Carbohydrate Research, 207 (1990) 167–175
Elsevier Science Publishers B.V., Amsterdam

Multivariate calibration in the determination of acetylated aldoses by g.l.c.

Markku D. Hämäläinen, Olof Theander*,
Department of Chemistry, The Swedish University of Agricultural Sciences, Box 7015, S-750 07 Uppsala (Sweden)

Erik Nordkvist,
National Laboratory of Agricultural Chemistry, Box 7004, S-750 07 Uppsala (Sweden)

and Ingrid Eva Ternrud
Department of Forest Products, The Swedish University of Agricultural Sciences, Box 7008, S-750 07 Uppsala (Sweden)

(Received December 18th, 1989; accepted for publication, February 28th, 1990)

ABSTRACT

An improved g.l.c. method for the analysis of acetylated aldoses is described. Fast acetylation (acetic anhydride–1-methylimidazole) can be performed directly on the acidic aldose hydrolysate in a one-vessel operation. Factors most likely to influence the ratios of isomers in the reaction solutions at the time of derivatisation were studied. The derivatisation procedure was modified to minimise these effects. Quantification can be based either on linear regression of peaks for single isomers of each sugar or on all peaks by designed calibration standards and partial least-squares regression (PLSR). The latter calibration gives slightly better reproducibility and stability. PLSR can also be used when the chromatographic separation is poor. The results and the reproducibility of the two calibration methods for the peracetates were compared with alditol acetate analysis for four plant materials.

INTRODUCTION

Analysis of neutral non-starch sugar constituents of plant polysaccharides is generally performed either by g.l.c., preferably of the alditol acetates[1], or by liquid chromatography[2]. In the former method, the plant material is hydrolysed, and the products are reduced and acetylated to give acetylated alditols[3,4], but the reduction is not always complete[5]. A modified method that omits the reduction step is now reported which involves more rapid derivatisation and more reliable identification.

The omission of the reduction step results in a more complex chromatogram, and problems arise in the calibration due to variations in the ratios of isomers, overlapping of peaks, and co-variation between peaks derived from isomers. Calibration based on linear regression (LR) and multiple linear regression is inadequate if the isomers are not resolved completely and/or if the variation in the ratios of isomers is not eliminated. Partial least-squares regression (PLSR), which is a generalised multivariate regression

* Author for correspondence.

0008-6215/90/$ 03.50

method[6-9], provides a solution to these problems. PLSR automatically handles additive relations and moderate non-linearities, and information on all peaks for isomers can be used if the standards are designed properly. The advantages of this type of design and calibration are discussed.

EXPERIMENTAL

Materials. — Pure acetic anhydride should be used in order to avoid interfering peaks in g.l.c., especially in the pentose area. The plant materials were peat of *Sphagnum fuscum*, defatted soya flour, lucerne (*Medicago sativa*), processed wheat fibres (Tricum AB, Sweden), and purified cotton-linters cellulose. Each sample had been analysed by current methods[4]. Each plant material was ground in a Tecator Cyclotec sample mill to pass a 0.5-mm screen.

Hydrolysis and derivatisation. — A mixture of the ground plant material (90–600 mg) and 12M sulfuric acid (3.0 mL) was agitated for 30 min at 30°, then stirred and left for an additional 30 min. Water (79 mL) and internal standard (mannitol, 5.00 mL, 2.00 mg/mL) were added, the sample was autoclaved[4] at 120° for 1 h and then filtered (Pyrex No. 2 glass filter) warm, and an aliquot (0.5 mL) was kept for 15 min at 20 ± 1°. After each of the following additions, the tubes were vibromixed. A portion (5.5 mL) of a freshly prepared mixture of 1-methylimidazole and acetic anhydride (1:10) was added followed, after 10 min, by ethanol (1 mL). After a further 10 min, the tube was cooled to ~ 10° in a water bath. In the following steps, it is important always to maintain the water level in the bath above the level of the solution in the test tube in order to prevent heating and the appearance of tarry deposits. Water (5 mL) was added and, after 5 and 10 min, 7.5M potassium hydroxide (5 mL each time). The upper layer, ethyl acetate containing the acetylated sugars, was withdrawn and stored over anhydrous sodium sulfate.

