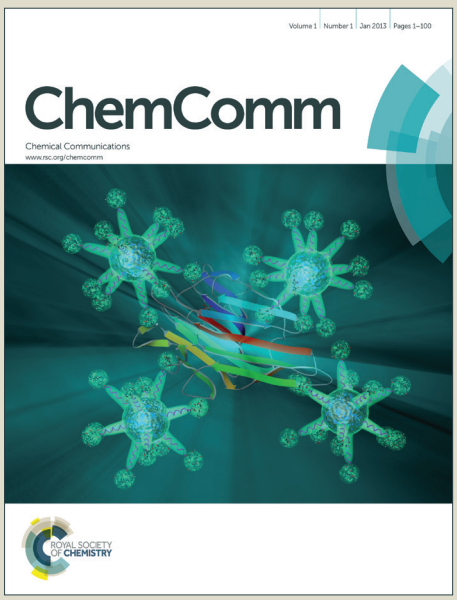
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

COMMUNICATION

Rapid Quantification of Starch Molecular Order Through Multivariate Modelling of ^{13}C CP/MAS NMR Spectra

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Bernadine M. Flanagan[a], Michael J. Gidley[a] and Frederick J. Warren[a]†

A partial least squares model has been generated enabling the rapid assessment of ordered molecular structure in a semi-crystalline polymer, starch, directly from solid state NMR spectra. Solid state NMR spectroscopy offers many advantages over conventional analysis tools being non-destructive and functional in complex mixtures.

Starch is typically the single largest component in most people's diet and represents an abundant and readily-processed renewable polymer source for industrial applications. The percentage helical order or crystallinity of starches is recognized as contributing to signature properties such as the integrity of granules, the mechanical properties of processed starches, and amylase digestibility.[1-7] The transition of starch from its native, semi-crystalline, granular form to an amorphous form through hydrothermal processing (cooking) results in a dramatic increase in susceptibility to α-amylase digestion.[4] The resultant increase in available glucose from starchy foods as a result of hydrothermal processing has been shown to be highly important to human nutrition,[8] and indeed the development of cooking, and subsequent increased availability of starchy foods as a glucose source, has been highlighted as a key step in human evolution.[9]

A number of methods have been used to characterize starch ordered structures such as X-ray diffraction (XRD),[10] differential scanning calorimetry (DSC),[11, 12] and rapid visco-analysis (RVA).[13] These techniques all have limitations in terms of experimental time, availability, amount of sample required and sample preparation. Solid state ^{13}C NMR spectroscopy[14] can be used to analyse short range order in starches due to the different chemical shift patterns characteristic of ordered (arising from A, B and V type crystalline polymorphs) and non-ordered materials. Using peak fitting software and amorphous material standards, a quantitative analysis of double helical, single-helical, and non-ordered starch content can be obtained.[14] The ^{13}C CP/MAS NMR spectrum of starch contains resonances which have been assigned to each of the carbons in the glucose ring (Figure 1). Approaches to determining ordered structure from starch NMR spectra involve subtracting non-ordered (amorphous) spectra and then deconvoluting to generate crystalline and amorphous sub-spectra, and assessing the relative contributions of these two spectra.[14-16] Generally, resonances arising as a result of the C1 carbon of the glucosidic ring are used, as they are particularly sensitive to changes in both relative order and crystalline polymorph.[14] The C1 peak is a triplet for the A type crystalline polymorph (indicated in Figure 1), and a doublet for B type crystalline starch.[14] Deconvolution approaches to evaluating ordered structure are of limited application, however, as they are time consuming and introduce a degree of subjectivity in the analysis of the data as amorphous reference spectra may vary depending on the botanical origin of starch and the preparation and drying methods.

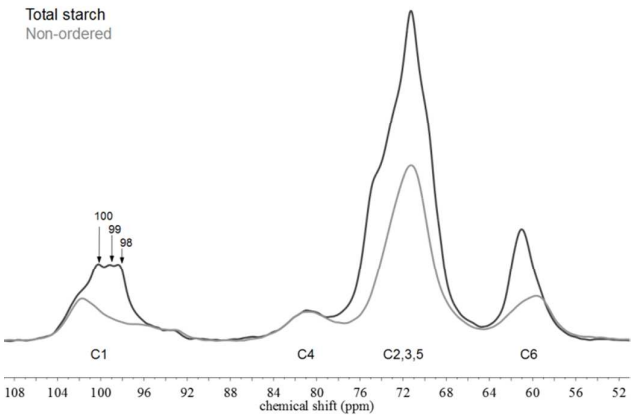


Figure 1. Exemplar 13C CP/MAS NMR spectra for semi-crystalline A-type starch (black) and non-ordered (amorphous) starch (grey) with the assignments of each of the peaks indicated.

