Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Polymers 57 (2004) 337–348
Carbohydrate Polymers
www.elsevier.com/locate/carbpol

Towards on-line monitoring of the composition of commercial carrageenan powders

M. Dyrby[a,*], R.V. Petersen[b], J. Larsen[b], B. Rudolf[b], L. Nørgaard[a], S.B. Engelsen[a]

[a]*Centre for Advanced Food Studies, Food Technology, The Royal Veterinary and Agricultural University, Rolighedsvej 30, 1958 Frederiksberg C, Denmark*
[b]*CP Kelco, Ved Banen 16, 4623 Lille Skensved, Denmark*

Received 28 October 2003; accepted 25 May 2004

Available online 21 July 2004

Abstract

In this study, a quantitative method for measuring the content of iota (ι), kappa (κ), lambda (λ), mu (μ) and nu (ν) carrageenan in powder samples from industrial production was developed using rapid spectroscopic techniques and chemometrics. The reference method for quantification was 600 MHz ^{1}H NMR spectroscopy, which was found superior to 150 MHz ^{13}C NMR due to higher sensitivity. Quantitative calibrations using partial least squares (PLS) regression on each carrageenan form were developed and compared for the three spectroscopic methods investigated, infrared, Raman and near-infrared spectroscopy. A large mixture design yielded near-perfect calibration models for all spectroscopic techniques, the best ones using the newly developed preprocessing method extended inverted signal correction (EISC) on Raman spectra yielding prediction errors (RMSECV) in the range 0.5–2.1% for the five different carrageenan forms. After variable selection using the new method backwards interval-PLS (bi-PLS), the models built on EISC treated Raman spectra of design samples as well as a few production samples yielded good predictions for all samples with RMSECV between 0.7 and 2.2% and correlations between 0.98 and 1.00.

Keywords: Carrageenan; Spectroscopy; Chemometrics; Quantification; IR; Raman; NIR; NMR; PLS; EISC; bi-PLS

1. Introduction

Carrageenans are sulfated polygalactans extracted from red algae (*Rhodophyceae*) mainly of the genus *Chondrus*, *Gigartina* or *Eucheuma*. There are several types of carrageenans, the major ones being iota (ι), kappa (κ) and lambda (λ). Minor types are mu (μ) and nu (ν), which are kappa- and iota-precursors, respectively, and are never found pure but always in coexistence with kappa and iota. Carrageenans are composed of alternating β-1,3-linked and α-1,4-linked D-galactopyranose sugars. The carrageenan type is determined by the amount and position of sulfate groups and anhydro-bridges as shown in Fig. 1. However, carrageenans are heterogeneous polymers consisting of a relatively broad distribution of chain lengths which each are mixed chains of two or more types of carrageenan (Therkelsen, 1993; van de Velde, Peppelman, Rollema, & Tromp, 2001).

In the food industry, carrageenans are widely used as stabilizing, thickening and gelling agents. The different carrageenan types have different functional properties, and therefore it is of great importance for the carrageenan industry to obtain detailed knowledge about the composition of their products in order to be able to design the functionality of the final product.

Nuclear magnetic resonance (NMR) spectroscopy has been a highly effective tool in structural and compositional analysis of carrageenans (Stortz, Bacon, Cherniak, & Cerezo, 1994; Turquois, Acquistapace, Arce Vera, & Welti, 1996; van de Velde et al., 2001). This technique uses the anomeric signal region of the ^{1}H or ^{13}C NMR spectra, where the different carrageenans contribute with a unique chemical shift signal and with an integrated intensity that is correlated to the amount of the individual carrageenan type (van de Velde et al., 2001). NMR measurements are usually performed on solutions of carrageenans and thus require extensive sample preparation due to the high viscosity of carrageenan solutions even at

* Corresponding author. Address: Department of Dairy and Food Science, Food Technology Group, The Royal Veterinary and Agricultural University, Rolighedsvej 30, 2000 Frederiksberg, Denmark. Tel.: +45-3528-3564; fax: +45-3528-3245.
E-mail address: md@kvl.dk (M. Dyrby).

0144-8617/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carbpol.2004.05.015

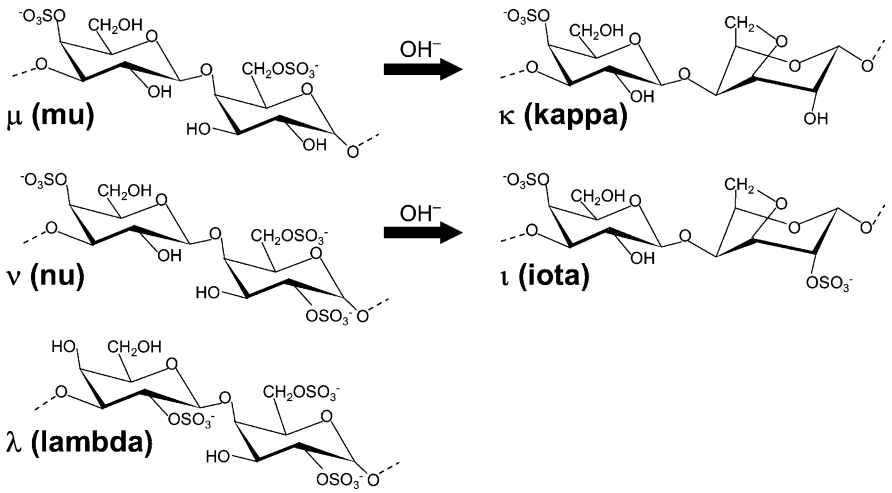

Fig. 1. The structure of the three main carrageenan types iota (ι), kappa (κ) and lambda (λ), and of mu (μ) and nu (ν) carrageenan, the precursors of kappa and iota, respectively.

elevated temperatures. Hence, the method is rather elaborate, time consuming, expensive and not suitable for at- or on-line monitoring of carrageenan composition.

Vibrational spectroscopy is widely used in the food industry and combined with chemometrics it is a very useful analytical tool for at- or on-line monitoring of the quality of food products (Pedersen & Engelsen, 2001). In pectin production, near-infrared has proved capable of at-line monitoring of the degree of esterification and amidation of pectin powders (Engelsen & Nørgaard, 1996). In carrageenan production, a replacement of the NMR reference method, which besides extensive sample preparation requires a strong external magnetic field, with a similar optical spectroscopic method, such as infrared (IR), Raman or near-infrared (NIR) spectroscopy combined with chemometrics would be both cost-effective and timesaving.

A number of structural studies of carrageenans have been made using IR (Chiovitti et al., 1995; Chopin & Whalen, 1993; Jacobsson & Hagman, 1993; Matsuhiro, 1996; Sekkal & Legrand, 1993; Turquois et al., 1996). IR spectroscopy in the ranges 800–850 and 930–1240 cm^{-1} has proven useful for distinguishing between the different main types (iota, kappa and lambda) of carrageenans because vibrational bands in these regions are due to stretching vibrations of the sulfate esters and the 3,6-anhydro ring. Detailed assignment of Raman spectra of carrageenans has also been made (Malfait, Van Dael, & van Cauwelaert, 1998; Sekkal & Legrand, 1993), where significant differences between the different carrageenan types can be observed in the range 1500–100 cm^{-1}. To our best knowledge, no investigations regarding spectral assignment of carrageenans by NIR spectroscopy have previously been reported.