The sugar solutions used for the calibrations were mixed with internal standard (mannitol), and sulfuric acid was added to 0.41M. The hydrolysis and derivatisation was performed as described above. G.l.c. of the alditol acetates was performed according to Theander and Westerlund[4], which is essentially as described above, with a reduction step prior to acetylation. G.l.c. was performed with a Packard model 427 gas chromatograph equipped with flame-ionisation detector and a CP-SIL 88 (9 m × 0.22 mm i.d.) capillary column with He as the carrier gas at 150 cm/min (split ratio 1:20). The detector and injector temperatures were 230 and 220°, respectively. Analyses were performed at 185° (see Fig. 1), and with the temperature programme 175° for 5 min, 4°/min to 220°, kept for 2 min at 220°.

Calibration and statistical analysis. — The most commonly occurring neutral sugars that occur in plant materials (Rha, Ara, Xyl, Man, Gal, and Glc) were included in the calibration, and the ranges of the calibration samples covered most of the polysaccharide compositions of different plant materials (Table I). Xylitol, mannitol, galactitol, and *myo*-inositol (retention times at 185°, 7.30, 11.61, 13.35, and 21.5 min, respectively) are suitable as internal standards and mannitol was selected.

Calibration was performed at 4 levels with 9 samples in a nearly orthogonal

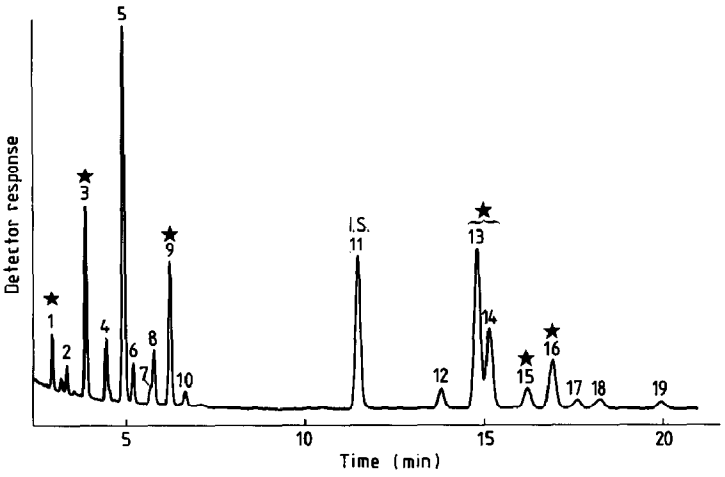

Fig. 1. Gas chromatogram of peracetylated aldoses prepared from an acid hydrolysate of peat sample: 1, Rha; 2, Rha; 3, Xyl; 4, Ara; 5, Xyl; 6, Ara; 7, Xyl; 8, Ara; 9, Ara; 10, Xyl; 11, mannitol (internal standard); 12, Gal; 13, Glc; 14, Glc; 15, Man; 16, Gal; 17, Gal; 18, Man (Glc); 19, Gal. Peaks denoted by * were used in the linear regression, and minor co-eluting peaks are in brackets.

TABLE I

The design matrix and the range of corresponding sugars used for the calibration[a]

Standard. No.	*Glc*	*Gal*	*Man*	*Xyl*	*Ara*	*Rha*	*Range (mg)*		
								−1	*+3*
1	+1	+1	−1	+1	+1	−1	Glc	3.01	100.02
2	−1	+1	+1	−1	+1	+1	Gal	1.00	23.97
3	+1	−1	+1	+1	−1	+1	Man	1.01	18.04
4	−1	−1	−1	−1	−1	−1	Xyl	0.99	15.96
5	0	0	0	0	0	0	Ara	1.00	15.96
6	0	0	0	0	0	0	Rha	1.00	9.98
7	+3	0	0	+3	0	0			
8	0	+3	0	0	+3	0			
9	0	0	+3	0	0	+3			

[a] An identical design was used for the two groups.

