Carbohydrate Research 344 (2009) 1833–1841

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Quantification of blockiness in pectins—A comparative study using vibrational spectroscopy and chemometrics

Hanne Winning [a,*], Nanna Viereck [a], Tina Salomonsen [a], Jan Larsen [b], Søren B. Engelsen [a]

[a] Quality and Technology, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark
[b] CP Kelco, Ved Banen 16, 4623 Lille Skensved, Denmark

ARTICLE INFO

Article history:
Received 23 July 2008
Received in revised form 7 October 2008
Accepted 9 October 2008
Available online 1 November 2008

Keywords:
Pectin
Infrared spectroscopy
Raman spectroscopy
Near infrared spectroscopy
Chemometrics
Blockiness

ABSTRACT

The gelling properties of pectins are related not only to the degree of esterification (DE), but also to the distribution of the ester groups. In this study, we have examined an experimentally designed series of 31 pectins originating from the same mother pectin and de-esterified using combinations of two different enzymatic mechanisms. The potential of using infrared (IR), Raman, and near infrared (NIR) spectroscopies combined with chemometrics for reliable and rapid determination of the DE and distribution patterns of methyl ester groups in a designed set of pectin powders was investigated. Quantitative calibration models using partial least squares (PLS) regression were developed and compared. The calibration models for prediction of DE obtained on extended inverse signal correction (EISC)-treated spectra of all three spectroscopic methods yielded models with cross-validated prediction errors (RMSECV) between 1.1%p and 1.6%p DE and correlation coefficients of 0.99. A calibration model predicting degree of random de-esterification (R) and block de-esterification (B) was developed for each spectroscopic method, yielding RMSECV values between 4.4 and 6.7 and correlation coefficients (r) between 0.79 and 0.92. Variable selection using interval PLS (iPLS) significantly improved the prediction of R for IR spectroscopy, yielding RMSECV of 3.5 and correlation coefficients of 0.95. All three spectroscopic methods were able to distinguish the spectral patterns of pectins with different enzyme treatments in simple classification models by principal component analysis (PCA). Extended canonical variate analysis revealed one specific signal in the Raman (1045 cm^{-1}) spectrum and one significant area (1250–1400 cm^{-1}) in the IR spectrum which are able to classify the pectin samples according to the four different enzyme treatments. In both Raman and IR spectra, the signal intensity decreased in the sequence R-B > B > B-R > R > re-methylated pectin.

1. Introduction

Pectins are a family of complex polysaccharides present in all plants, providing strength and flexibility to the plant cellwall.[1] Commercially, pectins are extracted from citrus peel and apple pomace and are utilized in the food industry because of their excellent gelling, thickening, and stabilizing properties.[2,3] Pectin molecules consist primarily of homogalacturonan (HGA) and rhamnogalacturonan I (RG-I) that are composed of unbranched α-1,4-D-linked galacturonic acid (GalA) residues and a backbone of repeating α-1,2-L-rhamnose(Rha*p*)-α-1,4-D-GalA disaccharide units, respectively.[4] Side chains consisting mainly of arabinan and/or galactan are attached to the RG-I backbone at the C-4 position of the Rha*p* residues. These RD-I parts are called 'hairy regions'.[1,5] The GalA residues in the HGA and RG-I backbones may be methyl esterified and/or O-acetylated.[1]

The degree and pattern of methyl esterification have important commercial implications and have been the subject of many studies because of the effect on the rheological and gel-forming properties of pectins.[6–9] Daas et al.[10] introduced the term 'degree of blockiness', expressed as the total amount of free GalA mono-, di-, and tri-GalA residues released by *endo*-polygalacturonase enzyme relative to the total amount of free GalA residues present in the pectin. The higher the 'degree of blockiness' of pectins with similar DE, the more blockwise the distribution of the methyl esters in the pectin. Previous investigations of the methyl ester distribution of pectins include enzymatic degradation combined with

Abbreviations: GalA, galacturonic acid; HGA, homogalacturonan; DE, degree of esterification; DP, degree of polymerization; RG-I, rhamnogalacturonan-I; Rha*p*, ramnopyranosyl unit; PME, pectin methyl esterase; B, block de-esterification; R, random de-esterification; IR, infrared; NIR, near infrared; FT, Fourier transformed; PC, principal component; LV, latent variables; PCA, principal component analysis; CV, cross-validation; RMSECV, root mean square of cross-validation; PLS, partial least square; iPLS, interval partial least square; ECVA, extended canonical variates analysis; EISC, extended inverted signal correction; r, correlation coefficient; %p, percent points.

* Corresponding author.
E-mail address: haw@life.ku.dk (H. Winning).

0008-6215/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2008.10.015

mass spectrometry or high performance chromatography.[9,11–18] These studies all concluded that alkaline de-esterification and fungus-pectin methyl esterase (PME) resulted in random distribution, and plant-PME resulted in blockwise distribution of methyl ester groups. Elaborate studies made by Limberg et al.[19] suggested that de-esterification using plant-PME is a process which selectively introduces block structure of adjacent free GalA units, whereas fungus-PME and alkali treatment lead to two different forms of homogeneous methyl esterification patterns. This finding led to the hypothesis that random de-esterification is more ordered than first assumed.[15] The terms 'random' and 'blocky' methyl ester distributions are merely interpretations, and there is no precise definition of how many contiguous free GalA units constitute a block. In addition, random distribution could be completely random or almost systematic (Fig. 1).

Because the distribution of the free ester groups is critical for the detailed functionality of pectins, development of a fast and reliable method able to assess the ester distribution of native or semi-refined pectins is desirable. Solution-state nuclear magnetic resonance (NMR) spectroscopy has proven to be a highly effective tool in compositional and structural analysis of pectins.[20–22] Previously, we have shown[20] that ^{1}H NMR is an effective method for the classification of pectins by enzyme treatment in a commercial set of designed pectins, as NMR is highly sensitive to conformational and spatial changes in the molecule. Moreover, we were able to show that NMR in combination with chemometrics was able to predict the degree of randomness. However, in order to be attractive to the manufacturing companies, the method has to be rapid and simple, preferably be able to measure close to the production facility (at-line) or directly in the production line (on-line or in-line). The highly complex nature of solution-state ^{1}H NMR makes it unsuitable for at-, on-, or in-line monitoring of pectin, as the equipment is very sensitive to vibrations and the magnet is disturbed by magnetic materials such as iron in the production facility. Thus, development of a faster and simpler method for characterization of the pectin composition is desired. Vibrational spectroscopy is widely used for compositional analysis of polymers that are used in the food industry, and in combination with multivariate data analysis (chemometrics) it is a powerful analytical screening method that can reveal detailed information concerning the composition and properties of material at a molecular level.[23–26] Vibrational spectroscopies, such as infrared (IR), Raman, and near infrared (NIR), are generally rapid, non-destructive, simple to operate, and require no sample preparation. Thus, these three techniques can be adapted to at-line analysis. Furthermore, Raman and NIR spectroscopies can also be implemented directly in the production line for in-line analysis using optical quartz fibers.

