Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Polymers 72 (2008) 730–739
Carbohydrate Polymers
www.elsevier.com/locate/carbpol

Chemometric prediction of alginate monomer composition: A comparative spectroscopic study using IR, Raman, NIR and NMR

Tina Salomonsen [a,b,*], Henrik Max Jensen [b], Dorthe Stenbæk [b], Søren Balling Engelsen [a]

[a] *University of Copenhagen, Faculty of Life Sciences, Department of Food Science, Quality & Technology, Rolighedsvej 30, 1958 Frederiksberg C, Denmark*
[b] *Danisco A/S, Edwin Rahrs Vej 38, 8220 Brabrand, Denmark*

Received 14 August 2007; received in revised form 23 October 2007; accepted 25 October 2007
Available online 7 November 2007

Abstract

The potential of using infrared (IR), Raman and near infrared (NIR) spectroscopy combined with chemometrics for reliable and rapid determination of the ratio of mannuronic and guluronic acid (M/G ratio) in commercial sodium alginate powders has been investigated. The reference method for quantification of the M/G ratio was solution-state ^{1}H nuclear magnetic resonance (NMR) spectroscopy. For a set of 100 commercial alginate powders with a M/G ratio range of 0.5–2.1 quantitative calibrations using partial least squares regression (PLSR) were developed and compared for the three spectroscopic methods. All three spectroscopic methods yielded models with prediction errors (RMSEP) of 0.08 and correlation coefficients between 0.96 and 0.97. However, the model based on extended inverted signal corrected (EISC) Raman spectra stood out by only using one PLS component for the prediction. The results are comparable to that of the experimental error of the reference method estimated to be between 0.01 and 0.08.

Keywords: Alginate; Spectroscopy; Chemometrics; NMR; IR; Raman; NIR

1. Introduction

Commercial alginates are extracted from brown seaweed (*Phaeophyceae*) and used as thickeners, stabilisers and gelling agents in the food and pharmaceutical industries. Chemically, they are a family of binary copolymers of (1–4) linked β-D-mannuronic acid (M) and its C-5 epimer α-L-guluronic acid (G). Alginates are typically described by their M/G ratio, distribution of M- and G-units along the chain and average molecular weight, since these parameters are closely related to the functionality of the alginates, i.e. solubility, interaction with metals, gel properties and viscosity (Haug, Myklesta, Larsen, & Smidsrød, 1967; Steginsky, Beale, Floss, & Mayer, 1992; Stokke, Smidsrød, Bruheim, & Skjåk-Bræk, 1991). The composition and sequential structure and thereby also the functionality of alginates vary according to season, age of population, species and geographic location (Haug, Larsen, & Smidsrød, 1974; Indergaard, Skjåk-Bræk, & Jensen, 1990; Stockton, Evans, Morris, Powell, & Rees, 1980). Because of these variations within the different alginate types, it is of great importance for the alginate industry to obtain detailed knowledge about the composition of its products in order to design the functionality of the final product.

Solution-state nuclear magnetic resonance (NMR) spectroscopy is a highly effective tool in compositional and structural analysis of alginates (Grasdalen, 1983; Grasdalen, Larsen, & Smidsrød, 1977, 1979, 1981; Penman & Sanderson, 1972). However, at the concentrations required for a good signal-to-noise ratio in the NMR analysis, alginate solutions are too viscous to give well-resolved spectra even

* Corresponding author. Address: University of Copenhagen, Faculty of Life Sciences, Department of Food Science, Quality & Technology, Rolighedsvej 30, 1958 Frederiksberg C, Denmark. Tel.: +45 8943 5285/+45 3533 3510; fax: +45 8925 1077.
E-mail addresses: tina.salomonsen@danisco.com, tisa@life.ku.dk (T. Salomonsen).

0144-8617/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carbpol.2007.10.022

at elevated temperatures (up to 90 °C). To reduce the viscosity, it is necessary to partially degrade the alginate chain by a mild acid hydrolysis prior to NMR analysis. This hydrolysis is relatively time-consuming and labour intensive. Hence, solution-state NMR of alginates is rather elaborate and time-consuming and certainly not suitable for at- or on-line monitoring of alginate composition. Thus, development of a faster and simpler method for characterisation of the alginate composition would be beneficial.

Vibrational spectroscopy, e.g. infrared (IR), Raman and near infrared (NIR) spectroscopy are, in combination with multivariate data analysis (chemometrics), useful analytical tools that can reveal detailed information concerning the composition and properties of material at a molecular level. Some of the advantages of these techniques compared to solution-state NMR are that they are rapid, non-destructive, easy to operate and in most cases require no sample preparation. Because of these qualities all three vibrational spectroscopic techniques are applicable for at-line analysis. However, only Raman and NIR spectroscopy can be implemented directly on the production line using optical quartz fibres for on-line analysis.

IR spectroscopy has proven useful for quantitative estimation of the M/G ratio of alginates (Filippov & Kohn, 1974; Mackie, 1971; Sakugawa, Ikeda, Takemura, & Ono, 2004). In these studies, it was found that the ratio of absorption band intensities at 1290 and 1320 cm^{-1} (Filippov & Kohn, 1974), 808 and 787 cm^{-1} (Mackie, 1971) and 1030 and 1080 cm^{-1} (Sakugawa et al., 2004) in the IR spectra of KBr tablets or films containing alginate gives a fairly good estimation of the M/G ratio. IR spectroscopy has also been used for determination of alginate concentration in solution (Bociek & Welti, 1975) and for identification of the type of polysaccharide (alginate, carrageenan or agar) derived from different seaweeds (Pereira, Sousa, Coelho, Amado, & Ribeiro-Claro, 2003). The literature concerning the analysis of alginate using Raman and NIR spectroscopy is sparse. Pereira et al. (2003) showed that Raman spectroscopy can be used to differentiate alginate from other seaweed polysaccharides (agar and carrageenan) and Horn, Moen, and Østgaard (1999) reported on the use of NIR spectroscopy to determine the alginate content of *Laminaria hyperborea* stipe during biodegradation. To the best of our knowledge, determination of the M/G ratio by Raman and NIR spectroscopy has not been carried out before.

Investigations of other polysaccharides have shown that the combination of vibrational spectroscopy and chemometrics is successful when the aim is to develop rapid, non-destructive and robust methods for monitoring different structural characteristics of polysaccharides (Dolmatova, Ruckebusch, Dupuy, Huvenne, & Legrand, 1998; Dyrby et al., 2004; Engelsen & Nørgaard, 1996). It has been shown that FT-IR, FT-Raman and NIR can be used for determination of the degree of esterification and amidation in pectins (Engelsen & Nørgaard, 1996) and of the content of different types of carrageenan in carrageenan powders (Dyrby et al., 2004) using partial least squares regression (PLSR). Moreover, FT-IR spectra of modified corn starches have been used for classification and recognition of their modifications using artificial neural network (ANN) processing (Dolmatova et al., 1998).

