ELSEVIER

PII: S0144-8617(96)00068-9

Carbohydrate Polymers **30** (1996) 9–24

0144-8617/96/$15.00 + .00

Comparative vibrational spectroscopy for determination of quality parameters in amidated pectins as evaluated by chemometrics

Søren Balling Engelsen* & Lars Nørgaard

The Royal Veterinary and Agricultural University, Food Technology, Department of Dairy and Food Science, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

(Received 12 February 1996; revised version received 16 April 1996; accepted 19 April 1996)

The potential of vibrational spectroscopy and chemometrics as a reliable and fast method for the determination of important gel-forming parameters in amidated pectins has been investigated.

For a set of 98 amidated pectin samples, six complete spectroscopic ensembles were recorded including NIR, FT-NIR, FT-IR and NIR FT-Raman spectroscopy. For each spectroscopic ensemble, quantitative models based on partial least squares regression (PLS) have been developed and compared. Chemometric models were constructed by dividing the spectroscopic ensembles up into a calibration set of 73 samples and an independent test set of 25 samples to evaluate the predictive ability of the models.

The 98 amidated pectin samples span a degree of esterification (%DE) between 20 and 55% and a degree of amidation (%DA) between 4 and 24 per cent. From all six spectroscopic ensembles quantitative PLS models were obtained for %DE and %DA with RMSEP (root mean square error of prediction) ranging between 1.5 and 2.1 and between 1.1 and 2.1 for %DE and %DA respectively. In both cases the results are comparable to that of the experimental error of the quantitative chemical determination.

Finally, different ways of selecting FT-IR spectral elements (variables) to correlate with %DE were compared. The method of principal variables (PV) was found to be superior to methods based on knowledge based selection and the resulting PV-based model was found to be slightly better than the PLS model using the full spectral information. Copyright © 1996 Elsevier Science Ltd

INTRODUCTION

Pectins are a family of polysaccharides found in the primary cell wall of most flowering plants. They are believed to embed the fundamental cellulose-xyloglucan framework (app. 50% of the wall mass) in a second domain of matrix pectic polysaccharides (app. 30% of the wall mass). These pectic polysaccharides are some of the most complex polymers known and they are believed to be involved in many functions of the cell wall: to determine cell wall porosity, to provide charged surfaces that modulate wall pH and ion balance and to serve as recognition molecules that signal appropriate developmental responses and to symbiotic organisms, pathogens and insects (Carpita & Gilbeaut, 1993).

*Corresponding author. Email: engelsen@muscadet. Foodsci. kvl.dk

Structurally the main pectic polysaccharide is primarily a linear polymer of (1,4)-linked α-D-galacturonic acid which may be partially methyl esterified. This backbone is interrupted by (1,2)-linked α-D-rhamnopyranose which serves as anchoring points for the attachment of lateral neutral oligosaccharides containing mainly D-galactopyranoses and L-arabinofuranoses ('hairy regions').

Although the relationships of structural details such as the degree of esterification, the distribution of rhamnopyranoses and the molecular weight on functional properties have been demonstrated it is still not fully elucidated why such a biopolymer, unlike other similar structures, develops a gel in the presence of water, sugars and acid (Cros *et al.*, 1992). At present our understanding of the 3D structures of pectins and the interactions between pectin molecules is highly simplified. However the still increasing knowledge about

pectins and pectin systems, combined with the arrival of new modeling tools specifically aimed towards complex carbohydrate structures (Engelsen *et al.*, 1996), is likely to provide more detailed insight in the structural arrangements of pectins.

The ability to gel and their non-toxicity (no acceptable daily intake value, ADI, has been specified for pectins (FAO, 1992)) has made pectins popular as sugar-based hydrocolloids in the confectionery industry. In the US alone the confectionery industry is growing at approximately 3% per year and the consumption has increased to approximately 11 lb. per capita (Carr *et al.*, 1995). Pectin gels are characterized by providing a very tender, short texture with excellent clarity combined with outstanding flavor release properties.

For commercial applications pectins are usually extracted from citrus peel, apple pomace and sugar beet and processed to yield high and low methoxy pectins which gel differently. High methoxy pectins are capable of forming gel networks at acid pH in the presence of soluble salts and the %DE controls their relative speed of gelation. Low ester pectins gelling properties are dependent on the presence of divalent ions such as calcium and are much less pH dependent than for the high methoxy pectins. In low ester pectins the %DE is the primary parameter in the control of their calcium reactivity. The presence of amide groups in low ester pectins also strongly affects the calcium reactivity and the gelation properties, adding a new dimension for the design of pectin gels.

Because of the importance of %DE in relation to pectin gelation properties it is desirable to obtain a fast and robust method to determine (predict) the %DE in pectin powders. Vibrational spectroscopy is a good candidate for the development of such methods as spectrophotometers and quantitative software algorithms (chemometric methods) become more and more reliable and advanced. The present work investigates the quantitative performance of different instruments, spectral regions, sampling techniques and software algorithms developed within the area of chemometrics.

For this investigation, six complete spectroscopic ensembles were recorded including NIR, FT-NIR, FT-IR and NIR FT-Raman spectroscopy on a set of 98 amidated pectin samples. The samples span a degree of esterification (%DE) between 20 and 55% with a mean value of 32.0%. The degree of amidation (%DA) of the samples range between 4 and 24% with a mean value of 17.4% (see Fig. 1). A priori it was expected that NIR and IR would be most suitable for the problem under investigation due to the strong absorbing and well-defined group frequencies of the ester bonds which are almost negligible in the corresponding Raman spectra. However due to the fact that the Raman technique usually provides much more 'clean' and highly resolved spectral data for carbohydrates (Goral & Zichy, 1990) we wanted to investigate the potential of application of Raman spectroscopy to future quantitative studies of this group of substances. As a light scattering technique Raman spectroscopy has been regarded primarily as a qualitative technique where the quantitative information was hidden in the varying scatter from the sample. However, the application of multivariate data-analysis methods such as principal component regression (PCR) and partial least squares regression (PLS) (Martens & Næs, 1993) has partially solved this problem wherefore quantitative Raman spectroscopy is now feasible although more attention has to be focussed on the reproducibility of the traditional sampling techniques. Furthermore, the rapid developments in the Laser-Raman technique including remote sampling using fiber-optics (ideal for NIR wavelengths) has made it possible to perform simple on-line or routine batch measurements suitable for industrial process plants.

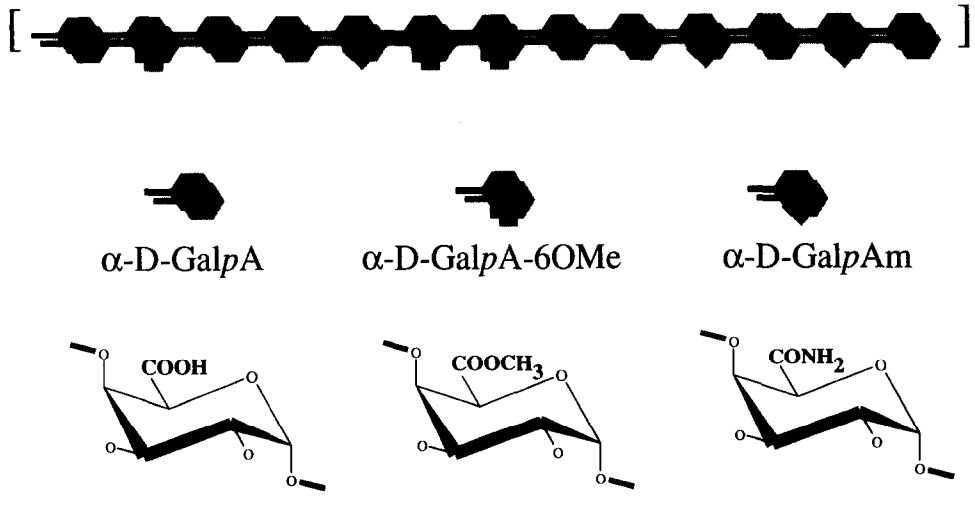

Fig. 1. A schematic model of a short sequence of the pectin backbone including α-D-Galactopyranuronic Acid, Methyl α-D-Galactopyrnanosiduronate and α-D-Galacto-pyranosiduronamide.

NIR spectra of pectins have previously been reported and used in quantitative studies. Horvárth *et al.* (1984) obtained a spectroscopic model for predicting pectin content in wheat bran using linear regression at two wavelengths (second derivative spectra) and Polesello *et al.* (1990) used NIR to obtain a model of the total galacturonic acid content and the total methoxy content also using linear regression at two wavelengths (second derivative spectra). However, to our knowledge there has been no investigations of the potential of IR and/or Raman spectroscopy for the quantitative determination of quality parameters in pectin powders or of the full potential of quantitative NIR determinations in pectins using multivariate methods such as PLS and PCR.

EXPERIMENTAL

Samples

Ninety eight finely powdered amidated pectin samples were provided by Copenhagen Pectin A/S (Hercules Inc.). The semi manufactured powder samples were examined without any form of pre-treatment such as drying and dilution.

Spectroscopic data

Two sets of dispersive near infrared data were collected using a NIRSystems Inc. spectrophotometer. One with a fixed sample cup with a diameter of 36 mm (the spectral ensemble hereafter designated NIR) and one with the same sample cup in a rotating device (spectral ensemble labeled NIR_R). Both used a split detector system with a Silicon (Si) detector between 400–1100 nm and a Lead Sulfide (PbS) detector from 1100 to 2500 nm. Angle of incident light was 180° and reflectance was measured at 45° angle. For further details see Table 1.

