tran, 1988; Imberty & Pérez, 1988; Imberty, Buléon, Tran & Pérez, 1991). Each chain cluster contains a branched, amorphous base from which a more compact crystalline region with parallel double helices emerges. The cluster dimension is conserved and independent of the botanical source, and as a result starch granules consist of lamellar structures of 9-nm periodicity. Natural amylopectin crystallise during biosynthesis as A-, B- or C-type polymorph (Hizukuri, 1985; Zobel, 1988). The A polymorph consists of densely hexagonally packed double helices and is generally found in cereal starches (Imberty et al., 1988), while the B polymorph contains channels of structured water and is mostly found in tuberous starches (Imberty et al., 1988). The C-type polymorph is a mixture of A- and B-type starches (Gidley & Bulpin, 1987).

A small fraction of the glucose units (0.1–1%) in most amylopectins also contains phosphoryl groups monoester linked to the O-2, O-3 or O-6 hydroxyl groups (Takeda & Hizukuri, 1971). Phosphate groups in starch are esterified to

* Corresponding author. Tel.: + 45-35-28-3334; fax: + 45-35-28-3333.
E-mail address: abl@kul.dk (A. Blennow)

0144-8617/00/$ - see front matter
PII: S0144-8617(99)00082-X

long chains of the amylopectin (Blennow, Bay-Smidt, Wischmann, Olsen & Møller, 1998; Takeda & Hizukuri, 1982), and positioned more than 9 glucose units from a branch point (Takeda & Hizukuri, 1981). The location of phosphate groups in relation to the clusters is not known but starch phosphorylation has been related to the distribution of neutral chains in amylopectin (Blennow et al., 1998) indicating a direct influence of the phosphate groups on the general structure of starch. The function of starch phosphate in the starch metabolic system is not clear but it has been shown that starch phosphorylation occurs concomitantly with starch biosynthesis (Nielsen, Wischmann, Enevoldsen & Møller, 1994) possibly involving starch branching enzyme (Viksø-Nielsen, Blennow, Nielsen & Møller, 1998).

Starch functionality is partly determined by variation in its glucosidic and phospho monoester covalent configuration. However, physicochemical properties of starch are not always easily predicted by analysis of separate molecular parameters e.g. phosphorylation, chain lengths or crystallinity (Kim, Wiesenborn, Orr & Grant, 1995; Lorberth, Ritte & Kossmann, 1998; Muhrbeck, Svensson & Eliasson, 1991) as a result of apparent co-variations between these factors. The relationships between chain length and phosphorylation (Blennow et al., 1998) as well as between chain length and polymorph type (Hizukuri, 1985), however, suggest starch functional properties to be predicted from either distribution or phosphorylation data. The main purpose of this work is to investigate the use of multivariate chemometric methods to reveal functional information hidden in the chain length profiles.

Classic curve-resolution techniques require many uncontrollable assumptions. In contrast, multivariate data-techniques require few or no assumptions. Decomposition of chromatographic information with a set of e.g. Gaussian bands without knowledge of the number of components contributing to the overall band-shape represents a usual, but uncontrollable approximation which can result in ambiguous or, in the worst case, erroneous results. As it is sensible to assume that observed polymodal chain length distributions of amylopectin is composed of a set of normally distributed populations of chains formed by "semi-specific" enzyme activity this technique is the de facto standard for analysing chromatograms of enzymatic degraded starch (Ong, Jumel, Tokarczuk, Blanshard & Harding, 1994). Using a two-dimensional chemometric evaluation method in which many different starch samples are included in the chromatographic analysis it is possible, in principle, to extract the underlying phenomena (latent variables) without making assumptions about the number of bands and their shape. The information is extracted from a matrix of chromatographic data defined by a number of different types of starches and their chromatograms.

In this work, the systematic variations in the polymodal chromatograms were analysed using three different multivariate data analysis methods: Principal Component Analysis (PCA); Partial Least Squares regression (PLS) and Non-Negative Alternating Least Squares Regression (NNALSR). When applying multivariate data analysis it is obligatory to have chromatograms available providing data covering the relevant variance space. To meet these requirements, we have collected and measured chain length distributions of 44 different starches with the aim of defining (1) the number of underlying DP (degree of polymerisation) profiles; (2) the shapes of these underlying DP profiles and (3) investigating the influence of phosphorylation upon the DP profiles. For reference a one-dimensional fitting approach (each chromatogram is treated individually) was carried out using decomposition into Gauss functions with forced integer maxima but with no a priori expectations to DP maxima of the Gauss bands.

The investigation documents a surprisingly simple structural organisation related to the degree of starch phosphorylation and the proposed general features of the amylopectin cluster dimension. The chemometric approach employed in this study is suggested to be useful as a method for predicting functional properties of starches from various sources, including in planta modified starches.

2. Material and methods

2.1. Plant material and starches

Gajutu starch was kindly provided by Dr K. Yamamoto (Hokuren Federation Agricultural Co-operatives, Japan) and shoti starch was from a local market. Maize starches were obtained from National Starch Chemical, and potato starches from lines 658, 659, 661 and 663 were kindly provided by Danisco Biotechnology (Denmark). Barley mutant starches PLP3A, PLP1A2, PLP2A and PLP1B were isolated from Hordeum vulgare cv. Pallas by Rasmussen and Hatzack (1998). These mutants are characterised by having low phytate content and high free phosphate content. Rice starch was from Sigma. All other starches were isolated and purified from stem tubers, rhizomes, kernels or leaves obtained from plants grown in a greenhouse as described elsewhere (Blennow et al., 1998).

2.2. Starch analyses

Starch samples (5 mg) were gelatinised and debranched using the amylolytic endo enzyme isoamylase (Megazyme, Sydney, Australia), as described earlier (Blennow et al., 1998). High performance anionic exchange chromatography (HPAEC) of neutral linear oligosaccharides were performed using a (Dionex DX 500) system fitted with a CarboPac PA-100 column and equipped with an S-3500 auto sampler, a GP40 pump, and an ED40 pulsed amperometric detector (PAD). Samples of 30 μl (150 μg of α-glucan), prepared as described above, were injected and separated using 1 ml/min flow rate, 150 mM isocratic NaOH and the following NaOAc gradient profile: 0–5 min

Table 1
Degree of phosphorylation and crystal polymorph of starches and peak DP of Gaussian bands