The purpose of this work was to develop a rapid quantitative method for measuring the M/G ratio of sodium alginate powder samples using vibrational spectroscopy and chemometrics. The method was developed using solution-state ^{1}H NMR spectroscopy on hydrolysed samples as the reference method.

2. Materials and methods

2.1. The samples

For this spectroscopic investigation, 100 different commercial sodium alginate samples were kindly provided by Danisco A/S (Brabrand, Denmark) as powders with a particle size of <106 μm. Chemical and physical properties of the samples were obtained from the analyses described below.

2.2. NMR spectroscopy

^{1}H NMR spectra were acquired on a Varian Mercury VX 200 MHz spectrometer (Varian, Inc., Palo Alto, California, USA). The average molecular weights of the alginates were reduced by partial acid hydrolysis (Grasdalen et al., 1979). The hydrolysed alginate was dissolved in D_2O (3.5% (w/v)), neutralised (pH 7) and filtered through a Whatman 13 mm syringe filter (pore size 0.45 μm) into a 5-mm NMR tube. 256 scans were acquired and averaged with a 5 s acquisition time after a 2 s relaxation delay and a 90° pulse. During acquisition the sample temperature was 90 °C. A pulse experiment with water suppression was used. The raw data were multiplied with a 1.0 Hz exponential line broadening function before Fourier transformation. Peak assignments and calculations of M/G ratio and the fractions of diad sequences (MM, GG and GM/MG) were performed as described by Grasdalen et al. (1979).

2.3. Water content

The water content of the alginate powders was determined as percent evaporated water after 2 h at 130 °C. Each sample was measured as true duplicates, and the average values were used in the data analysis.

2.4. Gel strength

For gel strength measurements (i.e. the force required to break the gel), homogeneous alginate gels were prepared by dissolving alginate (0.7% (w/w)) in demineralised water at room temperature. $CaHPO_4$ (0.15% (w/w)) was added as an inactivated form of calcium followed by addition of

the slowly hydrolysing D-glucono-δ-lactone (1.4% (w/w)), resulting in an *in situ* release of calcium cations from $CaHPO_4$. Subsequently, the solutions were filled into two 100 ml glass beakers and left at room temperature for 2 h. The gel strength of the alginate gels was measured using a Texture Analyser TX-XT2i (Stable Microsystems, Surrey, UK) mounted with a half inch cylindrical probe which penetrated 20 mm into the gel. A force–time curve was obtained at a crosshead speed of 0.5 mm/s. The gel strength was evaluated as the maximum resistance to the probe, i.e. the height of the force peak.

2.5. Viscosity

The apparent viscosity of alginate solutions (1% (w/w)) was measured using a Brookfield LVT viscosimeter (Brookfield, Essex, UK). The solutions were prepared with demineralised water and left to equilibrate for 2 h at 20 °C before measurement. The viscosity was read after 30 s at 60 rpm. LV1, LV2 and LV3 spindles were used for samples with a viscosity of <100, 100–500 and >500 cP, respectively.

2.6. Molecular weight

The average molecular weight (M_w) of the alginate samples was determined by size exclusion chromatography with multi-angle light scattering (SEC-MALS) using two columns (PSS SUPREMA-LUX 3000 Å and PSS SUPREMA-LUX 1000 Å, Polymer Standard Service GmbH, Mainz, Germany) in series as well as a DAWN EOS multiangle light scattering photometer and an Optilab rEX differential refractive index detector (both Wyatt Technology Corp., Santa Barbara, California, USA). The columns and the detectors were thermostated at 40 °C. The mobile phase was 0.05 M $LiNO_3$/200 ppm NaN_3 at a flow rate of 0.8 ml/min. Alginate was dissolved in the mobile phase (0.2% (w/v)) and 100 μl of this solution was injected into the columns. Data were analysed using Astra V software (Wyatt Technology Corp., Santa Barbara, California, USA). The refractive increment value (dn/dc) for alginate was set to 0.154 ml/g (Mackie, Noy, & Sellen, 1980). Each sample was measured as a true duplicates, and the average values were used in the data analysis.

2.7. FT-IR spectroscopy

IR spectra were collected on a Perkin Elmer Spectrum One FT-IR spectrometer (Perkin Elmer Instruments, Waltham, Massachusetts, USA). Spectra were acquired using an attenuated total reflectance (ATR) device equipped with a single-bounce diamond crystal. A total of 16 scans were averaged for each sample and the resolution was 4 cm^{-1}. The spectra were ratioed against a single-beam spectrum of the clean ATR crystal and converted into absorbance units. Data were collected in the range 4000–650 cm^{-1}. Each sample was measured three duplicates times and the average values were used in the data analysis.

2.8. FT-Raman spectroscopy

Raman spectra were collected on a Perkin Elmer System NIR FT-Raman interferometer (Perkin Elmer Instruments, Waltham, Massachusetts, USA) equipped with a Nd:YAG laser emitting at 1064 nm with a laser power of 200 mW. Data were collected using an InGaAs detector and stored as Raman shifts in the range 3600–200 cm^{-1}. A 180° back-scattering arrangement was used and no correction for the spectral response was applied. A total of 32 scans were averaged for each sample and the resolution was 32 cm^{-1}. Each sample was measured in two re-packed replicates, and the average values were used in the data analysis.

2.9. NIR spectroscopy

NIR spectra were collected using a NIR systems spectrometer model 6500 (Foss NIRSystem, Silver Springs, USA) in reflectance mode. The samples were packed in a small ring cup (internal diameter, 3.9 cm; depth, 1 cm) with a quartz window and measured using a spinning sample module. The range 1100–2498 nm was acquired using a lead sulphide detector. The angle of incident light was 180° and reflectance was measured at an angle of 45°. The number of scans was 32 and the interval between spectra data points was 2 nm. The NIR reflectance spectra were converted to $\log(1/R)$ units prior to data analysis using an internal ceramic reference. Each sample was measured in two re-packed replicates and the average values were used in the data analysis.

2.10. Multivariate data analysis

The results were evaluated by multivariate data analysis using principal component analysis (PCA) and partial least squares regression (PLSR).

PCA (Hotelling, 1933) is the primary tool for investigation of large bilinear data structures for the study of trends, groupings and outliers. By means of PCA it is possible to find the main variation in a multidimensional data set by creating new linear combinations from the underlying latent structures in the raw data. The two-dimensional data matrix (samples × variables) is decomposed into systematic variation and noise. The systematic variation is described by the principal components (PC1, PC2, etc.), each representing the outer product of scores and loadings. The scores contain information about the samples while the loadings contain information about the variables.

