



Descriptive and predictive assessment of enzyme activity and enzyme related processes in biorefinery using IR spectroscopy and chemometrics

Baum, Andreas

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

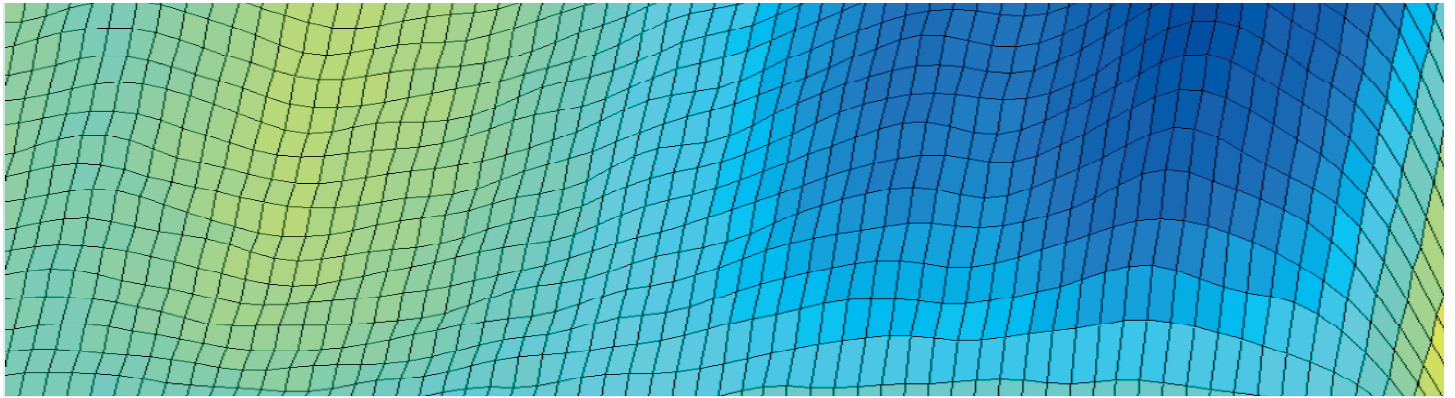
Citation (APA):
Baum, A. (2013). *Descriptive and predictive assessment of enzyme activity and enzyme related processes in biorefinery using IR spectroscopy and chemometrics*. Technical University of Denmark.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

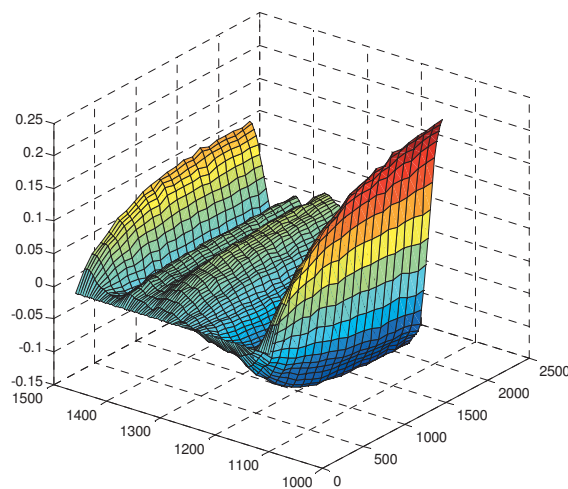
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

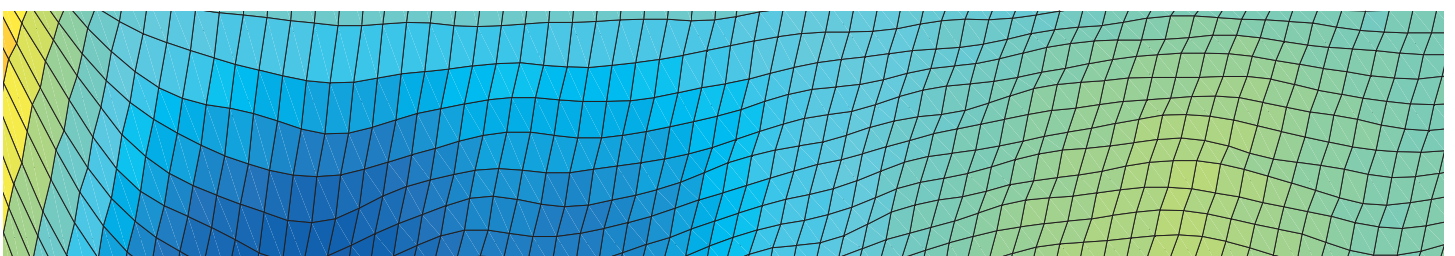


Descriptive and predictive assessment of enzyme activity and enzyme related processes in biorefinery using IR spectroscopy and chemometrics

Andreas Baum



FOSS



DESCRIPTIVE AND PREDICTIVE ASSESSMENT OF ENZYME
ACTIVITY AND ENZYME RELATED PROCESSES IN
BIOREFINERY USING IR SPECTROSCOPY AND
CHEMOMETRICS

Andreas Baum

Ph.D. Thesis

May 2013

Supervisors:

Prof. Jørn D. Mikkelsen

Per W. Hansen

Max Egebo

PREFACE

This Ph.D. thesis is submitted as a partial fulfillment of requirements for a Ph.D.-degree at Technical University of Denmark.

The research, which is the basis of this industrial Ph.D. project, has been carried out as collaboration between FOSS ANALYTICAL and Center for Bioprocess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark.

The project comprises work conducted under supervision of Prof. Jørn Dalgaard Mikkelsen (Department of Chemical and Biochemical Engineering, DTU), Per Waaben Hansen and Max Egebo (both from FOSS ANALYTICAL, Hillerød) between June 2010 and May 2013. The work was financed by the 7th Framework Program via the Marie Curie Initial Training Network, LeanGreenFood (EU-ITN 238084).

ACKNOWLEDGEMENTS

I sincerely thank my supervisors Per Waaben Hansen and Jørn Dalgaard Mikkelsen for inspiration, motivation, encouragement and support during the project work. During my time at FOSS, I had the pleasure to work with highly talented personalities which gave me a lot of inspiration to thrive in.

My acknowledgement goes also to Malgorzata Maria Dominiak, Silvia Vidal Melgosa, Anne Meyer and Jane Agger for the fruitful collaborations within and between Technical University of Denmark, University of Copenhagen, FOSS and Dupont. Without their enthusiastic effort this thesis would not have been possible.

Furthermore I highly appreciated helpful chemometric support from Rasmus Bro and his colleagues at University of Copenhagen. Lastly, I want to thank my family for the continuous support and especially Josephine for encouragement and love.

SUMMARY

Enzyme technology provides key strategies to green chemistry as many processes have undergone re-design to serve increasing demands towards being sustainable. While the population is rapidly increasing on our planet it is leading to accumulative problems in terms of production of waste, depletion of natural fossil resources and increasing demands for food and energy.

Biorefinery, in particular, deals with related challenges, as it is defined to deal with the conversion of biomass using enzyme technology to produce renewable energy, in terms of heat, power and fuel. Furthermore, biorefinery intends to extract value-added compounds from biomass to avoid down-cycling effects prior to e.g. biofuel production. Those value-added compounds are highly attractive to be utilized as food ingredients, bio-chemicals or precursors for pharmaceutical products and represent high market potentials for related industries.

However, as biorefinery concepts are implemented in many industrial processes an increasing demand for Process Analytical Technology (PAT) evolves to monitor, understand and steer processes optimally. Biomasses can be very diverse and are usually of complex chemical nature. Conventional univariate analytical methods therefore require time-consuming sample preparation which is mostly cumbersome to analyze biomass conversion related processes. Throughout this project alternative approaches will be presented to deal with the individual challenges. As outlined it seems obvious that more advanced techniques are necessary to monitor such difficult reactions as enzymatic biomass degradations. Such techniques should be of multivariate nature to capture and understand complex patterns in comparison to univariate techniques which can only capture information in a highly specific sense which does not allow interference of information. Vibrational spectroscopy (e.g. infrared) represents such multivariate techniques and is mostly used throughout the project. Data is analyzed by

chemometric methods to extract the underlying patterns from the complex datasets.

Hence, this project focuses on chemometric approaches utilizing mostly Fourier Transform Infrared (FTIR) spectroscopic data to provide descriptive and predictive insights into biomass conversion related processes.

Two main study fields are introduced to the reader. First, two-way chemometric methods are used to establish Process Analytical Technology (PAT) solutions for prediction of monosaccharide release efficiency of pretreated destarched corn bran using Near Infrared (NIR) spectroscopy (PAPER 1). Throughout this study predictive and descriptive models were established to evaluate the pre-treatment effect without the need to perform the subsequent enzymatic hydrolysis itself. Furthermore, the efficiency (and quality) of differently extracted pectin from lime peel could be predicted from measuring FTIR spectra (PAPER 2). The prediction models were compared to results retrieved from carbohydrate microarray analysis which additionally enhanced the understanding of the structural properties of the extracted pectin.

Secondly, enzyme kinetics of biomass converting enzymes was examined in terms of measuring enzyme activity by spectral evolution profiling utilizing FTIR. Chemometric multiway methods were used to analyze the tensor datasets enabling the second-order calibration advantage (reference Theory of Analytical chemistry). As PAPER 3 illustrates the method is universally applicable without the need of any external standards and was exemplified by performing quantitative enzyme activity determinations for glucose oxidase, pectin lyase and a cellolytic enzyme blend (Celluclast 1.5L). In PAPER 4, the concept is extended to quantify enzyme activity of two simultaneously acting enzymes, namely pectin lyase and pectin methyl esterase. By doing so the multiway methods PARAFAC, TUCKER3 and NPLS were compared and evaluated towards accuracy and precision.

RESUMÉ

Enzym teknologi er en af de nøgle strategier, der har stor betydning for forskning inden for området, ”den grønne kemi”. Mange processer i industrien er blevet ændret til at imødegå en bæredygtig grøn teknologi. Befolknings tilvæksten er meget stor og skaber problemer med hensyn til produktion af side-strømme (forurening), overudnyttelse af de naturlige fosile energi kilder, og et forøget krav til fødevarer og energi. Bio-rafinaderi kan være svaret på disse udfordringer, da denne teknologi fokuserer på omdannelse af biomassen ved hjælp af enzymer til produktion af grøn energi, der kan anvendes i form af varme, energi eller brændstof. Endvidere kan biorafinering udvinde værdiskabende forbindelser hvis en differentiell proces kan anvendes for biomassen nedbrydes til bio-brændstof produktionen. Disse værdiskabende forbindelser kan anvendes som fødevarer ingredienser (1), og repræsenterer et højt markedspotentiale i beslægtede industrier.

Biorafinerings konceptet er implementeret i mange industrielle processer og behovet for proces analytisk teknologi (PAT) for at måle, forstå og styre processen optimalt er meget værdifuld. Biomassen kan være meget forskelligt og er sædvanligvis meget komplekst opbygget med mange forskellige kemiske strukturer. Konventionel univariabel analytiske metoder er derfor meget tidskrævende både med hensyn til prøve forberedelse, og kompleksitet af de kemiske målemetoder. I dette PhD projekt er der anvendt alternative metoder for at måle disse komplekse biomasse processer. Der er adskillige udfordringer med at udskifte de klassiske teknikker med mere avancerede metoder, der kan måle de mange enzym reaktioner når biomassen nedbrydes. De multivariable metoder, der kan fange og forstå komplekse mønstre i modsætning til de univariable, som kun kan opfatte simple informationer, der ikke involverer interferens. Vibrations spektroskopi (f. eks. infrarød) repræsenterer multivariable teknikker, der er anvendt bredt i dette PhD studie. Data er blevet analyseret ved hjælp af

kemometriske metoder for at ekstrahere de komplekse mønstre fra de forskellige dataset.

Dette PhD projekt har fokuseret på kemometriske anvendelser baseret hovedsageligt på Fourier Transform Infrared (FTIR) spektroskopiske data, der kan beskrive og forudsige enzym baserede biomasse processer.

To hovedområder er blevet introduceret til læseren. Det første er en to-vejs kemometrisk metode, der er anvendt til at etablere Process Analytical Technology (PAT) løsninger til at forudsige frigørelse af monosaccharider fra forbehandlet ”de-starched corn bran” ved hjælp af Near Infrared (NIR) spektroskopi (artikel 1). I dette studie er der etableret modeller til at forudsige og beskrive effekter af forbehandling uden at udføre en egentlig enzymatisk hydrolyse. Endvidere kunne ekstraktion af pektin fra citrus skaller udført med forskellige processer forudsiges og beskrives ved en måling med FTIR spektrum (artikel 2). Modellen for forudsigelsen blev sammenlignet med data opnået fra kulhydrat micro-array analyser, som bidrog yderligere med analyser på kulhydraternes strukturelle egenskaber.

Det andet område er biomasse åbning, hvor enzymeres aktivitet og kinetik blev målt ved ”spectral evolution profiling” ved hjælp af FTIR. Kemometriske multivariable metoder blev anvendt til at analysere tensor datasets, der tillader en anden ordens kalibrerings fordel. Som artikel 3 illustrerer, er metoden universiel anvendelig uden det klassiske behov for eksterne standarder og dette er bevist ved at anvende kvantitative enzym aktivitets målinger for glucose oxidase, pektin lyase og cellulolytiske blandinger (Celluclast 1,5L). I artikel 4, er konceptet blevet udvidet til at kvantificere to enzym aktiviteter samtidigt, nemlig pektin lyase og pektin methylesterase. I dette forsøg blev de tre multiveje metoder, PARAFAC, TUCKER3 og NPLS sammenlignet og vurderet med hensyn til nøjagtighed og præcision.

LIST OF PUBLICATIONS

PAPER 1

Andreas Baum, Jane Agger, Anne S. Meyer, Max Egebo, Jørn D. Mikkelsen; *Rapid near infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatments*; New Biotechnology, Volume 29, Issue 3, 15 February 2012, Pages 293-301

PAPER 2

Andreas Baum, Malgorzata Dominiak, Silvia V. Melgosa, William G.T. Willats, Karen M. Sondergaard, Per W. Hansen, Jørn D. Mikkelsen; *FTIR and carbohydrate microarray analysis for prediction and characterization of enzymatic versus acidic extracted pectin*; Carbohydrate Polymers, submitted

PAPER 3

Andreas Baum, Anne S. Meyer, Javier L. Garcia, Max Egebo, Per W. Hansen, Jørn D. Mikkelsen; *Enzyme activity measurement via spectral evolution profiling and PARAFAC*; Analytica Chimica Acta, Volume 778, 17 May 2013, Pages 1-8

PAPER 4

Andreas Baum, Per W. Hansen, Anne S. Meyer, Jørn D. Mikkelsen; *Simultaneous measurement of two enzyme activities using infrared spectroscopy: comparative evaluation of PARAFAC, TUCKER and N-PLS modeling*; Analytica Chimica Acta, submitted

ABBREVIATIONS

ANN	Artificial Neural Networks
biPLS	backwards interval Partial Least Squares (regression)
CBM	Carbohydrate Binding Module
DCB	Destarched Corn Bran
DNA	Deoxyribonucleic Acid
FT	Fourier Transform
FTIR	Fourier Transform Infrared
GRAM	Generalized Rank Annihilation Method
HGI	Homogalacturonan I
HPLC	High Performance Liquid Chromatography
iPLS	interval Partial Least Squares (regression)
IR	Infrared
LDA	Linear Discriminant Analysis
LV	Latent Variable
mAb	Monoclonal Antibody
MLR	Multiple Linear Regression
MPCA	Multiway Principal Component Analysis
MSC	Multiple Scatter Correction
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
NPLS	Multiway Partial Least Squares (regression)
PARAFAC	Parallel Factor Analysis
PARAFAC2	Parallel Factor Analysis 2
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PCR	Principal Component Regression
PL	Pectin Lyase
PLS	Partial Least Squares (regression)
PLS-DA	Partial Least Squares - Discriminant Analysis
PME	Pectin Methyl Esterase
RGI	Rhamnogalacturonan I
RGII	Rhamnogalacturonan II
RMSEC	Root-Mean-Square-Error of Calibration
RMSECV	Root-Mean-Square-Error of Cross Validation
RMSEP	Root-Mean-Square-Error of Validation
SIMCA	Soft Independent Modeling of Class Analogy
SNV	Standard Normal Variate

TABLE OF CONTENTS

1.	Introduction	1
2.	Enzyme Technology and Biorefinery	2
3.	Infrared spectroscopy	3
3.1	Near Infrared Spectroscopy	3
3.2	Fourier Transform Infrared Spectroscopy	3
4.	Chemometrics	4
4.1	Two-way methods	4
4.2	Multi-way methods	6
4.3	Preprocessing	7
4.4	Validation	7
5.	Two-way chemometric methods for prediction of process relevant parameters (PAT)	8
5.1	Near Infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatments	8
5.2	FTIR and carbohydrate microarray analysis for prediction and characterization of enzymatic versus acidic extracted pectin	11
6.	Temporal Evolution Profiling and chemometric multiway analysis for determination of enzyme activity	15
6.1	Measuring enzyme activity using FTIR and PARAFAC	15
6.2	Simultaneous measurement of two enzyme activities: a comparative evaluation of PARAFAC, TUCKER3 and N-PLS modeling	19
7.	Conclusion and Perspectives	23
8.	References	24

1. Introduction

The beneficial use of enzymes for production of foods and beverages like cheese, yoghurt, bread, beer, vinegar, wine etc has been known for thousands of years. In fact paper and textiles were produced with the help of enzymes which were present in starting materials as early as 6000 BC in China, Sumer and Egypt [1].