Source	nmole Glc6 P/mg starch	Polymorph	Peak DP of Gauss band		
			1	**2**	**3**
Potato (*Solanum spp.*) *starches*					
S. t. N-82-ATH-2	7.8	B[a]	13	20	49
S. t. Dianella OP	8.5	B[a]	13	18	49
S. t. 90-BNT-2	8.8	B[a]	12	19	49
S. vernei 1642/2	12.2	B[a]	13	19	49
S. t. 663	13.4	B[a]	12	19	47
S. t. Dianella	14.0	B[a]	13	19	50
S. t. Kaptah	18.1	B[a]	14	20	47
S. t. 659	18.1	B[a]	14	19	46
S. t. Godiva	19.0	B[a]	14	20	49
S. t. 658	22.1	B[a]	13	18	48
S. t. 661	25.3	B[a]	13	19	46
S. phureja	28.6	B[a]	14	19	52
S. vernei (11)	29.2	B[a]	15	21	51
S. t. 90BKG-22	32.6	B[a]	14	20	53
S. t. 87BDN	33.1	B[a]	13	20	51
S. neorossi (511)	33.5	B[a]	14	20	48
Ginger (*Zingiberaceae*) *starches*					
Aframomum meleguata	0.40	m[b]	14	21	48
Hedychium spp. 2	0.73	m[b]	15	23	51
Curcuma domestica	1.8	m[b]	14	21	49
Boesenbergia rotunda	2.1	m[b]	14	19	47
Hedychium coronarium	2.4	B[c]	16	24	53
Hedychium spp. 1	6.7	m[b]	15	22	44
Curcuma armada	10.0	m[b]	13	21	50
Alpinia calcavata	22.8	m[b]	16	23	54
Curcuma zedoaria (Gajutu starch)	58.8	B[c]	14	31	37
Curcuma zedoaria (Shoti starch)	60.0	B[c]	17	24	54
Curcuma zedoaria	63.0	B[c]	17	24	53
Barley (*Hordeum vulgare*) *starches*					
H. v. PLP3A	0.0	A[d]	13	20	42
H. v. PLP1A2	0.0	A[d]	12	19	40
H. v. PLP2A	0.0	A[d]	12	19	42
H. v. PLP1B	0.0	A[d]	12	19	46
H. v. Bonus	0.0	A[d]	12	18	44
Maize (*Zea mays*) *starches*					
Z. m. Waxy CS	0.0	A[a]	13	19	41
Z. m. Waxy WT	0.0	A[a]	13	20	42
Z. m. Waxy WV	0.0	A[a]	13	20	42
Z. m. Waxy SA	0.0	A[a]	13	20	42
Z. m. Waxy WY	0.0	A[a]	13	20	42
Z. m. Waxy WR	0.0	A[a]	13	20	42
Miscellaneous starches					
Sorghum (Sorghum bicolor)	0.9	A[e]	13	19	47
Rice (Oryza sativa)	1.0	A[a]	13	19	41
Cassava (Manihot escuenta)	2.5	C[e,f]	13	19	46
Mung bean (Phaseolus aureus)	3.5	C[g]	13	20	48
Potato leaf starch (cv. Dianella)	4.5	m[b]	13	20	51
Arrow root (Maranata arundinacea)	4.6	C[a,d]	13	20	47

[a] Hizukuri (1985).
[b] m, information missing.
[c] Fujimoto et al. (1984).
[d] Hanashiro, Abe and Hizukuri (1996).
[e] Zobel (1988).
[f] Oates (1997).
[g] Hoover, Li, Hynes and Senanayake (1997).

linear gradient of 0–110 mM NaOAc; 5–130 min convex gradient (curve 4) of 110–350 mM NaOAc. Amylose and phosphorylated oligosaccharides present in the samples had a much higher affinity to the column than the neutral oligosaccharides and therefore did not interfere with the elution profile of neutral chains.

Preparation of phosphorylated amylopectin unit chains from potato and shoti starch was carried out as previously described (Blennow et al., 1998). Enzymic dephosphorylation of the phosphorylated chains (0.5–1 mg) was performed in 1 ml reaction mixtures containing 10 mM of Bicine–NaOH buffer, pH 8.5, and 80 μg of EIA-grade alkaline phosphatase (Boehringer Mannheim, Germany). Possible traces of amylolytic activity in the enzyme preparation were removed by passing the enzyme over a 5 ml α-cyclodextrin column (Viksø-Nielsen et al., 1998). The mixture was incubated for 15 h at 37°C and the reaction terminated by boiling for 3 min. Remaining phosphorylated chains which were resistant to dephosphorylation (20% of the chains in the potato starch sample and 80% of the chains in the shoti starch sample) were removed by DEAE–sepharose chromatography (Blennow et al., 1998). All chromatograms were integrated and the areas obtained were corrected for the decreasing detector response sensitivity with chain length (Blennow et al., 1998).

Starch phosphate content was determined as nmol glucose 6-phosphate/mg starch (Bay-Smidt, Wischmann, Olsen & Nielsen, 1994). For the maize samples, absence of phosphate in this starch was determined by HPAEC (Blennow, Bay-Smidt, Olsen & Møller, 1999).

Total sugar was determined using the phenol–sulphuric acid method as described by Dubois, Gilles, Hamilton, Rebers and Smith (1956).

2.3. Data analytical methods

The systematic variations in the polymodal chromatograms were analysed using three different chemometric methods: PCA, PLS and NNALSR. For reference a one-dimensional fitting approach (each chromatogram is treated individually) using Gaussian functions was applied.

2.3.1. Gaussian decomposition

In this approach polymodal distribution chromatograms were fitted successively to one, two and three Gaussian distributions according to the equation:

$$G(n) = \sum_i C_i \exp[-((n - \text{centre}_i)/(w_i/2))^2]$$

where n is the degree of polymerisation, i is the number of Gaussians, C_i is the maximum intensity (amount) of the ith Gaussian, centre_i is the peak centre of the ith Gaussian and w_i is the distance from centre to 2% of full height. In the fitting procedure new Gaussians are successively added using the maximum residual point as the new centre point guess. Firstly, Gaussians are least-squares fitted according to the formula described above, i.e. the centre point is fitted as a continuous floating point number. Secondly, when an optimal Gaussian is found the fitting is repeated using a penalty function forcing the initial guesses of the centre point to become integral. The curve fitting of the polymodal distribution profiles was carried out using a Simplex algorithm (Nelder & Mead, 1965) for the non-linear parameters centre_i and w_i combined with a least squares fit of the linear C_i parameters inside the Gauss function evaluation call. Although Gaussian decomposition is an inherently non-robust approach, the algorithm described above proved to exhibit fairly robust convergence behaviour and to provide fairly optimal and unique solutions. However, in few cases it was necessary to adjust the initial guess of the centre value of the third Gaussian to include a better description of the long chain length profiles.

2.3.2. Principal component analysis

Most of the latent variables methods used in multivariate data analysis are in one way or another based on PCA (Wold, Esbesnen & Geladi, 1987). The PCA method provides a simple and efficient technique for graphically describing systematic variations in complex data structures. The essence of the principal component methods lies in their construction of orthogonal latent factors (or principal components) from underlying latent structures in the original data. The multidimensional data set is resolved into orthogonal components whose linear combinations approximate the original data set in a least squares sense. In PCA the original two-dimensional data matrix (**X**) (samples × variables) is decomposed into a score matrix (**T**) and a loading matrix (**P**), and the residuals are collected in a matrix (**E**) : $\mathbf{X} = \bar{\mathbf{X}} + \mathrm{TP}^{\mathrm{T}} + \mathbf{E}$. The principle is illustrated in Fig. 4. The systematic variation is described by the principal components (PC1, PC2 etc.) which each represent the outer product of scores and loadings. The loading vectors for the principal components can be considered as pure hidden profiles that are common to all the measured chromatograms. What makes the individual raw chromatograms different is the amounts (scores) of hidden profiles.