PLSR (Wold, Martens, & Wold, 1983) is a multivariate calibration method by which two sets of data, X (e.g. spectra) and y (e.g. M/G ratio), are related by means of regression. The purpose of PLSR is to establish a linear model which enables the prediction of y from the measured spec-

trum X. PLSR is a two-block regression method, where the decomposition of X is performed under the consideration of y in a simultaneous analysis of the two data sets. Three different IR, Raman and NIR data sets were created prior to calibration: (1) raw spectra, (2) second derivative spectra (Savitzky–Golay with 15 point gap size) (Savitzky & Golay, 1964) and (3) spectra treated with extended inverted signal correction (EISC) (Martens, Nielsen, & Engelsen, 2003; Pedersen, Martens, Nielsen, & Engelsen, 2002). All data sets were mean centred (i.e. the mean of each variable was subtracted from the original measurement) prior to model development. In order to determine the important regions in the spectra with respect to prediction of the M/G ratio and to test the possibility of developing optimised local PLSR models using fewer variables, interval partial least squares regression (iPLSR) (Nørgaard et al., 2000) models were developed. iPLSR is an extension of PLSR, which develops PLSR models on equidistant subintervals of the full-spectrum region.

The relatively high number of samples available allowed for true test set validation. The calibration models were validated using a test set consisting of every fourth sample which were selected prior to development of the calibration model, but after the samples were sorted according to M/G ratio. Models were thus developed using internally segmented cross-validation (10 segments) on the calibration set (75 samples) and afterwards applied and tested using the independent test set (25 samples). The PLSR results are presented as number of PLSR components (# PC), squared correlation coefficients (r^2), root mean square error of cross-validation (RMSECV) and root mean square error of prediction (RMSEP) (test set validations).

The PCA and PLSR analyses were performed using LatentiX version 1.0 (Latent5, Copenhagen, Denmark, http://www.latentix.com), whereas EISC pre-processing and iPLSR regression were performed in MatLab 7.4 (Mathworks, Natick, Massachusetts, USA) using the command line EMSC/EISC toolbox (http://www.models.life.ku.dk) and the PLS-Toolbox 4.02 (Eigenvector Research, Manson, Washington, USA), respectively.

3. Results and discussion

3.1. The samples

The range and average values of the chemical and physical properties of the 100 different sodium alginates are listed in Table 1. Since the main purpose of this work was to find correlations between the M/G ratio and spectra obtained by vibrational spectroscopy, the samples were selected so that they varied as much as possible with respect to the M/G ratio. The samples spanned a M/G ratio range between 0.5 and 2.1. As described in Section 2, the relatively high number of samples available allowed for true test set validation of the calibration models. Fig. 1 shows the M/G ratio of the 75 calibration samples and 25 test samples ordered according to the M/G ratio. The samples

Table 1
The range, average values and average standard deviations (std.) of the chemical and physical properties of the 100 different sodium alginate samples

Chemical/physical property	Range	Average	Average std.
M/G ratio	0.5–2.1	1.3	0.06[a]
Fraction of GG (%)	9–44	23	1.0[a]
Fraction of MM (%)	10–49	34	1.8[a]
Fraction of MG/GM (%)	34–51	43	2.4[a]
Water content (%)	11.0–17.8	14.9	0.18[b]
Gel strength (g)	70–1139	566	15.5[c]
Viscosity (cP)	23–846	269	12.3[d]
M_w (kDa)	104–424	277	8.6[b]

[a] Based on true triplicates of nine different samples.
[b] Based on true duplicates of all samples.
[c] Based on the results from two gels of all alginates made from the same solution of the respective alginate.
[d] Based on measuring the viscosity of the same sample six times.

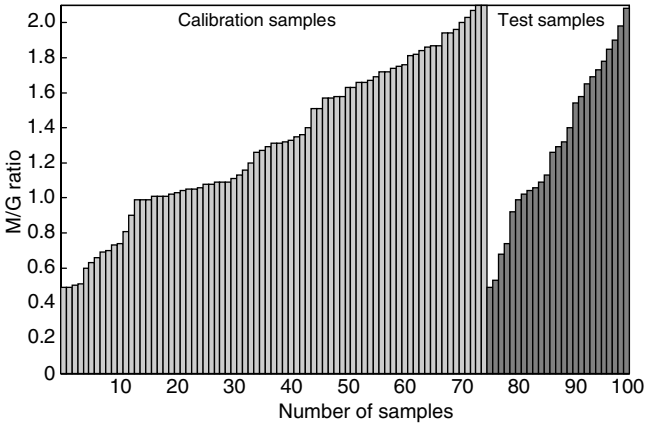

Fig. 1. M/G ratio of the 75 samples in the calibration set and the 25 samples in the test set.

did not only vary with respect to the M/G ratio, but also with respect to the other chemical and physical properties, indicating that we are dealing with a realistic data set representing the many variations found in commercial alginate samples. To investigate the variance structure in the chemical and physical data, a PCA was performed on the autoscaled data (mean centred + dividing each variable by its standard deviation) (Fig. 2). The PCA loading plot in Fig. 2a shows how the different chemical and physical parameters are correlated. The PCA score plot in Fig. 2b shows that the samples in both the calibration sample set and the test sample set spanned the variation within all the chemical and physical parameters.

3.2. The NMR reference method

Solution-state ^{1}H NMR was used as reference method for determination of the M/G ratio. The signals in the anomeric region of the ^{1}H NMR spectra (A (5.07 ppm), B (4.70 ppm) and C (4.46 ppm) in Fig. 3) contain specific information about the alginate composition and were assigned to the anomeric proton of G (A), the anomeric