Two sets of Fourier transform near infrared data were collected using a Perkin Elmer System 2000 interferometer. One with the standard diffuse reflection cell (spectral ensemble: FTNIR_DR) and one with the integrating sphere cell (spectral ensemble: FTNIR_IS). The interferometer was equipped with quartz windows and a quartz beamsplitter. The FTNIR_DR data set was collected between 12 000 and 4000 cm^{-1} using a liquid nitrogen cooled photovoltaic Indium Antimonide detector (InSb). The FTNIR_IS data set was collected between 12 000 and 4000 cm^{-1} using a dedicated lead sulfide detector (PbS). In the latter case a 4% attenuator was used to prevent saturation of the detector.

Finally, one set of Fourier transform infrared data (spectral ensemble: FTIR_DR) and one set of Fourier transform Raman data (spectral ensemble: Raman) were collected using a Perkin Elmer System 2000 interferometer. Norton-Beer medium apodization functions were used to convolute the Fourier transform as recommended by Griffiths (1995) for quantitative work. The Raman spectra were recorded using a Nd:YAG laser emitting at 1064 nm with a laser power of 200 mW. The Raman data set was collected using a InGaAs detector and stored as Raman shifts between 3600 and 0 cm^{-1}. The 180° scattering arrangement was used and no correction for the spectral response of the instrument was applied. The FTIR_DR data set was collected between 4000 and 700 cm^{-1} using the standard diffuse reflection cell and a liquid nitrogen cooled photoconductive HgCdTe detector (MCT). Prior to each measurement the sampling station was purged with dry air for 1.5 min. to keep water and carbondioxide concentrations constant. For these mid infrared measurements the interferometer was equipped with an optimized KBr beamsplitter. All spectra collected using the interferometer are ratioed against a single beam background spectrum recorded using twice as many spectral accumulations as with the sample spectra (see Table 1) and converted to absorbance or log(1/R) units.

As indicated by Table 1 we have chosen to preserve the wavelength unit provided by the instrument as default so as to keep constant x-spacing. It should be noted that a transformation of the x-units will have no influence on the multivariate analysis as long as the original data points are preserved.

Table 1. The spectroscopic data.

	NIR	NIR_R	FTNIR_DR	FTNIR_IS	FTIR_DR	Raman
Instrument	dispersive	dispersive	FT	FT	FT	FT
Sampling method	reflectance	reflectance (rot. cell)	diffuse reflection	integrating sphere	diffuse reflection	180° scattering
Reference	ceramic	ceramic	SiO_2	$BaSO_4$	KBr	none
X-variables	1050	1050	6001	6001	3301	2771
X-units	nm	nm	cm^{-1}	cm^{-1}	cm^{-1}	cm^{-1}
X-min	400	400	4000	4000	700	330
X-max	2500	2500	10 000	10 000	1900	3100
Resolution	—	—	8	8	8	8
Accumul.	16	16	16	16	64	32
X-sampling	2	2	1	1	1	1

Chemical data

Copenhagen Pectin A/S provided a set of 10 physico-chemical properties for the 98 pectin samples. Primarily the degree of esterification or methoxylation, **%DE**, the degree of amidation, **%DA**, and the degree of free acids, **%DFA**, of the galacturonic monosaccharide in the pectic backbone. **%DE** and **%DA** are determined by titration and **%DFA** is determined by difference. **pH** is measured in a 1% pectin water solution and **transparency** is a number for the color/darkness of the sample. **SAG** is a gel strength measurement and **Ca1** to **Ca4** are gel strength measurements at different levels of calcium ions (Rolin & De Vries, 1990).

CHEMOMETRICS

Principal Component Analysis (PCA)

Principal component analysis is a fundamental method in chemometrics. By this mathematical treatment one finds the main variation in a multidimensional data set by creating new linear combinations of the raw data (e.g. spectral variables) (Andersson, 1958; Wold *et al.*, 1987; Martens & Næs, 1993). The method is especially well suited when dealing with highly collinear variables as is the case in most spectroscopic techniques: two neighbor wavelengths show almost the same variation. In Fig. 2 the principle of PCA is demonstrated by the decomposition of a matrix (**X**) containing 98 *samples* × 1050 *wavelengths* (the NIR_R spectra). The decomposition results for three selected samples are shown. As a pretreatment the **X** matrix is centered by subtracting the mean spectrum from each sample spectrum. **X** is decomposed into a score matrix (**T**) and a loading matrix (**P**), and the residuals are collected in a matrix (**E**) with the same dimensions as **X**. Due to measurement noise **X** has a full mathematical (or statistical) rank, but only the significant number of principal components (PCs), in this example two, equal to the chemical rank of the **X** matrix is relevant in describing the systematic information in **X**. In this way the information in **X** is projected onto a lower dimensional sub-space. The loading vectors for the principal components can be considered as pure mathematical spectra that are common to all the measured spectra. The loadings are also called the hidden spectra. What makes the individual raw spectra different are the amounts (scores) of hidden spectra (loadings), see Fig. 2.

Partial least squares regression (PLS)

Partial least squares regression applies to the simultaneous analysis of two sets of variables on the same objects (Martens, 1991; Martens & Næs, 1993; Høskuldsson, 1994). It allows for the modeling of inter- and intra-block relationships from an **X**-block and **Y**-

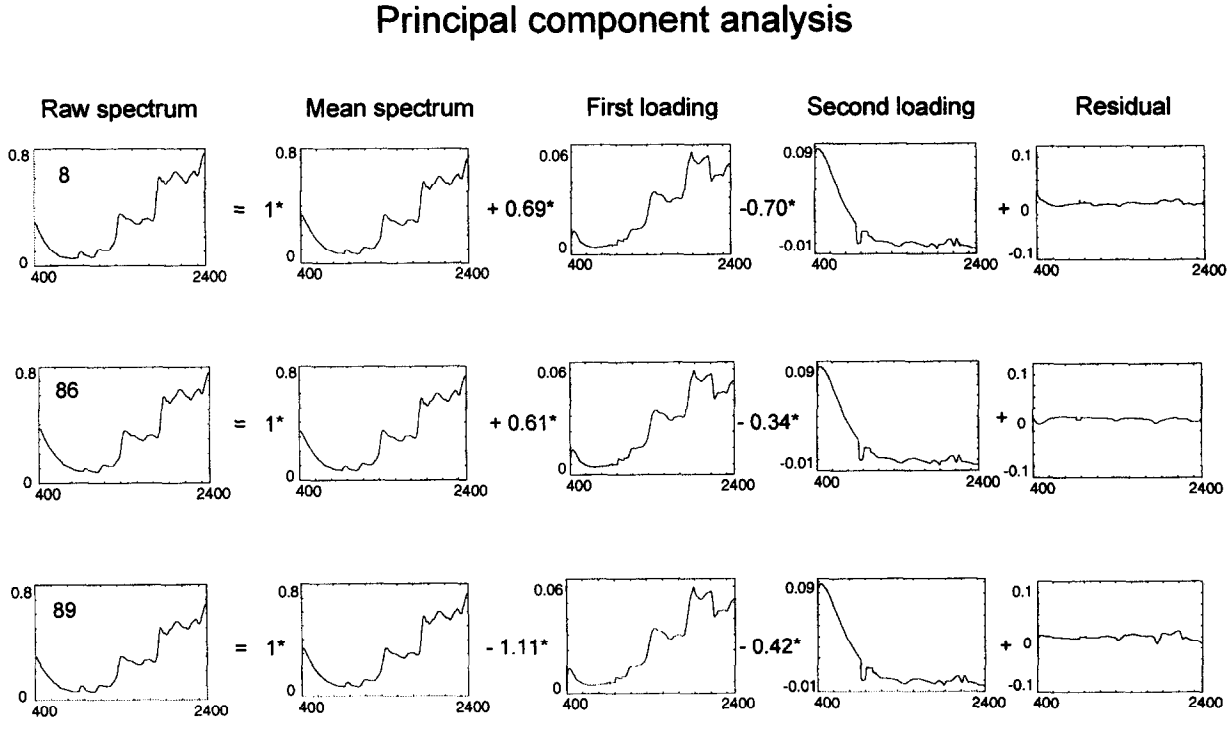

Fig. 2. PCA loadings and scores for sample numbers 8, 86 and 89 based on analysis of the mean centered NIR_R data set. The variance explained by the first two PC's is more than 90%. The mean spectrum, first and second loadings (collected in a matrix denoted **P**) are common to all the samples, while the scores (collected in a matrix denoted **T**) are different for each sample. The number in front of the loadings are the corresponding score values.

block of variables in terms of a lower-dimensional table of latent variables. PLS is commonly used in quantitative spectroscopy, and consequently applied in this work, to correlate the spectroscopic data (**X**-block — fast spectral measurements) with related physico-chemical data (**Y**-block — time consuming and laborious measurements). The main purpose of the regression is to build a predictive model as a screening tool enabling the prediction of a wanted characteristic (**y**) from a measured spectrum (**x**). In matrix notation we have the linear model $\mathbf{y} = \mathbf{Xb}$, where **b** contains the regression coefficients that are determined during the calibration step. It is noteworthy that the PLS-calculation starts with a covariance based calculation. When using mean centered data the elements of the first loading weight PLS-factor (calculated as **X'y**) are proportional to the covariance of the spectral intensities of **X** with respect to the **y** vector. This illustrates the balance between **X** and **y** during the mathematical analysis. The reader is referred to the literature for a thorough description of the algorithm (Geladi & Kowalski, 1986; Martens & Næs, 1993).