2.3.3. Partial least-squares regression

The standard multivariate calibration approach is to use PCR or PLS regression (Geladi & Kowalski, 1987), where the independent variables are decomposed into a set of scores. The dependent variables are then regressed on these scores instead of the original variables. The model structures of PCR and PLS are the same bi-linear structure as in PCA. The general purpose of PLS is multivariate calibration, i.e. to find a mathematical relation between two data sets e.g. a matrix of chromatograms (**X**) and a vector with a selected functionality (**y**). PLS performs a simultaneous decomposition of **X** and **y** in such a way that the information in the **y** vector is directly used as a guide for the decomposition of **X**. In matrix notation we have the linear model **y** =

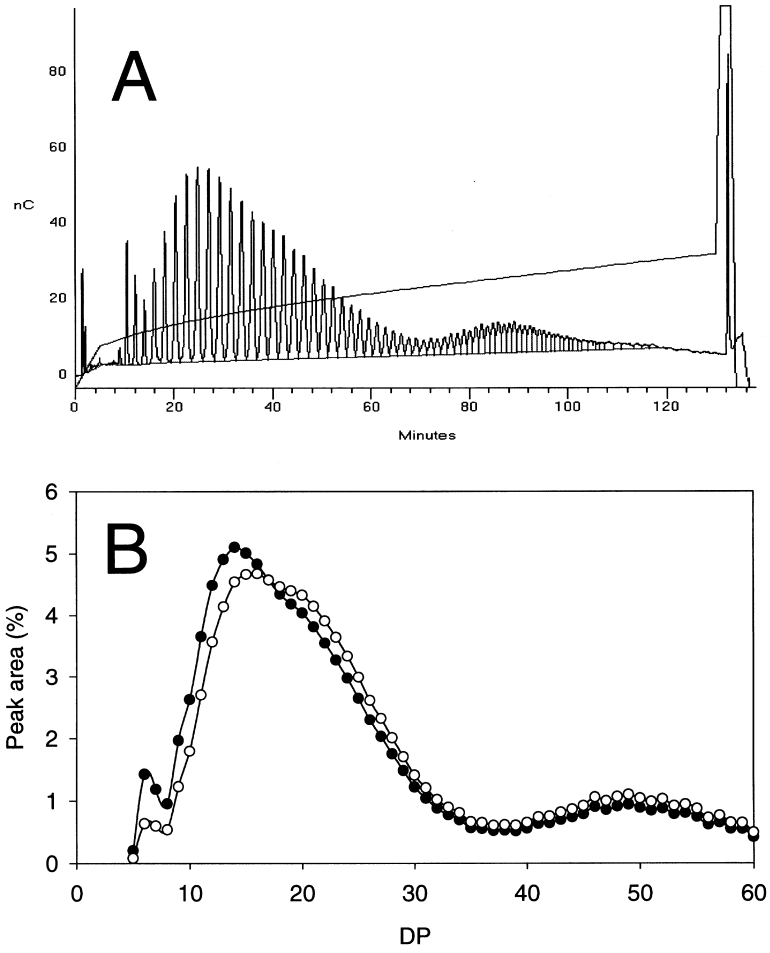

Fig. 1. Analysis of debranched starch. Starch granules were disintegrated in NaOH, treated with isoamylase at pH 4.0 for 2h and the sample analysed by HPAEC/PAD. (A). Original chromatogram of potato starch (cv Godiva). (B). Integrated peaks (filled symbols) and corrected (Blennow et al., 1998) peak areas (open symbols). Total peak area is set to 100%. (nC: nanocolumb).

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

2.3.4. Non-negative alternating least squares regression

Non-negative ALSR can be regarded as an analogue to PCA which does not require the latent factors to be orthogonal. It provides a simple repetitive solution to two regression problems. Given a few profiles (**S**), the amounts (**C**) required to reconstruct the original data (**X**) are calculated by the least squares solution: $\mathbf{C} = \mathbf{XS}(\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}$. Then an improved estimate of the profiles (**S**) is obtained by calculating: $\mathbf{S} = \mathbf{X}^{\mathrm{T}}\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}$ from which an improved estimate of the amount can be calculated and so forth. By continuing this iterative approach until no further improvement is observed, a best fit in the least squares sense is obtained to the original data matrix (**X**). In its native form ALSR does not impose any constraints on the latent variables for which reason the solution in a pure ALSR approach is often sensitive to the initial guesses of hidden profiles. A significant stabilisation of the ALSR solutions can often in practice be dealt with by adding constraints. As negative solutions have no sense in chromatographic data we applied non-negativity constraint implemented in the least squares sense (Bro & de Jong, 1997; Lawson & Hanson, 1974) with the sacrifice of a significant decrease in convergence speed.

2.3.5. Validation

In this work only the PCA and PLS results are validated. The validation was performed as full cross-validation (Martens & Dardenne, 1998) that successively leaves out one sample at a time from the calibration set and uses the rest for establishment of the calibration model. Throughout this study the parameter root mean square error of cross-validation (RMSECV) is used as an indicator for the overall prediction ability of the model.

2.3.6. Programs

Gaussian curve fitting and NNALSR was performed using in-house software written in Matlab (The Mathworks Inc., MA, USA), while PCA and PLS models were calculated by use of the programme Unscrambler (CAMO ASA, Trondheim, Norway).

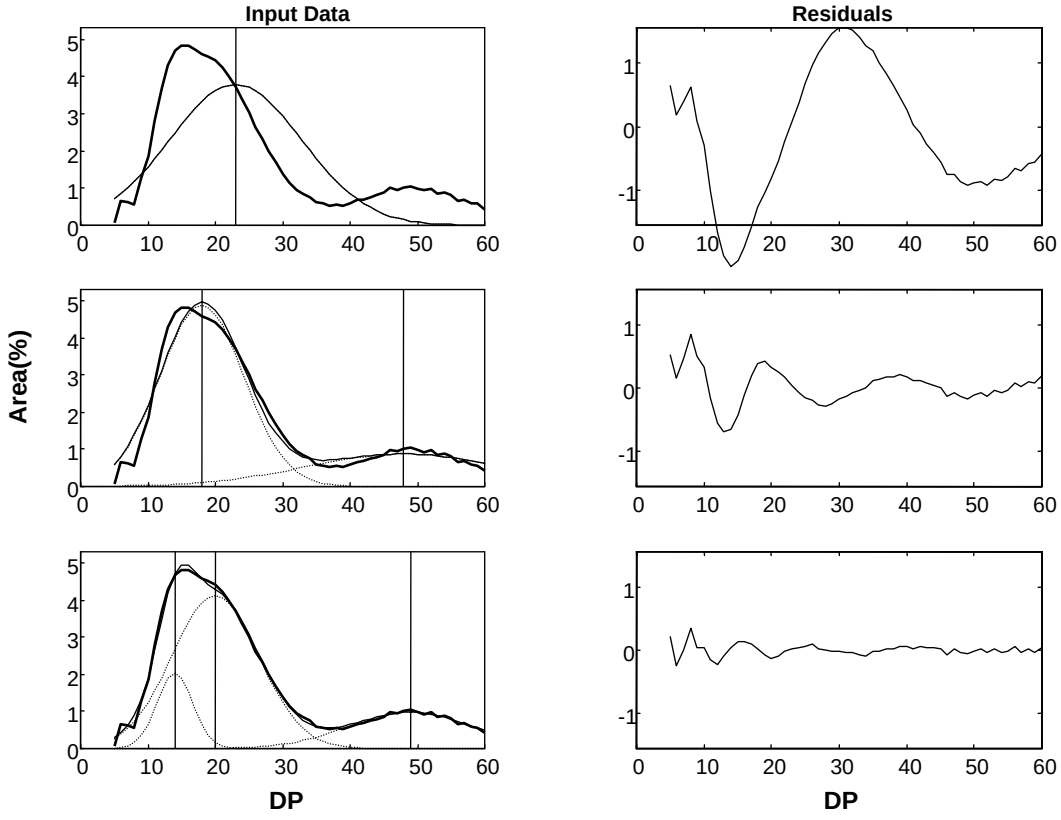

Fig. 2. Gauss deconvolution of a chain length distribution of debranched potato starch (cv. Godiva). Left: one, two and three Gaussian fits (dotted lines) were performed on the experimental data (thick lines) and their sum calculated (thin lines). Peak DP is indicated by vertical lines. Right: residual for each fit.