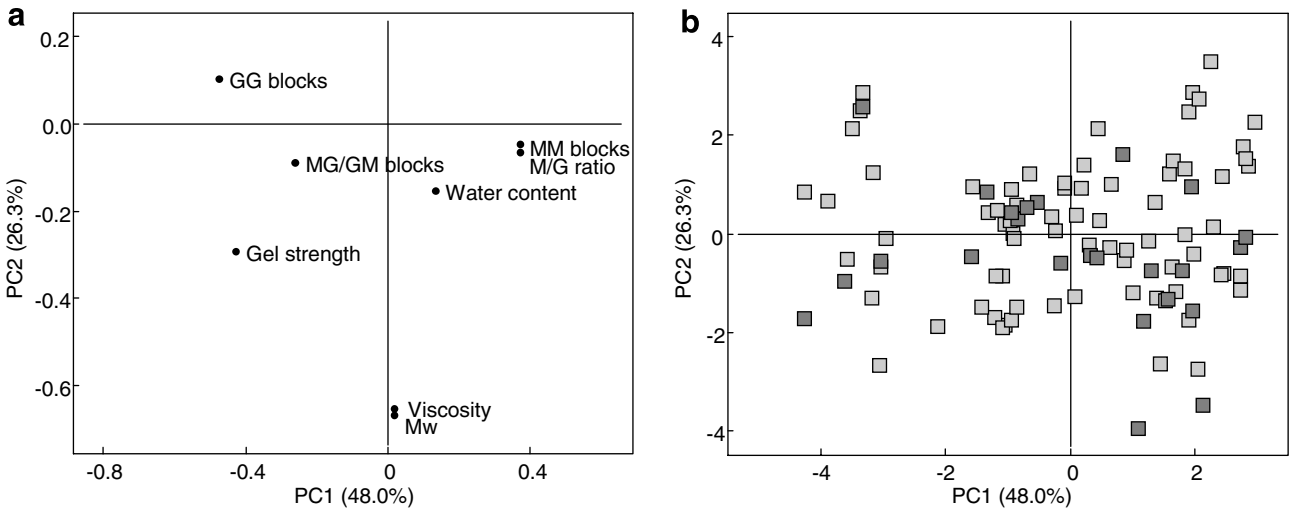

Fig. 2. PCA loading (a) and score (b) plot based on the chemical and physical properties of the 75 calibration (□) and 25 test (■) samples showing the first two principal components which explain 74.3% of the data variation.

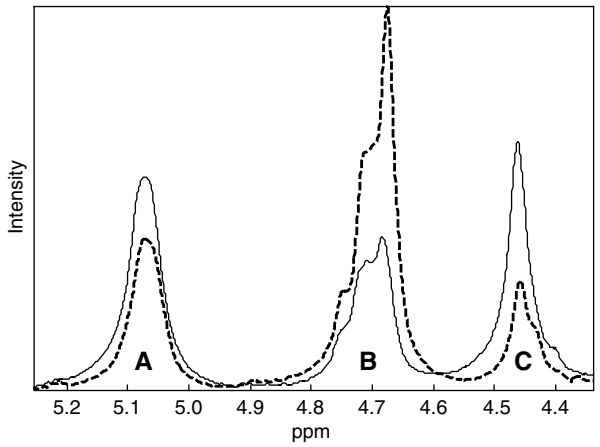

Fig. 3. ^{1}H NMR spectra of 3.5% (w/v) solutions of two hydrolysed sodium alginate samples with M/G ratio of 0.5 (—) and 2.1 (- -).

proton of M and H-5 of G-units adjacent to M (B), and H-5 of G-units adjacent to G (C) (Grasdalen, 1983; Grasdalen et al., 1979; Penman & Sanderson, 1972). Thus, from the relative areas of the three signals in this region the M/G ratio can be calculated, as described by Grasdalen et al. (1979). Furthermore, the fraction of the diad sequences (GG, MM and MG/GM) can be determined (Grasdalen et al., 1979), and if using a NMR spectrometer operating at 400 MHz or above, the fraction of triad frequencies can also be calculated due to the better peak separation at higher frequencies (Grasdalen, 1983).

In order to obtain comparable values of the M/G ratio, the relative integrals of the three signals were calculated using an in-house built routine where the same integration limits for all the spectra were applied. The limits for signal A, B and C (Fig. 3) were set to 5.18–4.96, 4.82–4.57 and 4.55–4.38 ppm, respectively. Changing these limits yielded slightly different M/G ratio values, emphasising the importance of using the same limits in all the calculations. For noise reduction, the spectra were smoothed using a Savitzky–Golay filter (15 point gap size) (Savitzky & Golay, 1964), resulting in more accurate signal integrations. Thus, using this routine, the ^{1}H spectra gave comparable results for the quantification of the M/G ratio, and hence the compositions of the calibration and validation samples used for PLSR models are based on the quantifications from the ^{1}H spectra.

The error of the quantification of the M/G ratio using the described ^{1}H NMR method was estimated to be between 0.01 and 0.08 (the standard deviation of quantifications of triplicates). The degradation of the alginate chain by acid hydrolysis to reduce the viscosity prior to NMR measurement may possibly contribute to this error, since the rate at which the different linkages are cleaved increases in the order: G–G < M–M < M–G < G–M (Grasdalen et al., 1979; Holme, Lindmo, Kristiansen, & Smidsrød, 2003; Smidsrød, Larsen, Painter, & Haug, 1969). Consequently, G-units and M-units will occur at higher frequencies than expected at the reducing and non-reducing ends, respectively, and the homopolymeric G-blocks will be less degraded than the others. Also, it has been shown that different hydrolysis procedures may result in different M/G ratio values (Penman & Sanderson, 1972; Shinohara et al., 2000). Another issue that must be considered when discussing error sources using solution-state NMR for the analysis of alginates is the possible selective micro-aggregation when divalent cations (e.g. calcium ions) are present. It is well known that calcium ions preferentially bind to G-blocks, resulting in gel formation (Draget et al., 2000; Grant, Morris, Rees, Smith, & Thom, 1973). Thus, the restricted motion of the G-blocks as by increase in viscosity reduces the relaxation times, giving rise to a broadening of the NMR lines beyond the level of detection. However, in this study the samples are expected to contain only small amounts of calcium (<0.25%), which will have a very limited influence on the NMR spectra.

3.3. Exploratory analysis

The raw IR, Raman and NIR spectra of the 75 calibration samples are shown in Fig. 4a, b and c, respectively. The spectra show some additive and multiplicative effects which can be explained by differences in physical properties of the samples (e.g. particle size, packing density and water content) and measurement conditions. The IR spectra were recorded using an attenuated total reflectance device where the alginate powder is pressed manually against a diamond. Thus, the intensity of the spectra depends on the contact to the diamond. In the Raman spectra, scatter effects are observed due to different particle sizes and fluorescence. The scatter effects observed in the NIR spectra are due to different particle sizes as well as differences in the water content of the samples. Conclusively, these differences in the spectra are a result of physical differences between the samples, which are not interesting with respect to the purpose of this work where the chemical composition of the alginates is in focus. Therefore, it is preferable to filter away these effects, so that they do not disturb the analysis. Different pre-processing techniques can be used for this purpose. The results of pre-processing the IR, Raman and NIR spectra with the EISC technique are shown in Fig. 4d, e and f, respectively. The spectra are coloured according to the M/G ratio and from a visual inspection of the EISC-corrected IR and Raman spectra it is clear that there are several bands in the spectra that contain information about the M/G ratio. Even though both IR and Raman spectroscopy are based on measurements of the fundamental molecular vibrations, they do not provide exactly the same information. While IR spectroscopy detects vibrations during which the electrical dipole moment changes, Raman spectroscopy is based on the detection of vibrations during which the electrical polarisability changes. Thus, as a rule of thumb IR is more sensitive to side group vibrations and Raman is more sensitive to skeleton vibrations. Therefore, changes in the M/G ratio will manifest differently. In the IR spectra the intensity of the bands at 1290, 1166, 882 and 814 cm^{-1} increase with increasing M/G ratio and the intensity of the bands at 1320, 1123, 1083, 950 and 899 cm^{-1} decrease with increasing M/G ratio. Filippov and Kohn (1974) also found that the bands at 1290 and 1320 cm^{-1} correlated with the M and G content, respectively. In the Raman spectra the intensity of the bands at 1412, 1090, 955 and 708 cm^{-1} increase with increasing M/G ratio and the intensity of the bands at 1313, 1232, 884 and 806 cm^{-1} decrease with