Principal variables

Principal variables (PV) is a method for the selection of variables/wavelengths in order to obtain predictive PLS-models based on a small number of variables and as an aiding tool in the interpretation of the models. The reader is referred to the literature for a thorough description (Höskuldsson, 1994; Nørgaard, 1995).

The method selects a few original variables (in this case columns corresponding to wavelengths) that describe as much of the total variance in the original data matrix as possible. Assuming that we have a matrix **X** ($s \times w$) and a vector **y** ($s \times 1$), then the first step in the PV method is to find the largest diagonal value of

$$\mathbf{X'yy'X}$$

with the dimension w × w. In this way the wavelength that co-varies most with the y-vector is chosen. This wavelength is the first principal variable, designated **w** ($s \times 1$). In the next step **X** is reduced (orthogonalised) with respect to the chosen variable

$$\mathbf{k} = (\mathbf{X'w})/(\mathbf{w'w})$$

$$\mathbf{X_{new}} = \mathbf{X} - \mathbf{wk'}$$

where **w** ($s \times 1$) and **k** ($w \times 1$) are column vectors. The next PVs are found by repeating the procedure from step 1 to 3 with $\mathbf{X_{new}}$ found in step 3. The number of PVs choosen depends on the actual problem to be solved. By analyzing the matrix **X'XX'X** in step 1 instead of **X'yy'X** the method of principal variables can be used for the selection of a small subgroup of **X** variables that correlate best with the rest of the **X** variables.

Programs

Chemometric calculations were performed using Matlab ver. 4.2c1 (Mathworks, Inc.) installed with the PLS_Toolbox ver. 1.5 (Wise and Gallagher; Eigenvector Technologies) and Unscrambler ver. 5.5 (CAMO A/S). Spectral data were collected and converted to ASCII using GRAM Research adopted for Perkin Elmer system 2000 (Galactic Industries Corporation/Perkin Elmer).

RESULTS AND DISCUSSION

Qualitative spectroscopic evaluation of the amidated pectin samples (NIR, IR, Raman)

The NIR spectra of the amidated pectins with extreme values of %DE are shown in Fig. 3a. The strongest co-variance between the NIR spectra and the %DE is found at 2228 nm (max-norm scaled co-variance of the first derivatives: *1.0*) most likely a NH combination vibration as also indicated by a negative covariance with the raw spectra. This spectral element is between the optimal wavelength found by both Horvárth *et al.* (1984) (2200 nm) and Polesello *et al.* (1990) (2256 nm) who both worked with non-amidated pectins. Two more spectral elements are found which co-vary negatively with the %DE namely a relatively broad band at about 1450 nm (*−0.3*), possibly related to the first overtone of OH vibrations, and a dual band at 1940 nm (*−0.5*) related to OH combination vibrations and 1974 nm (*−0.6*) related to NH combination vibrations.

The IR (Fig. 3b) and Raman (Fig. 3c) spectra of the amidated pectins show considerably more detail than the highly overlapping NIR spectra. Table 2 lists the most important spectral bands along with their covariance with the %DE and %DA. The co-variance is scaled to the max-norm to obtain numbers between zero and one. In the table we have also included tentative assignments primarily adapted from Séné *et al.* (1995).

Both the OH/NH-stretchings and the CH-stretchings are potential candidates for co-variate bands with respect to %DE and %DA because of the co-variance of the carboxylic OH stretching, the co-variance with the amide NH stretching and the co-variance of the methoxy CH stretching. However the CH stretching region of the Raman spectra displays only minor co-variance. In IR the strong and broadly absorbing OH- and NH-stretching vibrations, along with the water-gas vibrations, hides the relevant information.

In IR we find the strongest covariate group frequency at 1752 cm^{-1} due to the carbonyl CO stretching of the ester groups and also the carbonyl CO stretching at 1686 cm^{-1} (Amide I) and at 1650 cm^{-1} displays significant co-variation. In contrast the carbonyl stretching region in the Raman spectra displays zero co-variance.

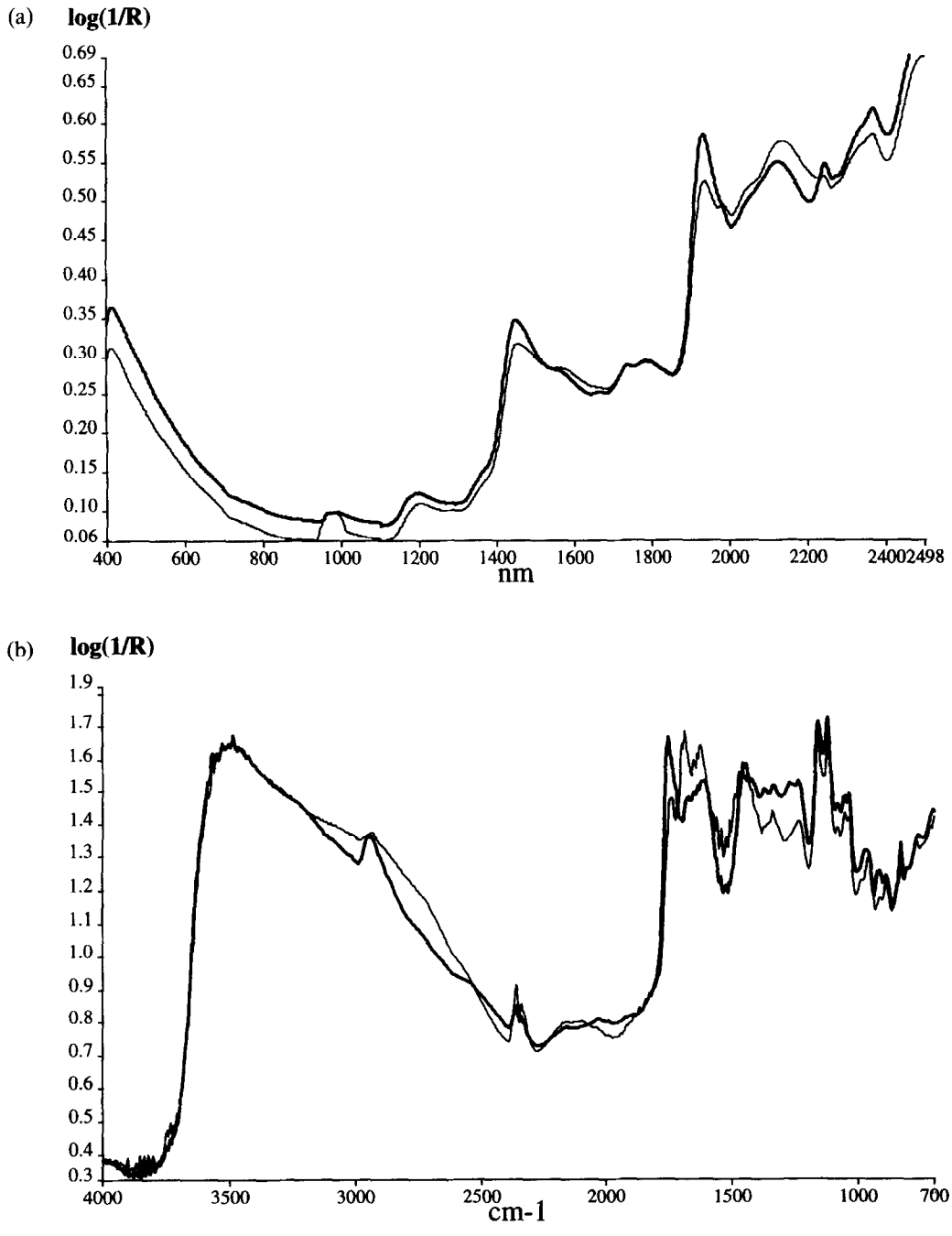

Fig. 3. (a) NIR (Rot. Cell) spectra of pectin samples with (thick) high %DE and (thin) low %DE. (b) FT-IR spectra of pectin samples with (thick) high %DE and (thin) low %DE. (c) NIR FT-Raman spectra (Raman shifted) of pectin samples with (thick) high %DE and (thin) low %DE.

In the spectral region between 1500 and 1200 cm^{-1}, dominated by complex in-plane C—O—H bendings and methyl and methylene CH bendings, we find no significant co-variate bands. More surprisingly the strong (IR) ester and ether bands at about 1160 and 1120 cm^{-1} show no significant co-variation with %DE or %DA. This could indicate that both are too strongly absorbing or alternatively that both arise from ether absorptions (see comparison with GalA, Table 2) from the glycosidic linkages and the ring oxygens and that the ester bonds are relatively too few to show absorptions.