Nonetheless, the industrial application of enzymes did not kick off before 1913 when Otto Roehm filed a patent on a crude protease mixture isolated from pancreases for use in laundry detergents [2]. The modern era of industrial enzymology had begun. Not least because of evolving recombinant biotechnological methods, many enzymes have been made commercially available up to today [3, 4]. Applications of enzymes can be found in biorefinery, textile, detergent, starch, paper and food industry, including the dairy, juice, brewing, wine and baking industries.

As the development of enzyme technology evolves fast, the refurbishment of industrial processes to be more environmentally clean and sustainable goes hand in hand. For many decades production processes were merely designed and optimized for cost efficient production using classical chemical methods which apply harsh chemicals accumulating and emitting a lot of waste to our environment. Furthermore those classical designed processes are not selective, give rise to side-streams of low quality, ask for rather extreme reaction conditions and are therefore energy inefficient. While many of such processes were replaced by more sustainable processes implying the concept of green chemistry [5], many resources are still down-cycled. To avoid such down-cycling effects Braungart and McDonough recently proposed a change in production philosophy towards a “cradle to cradle” perspective [6] which generally states the idea of re-designing industrial processes to *expand* our understanding of sustainability. Claiming the possibility of utilizing the, terrestrial spoken, infinite amount of renewable energy from the sun “Cradle to Cradle” processes should not be primarily optimized for energy consumption, but rather for being intrinsic meaning that all resources being used can and should be perfectly recycled instead of being lost at the end of a products lifetime.

In fact, enzyme technology presents a key strategy to such cradle-to-cradle processes since enzymes can bio-catalyze the formation of many industrial

products. The obvious advantage is that mild reaction conditions in aqueous media can be applied for direct conversions from substrates to the products while conventional approaches often involve several reaction steps possibly employing environmentally unfriendly organic solvents. Adding up to this, enzymes are highly chemo-, regio- and stereoselective enabling the possibility for design of targeted processes.

Those advantages are also very important for the concept of green chemistry [1]. An example for industrial application of green chemistry has been the production of penicillin [5]. While the conventional production of semi-synthetic penicillin requires several steps including a chemical hydrolysis using unattractive chemicals as CH_2Cl_2 at $-40\text{ }^\circ\text{C}$ the one step bio-catalytic conversion can be performed in water at $37\text{ }^\circ\text{C}$ using only light amounts of NH_3 for pH adjustment.

As by 2002 already more than 130 processes have been established using green chemical approaches as reported in [7]. The increasing industrial interest in green process design gives rise to high expectations of many more to come, as by now recombinant DNA techniques makes it, in principle, possible to produce virtually any enzyme for a commercially acceptable price [1]. However, it keeps challenging to avoid down-cycling of valuable resources.

The present project is part of a larger cross-European collaboration effort that focuses on production of value-added compounds in biorefinery relevant processes. As biorefinery is concerned about sustainable biomass conversion processes to produce fuels, power, and heat from biomass, especially the value-added components in diverse biomasses can elevate the economy of a process, e.g. the specific extraction of nutrition relevant biopolymers from biomass prior to biofuel refining.

The hypothesis of the project was that fast and non-invasive techniques, such as Fourier Transform Infrared (FTIR) and near infrared (NIR) spectroscopy, in combination with chemometrics create useful analytical solutions for qualitative and quantitative assessment of enzyme activity and enzyme related processes in biorefinery. The methodologies thereby provide much better understanding of the complex and obtrusive nature of biomass and offer valuable analytical alternatives, also with respect to the cumbersome reducing sugar analysis.

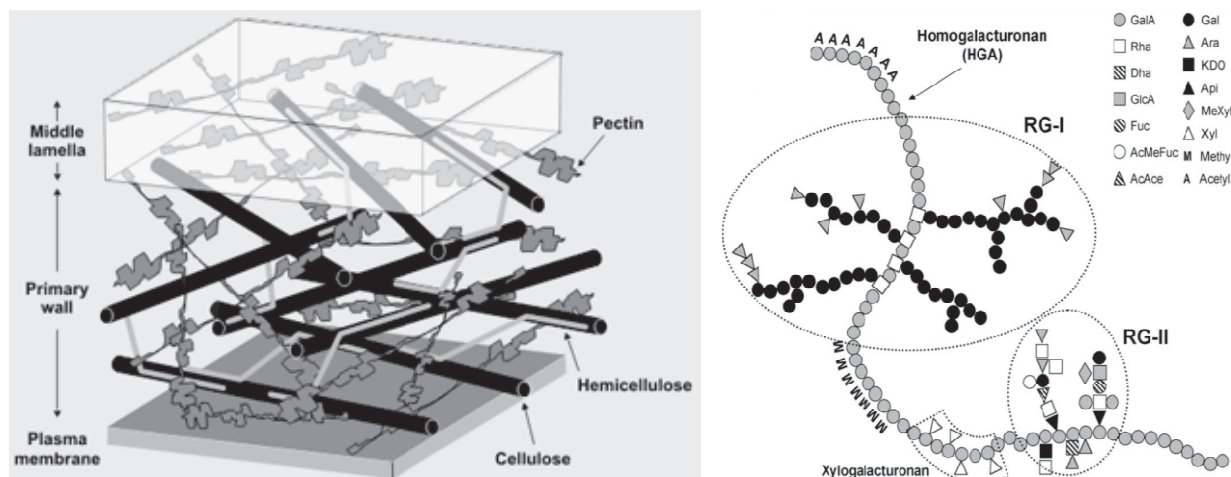


Figure 1. left: illustration of primary cell wall structure containing complex structural polysaccharides [10]; right: pectin is a structural heteropolysaccharide which binds cell wall components together [11].

The specific objectives were to apply:

1. Near infrared spectroscopy as a rapid and non invasive technique to evaluate biomass pretreatment effects towards saccharification efficiency;
2. Fourier transform infrared spectroscopy and carbohydrate microarray analysis as multivariate techniques to monitor pectin extraction in different media and to evaluate its structural properties;
3. Fourier transform infrared spectroscopy as a universal fingerprinting technique to measure enzyme activity of one or several enzymes without the need of environmental unfriendly chemicals.

Throughout this thesis four scientific articles will be introduced to the reader. Prior to summarizing the results of those individual findings a short description of the used concepts in enzyme technology and chemometrics shall be given. Finally a conclusion will be drawn followed by a discussion of future perspectives.

2. Enzyme Technology and Biorefinery

Biorefinery is a sub-discipline of biotechnology which deals with the bio-catalyzed conversion of biomass for production of fuels, heat and power to provide renewable energy sources in contrast to depletion of fossil resources. However, biorefinery also intends to produce high-added-value products, such as solvents, plastics and food

ingredients [8, 9] which promise high market potentials.

Biorefineries are often divided into first and second generation biorefineries. As first generation biorefineries are concerned about bioconversion of biomasses like corn, sugarcane or starch it has been controversy criticized regarding its competition with food production. On the other hand, second generation biofuel production utilizes biomasses derived from non-food crops or inedible waste products to avoid this competition. This leads to possible classification between biomass-producing and waste-material-utilization type refineries [12]. China, USA, Australia and other south-east asian countries are typical biomass-producing type countries while Japan, Denmark and other European countries represent waste-material-utilization type countries due to the lack of available agricultural space [9].

However, as mentioned earlier the utilized biomasses often contain polysaccharides of high value, such as pectin, xyloglucan, etc. (Figure 1). Biorefinery processes therefore increasingly focus on specific extraction of such biopolymers since they were found to be beneficial to human health when used as dietary fibers in food production [13]. Furthermore, polysaccharides as pectin can be used in the food ingredient industry to engineer desired product qualities, e.g. the gelling of jams.

The complex composition of cell wall material is therefore primarily degraded in a highly targeted sense to extract polysaccharides of interest using specific enzymes. Hereby, the used biocatalysts indicate advantages which help to extract

polysaccharides with desired structural properties. Additionally, enzymes are employed to break down residual and waste-material to release monosaccharides which can be used readily in bioethanol fermentation.

Because of their chemical complexity enzymatic biomass conversion processes require a comprehensive understanding and engineering. Enzyme technology helps to understand enzymatic catalysis, which involves studies concerning discovery and manufacturing of enzymes and its optimal working conditions (temp, pH, pressure) and immobilization and inhibition of enzymes. It also deals with investigations towards the mechanisms of enzymatic conversions and studies underlying kinetic phenomena.

However, many conventional analytical methods employed to investigate biomass conversions are of univariate nature and require time-consuming sample preparations utilizing diverse chemicals. Throughout this project alternative approaches will be presented to deal with the individual challenges. As outlined it seems obvious that more advanced techniques are necessary to monitor such difficult reactions as enzymatic biomass degradations. Such techniques should be of multivariate nature to capture and understand complex patterns in comparison to univariate techniques which can only capture information in a highly specific sense which does not allow interference of information.

3. Infrared spectroscopy

Infrared (IR) spectroscopy accounts for measuring the interactions of electromagnetic light from the infrared region with inter-atomic covalent bonds and intermolecular bondings of a sample. Since all organic molecules have intra- and intermolecular bonds that absorb in this range the IR spectrum obtained therefore reflects the chemical and structural composition of the sample being analyzed in a fingerprint sense. Further comprehensive introduction on Infrared spectroscopy is given by [14-16].

Hence, a spectral fingerprint measured by IR spectroscopy can be related to a certain thermodynamic state of a measured sample. In addition, the concept can be extended to measure time-resolved changes of such fingerprints enabling the possibility to study kinetics e.g. of enzymatic reactions. However, such analytical methods employing IR spectroscopy have been

proven to be difficult due to high water interferences originating from the aqueous reaction media and a high degree of overlapping information in the spectra.

During this project Near Infrared (NIR) spectroscopy and Fourier Transform Infrared (FTIR) Spectroscopy were utilized to study a) thermodynamic states of carbohydrate systems to extract PAT relevant parameters and b) kinetic changes in carbohydrate systems to quantify enzyme activity in a more universal fashion.

3.1 Near Infrared Spectroscopy

Near infrared (NIR) spectroscopy is concerned about IR spectroscopy in the range between 780 and 2500 nm. It typically contains highly overlapped absorption patterns due to overtone and combined vibrational modes. As water gives rise to very high interferences NIR is often used to characterize solid samples using, e.g., diffuse reflectance methods. The high degree of overlapping information made NIR somewhat inaccessible to chemical analysis until chemometrics evolved in the 1970s [17] which helped to understand the complex patterns in the spectra and to build calibration models. One advantage of NIR is that it can typically penetrate much farther into a sample than mid infrared radiation [18]. Near infrared spectroscopy is therefore directly applicable for analyzing bulk material without the need of complex sample preparation. These striking advantages of NIR gave rise to many industrial applications, especially to be found in routine control analyses in agricultural food, petrochemical, pharmaceutical, clinical and environmental industrial sectors [17, 19].

3.2 Fourier Transform Infrared Spectroscopy

Mid-infrared spectroscopy is concerned about IR spectroscopy in the range between 2.5 μm and 25 μm . While diverse NIR instruments are commercially available using monochromator and Fourier Transform (FT) technology for light dispersion, instruments for the mid-infrared region are dominantly using FT technology [20]. Besides being the cheaper solution the FT technology has several advantages, especially in the mid infrared region. Typically using a Michelson Interferometer, a spectrum is recorded for all wavelengths simultaneously resulting in an interferogram in contrast to scanning the wavelengths consecutively using a dispersive device. The interferogram is thereafter de-

Table 1. Possible classification of chemometric methods into exploratory methods (descriptive) and methods being suitable for calibration modeling (predictive). The individual methods should not be understood as strictly assigned to a one group since some methods can be used as descriptive and/or predictive tools. ANN can be used in diverse situations. However, ANN requires very large sample sizes to establish robust models.

	exploratory	calibration modeling
two-way	PCA, LDA, PLS-DA, SIMCA, ANN, dendrogram analysis	MLR, CLS, PCR, PLS, ANN
multi-way	PARAFAC, PARAFAC2, TUCKER3, MPCA	GRAM, NPLS

convoluted using Fourier Transformation [21]. This results in faster spectrum acquisition and enables the possibility for replicate measurements to encounter interfering noise. Hence, FTIR offers better sensitivity and a better signal-to-noise ratio than NIR spectroscopy.

FTIR is widely used in diverse kinds of scientific and industrial fields, e.g. in quality control analysis of milk and wine [22], as a PAT tool in bioprocesses [23], for characterization of plant cell wall material [24] or as a typical technique for sample classification [25]. In contrast to NIR spectra, which contain highly overlapped information, a mid-IR spectrum often indicates more resolved vibration bands which often enables the possibility for qualitative analysis of univariate kind, i.e. the identification of particular bands [26, 27]; although the full potential of FTIR is exploited by using multivariate methods to understand complete underlying spectral patterns.

4. Chemometrics

Chemometrics describes a scientific discipline which utilizes mathematical and computational tools to understand multi-parametric data in a descriptive and predictive way. It evolved as a consequence of the introduction of computational technology in chemical laboratories [28], which enabled acquisition of very big data sets. These data sets were difficult to understand with common univariate methods as they intend to extract correlations between single analytical responses and the investigated underlying phenomenon. However, those newly available data sets consisted of many variables encrypting mostly overlapped information, e.g. from spectroscopy. To extract the underlying patterns in the data which could lead to a comprehensive

understanding of its underlying (bio)-chemistry multivariate methods were necessary.

As previously mentioned chemometric methods could be divided into descriptive and predictive methods. However, they could also be distinguished in terms of being exploratory or being suitable for calibration modeling. A proposed overview is presented in table 1. Detailed descriptions of the chemometric methods used during the individual studies are given in the corresponding PAPERS 1-4.

4.1 Two-way methods

4.1.1 Descriptive chemometric methods

Our brain uses millions of receptors gathering tons of cues in every second of our life. This gathered information is filtered through neural networks to achieve very certain (and trained) pattern recognitions. Those highly complex (and non-linear) recognized patterns lead to every day's abilities as recognizing voices, tastes, scents, looks and textures of certain materials and living beings. Strikingly, we are able to recognize our friends' faces through billions of people on this planet using multivariate approaches, sorted out by our brain.

Similarly descriptive chemometric methods aim to extract patterns from chemical and biochemical data sets [29]. Due to limited available computational power and algorithms the amount of parameters to be analyzed simultaneously is much lower comparing it to neural networks of a human brain, but complexity evolves as algorithms and computational power are also being developed further [30]. The most principal descriptive chemometric method is Principal Component Analysis (PCA) which was initially described by Hotelling in 1933 and brought to the chemometric world by Wold in 1987 [31, 32]. By

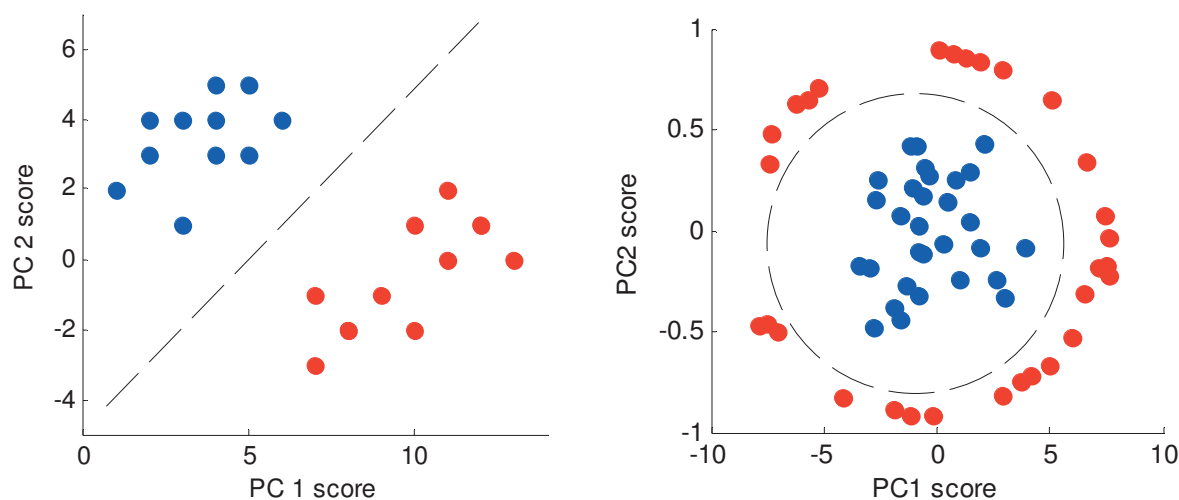


Figure 2. Sample scores derived from PCA can be used to distinguish between different sample groups using classification methods. To discriminate between two groups of samples linear discriminants (a) can be sufficient while other situations might demand non-linear discriminants (b). Discriminants are indicated by dashed lines.

performing a mathematical eigenvector decomposition on a data sets covariance matrix it finds suitable orthogonal projections to transform complex data sets, containing information along many variables, into spaces which have less significant dimension to represent the variance in the data more intuitively [33]. This is referred to as feature reduction and is one common property of chemometric methods.