The purpose of this work was to develop a quantitative method for measuring the concentrations of the three major carrageenan forms (iota, kappa and lambda) and the two precursors (mu and nu) in commercial production batches by vibrational spectroscopy combined with chemometrics. Only little research has been made in this area and only using mid-IR spectroscopy (Jacobsson & Hagman, 1993; Prado-Fernandez, Rodriguez-Vazquez, Tojo, & Andrade, 2003) and none directly on the powdered production batches. We set out to compare the quantitative performance of IR spectroscopy with that of Raman and NIR spectroscopy, which both in contrast to IR have the advantage that they can be implemented directly on the production line using optical fibers, thus enabling fast monitoring of the carrageenan content of industrial production batches.

2. Materials and methods

2.1. Production batches

Carrageenan production batches were obtained from CP Kelco (Lille Skensved, Denmark). Five production batches (P1–P5) were used for mixing calibration samples, and another five production batches were used as test samples (P6–P10). The five production batches used for calibration were chosen in order to make a design that contained the highest possible concentration for each of the five carrageenan forms. It was possible to achieve almost 100% pure iota and kappa production batches (P1 and P2). A production batch with approximately 90% lambda was chosen as the maximal amount of lambda carrageenan (P3). The maximum amount of the precursors mu and nu that can be found in production batches is around 10–20% and two samples (P4 and P5) were chosen with high content of mu and nu, respectively. The exact concentration of the different types of carrageenans in the production batches was determined by NMR spectroscopy and the composition of production batches P1–P5 is shown in Table 1. Production batches P1 and P5 originate from *Kappaphycus*

Table 1
The carrageenan composition determined by 600 MHz ^{1}H NMR of the five production batches used for mixing calibration samples

	Iota	Kappa	Lambda	Mu	Nu	Unidentified
P1	85.6	1.6	0.0	0.0	0.4	12.4
P2	4.8	93.9	0.0	0.0	0.0	1.3
P3	0.9	1.8	90.3	0.0	0.0	7.0
P4	11.8	60.0	0.0	21.2	2.1	4.5
P5	72.7	2.4	0.0	0.0	15.1	9.8

alvarezii also known as Cottonii in the industry, P2 and P4 originates from *Eucheuma denticulatum* also known as Spinosum in the industry. Differences are obtained by adjusting the level of alkaline treatment. P3 is made from hand-sorted tetrasporophytic plants of *Gigartina skottsbergii*. No special effort was done to control the counter ion composition.

2.2. Calibration samples

The calibration set was made of the five production batches P1–P5. The production batches were mixed to achieve a simplex-lattice design of degree five, where the concentration of P1–P5 was set to 0, 20, 40, 60, 80 and 100%. After determination of the exact carrageenan composition of P1–P5 using NMR as described below, the concentrations of iota, kappa, lambda, mu and nu in the calibration samples were calculated and these values were used in the PLS regression models.

Being a mixture design and due to the co-existence of mu with kappa and nu with iota the design is constrained, which implies that the components are not fully orthogonal, i.e. do not vary in a completely independent way. A measure of the lack of orthogonality is the condition number, which for this design was calculated to 5.5, which is an acceptable value (Eriksson, Johansson, Kettaneh-Wold, Wikström, & Wold, 2000). The design resulted in 128 calibration samples of which three were center points.

In order to reduce the effects of different moisture content of these highly hygroscopic powders, the carrageenan powders P1–P5 were dried at 60 °C for 4 h to remove all water. Immediately after drying, the carrageenan powders were incubated at constant humidity and when at equilibrium the calibration samples were mixed from these samples. The actual water content of P1–P5 after the incubation was measured and taken into account when calculating the carrageenan concentration of the calibration samples.

The mixing of the samples and the spectral measurements were performed randomized over five days. Ten of the calibration samples were prepared on two different days in order to investigate if spreading the experiment over several days lead to any systematic variance from e.g. instrumental drift or humidity. Furthermore, three true replicates were made on all samples to evaluate the repeatability of the mixing and weighing procedure. None of these investigations showed significant variations between days or repeated mixing, therefore the sample preparation was assumed not to have a significant effect on the spectra. Hence, all PLS models discussed in the results are based on average spectra of the three replicates.

2.3. Nuclear magnetic resonance spectroscopy

^{1}H and ^{13}C NMR spectra were acquired on a 14.1 Tesla Bruker AV600 (Bruker Biospin GMBH, Rheinstetten, Germany) operating at 600 MHz for ^{1}H and 150 MHz for ^{13}C. The carrageenan samples were prepared as 4% (w/w) solutions in D_2O using ultrasound for depolymerisation (van de Velde et al., 2001) and a temperature of 70 °C was used for both ^{1}H and ^{13}C NMR in order to obtain low viscosity solutions. For ^{1}H NMR spectroscopy, 128 scans were acquired and averaged in 1.7 s acquisition time after a 3 s relaxation delay and a 90° pulse with a 9615 Hz sweep width resulting in 16K complex data points. A pulse experiment with water suppression was used due to small amounts of residual water in the powders giving a total experiment time of 10 min. The spectra were internally referenced using TSP ($\delta = 0.0$ ppm). In order to obtain carrageenan compositions, the anomeric proton signals in the range 5.0–5.6 ppm were analyzed using an in-house routine for simultaneous least-squares curve fitting using Lorentzian functions. For ^{13}C NMR spectroscopy, 14,000 scans were averaged with a 2.2 s acquisition time after a 3.7 s relaxation delay and 90° pulse with a 15,015 Hz sweep width resulting in 32K complex data points. The total measurement time was 24 h. For carbon NMR, the anomeric range between 90 and 106 ppm was analyzed using standard integration of specific peaks.

2.4. IR spectroscopy

IR spectra were collected on an Arid-Zone MB100 FT-IR (Bomem, Quebec, Canada) interferometer. Spectra were acquired using an attenuated total reflectance (ATR) device equipped with a triple-bounce diamond crystal (Durascope, SensIR Technologies, Danbury, CT). The pressure applied to squeeze the powder towards the diamond was 5 N/cm^2. A total of 64 scans were averaged for each sample and the resolution was 8 cm^{-1}. The spectra were ratioed against a single-beam spectrum of the clean ATR crystal and converted into absorbance units. Data was collected in the range 4000–550 cm^{-1}.

2.5. FT-Raman spectroscopy

The spectra were acquired on a Perkin Elmer System 2000 NIR FT-Raman interferometer (PerkinElmer, UK) with a Nd:YAG laser emitting at 1064 nm. The power of the laser on the sample was set to 400 mW. The data were collected using an InGaAs detector and stored as Raman

shift in the range 3600–200 cm^{-1}. A 180° back-scattering arrangement was used and no correction for the spectral response was applied. The data were collected at a resolution of 32 cm^{-1} to reduce noise, and the number of scans was 64.

2.6. NIR spectroscopy

The NIR data were collected using a NIR systems spectrometer model 6500 (NIR systems, Inc., Silver Springs, USA) in reflectance mode. The range 1100–2500 nm was acquired using a lead sulfide detector. The angle of the incident light was 180° and reflectance was measured at an angle of 45°. The number of scans was 32 and the interval between spectral data points was 2 nm. A rotating sample cup with a quartz window was used for the measurements. Because of the relatively small amount of sample, a ring that reduced the effective diameter in the sample cup to 12 mm was used. The NIR reflectance spectra were converted to log(1/*R*) units prior to data analysis using an internal ceramic reference.