In the highly coupled and conformational specific region between 1100 and 700 cm^{-1}, where vibrations specific to pyranose ring and polysaccharide backbone structure and interactions reside, we find two interesting co-variate features. The first is found at about 960 cm^{-1} which co-varies significantly with both %DE and %DA in IR whereas the corresponding band in Raman does not. This band was assigned as methylene and methyl rocking vibrations by Séné *et al.* (1995), but, in our opinion, the frequency is too high for such a motion and we prefer to assign it to the second ether CO band

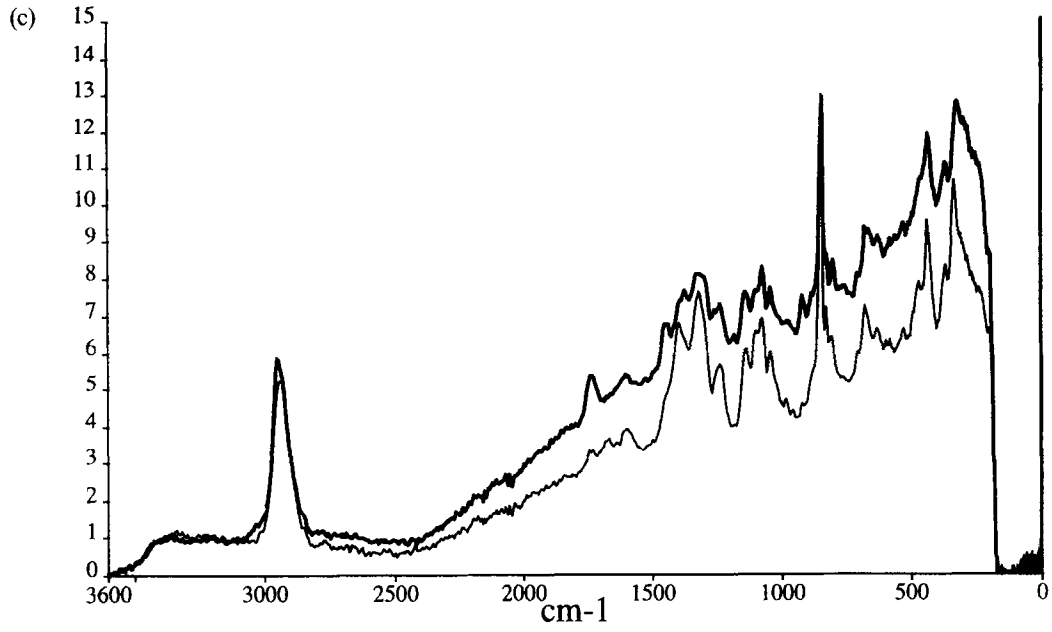

Fig. 3. Continued.

perhaps overlaid with the out-of-plane C—O—H vibrations arising from carboxyl dimers. However both possibilities can give rise to co-variance with %DE/%DA. The second interesting feature is the 850 cm^{-1} band which is found to be the strongest co-variate band in the Raman spectra whereas it is not found at all in the IR spectra. This band is known to be sensitive to the anomeric configuration (Cael *et al.*, 1974) and, in this case, is indicative of the α-form. Séné *et al.* (1995) assign this band to the symmetric C—O—C stretching of the glycosidic linkage but the presence of the band in the spectra of the monosaccharides contradicts this interpretation. Cael *et al.* (1974) performed normal coordinate analysis on D-glucose and considered this band to consist of a C1-H deformation coupled to a mixed CH_2 vibration. In a preliminary study we have found that the band at 843 cm^{-1} in α-D-glucose is not altered in methyl α-D-glucoside and that deuteration of the hydroxyl groups in methyl α-D-glucoside moves the band to 825 cm^{-1} (Engelsen, unpublished data). Accordingly this band is sensitive to both the anomeric configuration and to the substitution at C6 (CH_2OH versus CH_2OD and COOH) as is also indicated by the high co-variance between the band and the %DE. As Raman spectra in general are more sensitive to skeleton vibrations whereas IR spectra are sensitive to side groups we assign this vibration as a (C6-C5-O5-C1-O1) skeleton vibration as proposed by Sivchik and Zhbankov (1978) for α-D-glucose.

Below this frequency we find two more Raman active skeleton modes which co-vary with %DE. It is interesting and perhaps expected that all co-variant Raman bands are mainly skeleton modes only indirectly related to the group frequencies concerning the (methoxylated) carboxyl group. For this reason, if not for scattering effects, we expected relatively poor Raman models for the %DE and the %DA. In contrast IR co-variations rely mainly on the stable carbonyl stretching region suggesting good quantitative models for the %DE and the %DA.

Chemometric models

PCA of the raw physico-chemical data

To investigate the variance structure in the raw physico-chemical data a PCA was performed on the autoscaled Y-data. Fig. 4a shows a loading plot of the Y-data as a function of the 2 first PC's. In this case the first two PC's describe 57% (46% + 11%) of the total variance in the Y-matrix indicating a larger dimensionality (independent Y-variables). This is emphasized by the fact that 8 PC's are needed to describe 90% of the total Y-variance. Fig. 4a reveals that the highly complementary data %DE and %DA represent opposite and well separated directions on the maximum variance direction (the first PC) with only negligible influence of the second PC. Also the gel-strength measurements Ca1-Ca4 display a highly complementary behavior to the %DE with practically no influence of the second PC. In contrast transparency and pH are not influenced by the first PC but vary strongly with the second PC. The SAG measurements which vary in a negative direction with the second PC are also little influenced by the first PC. The corresponding scoreplot is shown in Fig. 4b indicating a gradient in the product type from '20' → '25'/'26'. A certain amount of class overlap is seen. When the scoreplot is compared with the loading plot (superimposed), it is seen that the product types '24–26' have

Table 2. Important fundamental vibrations of pectins and tentative assignments (Séné *et al.*, 1995). The covariance (cov) for the spectral elements of the spectra with respect to the %DE/%DA is indicated for both Raman and IR. Frequencies are in cm^{-1}. The letters δ, γ, ν and τ are used to designate respectively in-plane bending, out-of-plane bending, stretching and torsional motions

Frequency	Assignment	Activ.	IR	cov[1]	Raman	cov[2]	GalA[3]
~3500	ν(O-H) ν(N-H)	IR+/(R+)	3485	0/0		0/0	3360
2900±200	ν(C-H) (aliphatic)	IR+/R+	2942	±0.1/±0.1	2945–56	±0.2/±0.2	2982 2956 vs 2921
1745±10	ν(C=O) (ester)	IR+/R+	1752	±1/−1	1740	0/0	
~1720	ν(C=O) (carboxylic)	IR+/(R+)					1709 s
1670±20	ν(C=O) (Amide I)	IR+/(R+)	1686	+0.6/±0.6	1675	0/0	
1625±5	δ(NH2) (Amide II)	IR+/(R+)	1650	+0.3/−0.3			
~1455	δ(CH3) (ester)	(IR+)/R+	1458	±0.1/±0.1	1455	±0.1/±0.1	
1350±50	δ(C-O-H)	IR+/(R+)			1402	0/0	1400
1360±60	ν(COO-)	R+			1381	±0.1/0	
	δ(C-O-H)/ δ(CH3) def		1370	±0.1/±0.1			1363
~1355	δ(CH3) (acetate)		1339	0/±0.1	1332	0/0	
	δ(C-O-H)/ δ(CH2) twist				1323	±0.1/±0.1	1326
~1300	δ(CH2) wag	IR+/(R+)	1275	±0.1/±0.1			1248
~1230	Amide III/ ν(C-O) acetate	IR+/R+	1233-39	−0.1/0	1241	0/±0.1	
1200±50	ν(C-O) as (ester)	IR+	1160 vs	0/0	1186	0/0	1153
1100±50	ν(C-O) as (ether)	IR+	1120 vs	±0.1/±0.1	1144	0/0	1119
1160–970	Nonlocalized, highly coupled modes of polysaccharide backbones	IR+/R+	1082	0/0	1103	0/0	1095
			1055	±0.1/±0.1	1079	0/0	1077
			1038	±0.1/±0.1	1047	±0.2/±0.2	1056 1030
					987	±0.1/0	981
~950	ν(C-O)/ γ(COOH) (dimer)	IR+/(R+)	962	±0.3/±0.4	960	0/±0.1	
	δ(CH2)/ δ(CCH)/ δ(COH)	IR+/R+	919	0/±0.2	923	±0.2/±0.2	938
		IR+/R+	891	−0.2/±0.1	890	0/0	900
~850	(C6-C5-O5-C1-O1) skeleton	R+			851 vs	±1/±1	863 vs
~800	δ(NH2) wag	IR+/R+	834	±0.1/±0.1	833	0/±0.1	821
	δ(NH2) wag	R+			806	±0.2/±0.2	
	γ(R-O-H)	IR+/(R+)	771	±0.1/±0.1			773
700±50	γ(R-O-H)	IR+/(R+)			715	0/0	
684	ν(C-O-C)	R+			683	0/0	698 s
		R+			537	0/0	
		R+			477	0/±0.1	
441	τ(-C-O-C-) def.	R+			442	±0.2/±0.2	405
		R+			378	0/0	359
336	τ(-C-O-C-) def.	R+			339–330	±0.4/±0.4	

1 Scaled covariance between the IR intensities of the first derivatives of the spectral element and %DE/%DA
2 Scaled covariance between the Raman intensities of the first derivatives of the spectral element and %DE/%DA
3 The corresponding (Raman) spectral elements of D-Galacturonic acid • H_2O