Fourier transform (FT)-IR, FT-Raman, and NIR spectroscopies have proven efficient for the determination of the degree of esterification as well as of degree of amidation of pectin powders[27–29] using efficient multivariate regression methods such as partial least squares (PLS) regression.[30] Numerous publications have been made on the IR analysis of pectin.[26,28–35] IR and Raman spectroscopies measure the fundamental molecule vibrations and thus reveal many of the same properties. NIR measures over- and combination tones predominantly of the fundamental vibrations that involve hydrogen and are often more difficult to interpret. Most of the NIR signals from pectins stem from OH stretching vibrations. While the ester group is known to give rise to specific group frequencies in vibrational spectroscopy, signals relating to their specific distribution along the pectic polymer are bound to be more subtle. For this reason, it is necessary to apply multivariate data analysis in order to extract information about the specific distribution. The purpose of this study was to develop spectroscopic calibration models for IR, Raman, and NIR, and compare their ability to predict the amount of methyl esters removed with a random or block esterase. For this investigation, 31 pectin samples were produced by degradation of the same pectin by two different PMEs: a block esterase and a random esterase.

2. Materials and methods

2.1. The experimental design

High-methoxy pectin samples were designed by CP Kelco (Lille Skensved, Denmark) from a native (mother) citrus pectin (DE ~ 72.3%) which had been re-methylated to DE ~ 93.8%. The re-methylated pectin was produced from a 5 °C suspension of pectin in dry methanol, and thionyl chloride was added over a period of five days. The product was separated by filtration, washed with methanol, and dried. The re-methylated pectin was then de-esterified with two different pectin modifying enzymes: a random methyl esterase, Rheozym (now Novoshape®) (Novozymes, Denmark), and a block methyl esterase derived from the papaya fruit. The commercial Rheozyme is cloned from *Aspergillus aculeatus* and is believed to have the same de-esterification pattern as *Aspergillus niger*, which randomly removes the methyl groups using a multi-chain mechanism.[36] The de-esterification

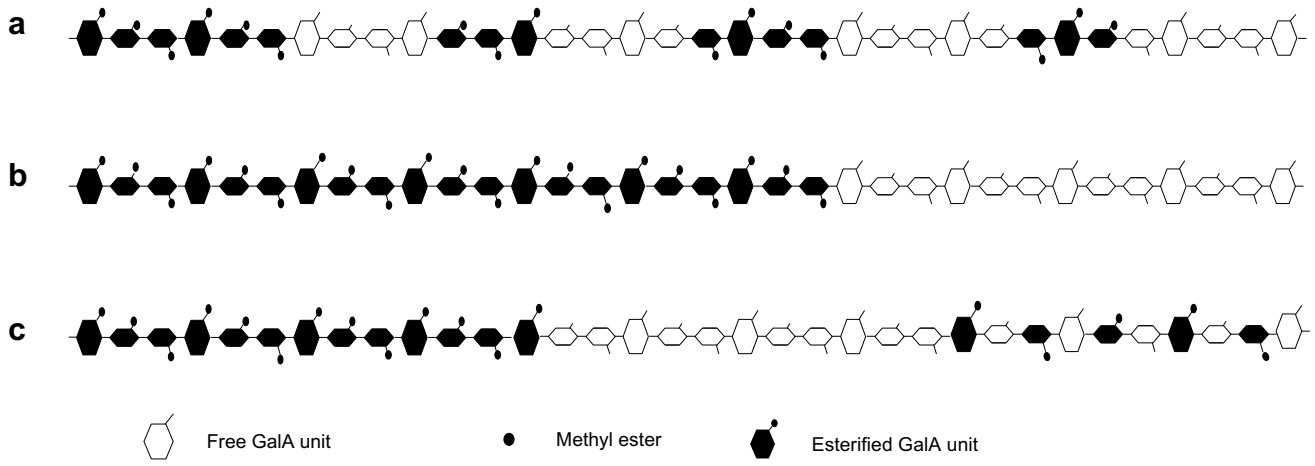

Figure 1. Schematic example of pGalA chains. HGA with methyl groups of (a) short blockwise distribution, (b) large blockwise distribution, and (c) block and systematic random distribution of free GalA units (modified from Ref. 20).

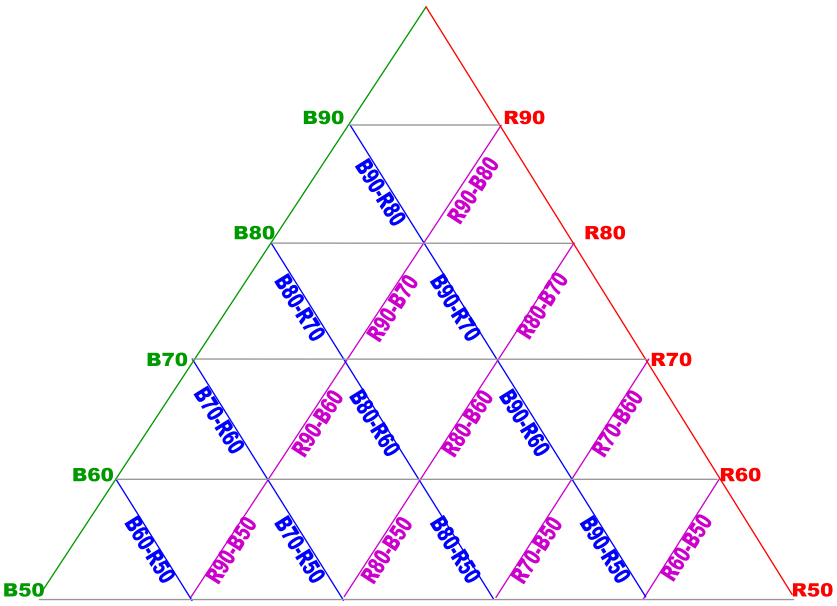

Figure 2. Schematic model of the triaxial experimental design. Reading downwards (left): De-esterification with block enzyme from DE ~ 100 to DE 90, 80, 70, 60, and 50. Reading downwards (right): De-esterification with random enzyme from DE ~ 100 to DE 90, 80, 70, 60, and 50. The horizontal lines indicate the total de-esterification.

pattern of the papaya block esterase is suggested as a single chain mechanism similar to the mechanism described for other esterases from, for example, orange.[37,38] Random de-esterification was preformed at pH 4.8 and the blockwise de-esterification at pH 5.5. The pectin samples were designed using a three-dimensional experimental design, illustrated in the tri-axial diagram (Fig. 2). One pectin series with blockwise methyl ester distribution and one pectin series with random methyl ester distribution were produced. This yielded two groups of five samples with varying DE from about 90% in 10% steps down to 50% DE (Block: B90, B80, B70, B60, and B50 and Random: R90, R80, R70, R60, and R50). A third series was made from the block series, which were subsequently de-esterified with the random enzyme. The fourth series was made from the random series and was subsequently de-esterified with the block enzyme. Thus, the total sample set consisted of four groups with different enzyme treatments: one pure random (R), one pure block (B), one random followed by block (R-B), and one block followed by random (B-R) (Fig. 2). One batch of each sample was produced in amounts of approximately 700 mg. The reference DE was determined by titration, and the de-esterification with block and random enzymes was thereby calculated. The designed pectins had a molecular weight of about 10–15 kDa, determined by a triple detector size exclusion chromatography module (Viscotek, Malvern Instruments, UK), which corresponds to a degree of polymerization (DP) of approximately 50 GalA units. A remarkable drop in molecular weight (from 180 to 13 kDa) was observed as a result of the re-methylation. This suggests a sever degradation of the pectin molecule during the re-methylation process.