2.7. Chemometric analysis

Multivariate data analysis was applied to obtain optimal quantitative and qualitative information from the measured spectra. Partial least squares (PLS) regression (Wold, Martens, & Wold, 1983) models were developed for each of the five investigated carrageenan types. For each spectroscopy, three different data sets were created prior to calibration: (1) raw spectra, (2) second derivative spectra (Savitzy-Golay) and (3) spectra treated with extended inverted signal correction (EISC) in its general form (Pedersen, Martens, Nielsen, & Engelsen, 2002).

2.8. Software

The Unscrambler version 8.0 (CAMO, Inc., Trondheim, Norway) was used for all PLS models. Software for preprocessing with EISC and for backwards interval-PLS (bi-PLS) can be found at www.models.kvl.dk and was performed in MATLAB version 6.5 (MathWorks, Inc., Natick, MA). Other spectral preprocessing was made in The Unscrambler.

3. Results and discussion

3.1. Incubation of carrageenan powders at constant humidity

In order to keep the water content at the same level during the mixing of calibration samples and the subsequent spectral measurements, it was chosen to incubate the production samples P1–P5 at constant humidity before the mixing. For this purpose, the water uptake of the carrageenan powders at different relative humidities was studied as shown in Fig. 2. A relative humidity of 43% was chosen, resulting in an equilibrium water content of 8–10%, which was close to the equilibrium water content when for example stored at room temperature under normal storage conditions.

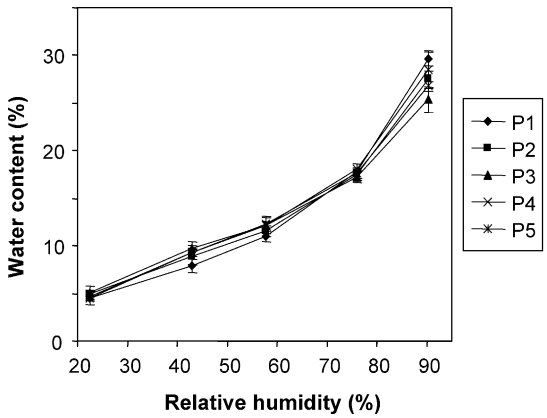

Fig. 2. Water uptake of the production samples P1–P5 at different relative humidity.

3.2. NMR reference method

The anomeric region of both ^{1}H and ^{13}C NMR spectra contains specific information about the different carrageenan types (van de Velde, Knutsen, Usov, Rollema, & Cerezo, 2002). The same samples, i.e. 4% solutions in D_2O, were used for the ^{1}H and ^{13}C NMR spectra that are shown in Fig. 3. As evident from the figure, the ^{1}H NMR spectra have significantly larger signal-to-noise-ratio (S/N ~ 1500) than the ^{13}C spectra (S/N ~ 50) due to the larger sensitivity of ^{1}H NMR spectroscopy despite an acquisition time of 24 h of the ^{13}C NMR spectra. In order to obtain a reasonable S/N of the ^{13}C spectra samples of higher concentration (7–10%) are necessary (van de Velde et al., 2002).

It was therefore concluded that with the sample preparation used yielding 4% solutions, the ^{13}C NMR spectra did not have sufficient S/N to quantify carrageenans present in low concentrations (lower than 5%), while ^{1}H NMR spectra obtained in only 10 min allowed for exact quantification (down to 0.5%) of the main types of carrageenans and even showed the presence of additional unidentified carrageenans as well as non-carrageenan substances such as Floridean starch (Yu et al., 2002).

In the ^{1}H NMR spectra, the width of the peaks is relatively large compared to the range of the anomeric proton signals. This results in some overlap, especially between the iota and mu peaks and between the lambda and nu peaks, and was dealt with applying simultaneous fitting with Lorentzian functions using an in-house built routine yielding the composition of the mixtures expressed as relative concentrations of the five carrageenan types as well as unidentified substances. Using this routine, the ^{1}H NMR spectra gave reliable results for the quantification of the five

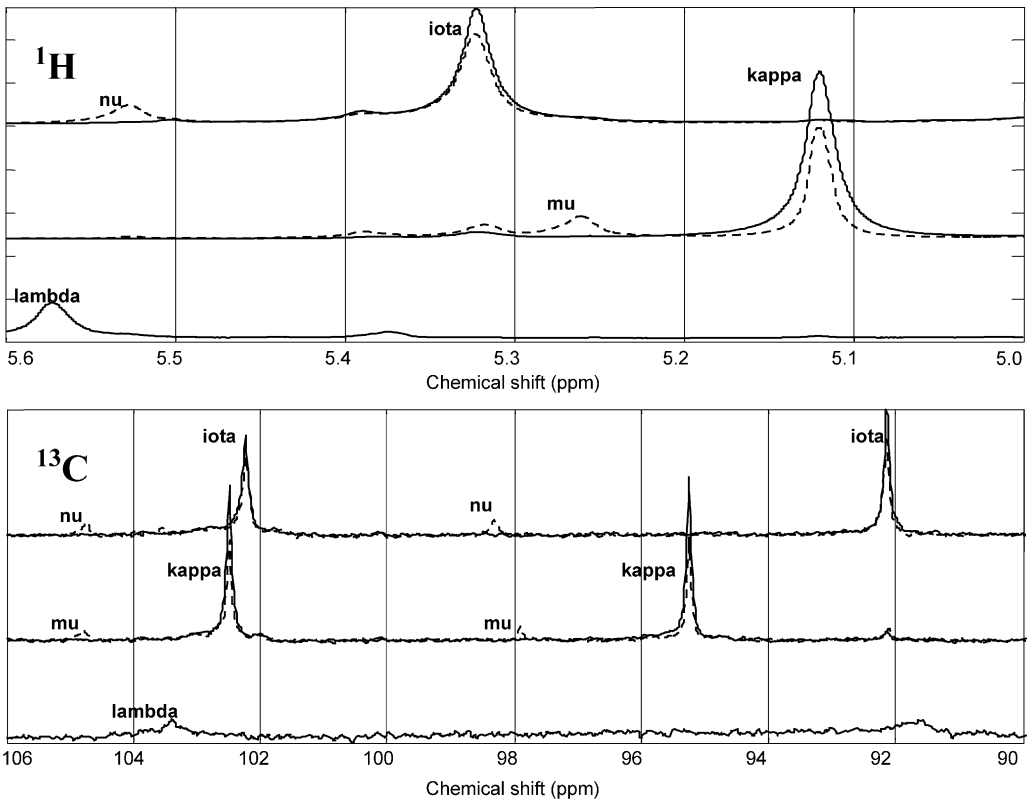

Fig. 3. The 600 MHz ^{1}H and 150 MHz ^{13}C NMR spectra of the production samples P1–P5. The spectra are from the top P1 (—) and P5 (- -), P2 (—) and P4 (- -), and P3 (—).

carrageenan types in the mixtures, and hence the compositions of the calibration and validation samples used for PLS models are based on quantifications made from the ^{1}H NMR spectra. The error of the quantification (sum of spectral error and fitting error) using the described ^{1}H NMR method was estimated to between 0.2 and 0.9% (the standard deviation of quantifications of replicates).