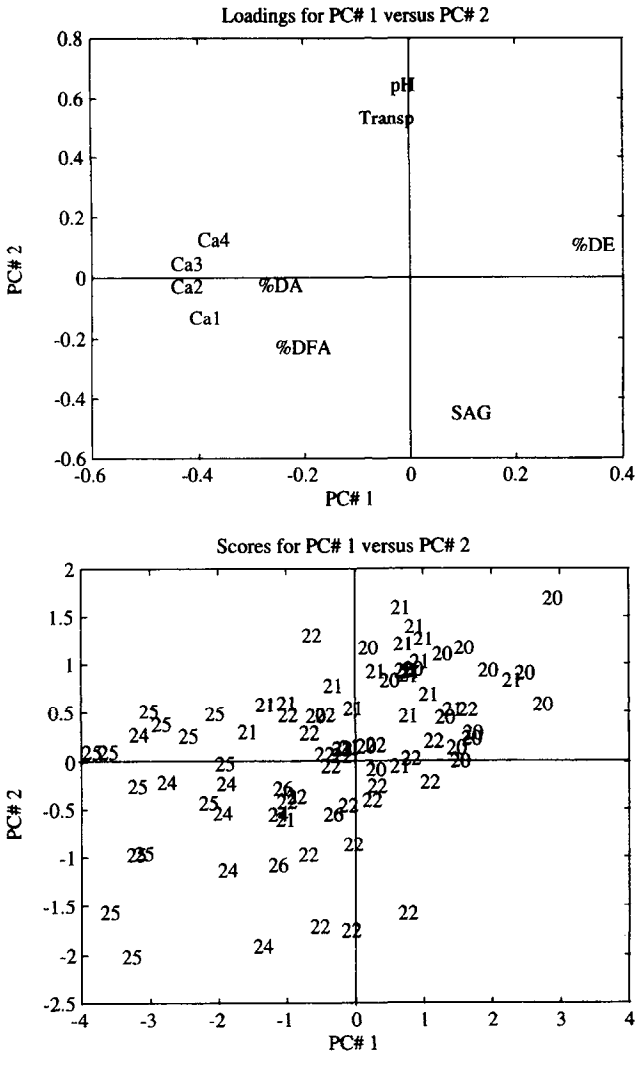

Fig. 4. (a) Loading plot from a PCA on chemical data measured on the pectins (%DE = degree of esterification, %DA = degree of amidation, %DFA = degree of free acid, Transp = transparency, Ca1-Ca4 and SAG are gel strength measures). (b) Score plot from a PCA on chemical data measured on the pectins showing different product sorts ('20'–'22', '24'–'26').

large levels of the gel strength measures (Ca1-Ca4) as well as %DE and %DA, while the opposite holds for the sample types labeled '20'.

PCA of the NIR spectra

A PCA on the NIR spectra (2nd derivative) shows that the spectral data can be used for quality control of the pectins. Analogue analysis can be performed using the other spectroscopic ensembles. The score plot in Fig. 5 shows a more distinct product sort separation than the analysis based on the chemical variables (Fig. 4b) indicating that the spectral technique(s) provide extremely relevant information about the samples. This is a quality of fast spectral techniques not widely appreciated, but, by the use of chemometrics, it will be in the future years. In Fig. 5 different product sorts ('20'–'22','24'–'26') are marked in the score plot of spectral NIR measurements. The score plots in Fig. 4b and 5 are zoomed plots focusing on the most difficult region; other product sorts are found in well separated classes outside the range of these plots.

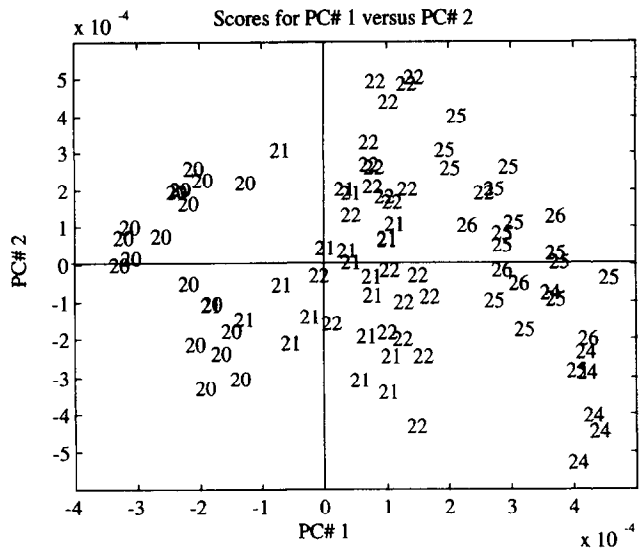

Fig. 5. Score plot from a PCA on NIR_R spectroscopic data showing different product sorts sorts ('20'–'22', '24'–'26').

The PLS models for %DE

The best PLS models obtained using the five different spectral sources are listed in Table 3. To construct the models we have used 73 samples (objects) with systematic cross validation (10 segments (Unscrambler manual)). We have tested the resulting models on an *independent* test set consisting of 25 samples. The test set was selected as 1 every 4th sample of the total set of 98 samples to minimize effects of sample variation. Prior to the application of PLS the X-matrices were mean centered. While building the PLS models, the performance can be optimized by removing sample spectra with defects and extreme samples not representative for the bulk samples. In this work it was decided not to remove concentration outliers to cover the maximal possible range of %DE and %DA. In general spectral PLS models will also benefit from the removal of noisy spectral regions. Accordingly the IR spectra were truncated to include only the frequency range from 1900 cm^{-1} to 700 cm^{-1} and the Raman spectra were truncated to include only the frequency range from 3100 cm^{-1} to 330 cm^{-1}. The improvements of these truncations were hardly significant but improved the speed of the calibration process. No truncation of the NIR spectra were performed.

Using the raw spectra the models performed well even though many PCs were needed to describe the spectral levels and scattering effects. Using instead the first or second derivative spectra, the models converged much faster and were often simpler (using a lower number of

Table 3. The best spectroscopic PLS models for the *degree of esterification* of amidated pectin powders. The models were based on 73 objects (samples) and the prediction ability tested on 25 independent objects. Systematic cross validation (10 subsets) were used to validate the models

	NIR	NIR_R	FTNIR_DR	FTNIR_IS	FTIR_DR	Raman
obj/var	*73/1050*	*73/1050*	*73/6001*	*73/6001*	*73/1201*	*73/2771*
derivative	2	1	1	1	1	1
outliers[1]	1	1	1	1	3	8
opt. PC[2]	2	3	3	2	4	4
y-var.[3]	96	96	93	94	93	90
RMSECV[4]	**1.19**	**1.17**	**1.54**	**1.54**	**1.55**	**1.90**
α(P/M)[5]	0.95	0.95	0.94	0.92	0.93	0.84
R[6]	0.98	0.98	0.97	0.97	0.97	0.95
obj/var	*25/1050*	*25/1050*	*25/6001*	*25/6001*	*25/1201*	*25/2771*
RMSEP[7]	**1.58**	**1.54**	**1.84**	**1.57**	**2.07**	**2.03**
α(P/M)[5]	0.96	0.95	0.98	0.92	0.91	0.84
R[6]	0.98	0.98	0.97	0.98	0.96	0.96

[1] Outliers are number of samples/spectra which deviate from the general model (not used in the final model)
[2] Number of principal components which is used to describe the model
[3] Per cent explained validated y-variance
[4] RMSECV is the root-mean-squared error of cross validation
[5] α(P/M) is the slope of the regression line in a predicted versus measured plot
[6] R is the correlation coefficient
[7] RMSEP is the root-mean-squared error of prediction

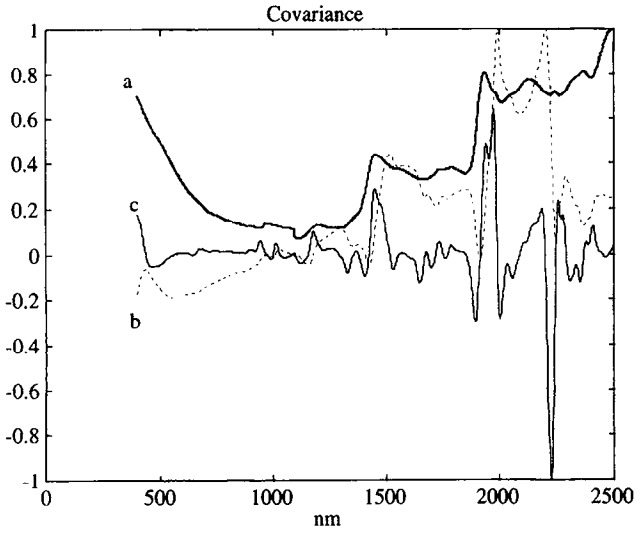

Fig. 6. The influence of derivative spectra on the covariance of the NIR_R ensemble with %DE. (a) the raw spectrum, (b) the covariance of the raw spectra with %DE and (c) the covariance of the first derivative of the spectra with %DE.

PLS factors). This derivative effect is due to the fact that the 1st derivative spectra lead to removal of constant scatter and 2nd derivatives lead to removal of 'growing' scatter. In first derivative spectra we observe that the co-variance is leveled out in (low) non co-variant spectral regions. This derivative effect is illustrated in Fig. 6. Although second derivative spectra are abundantly used in quantitative NIR spectroscopy we found only one case where the second derivative spectra performed better than the first derivative spectra.

