Available online at www.sciencedirect.com
SCIENCE DIRECT
Carbohydrate Polymers 64 (2006) 433–443
Carbohydrate Polymers
www.elsevier.com/locate/carbpol

Exploratory SAXS and HPAEC-PAD studies of starches from diverse plant genotypes

James S. Sanderson [a], Rob D. Daniels [a,1], Athene M. Donald [a,*], Andreas Blennow [b], Søren B. Engelsen [c]

[a] *Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK*
[b] *Department of Plant Biology, Center for Molecular Plant Physiology, The Royal Veterinary and Agricultural University, 40 Thorvaldsensvej, DK-1871 Frederiksberg C, Copenhagen, Denmark*
[c] *Centre for Advanced Food Studies (LMC), Rolighedsvej 30, DK-1958 Frederiksberg C, Copenhagen, Denmark*

Received 12 August 2005; received in revised form 10 December 2005; accepted 13 December 2005
Available online 2 February 2006

Abstract

A substantial amount of effort has been expended on developing a suitable model to describe the small angle X-ray scattering (SAXS) profiles of starch. The modelling work now offers good descriptions of the fine structure of the starch granule, with recent advances [Daniels, D.R., & Donald, A.M. (2003). An improved model for analysing the SAXS of starch granules. Biopolymers, 69, 165–175] having substantially improved the modelling of B-type crystal-structure starches. However, up till now physical evidence obtained experimentally to support the modelling results has been difficult to obtain. The work presented here shows experimental evidence in support of the hypotheses deduced from the modelling process. This support comes from results of work combining modelling of the SAXS profiles of a diverse range of starches with the physical characterisation technique of high performance anion exchange chromatography using pulsed amperometric detection (HPAEC-PAD). HPAEC-PAD has been used to characterise the chain-length profile of amylopectin for degrees of polymerisation (DP) 6–60 glucose units. This provides, for the first time, physical information of the fine structure of starch that can be compared directly with the results obtained from modelling the SAXS profile. To assist the identification of trends between the variables determined from SAXS profile modelling and those measured by HPAEC-PAD, a data mining approach, principal component analysis (PCA), was employed.

Keywords: Starch; SAXS; Principle components analysis; High performance anion exchange chromatography; Botanical diversity

1. Introduction

Starch is principally composed of two polymers of glucose: amylose and amylopectin, in addition to a number of other minor components. The amylopectin is an extremely large branched molecule, but there is extensive order of the fine structure of amylopectin molecules in the native starch granule and it is believed to adopt an efficiently packed, cluster structure (Robin, Mercier, Duprat, Charbonniere, & Guilbot, 1975). The cluster structure is dependent on many factors such as the distribution of chain-lengths and branch points. Indeed, the branch points are known to be non-random (Thompson, 2000) and clustering of branch points is observed in regions of low molecular order. Within the cluster structure, three specific types of chains can be identified, termed A, B and C (Peat, Whelan, & Thomas, 1954). The structure within the regions of low and high molecular order will be dependent on the distribution of branch points and chain-lengths, which in turn will give the length of the double helices. Recent work (Waigh et al., 2000) has described this system in terms of a liquid crystalline model, with the double helices identified as mesogen units. A mesogen is characterised as a solid rod-like molecule that can align to give liquid crystalline behaviour.

The A and short B chains can form double helices, which can crystallise into different crystal types. There are two distinct types, A-type and B-type (C-type starches such as legumes are now thought to be a mixture of A- and B-type crystal structures) (Buleon, Gerard, Riekel, Vuong, & Chanzy, 1998). It is now accepted that the length and ratio of the A and B chains of amylopectin are responsible for which crystal structure type is adopted (Hanashiro, Abe, & Hizukuri, 1996;

* Corresponding author. Tel.: +44 1223 337382; fax: +44 1223 337000.
E-mail address: amd3@cam.ac.uk (A.M. Donald).
[1] Present address: Department of Physics, University of Warwick, Coventry CV4 7A, UK.

0144-8617/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carbpol.2005.12.026

Hizukuri, 1985, 1986). This is because the crystal structures can only be formed by the packing of double helices and therefore the A and B chain-lengths involved in double helical formation must be of prime importance. The length between the branch point and the start of the double helix is termed the flexible spacer (Waigh et al., 2000), and is also of importance in determining the degree of crystallinity as this affects the degree of association allowed between double helices. Model studies of the branching geometry indicate that a short spacer is facilitated by aqueous hydration (Corzana et al., 2004). The size and ratio of A and B chains varies dramatically between genotypes and this influences the crystal structure type, with the A-type crystal structure commonly found in cereal species, and the B-type crystal structure generally found in tubers.

At a larger length scale, the cluster model of alternating more and less ordered regions stacks up to form a semi-crystalline ring. These semi-crystalline rings alternate with amorphous regions to form a periodic variation of organisation through the granule known as growth rings. The origin of these rings is still the subject of debate (Pilling, 2001). Fig. 1 shows the typical representation of a starch granule from the length scale of the granule to the length scale of the individual amylopectin chains. Some researchers have used AFM to identify so-called 'blocklets' in starches of different types, but have failed to demonstrate conclusively how these tie in with the other structures revealed by SAXS. It is thus difficult to relate this approach using blocklets to analysis based on (Gallant, Bouchet, & Baldwin, 1997; Ridout, Parker, Hedley, Bogracheva, & Morris, 2003, 2004) SAXS, and further discussion within this framework will not be attempted here.

The role of amylose in starch granules remains unclear. Some researchers have reported relationships between amylose content and amylopectin size and chain-length properties; however, much still remains unknown (Cheetham & Tao, 1997). There may be some relation through the activity of the biosynthetic enzymes responsible for extending and debranching polymer chains.

The sizes of the repeating lamellar structures, both of the amorphous growth rings and within the smaller semi-crystalline growth ring can be determined by scattering studies. The SAXS profile from hydrated starch granules shows a single peak, but the whole scattering pattern can be used to provide a set of parameters describing the granule structure (Cameron & Donald, 1992). Despite the apparently simple SAXS profile, modelling it has proved extremely difficult in light of the complexity of the fine structure of amylopectin (Donald, Kato, Perry, & Waigh, 2001). Particularly for B-type starches, it becomes important to take into account the fact that locally the lamellae may fluctuate, and a refined model to accommodate such fluctuations has been developed (Daniels & Donald, 2003, 2004). This refined model was originally used to investigate SAXS profiles from four starch species: maize, potato, tapioca and waxy maize. The same parameters are used as in the original model of Cameron & Donald (1992), but now additionally allowing the plane of the semi-crystalline lamellae to fluctuate. The original key parameters were the two electron density differences between pairs of the three types of regions: crystalline lamellae, amorphous lamellae and amorphous growth rings; the proportion of material involved in the lamellae which is crystalline, the lamellar repeat distance, the number of lamellae involved in each semi-crystalline stack and a parameter defining the distribution of lamellar sizes, β. In the refined model, the crystalline lamellae are treated as equivalent to smectic layers, which can fluctuate about a mean position. Such fluctuations introduce bending of the layers and splay of the double helical units, which are treated as mesogens (Daniels & Donald, 2003, 2004). β is reinterpreted not only as an indicator of the distribution of lamellae layer thicknesses, but also as an indicator of lamellar compressibility. New parameters ε, a measure of layer bending, and δ, related to splay of the mesogen units, are additionally introduced. By considering what must happen when layers of double helices are no longer flat, it can be seen that it is reasonable to assume a relationship between δ and ε such that if the mesogens splay from an ordered alignment, i.e. a high δ, then the lamellae