The measured DEs are shown in Table 1, together with the amount of ester groups removed with block and random enzymes, respectively. As the mechanisms of the two enzymes are not fully elucidated, the amount of ester groups removed with random or block esterase was used as reference values in the calibration models. The DE is calculated in percent of the number of esterified GalA units compared to the total number of GalA units (93.8% when fully esterified) and is denoted as % point (%p), which is also the unit for block and random de-esterification.

Table 1
Reference de-esterification determinations with random and block enzymes of design pectins

Target de-esterification code	1. De-esterification random enzyme	2. De-esterification block enzyme	R-enzyme treated %p	B-enzyme treated %p	Target de-esterification code	1. De-esterification block enzyme	2. De-esterification random enzyme	B-enzyme treated %p	R-enzyme treated %p
R90	93.8→88.9		4.9	0	B90	93.8→89.4		4.4	0
R90-B80		88.9→78.3		10.6	B90-R80		89.4→81.1		8.3
R90-B70		88.9→70.9		18	B90-R70		89.4→73.8		15.6
R90-B60		88.9→64.4		24.5	B90-R60		89.4→66.6		22.8
R90-B50		nd[a]	–	–	B90-R50		89.4→57.5		31.9
R80	93.8→80.0		13.8	0	B80	93.8→79.7		13.9	0
R80-B70		80.0→72.2		7.8	B80-R70		79.7→69.5		10.2
R80-B60		80.0→66.2		13.8	B80-R60		79.7→59.4		20.3
R80-B50		80.0→56.2		23.8	B80-R50		79.7→49.1		30.6
R70	93.8→71.9		21.9	0	B70	93.8→71.2		22.6	0
R70-B60		71.9→62.2		9.7	B70-R60		71.2→60.2		11
R70-B50		71.9→50.3		21.6	B70-R50		71.2→47.3		23.9
R60	93.8→64.9		28.9	0	B60	93.8→60.3		33.5	0
R60-B50		64.9→51.7		13.2	B60-R50		60.3→50.8		9.5
R50	93.8→58.8		35	0	B50	93.8→57.0		36.8	0
R50-B50		58.8→52		6.8	B50-R50		57.0→50.6		6.4

[a] nd: not determined.

2.2. FT-IR spectroscopy

IR spectra of pectin powders and pectin films (deposited from 1% w/w solution) were recorded using an Arid-Zone MB 100 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 powders were measured as single measurements due to the small sample amounts using 256 scans and a resolution of 4 cm^{-1}. Selected duplicate measurements showed a high degree of reproducibility. The pressure applied to squeeze the powder toward the diamond was 6 N/cm^2. The pectin films were prepared in duplicate from 4 μL 1% (w/w) pectin-water solution and were dried directly at the crystal for 15 min. The films were measured as duplicates using 256 scans and a resolution of 2 cm^{-1}, and the average spectra were used in the calculations. The spectra were ratioed against background scans collected as single-beam spectrum of the clean ATR crystal to obtain the spectra in absorbance units. All spectra were collected in the range of 4000–550 cm^{-1}.

2.3. FT-Raman spectroscopy

Raman spectra were collected on a Perkin Elmer System NIR FT-Raman interferometer (Perkin Elmer Instruments, Waltham, Massachusetts, USA) equipped with a Nd:YAG laser emitting at 1064 nm with a laser power of 200 mW. Data were collected using an InGaAs detector and were stored as Raman shifts in the range 3600–200 cm^{-1}. A 180° back-scattering arrangement was used and no correction for the spectral response was applied. A total of 256 scans were averaged for each sample and the resolution was 4 cm^{-1}. All measurements were carried out as duplicate and the average spectra were used in the calculations.

2.4. FT-NIR spectroscopy

NIR spectra were collected using a FT-NIR Bomem NetworkIR spectrometer in reflectance mode. The instrument was fitted with 2.1 μm InGaAs detectors using a custom-made fiber optical probe. Spectra were acquired in the range 13,500–4500 cm^{-1} using a custom-made fiber optical reflectance probe with a sapphire window. The number of scans was 32 and the interval between spectral data points was 16 cm^{-1}/s (0.77 cm^{-1}). An open sample cup from which the optic fiber probe emerged was used for the measurements. The NIR reflectance spectra were converted to log(1/R) units prior to data analysis by referencing to an external ceramic reference. All measurements were carried out as duplicate and the average spectra were used in the calculations.

2.5. Chemometric analysis and software

Multivariate data analysis in the form of principal component analysis (PCA)[39] and PLS[30] was applied to obtain optimal quantitative and qualitative information from the measured spectra. PCA is the primary tool for investigation of large bilinear data structures for the study of trends, groupings, and outliers. By means of PCA it is possible to find the main variation in a multidimensional data set by creating new linear combinations, principal components (PCs), from the underlying latent structures in the raw data. PLS is a multivariate calibration method by which two sets of data, X (e.g., spectra) and y (e.g., DE), are related by means of regression. The purpose of PLS is to establish a linear model of latent variables (LVs), which enables the prediction of a reference value y (slow measurement) from the measured spectrum X (fast measurement). PLS regression models were developed for DE, B, and R. Subsequent PLS models were made with interval PLS (iPLS).[40] iPLS is an extension of PLS, which develops local PLS models on a number of subintervals of the full spectrum region. Its main advantage is to provide an overall picture of the relevant information in different spectral subdivisions, thereby removing interferences from other regions.[40]

It is especially important to limit the number of LVs when the sample set is small, as in the present study, in order to avoid over-fitting. Validation of the calibration models is therefore crucial. All the calibration models were validated using cross-validation (CV)[41] with four segments, leaving one segment out at a time from which the root mean square error of cross-validation (RMSECV) is calculated. The relative predicting error ($RMSECV_{\%rel}$) is calculated as a percentage of the range. The PLS results are presented as number of LVs, correlation coefficients (r), and RMSECV.

As preliminary investigations led to the conclusion that the optimal pre-processing method was extended inverted signal correction (EISC)[42] for the three spectroscopic methods, it was decided to use EISC prior to all calculations. This is also the conclusion in a study using vibrational spectroscopy for the analysis of alginates also belonging to the hydrocolloid family.[23] In order to examine the dependence on indirect correlations to DE, the effect of orthogonalization of the spectral matrix with respect to DE was tested. Thus, the vector describing the variation due to DE was extracted from the spectra prior to model development. PCA and extended canonical variates (ECVA)[43] models were applied for classification of enzyme treatment. Canonical variates analysis (CVA)[44,45] is a method for estimation of directions in space that maximizes the differences between the groups. However, CVA cannot deal with highly collinear data such as spectroscopic data, where the number of variables is larger than the number of samples. The ECVA method is based on the standard CVA, and by a transformation of an eigenvector problem to a regression problem, it is possible to use PLS to solve the inner part of CVA and thereby allow for the analysis of collinear data.[43]

The multivariate data analysis was performed using LatentiX version 1.0 (Latent5, Copenhagen, Denmark, http://www.latentix.com), whereas EISC pre-processing, iPLS regression, and ECVA were performed in MatLab 7.6 (Mathworks, Natick, Massachusetts, USA) using the EMSC/EISC Toolbox, iToolbox, and the ECVA Toolbox version 2.02, respectively (all available at www.models.life.ku.dk).