The effect of using different pre-transformations (in this work: derivatives) and excluding outliers on the FTNIR_IS spectral set is shown in Table 4. In this special case one can argue whether the raw spectra or the first derivative spectra perform best. Actually, the model using the FTNIR_IS raw spectra produced the best predictions on the test set. In general there is no reason to reject the use of many PCs as long as the modeling on the test set performs well. Table 4 also reveals the effects of removing outliers from the model set. In column 3 some of the most extreme spectral and/or sample outliers have been removed but although the model improved to RMSECV = 1.20 the predicting ability on the test set deteriorated to RMSEP = 1.81. Apparently these outliers include some defects which are also represented in the test set. In column 4 also the concentration outliers were removed and the model improved little further to RMSECV = 1.18 but again the performance on the test set was worse, RMSEP = 1.66, than that of the less 'refined' model. In the last column we tried as an experiment to switch the test set (25 objects) and the validation set (73 objects). Surprisingly this model performed better than any of the other models on the test set, resulting in a RMSEP of 1.36. This result suggest that experimental errors plays an important role in the best models. Unfortunately the experimental errors on the reference %DE values were not measured. Taking into consideration the long time span of the sample collection and the involvement of different lab-workers the experimental error was estimated to be between 1 and 2% (Bjerrum, K., Copenhagen Pectin Factory). As the best models converge to an RMSEP of about 1.5%, further improvements of the calibration procedures cannot be validated.

Table 3 indicates that the dispersive NIR ensembles (NIR and NIR_R) result in the best cross validated

Table 4. The influence of *outliers and pre-transformations* of the spectroscopic data. PLS models for the degree of esterification of amidated pectin powders based on the FTNIR_IS spectra. The models were built using 73 objects (samples) and the prediction ability tested on 25 independent objects. Systematic cross validation (10 subsets) were used to validate the models. For symbol description see Table 3

	FTNIR_IS					
obj/var	*73/6001*	*73/6001*	*73/6001*	*73/6001*	*73/6001*	*25/6001*
derivative	0	1	1	1	2	1
outliers	1	1	4	10	—	1
opt. PC	6	2	2	2	5	2
RMSECV	**1.52**	**1.54**	**1.20**	**1.18**	**2.50**	**1.58**
α(P/M)	0.93	0.92	0.95	0.91	0.68	0.94
R	0.97	0.97	0.98	0.96	0.93	0.97
obj/var	*25/6001*	*25/6001*	*25/6001*	*25/6001*	*25/6001*	*73/6001*
RMSEP	**1.50**	**1.57**	**1.81**	**1.66**	**3.13**	**1.36**
α(P/M)	0.94	0.93	0.92	0.95	0.63	0.93
R	0.98	0.98	0.97	0.97	0.93	0.98

models. The potential advantages of Fourier transform spectroscopy (Felgett advantage, Jacquinot advantage, Connes advantage, constant resolution) (Griffiths & de Haseth, 1986) are in practice outnumbered by a more reproducible setup and sampling procedures. In the dispersive NIR measurements the background is scanned dynamically (shuttle bus) and in the 'small sample cup' the pectin powder is pressed up against a quartz window with a compressible 'paper lock'. Comparison of the spectra and the NIR and NIR_R columns in Table 3, indicates that the rotating device further enhances the reproducibility and thus the PLS modeling. Concerning the predicting ability the FTNIR_IS (Fig. 7) and the dispersive NIR models perform similarily. Presumably a design with dynamically background scans and a reproducible powder surface (pressure control) would further enhance the quantitative performance of this sampling technique. It is interesting to note that predicting performance (equal to the best dispersive NIR models) of the integrating sphere is close to the cross-validated performance. The standard diffuse reflection cell is in general too unstable to provide high reproducibility over a large time period as also indicated by the FTNIR_DR model performance. Finally, with regard to comparisons between the Fourier transform and dispersive procedures it should be noted that the concentration range of the esterification was between 20 and 55%. It is therefore possible that Fourier transform instruments, due to their higher throughput and signal to noise ratio, could perform better than dispersive instruments for smaller concentrations close to the detection limit. However, in standard applications, such as those investigated here, it is not the instrument sensitivity but rather the reproducibility of the sampling technique which is the limiting factor for precise quantitative work using chemometrics.

When considering the two spectral models in the region of the fundamental vibrations (FTIR_DR and Raman) we observe that more PC's are needed to level out noisy regions and describe the y-variation. The FTIR_DR model does provide an equally good cross-validated model as the FTNIR_DR but the predicting ability seems worse. The use of the diffuse reflection cell makes it difficult to evaluate the quantitative advantages or disadvantages of the two different spectral regions, but little is gained by measuring in the more difficult IR region. Although the predictive performance was equal to that of the FTIR_DR ensemble the Raman model was rather sensitive to spectral outliers, and proved to be the worst in comparison with the other ensembles. The lower signal to noise ratio in the Raman spectra caused by the low energies of the scattered light and the poor reproducibility of the spectra due to scattering effects are probably responsible for the poor performance of this spectral technique. However, it is possible that the sampling technique has a major influence on the result. The incident laser light is focused on a very tiny area on the sample and the depth focusing is not controlled nor is the pressure/density of the sample. We are currently taking steps to improve and control these factors to obtain a fairer assessment of the poten-

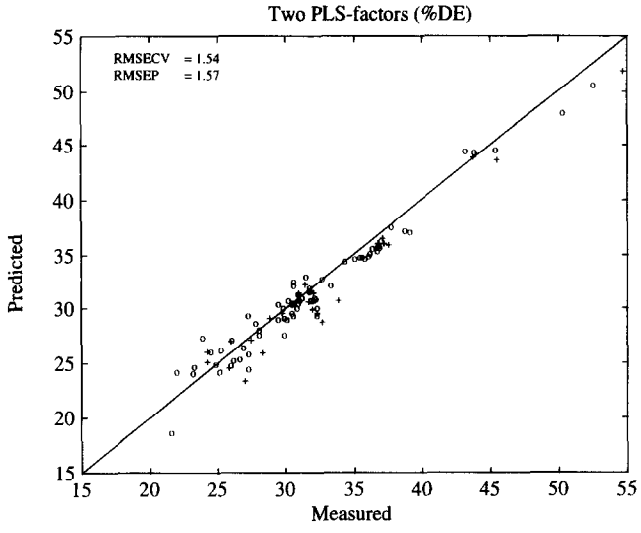

Fig. 7. The cross validated FTNIR_IS model of %DE (o) and the performance on the test set (+).

Table 5. The optimal spectroscopic PLS models for the *degree of amidation* of amidated pectin powders. The models were based on 73 objects (samples) and the prediction ability tested on 25 independent objects. Systematic cross validation (10 subsets) were used to validate the models. For symbol description see Table 3

	NIR_R	FTNIR_IS	FTIR_DR	Raman
obj/var	*73/1050*	*73/6001*	*73/1201*	*73/2771*
derivative	1	1	1	1
outliers	5	4	3	8
opt. PC	3	2	4	4
y-var. expl.	92	91	90	81
RMSECV	**1.17**	**1.26**	**1.36**	**1.73**
α(P/M)	0.95	0.92	0.91	0.78
R	0.96	0.95	0.95	0.90
obj/var	*25/1050*	*25/6001*	*25/1201*	*25/2771*
RMSEP	**1.11**	**1.20**	**1.29**	**2.08**
α(P/M)	1.16	0.95	0.92	0.75

tial of FT-NIR Laser-Raman spectroscopy in quantitative determinations.

The PLS models for %DA

The best PLS models obtained using the four different spectral sources are listed in Table 5. In this case the first derivative spectra also resulted in the best PLS models. The three spectral ensembles NIR_R, FTNIR_IS and FTIR_DR seem to provide models of quality equal to an RMSEP of about 1.20% using 3, 2 and 4 PCs respectively. The model created from the dispersive NIR_R ensemble (see Fig. 8) being slightly more precise. It is noteworthy that the RMSEPs in all three cases are improved in comparison to the RMSECVs in contrast to the best %DE models. This is probably a reflection of the problems of selecting a statistically representative calibration set and/or test set from a relatively small population. Furthermore, the %DA models display lower RMSEPs when compared to the %DE models indicating either more stable spectral information or more precise reference y-measurements. The latter possibility is supported by the fact that these models describe less of the y-variance when compared to the %DE models (Tables 3 and 5).

PLS models of pH and Ca2 gel strength

It was also possible to model **pH** and, to some extent, the gel strength **Ca2** from the spectroscopic ensembles. In these cases the FTIR_DR models were as good as the best NIR models whereas the Raman models were inferior (very poor for the **Ca2**). For this reason and because IR models give better possibilities of interpretation we have chosen only to present the FTIR_DR (first derivative) models of **pH** and **Ca2**. The sample points in the cross-validated PLS model of the **pH** from the FTIR_DR ensemble shown in Fig. 9 vary slightly

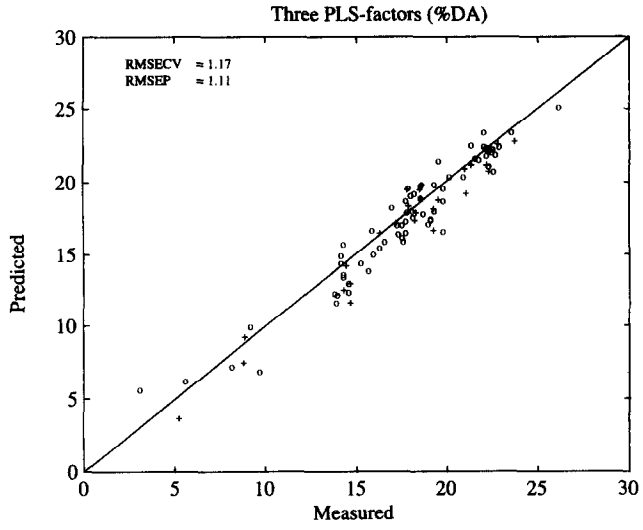

Fig. 8. The cross validated NIR_R model of %DA (o) and the performance on the test set (+).