More precisely, PCA decomposes data sets into loadings and scores. As loadings can be understood as orthogonal projections in space which describe the major variance in the data, the scores indicate how abundant the mentioned loadings are for each sample to be described in the feature reduced space. The scores are therefore a useful measure to compare samples with each other. Considering e.g. FTIR spectra of olive oils from two different regions, one can expect discrimination ability between the two origins, by simply plotting the scores of the major Principal Components against each other, assuming that the FTIR spectra contain discriminating spectral features. Such approaches can lead to classification of samples as they can form clusters in score plots. Clustering can be useful to discriminate between various properties in terms of origin, taste, (bio)-chemical properties, etc. As illustrated in Figure 2 clusters can form different shapes as they are colored in red and blue.

Since PCA is an unsupervised decomposition method which simply represents the over-all variance in the data it does not intend to separate the clusters to the highest extent, neither does it know about the group belonging of individual samples.

To establish models which are able to discriminate between pre-defined groups supervised classification methods are necessary. Dependent on the shape of the discriminant as shown in Figure 2 linear or non-linear classification methods are utilized. The most basic methods used for linear discrimination between sample groups (Figure 2a) are Linear Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLS-DA). On the other hand methods as Support Vector Machines (SVM) [34, 35] or Artificial Neural Networks (ANN) [36, 37] are used for non-linear cases as illustrated in Figure 2b.

4.1.2 Predictive chemometric methods

Analytical chemistry often requires quantitative estimations of certain properties of (bio)-chemical system possibly using fast and non-invasive techniques [38].

Those properties may be any kind of quality measure of a final product, intermediate product or starting material. During a process, relevant parameters as temperature, pH, time, etc. could be of high interest while often simply the yield of a process is to be predicted; in simple words: any

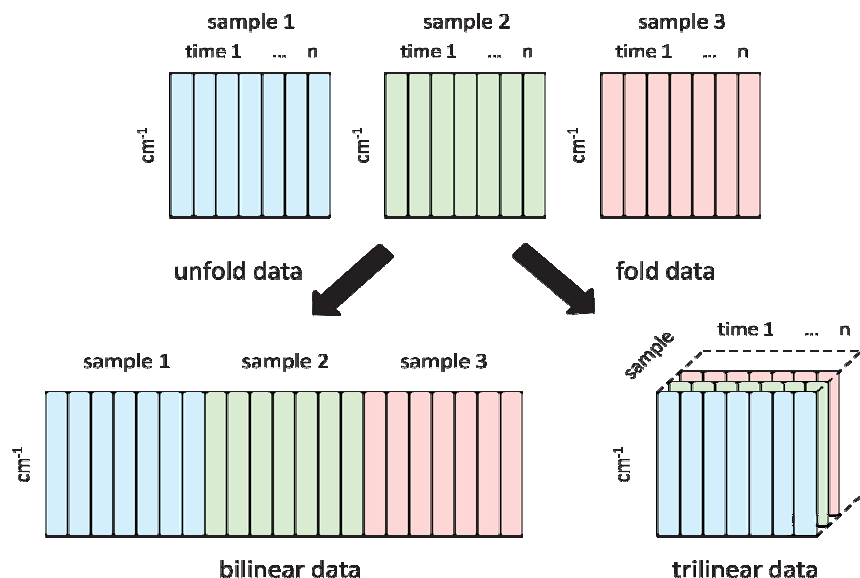


Figure 3. Samples which are recorded using coherent time steps and coherent spectral window can be analyzed either by two-way methods using unfolded data or by multiway methods using folded data in tensor format.

kind of quantitative measure regarding a process could be of interest.

While determining those properties by traditional univariate analytical techniques has been cumbersome for decades predictive chemometric methods, as given in Table 1, have shown to be the key to provide fast and non-invasive analytical solutions, especially in industry, where one intends to reduce efforts concerning complex sample preparation [39, 40].

Predictive chemometric models are usually established from two data blocks. Multivariate measurements techniques as infrared (IR) spectroscopy, mass spectrometry, Nuclear Magnetic Resonance (NMR), etc. are utilized to generate arrays of data for various samples which are typically stored as a combined matrix in the so-called X-block. For each vector in such a multivariate data matrix the known reference value of the property of interest is typically determined by conventional analysis (e.g. HPLC) and stored in the Y-block.

While descriptive chemometric methods usually only decompose X-blocks it is now possible to use the information in the Y-block to find suitable calibration projections in the multivariate space of the X block. The simplest approach for this is known as Principal Component Regression (PCR) where Principal Components obtained from PCA are regressed against the Y-block. Another, although related, approach is called Partial Least Squares (PLS) [41] regression where Principal

Components are rotated in space to find new projections which indicate maximal covariance with the Y-block data. Hence, the difference to PCR is that the orthogonality of the individual loadings in the principal components (results from PCA) is omitted.

Once a prediction model is obtained it can be used to predict the wanted property from a single multivariate measurement without the need of reference analysis. Handy introductions to the topic of multivariate calibration are given in [39, 41-43].

4.2 Multi-way methods

Previously mentioned methods are referred to as two-way methods, meaning that their bilinear data can be expressed in matrices, where each sample is represented by a vector.

On the other hand chemometric multi-way analysis is concerned about higher order data [44]. This kind of data demands new methods which are capable to deal with data sets expressed in three dimensional arrays, as we can imagine it for the simplest, tri-linear case. In this particular case each sample is represented by a two dimensional matrix of analytical responses. The individual elements of each response matrix therefore depend on an additional measurement dimension, e.g. time. Accordingly quadri-linear data would consist of four measurement dimensions, where each sample would contain a cube of data itself. First attempts to analyze higher order data were made in the 1960s and 1970s in the field of

psychometrics [45-47] which led to the introduction of multiway analysis in chemometrics in the 1980s [48, 49]. As reviews on multi-way chemometrics were published in 1995 and 2005 by Bro et al. the number of publications in the field of multi-way analysis increased significantly from around 120 to more than 300 with many more to come after 2005 [49, 50].

So far, multiway analysis has been applied using diverse data sets derived from fluorescence spectroscopy [51, 52], chromatography [53, 54], flow injection analysis [55], magnetic resonance spectroscopy [56, 57], NIR [58], UV-VIS [59] and other types of data. The rising number of applications clearly shows the high potential of multiway methods in many scientific disciplines.

To introduce the concept of multiway methods in more detail, the difference between two-way and multi-way data structures is illustrated in Figure 3. When talking about multiway methods the borderline between descriptive vs. predictive methods (Table 1) is more blurry than for two-way methods. An example could be Parallel Factor Analysis (PARAFAC) [47, 60]. When performing a PARAFAC one obtains a unique solution, meaning that the solution has no rotational ambiguity [60], but rather characterizes common unique underlying profiles for all samples. This leads to the so called second-order calibration advantage, meaning that the obtained calibrations retrieved from PARAFAC are robust against unknown interferences when predicting future test samples. To benefit from the second-order advantage the interfering samples must, however, indicate loadings in mode 2 and 3 which differ from the calibration model [60, 61].

In addition, the ability to obtain calibration models from unsupervised methods like PARAFAC leads to another advantage. While supervised regression methods like multiway Partial Least Squares regression (NPLS) [62] open up possibilities for biasing the results of the models, unsupervised tensor decomposition methods as PARAFAC and TUCKER3 [46, 63] can find prediction models without the need for reference analysis. However, since no scaling of the scores would be available the results would only contain valuable information when comparing samples in relation to each other. Nonetheless, this can be very valuable when relative results are desirable.

4.3 Preprocessing

Prior to application of any kind of chemometric method the data is usually preprocessed. The purpose is to remove any kind of irrelevant systematic variation in the data to gain better prediction performance and to decrease the necessary number of factors for modeling. The major two types of preprocessing are centering and scaling [64, 65]. Centering typically removes offsets in spectral data and therefore reduces the rank of the dataset. A PCA applied to non mean-centered spectroscopic data would therefore typically result in an additional Principal Component representing the common spectral offset in relation to the origin of the coordinate system. In addition, all other Principal Components indicating spectral features in the dataset would therefore be biased by the offset since PCA finds Principal Components being orthogonal to each other. It is therefore important to remove such offsets using appropriate centering methods.

Secondly, scaling methods are concerned about leveling the impact of the individual variables during modeling since the individual leverages of the used variables might be biased due to different measurement units (e.g. units of temperature, pressure, concentration). To correct for such effects auto-scaling is typically used. In spectroscopic data such biasing effects are not present since all variables are measured using the same units. In this case it can be believed that variables with a great deal of variation are indeed more important to describe a dataset. However, other non-linear effects in spectroscopic data, such as unwanted light scattering effects, need to be considered. Those effects are typically corrected for using preprocessing methods as Standard Normal Variate (SNV) [66] or Multiple Scatter Correction (MSC) [67, 68] on a per-spectrum basis.

4.4 Validation

Chemometric modeling is usually supported by strong concepts of validation. These models are often established using factors derived from principal components and the question arises of how many are necessary to describe a data set sufficiently without implementing random variation as noise in the model (overfitting). Cross validation and prediction errors as the root-mean-square-error of prediction (RMSEP) provide useful indications to find the right number of factors to be used in a model.

When using cross-validation the data is usually split into equally sized groups and a model is calculated several times excluding an alternating group at a time. Thereafter the established models are used to predict exactly those left-out groups and the mean error is calculated accordingly. This procedure can be repeated using different numbers of factors. The resulting root-mean-square-errors of cross validation (RMSECV) are then plotted against the used number of components. Thus, this plot is able to indicate overfitting by an increasing RMSECV when an increasing number of factors is used.

On the other hand, the RMSEP derives from test set validation, meaning that an established model is used to test for prediction ability towards an independent set of samples. By doing so the RMSEP also indicates additional concerns as poor accuracy of a model or the presence of potential systematic bias by the test set samples themselves. If a systematic bias is present the responsible chemometrician should consider that the used data for the calibration did not contain variance as it was only present in the test set. Hence, the RMSEP represents a measure for model accuracy.

5. Two-way chemometric methods for prediction of process relevant parameters (PAT)

Prediction of valuable process analytical parameters offers high market potentials as customers increasingly demand more certified and standardized quality measures for industrial products. The pharmaceutical industry introduced process analytical technology (PAT) in 2003 by defining PAT's as systems for design, analysis, and control of manufacturing processes. Those systems are utilized to assure high quality through timely measurements of critical quality and performance attributes of raw materials, in-process materials, and final products [69, 70].

Industrial implemented PAT in e.g. industrial fermentation of pharmaceutical products are exemplified by studies as "Batch-to-batch reproducibility of fermentation processes by robust process operational design and control" [71] or "Chemometrics in bioprocess engineering: process analytical technology (PAT) applications" [72].

While in the past cumbersome and highly specific univariate chemical methods have been used to

characterize such chemical systems multivariate methods combined with spectroscopy offer fast and high through-put capable approaches to measure analytes or properties of interest which typically indicate no need to interrupt the process. By doing so the major aim is to design, develop and operate processes consistently to ensure a predefined quality at the end of the manufacturing process [72].

5.1 Near Infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatments

5.1.1 Background and hypothesis

Enzymatic conversion of biomasses is an important step in the exploitation of biomass residuals in the food and biofuel industry and viable processes are dependent on reliable, fast and low-cost process conditions. It is of great importance to know the basic monosaccharide composition and the distribution in polysaccharide forms to target and develop enzymatic conversion, and this traditionally represents a demanding task. In addition, most low-priced biomass resources are of a recalcitrant nature, which is why various hydrothermal or thermo-chemical pretreatment forms are often applied to maximize the yields of enzymatic conversion [73-75]. Such pretreatments are costly and the efficiency of the pre-treatment step in generating a substrate which is enzymatically degradable needs utmost optimization. Evaluations of pretreatment methods are commonly done by large experimental setups and require time-consuming and advanced analytical methods [76, 77].

In the present work differently pretreated samples (varying time, pH, temp) originating from destarched corn bran (DCB), an agro-industrial residue from corn starch processing, have been evaluated using NIR spectroscopy. Destarched Corn bran is a fibrous, heteroxylan rich side-stream from the starch industry which may be used as a feedstock for bioethanol production [78, 79] or as a source of xylose for other purposes (flavors, pharmaceuticals, animal feed) [80].

During this study the specific hypothesis was that NIR spectroscopy offers a rapid and non-invasive technique to predict the saccharification efficiency of destarched corn bran after various pretreatments, but prior to enzymatic hydrolysis.

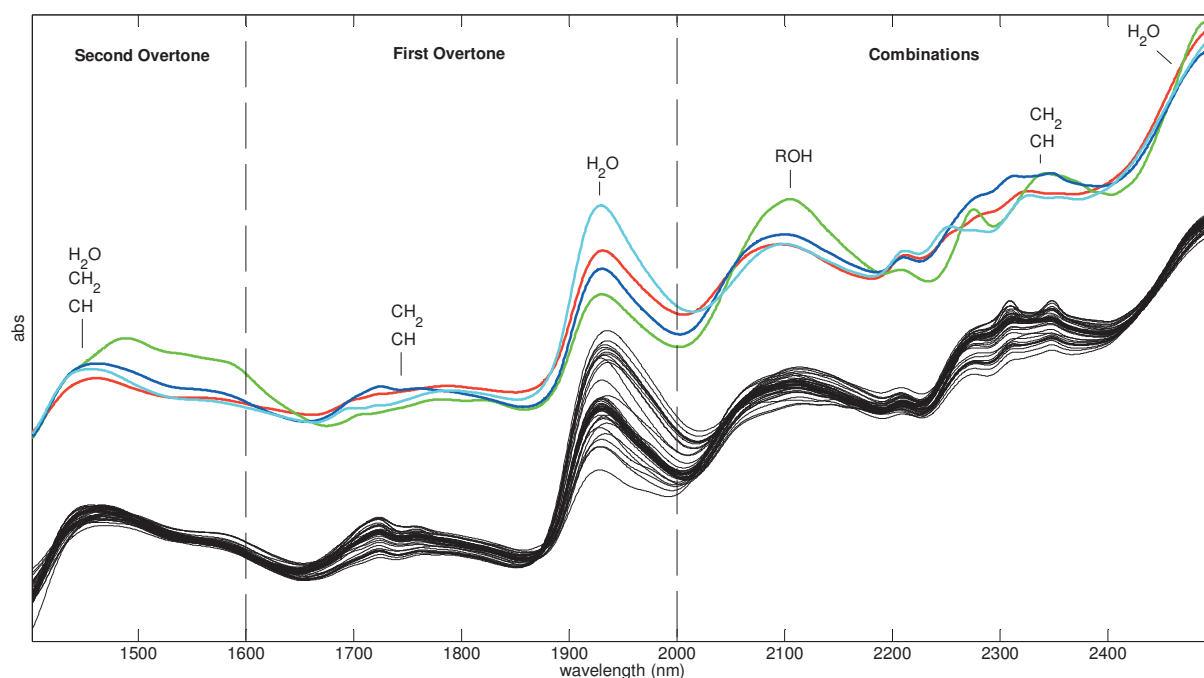


Figure 4. NIR spectra of 43 differently pretreated DCB samples. Standard spectra for arabinoxylan soluble (red), arabinoxylan insoluble (blue), cellulose (green) and xylan (cyan) have been included with an offset for comparison. Various combination, first overtone and second overtone vibration bands have been highlighted for chemical interpretation.