[a] *Centre for Nutrition and Food Sciences, ARC Centre of Excellence in Plant Cell Walls, Queensland Alliance Agriculture and Food Innovation, The University of Queensland, St Lucia, Qld 4072, Australia*
† Tel ++61(0)733467373 f.warren@uq.edu.au
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

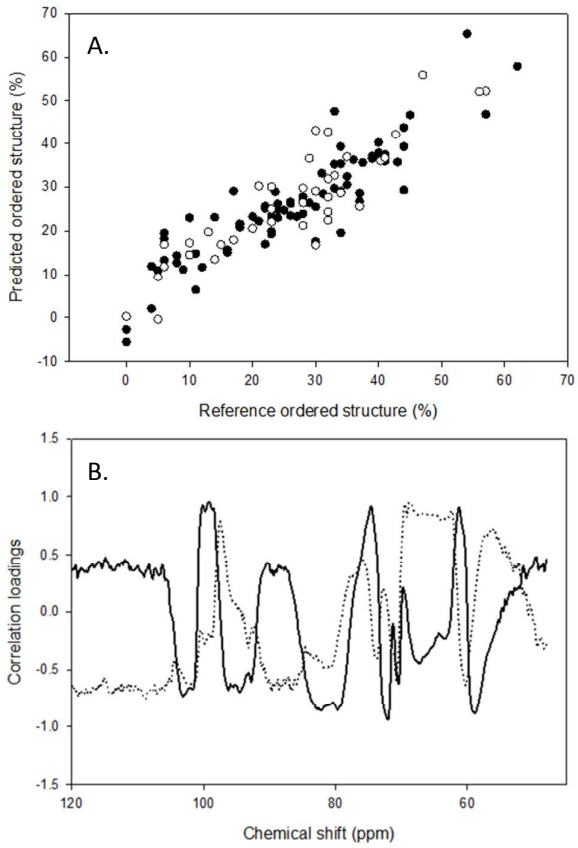


Figure 2. PLS model constructed using the full 13C CP/MAS NMR spectra for starch. a. Reference vs. predicted ordered helical structure for calibration samples (filled circles) and validation samples (open circles). b. Model correlation loadings for principle component 1 (solid line) and principle component 2 (broken line)

A significant advantage of NMR over alternative methodologies such as XRD is that impurities in the samples such as lipid, protein or polyphenols do not interfere with the spectrum of starch making it possible to study the molecular order of starches even in processed foods and complex mixtures, so improved data analysis tools are urgently needed. Using a large reference set of starch spectra including different varieties of wheat, barley, rice and maize, the present study aimed to use chemometric approaches to analyse ordered structure in starch from ^{13}C CP/MAS NMR spectra. The starches were used either in a granular form or following processing such as: extrusion, digestion or boiling. Freeze drying, oven drying and air drying were also used to introduce diversity to the sample set. A total of 114 starch spectra were used, randomly split into a calibration set consisting of 76 spectra and a validation set comprising 38 spectra. These spectra were subjected to Partial Least Squares (PLS) modelling. PLS is a statistical method whereby two matrices X- (in this case NMR spectra) and Y- (molecular order values from spectral deconvolution) are modelled to find the latent variables in X that best predict the latent variables in Y.

The spectral deconvolution method of Tan *et al.*[14] was applied to all 114 spectra to produce reference data values, and the starches were found to cover a range of molecular order from amorphous to 50% order and all crystalline polymorphs (A, B and V) were represented (Supplementary Table T1). This provides a reference data set of spectra for analysis by multivariate modelling covering the full range of ordered structure and crystalline polymorph that may be expected to be encountered during the analysis of native and processed starches.[6, 10, 15, 17] DSC and XRD was also carried out on a subset of the starch samples (Supplementary Table T1) to allow comparison of the performance of the conventional deconvolution method and new multivariate analysis methods with alternative measures of starch ordered structure.