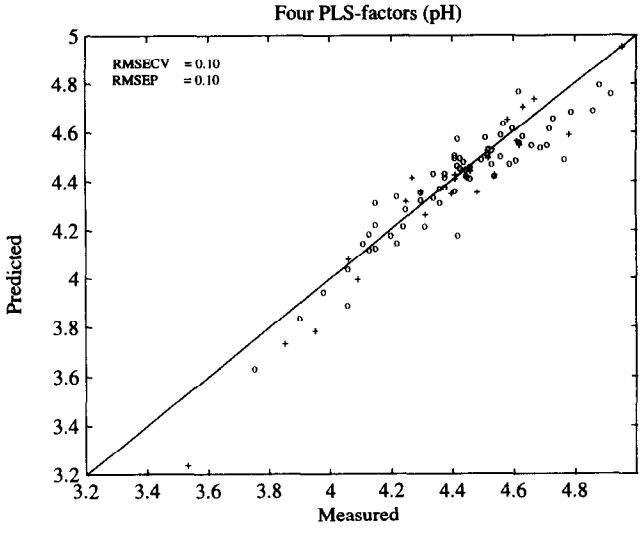

Fig. 9. The cross validated FTIR_DR model of pH (o) and the performance on the test set (+).

from the ideal line. Such behavior indicates that the PLS modeling will benefit from a pre-transformation of the y-data to produce data with a more normal (or uniform) distributed profile. However, since this model was of minor concern to us we accepted the model with a RMSEP of about 0.1 pH using four PCs accounting for 84% of the total y-variance.

Finally we found a PLS model for **Ca2** gel strength (Fig. 10) describing 60% of the y-variance using 3 PCs. This is certainly not an impressive model for the gelation potential, but we must emphasize that the pectin sample set was not designed for gel strength modeling and included several classes which in the Ca2 test, had to be treated with different levels of sugar contents. This together with the fact that the different pectin systems may have their optimum gel strength at different calcium levels, might be responsible for the relatively

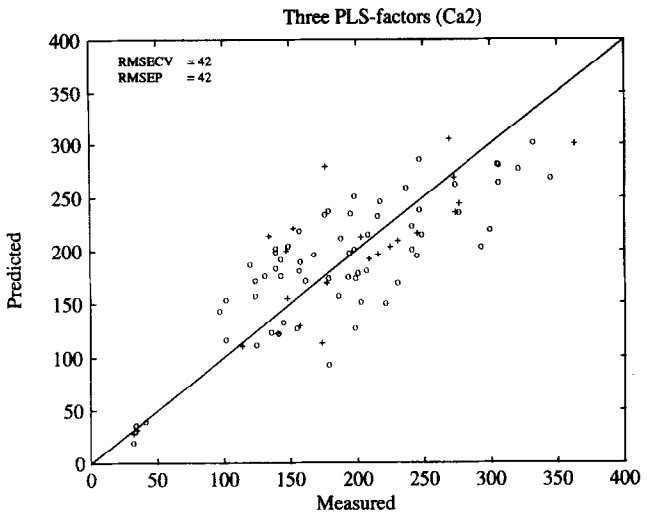

Fig. 10. The cross validated FTIR_DR model of Ca2 (o) and the performance on the test set (+).

poor model which will have to be investigated in future experiments. However, it is perhaps more likely that the remaining 40% of the y-variance not accounted for in the model is hidden information not available in the vibrational spectra of the amidated pectins. One possible explanation of the fact that the gelation potential is not fully present in the spectral information may be the importance of the distribution of free acid groups with respect to methoxylated/amidated galacturonic acids. This explanation could be supported by the fact that blockwise distribution of ester groups in pectins results in higher calcium reactivity than pectins with the random distributed estergroups at the same degree of esterification (Bjerrum, K., Copenhagen Pectin Factory, unpublished data). This **Ca2** PLS model which can describe 60% of the y-variance and provide stable results on an independent test set can be useful if the resulting RMSEP is acceptable. There is no *a priori* reason to expect such a model to be less stable than a model which can describe say 95% of the y-variance.

Spectral elements influencing the FT-IR PLS models

The PLS model allows evaluation of the most influential spectral elements by constructing a so-called residual variance versus leverage (distance from the orthogonal projection of the x-variable on the PC model to the center of the model) plot of the x-variables. In this plot the more influential wavenumbers will be distinguished by having a large leverage and a small residual variance.

Table 6 lists the seven most influential wavenumbers in the FTIR_DR ensemble with respect to the %DE, %DA, %DFA, pH and Ca2 models. The table demonstrates the importance of the carbonyl stretching region to quantitative IR models and all models use the derivative carbonyl stretches at 1774, 1703–1715 and 1656 cm^{-1}. The most influential spectral element in all models is centered at 1774 cm^{-1} and can be interpreted as the left side lobe of the carbonyl stretching of saturated esters found in the raw pectin spectra at 1752 cm^{-1}. The %DA and %DFA models are strongly influenced by this spectral element suggesting that these models get their primary information from the variation in %DE. The second most influential spectral element is centered at 1703 cm^{-1} except for the %DFA model. This band did not appear as a peak in the raw spectra (see Table 2) but is most probably the right side lobe of the strongly IR active out-of-phase carbonyl stretch expected at about 1720 cm^{-1} for the free acid form. Carboxylic dimers which are usually formed in condensed phases will in theory give rise to a splitting of the carbonyl levels in a Raman active in-phase stretch at about 1650 cm^{-1} and a strongly IR active out-of-phase stretch around 1720 cm^{-1}. The third most influential spectral element centered at 1656 cm^{-1} is located in a region where the in-phase carbonyl stretchings characteristic of carboxyl groups are expected (see Table 2). However, since the in-phase stretch is only Raman active (for symmetric molecules) it is more likely related to the right sidelobe of the amide I band at 1686 cm^{-1} or to the corresponding amide II band expected to be found around 1630–1620 cm^{-1}. Of the less influential spectral elements it is interesting that the band at 937 cm^{-1} influences the %DE, pH and Ca2 models and can perhaps, on this basis, be attributed to the out-of-plane C—O—H vibrations in carboxylic dimers.

It is perhaps interesting to note that all the corresponding Raman models were most strongly influenced by the 850 cm^{-1} spectral element. For this reason, it is possible that this strong and relative sharp band can be very useful in future quantitative Raman measurements of carbohydrates.

Table 6. The most influencial wavenumbers in the FTIR_DR models as determined by variance versus leverage plots. Wavenumbers are in reciprocal centimeters. The sequence σ_1, σ_2, σ_3 etc. indicates decreasing influence

model	PC	y-var	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
%DE	4	93	1774	1703	1654	1711	1685	937	1751
%DA	4	90	1774	1703	1655	1686	1711	1544	1164
%DFA	7	73	1774	1708	1656	716	1686	1465	1549
pH	4	84	1775	1703	1656	1713	1680	1753	937
Ca2	3	60	1774	1703	1656	1715	1167	1614	937

Selection of spectral FT-IR wavenumbers correlating with %DE

To investigate both the spectroscopic and chemometric selection of spectral variables correlating with %DE, four different selection approaches were chosen on the FTIR_DR data set: (a) knowledge based (KB) selection of wavenumber ranges proposed in the literature for relevant group frequencies (see the above section on qualitative evaluation of the spectral information); (b) co-variance based (CB) selection of single wavenumbers by visual inspection of the spectra and co-variances given in Table 2; (c) selection of single wavenumbers by the chemometric method principal variables (PV), and (d) selection of single wavenumbers based on PLS influence plots (IP) (see above section). In Table 7 the wavenumber ranges are given for comparison. PLS %DE models were calculated for the chosen variables (Table 7) as above by mean centering the data. The PV and IP methods retain the predictive ability for %DE; the PV method being superior even compared with the full spectrum model (see Table 3) demonstrating the advantage of chemometric selection of variables.

Table 7. Selected wavelength models. The models are as follows: (KB) Knowledge based: 1150–1250, 1345–1365, 1445–1465 and 1735–1755 cm^{-1}; (CB) covariance based: 1160, 1339, 1458 and 1752 cm^{-1}; (PV) principal variables: (selected order) 1774, 1704, 1657, 1684, 1748, 1669, 1469, 1616, 744 and 1644 cm^{-1} and (IP) influence based: (selected order) 1774, 1703, 1654, 1711, 1685, 937 and 1751 cm^{-1}. For symbol description and comparison with the full spectral model see Table 3

Selection	KB	CB	PV	IP
obj/var	*73/164*	*73/4*	*73/10*	*73/7*
derivative	1	1	1	1
outliers	3	3	3	3
opt. PC	3	4	5	6
RMSECV	**2.31**	**5.31**	**1.44**	**1.72**
R	0.92	0.46	0.97	0.96
obj/var	*25/164*	*25/4*	*25/10*	*25/7*
RMSEP	**3.34**	**6.73**	**1.91**	**2.15**
R	0.88	0.37	0.96	0.95

2D IR-NIR correlations of amidated pectins

2-D NIR/IR correlation spectra have been published by Barton *et al.* (1992) using five ryegrass silage cell wall fiber samples and by Windig *et al.* (1995) using four component constructed samples. The availability of NIR and IR spectral ensembles of 98 samples measured using the same sampling technique with the same instrument allowed us to construct a 2D correlation spectrum from the FTIR_DR and FTNIR_DR ensembles. The resulting 2D correlation spectrum displaying only correlations (R^2) above 0.5 is shown in Fig. 11. This spectrum reveals that combination tones rather than overtones are predominant in the NIR spectra as indicated by the lines showing double (n = 2) and quadrapole overtones (n = 4). Furthermore the spectrum shows that practically all the information in the NIR spectra is found below 7500 cm^{-1} (1333 nm) so that for practical purposes wavenumbers under this limit can be discounted. Finally, it is interesting to note that the most influential wavenumber in the spectroscopic PLS models: 2228 nm (4488 cm^{-1}) is a very complex combination tone band with strong correlations with the IR regions: 850–1000 cm^{-1}, 1250–1300 cm^{-1}, 1650–1800 cm^{-1} and 2900–3000 cm^{-1} and as such includes the most informative bands regarding ester, carbonyl and aliphatic stretches.