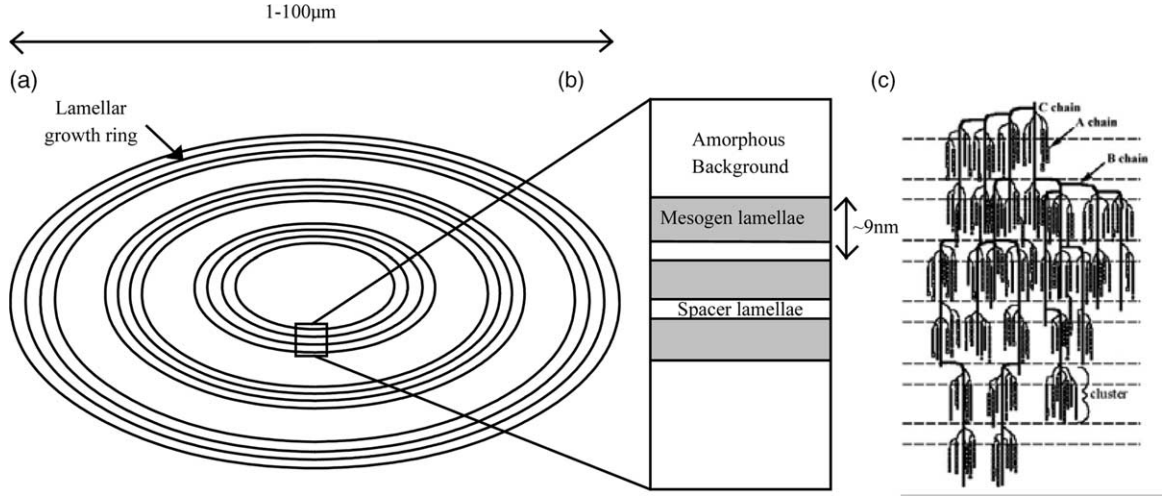

Fig. 1. Schematic structure of a starch granule, showing the arrangement of the amylopectin polymer within semi-crystalline growth ring within the starch granule at different length scales.

layers may need to bend to accommodate this, yielding a finite ε. From the data of the four starches previously investigated, a hypothesis was proposed that there was a relationship between β and ε, such that a high β value (high degree of layer compressibility) was observed concurrently with a low value of ε (layer bending) (Daniels & Donald, 2003). In other words, if layer bending was not observed, this suggested the lamellae structure was compressible.

This correlation is a reasonable relationship to suggest when one considers the fine structure within the lamella system. Longer mesogen units are likely to be more stable and will tolerate less decoupling from the backbone. In this situation, there would be little room for compression across the lamellae structure and therefore overall layer bending needs to occur to accommodate any internal stresses. The alternative situation of small mesogen units with long flexible spacers will decouple the amylopectin backbone from the double helices effectively. Here compression could be accommodated within the lamellae structure, due to the long flexible spacers. This would support the inverse relationship of β and ε, i.e. high compressibility correlates to low layer bending, and this would allow the approximate 9 nm repeat within the lamellar structure, found in all starches so far investigated (Donald et al., 2001) to be maintained.

With the recent advances in chromatographic separation methods coupled to powerful detection capabilities of systems such as pulsed amperometry, studies of carbohydrate mixtures, for example the determination of chain-length profiles of amylopectin, have become routine. With appropriate systems to cover a detection range of DP 6–60, the fine structure of amylopectin can be analysed and correlated with the approximate length scale of the lamellae regions investigated by SAXS. Therefore, such analysis provides the opportunity to obtain physical evidence for the model derived from SAXS data. To enable the identification of correlations between amylopectin chain-length properties and SAXS parameters, the multivariate data analytical method of Principal Component Analysis (PCA) can be utilised. PCA is a bi-linear mathematical technique, which enables reduction of the original data matrix into latent so-called 'principal components' (PC) that describes the data (Anderson, 1984). This approach facilitates identification of relationships between variables and the parameters that have been measured to generate the data matrix, in this case relationships between the parameters of the SAXS model and the properties of the chain-lengths of the amylopectin molecules. The PCs are plotted graphically into score and loading plots to allow visual identification of relationships.

The work presented here is a study based on 27 different starches from a range of plant genotypes. Included are traditional examples such as maize and potato, as well more unusual examples, such as mung bean and water chestnut starches. In addition, a number of transgenic potato starches and chemically modified maize starches are included.

2. Experimental

This paper presents the results of a study of 27 starch samples of various botanical species, including examples of C-type crystal structure starches, in addition to examples of chemically and genetically modified starches.

The starches used in this study were either obtained pre-extracted or extracted from the storage organ, i.e. tuber, cereal, or legume seed. Details of the starches used and the extraction method where appropriate are given in Table 1. In the case of extraction from the storage organs, a standard extraction method was adopted as described (Section 2.1). Some pre-extracted starches were purchased from commercial suppliers and information concerning the extraction technique is unavailable (noted as N/A, not available in Table 1). All extraction methods are regarded as mild and the data obtained was not dependent on the extraction method used.

2.1. Extraction method '1'

Storage organs were soaked in twice the weight of starchy material of 0.25% NaOH at 4 °C for 24 h. For sweet potato, the tubers were washed and peeled and cut into one-centimetre cubed pieces, prior to soaking. After the soaking period the material was repeatedly washed with fresh water. The washed material was blitzed in a food processor and passed through a series of sieves with pores of reducing size from 350 to 70 μm. The starchy material that passed through the sieve was washed several times with water. Centrifugal force ($1000\times g$, 10 min) enabled separation of the starch from the debris top layer, which was mechanically removed. The material was re-suspended in clean water and this process was repeated until no debris layer was observed. At this stage, the starch material was suspended in acetone and dried overnight at room temperature before being stored in airtight containers ready for characterisation.

2.2. Extraction method '2'

The following method is described in detail in Bay-Smidt, Wischmann, Olsen, and Nielsen (1994). Three rods from within the tuber were punched out of the perimedulla with a 5 mm cork borer and sliced into 3 mm thick discs and immediately frozen in liquid nitrogen. Six discs (approximately 350 mg) were homogenised in 2 ml ice-cold water and the homogenate was filtered through two layers of gauze and the filtrate washed successively with 20 ml H_2O, 20 ml 80% ethanol and 20 ml H_2O. After each wash, the starch granules were collected by centrifugation ($4000\times g$, 10 min).