3.3. Exploratory PCA models

Before the development of calibration models, PCA was used to investigate to what extent the investigated spectroscopic methods used in this work could contrast the five different carrageenan forms. Since spectral preprocessing strongly affects the type of information extracted in the first loadings, PCA was performed on both raw spectra and spectra preprocessed with Savitzy-Golay (second derivative) and EISC (Pedersen et al., 2002).

For all spectroscopies, the design was partly represented in the scoreplot of the first two principal components of the PCA models on calibration samples. The best results were observed using EISC treated spectra and the best representation of the pentanary design was seen with IR (Fig. 4a) and Raman (Fig. 4b) spectra. For IR a quaternary structure involving iota, kappa, lambda and mu is seen in the scoreplot of the first two principal components, and for Raman a ternary structure involving iota, kappa and lambda can be seen. NIR showed similar trends albeit not as systematic.

3.4. Qualitative evaluation of carrageenan spectra

Infrared absorbance. The spectral assignment of carrageenan IR spectra has recently been reviewed and reexamined by Prado-Fernandez et al. (2003) for the iota, kappa and lambda types. The challenging part of carrageenan assignments is that the normal modes that are able to distinguish amongst the different types of carrageenans are complex pyranosidic skeleton modes or sulphate modes which both reside in the crowded fingerprint region and are sensitive to the molecular environment (see Fig. 5A). Selective bands do exist for iota/nu carrageenans at 719 and 806 cm^{-1}, for kappa carrageenans at 734 cm^{-1} and for lambda carrageenans at 810–830 cm^{-1} (broad) and at approximately 1010 cm^{-1}, but none of these bands are even close to being baseline separated and it is thus the fine spectral details (patterns) that enable IR characterization of the carrageenans. For this reason, it is necessary to apply pattern recognition methods such as PLS to develop robust calibrations. Similarly the mu and nu spectra are qualitatively very close to their parental kappa and iota spectra. Of the two precursors only mu differ visually from kappa by the apparent lack of intensity of the 840 and 890 cm^{-1} bands and by a small shift of the former band to 850 cm^{-1}.

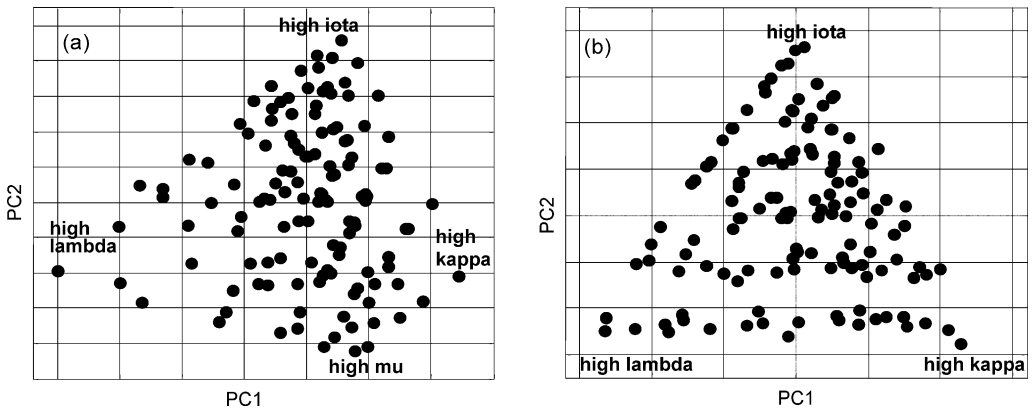

Fig. 4. PCA scoreplot of the first two principal components for EISC treated IR spectra (a) and EISC treated Raman spectra (b).

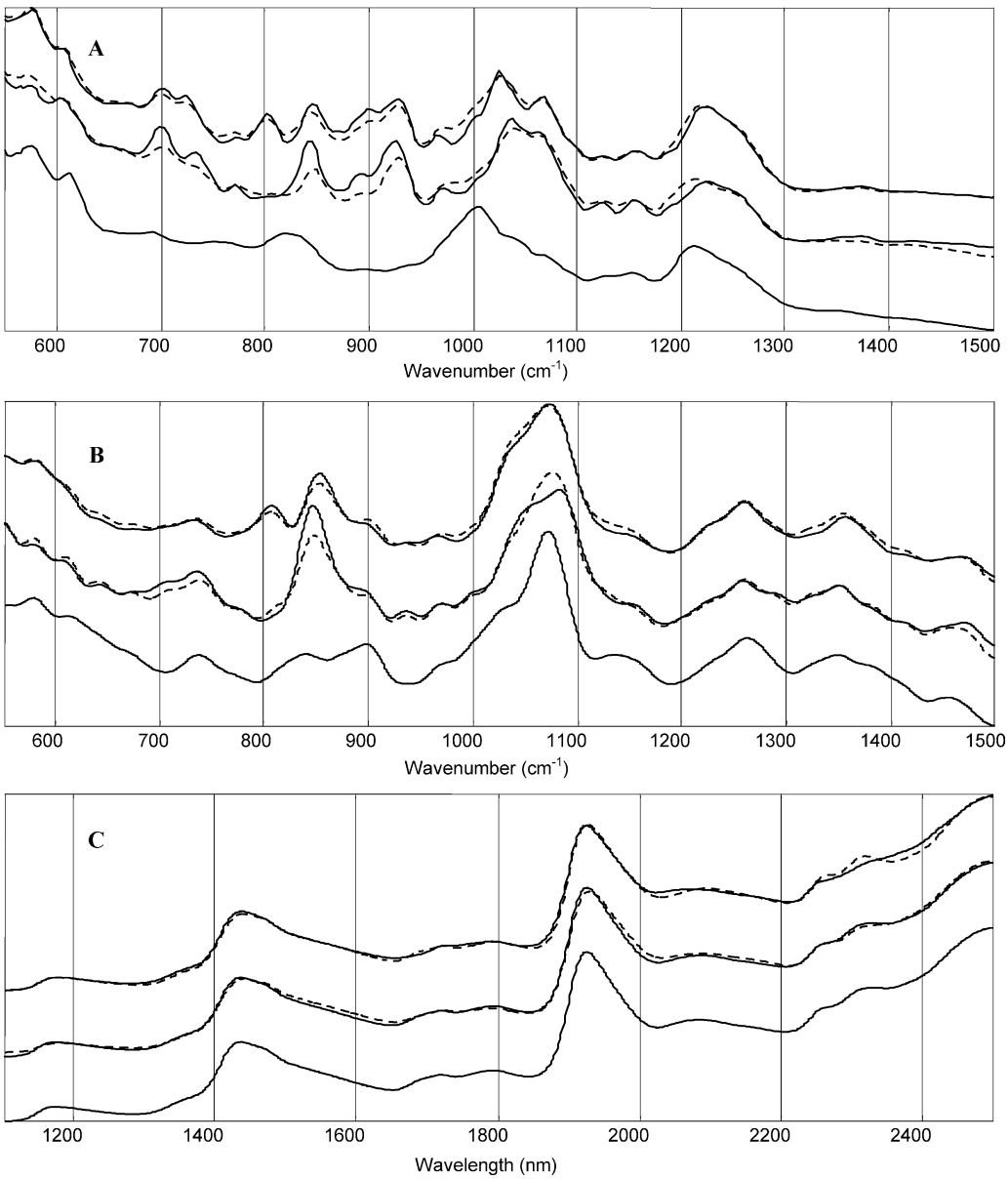

Fig. 5. EISC treated IR, Raman and NIR spectra of the production samples P1–P5. The spectra are from the top P1 (—) and P5 (- -), P2 (—) and P4 (- -), and P3 (—).