A preliminary model (Figure 2) was built using the full spectral range covering all the main resonances associated with the glucan ring, from 48 to 122 ppm, using single normal variate (SNV) normalised spectra, to ascertain the validity of using PLS to model NMR spectra of starch, and to explore the main regions of variance in the spectrum correlated with the presence of ordered starch structure. As shown in Figure 1a,

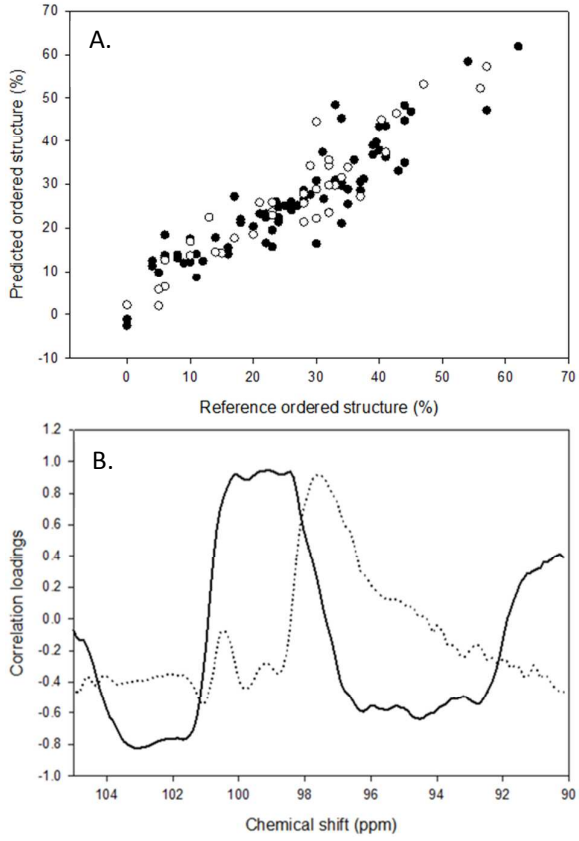


Figure 3. PLS model constructed using the C1 region of the 13C CP/MAS NMR spectra for starch. a. Reference vs. predicted ordered helical structure for calibration samples (filled circles) and validation samples (open circles). b. Model correlation loadings for principle component 1 (solid line) and principle component 2 (broken line)

using 2 principle components, a predictive model can be built for total ordered structure arising from A or B type crystallinity, with a root mean square error (RMSE) of 5.5%, showing very similar performance for the calibration and validation datasets. The correlation loadings (Figure 2b) demonstrate that the model is fitting to variation in all of the starch associated peaks in the chosen spectral region. It should be noted that while A/B ordered structure could be predicted from the model, it was not possible to predict V-type ordered structure from this model (or any of the models presented in this paper), as it was found that there was too little variability in V-type crystallinity between the different samples, and the overall amount of V-type present in each of the samples was very low (<10%).

Table 1. Fitting parameters for PLS models using calibration and validation data sets, for the full spectrum, and each individual peak.

Spectral region	Principle Components	r^2 (calibration)	RMSE (calibration)	r^2 (validation)	RMSE (validation)
Full Spectrum	2	0.83	5.51%	0.84	5.39%
C1	2	0.85	5.20%	0.87	4.86%
C2,3,5	2	0.81	5.95%	0.72	7.18%
C4	2	0.76	6.67%	0.86	5.21%
C6	2	0.85	5.23%	0.82	5.74%

To further explore the suitability of each of the major peaks in the starch spectrum for modelling ordered structure, PLS models were constructed using spectral regions relating to each of the major peaks present in the spectrum of starch (C1, 90-105 ppm; C2,3,5, 65-78 ppm; C4, 77-88 ppm; C6, 56-64 ppm). The model performance for each of the major peaks, and for the full spectrum, is summarised in Table 1. As would be expected, given that Figure 2b demonstrates that there are correlations between all the major peaks in the spectrum and the level of ordered structure in starch, and Figure 1 demonstrates that there are significant contributions from the ordered sub-spectrum of starch to each of the major peaks in the NMR spectrum of starch, it was possible to construct and validate models using each of the peaks individually (Supplementary Figure S1 and Figure 3). There was, however, variation in model performance between different peaks. The worst calibration performance was observed for the C4 peak, which is to be expected as this peak has the least difference between amorphous and crystalline sub-spectra (Figure 1), therefore the least variation for the model to fit to. The best model performance was observed for the C1 peak. This is in agreement with previous studies which have focussed on the C1 peak for determining ordered helical structure from starch NMR spectra.[10, 14, 16, 18] A closer look at the model for the C1 peak (Figure 3a) shows that it is capable of predicting ordered structure from between 5% and 55% ordered structure (below 5% ordered structure the model becomes unreliable due to the small number of available calibration samples). The regression loadings (Figure 3b), show that the model is fitting to the triplet C1 peak, with peaks in the regression coefficients at 98, 99 and 100 ppm, matching those in the crystalline sub-spectrum of A-type crystalline starch. The performance of the model constructed using the C1 peak is comparable to that of the full spectra, but using only the C1 peak has the advantage that the model is more robust to the presence of other spectral signals. Thus, the model is applicable to complex mixtures containing other components, as long as there is no overlap with the C1 spectral region.