5.1.2 Experimental

Diffuse reflectance near Infrared spectroscopy (NIR) was used as a rapid and non-destructive analytical tool for evaluation of the pretreatment effects. Spectra were acquired between 1400 and 2500 nm using a monochromator based XDS instrument connected to a Rapid Content Analyzer (FOSS ANALYTICAL, Hillerød). The 43 destarched corn bran samples were measured directly as powders with no further physical pretreatment using a round quartz cuvette (Hellma, Sussex, UK). The 43 samples were divided into five groups due to their pretreatment conditions (indicated colors refer to Figure 5):

- A) raw destarched corn bran (green color)
- B) elevated temperature pretreatment (red color)
- C) mid-range pH catalyzed pretreatments at intermediate temperatures (blue color)
- D) low pH catalyzed pretreatment at intermediate temperature (light blue color)
- E) pH catalyzed pretreatments with extended incubation time (black)

5.1.3 Chemometric Analysis

Prior to any modeling all spectra were pretreated by Multiple Scatter Correction (MSC) and mean-centered to correct for unwanted light scattering

effects and to remove obstructive background signals.

Multivariate analysis of the 43 acquired spectra was performed to identify classification criteria between the five pretreatment groups. To do so, PCA and hierarchical clustering methods (Dendrogram) were used to distinguish between the different pretreatment groups and to identify outliers clearly.

Secondly, backwards interval Partial Least Squares regression (biPLS) and normal PLS were used to establish quantitative multivariate prediction models towards the monosaccharide release after standardized enzymatic hydrolysis. The three monosaccharides which were determined by HPLC reference analysis after hydrolysis were glucose, arabinose and xylose. Leave-one-out cross validation was used during modeling to validate the results and to evaluate the necessary number of components. By using an interval selection algorithm [81, 82] important wavelengths could be highlighted for modeling the individual monosaccharides. However, the small sample size did not leave abilities for further validation in terms of prediction test set validation.

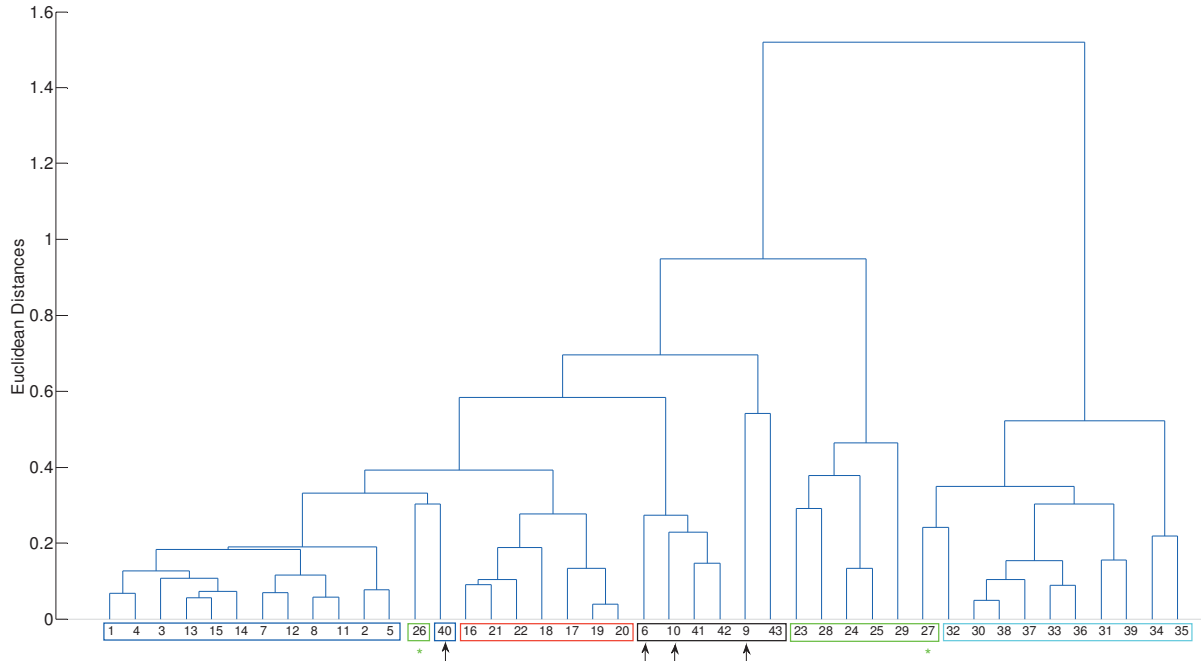


Figure 5. A dendrogram of 43 DCB samples led to clear classification and identified outliers (*). Arrows additionally indicated samples which belonged rather to another pretreatment group than expected. They were therefore merged into their new groups (PAPER 1).

5.1.4 Results

The 43 acquired NIR spectra are shown in Figure 4 together with relevant spectra of biopolymer standards. When looking at the spectra it can be seen that the DCB samples are most comparable with the blue standard spectrum, indicating insoluble arabinoxylan.

Using descriptive chemometric methods as PCA resulted not only in formation of clusters according to pretreatment grouping, but also in identification of outliers. A score plot is given in PAPER 1. Additional evidence to identify certain samples as outliers was given by dendrogram plotting (marked as stars in Figure 5). To establish classification using the dendrogram scores from five principal components have been used to calculate the Euclidean distances using an average linkage algorithm [26]. In fact, some samples

could be identified to belong to another group than originally expected (arrows Figure 5). The achieved classification therefore introduced useful quality measures of the pretreatment. The sample groups in Figure 5 were marked using colors according to pretreatment group belonging as indicated in the experimental section of the chapter.

PLS and biPLS modeling parameters are given in

Table 2. As the correlation coefficients and the RMSECV values improved significantly for the biPLS models calibration performance increased by deselecting disturbing spectral ranges. All three relative monosaccharide amounts after enzymatic hydrolysis could be predicted well. Further calibration details are given in PAPER 1.

Table 2. Modeling parameters as RMSECV, RMSEC, correlation coefficients R^2 and number of latent variables (LV) are presented for the PLS/biPLS prediction models using the full spectra and the selected spectral intervals, respectively.

	RMSECV (%)		RMSEC (%)		R^2		number of LVs	
	full spectra	biPLS	full spectra	biPLS	full spectra	biPLS	full spectra	biPLS
arabinose	3.12	1.39	1.07	0.99	0.82	0.96	7	6
xyllose	1.16	0.77	0.90	0.62	0.95	0.98	3	3
glucose	8.68	5.69	6.37	3.83	0.84	0.93	5	5

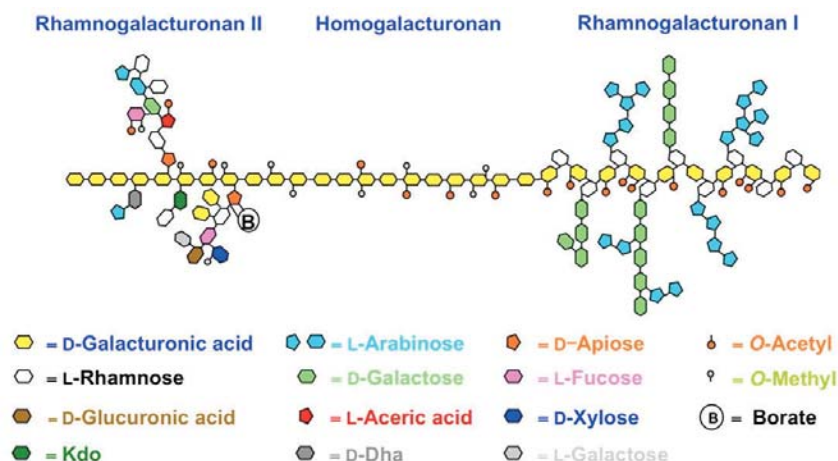


Figure 6. Main structural domains being present in pectin. The figure was retrieved from [87] and subsequently modified

5.1.5 Conclusion

As the results show NIR spectroscopy is a useful technique to monitor the pretreatment of recalcitrant biomass like destarched corn bran to elevate monosaccharide release for efficient production of biofuels. Using chemometric classification outliers could easily be detected. Secondly, iPLS and PLS have proven to be useful prediction tools in respect to the final monosaccharide yield, even without performing the enzymatic hydrolysis itself. This in particular represents a useful PAT tool to monitor the quality of incoming biomass batches at the beginning of the process chain and enables possibilities to react (e.g. change process parameters) when dissatisfying pretreatment has undergone.

5.2 FTIR and carbohydrate microarray analysis for prediction and characterization of enzymatic versus acidic extracted pectin

5.2.1 Background

Enzymes are used in many processes to release fermentable sugars for production of biofuel, but also for refinery of biomass to extract functional food ingredients such as pectin and other prebiotic oligosaccharides. The complex biomasses may, however, require a multitude of specific enzymes which enhance the biomass opening and help to specifically extract a biopolymer of interest. In the present study it was demonstrated that FTIR and carbohydrate microarray analysis represent useful PAT tools for industrial pectin extraction from lime peel, i.e. the

prediction of pectin yield and characterization of its quality. The methods also intend to enable continuous process monitoring and feedback control, identifying causes of process deviation and process failure. This opposes post-process product testing which is especially time-consuming and does not provide any continuous process control.

Pectin, mainly being produced from citrus peel, is a valuable resource used as gelling agent in jams and jellies, in medicines, sweets, as a stabilizer in fruit juices and milk drinks, and as a source of dietary fiber [83, 84]. Its efficient industrial production relies on acidic or enzymatic extraction methods which insist on optimization to increase the pectin yield and to obtain desired structural properties. Pectin is a complex biopolymer which is commonly understood to be composed by three main structural domains, namely Homogalacturonan I (HGI), Rhamnogalacturonan I (RGI) and Rhamnogalacturonan II (RGII). While HGI regions are generally described as linear backbone regions of α -(1–4)-linked D-galacturonic acid molecules, RGI and RGII regions indicate backbones with high degree of side-chains containing neutral sugars as arabinose, galactose and xylose. An illustration of typical pectic structures is given in Figure 6. Comprehensive introduction on carbohydrate microarrays can be found in [85, 86].

The hypothesis of the particular study was that FTIR spectroscopy and carbohydrate microarray analysis are suitable techniques for monitoring the pectin yield during the extraction process and that they are furthermore able to characterize the

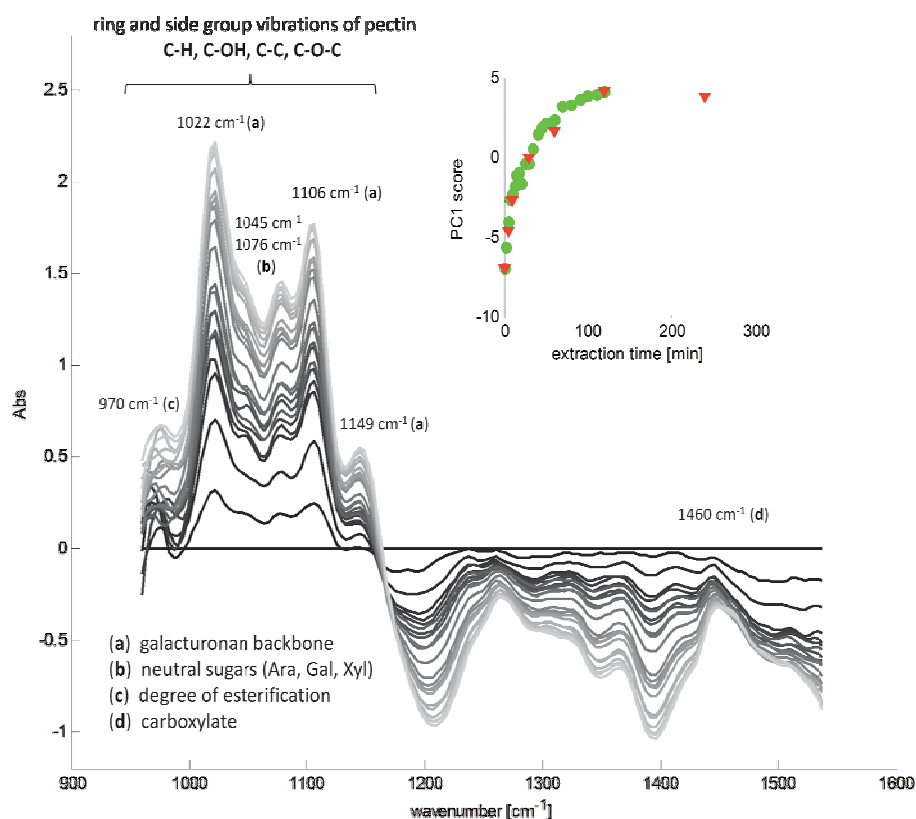


Figure 7. Extraction I spectra at different time points illustrated by changing grey-scale (from dark to light). Various structural pectin features are identified in the spectra. Upper right: Scores from a one component PCA plotted against elapsing extraction time for extraction I (green dots) and extraction II (red triangles). Extraction time could be optimized as the scores did not change after 120 min.

structural properties of the obtained pectins from different extraction media.

5.2.2 Experimental

Three time-resolved pectin extractions have been conducted. In each case samples were taken at specific time points and individual pectin yields were determined by propan-2-ol precipitation (reference analysis). Pectin extractions I and II were conducted using the commercial enzyme preparation Laminex C2K (Dupont, Leiden, Netherlands) to degrade the lime peel while extraction III was carried out using 42% HNO_3 . FTIR spectroscopy and carbohydrate microarray analysis have been utilized for separate data acquisition.

One aliquot of all samples was used for FTIR spectra acquisition in the spectral range between 950 and 1550 cm^{-1} . All spectra were acquired using a MilkoScan FT2 (FOSS ANALYTICAL, Hillerød, Denmark). The instrument consisted of

an interferometer which scanned the IR spectrum using a cuvette with a path length of 50 μm .

Carbohydrate microarray analysis was carried out at the Department of Plant and Environmental Sciences, University of Copenhagen. For a detailed procedure the reader is referred to PAPER 2. In brief, one aliquot of all samples was printed on a nitrocellulose membrane using a piezoelectric robot, ArrayjetSprint (Arrayjet, Roslin, UK) followed by immobilization and probing with a number of monoclonal antibodies (mAb) and carbohydrate binding modules (CBM). The microarrays were developed using a solution of 5-bromo-4-chloro-3-indolylphosphate and nitro blue tetrazolium. Finally the developed microarrays were quantified by scanning at 2400 dpi using CanoScan 8800F (Canon, Søborg, Denmark). The results were quantified using microarray software Array-Pro Analyzer 6.3 (Media Cybernetics, Rockville, USA) and a heatmap was printed.

Table 3. Figures of merit for PLS models which were established using pectin yields ranging from 0.07 to 0.34 g.

	FTIR		Microarray	
	model 1	model 2	model 3	model 4
R_{cal}	0.97	0.93	0.89	0.88
R_{pred}	0.99	0.98	0.91	0.73
RMSEP [g]	0.016	0.023	0.058	0.064
LVs	2	4	2	5

5.2.3 Chemometric analysis

The data obtained from both techniques were chemometrically analyzed using PCA and PLS. Prior to analysis spectra were pretreated using SNV and mean centering while the heatmap data from the carbohydrate microarray was pretreated using auto-scaling and mean centering.

PCA was used to monitor the FTIR spectral changes. Two PCA models were established a) for enzymatic extraction I and II; and b) for the acidic extraction III.

Secondly, PLS was used to establish quantitative models to predict the pectin yield using data from both techniques. To do so, two sets of models were established. For FTIR, model 1 and 2 were established on extraction I and extraction I+III samples, respectively. Both models were validated and evaluated by extraction II samples which served as prediction test set. Models 3 and 4 were established for the carbohydrate microarray data in a similar fashion which enabled the possibility to compare prediction performance of both techniques. The numbers of latent variables were determined by leave-one-out cross validation.

Additionally, FTIR spectra and carbohydrate microarray binding patterns of acidic versus enzymatic extracted pectin were analyzed towards structural properties using spectral PCA loadings and bi-plots derived from PLS.

5.2.4 Results

The FTIR spectra for extraction I are shown as difference spectra in Figure 7 while extraction II and III difference spectra are given in PAPER 2. Monosaccharide standard spectra for xylose, arabinose and galactose are also given in PAPER 2 for qualitative interpretation of the spectra. Main spectral features due to Homogalacturonan backbone, degree of esterification and neutral sugar content could be identified as indicated in Figure 7. While spectra for enzymatic extracted pectin (extraction I and II) indicated similar characteristics spectra from acidic extracted pectin showed lower bands at 1045 and 1076 cm^{-1}

indicating lower amounts of neutral sugars released from the lime peel (see PAPER 2).