3. Results

3.1. HPAEC distribution of potato neutral amylopectin unit chains

Amylopectin chain length distributions of completely debranched starch were determined for 44 different starches (Table 1) using high performance anion exchange chromatography with pulsed amperometric detection. This method is superior compared to, for instance, size exclusion chromatography for analysis of moderately long α-glucan chains, as it enables quantification of single linear α-glucan chains within a sample, thus providing reliable and detailed raw data. With the Na-acetate gradient employed, chains from DP 5 to DP 60 eluted between 10 and 120 ml and baseline separation was obtained for chains from DP 5 to DP 35 as demonstrated in the original chromatogram of potato starch (c.v. Godiva) in Fig. 1(A). For chains longer than DP 35 the peaks were clearly identified and could be readily integrated in spite of weak overlaps (Fig. 1(A)). The peaks were integrated and correction was made for PAD detector response variations depending on α-glucan chain length as described earlier (Blennow et al., 1998). The correction factors are significant only for chains less than approx. DP 25 (Fig. 1(B)).

The starches analysed represent crystalline polymorphs of type-A, -B and -C and degrees of phosphorylation ranging from 0 to 63 nmol G6P/mg starch (Table 1). Starch prepared from *Curcuma zedoaria* has the highest natural degree of phosphorylation (63 nmol G6P/mg) presently known in nature. Starches prepared from tuberous tissues are highly phosphorylated whereas starches prepared from seeds have few or no covalently linked phosphate groups (Blennow et al., 1998; Jane, Kasemsuwan & Chen, 1996). Exceptions in this material are mung bean (*Phaseolus aureus*) starch, which has a relatively high degree of phosphorylation (4.5 nmol/mg starch) and the rhizome starches prepared from *Aframomum meleguata* (0.40) and *Hedychium spp.* 2 (0.73) that are very low.

3.2. Decomposition of chromatograms to Gauss bands

By convention, the distribution characteristics of amylopectin unit chains are interpreted by quantification of overlapping peaks in the polymodal chromatograms after inspection of minima and inflection points (Hizukuri, 1986). More detailed underlying structural phenomena was revealed by Ong et al. (1994) by decomposition of chain length distribution data obtained by gel permeation chromatography of debranched starches to Gauss bands. The great number of factors obtained (6–8) with varying positions and relative contributions did not simplify the

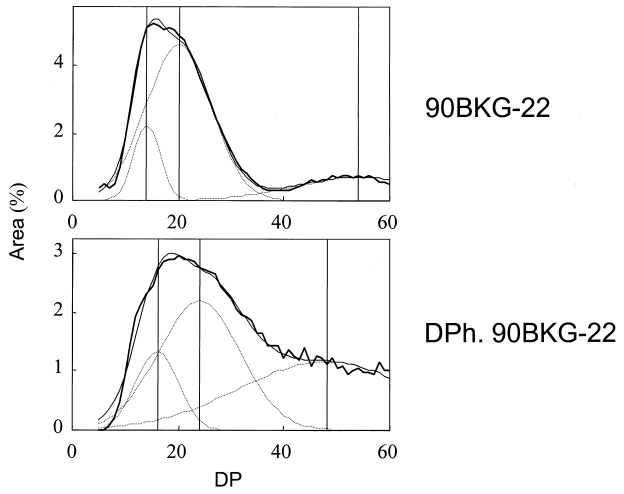

Fig. 3. Gauss deconvolution of chain length distributions of neutral oligosaccharides (top) and dephosphorylated phospho-oligosaccharides (bottom) prepared from potato starch cv 90BKG-22.

interpretation of the results with respect to general structural features of starch. In this study the chromatograms were resolved into three Gauss bands peaking in the regions DP 12–17, DP 18–24 and DP 37–53 and with LS residuals for the fits between 0.51 and 2.01 for all the 44 starches (Table 1). The relative contribution for these three bands varied between 20–67%, 41–66% and 4–16%, respectively, in the data set and the region at DP 18–42 contained on the average the most dominating factors. A typical chromatogram deconvoluted into 1, 2 and 3 Gauss bands is shown in Fig. 2. Corresponding sub-populations of chains were identified by Hizukuri (1986) in five investigated starches using gel permeation chromatography and defined as A chains (peak DP 11–16), B1 chains (peak DP 16–19) and B2 chains (peak DP 38–45), respectively.

Isolated phosphorylated chains that were enzymically dephosphorylated by alkaline phosphatase and analysed by the same method as the neutral chains showed poor relation to the neutral chains (Fig. 3). However, the major factors peaked at DP 16 and DP 24 indicating an arrangement of these chains in the amylopectin clusters similar to that of the neutral chains. As only a part of the chain population, most likely the shorter chains, could be analysed because of the inability of alkaline phosphatase to completely dephosphorylate all chains, these results should be interpreted with some caution. In contrast to an earlier study using gel permeation chromatography (Takeda & Hizukuri, 1982) where no phosphorylated chains shorter than approximately DP 20 were found in potato starch, HPAED/PAD enabled detection of chains as low as DP 10 (Fig. 3). This result indicates the existence of phospho oligosaccharides with phosphate groups attached close to or at the glucose residue at the non-reducing end as the phosphate groups are never closer than approximately nine glucose units from a branch point (Takeda & Hizukuri, 1981).

3.3. Principal component analysis

The principle of PCA is demonstrated for chain length distributions of barley (H.v. PLP1A2), potato (S.t. 90BNT-2) and shoti (*Curcuma zedoaria*) starch (Fig. 4). In general, the 1st loading, representing short chains, showed high scores in starches with low phosphate content, e.g. cereal starches. The 2nd loading with somewhat higher peak DP