Raman scattering. Like infrared spectroscopy Raman scatter measures the fundamental molecular vibrations, but with different selection rules. As a general rule Raman is more sensitive to skeleton vibrations and less sensitive to side group (e.g. sulphate) vibrations. The main variations related to carrageenan types observed in the Raman spectra (see Fig. 5B) is found in the most intensive bands around 850 cm^{-1} and the dual peak in the range from 1000 to 1150 cm^{-1}. The lambda carrageenans are characterized by almost complete lack of the 850 cm^{-1} band. Like in the infrared spectra iota/nu carrageenans are characterized by the presence of the 805 cm^{-1} band and by comparison to the kappa/mu carrageenans by a less intense and slightly shifted 840–850 cm^{-1} band. The shape (relative peak intensity and position) of the dual peak in the range 1000–1100 cm^{-1} appear to be almost diagnostic amongst all five carrageenan types investigated.

Near-infrared reflectance. Unlike the infrared and Raman spectroscopies NIR spectroscopy measures over- and combination-tones predominantly of the fundamental vibrations that involve hydrogens. As expected when the variations amongst the samples do not involve hydrogens the spectra appear almost similar. However, when scrutinizing the spectra in Fig. 5C small spectral differences amongst the different carrageenans can be revealed but evidently multivariate calibration techniques are required to relate these highly overlapped and almost similar spectral features to the concentrations of the different carrageenan types.

3.5. Calibration models

Calibration models based on IR, Raman and NIR spectra of the 128 calibration samples from the experimental design were developed using PLS. Calibration models were based on raw, second derivative and EISC pretreated spectra. For IR and Raman data only the range 1500–560 cm^{-1} was used in calibrations in order to avoid the strong water bending mode centered around 1640 cm^{-1} and the noisy low energy range below 560 cm^{-1}. For NIR the whole recorded spectral range 1100–2500 nm was used.

Table 2 shows the results of PLS models on calibration samples using raw, second derivative and EISC treated IR, Raman and NIR spectra. All calibration models were validated using cross-validation with five segments representing five days of random sample mixing and measurement. For each calibration model, three values are given: the root-mean-squared error of cross-validation (RMSECV), which is a measure of the prediction error, the correlation coefficient (R) between the measured and predicted values, and the number of PLS components (#PC) used in the model.

For IR between four and seven PLS components were found optimal in the models, but most often five components. For Raman between two and eight PLS components. For NIR between five and eight PLS components were needed, most often six. For all three spectroscopic techniques, the raw spectra yielded the poorest calibration models while preprocessing with both second derivative and EISC significantly improved the quality of the calibrations. For IR and NIR, the overall performance of EISC is more or less equal to that of second derivative, but for Raman spectra the prediction uncertainty is reduced with up to 66% using EISC compared to second derivative and generally using fewer components, i.e. simpler models.

Fig. 6 shows raw and EISC treated IR, Raman and NIR spectra of the calibration samples. To our knowledge, this is the first application of EMSC or EISC to Raman spectra and the results are most promising and significantly improved from standard MSC, which also was demonstrated to be the case with the analysis of NIT spectra measured on single

Table 2
Results of cross-validated prediction models on IR, Raman and NIR spectra of calibration samples using raw spectra, second derivative and EISC treated spectra

			Iota	Kappa	Lambda	Mu	Nu
IR	Raw	RMSECV[a]	5.2	5.1	6.7	1.2	1.5
		R[b]	0.97	0.98	0.95	0.97	0.91
		#PC[c]	5	5	5	5	6
	Second derivative	RMSECV	4.5	4.3	6.3	0.9	0.8
		R	0.98	0.98	0.95	0.98	0.97
		#PC	5	5	5	4	7
	EISC	RMSECV	4.8	4.0	4.6	1.0	1.1
		R	0.97	0.98	0.98	0.98	0.94
		#PC	5	5	5	5	5
Raman	Raw	RMSECV	4.4	4.3	5.3	1.3	0.7
		R	0.98	0.98	0.97	0.97	0.98
		#PC	4	6	8	8	8
	Second derivative	RMSECV	3.8	4.7	5.0	1.2	0.8
		R	0.98	0.98	0.97	0.97	0.97
		#PC	4	3	5	8	5
	EISC	RMSECV	**1.5**	**1.6**	**2.1**	0.9	0.5
		R	1.00	1.00	0.99	0.98	0.99
		#PC	3	2	4	4	6
NIR	Raw	RMSECV	4.5	3.4	4.4	0.8	1.6
		R	0.98	0.99	0.98	0.99	0.88
		#PC	6	6	6	6	6
	Second derivative	RMSECV	2.7	2.4	3.3	0.8	0.6
		R	0.99	0.99	0.99	0.99	0.98
		#PC	6	6	6	6	6
	EISC	RMSECV	2.8	2.6	2.2	**0.6**	**0.5**
		R	0.99	0.99	0.99	0.99	0.99
		#PC	5	5	5	5	8

The overall lowest prediction errors are printed in bold.
[a] RMSECV is the root-mean-squared error of cross-validation in percent.
[b] R is the correlation coefficient.
[c] #PC is the number of PLS components used.

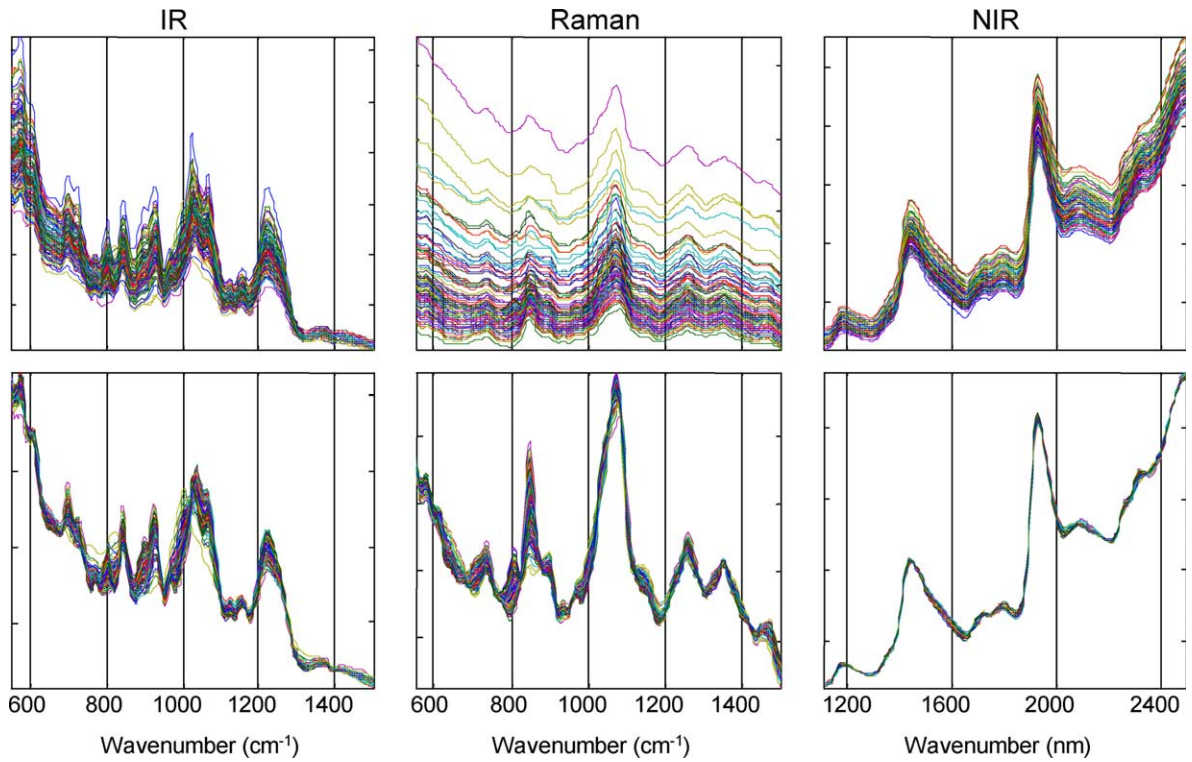

Fig. 6. Raw (top) and EISC treated (bottom) IR, Raman and NIR spectra of the 128 calibration samples.