Additionally, FTIR measurement indicated lower degree of esterification for acidic extracted pectin. However, at this point FTIR could not determine the neutral sugars to be part of RGI or RGII pectic regions or to be free in solution. In fact, the carbohydrate microarray results (Figure 8) suggested lower abundances of RGI regions and neutral sugar oligosaccharide like arabinan and galactan for enzymatic extracted pectin. This lead to the conclusion that the high amounts of neutral sugars, being enzymatically extracted from lime peel, were free in solution and therefore measured by FTIR while they were not immobilized on the highly specific microarrays. When looking at the heatmap in Figure 8 it could also be seen that antibody responses regarding RGI pectic regions and galactan decreased throughout the enzymatic extraction. This indicates the presence of side-chain cleaving enzymes in the commercially preparation Laminex C2K which led to degradation of RGI regions during the extraction period. Further elaboration is given in PAPER 2.

Using the FTIR data, one component PCA models for a) the enzymatic extraction I and II (Figure 7); and b) the acidic extraction III (PAPER 2) indicated optimal extraction times at 120 min in both cases since spectra did not change significantly after that time point. This finding could be confirmed by pectin yield determination (reference analysis).

PLS prediction models were established using data from both techniques as described in the previous section. Results towards accuracy (RMSEP), precision (correlation coefficients) and number of latent variables are given in Table 3.

PLS modeling of the carbohydrate microarray data, i.e. the interpretation of bi-plots, gave a comprehensive understanding of the antibody binding pattern being related to pectin release.

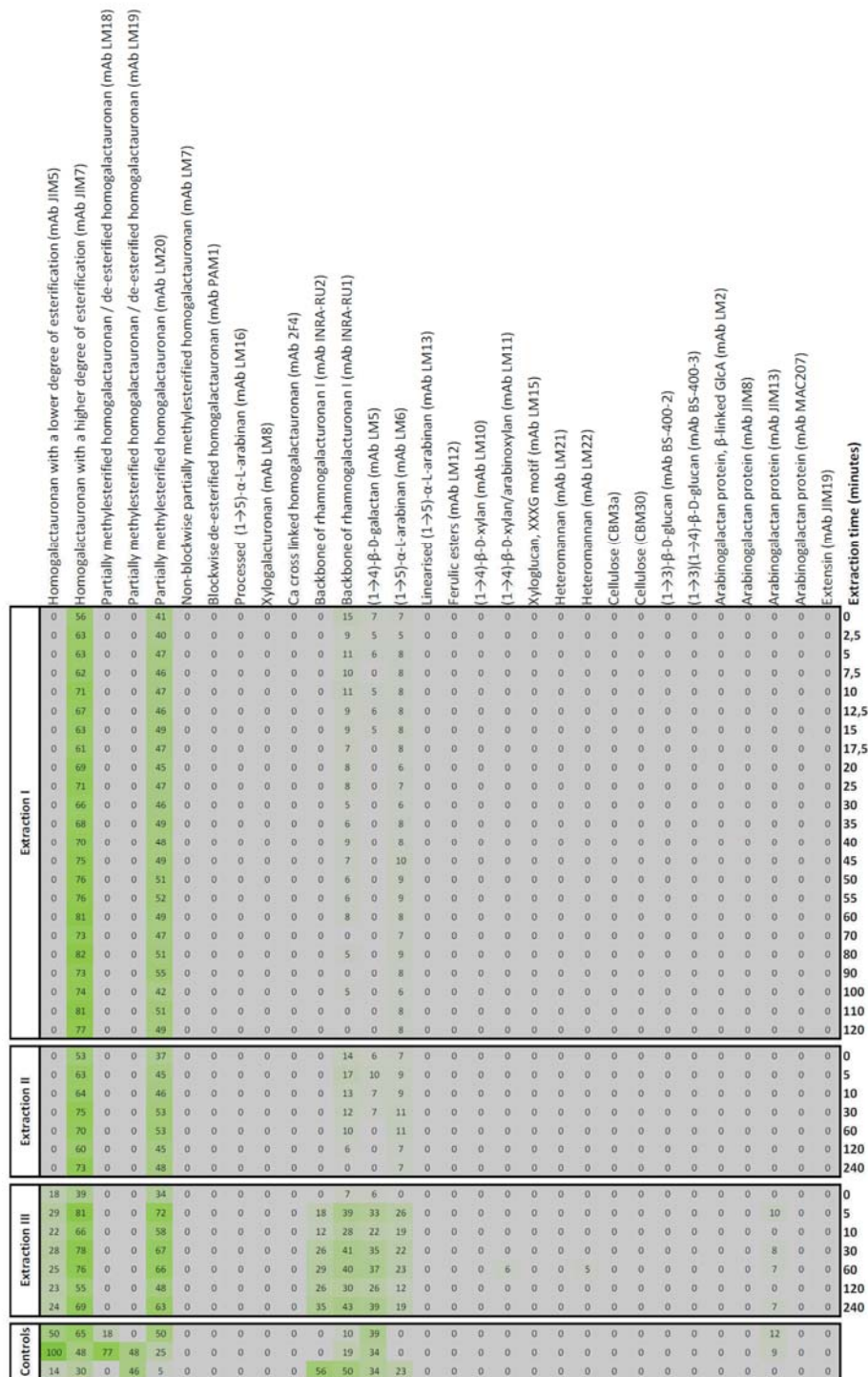


Figure 8. Screening of polysaccharides from lime peel extractions using microarray technology. Extracted pectin microarrays were probed with a panel of 30 cell wall glycan directed monoclonal antibodies and CBMs. Binding results were presented in a heatmap in which colour intensity was correlated to mean spot signal values. The highest mean spot signal value was set to 100 and the rest of values were normalized accordingly. At the top of the heatmap the used probes can be seen between brackets preceded by a short explanation of the epitope recognized by the particular probe. At the right part of the heatmap extraction times are shown. The sub-heatmap at the bottom, “controls”, includes some defined pectic samples with various degrees of methylation. At the top, a commercially produced pectin with degree of esterification (DE) 81%, in the middle a pectin with DE 31% and at the bottom, rhamnogalacturonan I. (PAPER 2)

Those binding patterns were rather difficult to identify without multivariate methods since none of the 30 used antibodies or CBMs led to prediction ability itself. More information is given in PAPER 2.

When looking at the figures of merit of all four PLS models (Table 3) it can be concluded that models 2 and 4, which included the acidic extraction III samples, decreased in prediction performance for both techniques. However, yield prediction ability for carbohydrate microarray data did decrease more severe when including the acidic extraction III as it can be seen from the increasing RMSEP for model 4 and the decreasing correlation coefficients. Noticeably, the higher RMSEP (lower accuracy) did not only result from worse precision, as it can be seen when comparing the raw data in Figure 7 and Figure 8, but also from a bias induced by the prediction samples itself. Further elaboration, including calibration curves, is given in PAPER 2.

5.2.5 Conclusion

Two multivariate techniques, namely FTIR and carbohydrate microarray analysis showed predictive and descriptive abilities towards the extraction of pectin from lime peel. Both methods could predict the pectin yield well using PLS modeling. However, while FTIR predicted the pectin yield better in terms of accuracy and precision, the highly specific carbohydrate microarray analysis contributed with valuable information regarding the structural properties of the extracted pectin.

Thus, both techniques represent rapid (FTIR) and high-throughput capable (carbohydrate microarray) tools which can help to understand the extraction of pectin in different media. Both techniques are furthermore recommended to be implemented as industrial PAT solutions. Especially FTIR has shown to be highly suitable for online-monitoring of the pectin extraction process which could lead to optimization of parameters regarding process routine and product quality.

6. Temporal Evolution Profiling and chemometric multiway analysis for determination of enzyme activity

In the previous chapter 5 spectral fingerprints derived from FTIR and NIR have been used to characterize certain thermodynamic states of carbohydrate systems. A constant thermodynamic state was hereby directly linked to the acquired spectrum which also reflected qualitative and quantitative properties of its underlying chemical constituents. Two-way chemometric methods have been used to analyze the corresponding data sets.

However, in this chapter multilinear data sets are handled which are constructed in a way that they contain an additional measurement dimension, namely time. This new measurement dimension enables the possibility to observe and calibrate for changes of thermodynamic states. In this chapter we therefore deal with data structures which are defined by its dimensions of samples, wavenumbers and time points (as described in Figure 3). Those multiway data structures can be used to study kinetic phenomena, as determining the enzyme activity of one or several enzymes.

An interesting property of such data sets is that the underlying enzyme activity can be extracted independently from the interfering background as illustrated in Figure 9. This means that calibration models are still valid, even if a new sample indicates a different chemical matrix. The interfering background is simply removed by chemometric preprocessing as shown in Figure 9. To construct multiway data structures, mathematically denoted as tensors, time-resolved measurements are necessary. During the two studies, which are introduced in chapters 6.1 and 6.2, these measurements have been carried out by continuously monitoring FTIR spectra of various samples which differed by the fact that they contained different amounts of active enzyme during the time-course of the enzymatic reaction.

6.1 Measuring enzyme activity using FTIR and PARAFAC

6.1.1 Background

A key point to investigate enzyme related processes is the estimation of enzyme activity which is a fundamental measurement required for determining the amount of active enzyme.

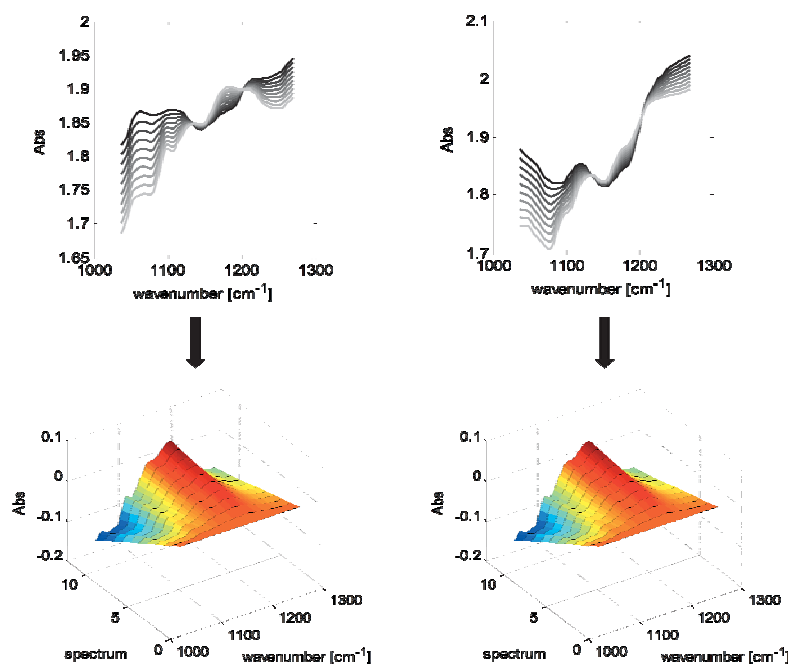


Figure 9. The figure shows simulated data. Top: A time-resolved change of spectral fingerprints is illustrated by changing grey-scale. Bottom: The observed kinetic phenomenon, e.g., an enzymatic reaction, can be extracted independently from the sample matrix as the background can be removed by simple data preprocessing. In this case, the initial spectrum was subtracted from all spectra of the series yielding similar evolution profiles in both cases.

Formally, enzyme activity is defined as the amount of active enzyme which will catalyze the transformation of a particular molar amount of substrate per time unit under optimal conditions. Its measurement is decisive for defining enzyme dosage, reaction time, substrate use and product yields in practical enzyme catalyzed reactions. A quantitative enzyme activity assay - optimally measured directly as the reaction rate defining the initial substrate consumption rate - is thereby also a fundamental measurement used for identifying a specific enzyme.

However, according to this it is difficult to define the enzyme activity of enzymes which can biocatalyze more than one substrate. Often the corresponding products are even able to serve as substrates themselves. A particular example is given by pectin lyase which catalyzes the depolymerization of homogalacturonan regions in pectin which can serve as substrate several times and therefore produce a distribution of products with different chemical and structural properties. Hence, the question arises of which unique product or substrate to use for the quantification of the enzyme activity although the catalytic reaction is best described by all of them?

It appears to be equally difficult to define the overall enzyme activity of cellolytic enzyme blends since they contain cellulases and several other accompanying enzymes which actively help to break down recalcitrant biomasses efficiently. The IUPAC defined a standardized method for cellulase related enzyme activity determination using Filter Paper units [88]. This method employs standardized filter paper as a substrate to be degraded by the enzymes while conventional reducing sugar assays [88, 89] are used to monitor the enzymatic degradation by screening the formation of reducing sugars during the time-course of the catalyzed reaction. However, those assays are usually referenced to only one standard reducing sugar as, e.g., glucose and are therefore strongly depend on the used reference monosaccharide itself. Hence, it is questionable how well those filter paper units can describe the actual overall enzyme activity as it might vary in dependence of the present biomass (substrate) and the used reducing sugar standard. Additionally, reducing sugar assays as described in [88, 89] are very cumbersome, time-consuming and often employ toxic and environment unfriendly chemicals.

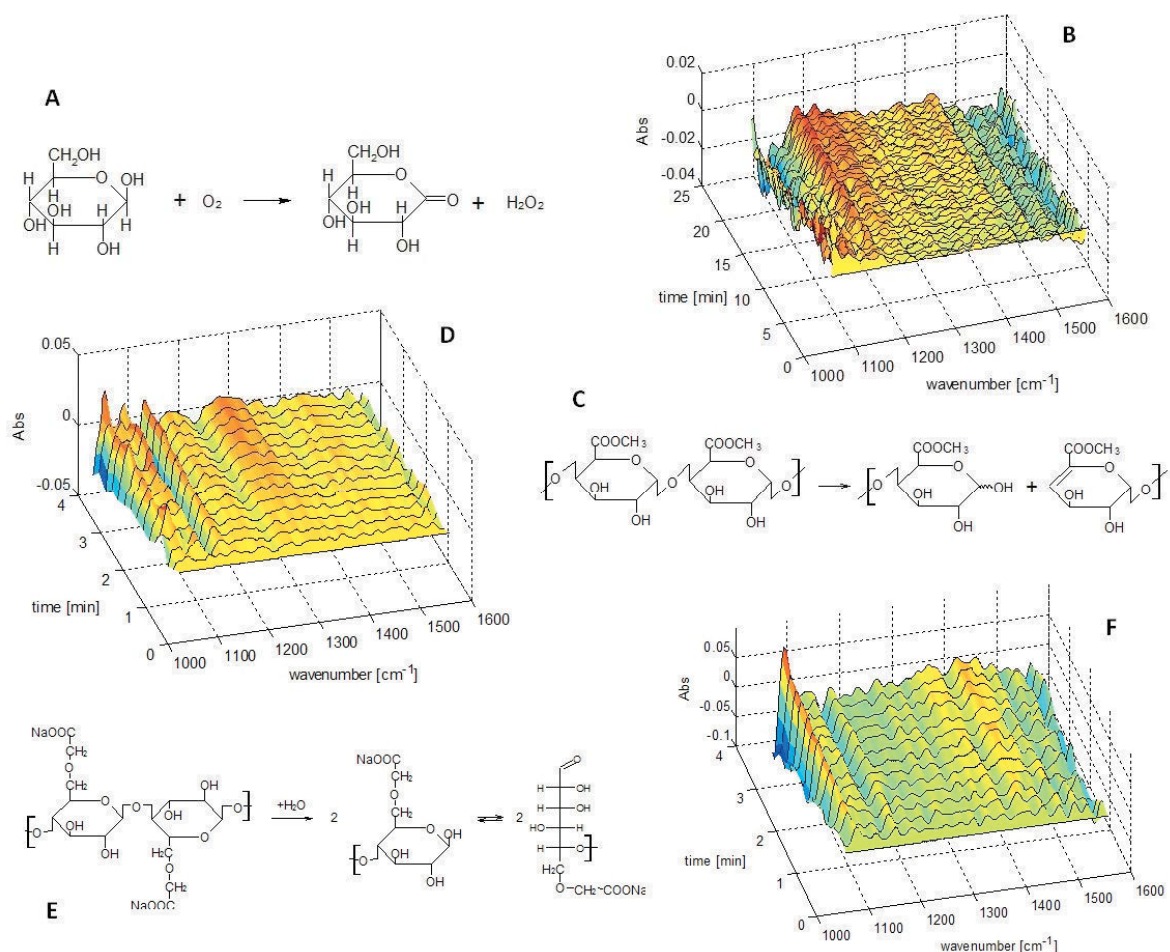


Figure 10. Each enzymatic reaction results in a specific spectral evolution profile, representing both, substrate depletion and product evolution. The enzymes used were glucose oxidase (A, B), pectin lyase (C, D) and Celluclast 1.5L (E, F).

During this study (PAPER 3) the hypothesis is tested, that FTIR and chemometric multiway analysis can be used as an universally applicable approach for rapid assessment of enzyme activity without using any external standards, even on genuine, complex substrates as biopolymers (which may not be chromogenic). The methodology is demonstrated by spectral evolution profiling of Fourier Transform Infrared (FTIR) spectral fingerprints using parallel factor analysis (PARAFAC) for pectin lyase, glucose oxidase, and a cellulase preparation.