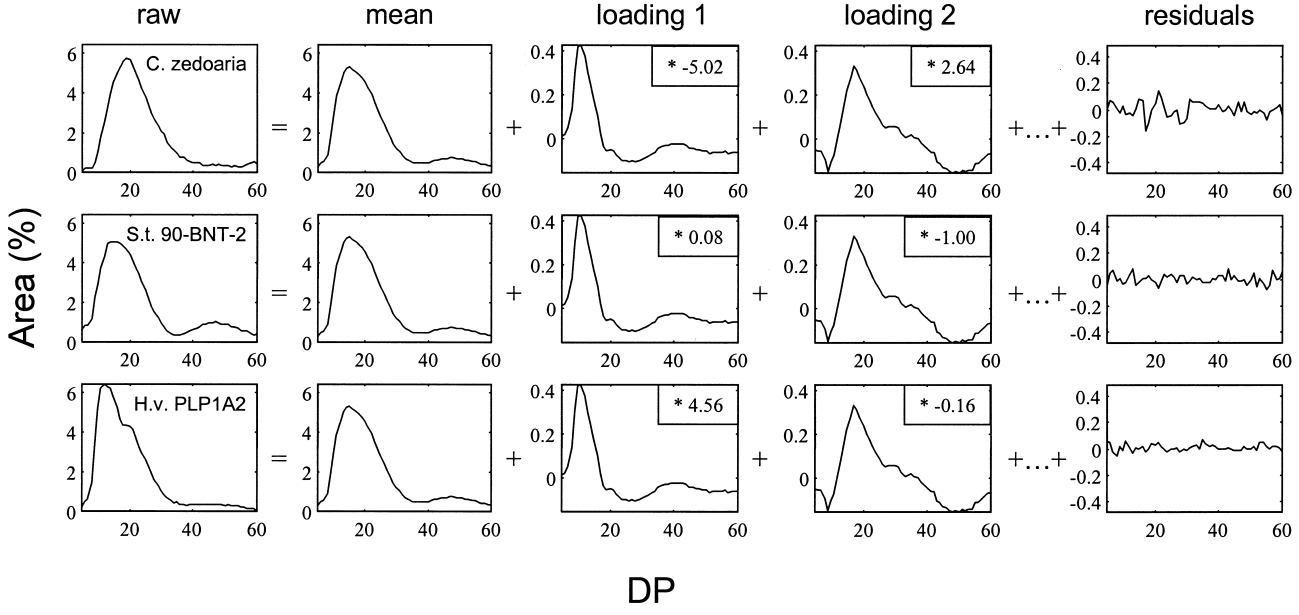

Fig. 4. Principal component analysis (PCA, 6 component solution) of neutral chain length distributions. Barley (H.v. PLP1A2), potato (S.t. 90-BNT-2) and shoti starch raw data (right) and their corresponding residuals (left) represent three differently phosphorylated starches (Table 1). The middle three columns show mean, 1st and 2nd loading for all the distributions. The scores (parameters of each loading) are indicated in boxes. PC1 and PC2 explain 76 and 11%, respectively, of the variation. PC3–PC6 contributes significantly to the model as determined by cross validation (not shown).

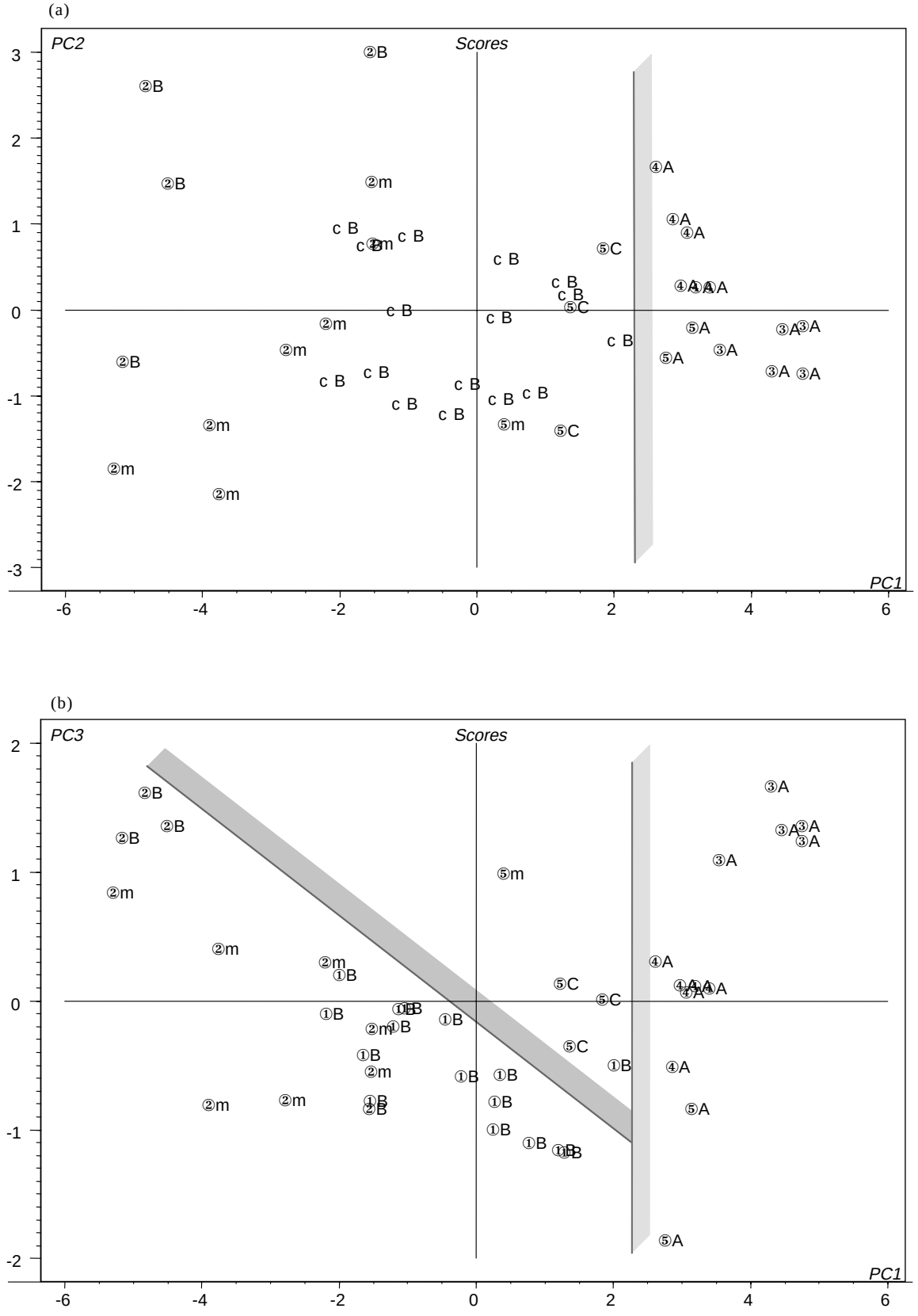

Fig. 5. PCA score plots of the whole neutral chain distribution data set described in Table 1. The starches are classified according to crystal polymorph A, B and C, respectively. For starches labelled m, information with respect to polymorph is missing. (1): potato, (2): ginger, (3): barley, (4): maize, (5): miscellaneous. Shaded bars indicate separation of polymorphs. Top: PC1 vs. PC2, bottom: PC1 vs. PC3.