design[10] (Table I). The differences and advantages of PLSR compared to conventional calibration have been reviewed[11] and a textbook on multivariate calibration is available[6]. Raw data were integrated with the Nelson 2600 Chromatography Software System on an IBM-AT compatible computer. The areas were normalised with respect to the internal standard and collected into two X-matrices (1, Rha, Ara, Xyl; 2, Man, Gal, Glc). As the accuracy in the areas determined was better for the larger peaks, except for that of xylose, the X-matrices were not scaled. This allows the large peaks to have more influence on the PLSR models. The Y-matrices (the amounts in the calibration stan-

dards) were autoscaled (*i.e.*, the areas were multiplied with the inverse of the standard deviation), in order to give each aldose an equal impact on the models. All matrices were mean centered, *i.e.*, the mean value was subtracted from each variable. Three significant PLSR factors were used in each model. Each factor accounted primarily for the variation of one sugar.

In order to find the best conditions for a reproducible derivatisation procedure, the following three factors were studied. (*a*) The delay time between the additions of 1-methylimidazole and acetic anhydride to the mixture of aldoses in 0.41M H_2SO_4 was varied (0, 1, 5, and 10 min) at constant temperature (20°). (*b*) After hydrolysis, the mixture was cooled to different temperatures (90, 40, 30, and 20°), and an aliquot was added to a freshly prepared mixture of 1-methylimidazole and acetic anhydride. (*c*) After hydrolysis, the acidic mixture of aldoses was cooled to 20° and derivatised after 0, 15, 30, 60, and 90 min. The correlation between peak areas (X-matrix) and the different factors (y) were subjected to three separate PLSR analyses.

For the multivariate calibrations, the commercially available SIRIUS package (Pattern Recognition Systems A/S, N-5015 Bergen, Norway) and the SIMCA 3-B Basic program package (Sepanova AB, Östrandsv. 14, S-122 45 Enskede, Sweden) were used, each of which is intended for use with IBM-PC compatible computers.

RESULTS AND DISCUSSION

The sample chromatogram shown in Fig. 1 consists of two groups of acetylated sugars that correspond to anomers and isomeric ring forms of (*a*) pentoses and deoxyhexose, and (*b*) hexoses. Since each sugar gives at least two peaks, the identification was easier and more secure than with the alditol acetate method, where each sugar gives only one derivative.

Reduction of the variations in the ratios of isomers during the derivatisation procedure. — In the previous derivatisation procedure[3–5], there is an undefined delay between the additions of 1-methylimidazole and acetic anhydride to the solution of the aldose. This is the main factor that influences the ratios of isomers, since the addition of 1-methylimidazole changes the solvent properties of the mixture. Displacement between the different isomers resulting from different delay times is shown in Fig. 2, as the described variances in the two-factor PLSR model. The effect was most pronounced for glucose (peak 13→14), rhamnose (2→1), galactose and mannose (18→15). In order to minimise this influence, a mixture of *freshly* prepared 1-methylimidazole and acetic anhydride was added to the acidic mixture of aldoses, *i.e.*, there was no delay time.

Figure 3 shows the effect of temperature on the equilibrium mixture analysed by PLSR. The effect of temperature on the areas was low but significant (0.7%/°, mean area variation). Arabinose (peak 6→8,9), xylose (7,10→3,5), galactose (19,10→16,12), and mannose (18→15) showed clear displacements between the different isomers. Increase in the total variation for the method could not be seen if the temperature was kept within ±1°.