6.1.2 Experimental

All spectral evolution profiles were obtained using a MilkoScan™ FT2 (FOSS ANALYTICAL, Hillerød, Denmark) and reference enzyme activity determinations were carried out by conventional enzymatic assays as described in PAPER 3. Spectra were acquired in the range between 1000 cm^{-1} and 1600 cm^{-1} while the temperature of the

cuvette (path length 50 μm) was equilibrated at 42 °C during all measurements. Sets of evolution profiles were acquired using equal substrate concentrations and different dosages of enzyme. Depending on the nature of the enzymatic reaction, measurements have been carried out in flow-back mode where the reaction mixture was continuously led back to the reaction container to ensure access of gases as oxygen which was necessary for the reaction of glucose oxidase. The other enzymatic reactions have been pumped into the cuvette only once for continuous measurements (pectin lyase and Celluclast 1.5L). In this case the reaction mixture stayed inside the cuvette during the whole acquisition period. Spectra for each evolution profile were measured consecutively using time steps of 16.6 s for pectin lyase and Celluclast 1.5L and 31.0 s for glucose oxidase, respectively.

6.1.3 Chemometric analysis

The acquired evolution profiles were transformed into tri-linear data structures using three-dimensional arrays as illustrated in Figure 11.

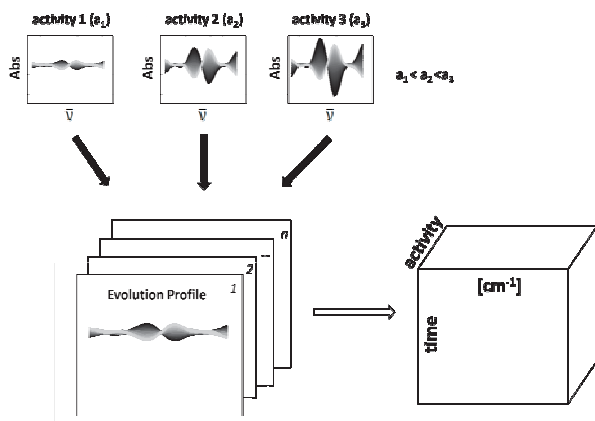


Figure 11. Conversion from two-way to a three-way data structure. All data matrices containing time-resolved spectra for a certain enzyme activity are stacked behind each other to form a tensor. The cubic data structure fulfills trilinearity since all evolution profiles are recorded using the same consecutive time steps.

The three-dimensional arrays for all three enzymatic reactions contained evolution profiles where each enzyme activity was represented by triplicate measurements. Some outliers were taken out during the calibration: The final tensors indicated the following dimensions:

Glucose oxidase: 29 samples x 40 time points x 121 spectral variables; Pectin lyase: 32 samples x 15 time points x 128 spectral variables; Celluclast 1.5L: 33 samples x 15 time points x 148 spectral variables.

Wavenumber ranges have been selected to optimize the calibration performance. Prior to tensor formation spectra were normalized for each sample slab of the tensor using SNV (PAPER 3).

Parallel Factor Analysis was used to resolve the common underlying spectral substrate and product evolutions by decomposing the tensor into spectral and kinetic loadings using an alternating least squares algorithm [60]. Decomposition of the trilinear data structure therefore could be understood as if common kinetic phenomena were fitted to all samples simultaneously (see PAPER 3).

6.1.4 Results

Individual evolution profiles for the three enzymatic reactions of glucose oxidase, pectin lyase and Celluclast 1.5L are shown in Figure 10 as difference spectra.

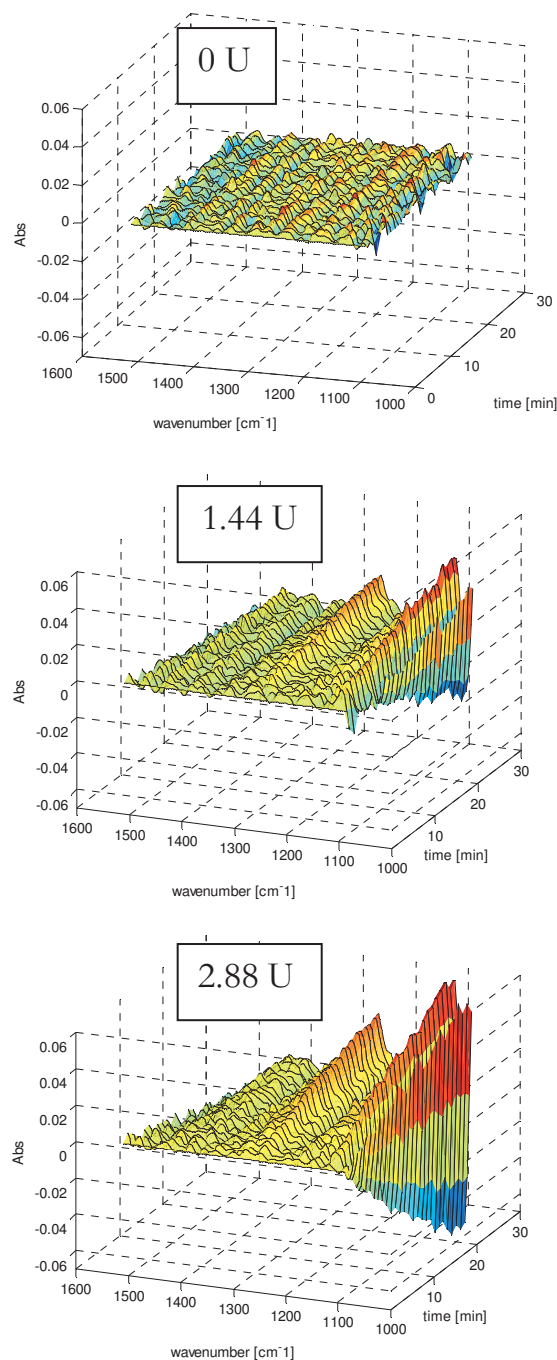


Figure 12. Spectral evolution profiles measured by FTIR containing different amounts of pectin lyase bio-catalyzing pectin degradation

Table 4. Calibration parameters for the three modeled enzymatic reactions

	pectin lyase	glucose oxidase	Celluclast1.5L
R ² substrate calibration	0.998	/	0.99
R ² product calibration	0.995	0.97	0.98
Time of spectral evolution	4.2 min	20.7 min	4.2 min
Number of spectra in each Evolution Profile	15	40	15
Calibration range (per ml Substrate)	0-200 mU/ml	0-6 U/ml	0-80 mU/ml
LOD*	9 mU/ml	277 mU/ml	3.79 mU/ml

* Limit of Detection (LOD) only valid for used time of spectral evolution. LOD decreased for extended observation time.

The figure shows that each enzymatic reaction results in its own specific temporal evolution due to the complex IR fingerprints of substrates and products. In addition, the intensity of the spectral evolution depended on the used enzyme amount. Several evolution profiles indicating different amounts of pectin lyase which bio-catalyzed the degradation of pectin are shown in Figure 12. For enzyme activity calibrations the scores retrieved from the three PARAFAC decompositions were plotted against the determined reference values. Since the evolution profiles monitored both, substrate depletion and product evolution simultaneously, the PARAFAC decomposition resulted in two calibrations for each enzymatic reaction. Correlation coefficients of the calibrations are given in Table 4. However, the glucose oxidase calibration showed poorer performance than the calibrations of the other two enzymatic reactions due to a lower signal-to-noise ratio in the spectra. It therefore resulted in only one calibration. Calibration relevant parameters as limit of detection (LOD), spectral calibration range, time of spectral evolution and number of spectra in the individual evolution profiles are given in Table 4.

In addition, the calibration performance was investigated in dependence of observation time used in the individual evolution profiles. Further information on this and an application example of the method to investigate the thermal stability of pectin lyase is given in PAPER 3.

6.1.5 Conclusion

The method presents a universal and straight forward approach to quantify enzyme activity without using any external standards. Three enzyme activities, namely glucose oxidase, pectin lyase and the overall activity of a cellolytic enzyme blend, could be determined using spectral evolution profiling and PARAFAC analysis. The suggested approach enables possibilities to monitor and create a better understanding of

complex biomass degrading enzymatic reactions and therefore offers attractive alternatives to the cumbersome reducing sugar analysis. Although FTIR serves as a very versatile and non-invasive technique to monitor both, substrate depletion and product evolution, its major drawback can be found in the low sensitivity as detection limits are relatively high as indicated in Table 4.

6.2 Simultaneous measurement of two enzyme activities: a comparative evaluation of PARAFAC, TUCKER3 and N-PLS modeling

6.2.1 Background

Chapter 6.1 dealt with enzyme activity determinations of bio-catalytic systems which contained only one enzyme or a blend of enzymes whose amounts relative to each other were constant in all reactions as if they would represent a single “overall enzyme activity”.

This chapter focuses on extending the concept to determine enzyme activity of two co-acting enzymes simultaneously, i.e. pectin lyase and pectin methyl esterase acting on the same substrate, i.e. pectin from citrus peel. Both enzymatic reactions are given in Figure 14. While assuming that both enzymes do not interact with each other in an initial phase of the reaction, a superimposed nature of the individual evolution profiles can be expected as illustrated in Figure 13 (simulated data). In addition, substrate depletion during spectral acquisition should be avoided by choosing high initial concentrations. When mixing different amounts of the enzymes the overall absorption pattern of the spectral evolution is therefore expected to change indicating simple superimposition (Figure 13).

Throughout this study the hypothesis shall be tested if mid-infrared spectral evolution profiling and chemometric multiway analysis can determine two enzyme activities simultaneously. By doing so, three multiway methods shall be compared

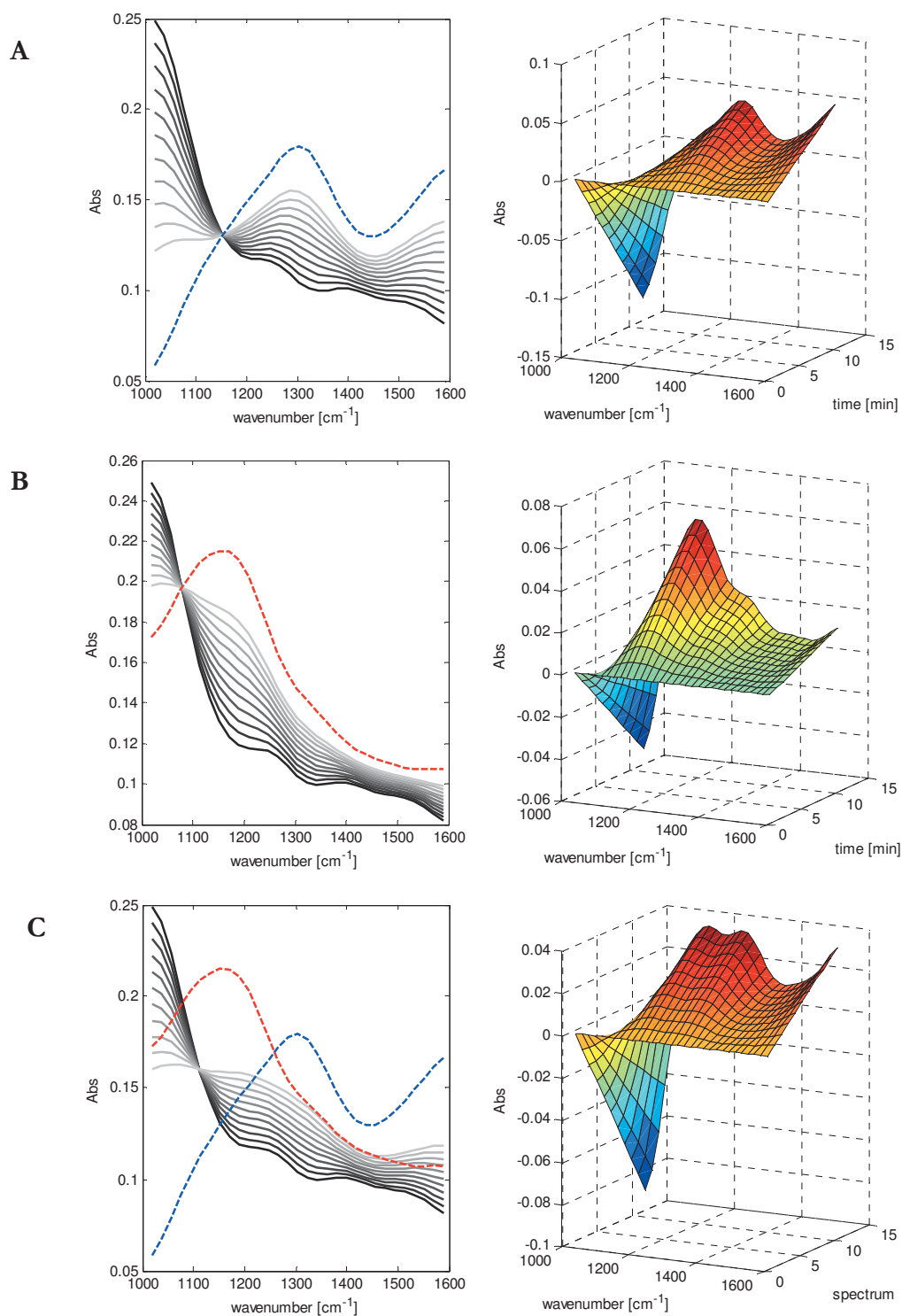


Figure 13. Simulated data for three enzymatic reactions. A) Enzyme 1 catalyzes the conversion of the substrate to product P1; B) Enzyme 2 converts the substrate to product P2; C) Enzyme1+Enzyme2 (half dose enzyme1 + half dose enzyme2) convert the Substrate to both Products (P1 and P2) at the same time. The spectral changes from Substrate fingerprint to product fingerprint are illustrated on the left side, whereas difference spectra using surf plots are shown on the right side. Individual final product fingerprints are given by red and blue dashed lines, respectively.

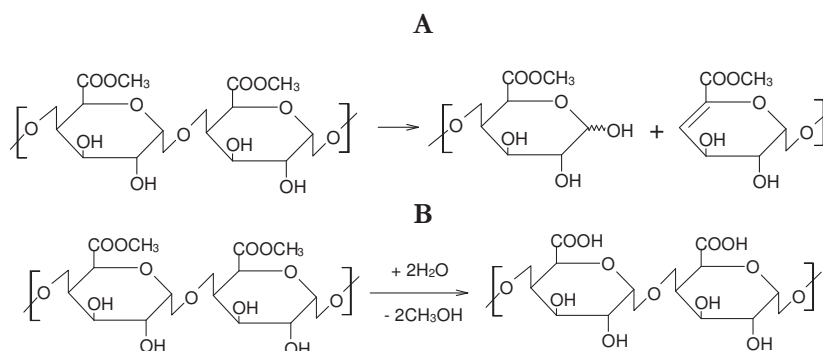


Figure 14. A) Enzymatic reaction of pectin lyase; B) enzymatic reaction of pectin methyl esterase; although both enzymes act on the same substrate, namely pectin, they cleave the polymer at different specific sites.

towards modeling performance, namely PARAFAC, TUCKER3 and NPLS.

6.2.2 Experimental

All spectral evolution profiles were acquired between 1000 cm^{-1} and 1600 cm^{-1} using FTIR in a similar fashion as described in chapter 6.1. To establish calibration models for both enzymes, PL and PME, an appropriate experimental design was necessary. It should take into account that e.g. samples with one particular PL activity can be superimposed by PME activities which span from zero to a certain maximum value and vice versa. To establish robust calibration models all possible degrees of superimposition had therefore to be considered as shown in Figure 15.

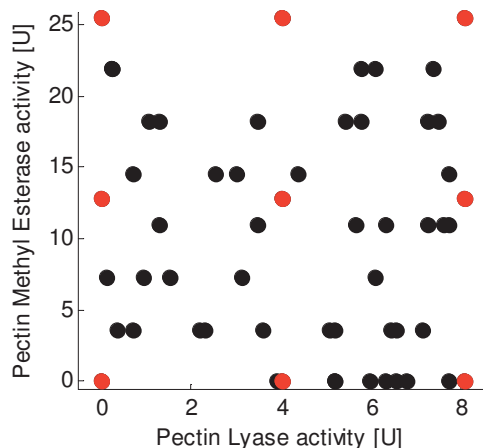


Figure 15. Experimental Design: Red samples define calibration ranges, while black samples are randomly placed to create a normally distributed sampling scenario.

Thus, a total number of 56 evolution profiles were acquired resembling diverse scenarios of superimposed PL and PME activities being added to 20 ml of 1g/l pectin solution. The spectra were

acquired consecutively using time steps of 16.6 seconds which resulted in eleven minutes observation time for one evolution profile (40 spectra).