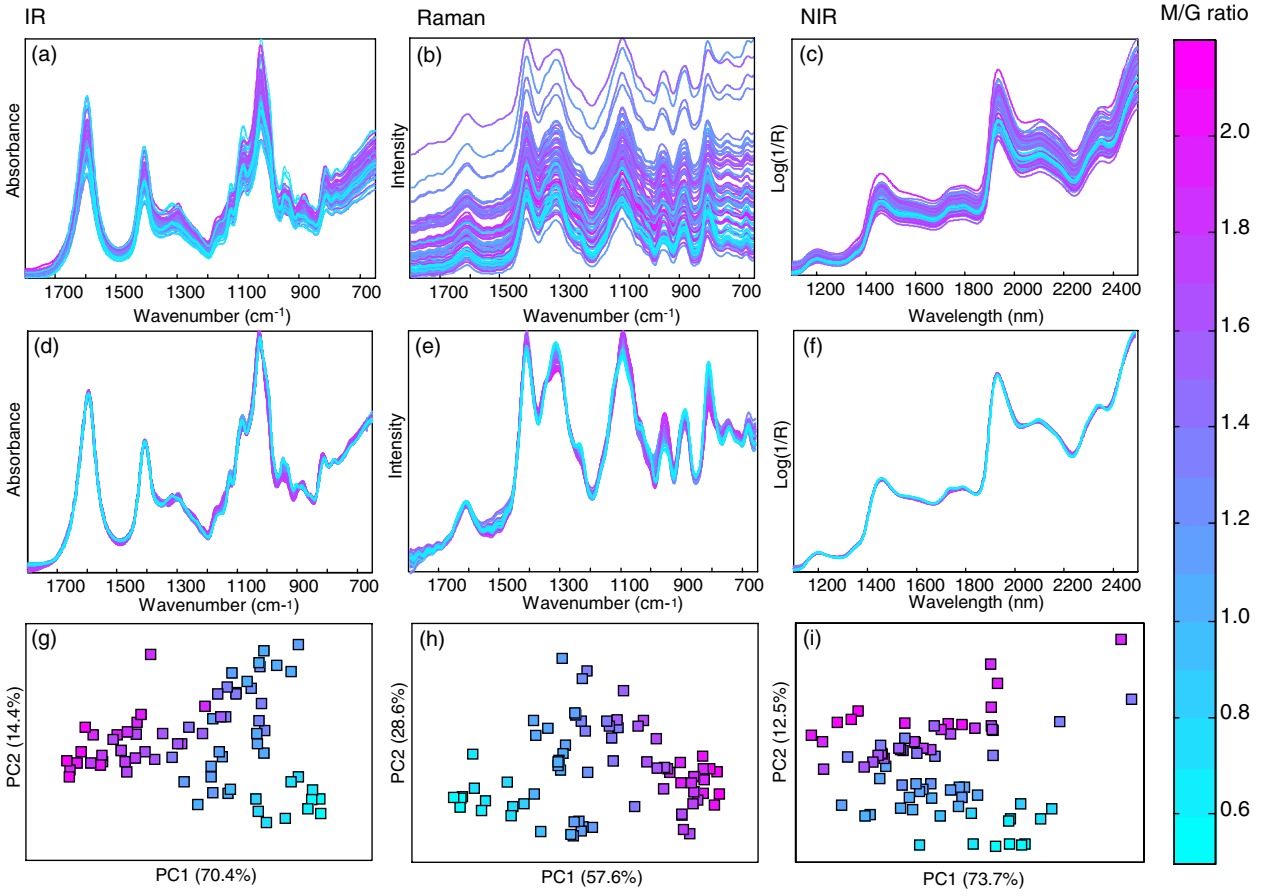

Fig. 4. Raw IR (a), Raman (b) and NIR (c) spectra as well as EISC-corrected IR (d), Raman (e) and NIR (f) spectra of the 75 calibration samples. PCA score plot of the EISC-corrected IR (g), Raman (h) and NIR (i) spectra showing the first two principal components (PCs) which explain 84.8%, 86.2% and 86.2%, respectively, of the data variation. The spectra and the scores are coloured according to the M/G ratio.

increasing M/G ratio. Thus, in the IR and Raman spectra the bands related to the concentration of M and G are all found in the fingerprint region, whereas the asymmetric stretching band from the carboxylate ion centred around 1600 cm^{-1} is not related to the M/G ratio.

In contrast to IR and Raman, NIR spectroscopy measures over- and combination-tones predominantly of the fundamental vibrations that involve hydrogens. Thus, when the variations amongst the samples do not involve hydrogens, the NIR spectra appear almost similar. This explains why the curvature of all the NIR spectra is so alike. However, a slight intensity increase in the 1700–1800 nm region (first overtone C–H stretching modes) and at 2250 nm (combination bands) for samples with low and high M/G ratio, respectively, can be observed in the EISC-corrected spectra (Fig. 4f). Since the spectra are so alike, multivariate data analysis techniques are required to release these highly overlapped and almost similar spectral features related to the M/G ratio.

To obtain an overview of the variation in the multivariate spectra before the development of calibration models, PCA was used to investigate the extent to which the spectroscopic methods allowed for differentiation between alginates with different M/G ratios. The PCA score plot based on the EISC-corrected IR, Raman and NIR spectra of the 75 calibration samples are shown in Fig. 4g, h and i, respectively. The samples are coloured according to the M/G ratio, clearly showing that the main variation in the EISC-corrected IR (Fig. 4g) and Raman (Fig. 4h) spectra are correlated with differences in the M/G ratio of the alginate samples, whereas the main variation of the NIR spectra cannot be explained by the variation in M/G ratio. NIR absorbance is strongly influenced by the different contents of water (Table 1) in the samples. Thus, PC1 from the PCA of the EISC-corrected NIR spectra is modelling the water content (Fig. 4i). However, the variations in M/G ratio are explained along PC2, indicating that there is information about the M/G ratio in the NIR spectra.

In conclusion, the spectra obtained by the three different vibrational techniques all contain information about the M/G ratio. Thus, this preliminary exploratory analysis indicates that it is possible to develop PLSR models with a good prediction performance.