3. Results and discussion

3.1. Explorative analysis

The EISC pre-processed IR, Raman, and NIR spectra of the 31 pectin samples are shown in Figure 4a–c together with the corresponding PCA score plots (d–f). The PCA score plots reveal that the spectra contain information which can differentiate between the four different enzyme treatments. This enzyme treatment differentiation is observed in the first two PCs for the PCA model obtained on the IR spectra (d), in PC1 versus PC3 for Raman spectra (e), and to a lesser extent in PC2 versus PC3 for the PCA model obtained on the NIR spectra (f).

In the PCA score plot of the IR spectra (d), PC1 divides the samples by their enzyme treatment, whereas PC2 explains the variation in the spectra due to DE. The block samples (B) are grouped together with the samples de-esterified with random enzyme followed by block enzyme (R-B). In the score plot from the PCA of the Raman spectra (e), the PC1 direction again divides the pectins according to the enzyme treatment, grouping the R and the B-R into two very distinct groups and again with the B and the R-B grouped together. The PCA score plot obtained on NIR spectra (f) shows not as distinct groups as PCA models obtained on IR and Ra-

man data. The DE is explained along PC2, but the complexity of the methyl ester distributions is less reflected in the PC3 direction. Similarly to the PCA of the IR and Raman spectra, the score plot obtained on the NIR spectra shows that pure block (B) and B-R are more alike and pure random (R) and R-B are more alike. The PCA score plots of the IR and Raman spectra show that the samples de-esterified with two enzymes are similar to the samples which have been de-esterified with the second enzyme, that is, the B-R samples lie near the R samples and the R-B samples are located near the B samples (Fig. 4). This indicates that the nature of the samples primarily reflects the last enzyme treatment. This is especially the case for the R-B samples. Thus, it is likely that the methyl ester pattern made by the random enzyme simply is stripped off by the block enzyme, as we also concluded in a previous study.[20] As a result, the R-B pectins lose the initially induced 'randomness'.

3.2. Calibration models

Calibration models for prediction of the DE, B, and R from the IR, Raman, and NIR spectra of the pectin powders and film were developed using PLS regression. The modeling of the IR and Raman spectra was performed using the spectral ranges 1800–650 $cm^{-1}$and 1800–250 cm^{-1}, respectively. For the NIR calibrations, the spectral range from 9000 to 4500 cm^{-1} (1111–2222 nm) was used. Furthermore, a new variable was introduced, (R/B+R), to obtain a reference value for the ester distribution as the ratio between one enzyme treatment and the total de-esterification by both enzymes.

The calibration results of the three methods are shown in Table 2. The calibration models for prediction of DE by all three spectroscopic methods yielded very good models with RMSECV between 1.1%p and 4.1%p DE (range ~ 47–94%p) and correlation coefficients of 0.99 or above. The best PLS regression results were obtained using Raman spectra with a model using 6 LVs, resulting in a correlation coefficient of 1.0 and a prediction error (RMSECV) of 1.1%p (Table 2). Calibration models predicting the distribution of free GalA units, expressed as B and R, were developed for each spectroscopic method, yielding RMSECV values between 4.4%p and 8.5%p R (or B) (range ~ 0–35%p) and correlation coefficients between 0.68 and 0.92. In the prediction of B and R, PLS models using the spectra of the pectin films were slightly better than the corresponding PLS models made on powder. Generally, random de-esterification (R) was better predicted. This was also found in our previous study, where solution-state ^1H NMR spectra of the pectins were used for the predictions.[20] Calibration models obtained using IR and Raman spectra required at least six components to yield acceptable prediction errors of 6–7%p R. The NIR calibration models turned out to be the worst of the three methods with respect to prediction of R, with a large prediction error of 8.5%p and a poor correlation coefficient of 0.68. This apparent discrepancy underlines the difficulty in obtaining a useful blockiness descriptor that both relates to functionality and can be predicted by spectroscopic fingerprints. As an example, the variable R/(B+R) is generally better predicted by the spectroscopic methods than B or R (Table 2).

3.3. Orthogonalization by DE

To test if the predictions of R are due to an indirect correlation with DE, the spectroscopic data ensembles were orthogonalized according to DE. Using the DE-orthogonalized spectra in the PLS regression resulted in extremely poor DE predictions (Table 2). The models predicting R and B were also seriously deteriorated by the DE-orthogonalization, while the predictions of the ratio reference value (R/B+R) variable were significantly improved. This proves that the information related to randomness and blockiness is not significantly confounded with the DE information and that the variations in the spectra are simplified by the removal of the DE vector with regard to R/(B+R). This implies that there is variance, especially in the IR and Raman spectra, which correlates with blockiness and randomness independent of the DE.

Orthogonalization of the spectra also affected the grouping by enzyme treatment. Figure 4a and b shows PCA score plots of the orthogonalized IR and Raman spectra, which revealed the same or a slightly improved grouping of the samples compared to the PCA score plot of the non-orthogonalized spectra (Fig. 3d and e. However, the DE-orthogonalization of the NIR spectra had worsened the grouping of the samples compared to the non-orthogonalized NIR spectra (Fig. 3f).

3.4. Variable selection using iPLS

iPLS was applied to IR, Raman, and NIR spectra in order to possibly improve the prediction of B and R and to encircle the areas of the spectra which hold information about the methyl ester pattern. No intervals were found to significantly improve the prediction performance of B of any of the methods. However, the prediction of R was significantly improved when using the 970–870 cm^{-1} interval in the IR spectra. This interval yields a RMSECV value of 3.5%p R and a correlation coefficient of 0.95. Moreover, iPLS showed that the 7400–7200 cm^{-1} interval in the NIR spectra was significant with respect to the random ester distribution (R), yield-

Table 2
Results of cross-validated prediction models of EISC-treated IR (1800–650 cm^{-1}), Raman (1800-230 cm^{-1}), and NIR (9000–4500 cm^{-1}) spectra

Method		DE	R	B	R/(B+R)	Ortho		DE	R	B	R/(B+R)
IR powder	LV	4	5	5	8	IR powder	LV	1	1	1	8
	$RMSECV_{abs}$	1.4	6.5	8.0	0.19		$RMSECV_{abs}$	13.8	11.3	11.6	0.15
	$RMSECV_{\%rel}$	3.0	18.8	22.9	19		$RMSECV_{\%rel}$	29.3	32.3	33.1	15
	r	0.99	0.81	0.71	0.84		*r*	−0.50	0.10	0.13	0.91
IR film	LV	7	6	6	6	IR film	LV	1	4	4	6
	$RMSECV_{abs}$	1.6	6.1	6.7	0.21		$RMSECV_{abs}$	13.0	8.0	8.4	0.17
	$RMSECV_{\%rel}$	3.4	17.4	19.1	21		$RMSECV_{\%rel}$	27.6	22.9	24.0	17
	r	0.99	0.84	0.81	0.79		*r*	−0.48	0.70	0.68	0.86
Raman powder	LV	6	6	6	8	Raman powder	LV	1	5	3	7
	$RMSECV_{abs}$	1.1	5.9	6.6	0.13		$RMSECV_{abs}$	13.5	9.9	9.6	0.09
	$RMSECV_{\%rel}$	2.3	16.9	18.8	13		$RMSECV_{\%rel}$	28.7	28.3	27.4	9
	r	1.0	0.84	0.81	0.92		*r*	−0.54	0.54	0.52	0.95
NIR powder	LV	4	4	4	4	NIR powder	LV	1	3	2	3
	$RMSECV_{abs}$	1.3	8.3	8.5	0.28		$RMSECV_{abs}$	12.6	10.0	10.9	0.25
	$RMSECV_{\%rel}$	2.8	23.7	24.2	28		$RMSECV_{\%rel}$	26.8	28.5	31.1	25
	r	0.99	0.68	0.68	0.61		*r*	−0.31	0.45	0.35	0.70

The predicting error is displayed as the absolute error, RMSECV (abs), and the relative error RMSECV (%rel).