wheat seeds (Martens, Nielsen, & Engelsen, 2003; Pedersen et al., 2002). The main reason for the improved EISC Raman model is the inclusion of the quadratic wavelength term, which apparently is able to compensate for subtle effects related to the Raman scatter wavelength dependency and for varying exposure of fluorophores.

The best calibration models for iota, kappa and lambda are found using EISC treated Raman spectra, while for mu and nu the best model uses EISC treated NIR spectra. However, the difference between NIR and Raman for mu and nu is very small and thus the overall best models are found using EISC treated Raman spectra. These models yield prediction errors of 1.5, 1.6 and 2.1% for iota, kappa and lambda, respectively, for which the range of concentrations in the calibration set is approximately 0–95%. Calibration models of the mu and nu content, which varies between approximately 0 and 20%, yield prediction errors of 0.9 and 0.5%, respectively.

The uncertainty of the ^{1}H NMR measurement of carrageenan composition used as reference values was 0.5, 0.3, 0.9, 0.4 and 0.2% for iota, kappa, lambda, mu and nu, respectively. The uncertainty of the reference method contributes to the RMSECV in the models and gives the lower limit of the prediction error in the calibration models. Although relatively low, the obtained prediction errors are somewhat higher than the uncertainty of the reference method, and thus contain a certain degree of error due to noise in the spectral data and modeling error, which can be ascribed mainly to the high degree of impurities in the production samples.

The highest prediction error is found for the lambda model (2.1%), which could be partly due to a relatively high quantification error (0.9%) of the NMR method. This is probably due to the high viscosity of the nearly pure lambda sample (P3), which in turn resulted in a lower S/N for the ^{1}H NMR spectrum for this sample than for the other samples.

In literature, there are only few studies to compare our quantitative models against. The first attempt to develop a method for determining the composition of carrageenan samples using multivariate analysis and spectroscopy was by Jacobsson and Hagmann (1993). They applied PLS and neural networks to several selected peak heights from IR spectra of mixtures of pure iota, kappa and lambda carrageenan prepared as KBr tablets. The study was rather small and prediction errors of more than 20% were obtained, which is ascribed to the fact that relatively few IR peaks were included in the data analysis.

In 1999, Hansen and Wichmann presented an IR method for measuring the concentration of carrageenans at the conference *Gums and Stabilizers for the Food Industry* but their results have not been published. They used purified commercial iota, kappa and lambda carrageenan to produce a number of standard mixtures prepared as KBr tablets. The RMSECV gained was 1.1–1.8%, which is slightly lower than in this work, but their model was not able to predict lambda when a sample contained mu or nu, and samples known to contain mu and nu but no lambda were predicted to high lambda contents. Furthermore, the performance of the models on production batches was not shown. In our work, the carrageenan samples are production

batches that contain impurities, which complicates the building of PLS models.

Most recently, a study was published using IR and PLS on standard mixtures of iota, kappa and lambda carrageenan (Prado-Fernandez et al., 2003). The carrageenan mixtures were presented to the IR instrument as thin films and the method is thus not suitable for on-line production control. Prediction errors using leave-one-out cross-validation of 4.2, 3.3 and 3.3% were obtained for iota, kappa and lambda, respectively. Using a test set (also consisting of standard mixture samples) errors of approximately 4, 3 and 4% were obtained. The standards were purchased as pure commercial standards and were further purified using repeated partial solubilization and precipitation, dialysis to remove the gel promoting counter ions and finally lyophilization. The prediction of several industrial blends was performed with a maximum of 30% deviation from reference values. The sample preparation for industrial blends was around 1 h excluding drying before weighing.

In order to evaluate the performance of our calibration models on industrial production batches, the composition of production samples P6–P10 was predicted from their IR, Raman and NIR spectra, but with very poor agreement with the composition obtained from the NMR measurement. The reason for this was investigated and is ascribed to the fact that these production samples have different amounts of various counter ions (concentrations of K^+, Ca^{2+}, Na^+ and Mg^{2+} were measured), which cause shifts of the peaks in the vibrational spectra. Furthermore, in the calibration samples, we could not avoid a high correlation between the content of specific ions and specific carrageenans, and thus the calibration models are possibly based on spectral information of a combination of counter ion type and carrageenan type. As the counter ions seemed to affect peaks over the whole spectral range, it was not possible to manually remove single entire peaks responsible for the observed differences and therefore variable selection was applied.

3.6. Variable selection using bi-PLS

The assumption behind this approach to improving the prediction of production samples was that the spectra contain regions with carrageenan information that is 'corrupted' by the counter ions as well as regions with carrageenan information, which is *not* corrupted by the ions. Thus, variable selection on a data set containing both calibration samples and production samples, where the ion and carrageenan concentrations are not as strongly correlated, should be able to produce models that are better suited for the prediction of production samples.

To this end, we use the new approach to variable selection called bi-PLS developed by Leardi and Nørgaard (in preparation), which is an extension of interval-PLS (Nørgaard et al., 2000). It is based on splitting the spectra up into a number of intervals, and calculating PLS models leaving one interval out at a time. The model that results in the smallest RMSECV is retained and new models are calculated again leaving one interval out now starting with one interval less. Finally, the model that results in the smallest RMSECV overall is selected, and the intervals that are used in this model (as well as the intervals that were left out) can be investigated for spectral information.

Variable selection using bi-PLS was applied to the data set consisting of EISC treated Raman spectra of the calibration and production samples (giving a total of 133 samples) in order to optimize the prediction of production samples. The EISC treated Raman spectra were chosen due to their superiority in the PLS models on calibration samples. In order to direct the selection of intervals to improve the prediction of the production samples, cross-validation with manually selected segments was performed with the production samples (P1–P10) as a separate segment. Different numbers of spectral intervals were tested (5, 10, 15 and 20) to explore the spectral regions. The number of PLS components was restricted to five to avoid over-fitting.

Table 3 shows the bi-PLS results for EISC pretreated Raman spectra of calibration and production samples as well as the corresponding models on the entire spectral range (denoted global PLS) consisting of 950 data points. For bi-PLS, the lowest values of RMSECV were found dividing the spectral range into 15 intervals and using between 5 and 10 intervals in the calibrations, yielding models based on 315–634 variables. Note that the larger prediction errors for these model (Table 3) compared to the models on calibration samples only (Table 2) arise from the different validation method as well as from the inclusion of the production samples into the calibration set.