2.3. Small angle X-ray scattering studies (SAXS)

The SAXS experiments were carried out on station 8.2, Synchrotron Radiation Source, Daresbury, Cheshire, UK. Starch material was hydrated with pure water to 40% starch/water by weight prior the SAXS experiment, and suspended within sample cells designed in-house (Perry,

1999). The high-intensity beam was collimated to a spot size of approximately 3 mm × 2 mm, with the SAXS profiles being recorded on a quadrant detector. This experimental set-up has extensively been used in other studies and further details are reported elsewhere (Perry, 1999).

2.4. Modelling the SAXS profiles

Previous investigators have established the suitable modelling methods for SAXS patterns obtained from SAXS experiments under the conditions employed in these experiments (Cameron & Donald, 1992), and the mathematical and physical explanation for the SAXS model are described above and in more detail elsewhere (Daniels & Donald, 2003). Table 2 identifies the relevant parameters. It has previously been shown that cross correlations involving the parameters of N, $\Delta\rho$ and $\Delta\rho_u$ can be ignored, as the model is not very sensitive to these parameters and values of $\Delta\rho$ and $\Delta\rho_u$ are relative rather than absolute. The value of N has little impact on the best fit observed for the starches, and indeed in the original data (not shown) the best fit to the four starches had only a slight variation in the value of N.

The SAXS profiles from all 27 starches were fitted with the model described in (Daniels & Donald, 2003); the values of the parameters obtained are given in Table 3. Prior to modelling, all starches were scaled to equal intensity at high q ($q = 0.2$ Å^{-1}) to account for variations in sample concentration. A typical example of a best fit to the data is shown in Fig. 2.

Table 1
Starches used with extraction method identified (GMP, genetically modified potato; N/A, not available)

Starch identity	Latin name	Extraction method	Source	Comments
926.28.3	*Solanum tuberosum*	2	KVL, Copenhagen, Denmark (Vikso-Nielsen et al., 2001)	Tuber. GMP anti-sense GWD (high amylose, low phosphate)
Black eye bean	*Vigna unguiculata*	1	TRS Wholesale Co. Ltd, Southhall, Middlesex, UK	Seed
Buckwheat	*Eriogonum microthecum*	1	Arjuna Wholefoods, Cambridge, UK	Grain
Chickpea	*Cicer arietinum*	1	TRS Wholesale Co. Ltd, Southhall, Middlesex, UK	Seed
Colflo67	*Zea mays*	N/A	DuPont Chemical Research Centre, Cambridge, UK (Gift)	Grain. Chemically modified maize, organic ester linkages
Dry foul bean	*Phaseolus vulgaris*	1	Cypressa, Cypressa House, 38 Drayton Park, London, UK	Seed
FridgeX	*Zea mays*	N/A	DuPont Chemical Research Centre, Cambridge, UK (Gift)	Grain. Chemically modified maize, phosphate cross-linked
Ginger starch	*Curcuma zedoaria*	2	Dr K. Yamamoto, Hokuren Federation Agricultural Cooperatives, Japan)	Root (C. zedoaria)
H943.31.1	*Solanum tuberosum*	N/A	Danisco, Denmark	Tuber. GMP anti-sense SBE I+II, high phosphate and amylose
H944.16.1	*Solanum tuberosum*	N/A	Danisco, Denmark	Tuber. GMP anti-sense SBE I+II, high phosphate and amylose
KMC	*Solanum tuberosum*	N/A	KMC, Denmark	Tuber. Bulk potato starch
Maize	*Zea mays*	N/A	Sigma Chemicals, P.O. Box 14508, St Louis, USA	Grain
Millet	*Panicum miliaceum*	1	Arjuna Wholefoods, Cambridge, UK	Grain
Mung bean	*Vigna radiata*	1	Sitthinan Co. Ltd, Sathorn Thani Building, Bangkok, Thailand	Seed
Pinto	*Phaseolus vulgaris*	1	Cypressa, Cypressa House, 38 Drayton Park, London, UK	Seed
Potato	*Solanum tuberosum*	N/A	BDH Lab Suppliers, Poole, Dorset, UK	Tuber
Quinoa	*Chenopodium quinoa*	1	www.quinoa.com	Grain
Rice	*Oryza sativa*	N/A	Sigma Chemicals, P.O. Box 14508, St Louis, USA	Tuberous root
Sago	*Cycas circinalis*	N/A	DuPont Chemical Research Centre, Cambridge, UK (Gift)	From trunk
Sweet potato	*Ipomoea batatas*	1	Cambridge Market, Cambridge UK	Tuber
Tapioca cam	*Manihot*	N/A	Thye Huat Chan Sdn Bhd, 368-3-12A, Bellisa Row, Malaysia	Root
Tapioca ku	*Manihot*	N/A	Foodex Co. Ltd, 3/164–165, Soi Rachada, Bangkok, Thailand	Root
Urid dal	*Phaseolus vulgaris*	1	TRS Wholesale Co. Ltd, Southhall, Middlesex, UK	Seed
Water chestnut	*Trapa natans*	N/A	China Int. Native Produce & Animal By-products, China	Corm
Waxy maize	*Zea mays*	N/A	Sigma Chemicals, P.O.Box 14508, St Louis, USA	Grain
Amylopectin potato	*Solanum tuberosum*	NA	Lyckeby Stärkelsen, Sweden	High amylopectin tuber
White kidney bean	*Phaseolus vulgaris*	1	Cypressa, Cypressa House, 38 Drayton Park, London, UK	Seed

Table 2
Identification and explanation of the parameters of the improved SAXS model

Parameter	Explanation
β	An indication of the distribution of lamellae layer thicknesses
d	The combined distance of one crystalline and one amorphous region within the semi-crystalline growth ring (units = Å)
Φ	An indication of the relative lengths of the rigid mesogen units
N	The number of repeats within a semi-crystalline growth ring
$\Delta\rho$	The difference in the scattering density of the crystalline and amorphous lamellae within the semi-crystalline growth ring
$\Delta\rho_u$	The difference in the scattering density between the amorphous material and the background material
ε	An indication of layer bending, i.e. an indication of how the lamellae buckle
δ	An indication of splay distortions of the mesogen units away from the preferred orientation direction

2.5. Amylopectin chain-length distribution analysed by high performance anion exchange chromatography-pulsed amperometric detection (HPAEC-PAD)

The method followed in these experiments is based on previous work by Blennow, Engelsen, Munck, and Moller (2000). Five milligrams of starch was gelatinised in 1 ml H_2O at 95 °C for 3 min and then allowed to cool to 25 °C before 4 μl isoamylase (350 U/ml, Megazyme, Ireland) was added allowing for complete specific debranching of the α-1,6-linkages (debranching) in the amylopectin. After incubation at 40 °C for 2 h the sample was centrifuged and 200 μl of the supernatant injected into the HPAEC-PAD system.

The HPAEC-PAD system used was a DIONEX DX500 system (Dionex Corporation, USA), equipped with an S-3500 auto sampler, a GP40 pump and an ED40 PAD system. The column was a 300 mm CarboPac PA100 with a PA100 guard column. Amylose had an extremely strong affinity for the column and did not interfere with the determination of the amylopectin chain-length profile. The PAD system was calibrated with a number of standards of known DP as the detector response was not linear for short chain-lengths (Blennow, Bay-Smidt, Wischmann, Olsen, & Moller, 1998). Chain-lengths between DP 6 and 60 were recorded and analysed.