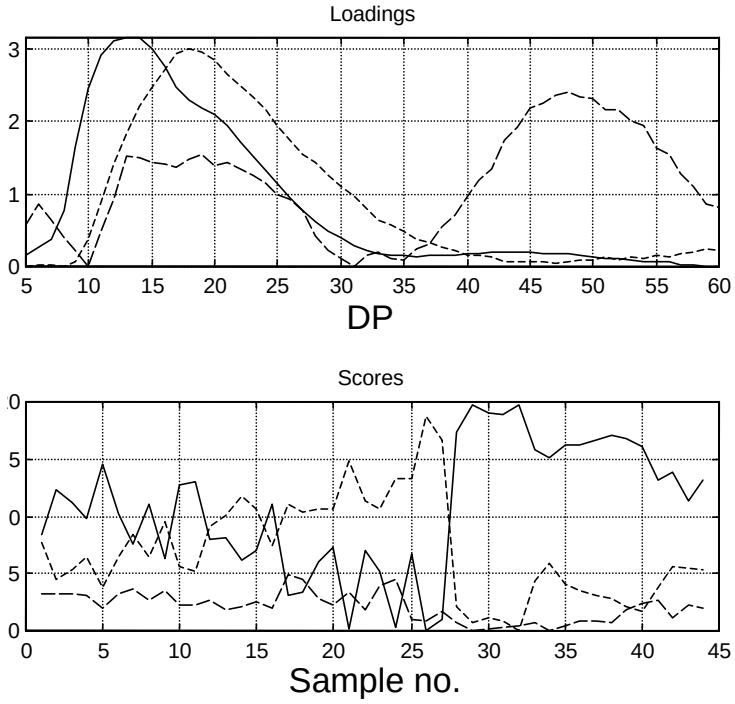

Fig. 6. Loadings and scores (in the same order as in Table 1) obtained from a three component solution of Non-Negative Alternating Regression (NNALSR) of the whole chain length distribution data set. The three loadings together describe 99.6% of the variation. Each score and its corresponding loading is indicated by identical curve type.

scored high in highly phosphorylated starches. This loading also provided a negative contribution at approximately DP 50. This effect may be caused by the absence, in the chain population, of phosphorylated chains with similar peak DP (Blennow et al., 1998). The optimal cross-validated PCA model consisted of six PCs suggesting that six underlying latent factors are present in the data material. These factors had in several cases the shape of Gaussian distributions.

A score plot of PC1 vs. PC2 demonstrates two major features of the starch (Fig. 5(a)). First, PC1 with only one exception efficiently separates starches of A-, B- and C-type crystalline polymorph. Starches with the C-type polymorph consisting of a mixture of A- and B-type crystal polymorphs are located at the border between the A- and B-type starches. For starches denoted “m”, no information regarding crystal polymorph was available. However, all these starches, except for the one labelled 5 which is a potato leaf starch, were prepared from rhizomes of plants from the ginger family and most probably have the B-type polymorph. The 1st loading, peaking at DP 10-11 (Fig. 4), i.e. rather short chains, to a high degree explained these features. As a consequence, the highly phosphorylated starches are located to the left in the diagram, while the non-phosphorylated or slightly phosphorylated starches are located to the right. This effect is probably related to the correlation between chain length characteristics and crystal polymorph as demonstrated by Hizukuri (1985). According to this study, B-type starches were enriched in a population with longer chains. These starches generally also show high degree of phosphorylation (Blennow et al., 1998).

Secondly, the botanical source had a major influence on the structural properties of the starch as revealed by grouping of the samples effectuated by the 2nd loading, peaking at DP 17 (Fig. 4). The pronounced scattering of the ginger group reflects the variation within a plant family which is expected to be larger than within species/variety. Thus, the barley, the maize and the potato groups, each representing species/variety variation, were more condensed than the ginger group.

The PC1 vs. PC3 plot (Fig. 5(b)) even more efficiently separated the different classes of starch. Only two of the starches, i.e. potato leaf (0.6,1.0) and potato S.t. 663 (2.2,−0.5) (co-ordinates in parenthesis), deviated from the general pattern. Leaf starch is daily recycled in the plant, and line #663 is a potato with very low amounts of starch branching enzyme. These features may explain the extreme behaviour of these two starches.

3.4. Non-negative alternating least squares regression

Using NNALSR three major components explained to a high degree the variation in the data (Fig. 6). The factor that peaks at DP 18 was a major constituent in the highly phosphorylated starches while the factor peaking at DP 13–14 was pronounced in starches with a low degree of phosphorylation. This phenomenon is clearly demonstrated for the extremely phosphorylated shoti starch, which was explained by 90% by the DP 18 factor. This is in agreement with Gauss deconvolution (see above) where the bands peaking at DP 18–25 were dominating the highly phosphorylated

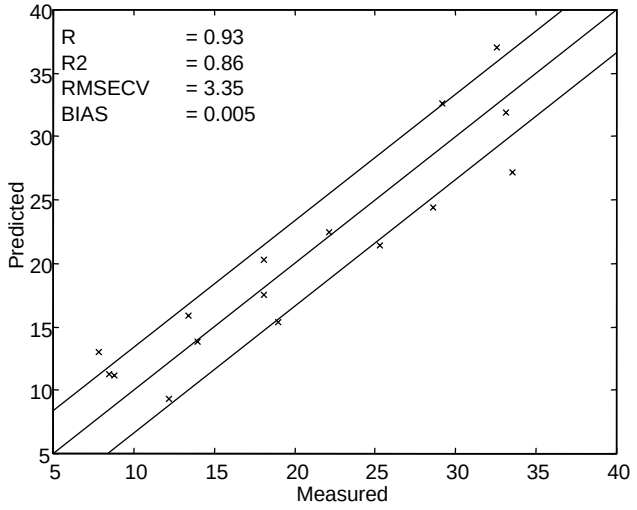

Fig. 7. Prediction of phosphorylation of potato starch from chain length distributions by Partial Least Squares regression (PLS). Offset line is indicated.

starches. It can therefore be concluded that chains of approx. DP 20 constitute a major part of amylopectin, especially in those amylopectins that are highly phosphorylated.

3.5. Prediction of starch phosphorylation

The results from Gauss decomposition, PCA and NNALSR suggest that starch phosphorylation can be predicted from distribution data. A robust technique by which predictions can be calculated from this type of data is PLS. As revealed for the potato starches, there was a strong correlation ($R = 0.93$, RMSECV = 3.35, 2 components used) between starch phosphorylation and the molecular structure (Fig. 7). A more general but much weaker correlation to phosphorylation was observed using the entire data matrix ($R = 0.71$, RMSECV = 10.98, 4 components used), indicating that additional structural factors related to botanical source (see PCA) gave significant contributions to the chain length distribution. Chains of DP 17-18 were strong determinants in this model. Individual Gauss bands failed to predict starch phosphorylation (the three bands were correlated to phosphorylation with $R = 0.14$, $R = 0.13$ and $R = 0.41$, respectively), indicating that starch phosphorylation is not easily explained by separate events resulting in normally distributed chains.

The efficiency of reconstruction of the original data by Gauss deconvolution, PCA and NNALSR using 3 components for each method is demonstrated in Fig. 8. It can be clearly seen from the LS residuals that PCA most perfectly describes the variation in the data. That Gauss and NNALSR give significantly higher residuals partly results from the imposed restrictions e.g. the non-negativity which suggests that 3 factors are not sufficient to describe the data. However, none of these methods provided more robust or unique solutions by fitting more factors, e.g. 6 factors, as the PCA solution would suggest.

4. Discussion

Previous investigations carried out on phosphorylated starches have revealed relations between the degree of phosphorylation and the unit chain length of the amylopectin i.e.