The reduction in prediction error was between 27 and 67% using bi-PLS compared to the global PLS. The significant reduction in prediction error compared to the global PLS model supports our assumption that certain regions of the spectra containing carrageenan information indeed were corrupted by the presence of different counter ions, and hence the removal of these regions markedly improved the prediction of the production samples.

Table 3
bi-PLS results for EISC pretreated Raman data of calibration and production samples

		Iota	Kappa	Lambda	Mu	Nu
Global PLS	RMSECV[a]	2.6	3.5	6.0	1.6	1.2
	#PC[b]	4	4	5	5	4
	#Var.[c]	950	950	950	950	950
bi-PLS	RMSECV	1.9	2.1	1.9	0.8	0.7
	#PC	5	5	5	5	5
	#Var.	379	315	634	634	570

[a] RMSECV is the root-mean-squared error of cross-validation in percent.
[b] #PC is the number of PLS components used.
[c] #Var. is the number of variables in the model.

3.7. Mixed calibration models

As mentioned previously, PLS models based on the calibration samples alone did not perform well on the production samples due to the differences in counter ion content and possibly impurities. To obtain predictions of production samples with reasonable error the samples included in the calibrations should span all possible variation, thus including several production samples with a representative variation in counter ion content as well as carrageenan content.

The sample set available for this study only contains 10 production samples (P1–P10) and of these only one sample contains lambda and only two contain mu. These samples cannot be said to be representative of all future samples to be predicted. However, in order to evaluate the potential of expanding the models built on calibration samples with production samples, and thus to be able to predict the carrageenan content in future production samples, PLS models were built on EISC treated Raman spectra of the calibration samples and the production samples together using the spectral regions found with bi-PLS. Cross-validation with 10 segments each containing one production sample is used to represent future situations where production samples are predicted in a model already containing production samples.

Fig. 7 shows the regression coefficients for the models of each of the five carrageenans as well as EISC treated Raman spectra of P1–P3. All five models use information that is spread over the entire spectral range. The interval containing the 840 and 850 cm^{-1} peaks are left out in all but the mu model, suggesting that this region contains information corrupted by the counter ions. The broad peak around 1000–1150 cm^{-1} is important in all models, but is weighed very differently in the five models. The peak around 1040 cm^{-1} has a high positive regression coefficient in the iota and nu models, but a high negative regression coefficient for the kappa model. The peak around 1075 cm^{-1} has positive regression coefficients for the kappa and mu models, but is left out for all other models. The lambda model only seems to use the width of the peak having a high negative correlation coefficient at 1120 cm^{-1}. Furthermore, characteristics of individual carrageenan types can be seen in the regression coefficients, e.g. the 805 cm^{-1} peak for the iota model and 900 cm^{-1} for the lambda model. The models for the precursors mu and nu seem to use information complementary to the kappa and iota models, e.g. for iota/nu the region 1100–1500 cm^{-1} and for kappa/mu the region 550–850 cm^{-1}.

Fig. 8 shows the predicted versus measured values for the PLS models built on calibration and production samples together. The prediction errors are for iota 1.9%, for kappa 2.2%, for lambda 2.1%, for mu 0.8% and for nu 0.7% all using five PLS components. It is clear from the figure that the prediction of production samples is still made with somewhat higher error than for the calibration samples, but the error of prediction is significantly lower than using models built on calibration samples alone (results not shown).

The predictions show that it is possible to predict the carrageenan content in production samples with reasonable

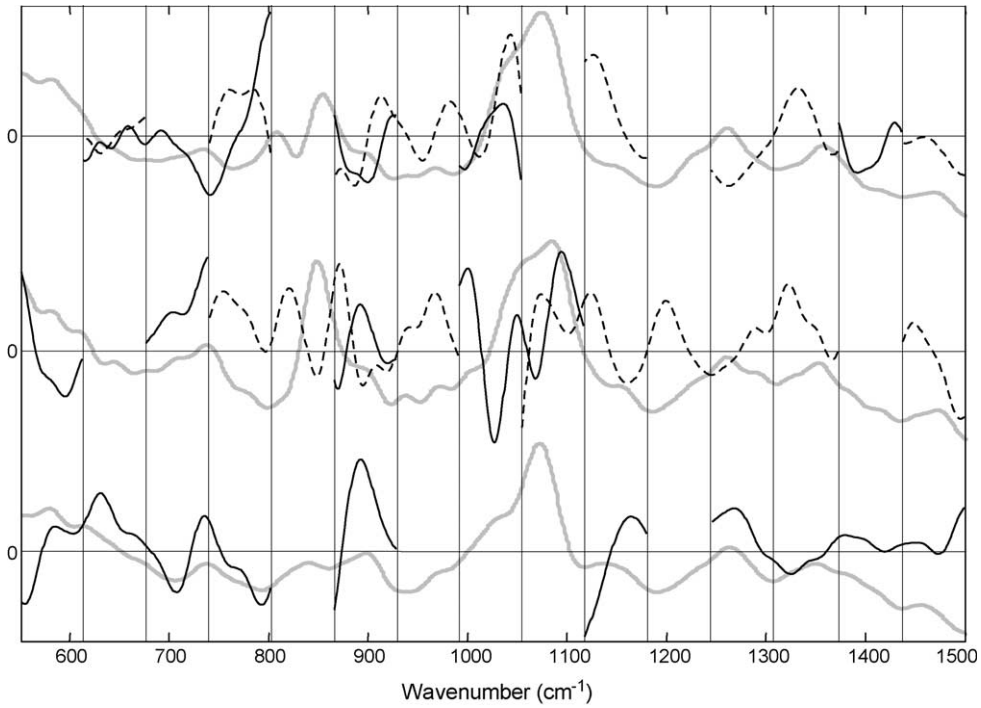

Fig. 7. PLS regression coefficients for the final mixed calibration models based on EISC treated Raman spectra using the spectral regions chosen by bi-PLS. Top: regression coefficients for the iota (—) and nu (- -) models and P1 spectrum (gray, arbitrary intensity); Middle: regression coefficients for the kappa (—) and mu (- -) models and P2 spectrum (gray); Bottom: regression coefficient for the lambda (—) model and P3 spectrum (gray).