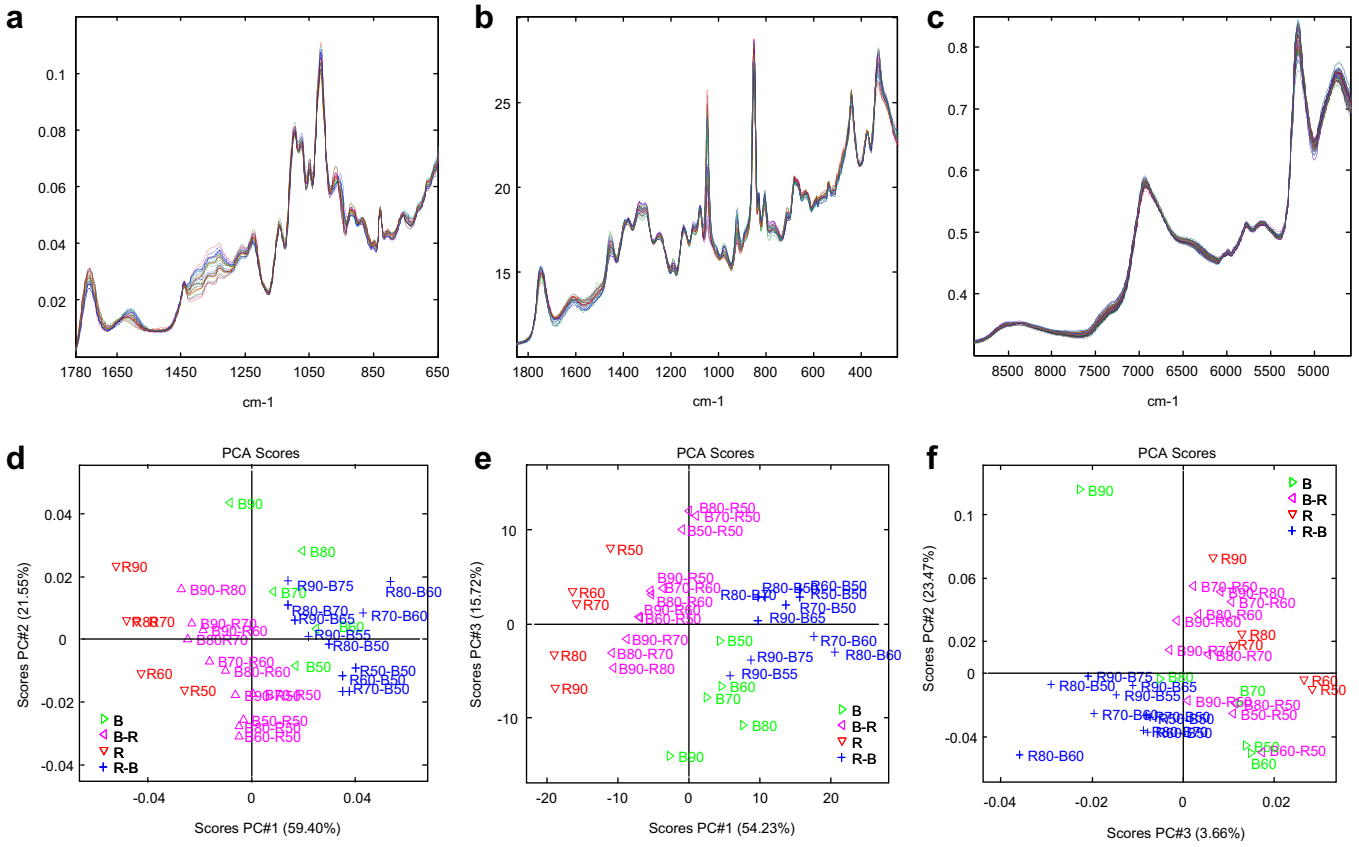

Figure 3. EISC pre-processed IR (a), Raman (b), and NIR spectra (c) and their corresponding PCA score plots (d–f), colored according to the enzyme treatment, showing the ability to differentiate between the four different enzyme treatments.

ing RMSECV of 6.2%p R and correlation coefficient of 0.84. Absorbance in this region of NIR is exclusively due to first overtones of combination bands (C–H stretch and C–H deformations) originating from CH_3 groups, which makes very good sense. In the case of the Raman spectra, only the region which includes the strong α-anomer-sensitive pyranosidic band at 850–780 cm^{-1} comes close in performance to the global prediction of R.

The most influential spectral area of the IR spectrum (970–870 cm^{-1}) found by iPLS is dominated by (C–O–H) and (CH_3) in-plane bending.[28,33] This signal has good potential to be affected by the randomness, since these vibrations are correlated to DE.

The predicted plot versus measured plot in Figure 5d–f show how well the best iPLS models perform. These local iPLS models contain less interference, and are thus more simple and easier to interpret. The distributions of the samples along the regression line as well as the sample distance to the regression line are very important and the three models display good linearity. An example is the sample in the NIR predicted plot versus measured plot in the lowest corner to the left, which has the value zero (measured) but is predicted to minus twenty. This is presumably because the R-values are constructed target values rather than real quantitative values.

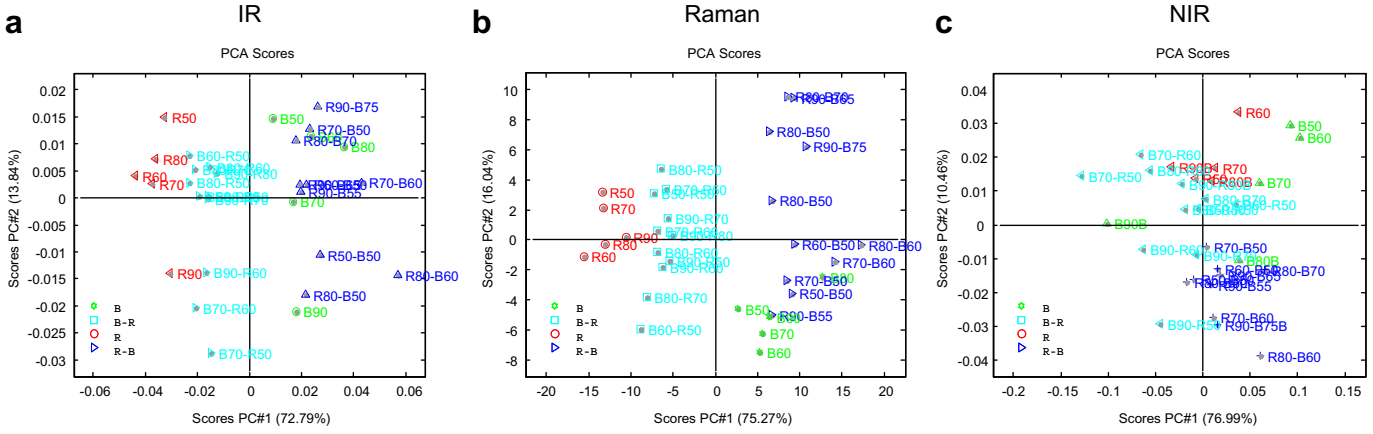

Figure 4. PCA score plots (PC1 vs PC2) of (a) IR, (b) Raman, and (c) NIR spectra. The scores are colored according to enzyme treatment.