HPAEC-PAD analysis was carried out on all 27 starches. The system was optimised to assess the chain-length distribution of amylopectin between DP 6 and 60. A typical chain-length profile is shown in Fig. 3.

The outputs from the HPAEC-PAD system were normalised to allow quantification of the chain-length populations between DP 6 and 60. This was achieved by defining a baseline for each sample and calculating each peak area as a percentage of the total area from the detector response to the baseline. Therefore,

Table 3
Data for five of the parameters obtained from fitting the SAXS profiles of all starches with the improved model

Starch	β	d (nm)	Φ	ε	δ
Buckwheat (A)	0.27	9.3	0.735	0.01	0.33
Colflo67 (A)	0.38	9.0	0.71	0.01	0.01
FridgeX (A)	0.40	8.9	0.7	0.01	0.01
Maize (A)	0.33	9.4	0.7	0.1	0.33
Millet (A)	0.26	9.3	0.75	0.01	0.33
Quinoa (A)	0.35	9.0	0.68	1×10^{-6}	0.1
Rice (A)	0.28	8.6	0.65	0.2	0.5
Sago (A)	0.225	9.2	0.66	0.4	0.5
Sweet potato (A)	0.275	9.7	0.745	0.01	0.33
Tapioca cam (A)	0.27	9.6	0.765	0.01	0.2
Tapioca ku (A)	0.285	9.6	0.7	0.01	0.2
Waxy maize (A)	0.37	9.0	0.72	0.01	0.01
Water chestnut (A)	0.29	9.8	0.79	0.1	0.2
Ginger starch (B)	0.3	9.1	0.76	0.3	0.29
H943.31.1 (B)	0.32	9.4	0.73	0.2	0.33
H944.16.1 (B)	0.4	9.2	0.68	0.2	0.2
KMC (B)	0.25	9.4	0.83	0.8	0.25
Potato (B)	0.21	9.3	0.78	0.4	0.33
Waxy potato (B)	0.22	9.1	0.77	0.5	0.4
926.28.3 (B)	0.30	9.1	0.72	0.6	0.14
Black eye bean (C)	0.28	9.8	0.74	0.2	0.5
Chickpea (C)	0.265	9.9	0.79	0.2	0.5
Dry foul bean (C)	0.33	9.4	0.735	0.8	0.4
Mung bean (C)	0.39	9.3	0.72	0.01	0.07
Pinto (C)	0.305	9.5	0.735	0.2	0.5
Urid Dal (C)	0.275	9.3	0.72	0.3	0.67
White kidney bean (C)	0.33	9.5	0.70	0.1	0.33

Note. d spacing value is corrected for effect of additional two parameters used in the improved model, as described previously (Daniels & Donald, 2003).

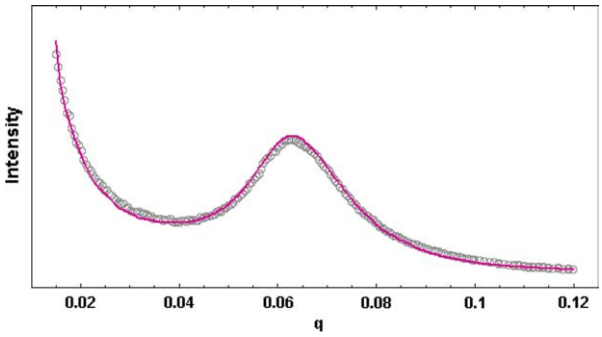

Fig. 2. A typical SAXS pattern (raw data as open circles) with the fit from the SAXS model (solid line). This example is buckwheat starch.

a population distribution for each starch for each chain-length of amylopectin between DP 6 and 60 was obtained, i.e. 54 chain-length populations per starch. However, to aid identification of trends, the chain-lengths were grouped into sets of five DP, i.e. DP 6–10, DP 11–15, etc. This data is shown in Table 4, and used in the Principal Component Analysis statistics described in Section 3.

2.6. Principal component analysis (PCA)

One of the main advantages of applying multivariate chemometric data analysis to large data sets is the possibility of undertaking an explorative inductive investigation (Munck, Nørgarrd, Engelsen, Bro, & Andersson, 1998). Principal Component Analysis (PCA) is a most useful tool for analysis of large data sets, with the aim of identifying underlying relationships that may not be directly obvious (Anderson, 1984). In PCA, the data matrix is decomposed by consecutive orthogonal extraction of the largest variation (principal components, PCs) in data until the variation left is unsystematic (noise). The PCA model can be written as: $\mathbf{X}=\mathbf{X}+\mathbf{TP}^{\mathrm{T}}+\mathbf{E}$, where $\mathbf{P}$ is the loading matrix and $\mathbf{T}$ is the score matrix. The loading vectors can be considered as pure hidden profiles that are common to all the measured samples. The scores are the amounts of each hidden profiles that makes up the individual samples. In this study, PCA was employed to identify correlations between the SAXS parameters and the HPAEC-PAD data for the amylopectin chain-length data. Only full cross-validated models are reported.

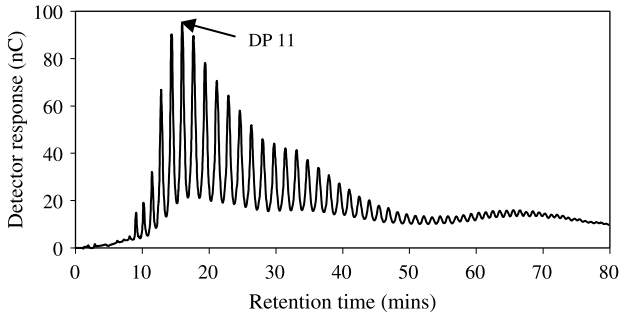

Fig. 3. Typical debranched amylopectin chain-length profile (buckwheat starch).

Table 4
Percentage of amylopectin chain-lengths between DP 6–15, DP 16–24 and DP 25–60, from the chain-length data between DP 6 and 60 for all 27 starches, grouped into crystal-type

Starch	DP 6–15	DP 16–24	DP 25–60
Buckwheat (A)	32	35	33
Colflo67 (A)	33	33	34
FridgeX (A)	50	26	24
Maize (A)	30	38	32
Millet (A)	30	39	31
Quinoa (A)	35	35	30
Rice (A)	31	38	31
Sago (A)	29	41	30
Sweet potato (A)	28	36	36
Tapioca cam (A)	29	34	37
Tapioca ku (A)	30	34	36
Waxy maize (A)	32	36	32
Water chestnut (A)	30	37	33
Ginger starch (B)	18	40	42
H943.31.1 (B)	18	37	45
H944.16.1 (B)	14	38	48
KMC (B)	24	35	41
Potato (B)	25	37	48
Waxy potato (B)	23	36	41
926.28.3 (B)	23	36	41
Black eye bean (C)	26	35	36
Chickpea (C)	26	39	35
Dry foul bean (C)	26	40	34
Mung bean (C)	27	38	35
Pinto (C)	36	40	34
Urid dal (C)	24	40	36
White kidney bean (C)	26	40	34