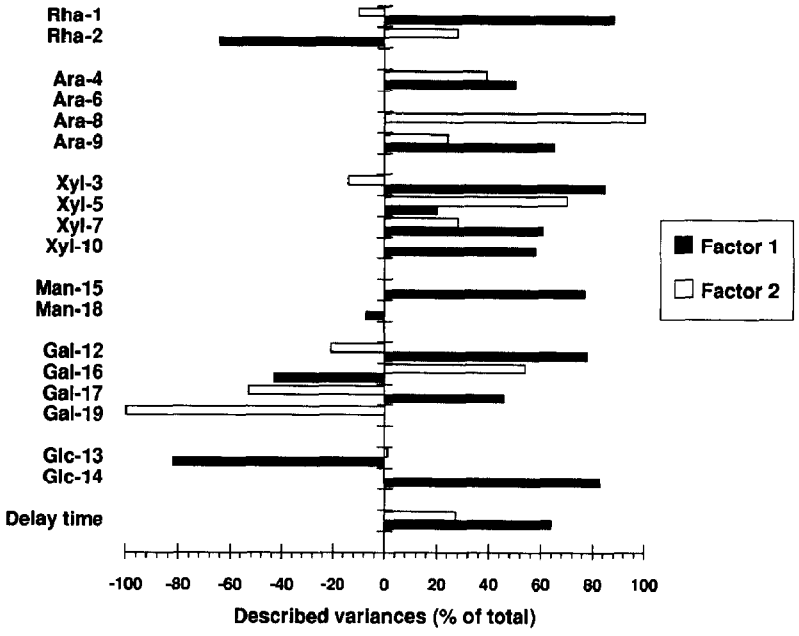

Fig. 2. The effect of time on the isomer ratios after a change in solvent properties displayed with PLSR (X=peak areas, Y=delay time). The negative and positive bars indicate decreasing and increasing areas with increasing time, respectively, in the two-factor PLSR model. The numbers after each sugar refer to peak numbers in Fig. 1.

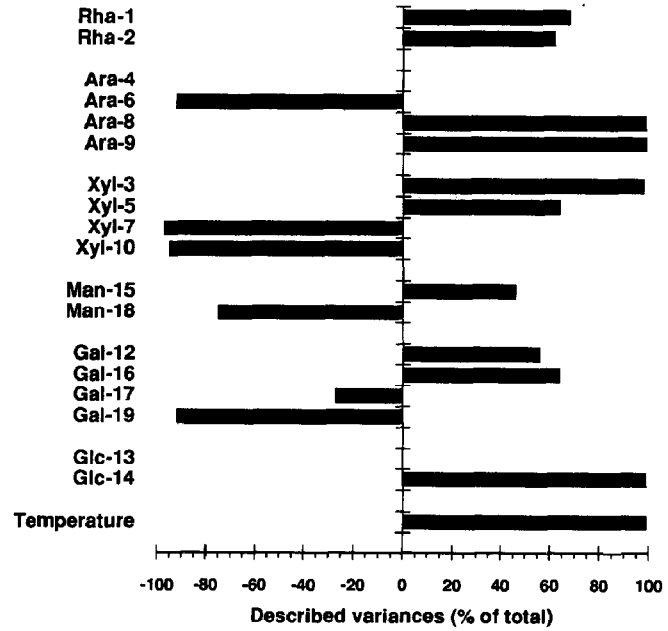

Fig. 3. The effect of temperature on the isomer ratios, displayed with one PLSR factor (X=peak areas, Y=temperature). The negative and positive bars indicate decreasing and increasing areas with increasing temperature, respectively, in the one-factor PLSR model. The numbers after each sugar refer to peak numbers in Fig. 1.

There was no significant variation in peak areas related to the delay time after hydrolysis.

It is not only convenient for derivatisation to be performed directly in the sulfuric acid hydrolysate, but the highly acidic conditions also ensure that the equilibrium of the isomers is achieved rapidly. Thus, it is sufficient to establish a constant temperature ($\pm 1°$) prior to derivatisation. The mixture of aldoses will then be stable and, finally, by mixing the 1-methylimidazole and acetic anhydride immediately before addition to each sample, the effect of a change in solvent properties is minimised.

Calibration based on single peaks. — After the derivatisation had been standardised, it was possible to find single peaks with base-line separation (Fig. 1) and sufficient precision to use with LR for all the common aldoses (Fig. 4). These peaks, which corresponded to the most preponderant isomers of the equilibrium mixtures, were quantified with the greatest precision and therefore gave the best correlations. The two main peaks for glucose were not base-line separated and the precision was improved by using their sum. The main peak for xylose (peak 5, Fig. 1) contained impurity, so that the quantification was based on the second largest peak (peak 3).