The molar enzyme activities of PL and PME for reference analysis were determined by conventional assays as described in PAPER 4.

6.2.3 Chemometric analysis

The acquired 56 evolution profiles were transformed into trilinear data structures as shown in Figure 11. 37 samples were used for calibration modeling and the 19 remaining samples served as a prediction test set for validation. Prior to any multiway modeling all spectra were normalized using SNV for each individual sample slab of the three-way array (PAPER 4). The final tensor contained data in the format 56 samples \times 40 time points \times 128 spectral variables. The three-way array was consecutively double centered using multiway centering towards mode 1 and mode 2 [64].

PARAFAC, TUCKER3 and NPLS analysis were carried out using PLS toolbox 6.0.1 (Eigenvector Research Inc., WA, USA). The mean prediction error (RMSEP) and the pooled standard deviation (s_p) were calculated as described PAPER 4 and represent measures for accuracy and precision.

6.2.4 Results

The calibration results for PARAFAC, TUCKER3 and NPLS modeling are given in Table 5 and Figure 16. Individual evolution profiles containing different amounts of PL and PME are given in PAPER 4.

Using leave-one-out cross-validation the number of factors for the PARAFAC model was determined to be 2 while the individual ranks for TUCKER3 modeling towards the modes 1 (sample mode), 2 (kinetic mode) and 3 (spectral

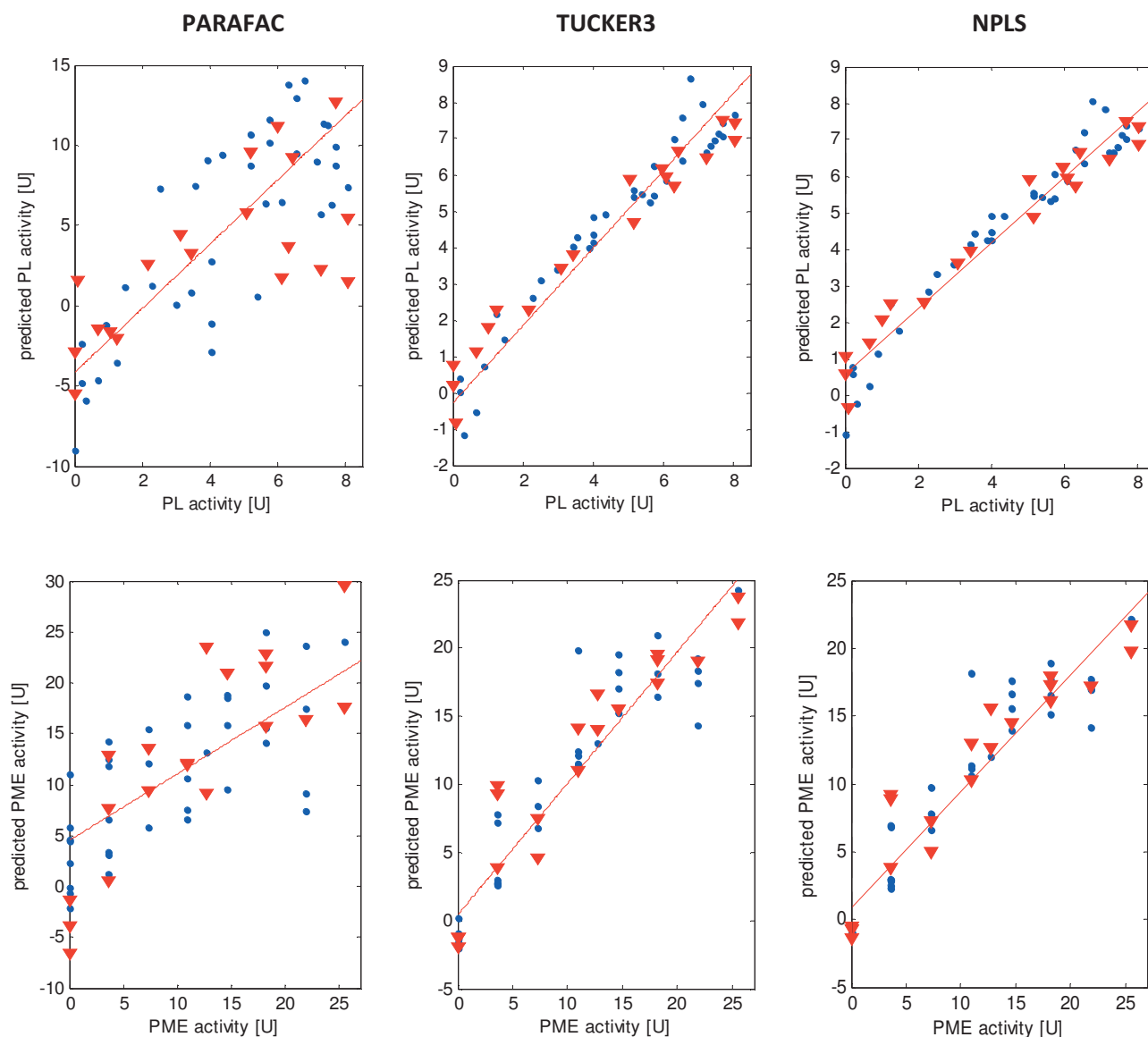


Figure 16. PL and PME calibrations derived from PARAFAC, TUCKER3 and NPLS models. Top: PME, bottom: PL; The blue points indicate leave-one-out cross validated calibration samples while red triangles indicate predictions for the test set samples.

mode) were found to be 2, 1 and 2, respectively. The kinetic mode loadings of the TUCKER3 and NPLS models indicated that the rank in this mode was indeed 1 since both enzymatic reactions evolved identically, i.e. the initial rates changed in a similar fashion. The reduction of the rank in the kinetic mode (mode 2) was therefore assumed to be beneficial for calibration modeling as PARAFAC had problems to model the two enzymatic systems simultaneously (Figure 16). Further elaboration and loadings from all modes are presented in PAPER 4. Individual NPLS models for PL and PME indicated only one significant factor (one underlying evolution profile) for each activity calibration because leave-

one-out cross validation indicated no prediction improvement for NPLS models when using more than one factor, respectively.

Surprisingly, TUCKER3 calibration performance was comparably good with respect to the result retrieved from the supervised regression method NPLS. This study therefore indicated a useful application for TUCKER3, especially in contrast to PARAFAC.

In addition the high calibration accuracy of the unsupervised result from TUCKER3 indicated that no significant disturbing interaction between the enzymes was present since it would have led to decrease in calibration performance.

Table 5. Regression coefficients for calibration and prediction set samples are given together with pooled standard deviation s_p and root-mean-square-error of prediction RMSEP

	PARAFAC		TUCKER3		N-PLS	
	PL	PME	PL	PME	PL	PME
R^2_{cal}	0.67	0.52	0.92	0.88	0.95	0.89
R^2_{pred}	0.48	0.70	0.96	0.89	0.96	0.89
S_p	1.10 U	1.62 U	0.26 U	1.19 U	0.24 U	1.10 U
RMSEP	3.60 U	5.31 U	0.62 U	2.72 U	0.70 U	2.85 U

6.2.5 Conclusion

FTIR spectroscopy was employed to follow the bioconversion of two enzymatic reactions time-resolved. All three multiway methods, namely PARAFAC, TUCKER3 and NPLS, could predict the enzyme activities of the two co-acting enzymes, PL and PME, simultaneously. When comparing the three multiway models mainly TUCKER3 surprised with its outstanding performance. As both, accuracy and precision were comparable to the regression results from N-PLS, although resulting from a non-supervised analysis; it shows clearly that TUCKER3 has great potential when analyzing kinetic data, especially when PARAFAC fails to resolve the underlying phenomena as it could be seen during this study. Since many enzymatic reactions still rely on monitoring only one substrate or product the here presented approach enables much better observation of the whole reaction pattern. This is particularly important when enzyme kinetics is to be investigated on genuine substrates where two (or more) enzymes act concerted on the same substrate. At present those reactions can hardly be monitored by univariate measurement principles as they result in very complex reaction patterns. Time-resolved multivariate analysis and FTIR spectroscopy therefore inaugurate new possibilities to monitor such reactions to gain a better understanding of their complex underlying nature.

7. Conclusion and Perspectives

This thesis dealt with descriptive and predictive assessment of enzyme activity and enzyme related processes in biorefinery using FTIR and chemometrics. Within the project two major research fields were dealt with, i.e. the establishment of PAT tools for prediction of relevant parameters in biorefinery processes and the establishment of universally applicable enzyme

activity assessment methods. The developed methodologies complied with concepts like “cradle-to-cradle” [6], meaning that no resources are meant to be wasted and toxic chemicals should be avoided. The major results are summarized in the following:

1. Hydrolysis efficiency of destarched corn bran after various pretreatment can be predicted by non-invasive NIR spectroscopy and chemometric modeling prior to enzymatic conversion. Prediction models towards the monosaccharide release enable a useful PAT to monitor the quality of incoming biomass batches and to evaluate the effect of pretreatment while outliers are easily detected. (PAPER 1)
2. FTIR, carbohydrate microarray analysis and chemometric data analysis are suitable techniques for the qualitative and quantitative analysis of time-resolved pectin extraction. FTIR represents a useful PAT tool to monitor the extraction process online using predictive models which also leads to optimization of extraction time. Carbohydrate microarray analysis gives strong support to determine the obtained pectin structure, i.e. to determine the degree of esterification and neutral sugar side-chain abundance. (PAPER 2)
3. FTIR and chemometric multiway analysis, i.e. PARAFAC, give better and more objective determination abilities towards enzyme activity. The method is universally applicable to a wide range of enzymes and does not require use of any external chemicals. (PAPER 3)
4. FTIR and chemometric multiway chemometrics, i.e. PARAFAC, TUCKER3 and NPLS, can determine the enzyme activities of the co-acting enzymes pectin

lyase and pectin methyl simultaneously.
(PAPER 4)

The results of the project were presented in chapter 5 and chapter 6. The division of the chapters was hereby based on application of two-way versus multiway chemometric methods. On one hand, two-way chemometric methods primarily enabled opportunities for analysis of thermodynamic states, while on the other hand multiway methods were employed to establish models towards time-resolved changes of thermodynamic states.

While chemometric two-way methods, in combination with rapid and non-invasive techniques which required almost no sample preparation, showed great benefit when characterizing (bio)-chemical systems the study indicated some outstanding advantages of multiway analysis. Enzyme activity was hereby found to form specific spectral evolution profiles, being composed of spectral substrate and product evolutions, whose intensity varied in dependence of the used enzyme amount. Enzyme activities could therefore be quantified and identified according to the shape of their specific evolutions measured by FTIR.

The inherent uniqueness property of e.g. unsupervised PARAFAC models paves the road for establishment of enzyme activity calibrations, even without reference analysis which is in particular interesting as relative enzyme activities are desired when characterizing an enzyme in terms of its thermal stability.

Secondly, the second order advantage makes models derived from multiway methods robust against unknown interferences in future prediction samples. This means that enzyme activity calibrations can be applied to different sample matrices, even containing possible additional interfering substrates. However, this was not investigated throughout this project and, theoretically, only holds when kinetic and fingerprint loadings of the interfering spectral evolution differ from the loadings of the calibrated enzymatic spectral evolution.

Outlining these striking advantages, hence, promises further application areas of the proposed multiway methods regarding enzyme activity determination. In terms of future perspectives, studies could concern investigation of more than two enzymes acting simultaneously on common substrates to evaluate the limits of

de-convoluting underlying evolution profiles of the individual enzymatic reactions. In addition, the interaction between enzymes can be investigated as its appearance would be expected to disturb calibration performance of the unsupervised multiway models. The method also enables descriptive possibilities to monitor the quality of enzyme production batches in terms of side-activity abundance as the specific spectral evolution profiles indicate additional spectral evolution easily. The major advantage is hereby, that the method utilizes the fingerprinting technique FTIR which captures information concerning literally all changes in (bio)-chemical systems making it universally applicable, even considering recalcitrant sample matrices as, e.g., milk and juice.

In terms of biorefinery, the proposed methodologies offer outstanding opportunities to screen enzymatic reactions, also in a qualitative sense, and could therefore have significant impact on analyzing bio-catalytic processes as they have been cumbersome to understand for decades because of their obtrusive complexity.

8. References

- [1] R. Rastall, Novel enzyme technology for food applications, Woodhead Publishing Ltd, 2008.
- [2] J.T. Wood, Das Entkälken und Beizen der Felle und Häute, Springer, 1914, p. 104.
- [3] K. Chen, F.H. Arnold, Proceedings of the National Academy of Sciences, 90 (1993) 5618.
- [4] W.P. Stemmer, Proceedings of the National Academy of Sciences, 91 (1994) 10747.
- [5] R.A. Sheldon, I. Arends, U. Hanefeld, Green chemistry and catalysis, Wiley-VCH, 2007.
- [6] W. McDonough, M. Braungart, Cradle to cradle : remaking the way we make things, North Point Press, New York, 2002.
- [7] A.J. Straathof, S. Panke, A. Schmid, Current opinion in biotechnology, 13 (2002) 548.
- [8] K. Buchholz, V. Kasche, U.T. Bornscheuer, Biocatalysts and enzyme technology, Wiley-Blackwell, 2012.
- [9] H. Ohara, Applied microbiology and biotechnology, 62 (2003) 474.
- [10] M. McCann, K. Roberts, The cytoskeletal basis of plant growth and form (1991) 109.
- [11] S. Pérez, M. Rodríguez-Carvajal, T. Doco, Biochimie, 85 (2003) 109.

- [12] H. Ohara, Preprints of zero emission symposium, 2000, p. 81.
- [13] F. Guillon, M. Champ, J.-F. Thibault, Functional foods 315.
- [14] N. Colthup, L. Daly, S. Wiberley, (1990).
- [15] B.H. Stuart, Infrared spectroscopy: fundamentals and applications, Wiley, 2004.
- [16] M. Paradkar, S. Sakhamuri, J. Irudayaraj, Journal of food science, 67 (2002) 2009.
- [17] M. Blanco, I. Villarroya, TrAC Trends in Analytical Chemistry, 21 (2002) 240.
- [18] J. Lammertyn, A. Peirs, J. De Baerdemaeker, B. Nicolai, Postharvest biology and technology, 18 (2000) 121.
- [19] H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, Near-infrared spectroscopy: principles, instruments, applications, Wiley-Vch, 2008.
- [20] J. Russell, A. Fraser, Clay Mineralogy: Spectroscopic and chemical determinative methods, Springer, 1994, p. 11.
- [21] P. Griffiths, J.A. De Haseth, Fourier transform infrared spectrometry, Wiley-Interscience, 2007.
- [22] S.K. Andersen, P.W. Hansen, H.V. Andersen, Handbook of Vibrational Spectroscopy (2002).
- [23] S. Sivakesava, J. Irudayaraj, A. Demirci, Journal of Industrial Microbiology and Biotechnology, 26 (2001) 185.
- [24] M. McCann, L. Chen, K. Roberts, E. Kemsley, C. Sene, N. Carpita, N. Stacey, R. Wilson, Physiologia Plantarum, 100 (1997) 729.
- [25] O. Faix, Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood, 45 (1991) 21.
- [26] N. Wellner, M. Kačuráková, A. Malovíková, R.H. Wilson, P.S. Belton, Carbohydrate research, 308 (1998) 123.
- [27] M. Kacurakova, P. Capek, V. Sasinkova, N. Wellner, A. Ebringerova, Carbohydrate Polymers, 43 (2000) 195.
- [28] K.J. Astrom, B. Wittenmark, Computer-Controlled systems: theory and design, Prentice hall, 1999.
- [29] T.W. Anderson, An introduction to multivariate statistical analysis, Wiley New York, 1958.
- [30] R. Bro, S. De Jong, Journal of chemometrics, 11 (1997) 393.
- [31] S. Wold, K. Esbensen, P. Geladi, (Vol. 1), Elsevier, 1987, p. 37.
- [32] H. Hotelling, 1933, p. 498.
- [33] B.M. Wise, N.B. Gallagher, (Vol. 6), Elsevier, 1996, p. 329.
- [34] M.A. Hearst, S. Dumais, E. Osman, J. Platt, B. Scholkopf, Intelligent Systems and their Applications, IEEE, 13 (1998) 18.
- [35] C.J. Burges, Data mining and knowledge discovery, 2 (1998) 121.
- [36] D.W. Patterson, Artificial neural networks: theory and applications, Prentice Hall PTR, 1998.
- [37] A.K. Jain, J. Mao, K.M. Mohiuddin, Computer, 29 (1996) 31.
- [38] K.S. Booksh, B.R. Kowalski, Analytical Chemistry, 66 (1994) 782A.
- [39] K.R. Beebe, B.R. Kowalski, Analytical Chemistry, 59 (1987) 1007A.
- [40] B.R. Kowalski, Analytical Chemistry, 52 (1980) 112R.
- [41] S. Wold, M. Sjöström, L. Eriksson, Chemometrics and intelligent laboratory systems, 58 (2001) 109.
- [42] E. Sanchez, B.R. Kowalski, Journal of chemometrics, 2 (1988) 247.
- [43] E. Sanchez, B.R. Kowalski, Journal of chemometrics, 2 (1988) 265.
- [44] H.A. Kiers, Journal of chemometrics, 14 (2000) 105.
- [45] L.R. Tucker, Contributions to mathematical psychology (1964) 109.
- [46] L.R. Tucker, Psychometrika, 31 (1966) 279.
- [47] R.A. Harshman, (1970).
- [48] A. Smilde, R. Bro, P. Geladi, Multi-way analysis: applications in the chemical sciences, Wiley, 2005.
- [49] R. Bro, J.J. Workman JR, P.R. Mobley, B.R. Kowalski, Applied Spectroscopy Reviews, 32 (1997) 237.
- [50] R. Bro, Critical reviews in analytical chemistry, 36 (2006) 279.
- [51] C.M. Andersen, R. Bro, Journal of chemometrics, 17 (2003) 200.
- [52] L. Moberg, G. Robertsson, B. Karlberg, Talanta, 54 (2001) 161.
- [53] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, Journal of Chromatography A, 1027 (2004) 269.
- [54] B.J. Prazen, K.J. Johnson, A. Weber, R.E. Synovec, Analytical Chemistry, 73 (2001) 5677.
- [55] M. Pistonesi, M.E. Centurión, B.S.F. Band, P.C. Damiani, A.C. Olivieri, Journal of pharmaceutical and biomedical analysis, 36 (2004) 541.