3. Results and discussion

The first set of correlations examined, were to confirm the anticipated correlation between chain-length, from HPAEC-PAD data, and crystal type. Fig. 4 shows that a scores plot, based on the percentage of each chain-length from DP 6 to 60, clearly differentiates the three crystal

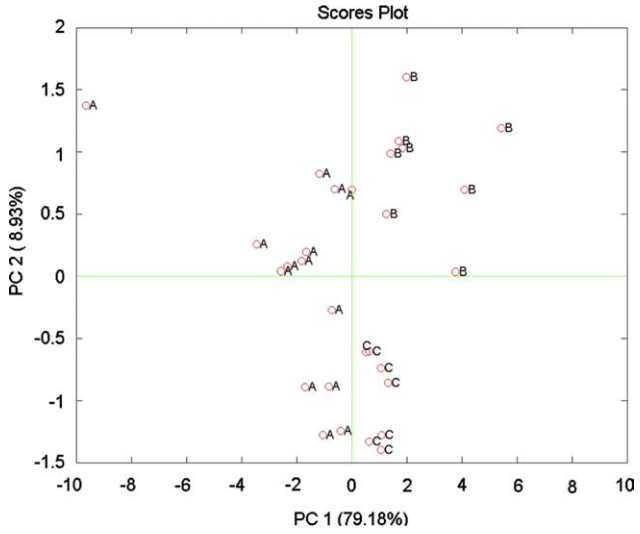

Fig. 4. PCA plot of amylopectin chain-length profile (DP 6–60) for all starches. Letters correspond to starch crystal-type.

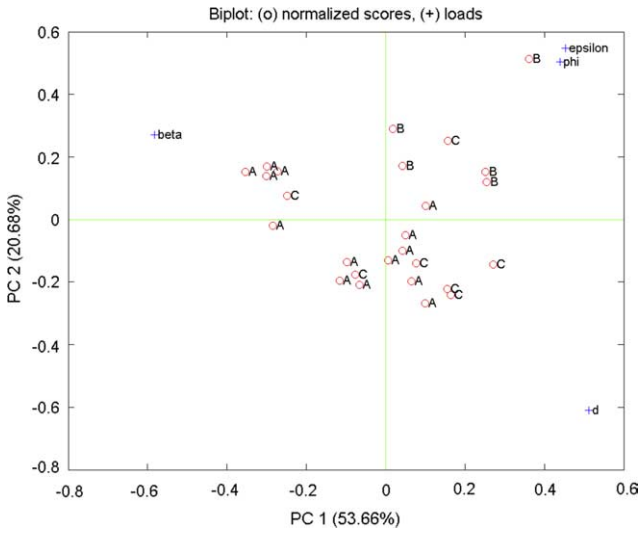

Fig. 5. Bi-plot PCA plot of PC1 vs. PC2 of SAXS model parameters β, d, Φ and ε. Letters corresponds to starch crystal-type.

types. This confirms the validity of this approach, substantiating the expectation that chain-length distribution can be correlated with crystal type (Hanashiro et al., 1996; Hizukuri, 1985, 1986). Next, an attempt was made to carry out a bi-plot analysis based on the SAXS data, using the key variables β, d, Φ, and ε (Fig. 5). This figure clearly separates the A and B type starches, whilst the C type mainly cluster with the A type (in this figure, the genetically modified potato starches H943.31.1. and H944.16.1 are omitted; if these are included the figure becomes much less clear due to their highly unusual SAXS profiles which were hard to fit accurately). The exception to this rule is dry foul bean, which is C-type but sits amidst the B-type starches. The bi-plot allows identification of the principal effects of the four variables β, d, Φ, and ε and the correlations between them. Most importantly, β and d plot opposite one another, suggesting a negative correlation. This is unsurprising since a low β implies the structure is relatively incompressible, which would be expected to lead to a larger d and conversely. This inverse correlation can be seen more clearly in Fig. 6a, which shows a correlation which is statistically significant at the $p<0.01$ level[2].

To enable identification of hidden relationships between the HPAEC-PAD data and the SAXS parameters, Principal Component Analysis (PCA) was employed on the combined information from SAXS parameters and the amylopectin chain-length distributions, using the grouped amylopectin chain-length populations (i.e. DP 6–10, DP 11–15, DP 16–20, etc.). The PCA results for this analysis are shown in the bi-plot (combined score and loading plot) of the complete data matrix is shown in Fig. 7 which shows a clear separation amongst the three starch types A, B and C. The biplot, which includes more than 65% of the variation in the total data set, also suggests that the C-types is more closely related to the

[2] Note. Statistical analysis for correlations from fisher analysis, details in [24].

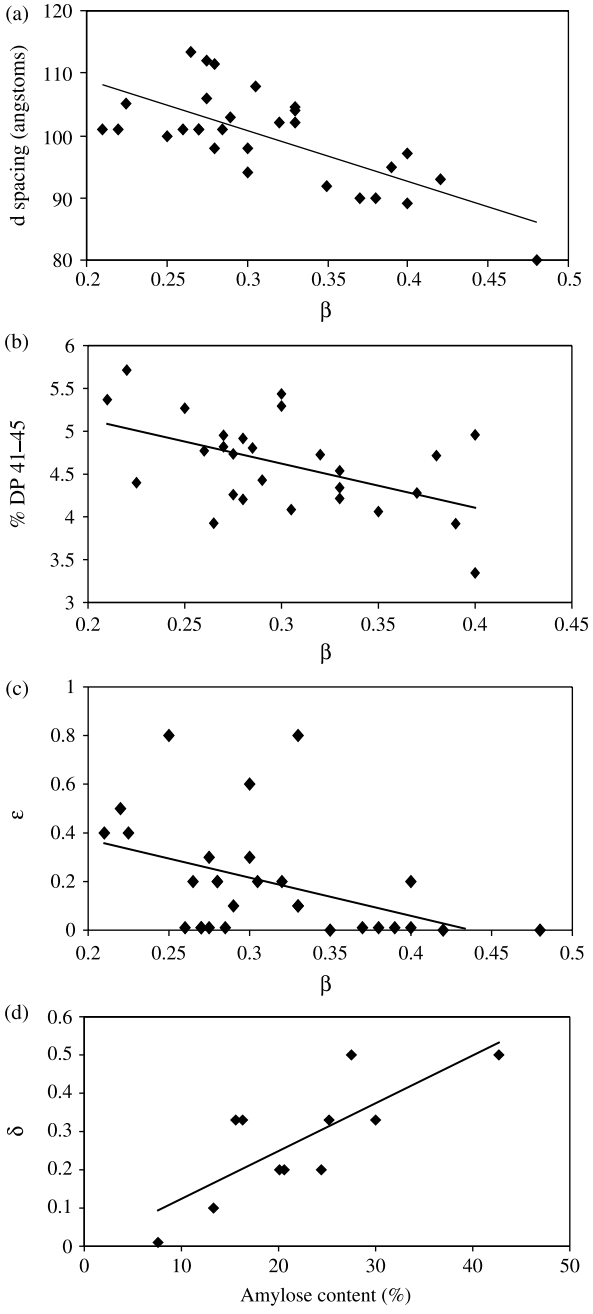

Fig. 6. (a) Test for negative correlation between β and d, suggested from Section 5.2 and from the PCA bi-plot in Fig. 7 (linear regression $R^2=0.50$, statistically significant at the $p<0.01$ level); (b) Positive correlation observed between amylose content and δ from 11 native A-type crystal structure starches (linear regression $R^2=0.61$, statistically significant at the $p<0.05$ level); (c) Negative correlation observed between β and ε SAXS parameters from the 27 starches investigated, data from Table 3 (linear regression $R^2=0.19$, statistically significant at the $p<0.05$ level); (d) Positive correlation observed between amylose content and δ from 11 native A-type crystal structure starches (linear regression $R^2=0.61$, statistically significant at the $p<0.05$ level).