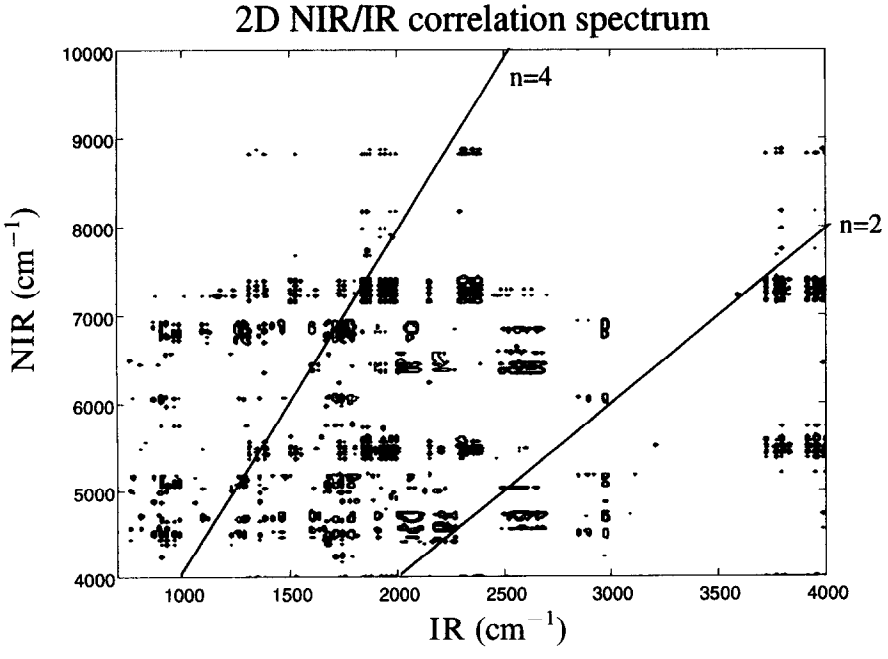

Fig. 11. The 2D NIR/IR correlation-spectrum of amidated pectins made from the first derivative spectra.

CONCLUSIONS

The present investigation has shown the possibility of creating a fast spectroscopic method for determining quality parameters in amidated pectins. The samples were collected from a broad range of different batches of raw materials and covered a long time period. Both NIR and IR regions in the electromagnetic spectrum provided sufficient information to create good predictive PLS models for %DE and %DA in amidated pectins. These PLS models all fall in the range of the estimated experimental uncertainty of the reference measurements. The results also indicate that the Raman effect contains less information (as indicated by the difference in explained y-variance of the models). The latter result is attributed to the fact that Raman spectra are more sensitive to skeleton vibrations than to side group vibrations. In addition modeling studies have shown the effect on 6-*O* methyl ester substitution upon the pectic backbone structure is negligible (Cros *et al.*, 1992). However, it is also possible that the less quantitative information in the Raman spectra is due to the analytical sampling procedure or to the low signal to noise ratio inherent to Raman spectroscopy.

It should be possible to improve the models and the calibration if a more homogeneous sample set is selected and if the chemical reference determinations are performed in a continuous series including determination of standard deviations. In order to obtain perhaps even better predictive PLS models future investigations should focus on the connection between different types of spectroscopy (NIR, IR and Raman) by methods such as 2D-correlation spectroscopy (Barton *et al.*, 1992; Windig *et al.*, 1995), PLS2 modeling (Martens & Næs, 1993), canonical correlation (Devaux *et al.*, 1993) and Procrustes rotation (Vigneau *et al.*, 1995) as well as the combination of the data from some of these spectroscopies.

A most important result of the present investigation is the observation, that models built with a few (10) wavenumbers selected by PV (principal variables) methods were found to be superior in performance compared to models with knowledge based selection of wavenumbers (and wavenumber ranges) and equal to (or even slightly better than) the full spectral PLS models. Unlike the knowledge based selection methods the unbiased PV method is able to 'navigate' through the complex spectra of the pectin samples and extract the relevant information. The fact that knowledge based selections are inferior to the PV selection can be attributed mainly to the complex pectin samples. These are semi-manufactured pectins extracted from different batches of raw materials and, as such, are assumed to be micro-heterogeneous with respect to parameters such as degree of polymerization, hairy region contents and trace materials. In future research it will be interesting to compare the PV selection method with more elaborate search (and selection) methods such as genetic algorithms (Hibbert, 1993) and simulated annealing methods (Hörchner & Kalivas, 1995).

Finally, it is interesting that it proved possible to develop a PLS model (based on the FTIR_DR ensemble) describing 60% of the y-variance in the Ca2 gel strength from such a heterogeneous material. We hope to pursue this result with further experiments with a much more homogeneous sample set, perhaps including spectroscopic sampling in gel phase.

ACKNOWLEDGEMENTS

We are much indebted to Klaus Bjerrum, Copenhagen Pectin A/S (Hercules Inc.), who provided us with the pectin samples and related physico-chemical data and who from the beginning saw the possibilities in the chemometric approach to batch or process quality control of pectin manufacturing. This investigation was made possible by funds from the Føtek programme to professor Lars Munck and by funds from the Danish Research Councils and the Department of Education to the Danish Food Research Center for Advanced Studies (LMC).

REFERENCES

Anderson, T.W. (1958). *An Intruduction to Multivariate Statistical Analysis.* Wiley Publications in Statistics, John Wiley & Sons, Inc.

Barton II, F.E., Himmelsbach, D.S., Duckworth, J.H. & Smith, M.J. (1992). *Applied Spectroscopy,* **46,** 420–429.

Cael, J.J., Koenig, J.L. & Blackwell, J. (1974). *Carbohydr. Res.,* **32,** 79–91.

Carpita, N.C. & Gilbeaut, D.M. (1993). *The Plant Journal,* **3,** 1–30. And references therein.

Carr, J.M., Sufferling, K. & Poppe, J. (1995). *Food Technology,* **7,** 41–44. And references therein.

Cros, S., Hervé du Penhoat, C., Bouchemal, N., Ohassan, H., Imberty, A. & Pérez, S. (1992). *Int. J. Biol. Macromol.,* **14,** 313–320. And references therein.

Devaux, M.F., Robert, P., Quannari, A., Safar, M. & Vigneau, E. (1993). *Applied Spectroscopy,* **47,** 1024–1029.

Engelsen, S.B., Cros, S., Mackie, W. & Pérez, S. (1996). *Biopolymers,* **39,** 417–433.

FAO (1992). *Food and Nutritional Paper,* **52.**

Geladi, P. & Kowalski, B.R. (1986). *Analytica Chimica Acta,* **185,** 1–17.

Góral, J. & Zichy, V. (1990). *Spectrochimica Acta,* **46A,** 253–275.

Griffiths, P.R. & de Haseth, J.A. (1986). *Fourier Transform Infrared Spectroscopy,* Chemical Analysis Vol. 83, Wiley – Interscience.

Griffiths, P.R. (1995). Lecture Notes given at the Bowdoin College International Infrared Course, Perkin Elmer, Lerum, Sweden.

Hibbert, D.B. (1993). *Chemometr. Intell. Lab. Syst.,* **19,** 277–293.

Horvárth, L., Norris, K.H., Horvárth-Mosonyi, M., Rigó, J. & Hegedüs-Völgyesi, E. (1984). *Acta Alimentaria,* **13,** 355–382.

Hörchner, U. & Kalivas, J.H. (1995). *J. Chemometrics,* **9,** 283–308.

Höskuldsson, A. (1994). *Chemometr. Intell. Lab. Syst.*, **23,** 1–28.

Martens, H. (1991). *Spectroscopy World,* **34,** 26–27.

Martens, H. & Næs, T. (1993). *Multivariate Calibration*, Wiley, New York.

Nørgaard, L. (1995). *Talanta,* **42,** 1305–1324.

Polesello, A., Giangiacomo, R., Forni, E. & Braga, F. (1990). *Carbohydrate Polymers,* **12,** 27–38.

Rolin, C. & De Vries, J. (1990). In *Food Gels* (Ed. P. Harris), Elsevier Applied Science, London, pp. 401–434. And references therein.

Séné, C.F.B., McCann, M.C., Wilson, R.H. & Grinter, R. (1995). *Plant Physiol.,* **106,** 1623–1631. And references therein.

Sivchik, V.V. & Zhbankov, R.G. (1978). *Zhurnal Prikladnoi Spektroskopii,* **28,** 1038–1045.

Windig, W., Margevich, D.E. & McKenna, W.P. (1995). *Chemometr. Intell. Lab. Syst.,* **28,** 109–128.

Vigneau, E., Devaux, M.F. & Safar, M. (1995). *J. Chemometrics,* **9,** 125–135.

Wold, S., Esbensen, K. & Geladi, P. (1987). *Chemometr. Intell. Lab. Syst.*, **2**, 37–52.