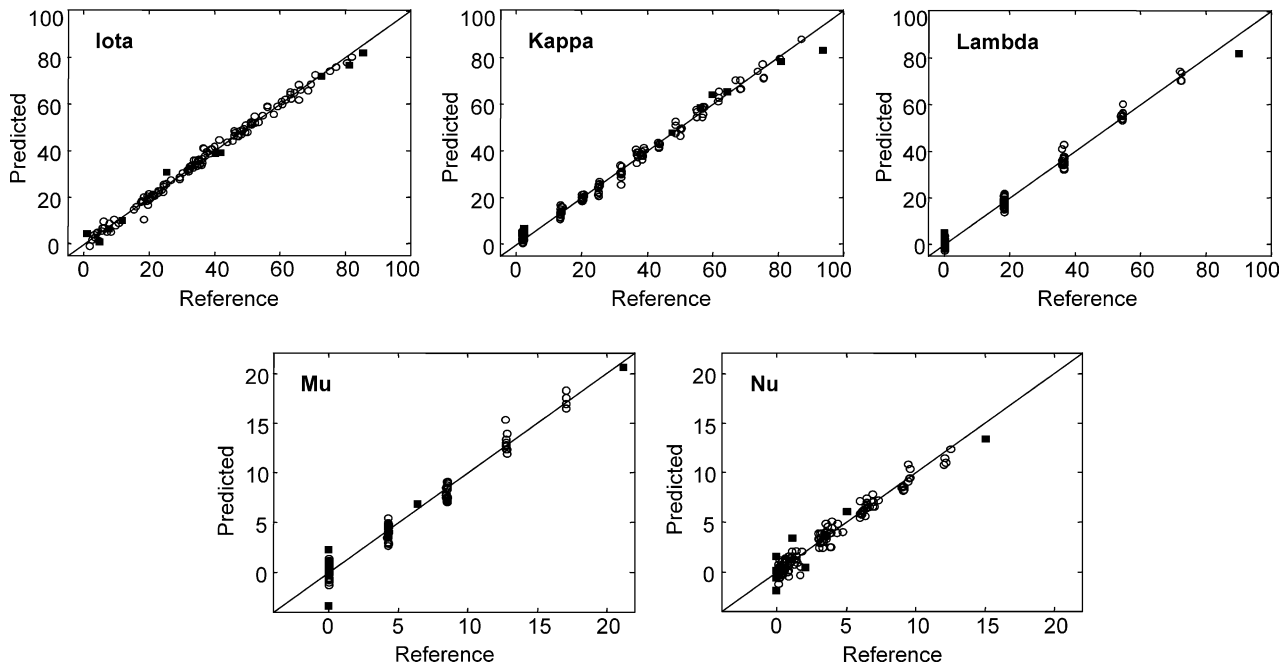

Fig. 8. Cross-validated predictions versus reference carrageenan content for calibration samples (○) and production samples (■). The models are the final mixed calibration models of iota, kappa, lambda, mu and nu content based on EISC treated Raman spectra using the regions chosen by bi-PLS.

accuracy using these models. However, for a more robust model is would be recommended that more production samples were included in the calibration set as only 10 production samples (P1–P10) are included in these models. More production samples would make the counter ions content less correlated to carrageenan content and thus allow for better prediction of new production samples with a different counter ion content.

4. Conclusion

Quantitative and non-destructive methods determining the carrageenan composition from untreated powders were developed using IR, Raman and NIR spectroscopy and chemometrics, using ^{1}H NMR as the reference method. The developed methods were able to handle a pentanary design consisting of iota, kappa, lambda, mu and nu carrageenan yielding very good calibrations with correlations between reference concentrations and predicted values in the range 0.94–1.00.

The preprocessing method EISC generally improved the predictive capabilities of the three spectroscopic methods IR, Raman and NIR. Especially, the Raman results were improved to an extent that this method provided the best overall models with lowest prediction error (expressed as RMSECV). Moreover, for the first time, we have developed good quantitative models from the favorite industrial technique, NIR spectroscopy, despite its apparent lack of sensitivity to the sulphate decoration of carrageenans.

The prediction of a few industrial production samples from the calibration models based on design samples only was unsuccessful due to differences in counter ion contents. Using the variable selection method bi-PLS on the Raman data, spectral regions were found that significantly improved the prediction of the production samples. Final PLS models on the calibration samples and production samples together using the regions found with bi-PLS, yielded cross-validated prediction errors (RMSECV) of approximately 2% for iota, kappa and lambda carrageenan, and 1% for mu and nu carrageenan. Including a larger number of production samples that span the range of the design properly with respect to both carrageenan and counter ion content is likely to further improve the models.

Acknowledgements

The authors would like to thank U.G. Sidelmann and J. Breinholt at Novo Nordisk A/S, Denmark for access to 600 MHz NMR facilities and instrumental support.

References

Chiovitti, A., Liao, M. L., Kraft, G. T., Munro, S. L. A., Craik, D. J., & Bacic, A. (1995). Cell wall polysaccharides from Australian red algae of the family Solieriaceae (Gigartinales, Rhodophyta): Iota/kappa/beta-carrageenans from *Melanema dumosum*. *Phycologia*, *34*(6), 522–527.

Chopin, T., & Whalen, E. (1993). A new and rapid method for carrageenan identification by FT IR diffuse reflectance spectroscopy directly on dried, ground algal material. *Carbohydrate Research*, *246*, 51–59.

Engelsen, S. B., & Nørgaard, L. (1996). Comparative vibrational spectroscopy for determination of quality parameters in amidated pectins as evaluated by chemometrics. *Carbohydrate Polymers*, *30*(1), 9–24.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wikström, C., & Wold, S. (2000). *Design of experiments. Principles and applications*. Umeå, Sweden: Umetrics AB.

Jacobsson, S. P., & Hagman, A. (1993). Chemical composition analysis of carrageenan by infrared spectroscopy using partial least-squares and neural networks. *Analytica Chimica Acta*, *284*, 137–147.

Malfait, T., Van Dael, H., & van Cauwelaert, F. (1998). Molecular structure of carrageenans and kappa oligomers: A Raman spectroscopic study. *International Journal of Biological Macromolecules*, *11*, 259–264.

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*(3), 394–404.

Matsuhiro, B. (1996). Vibrational spectroscopy of seaweed galactans. *Hydrobiologia*, *326/327*, 481–489.

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*(3), 413–419.

Pedersen, D. K., & Engelsen, S. B. (2001). Monitoring industrial food processes using spectroscopy and chemometrics. *New Food*, *4*(2), 9–13.

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*(9), 1206–1214.

Prado-Fernandez, J., Rodriguez-Vazquez, J. A., Tojo, E., & Andrade, J. M. (2003). Quantitation of κ-, ι- and λ-carrageenans by mid-infrared spectroscopy and PLS regression. *Analytica Chimica Acta*, *480*(1), 23–37.

Sekkal, M., & Legrand, P. (1993). A spectroscopic investigation of the carrageenans and agar in the 1500–100 cm^{-1} spectral range. *Spectrochimica Acta*, *49A*(2), 209–221.

Stortz, C. A., Bacon, B. E., Cherniak, R., & Cerezo, A. S. (1994). High-field NMR spectroscopy of cystocarpic and tetrasporic carrageenans from *Iridaea undulosa*. *Carbohydrate Research*, *261*(2), 317–326.

Therkelsen, G. H. (1993). Carrageenan. In R. L. Whistler, & J. N. BeMiller (Eds.), *Industrial gums* (pp. 145–180). San Diego: Academic Press, Inc.

Turquois, T., Acquistapace, S., Arce Vera, F., & Welti, D. H. (1996). Composition of carrageenan blends inferred from ^{13}C NMR and infrared spectroscopic analysis. *Carbohydrate Polymers*, *31*, 269–278.

van de Velde, F., Knutsen, S., Usov, A. I., Rollema, H. S., & Cerezo, A. S. (2002). ^{1}H and ^{13}C high resolution NMR spectroscopy of carrageenans: Application in research and industry. *Trends in Food Science and Technology*, *13*, 73–92.

van de Velde, F., Peppelman, H. A., Rollema, H. S., & Tromp, R. H. (2001). On the structure of kappa/iota-hybrid carrageenans. *Carbohydrate Research*, *331*, 271–283.

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.

Yu, S. K., Blennow, A., Bojko, M., Madsen, F., Olsen, C. E., & Engelsen, S. B. (2002). Physico-chemical characterization of floridean starch of red algae. *Starch/Stärke*, *54*(2), 66–74.