A-type starches, than B, as was also suggested in Fig. 4. The variables of both ε and Φ plot in the top right hand corner of Fig. 7, located in the region of the long chain-length values and the B-type crystal structure starches, suggesting correlations between these parameters. From previous work (Daniels and Donald, 2003) it was expected that ε should correlate with

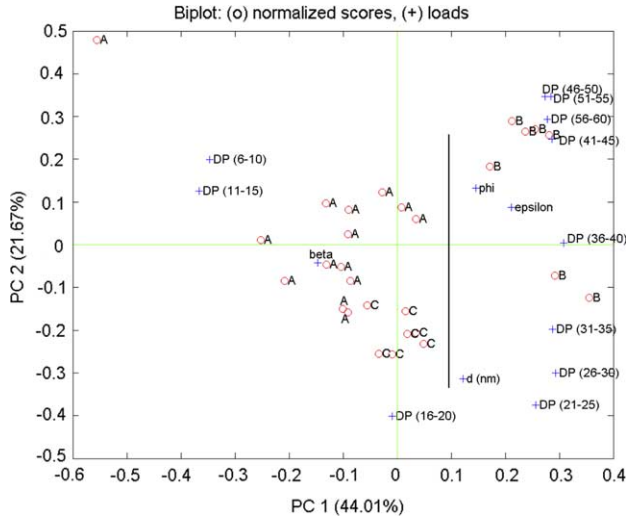

Fig. 7. PCA bi-plot of modelled SAXS parameters and grouped amylopectin chain-length data.

longer mesogen units, and these in turn correlate with a larger Φ. This PCA work thus confirms a direct correlation of ε and Φ with the longer amylopectin chain-lengths. Furthermore, in Fig. 7, we can see that β shows a direct negative correlation (plotting opposite) to the long amylopectin chain-lengths (DP > 40). These longer chain-lengths are likely to pass through several lamellae within the semi-crystalline growth ring and this is likely to place significant limitations on any compressibility of the lamellae structure. An example of the negative correlation of β and the percentage of amylopectin chain-lengths, in this case between DP 41 and 45, is shown in Fig. 6b.

In the PCA landscape, clear clustering between the parameters obtained from the SAXS modelling and the amylopectin chain-lengths was observed. Moreover, the variables representing the shorter amylopectin chain-lengths plot in the region of the A-type crystal structure starches (Fig. 7). This finding suggests that there are a higher proportion of shorter chains in the A-type crystal structure starches than the B-type or C-type crystal structure starches. This is confirmed when considering the original data on the chain-length properties of amylopectin in Table 4. The average proportion of amylopectin chain-lengths of DP 6–15 (as a percentage) for A-, B- and C-type crystal structure starches is 32, 20 and 26%, respectively. This confirms previously published work (Blennow et al., 2000). By considering the other groupings of the amylopectin chain-lengths to the crystal structure types it is possible to observe further correlations. The average proportions of the amylopectin chain-lengths in the three grouped chain-length regions of DP 6–15, DP 16–30 and DP 31–60 for the three crystal type structure starches are shown in Table 4. From Table 4, it can be seen that C-type crystal starches have a higher proportion of amylopectin chain-lengths of DP 16–30 than A- or B-type crystal starches. Additionally, B-type crystal starches have a higher proportion of amylopectin chain-lengths of DP 31–60 than A- or C-type crystal starches.

Previous studies (Blennow et al., 2000; Thompson, 2000) on debranched amylopectin chain-lengths in the region of DP 6–60 have suggested correlations between the chain-length populations and the starch crystal type. It is typically observed that A-type crystal starches have a higher proportion of shorter chain-lengths between DP 6 and 60 than B- and C-type crystal starches, with B-type crystal starches having the highest proportion of longer chain-lengths. The chain-length data obtained by the HPAEC-PAD experiments on the starches used in this study is in agreement with these predictions. Taking the cumulative area of the chain-lengths for DP 6–15, DP 16–24 and DP 25–60 allows separation of the starches based on their crystal structure. A-type crystal structure starches have a greater proportion of chain-length between DP 6 and 15 and B-type crystal structure starches have a greater proportion of chain-lengths between DP 25 and 60. C-type crystal structure starches do not have noticeably larger populations in any range. In addition to being able to consider the individual chain-lengths of amylopectin from starches after debranching from DP 6 to 60, one can also calculate an average chain-length value of the amylopectin, CL. It has been previously reported (Hizukuri, 1985) that B-type crystal structure starches have higher average chain-lengths than A- and C-type crystal structure starches, with the A-type crystal structure starches having shorter CL than C-type crystal structure starches (since, C-type crystals starches are a combination of A and B polymorphs (Hizukuri, 1985)). The proportion of chain-lengths in the groups DP 6–15, DP 16–24 and DP 25–60, and the calculated CL from the chain-length data between DP 6 and 60 for all 27 starches, grouped into crystal-type, is shown in Table 5.

All the B-type crystal starches showed longer average chain-lengths than the A- and C-type crystal starches. Generally speaking, the C-type crystal starches have slightly higher CL than the A-type crystal structure starches, which is expected as C-type crystal starches are a combination of A and B polymorphs, and therefore should have a higher CL than pure A-type crystal structure starches.

As already noted (Fig. 7), the SAXS parameter β plots in the region of the A-type crystal structure starches. This is in agreement with the above discussion on the effects of β on the SAXS modelling results. A higher value of β will typically correlate to A-type crystal structure starches. It is interesting to note that whilst mung bean, C-type crystal structure starch, had the highest value of β, in general the C-type crystal structure starches do not plot immediately close to β. This, therefore, is a good example the usefulness of PCA at being able to manage

Table 5
Average of amylopectin chain-lengths between DP 6–15, DP 16–24 and DP 25–60, and CL by starch crystal type

Starch type crystal	DP 6–15 (%)	DP 16–30 (%)	DP 31–60 (%)	CL
A-type	32	36	32	23.8
B-type	20	37	43	27.2
C-type	27	38	35	24.6

and distinguish unexpected single values that differ from the expected norm.