3.4. Calibration models

Calibration models for prediction of the M/G ratio from the raw, second derivative and EISC-corrected IR, Raman and NIR spectra of the alginate powders were developed using PLSR. The models were developed using the 75 calibration samples and tested using the 25 test set samples. For IR and Raman data only the range 1800–650 cm^{-1} was used in the calibrations in order to avoid the noisy low energy range below 650 cm^{-1} and the strong O–H stretching modes in the IR spectra centred around 3300 cm^{-1}. For NIR the whole recorded spectral range 1100–2498 nm was used.

The PLSR models obtained using the three different spectroscopic techniques and different pre-processing techniques are listed in Table 2. Using the raw spectra, the models performed relatively well compared to the upper limit of the experimental error of the M/G ratio measurements, which was 0.08. However, relatively many PLSR components (4–6) were needed in order to deal with the additive and multiplicative effects of the spectra due to the different factors described in the previous section. Using instead the second derivative or EISC-corrected spectra in order to remove these effects and only model the chemical information in the spectra, simpler models with lower number of PLSR components and lower prediction errors were obtained. The models based on the EISC-corrected IR and Raman spectra and the second derivative NIR spectra all resulted in a prediction error of 0.08. However, it is noteworthy that the model based on EISC-corrected Raman spectra only needs one PLSR component to explain the M/G ratio, whereas the best IR and NIR models need four and three, respectively. It is worth noting that the calibration error (RMSECV) and the prediction error (RMSEP) are the same for the best model of all the spectroscopic techniques. This indicates that the models are very robust, i.e. the model can predict new samples that have not been involved in the modelling with high precision. Since it is possible to develop a model with a prediction error comparable to the experimental error of the reference method using all three vibrational spectroscopic techniques, it can be concluded that IR, Raman and NIR are suitable methods for a rapid determination of the M/G ratio of commercial alginate samples. However, it should be noted that the model based on the EISC-corrected Raman spectra is the simplest and therefore probably also more robust when it comes to predicting new samples.

Fig. 5a shows the measured versus predicted plot from the model based on the EISC-corrected Raman spectra ($r^2 = 0.97$). The model is based on only one PLSR component and the prediction error is 0.08. The regression coeffi-

Table 2
Results of PLSR models based on raw, second derivative and EISC-corrected IR, Raman and NIR spectra of sodium alginate powder for prediction of the M/G ratio

	Pre-processing	#PC	r^2	RMSECV	RMSEP
IR	Raw	5	0.95	0.10	0.09
	Second derivative	3	0.95	0.10	0.10
	EISC	4	0.97	0.08	0.08
Raman	Raw	4	0.93	0.12	0.12
	Second derivative	3	0.94	0.11	0.12
	EISC	1	0.97	0.08	0.08
NIR	Raw	6	0.96	0.09	0.11
	Second derivative	3	0.96	0.08	0.08
	EISC	3	0.95	0.10	0.13

The models were developed using the 75 calibration samples and the prediction ability was tested using the 25 test samples. The results are presented as number of PLSR components (#PC), squared correlation coefficients (r^2), root mean square error of cross-validation (RMSECV) and root mean square error of prediction (RMSEP) (test-set validations).

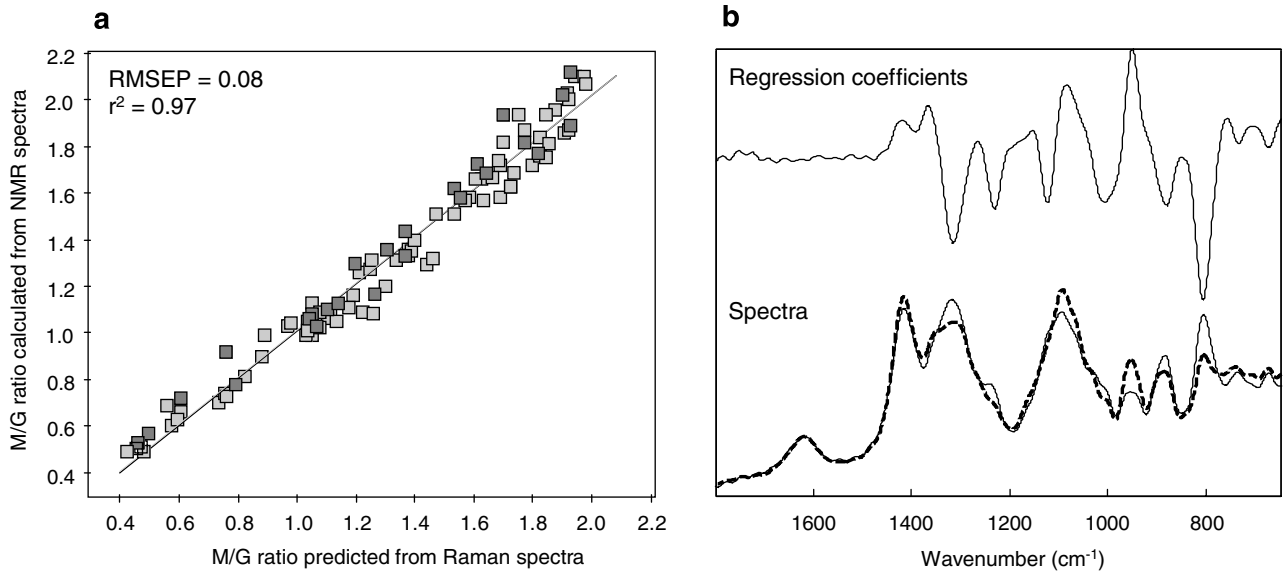

Fig. 5. (a) M/G ratio measured by NMR versus M/G ratio predicted from the EISC-corrected Raman spectra of the 75 calibration (□) and 25 test (■) samples using one PLSR component. (b) The regression coefficients are plotted together with the spectra of two alginates with M/G ratios of 0.5 (—) and 2.1 (- -).

cients shown together with the spectra of a sample with low (0.5) and high (2.1) M/G ratio tell which parts of the spectra that are important for the prediction of the M/G ratio (Fig. 5b). As expected from the visual inspection of the EISC-corrected spectra (Fig. 4e), the carbonyl stretching region at 1607 cm^{-1} does not contribute to the prediction. The most influential spectral band is the band at 806 cm^{-1}, which most likely arises from the α-configuration of the G-units. The second most influential spectral band is centred at 955 cm^{-1}. Actually, the ratio between the intensities 955 cm^{-1} (M) and 806 cm^{-1} (G) correlated well ($r^2 = 0.95$) to the M/G ratio calculated from the NMR spectra. Moreover, the ratio between the intensities of 1290 cm^{-1} (M) and 1320 cm^{-1} (G) in the EISC-corrected IR spectra correlated very well ($r^2 = 0.96$) to the M/G ratio, as also found by Filippov and Kohn (1974). While these bi-variate methods are almost as good as the multivariate calibration methods, they are more sensitive to noise and interferences and will thus be less robust for industrial quality control.