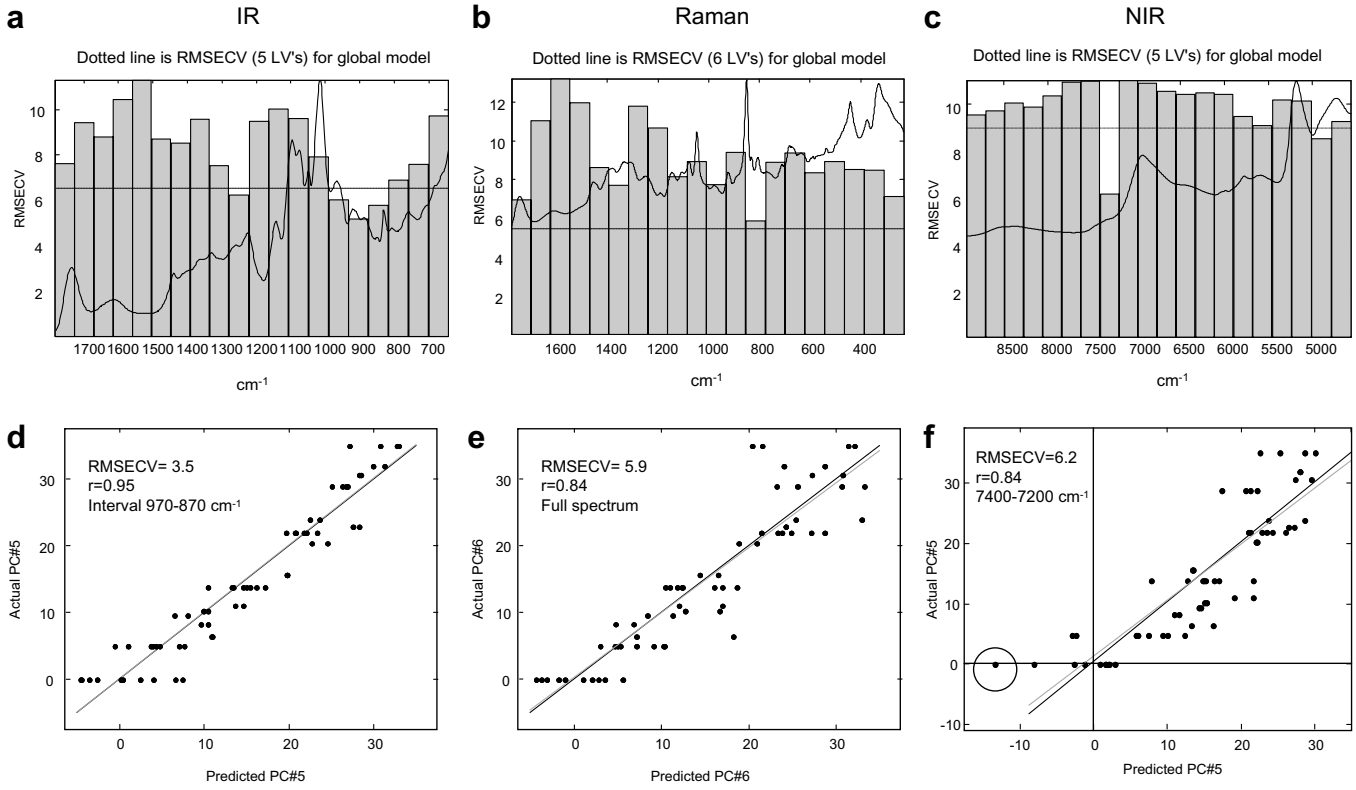

Figure 5. IR (left), Raman (middle), and NIR (right) iPLS models (top) obtained on EISC pre-processed spectra for prediction of R. Predicted plot versus measured plot of calibrations models obtained on the single best interval selected by the iPLS models.

3.5. IR assignment

In IR, it is possible to measure dried pectin solution (deposited from 1% w/w solution) on the ATR crystal which develops into a pectin film, which is why both pure powder and the pectin film were measured. The differences between the two techniques are illustrated in Figure 6, where the average pectin film and the average pectin powder are plotted. The interval found by iPLS (970–870 cm^{-1}) is dominated by (C–O–H) and (CH_3) in-plane bending.[28,33] The assignment of the IR spectrum is illustrated in Figure 6.

3.6. Raman

As there were no intervals in the iPLS of the Raman spectra which were able to improve the predictions of R, extended canonical variates analysis (ECVA)[43] was used to investigate wether the Raman spectra contained other distinct information related to the enzyme treatments.

Figure 7a shows the canonical variates and reveals a promising discrimination between the samples with respect to the enzyme treatment. According to this model, only one sample, B90-R50, is misclassified. Apparently, this sample resembles the blockiness

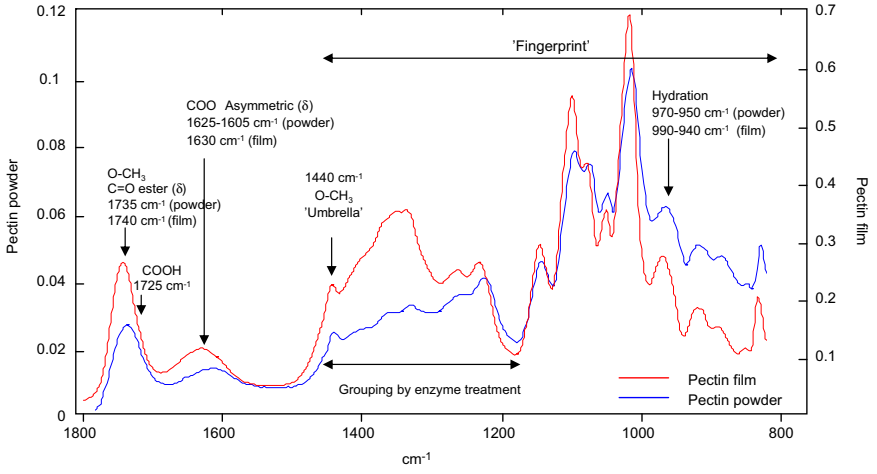

Figure 6. Mean IR spectra of all the pectin powders and pectin film, with arrows indicating the group frequency assignments.