To allow wider use of this model, a spread sheet (Excel) macro was created to allow prediction from the model parameters (Supplementary data file). ^{13}C CP/MAS NMR spectra of starch can be copied and pasted into the spread sheet as x-y data. The macro interpolates the C1 region of the spectrum, and applies SNV normalisation and mean centring. It then applies multiple linear regression using β-factors extracted from the C1 PLS model to give a prediction of percentage ordered double helical structure present in the sample from the spectrum.

A subset of 56 samples had ordered double helical structure predicted using the prediction macro, and using the conventional deconvolution method, and the results were correlated (Supplementary Figure S2) against two alternative measures of starch ordered structure (XRD and DSC). It can be seen that there is a significant ($p < 0.0001$) correlation between the values predicted by the macro and both XRD and DSC data, and that the predication macro provides very similar performance to the deconvolution method. Close inspection suggests that the macro may actually provide somewhat better performance, as the data points are more evenly spaced. For comparison, a correlation is also included between XRD and DSC data.

The new prediction equation removes the need for the lengthy iterative fitting process and the need for an amorphous standard. The NMR method requires no sample preparation, the spectra used for prediction can be collected in as little as one hour, and the method is non-destructive. Using the provided spread sheet macro, the present work demonstrates the rapid and efficient determination of ordered double helix starch structure from ^{13}C CP/MAS NMR spectra, with reduced subjectivity. This will allow this powerful and flexible tool to become more widely and readily applied in studies of starch structure in complex, heterogeneous systems.

References

1 F. J. Warren, P. G. Royall, S. Gaisford, P. J. Butterworth and P. R. Ellis, Carbohydr. Polym., 2011, 86, 1038-1047.

2 R. Tahir, P. R. Ellis and P. J. Butterworth, Carbohydr. Polym., 2010, **81**, 57-62.
3 M. Li, T. Witt, F. Xie, F. J. Warren, P. J. Halley and R. G. Gilbert, Carbohydr. Polym., 2015, **122**, 115-122.
4 A. J. Baldwin, D. L. Egan, F. J. Warren, P. D. Barker, C. M. Dobson, P. J. Butterworth and P. R. Ellis, Biomacromolecules, 2015, **16**, 1614-1621.
5 S. Hizukuri, Carbohydrate research, 1985, **141**, 295-306.
6 A. Lopez-Rubio, B. M. Flanagan, A. K. Shrestha, M. J. Gidley and E. P. Gilbert, Biomacromolecules, 2008, **9**, 1951-1958.
7 F. Bertocchi and M. Paci, J. Agric. Food Chem., 2008, **56**, 9317-9327.
8 J. Holm, I. Lundquist, I. Björck, A. Eliasson and N. Asp, Am. J. Clin. Nutr., 1988, **47**, 1010-1016.
9 K. Hardy, J. Brand-Miller, K. D. Brown, M. G. Thomas and L. Copeland, Q. Rev. Biol., 2015, **90**, 251-268.
10 A. Lopez-Rubio, B. M. Flanagan, E. P. Gilbert and M. J. Gidley, Biopolymers, 2008, **89**, 761-768.
11 D. Cooke and M. J. Gidley, Carbohydr. Res., 1992, **227**, 103-112.
12 P. J. Jenkins and A. M. Donald, Carbohydr. Res., 1998, **308**, 133-147.
13 X.-Z. Han and B. R. Hamaker, J. Cereal Sci., 2001, **34**, 279-284.
14 I. Tan, B. M. Flanagan, P. J. Halley, A. K. Whittaker and M. J. Gidley, Biomacromolecules, 2007, **8**, 885-891.
15 T. Y. Bogracheva, Y. Wang and C. Hedley, Biopolymers, 2001, **58**, 247-259.
16 M. J. Gidley and S. M. Bociek, J. Am. Chem. Soc., 1985, **107**, 7040-7044.
17 T. Bogracheva, Y. Wang, T. Wang and C. Hedley, Biopolymers, 2002, **64**, 268-281.
18 M. Paris, H. Bizot, J. Emery, J. Buzare and A. Buleon, Int. J. Biol. Macromol., 2001, **29**, 127-136.