When considering the SAXS data and HPAEC-PAD data separately, the previously established correlations within the individual data sets are also confirmed. The previous study postulated a negative correlation between β and ε from the four starches that were investigated (Daniels & Donald, 2003). As can be seen in Fig. 6c, the same negative correlation between β and ε is seen from the 27 starches investigated. The correlation is statistically significant at a level of $p < 0.05$. This tells us that a low β (compressibility) typically accompanies a high ε value (degree of layer bending), and supports the earlier hypothesis (Daniels & Donald, 2003) based on only four native starches. The negative correlation between β and ε does not appear to be particularly good. The reason is found in that good fits to the SAXS profile of A-type crystal structure starches were insensitive to changes in the ε variable as long as it was small. Therefore, often the best fits to the model had ε values close to zero. For some other starches such as C-type crystal structure starches and the modified starches the ε parameter was much more important.

Let us now look in more detail at the lamellar distortions. The weak negative correlation between β and ε (Fig. 6c) suggests that there may be a positive correlation between ε (degree of layer bending) and δ (splay of mesogens away from a preferred ordered direction), as both of these parameters relate to the magnitude of the distortions. Although the correlation is weak, it is improved when the modified starches and C-type crystal structure starches were not included (results not shown). There was some separation of the starches based on their crystal structure type. The A-type crystal structure starches typically showed little sign of splay (low δ), with the three starches with the lowest δ values being colflo67 and fridgeX (both chemically modified maize starches) and waxy maize starch. This implies, for the A-type crystal structure starches, that there is little splay of the mesogens away from the ideal, ordered parallel arrangement, and explains why the flat layer model originally described in the Cameron model (Cameron & Donald, 1992) worked well as long as only A-type starches such as wheat and maize were examined. The C-type crystal structure starches typically had the highest δ values, implying the greatest splay. This may be explained as C-type crystal structure starches are a combination of A- and B-type crystal structures and this combination may require greater splay within the structure in certain areas of the granule. Although the packing of A- and B-type crystal structures is thought to be spatially separate, with B-type crystals located at the centre of the granule and A-type crystals located towards the periphery (Bogracheva, Morris, Ring, & Hedley, 1998; Gidley, 1987) there must be some overlap and there must be a junction. The X-ray beam size used in these experiments was focused to approximately 3 mm × 2 mm, and therefore the resulting SAXS profiles were an average over many granules. Thus, over the transition range from A- to B-type crystal structures a higher local δ might indeed be expected to be observed.

Much is still unknown about the location and role of amylose within the starch granules. It has been postulated that amylose can co-crystallise with the amylopectin within the semi-crystalline growth ring and that this could disrupt the packing of the mesogen units (Jenkins & Donald, 1995). This disruption could be explained as causing mesogen splay from an ideal packing arrangement. The overall correlation for the 27 starches investigated here between δ and amylose content was very weak (results not shown), but if only starches of a given crystal structure are considered stronger correlations are seen. It is not surprising that weaker correlations will be observed when considering all crystal structure types together as there are likely to be other factors that are crystal-structure type specific that need be taken into account. When considering the A-type crystal structure starches only, which are known to have the highest crystallinity, since the shorter mesogen unit lengths permit good alignment, it may be expected that an increase in amylose content amongst these starches will have a significant effect on reducing the ability of the mesogen units to associate, pack and crystallise. The effect of increasing amylose content on the packing of the mesogens can be shown by considering the δ values. Fig. 6d shows the positive correlation between δ and the amylose content of the native A-type crystal structure starches; the positive correlation is obvious.

This suggests that increasing amylose content in A-type crystal structure starches may result in interference of the packing of the mesogen units in some way, as suggested by different models in earlier work (Jenkins et al., 1995; Yuryev et al., 2004). However, this is not a unique interpretation and the apparent correlation could arise from an independent effect that alters the amylose content. For example, some literature results suggest a correlation of increasing mesogen length with increasing amylose content due to biosynthetic growth pathways of starch (Shi, Capitani, Trzasko, & Jeffcoat, 1998), and this may cause a reduction in the quality of packing between the mesogen units by altering their lengths and therefore their interactions. More investigation is obviously required before this argument can conclusively be resolved.

From the data in Table 3, it is apparent that C-type crystal structure starches typically had the largest d spacing, indeed the biggest d spacing of all was 9.9 nm for chickpea. The lowest d spacing was 8.6 nm for rice, an A-type crystal structure starch. Overall, the d spacing values do not vary much between the starches, 8.6–9.9 nm (rice and chickpea, respectively), and therefore the essentially constant 9 nm repeat distance found previously, and hypothesised to be universal amongst all starches (Jenkins, Cameron, & Donald, 1993), is supported here. The diverse range of 27 starches considered in this study include both chemically and genetically modified starches, and these too were found to have a d spacing of around 9 nm (see Table 3). A definitive explanation for this consistent 9 nm repeat has yet to be confirmed. As indicated earlier (Waigh, Perry, Riekel, Gidley, & Donald, 1998), one physical explanation is that the combined effects of a long flexible spacer stabilising short mesogens, and conversely short flexible spacers being sufficient to stabilise long mesogens both lead to

a constant overall lamellar repeat distance. Such physical effects may combine with some aspect of the biosynthetic pathway through enzymatic activity (Fulton et al., 2002).

The Φ parameter gives an indication of the proportion of the rigid mesogen units within the crystalline region, such that a value of 1 indicates no flexible spacers within the lamellae, and a value of 0 corresponds to no mesogen material. The average Φ values for A-, B- and C-type crystal structure starches were 0.71, 0.75 and 0.73, respectively. The lowest Φ value corresponded to an A- crystal structure starch (rice, 0.65) and the highest value corresponded to a B-type crystal structure starch (KMC, 0.83). However, not all starches followed this 'trend'. For example, water chestnut (an A-type crystal structure starch) had a very high Φ value of 0.79 and, sweet potato and tapioca also have relatively (in comparison to the other A-type crystal structure starches) high Φ values. One hypothesis to explain this apparent variation in the Φ values of the A-type crystal structure starches is to consider the storage organ from which the starch was extracted. These data suggest that A-type crystal structure starches extracted from the endosperms of cereals will have a low Φ value, whilst A-type crystal structure starches extracted from other sources may have a higher Φ value, examples being sweet potato, tapioca and water chestnut that were extracted from tubers, roots and corms, respectively. However, more research would be required on a wider range of A-type crystal structure starches from different storage organs to confirm or reject this hypothesis.

There was also wide variation in the Φ values of the B-type crystal structure starches. However, the B-type crystal structure starches that had low values of Φ corresponded to the genetically modified potato starches, i.e. H944.16.1 (0.68), 926.28.3 (0.72), and H943.31.1 (0.73). As noted above for the case of H944.16.1, there is reason to think that the genetic modifications of these starches result in abnormal amylopectin chain-length properties and a corresponding abnormal fine structure of amylopectin. It is therefore probable that the amylopectin of these starches will be poorly described by the SAXS model used here and further refinement would be required to deal accurately with these genetically modified starches.