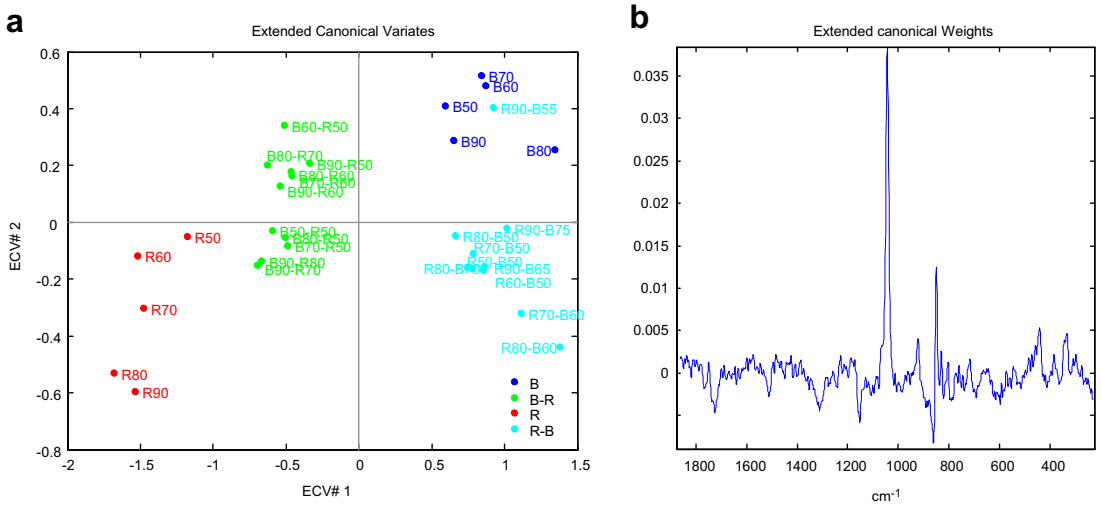

Figure 7. ECV#1 versus ECV#2 for the Raman data discriminating between the four different enzyme treatments (a). Only one sample, R90-B55, is misclassified. Extended canonical weight for the first direction showing an important signal at 1045 cm^{-1} (b).

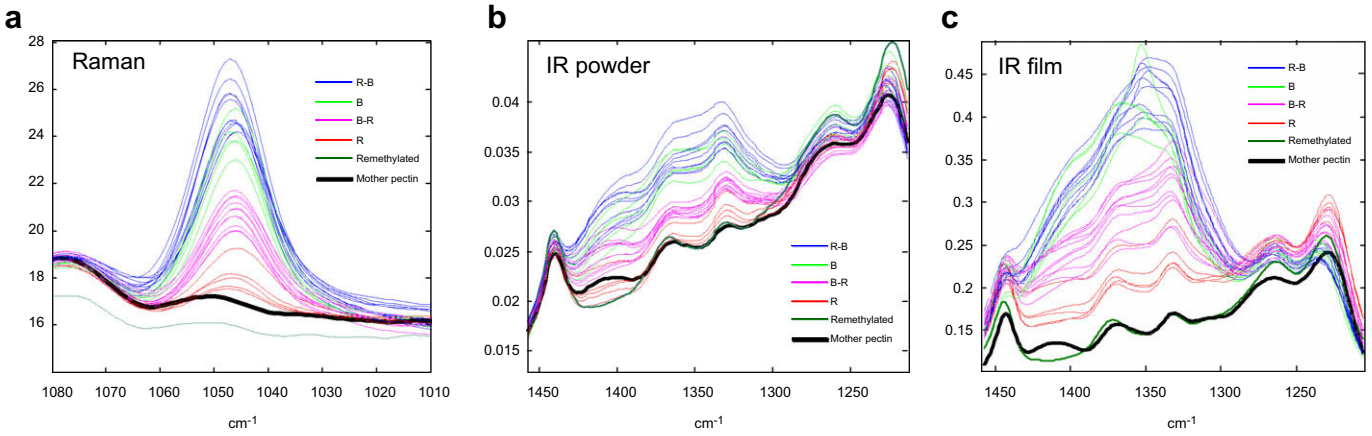

Figure 8. Raman spectra of pectin powder containing the signal at 1045 cm^{-1} (a), and IR spectra of powder (b) and film (c) of the region from 1400 to 1300 cm^{-1} colored according to the four different enzyme treatments. A re-methylated pectin and the native mother pectin are superimposed together with the 31 enzyme treated samples.

group more than the random block de-esterification. The fact that sample B90-R50 has almost been de-esterified 25%p by block enzyme and only 5%p by random enzyme (Table 1) explains the misclassification of this sample as having a more blockwise distribution than random ester distribution.

The right plot in Figure 7b shows the canonical weight vector for the EVCA model of the Raman spectra based on five PLS components in the inner relation for the ECVA method. This weight shows high value in the spectral range 1050–1040 cm^{-1}, indicating that this peak is important for the separation of the enzyme treatment groups. A similar ECVA model calculated on the basis of the IR spectra (powder and film) resulted in the same misclassification of sample B90-R50. However, the canonical weight vectors for the IR ensemble were not as obvious to interpret as the first weight vector in the Raman spectra, as several areas show high weight values. One of these areas in IR from 1420 to 1250 cm^{-1} is illustrated in Figure 8b and c together with the specific area (1050–1040 cm^{-1}) of the Raman spectra (Fig. 8a).

The spectra in Figure 8 are colored according to the different enzyme treatments. In both Raman and IR spectra, the signal intensity decreases in the sequence R-B > B > B-R > R > re-methylated pectin (the starting point for all the pectins before enzyme treatment). The most intense signals stem from the random-block enzyme treatment, which implies that contiguous smaller blocks of galacturonic acid units give rise to this specific signal. We assume that the signal at 1045 cm^{-1} in the Raman spectrum is due to changes in the galacturonic acid backbone structure when the substitution of the galacturonic acid chain with ester groups is changed.

The IR film spectra confirm the intensity relations: R-B > B > B-R > R > re-methylated pectin, and establish that the B samples resemble the R-B samples. The latter indicate that block enzymes remove the 'randomness', as the samples with the highest intensity are B60 and R50-B50 and the samples with the lowest intensity are sample R90 in the IR film spectrum. However, the interpretations of signals in the IR spectrum are more difficult to assign.

4. Conclusion

This investigation has shown that it is possible to develop a rapid, non-destructive, and robust quantitative method for measuring the degree of esterification and characterize the distribution of methyl ester and carboxyl groups of enzyme-treated pectin using vibrational spectroscopies (IR, Raman, and NIR) and chemometrics. The degree of esterification, DE, can be predicted very precisely from calibration models obtained on spectra of all three

spectral sources. When it comes to information about the distribution of the ester groups, the random ester distribution (expressed as the amount of esters removed with random enzyme) was generally better predicted than blockwise ester distribution. The best full spectrum predicting model with respect to R was obtained on the IR pectin film. IR on pectin film compared to that on pectin powders generally yielded better prediction models with lower prediction errors.

The experimental design, which consisted of pectin with block de-esterification, random de-esterification, block-random de-esterification, and random-block de-esterification, could be reflected in PCA score plots of all three methods. Using the variable selection method iPLS on the spectra of IR and NIR, regions were found where the prediction models R could be significantly improved. By employing the classification method ECVA, one specific signal was found in the Raman (1045 cm^{-1}) spectrum and one significant area (1250–1400 cm^{-1}) in the IR spectrum which are able to classify the pectin samples according to the four different enzyme treatments. In both Raman and IR spectra, the signal intensity decreases in the sequence R-B > B > B-R > R > re-methylated pectin.

Acknowledgments

We thank the Danish National Advanced Technology Foundation for financial support for the project 'Valorization of potato byproducts: New food ingredients by an enzyme-mediated extraction process'. Lisbeth T. Hansen and Gilda Kischinovsky from Faculty of Life Sciences, University of Copenhagen, are greatly acknowledged for help with the experimental work and for proofreading the manuscript, respectively. CP Kelco is acknowledged for supplying the pectin samples.