The hydrolysis–derivatisation procedure and the LR model were also tested with 4 replicate analyses of purified cotton-linters cellulose, and gave a 100.6% mean dry-matter content of "anhydroglucose" with a coefficient of variation of 0.6%.

Multivariate calibration based on all peaks. — Multivariate calibration by PLSR gave better results compared with LR (Fig. 4). Overlapping peaks can be utilised by the

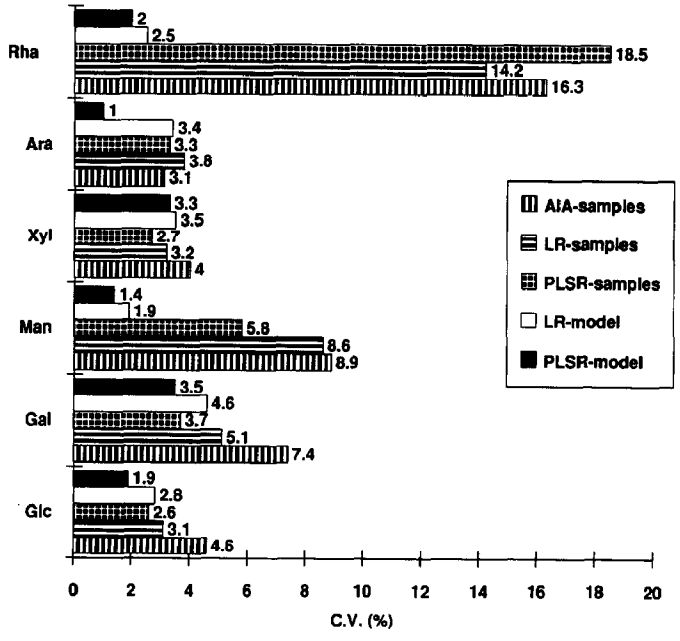

Fig. 4. The coefficient of variation (c.v.) for aldoses analysed as acetates and calibrated by linear regression (LR) and partial least-squares regression (PLSR), or analysed as alditol acetates (AlA).

PLSR, by preparing standards according to an orthogonal design[10] or a nearly orthogonal design[12] (Table I), *i.e.*, no or negligible correlation exists between the amounts in the mixtures. The above-mentioned separation between groups of sugars makes it possible to use the same calibration design for pentoses–deoxyhexose and hexoses (*i.e.*, a smaller number of calibration samples are needed to create a nearly orthogonal design). Two separate PLSR models were also calculated, one for each group.

The effect of PLSR analysis on an unresolved peak is illustrated in Fig. 5. The mean-centered area of peak 18, derived mainly from mannose with a minor contribution from glucose, was plotted against the content of mannose (mean-centered and autoscaled) in corresponding standard solutions. Due to the influence of glucose, the correlation between mannose and the peak area was too low for quantification with LR. The first PLSR factor extracted the variation caused by glucose (using also peak 13 and 14), and left residuals which were well correlated with mannose (the second factor did not affect this peak.) The third factor extracted the variation caused by mannose (using also peak 15) and the residuals were left with little and unstructured variation.

When new samples with different patterns in the X-matrix (new sugar, impurity, *etc.*) are fitted to the PLSR model, the X-residuals can be used to find deviating samples.