3.5. Variable selection using interval PLSR (iPLSR)

Variable selection by interval PLSR (iPLSR) was applied to encircle the significant regions in the Raman spectra with respect to the M/G ratio and to test the possibility of developing optimised local PLSR models for prediction of the M/G ratio using fewer variables. The iPLSR algorithm is designed to develop local PLSR models on spectral subintervals of equal width of the full-spectrum region. The advantage of the iPLSR approach is that smaller intervals will contain less interference and thus result in simpler, more precise and more easily interpretable models (Nørgaard et al., 2000).

The iPLSR approach was applied to the EISC-corrected IR and Raman spectra and second derivative NIR spectra due to the good performance of the full-spectrum models based on these data sets. Different numbers of variables in the spectral intervals were tested (5, 10, 25 and 50) and the number of components in the local models was restricted to be equal to or less than the optimal number of components in the global model (Table 2). The results of iPLSR of the EISC-corrected IR and Raman of the 75 calibration samples with 50 variables in each interval (23 intervals) are presented in Fig. 6a and b, respectively, together with the spectra of two samples with extreme M/G ratio values. The models were validated with segmental cross-validation using 10 segments. The calibration errors (RMSECV) for each of the 23 intervals are presented as bars and the calibration error of the global model is presented as a horizontal dotted line. All IR models were calculated using four PLSR components with exception of the model based on the spectral region 1150–1100 cm^{-1} (C–O–C pyranose skeleton modes), which is based on only three PLSR components. Thus, the model based on the IR spectral region 1150–1100 cm^{-1} is simpler and has approximately the same calibration error (RMSECV) as the full-spectrum model. Also, the models based on the spectral region 1350–1300 cm^{-1} yield a quantitative performance equal to the full-spectrum model. However, this model uses four PLSR components as the full-spectrum model. All Raman models were calculated using one PLSR component and only one spectral region (850–800 cm^{-1}) containing the band due to the α-configuration of the G-units, resulted in a calibration error equal to the error of the full-spectrum model. This region was also identified as the most influential spectral band in the inspection of the regression coefficients (Fig. 5b) and could be identified by visual inspection of the spectra (Fig. 4e) as an important region for differentiating between alginates with high and low M/G ratio. From iPLSR of the second derivative

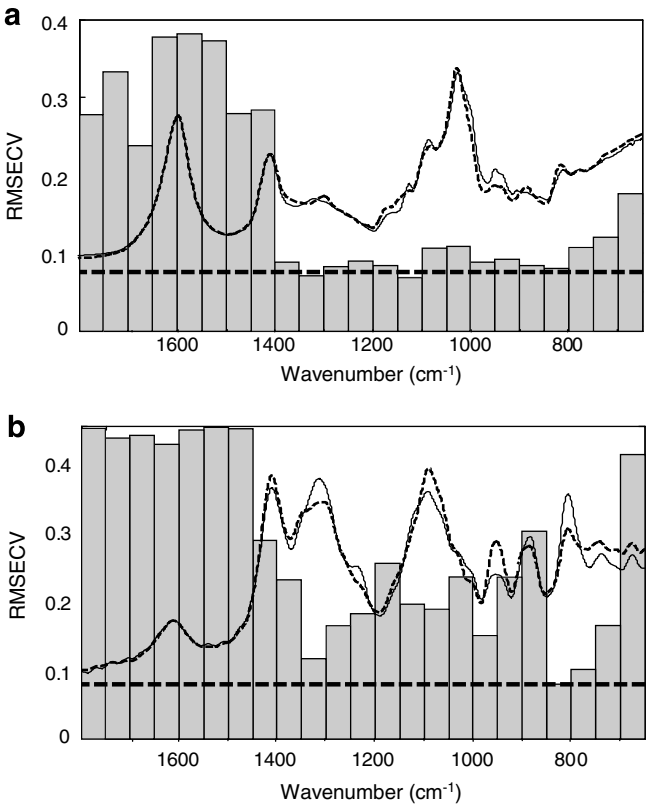

Fig. 6. Interval PLSR results. Cross-validated prediction performance (RMSECV) for 23 PLSR interval models (bars) and for the full-spectrum model (- -) based on the EISC-corrected IR (a) and Raman (b) spectra of the 75 calibration samples plotted together with the spectra of two alginates with M/G ratios of 0.5 (—) and 2.1 (- -). All IR models use four PLSR components with exception of the model based on the spectral region 1150–1100 cm^{-1} using three PLSR components. All Raman models are based on one PLSR component.

NIR spectra the regions 1750–1800 nm performed the same as the full-spectrum model using the same number of PLS components.

The regions in the IR, Raman and NIR spectra identified above using intervals containing 50 variables were also identified as the most important regions using intervals with 5, 10 and 25 variables. However, when testing the prediction performance of the models based on the smaller intervals using the 25 test samples, it resulted in higher prediction errors (0.09–0.11) compared to the full-spectrum model (0.08). Only the model based on the IR spectral region 1150–1100 cm^{-1} using three PLSR components resulted in a prediction error lower than the full-spectrum model when tested (0.08).

In summary, iPLS has been used to show and confirm the relevant areas of the spectra. However, it was not possible to develop models based on fewer variables with a better prediction performance than the full-spectrum model.

4. Conclusion

The present investigation has shown that it is possible to develop a rapid, non-destructive and robust quantitative method for measuring the monomer composition (M/G ratio) of commercial alginate powders using vibrational spectroscopy (IR, Raman and NIR) and chemometrics. The M/G ratio can be predicted from spectra of all three spectral sources investigated with an error comparable to that of the solution-state ^{1}H NMR reference method. This is a valuable achievement, since vibrational spectroscopic techniques have a high industrial potential for at- or on-line quality control as well as screening of large numbers of samples.

Spectral pre-processing was applied in order to remove the additive and multiplicative effects not related to the chemical composition of the samples. This generally improved the predictive capabilities of the three spectroscopic methods IR, Raman and NIR. Especially the Raman results were improved when pre-processing the spectra using the EISC method. This resulted in a very simple model using only one PLSR component and with a prediction error (RMSEP) of 0.08 and a correlation of 0.97, which is comparable to the standard deviation of the NMR reference method (0.01–0.08).

For an even better fit between the measured and predicted values, the error from the reference method must be lowered, since it is not possible to predict values with better accuracy than that of the reference method. We believe that a reduction in the error from the reference method deserves further attention. Therefore, we are currently looking into the possibility of using the NMR techniques ^{1}H high resolution magic angle spinning (HR MAS) NMR and solid-state ^{13}C cross-polarisation (CP) MAS NMR for determination of the M/G ratio. By using these techniques the destructive and time-consuming hydrolysis procedure necessary for solution-state NMR can be avoided.