References

1. Carpita, N. C.; Gibeaut, D. M. *Plant J.* **1993**, *3*, 1–30.
2. Rolin, C. Pectin. In *Industrial Gums: Polysaccharides and their Derivatives*; Whistler, R. L., BeMiller, J. N., Eds., 3rd ed.; Academic Press: San Diego, 1993; pp 257–293.
3. Ralet, M. C.; Bonnin, E.; Thibault, J. F. Pectins. In *Polysaccharides II*; Vandamme, E. J., De Baets, S., Steinbüchel, A., Eds.; WILEY-VCH Verlag GMbH: Weinheim, 2002.
4. Pilnik, W.; Voragen, A. *Food Enzymology*; Elsevier: London, 1991.
5. Ridley, B. L.; O'Neill, M. A.; Mohnen, D. A. *Phytochemistry* **2001**, *57*, 929–967.
6. Rolin, C. Pectin. In *Industrial Gums: Polysaccharides and their Derivatives*; Whistler, R. L., BeMiller, J. N., Eds.; 3rd ed.; Academic Press: San Diego, 1993; pp 257–293.
7. Ralet, M. C.; Dronnet, V.; Buchholt, H. C.; Thibault, J. F. *Carbohydr. Res.* **2001**, *336*, 117–125.
8. Ralet, M. C.; Crepeau, M. J.; Buchholt, H. C.; Thibault, J. F. *Biochem. Eng. J.* **2003**, *16*, 191–201.
9. Willats, W. G. T.; McCartney, L.; Mackie, W.; Knox, J. P. *Plant Mol. Biol.* **2001**, *47*, 9–27.
10. Daas, P. J. H.; Meyer-Hansen, K.; Schols, H. A.; De Ruiter, G. A.; Voragen, A. G. J. *Carbohydr. Res.* **1999**, *318*, 135–145.
11. Daas, P. J. H.; Voragen, A. G. J.; Schols, H. A. *Biopolymers* **2001**, *58*, 195–203.
12. Daas, P. J. H.; Voragen, A. G. J.; Schols, H. A. *Carbohydr. Res.* **2000**, *326*, 120–129.
13. Daas, P. J. H.; Meyer-Hansen, K.; Schols, H. A.; De Ruiter, G. A.; Voragen, A. G. J. *Carbohydr. Res.* **1999**, *318*, 135–145.
14. Körner, R.; Limberg, G.; Christensen, T. M. I. E.; Mikkelsen, J. D.; Roepstorff, P. *Anal. Chem.* **1999**, *71*, 1421–1427.
15. Limberg, G.; Korner, R.; Buchholt, H. C.; Christensen, T. M. I. E.; Roepstorff, P.; Mikkelsen, J. D. *Carbohydr. Res.* **2000**, *327*, 293–307.
16. Limberg, G.; Korner, R.; Buchholt, H. C.; Christensen, T. M. I. E.; Roepstorff, P.; Mikkelsen, J. D. *Carbohydr. Res.* **2000**, *327*, 321–332.
17. Ralet, M. C.; Thibault, J. F. *Biomacromolecules* **2002**, *3*, 917–925.
18. Ralet, M. C.; Dronnet, V.; Buchholt, H. C.; Thibault, J. F. *Carbohydr. Res.* **2001**, *336*, 117–125.
19. Limberg, G.; Körner, R.; Buchholt, H. C.; Christensen, T. M. I. E.; Roepstorff, P.; Mikkelsen, J. D. *Carbohydr. Res.* **2000**, *327*, 293–307.
20. Winning, H.; Viereck, N.; Nørgaard, L.; Larsen, J.; Engelsen, S. B. *Food Hydrocolloids* **2007**, *21*, 256–266.
21. Andersen, A. K.; Larsen, B.; Grasdalen, H. *Carbohydr. Res.* **1995**, *273*, 93–98.
22. Grasdalen, H.; Andersen, A. K.; Larsen, B. *Carbohydr. Res.* **1996**, *289*, 105–114.
23. Salomonsen, T.; Jensen, H. M.; Stenbaek, D.; Engelsen, S. B. *Carbohydr. Polymers* **2008**, *72*, 730–739.
24. Pedersen, D. K.; Engelsen, S. B. *New Food* **2001**, *4*, 9–13.
25. Dyrby, M.; Petersen, R. V.; Larsen, J.; Rudolf, B.; Nørgaard, L.; Engelsen, S. B. *Carbohydr. Polym.* **2004**, *57*, 337–348.
26. Synytsya, A.; Čopîková, J.; Matějka, P.; Machovič, V. *Carbohydr. Polym.* **2003**, *54*, 97–106.
27. Zachariassen, C. B.; Larsen, J.; van den Berg, F.; Engelsen, S. B. *Chemometrics Intell. Lab. Syst.* **2005**, *76*, 149–161.
28. Engelsen, S. B.; Nørgaard, L. *Carbohydr. Polym.* **1996**, *30*, 9–24.
29. Engelsen, S. B.; Mikkelsen, E.; Munck, L. *Prog. Colloid Polym. Sci.* **1998**, 166–174.
30. Wold, S.; Martens, H.; Wold, H. *Lecture Notes Math.* **1983**, *973*, 286–293.
31. Černá, M.; Barros, A. S.; Nunes, A.; Rocha, S. M.; Delgadillo, I.; Čopîková, J.; Coimbra, M. A. *Carbohydr. Polym.* **2003**, *51*, 383–389.
32. Filippov, M. P.; Shkolenko, G. A.; Kohn, R. *Chem. Zvesti* **1978**, *32*, 218–222.
33. Séné, C. F. B.; McCann, M. C.; Wilson, R. H.; Grinter, R. *Plant Physiol.* **1994**, *106*, 1623–1631.
34. McCann, M. C.; Bush, M.; Milioni, D.; Sado, P.; Stacey, N. J.; Catchpole, G.; Defernez, M.; Carpita, N. C.; Hofte, H.; Ulvskov, P.; Wilson, R. H.; Roberts, K. *Phytochemistry* **2001**, *57*, 811–821.
35. McCann, M. C.; Hammouri, M.; Wilson, R.; Roberts, K.; Belton, P. *Plant Physiol.* **1992**, *100*, 1940–1947.
36. van Alebeek, G. J. W. M.; van Scherpenzeel, K.; Beldman, G.; Schols, H. A.; Voragen, A. G. J. *Biochem. J.* **2003**, *372*, 211–218.
37. Kohn, R.; Machovič, O.; Machova, E. *Coll. Czech. Chem. Commun.* **1983**, *48*, 790–797.
38. Kohn, R.; Furda, I.; Kopec, Z. *Coll. Czech. Chem. Commun.* **1968**, *33*, 264.
39. Hotelling, H. *J. Educ. Psychol.* **1933**, *24*, 417–441.
40. Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B. *Appl. Spectrosc.* **2000**, *54*, 413–419.
41. Wold, S. *Technometrics* **1978**, *20*, 397–405.
42. Martens, H.; Nielsen, J. P.; Engelsen, S. B. *Anal. Chem.* **2003**, *75*, 394–404.
43. Nørgaard, L.; Bro, R.; Westad, F.; Engelsen, S. B. *J. Chemometrics* **2006**, *20*, 425–435.
44. Krzanowski W. J. Principles of Multivariate Analysis: A User's Perspective. Revised, edition ed. Oxford Science Publication: New York, 2000.
45. Rao C. R. Advanced Statistical Methods in Biometric Research. New York, 1952.