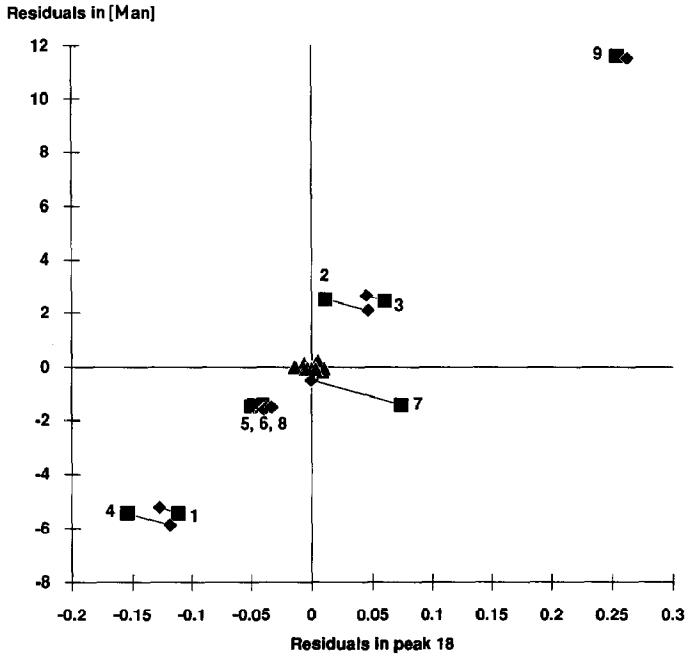

Fig. 5. The effect of extraction of PLSR factors from the variation of an overlapped peak and mannose. The residuals of peak 18 are plotted against the residuals in mannose (y). The data are mean-centred ($x_i - x_{mean}$, $y_i - y_{mean}$). The numbers 1–9 refer to the calibration samples in Table I. Mean-centred residuals in unmodelled data (■); residuals after one PLSR-factor, the variation related to glucose is modelled (eliminated), and the correlation is improved (◆); residuals after modelling the variation related to mannose content, leaving residuals which are mainly random noise (▲).

The effect of this noise is, at least partly, left in the residuals and not used in the prediction. PLSR is a soft regression method, *i.e.*, projections from the X-matrix (latent variables, factors) are used for building the regression model. This procedure gives more robust models and also a reduction of noise. A linear combination of peaks 15 and 18 gave a noise-reduction step which was revealed in the lower coefficients of variation of PLSR compared with LR based on peak 15 (Fig. 4). Although peak 18 contained more noise than peak 15, the combined information used by PLSR gave a more robust model than LR. When analysing the plant samples, this procedure decreased the coefficients of variation from 8.6% (LR) to 5.8% (PLSR). Thus, PLSR is an excellent tool for multivariate calibration in equilibrium systems.

The multivariate calibration methodology has been applied to a more complex system, namely, methanolysis[13] of sugars and acetylation of the resulting methyl glycosides for Rha, Fuc, Ara, Xyl, Man, Gal, and Glc and the methyl ester methyl glycosides of GlcA and GalA. When there was overlap between peaks from hexoses and uronic acids, conventional calibration was not applicable. However, 98.6% of the variation in designed standards could be described by a PLSR analysis, to give relative standard deviations of 2–4% for pentoses, hexoses, and deoxyhexoses, and 10–13% for the uronic acids.

Analysis of plant samples and reproducibility. — Four aldose-containing samples, namely, peat of *Sphagnum fuscum*, processed wheat fibre, lucerne, and soya flour, were