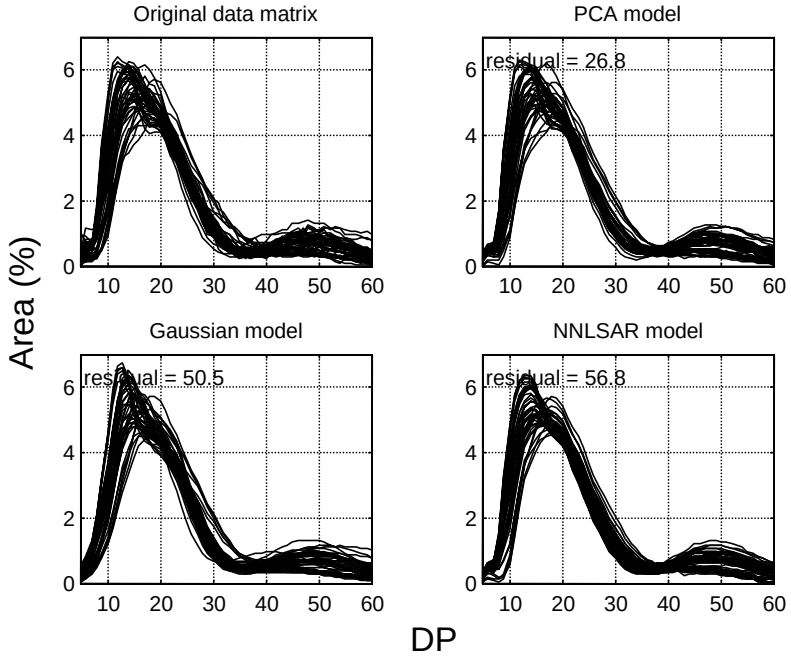

Fig. 8. The original data matrix and models created by Gauss deconvolution, 3 component solution, PCA, 3 component solution and NNALSR, 3 component solution. LS residual for each method is indicated.

highly phosphorylated starches have long chains (Blennow et al., 1998; Cooke et al., 1996; Fujimoto, Kubo, Yonemoti, Sugnuma & Nagahama, 1984). This work reveals underlying features of the amylopectin chain length data and identifies specific factors related to starch phosphorylation using chemometrics.

Variation in the dimension of the crystalline region of the cluster of 6–8 nm has been demonstrated elsewhere (Jenkins, Cameron & Donald, 1993; Jenkins & Donald, 1995). These dimensions would correspond to parallel double helices built from chains between DP 17 and 23, i.e. a range in which we found a strong positive correlation to starch phosphorylation (Blennow et al., 1998). The importance of a chain population of approx. DP 20 for the general starch structure is demonstrated by the major Gaussian band peaking at DP 19–24 found in all the starches investigated. Moreover, the 2nd PCA loading peaking at DP 17 and the 2nd NNALSR loading with peak DP of 18 manifests this chain length pool as a general determinant for starch structural variation. That the 1st PCA loading peaking at DP 10–11 determines the A-type crystal polymorph in native starch granules is of major interest as linear chains of this length readily crystallise as A-type polymorph (Gidley & Bulpin, 1987). Moreover, maize starches which are enriched in DP 10–13 chains show the A-type crystalline polymorph (Cheetham & Tao, 1998). From a conceptual point of view, it is intriguing that information about the crystalline polymorph of the starch granule can be extracted from chain length distributions of enzymically debranched starch. This fact suggests that the DP 10–11 chain population dictates the initial steps of amylopectin packing and crystallization in the granule. However, whether there is a direct relationship between chain length distribution data and crystalline dimensions of the cluster remains to be clarified.

Both PCA and NNALSR reveal that amylopectin molecular structure can be precisely explained by a few factors, a major one of these is positively correlated to starch phosphorylation. The fact that the most extremely phosphorylated starch, i.e. those from *Curcuma zedoaria*, are almost totally explained by one single factor peaking at DP 18 indicates that a maximal phosphorylation degree for amylopectin has been reached. Additional factors at high DP values may exist in, for example, high amylose starches containing long-chain amylopectin potentially resulting in even higher degrees of starch phosphorylation. However, the presence of large amounts of these long chains in starch granules seems to result in defect starch granules (Lloyd, Hedley, Bull & Ring, 1996). This is further indicated by inspection of the proposed model for amylopectin clusters (Gallant et al., 1997) as a significant proportion of chains with DP higher that approximately 25 would have the potential of breaking the constant 9-nm cluster dimension.

The prediction of a significant part of amylopectin molecular structure, i.e. its degree of phosphorylation by only few structural factors (3–6) indicates the existence of strict rules for the relationship between starch molecular structure and glucan packing. The potential physiological importance of starch phosphorylation was demonstrated in genetically engineered potato plants which have lost the ability of proper starch mobilisation caused by impaired starch phosphorylation (Lorberth et al., 1998). Thus, the balance between efficient α-glucan packing and metabolic availability in the plant is determined by the molecular structure of the starch which can be easily monitored by chain length distribution or phosphorylation. Enzyme-catalysed reactions which potentially could result in the observed correlation between chain distribution and phosphorylation could be that phosphorylated α-glucan chains are less suitable substrates for branching enzyme, stimulated elongation as shown for glycogen synthase (Lomako, Lomako, Kirkman & Whelan, 1994) or a result of altered degradation/recycling of starch (Lorberth et al., 1998). The exact relation between the factors revealed in this investigation and possible enzyme activities involved as well as relations to precise functional properties of the starch remains to be resolved.

5. Conclusions

The accelerating production of genetically modified plants, some of which having novel or even unexpected alterations in starch structure, increases the demand for new combinations of analytical systems and statistical tools for prediction of starch functionality. This investigation demonstrates a novel approach based on HPAEC/PAD combined with chemometrics to meet these demands. The results reveal a surprisingly simple system, with only a few (3–6) underlying structural motifs/effects hidden in the polymodal distribution of debranched starch. Moreover, information of the starch granule crystalline polymorph is conserved after gelatinization and enzymic debranching indicating that unit amylopectin unit chains of DP 10–11 are determining factors for starch granule formation. Finally, phosphorylation can be predicted from the polymodal chromatograms of the *neutral* (non-phosphorylated) chains suggesting that phosphate groups play an important role in the size distribution of the amylopectin side chains. The chemometric approach may be further refined to predict functional properties of starches of various botanical origins, or of starches that are genetically modified, by quantification of specific underlying factors in amylopectin chain structure.

Acknowledgements

We thank the Danish Potato Breeding Foundation and Danisco Biotechnology for providing the potato plant material. Barley lines were provided by Dr Søren Rasmussen, Plant Biology and Biogeochemistry Department, Risø National Laboratory, Denmark. The Botanical Gardens at the University of Copenhagen, Denmark, at The Royal Agricultural and Veterinary University, Denmark and at the

University of Lund, Sweden generously provided us with Zingiberaceae plant material. Biogram AB, Sweden is acknowledged for providing maize starches. The Danish Food Biotechnology Programme (Føtek II), the Danish Directorate for Development (non-food) and the Danish Biotechnology Programme financially supported this work. SBE was supported by The Danish Food Research Center (LMC) in a project concerning molecular modelling of carbohydrates.

References

Bay-Smidt, A. M., Wischmann, B., Olsen, C. E., & Nielsen, T. H. (1994). Starch bound phosphate in potato as studied by a simple method for determination of organic phosphate and ^{31}P NMR. *Starch/Stärke*, *46*, 167–172.

Blennow, A., Bay-Smidt, A. M., Wischmann, B., Olsen, C. E., & Møller, B. L. (1998). The degree of starch phosphorylation is related to the chain length distribution of the neutral and the phosphorylated chains of amylopectin. *Carbohydrate Research*, *307*, 45–54.

Blennow, A., Bay-Smidt, A. M., Olsen, C. E., & Møller, B. L. (1999). Analysis of starch bound glucose 3-phosphate and glucose 6-phosphate cusing controlled acid treatment combined with high performance anion-exchange chromatography. *Journal of Chromatography A*, *829*, 385–391.