4. Conclusion

From the modelling work there is evidence that the starches can be separated by crystal type on the basis of their SAXS parameter values. The B-type crystal structure starches typically had high values of Φ and ε, implying long mesogen units and high layer bending, which also correlated with high δ values (degree of mesogen splay away from the preferred ordered direction). The A-type crystal structure starches typically had high values of β, implying the lamellae were highly compressible. This occurs when there are a high proportion of short mesogen units.

Significant correlations were found when comparing the SAXS parameter values and amylopectin chain-length data, which provided, for the first time, real experimental data to support the conclusions drawn from the SAXS model. A large Φ (long mesogen length) and a large ε (degree of layer bending) correlated to long amylopectin chain-lengths, whilst a high β value (layer compressibility) correlated to short amylopectin chain-lengths. A negative correlation was observed between Φ and β. Additionally, a negative correlation was observed between ε and β as expected. The application of PCA to a combined data set of SAXS parameters and experimentally determined amylopectin chain-length properties confirmed correlations that had previously been suggested on the grounds of intuition. Additionally, many of these intuitive correlations previously identified from small data sets were confirmed across a much larger, botanically diverse data set, which included examples of chemically and genetically modified starches. With the identification of clear correlations comes the suggestion that the amylopectin fine structure must be highly regulated. However, there are obvious species-specific properties, for example the B-type crystal structure starches having a higher proportion of longer amylopectin chain-lengths. This high degree of regulation is most likely to occur during formation through the biosynthetic process. This work, therefore, supports previous suggestions that there must be highly specific differences between the biosynthetic pathways of different starch producing plants (Waigh et al., 1998). These differences may arise either from the three groups of the A-, B- and C-type crystal structure starches, or alternatively from the storage organs utilised by starch producing plants.

Acknowledgements

The authors would like to acknowledge the financial support of the BBSRC, a CASE award from DuPont, and the Marie Curie Fellowship Award scheme. Thanks are also made for support and use of the facilities at The Centre of Molecular Plant Physiology (PlaCe), KVL, Denmark. The authors would like to thank Dr Nick Terrill, 8.2 station scientist, SRS, Daresbury Cheshire, UK for his assistance with the SAXS experiments. Finally, the authors would like to thank all the suppliers of the starches as indicated in Table 1.

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. New York: Wiley.

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 31P NMR. *Starch/Stärke*, *46*, 167–172.

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

Blennow, A., Engelsen, S. B., Munck, L., & Moller, B. L. (2000). Starch molecular structure and phosphorylation investigated by a combined chromatographic and chemometric approach. *Carbohydrate Polymers*, *41*, 163–174.

Bogracheva, T. Y., Morris, V. J., Ringer, S. G., & Hedley, C. L. (1998). The granular structure of C-type pea starch and its role in gelatinization. *Biopolymers*, *45*, 323–332.

Buleon, A., Gerard, C., Riekel, C., Vuong, R., & Chanzy, H. (1998). Details of the crystalline ultrastructure of C-starch granules revealed by synchrotron microfocus mapping. *Macromolecules*, *31*, 6605–6610.

Cameron, R. E., & Donald, A. M. (1992). A small-angle X-ray scattering study of the annealing and gelatinisation of starch. *Polymer*, *33*, 2628–2636.

Cheetham, N. W. H., & Tao, L. (1997). The effects of amylose content on the molecular size of amylose, and on the distribution of amylopectin chain length in maize starches. *Carbohydrate Polymers*, *33*, 251–261.

Corzana, F., Motowia, M. S., Hervé du Penhoat, C., van den Berg, F., Blennow, A., Perez, S., et al. (2004). Hydration of the amylopectin branch point. Evidence of restricted conformational diversity of the α(1–6) linkage. *JACS*, *126*, 13144–13155.

Daniels, D. R., & Donald, A. M. (2003). An improved model for analysing the SAXS of starch granules. *Biopolymers*, *69*, 165–175.

Daniels, D. R., & Donald, A. M. (2004). Soft material characterization of the lamellar properties of starch: Smectic side-chain liquid–crystalline polymeric approach. *Macromolecules*, *37*, 11312–11318.

Donald, A. M., Kato, K. L., Perry, P. A., & Waigh, T. A. (2001). Scattering studies of the internal structure of starch granules. *Starch*, *53*, 504–512.

Fulton, D. C., Edwards, A., Pilling, E., Robinson, H. L., Fahy, B., Seale, R., et al. (2002). Role of granule-bound starch synthase in determination of amylopectin structure and starch granule morphology in potato. *The Journal of Biological Chemistry*, *277*, 10834–10841.

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.

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.

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

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

Munck, L., Nørgaard, L., Engelsen, S. B., Bro, R., & Andersson, C. A. (1998). Chemometrics in food science—a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics and Intelligent Laboratory Systems*, *44*, 31–60.

Peat, S., Whelan, W. J., & Thomas, G. J. (1954). *Journal of the Chemical Society* , 4440.

Perry, P. (1999). *Plasticisation and thermal modification of starch*. PhD thesis, University of Cambridge.

Pilling, E. M. (2001). *The origins of growth rings in starch granules*. PhD thesis, University of East Anglia.

Ridout, M. J., Parker, M. L., Hedley, C. L., Bogracheva, T. Y., & Morris, V. J. (2003). Atomic force microscopy of pea starch granules: Granule architecture of wildtype parent, *r* and *rb* single mutants, and the *rrb* double mutant. *Carbohydrate Research*, *338*, 2135–2147.

Ridout, M. J., Parker, M. L., Hedley, C. L., Bogracheva, T. Y., & Morris, V. J. (2004). Atomic force microscopy of pea starch: Origins of image contrast. *Biomacromolecules*, *5*, 1519–1527.

Robin, J. P., Mercier, D., Duprat, F., Charbonniere, R., & Guilbot, A. (1975). Starch lintnerization. *Starch/Stärke*, *27*, 36.

Shi, Y. C., Capitani, T., Trzasko, P., & Jeffcoat, R. (1998). Molecular structure of a low amylopectin starch and other high amylose maize starches. *Journal of Cereal Science*, *27*, 289–299.

Thompson, D. B. (2000). On the non-random nature of amylopectin branching. *Carbohydrate Polymers*, *43*, 223–239.

Vikso-Nielsen, A., Blennow, A., Jorgensen, K., Kristensen, K., Jensen, A., & Moller, B. L. (2001). Structural, physicochemical and pasting properties of starches from potato plants with repressed r1-gene. *Biomacromolecules*, *2*, 836–843.

Waigh, T. A., Kato, K. L., Donald, A. M., Gidley, M. J., Clarke, C. J., & Riekel, C. (2000). Side-chain liquid–crystalline model for starch. *Starch/Stärk*, *52*, 450–460.

Waigh, T. A., Perry, P. A., Riekel, C., Gidley, M. J., & Donald, A. M. (1998). Chiral side chain liquid crystalline properties of starch. *Macromolecules*, *31*, 7980–7984.

Yuryev, V. P., Krivandin, A. V., Kiseleva, V. I., Wasserman, L. A., Genkina, N. K., Fornal, J., et al. (2004). Structural parameters of amylopectin clusters and semi-crystalline growth rings in wheat starches with different amylose content. *Carbohydrate Research*, *339*, 2683–2691.