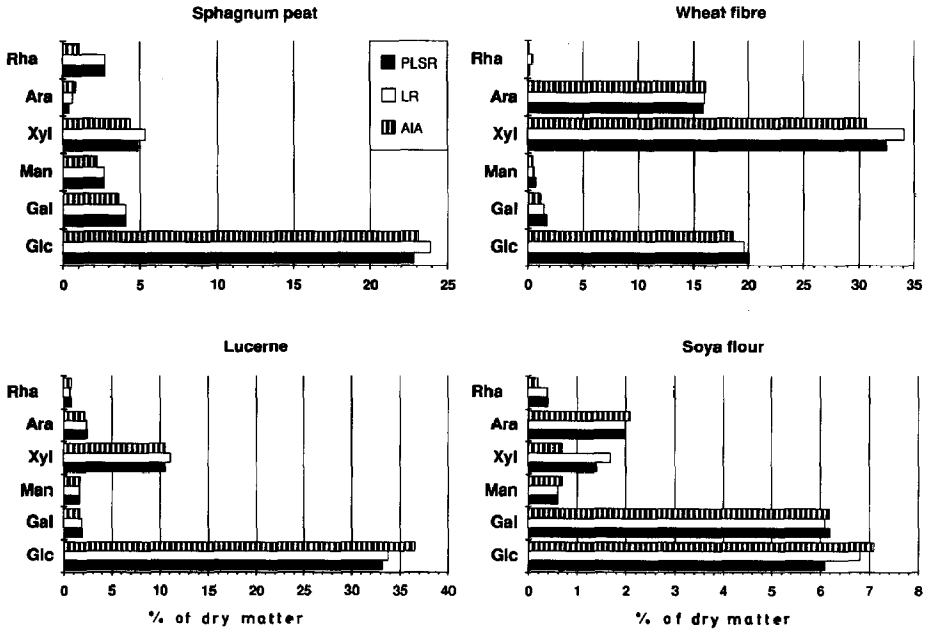

Fig. 6. Mean composition of test samples. Results are calculated with partial least-squares regression (PLSR) and linear regression (LR) for the acetates or analysed as alditol acetates (AlA). Carbohydrates given as "anhydro" residues.

analysed in quadruplicate during 4 days. For purposes of comparison, the samples were also analysed four times as the alditol acetates. The aldose contents and coefficients of variation, calculated from the single peak regression (LR), PLSR models (PLSR), and the alditol acetates analysis (AlA), are presented in Fig. 6. The two calibration methods and the alditol acetate analysis gave similar contents of constituents. However, the PLSR model finally gave a better reproducibility, as shown by the coefficients of variation (Fig. 4). The high coefficients of variation of rhamnose in all three methods are due to the low contents in the test samples.

The method described was conducted in one vessel. Since derivatisation is effected immediately after hydrolysis (when constant temperature is achieved) and the procedure takes ~35 min, the capacity of the method depends largely on the equipment for hydrolysis and g.l.c. Some 40–50 samples a day can be processed if the g.l.c. equipment has an auto-injector and the calculations are made automatically.

ACKNOWLEDGMENTS

We thank Mr. Peter Brandt for skilful experimental work, and the Swedish National Energy Administration and the Swedish Council for Building Research for financial support (to M. H.).

REFERENCES

1 G. C. McGinnis and C. H. Biermann, in C. H. Biermann and G. C. McGinnis (Eds.), *Analysis of Carbohydrates by GLC and MS*, CRC Press, Boca Raton, 1989, pp. 87–118.
2 G. C. McGinnis, M. L. Laver, and C. H. Biermann, in C. H. Biermann and G. C. McGinnis (Eds.), *Analysis of Carbohydrates by GLC and MS*, CRC Press, Boca Raton, 1989, pp. 19–26.
3 A. S. Bittner, L. E. Harris, and W. F. Campbell, *J. Agric. Food Chem.*, 28 (1980) 1242–1245.
4 O. Theander and E. A. Westerlund, *J. Agric. Food Chem.*, 34 (1986) 330–336.
5 A. B. Blakeney, P. J. Harris, R. J. Henry, and B. A. Stone, *Carbohydr. Res.*, 113 (1983) 291–299.
6 H. Martens and T. Naes, *Multivariate Calibration*, Wiley, New York, 1989.
7 H. Martens, *Dr. Techn. Thesis*, Technical University of Norway, Trondheim, 1985.
8 P. Geladi and B. R. Kowalski, *Anal. Chim. Acta*, 185 (1986) 1–17.
9 S. Wold, C. Albano, W. J. Dunn, III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, and M. Sjöström, in B. R. Kowalski (Ed.), *Proceedings of the NATO Advanced Study Institute on Chemometrics*, Reidel, Dordrecht, The Netherlands, 1983, pp. 17–95.
10 G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, Wiley Interscience, New York, 1978, pp. 291–433.
11 W. J. Dunn, III, S. Wold, U. Edlund, S. Hellberg, and J. Gasteiner, *Quant. Struct. Act. Relat.*, 3 (1984) 131–137.
12 B. Harath, P. Geladi, and C. Albano, *Filtr. Sep.*, (1989) 53–56.
13 Y. W. Ha and R. L. Thomas, *J. Food Sci.*, 53 (1988) 574–577.