Bro, R., & de Jong, S. (1997). A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, *11*, 393–401.

Cheetham, N. W. H., & Tao, L. (1998). Variation in crystalline type with amylose content in maize starch granules: an X-ray powder diffraction study. *Carbohydrate Polymers*, *36*, 277–284.

Cooke, D., Debet, M., Gidley, M. J., Jobling, S. A., Safford R., Sidebottom, C. M., Westcott, R. J., (1996). Patent WO 96/34968.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Analytical Chemistry*, *28*, 350–356.

Ellis, R. P., Cochrane, M. P., Dale, M. F. B., Duffus, C. M., Lynn, A., Morrison, I. M., Prentice, R. D. M., Swanston, J. S., & Tiller, S. A. (1998). Starch production and industrial use. *Journal of Food Science*, *77*, 289–311.

French, D. (1972). Fine structure of starch and its relationship to the organization of starch granules. *Journal of Japanese Society of Starch Sciences*, *19*, 8–25.

Fujimoto, S., Kubo, Y., Yonemoti, A., Suganuma, T., & Nagahama, T. (1984). Starches from Gajutsu (*Curcuma zedoaria*) Hana-Shukusha (*Hedychium coronarium*) and Hana-kanna (*Canna generalis*) (studies on starches of wild plants in Japan. Part 6). *Journal of Japanese Society of Starch Sciences*, *31*, 134–138.

Gallant, D. J., Bouchet, B., & Baldwin, P. M. (1997). Microscopy of starch: evidence of a new level of granule organization. *Carbohydrate Polymers*, *32*, 177–191.

Geladi, P., & Kowalski, B. R. (1987). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, *185*, 1–17.

Gidley, M. J., & Bulpin, P. V. (1987). Crystallisation of malto-oligosaccharides as models of the crystalline forms of starch. *Carbohydrate Polymers*, *161*, 291–300.

Hanashiro, I., Abe, J., & Hizukuri, S. (1996). A periodic distribution of the chain length of amylopectin as revealed by high-performance anion-exchange chromatography. *Carbohydrate Research*, *283*, 151–159.

Hizukuri, S. (1985). Relationship between the distribution of the chain length of amylopectin and the crystalline structure of starch granules. *Carbohydrate Research*, *141*, 295–306.

Hizukuri, S. (1986). Polymodal distribution of the chain lengths of amylopectins, and its significance. *Carbohydrate Research*, *147*, 342–347.

Hoover, R., Li, X. Y., Hynes, G., & Senanayake, N. (1997). Physicochemical characterization of mung bean starch. *Food Hydrocolloids*, *11*, 401–408.

Imberty, A., Chanzy, H., Pérez, S., Buléon, A., & Tran, V. (1988). The double-helical nature of the crystalline part of A-starch. *Journal of Molecular Biology*, *201*, 365–378.

Imberty, A., & Pérez, S. (1988). A revisit to the three-dimensional structure of B-amylose. *Biopolymers*, *27*, 1215–1221.

Imberty, A., Buléon, A., Tran, V., & Pérez, S. (1991). Recent advances in knowledge of starch structure. *Starch/Stärke*, *43*, 375–384.

Jane, J., Kasemsuwan, T., & Chen, J. F. (1996). Phosphorous in rice and other starches. *Cereal Foods World*, *41*, 827–832.

Jenkins, P. J., Cameron, R. E., & Donald, A. M. (1993). A universal feature in the structure of starch granules from different botanical scources. *Starch/Stärke*, *45*, 417–420.

Jenkins, P. J., & Donald, A. M. (1995). The influence of amylose on starch granule structure. *International Journal of Biological Macromolecules*, *17*, 315–321.

Kim, Y. S., Wiesenborn, D. P., Orr, P. H., & Grant, L. A. (1995). Screening potato starch for novel properties using differential scanning calorimetry. *Journal of Food Science*, *60*, 1060–1065.

Lawson, C. L., & Hanson, R. J. (1974). *Solving Least Squares Problems*, Englewood Cliffs, NJ: Prentice-Hall.

Lloyd, R. J., Hedley, C. L., Bull, V. J., & Ring, S. G. (1996). Determination of the effect of r and rb mutations on the structure of amylose and amylopectin in pea (*Pisum sativum L.*). *Carbohydrate Polymers*, *29*, 45–49.

Lomako, J., Lomako, W. M., Kirkman, B. R., & Whelan, W. J. (1994). The role of phosphate in muscle glycogen. *BioFactors*, *4*, 167–171.

Lorberth, R., Ritte, G., & Kossmann, J. (1998). Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nature Biotechnology*, *16*, 473–477.

Manners, D. J. (1989). Recent developments in our understanding of amylopectin structure. *Carbohydrate Polymers*, *11*, 87–112.

Martens, H., & Dardenne, P. (1998). Validation and verification of regression in small data sets. *Chemometr. Intell. Lab. Syst.*, *44*, 99–121.

Muhrbeck, P., Svensson, E., & Eliasson, A. C. (1991). Effect of the degree of phosphorylation on the crystallinity of native potato starch. *Starch/Stärke*, *43*, 466–468.

Nelder, J. A., & Mead, R. (1965). Simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Nielsen, T. H., Wischmann, B., Enevoldsen, K., & Møller, B. L. (1994). Starch phosphorylation in potato tubers proceeds concurrently with de novo biosynthesis of starch. *Plant Physiology*, *105*, 111–117.

Oates, C. G. (1997). Towards an understanding of starch granule structure and hydrolysis. *Trends in Food Science and Technology*, *8*, 375–382.

Ong, H., Jumel, K., Tokarczuk, P. F., Blanshard, J. M., & Harding, S. E. (1994). Simultaneous determinations of the molecular weight distributions of amyloses and the fine structures of amylopectins of native starches. *Carbohydrate Research*, *260*, 99–117.

Oostergetel, O. G. T., & Van Bruggen, E. F. J. (1993). The crystalline domains in potato starch granules are arranged in a helical fashion. *Carbohydrate Polymers*, *21*, 7–12.

Rasmussen, S., Hatzack, F., (1998). Identification of two low-phytate barley (*Hordeum vulgare L.*) grain mutants by TLC and genetic analysis, Hereditas, in press.

Takeda, Y., & Hizukuri, S. (1971). Studies on starch phosphate: Part 2. Isolation of glucose 3-phosphate and maltose phosphate by acid hydrolysis of potato starch. *Starch/Stärke*, *23*, 267–272.

Takeda, Y., & Hizukuri, S. (1981). Re-examination of the action of sweet-potato beta-amylase on phosphorylated (1–4)-alpha-D-glucan. *Carbohydrate Research*, *89*, 174–178.

Takeda, Y., & Hizukuri, S. (1982). Location of phosphate groups in potato amylopectin. *Carbohydrate Research*, *102*, 321–327.

Viksø-Nielsen, A., Blennow, A., Nielsen, T. H., & Møller, B. L. (1998). Phosphorylated a(1 → 4) glucans as substrate for potato starch branching enzyme I. *Plant Physiology*, *117*, 869–875.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principle component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*, 27–52.

Zobel, H. F. (1988). Molecules to granules: a comprehensive starch review. *Starch/Stärke*, *40*, 44–50.