Acknowledgements

The authors thank the Ministry of Science, Technology and Innovation for partly sponsoring the Industrial PhD project conducted by Tina Salomonsen in co-operation with Danisco A/S and Quality & Technology, Department of Food Science, Faculty of Life Sciences (LIFE) at the University of Copenhagen. At Danisco, many colleagues have contributed to this work. Special credit is given to Bente Høj Andersen, Keld Hartwig and Anita Beck-Rasmussen. Mette Mortensen and Gilda Kischinovsky from LIFE are also acknowledged for help with parts of the experimental work and for proofreading the manuscript, respectively.

References

Bociek, S. M., & Welti, D. (1975). Quantitative analysis of uronic acid polymers by infrared spectroscopy. *Carbohydrate Research, 42*, 217–226.

Dolmatova, L., Ruckebusch, C., Dupuy, N., Huvenne, J. P., & Legrand, P. (1998). Identification of modified starches using infrared spectroscopy and artificial neural network processing. *Applied Spectroscopy, 52*, 329–338.

Draget, K. I., Strand, B., Hartmann, M., Valla, S., Smidsrød, O., & Skjåk-Bræk, G. (2000). Ionic and acid gel formation of epimerised alginates; the effect of AlgE4. *International Journal of Biological Macromolecules, 27*, 117–122.

Dyrby, M., Petersen, R. V., Larsen, J., Rudolf, B., Nørgaard, L., & Engelsen, S. B. (2004). Towards on-line monitoring of the composition of commercial carrageenan powders. *Carbohydrate Polymers, 57*, 337–348.

Engelsen, S. B., & Nørgaard, L. (1996). Comparative vibrational spectroscopy for determination of quality parameters in amidated pectins as evaluated by chemometrics. *Carbohydrate Polymers, 30*, 9–24.

Filippov, M. P., & Kohn, R. (1974). Determination of composition of alginates by infrared spectroscopic method. *Chemicke Zvesti, 28*, 817–819.

Grant, G. T., Morris, E. R., Rees, D. A., Smith, P. J. C., & Thom, D. (1973). Biological interactions between polysaccharides and divalent cations: The egg-box model. *FEBS Letters, 32*, 195–198.

Grasdalen, H. (1983). High-field ^{1}H spectroscopy of alginate: Sequential structure and linkage conformations. *Carbohydrate Research, 118*, 255–260.

Grasdalen, H., Larsen, B., & Smidsrød, O. (1977). ^{13}C NMR studies of alginate. *Carbohydrate Research, 56*, C11–C15.

Grasdalen, H., Larsen, B., & Smidsrød, O. (1979). NMR study of the composition and sequence of uronate residues in alginates. *Carbohydrate Research, 68*, 23–31.

Grasdalen, H., Larsen, B., & Smidsrød, O. (1981). ^{13}C NMR studies of monomeric composition and sequence in alginate. *Carbohydrate Research, 89*, 179–191.

Haug, A., Larsen, B., & Smidsrød, O. (1974). Uronic acid sequence in alginate from different sources. *Carbohydrate Research, 32*, 217–225.

Haug, A., Myklesta, S., Larsen, B., & Smidsrød, O. (1967). Correlation between chemical structure and physical properties of alginates. *Acta Chemica Scandinavica, 21*, 768–778.

Holme, H. K., Lindmo, K., Kristiansen, A., & Smidsrød, O. (2003). Thermal depolymerization of alginate in the solid state. *Carbohydrate Polymers, 54*, 431–438.

Horn, S. J., Moen, E., & Østgaard, K. (1999). Direct determination of alginate content in brown algae by near infra-red (NIR) spectroscopy. *Journal of Applied Phycology, 11*, 9–13.

Hotelling, H. (1933). Analysis of complex statistical variables into principal components. *Journal of Educational Psychology, 24*, 417–441.

Indergaard, M., Skjåk-Bræk, G., & Jensen, A. (1990). Studies on the influence of nutrients on the composition and structure of alginate in *Laminaria saccharina* (L) Lamour (Laminariales, Phaeophyceae). *Botanica Marina, 33*, 277–288.

Mackie, W. (1971). Semi-quantitative estimation of composition of alginates by infra-red spectroscopy. *Carbohydrate Research, 20*, 413–415.

Mackie, W., Noy, R., & Sellen, D. B. (1980). Solution properties of sodium alginate. *Biopolymers, 19*, 1839–1860.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry, 75*, 394–404.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy, 54*, 413–419.

Pedersen, D. K., Martens, H., Nielsen, J. P., & Engelsen, S. B. (2002). Near-infrared absorption and scattering separated by extended inverted signal correction (EISC): Analysis of near-infrared transmittance spectra of single wheat seeds. *Applied Spectroscopy, 56*, 1206–1214.

Penman, A., & Sanderson, G. R. (1972). Method for determination of uronic acid sequence in alginates. *Carbohydrate Research, 25*, 273–282.

Pereira, L., Sousa, A., Coelho, H., Amado, A. M., & Ribeiro-Claro, P. J. A. (2003). Use of FTIR, FT-Raman and ^{13}C NMR spectroscopy for identification of some seaweed phycocolloids. *Biomolecular Engineering, 20*, 223–228.

Sakugawa, K., Ikeda, A., Takemura, A., & Ono, H. (2004). Simplified method for estimation of composition of alginates by FTIR. *Journal of Applied Polymer Science, 93*, 1372–1377.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*, 1627–1639.

Shinohara, M., Nishida, R., Aoyama, T., Kamono, H., Bando, H., & Nishizawa, M. (2000). Comparison of guluronate contents of alginates determined by different methods. *Fisheries Science, 66*, 616–617.

Smidsrød, O., Larsen, B., Painter, T., & Haug, A. (1969). The role of intramolecular autocatalysis in acid hydrolysis of polysaccharides containing 1,4-linked hexuronic acid. *Acta Chemica Scandinavica, 23*, 1573–1580.

Steginsky, C. A., Beale, J. M., Floss, H. G., & Mayer, R. M. (1992). Structural determination of alginic acid and the effects of calcium-binding as determined by high-field NMR. *Carbohydrate Research, 225*, 11–26.

Stockton, B., Evans, L. V., Morris, E. R., Powell, D. A., & Rees, D. A. (1980). Alginate block structure in Laminaria digitata: Implications for holdfast attachment. *Botanica Marina, 23*, 563–567.

Stokke, B. T., Smidsrød, O., Bruheim, P., & Skjåk-Bræk, G. (1991). Distribution of uronate residues in alginate chains in relation to alginate gelling properties. *Macromolecules, 24*, 4637–4645.

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. *Lecture Notes in Mathematics, 973*, 286–293.