- [56] W. Windig, B. Antalek, *Chemometrics and intelligent laboratory systems*, 37 (1997) 241.
- [57] H.T. Pedersen, R. Bro, S.B. Engelsen, *Journal of Magnetic Resonance*, 157 (2002) 141.
- [58] P. Geladi, P. Åberg, *Journal of near infrared spectroscopy* (2001).
- [59] A.M. Antunes, M. Ferreira, P.L. Volpe, *Journal of chemometrics*, 16 (2002) 111.
- [60] R. Bro, *Chemometrics and intelligent laboratory systems*, 38 (1997) 149.
- [61] Y. Wang, O.S. Borgen, B.R. Kowalski, M. Gu, F. Turecek, *Journal of chemometrics*, 7 (1993) 117.
- [62] R. Bro, (1996).
- [63] C.A. Andersson, R. Bro, *Chemometrics and intelligent laboratory systems*, 42 (1998) 93.
- [64] R. Bro, A.K. Smilde, *Journal of chemometrics*, 17 (2003) 16.
- [65] R.A. van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, *BMC genomics*, 7 (2006) 142.
- [66] R. Barnes, M. Dhanoa, S.J. Lister, *Applied spectroscopy*, 43 (1989) 772.
- [67] S. Jensen, H. Martens, *Cereal Chem*, 60 (1983) 172.
- [68] P. Geladi, D. MacDougall, H. Martens, *Applied spectroscopy*, 39 (1985) 491.
- [69] M.L. Balboni, *Pharmaceutical Technology* (2003).
- [70] L.X. Yu, R.A. Lionberger, A.S. Raw, R. D'Costa, H. Wu, A.S. Hussain, *Advanced Drug Delivery Reviews*, 56 (2004) 349.
- [71] S. Gnath, M. Jenzsch, R. Simutis, A. Lübbert, *Journal of biotechnology*, 132 (2007) 180.
- [72] J.A. Lopes, P.F. Costa, T.P. Alves, J.C. Menezes, *Chemometrics and intelligent laboratory systems*, 74 (2004) 269.
- [73] S.D. Mansfield, C. Mooney, J.N. Saddler, *Biotechnology Progress*, 15 (1999) 804.
- [74] B.C. Saha, R.J. Bothast, *Applied biochemistry and biotechnology*, 76 (1999) 65.
- [75] M. Pedersen, A.S. Meyer, *New biotechnology*, 27 (2010) 739.
- [76] J. Agger, K.S. Johansen, A.S. Meyer, *New biotechnology*, 28 (2011) 125.
- [77] N. Qureshi, B.C. Saha, B. Dien, R.E. Hector, M.A. Cotta, *biomass and bioenergy*, 34 (2010) 559.
- [78] K. de Carvalho Lima, C. Takahashi, F. Alterthum, *Journal of industrial microbiology & biotechnology*, 29 (2002) 124.
- [79] B. Dien, R. Hespell, L. Ingram, R. Bothast, *World Journal of Microbiology and Biotechnology*, 13 (1997) 619.
- [80] J. Schutte, J. De Jong, E. Van Weerden, M. Van Baak, *British Poultry Science*, 33 (1992) 89.
- [81] C.M. Andersen, R. Bro, *Journal of chemometrics*, 24 (2010) 728.
- [82] X. Zou, J. Zhao, Y. Li, *Vibrational spectroscopy*, 44 (2007) 220.
- [83] W.G. Willats, J.P. Knox, J.D. Mikkelsen, *Trends in Food Science & Technology*, 17 (2006) 97.
- [84] I. Bottger, *Gums and stabilisers for the food industry*, 5 (1990) 247.
- [85] I. Moller, S.E. Marcus, A. Haeger, Y. Verhertbruggen, R. Verhoef, H. Schols, P. Ulvskov, J.D. Mikkelsen, J.P. Knox, W. Willats, *Glycoconjugate journal*, 25 (2008) 37.
- [86] T. Kristensen, J.D. Mikkelsen, J.P. Knox, *Proteomics*, 2 (2002) 1666.
- [87] J. Harholt, A. Suttangkakul, H.V. Scheller, *Plant physiology*, 153 (2010) 384.
- [88] T.M. Wood, K.M. Bhat, *Methods in enzymology*, 160 (1988) 87.
- [89] G.L. Miller, *Analytical Chemistry*, 31 (1959) 426.

PAPER 1

Andreas Baum, Jane Agger, Anne S. Meyer, Max Egebo, Jørn D. Mikkelsen; *Rapid near infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatments*; New Biotechnology, Volume 29, Issue 3, 15 February 2012, Pages 293-301



Rapid near infrared spectroscopy for prediction of enzymatic hydrolysis of corn bran after various pretreatments

Andreas Baum^{1,2}, Jane Agger³, Anne S. Meyer¹, Max Egebo² and Jørn Dalgaard Mikkelsen¹

¹ Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

² Foss Analytical, Slangstrupsgade 69, DK-3400 Hillerød, Denmark

³ Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

Efficient generation of a fermentable hydrolysate is a primary requirement in the utilization of fibrous plant biomass as feedstocks in bioethanol processes. The first biomass conversion step usually involves a hydrothermal pretreatment before enzymatic hydrolysis. The purpose of the pretreatment step is to increase the responsivity of the substrate to enzymatic attack and the type of pretreatment affects the enzymatic conversion efficiency. Destarched corn bran is a fibrous, heteroxylan-rich side-stream from the starch industry which may be used as a feedstock for bioethanol production or as a source of xylose for other purposes. In the present study we demonstrate the use of diffuse reflectance near infrared spectroscopy (NIR) as a rapid and non-destructive analytical tool for evaluation of pretreatment effects on destarched corn bran. NIR was used to achieve classification between 43 differently pretreated corn bran samples using principal component analysis (PCA) and hierarchical clustering algorithms. Quantification of the enzymatically released monosaccharides by HPLC was used to design multivariate calibration models (biPLS) on the NIR spectra. The models could predict the enzymatic release of different levels of arabinose, xylose and glucose from all the differently pretreated destarched corn bran samples. The present study also demonstrates a generic, non-destructive solution to determine the enzymatic monosaccharide release from polymers in biomass side-streams, thereby potentially replacing the cumbersome HPLC analysis.

Introduction

Enzymatic conversion of biomasses is an important step in the exploitation of biomass residuals in the food and biofuel industry and viable processes are dependent on reliable, fast and low-cost process conditions. It is of great importance to know the basic monosaccharide composition and the distribution in polysaccharide forms to target and develop enzymatic conversion, and this traditionally represents a demanding task. In addition, most low-priced biomass resources are of a recalcitrant nature, which is why various hydrothermal or thermo-chemical pretreatment forms are often applied in to maximize the yields of enzymatic conversion

[1–3]. Such pretreatments are costly and the efficacy of the pretreatment step in generating a substrate which is enzymatically degradable needs utmost optimization. Evaluations of pretreatment methods are commonly done by large experimental setups and require time-consuming and advanced analytical methods [4,5].

Near infrared spectroscopy (NIR) as an analytical tool offers benefits, such as high-throughput screening, which are highly valuable to the research field of biomass conversion and various studies have focused on applying NIR spectroscopy as an alternative method for rapid and reliable screening of biomass composition and digestibility. Such diverse cases include prediction of polysaccharide composition of several botanical fractions of

Corresponding author: Mikkelsen, J.D. (jdm@kt.dtu.dk)

switchgrass and corn stover [6], wheat straw digestibility [7], characterization of feedstock and fermentation yields of rye, wheat and triticale kernels for bioethanol production [8] and characterization of starch waxiness in wheat breeding programs [9]. Besides being rapid it is a low cost, highly versatile analytical technique and has the special advantage of being capable of analyzing insoluble, powder samples and suspensions, which are features particularly relevant for recalcitrant biomasses.

NIR analysis involves acquisition of large numbers of absorbencies in a spectral region typically ranging from 1000 to 2500 nm, where organic matter has distinct spectral fingerprints related to functional groups [10]. The large amounts of data generated by such methods are generally combined with multivariate quantitative and qualitative analysis by which predictive models and interpretations can be made [10]. The combination of NIR spectroscopy and multivariate analytical techniques represents an exceptionally feasible method for building reliable models to describe biomasses. Because of the capacity of the multivariate analysis such models can be based on comprehensive data amounts, which may include variables like cultivars, harvest time, plant tissues, storage conditions, growth conditions and locations [6–12].

In the present work, NIR spectroscopy has been applied to evaluate differently pretreated samples originating from corn bran, an agro-industrial residue from corn starch processing [4,11]. The aim of the study was to examine the potential for employing NIR spectroscopy and chemometrics to predict the efficiency of different pretreatments on enzymatic degradation of corn bran. Corn bran represents a complex and recalcitrant biomass stream particularly rich in C-5 carbohydrates mainly present as the structural polysaccharide arabinoxylan, which constitutes approximately 56% of the dry matter (DM) [11]. Corn bran also contains significant amounts of cellulose and is therefore an interesting substrate for enzymatic biomass conversion into readily available monosaccharides.

The spectral data originating from corn bran samples after various pretreatments were compared to results of enzymatic hydrolysis obtained by HPLC analysis of enzymatically released monosaccharides. It was hypothesized that: (i) NIR spectroscopy can hierarchically distinguish insoluble corn bran according to pretreatment conditions and (ii) the extent of enzymatic hydrolysis of corn bran after a particular set of pretreatment conditions can be predicted from the insoluble corn bran's NIR spectra.

Materials and methods

Experimental setup

The experiment was designed as illustrated in Fig. 1. Destarched corn bran was subjected to different pretreatments with varying temperature, pH and time as outlined in 'Biomass' section. Freeze dried samples from each pretreatment experiment were used in powder form for NIR measurements. Concomitantly the pretreated DCB samples were enzymatically hydrolyzed to release monosaccharides, which were quantified by HPLC techniques as described in 'Enzymatic hydrolysis and monosaccharide analysis' section.

First NIR spectroscopy was used to achieve classification among the different pretreatment conditions in relation to carbohydrate composition and to detect outliers in the sample sets by using

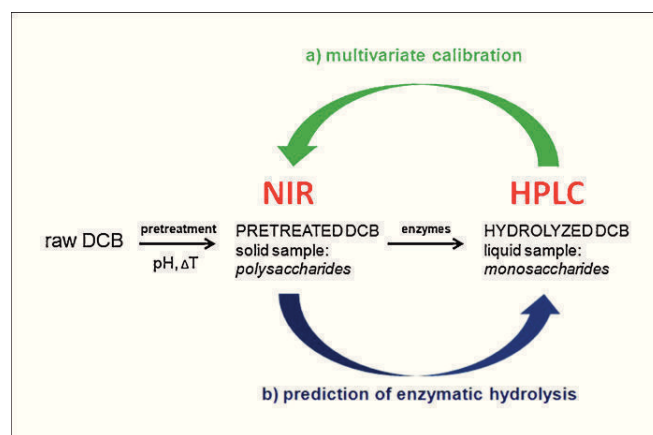


FIGURE 1

Principle of multivariate calibration. HPLC and NIR spectroscopy were used to determine carbohydrate composition of pretreated DCB samples: (a) HPLC data from enzymatic hydrolyzed samples were used to establish multivariate calibration on the NIR spectra; (b) The prediction model established from (a) can then be used to predict the outcome of the enzymatic hydrolysis directly from NIR spectra without need for hydrolysis.

hierarchical clustering algorithms in connection with PCA Analysis. Secondly, the determined quantities of glucose, xylose and arabinose from HPLC were employed to establish multivariate calibrations using Partial Least Squares modeling (Fig. 1a). The established models could then be used for prediction of monosaccharide release (Fig. 1b) by just measuring the NIR spectrum of the pretreated DCB samples (validation).

Biomass

Destarched corn bran after various pretreatments was used as substrate in all enzymatically catalyzed conversions and all NIR spectroscopy measurements. Raw corn bran was obtained from Archer Daniel Midlands Company, Decatur Illinois, USA. The material was milled and enzymatically destarched with α -amylase and amyloglucosidase as previously described [11] before all pretreatments, measurements and hydrolysis experiments.

Pretreatment conditions were categorized into five different setups (A–E) as follows:

- A: *Raw destarched corn bran, sorted according to particle sizes.* This substrate was identical to group B substrate except that no pretreatment other than destarching had occurred. After destarching, the material was freeze dried, milled and sorted into four particle size fractions based on size by sieving, namely from [1000;710], [710;355], [355;250], [250;150] μm [4]. These four particle size fractions were analyzed by NIR and enzymatically hydrolyzed.
- B: *Elevated temperature pretreatment, sorted according to particle sizes.* Destarched corn bran in an aqueous non-buffered 6% DM weight/volume (w/v) slurry was pretreated in a loop autoclave at Risø DTU, National Laboratory for Sustainable Energy, Roskilde, Denmark as described previously [4,12]. The pretreatment consisted of heating to 190°C for 10 min. No chemicals were added. After pretreatment the pH was 4.2 in the total slurry. Hereafter the material was separated into a soluble and an insoluble fraction by filtration and the

insoluble residue was washed in Milli-Q water and freeze dried. Finally, the freeze dried insolubles were milled and sorted into four fractions of different particle size as described for group A. These four particle size fractions were analyzed by NIR and enzymatically hydrolyzed.

- C: *Mid-range pH catalyzed pretreatments at intermediate temperatures.* In these experiments destarched corn bran was suspended in a 2% DM (w/v) aqueous non-buffered slurry before heat incubation in a custom-built pretreatment reactor as described in [4]. Before incubation pH was adjusted by either HCl or NaOH to meet initial pH conditions of either 2, 7 or 12 in the pretreatment. The pre-pH adjusted corn bran slurry was incubated at temperatures ranging from 100 to 150°C for 10–120 min. The pretreated slurries were enzymatically hydrolyzed and an aliquot of the material was then filtered and the insoluble residues were washed in Milli-Q water before freeze drying for NIR measurements.
- D: *Low pH catalyzed pretreatment at intermediate temperature.* This material was pretreated at conditions similar to those applied for group C as the pretreatment reactor, DM loadings and handling after pretreatment were the same. In contrast to group C the pH was adjusted between 1 and 2, all experiments were performed at 150°C and the pretreatment time was narrowed down to be between 10 and 65 min [4].
- E: *pH catalyzed pretreatments with extended incubation time.* The last pretreatment group included destarched corn bran samples which had been pretreated similarly as the samples in group C and D, but either pre-adjusted to pH 2 or 12 and then incubated for extended time intervals (from 120 to 240 min) as described in [4].

Enzymatic hydrolysis and monosaccharide analysis

All enzymatic hydrolysis experiments were conducted according to the same scheme. For dry samples a 2% DM (w/v) suspension was made in 0.1 M succinate buffer pH 5. Slurry samples were pH adjusted to 5 before hydrolysis. A basic set of enzymes was applied to all samples, one endo-(1,4)- β -xylanase (GH10, *Humicola insolens*), one β -xylosidase (GH3, *Trichoderma reesei*), two α -L-arabinofuranosidases (GH43 from *H. insolens* and GH51 from *Meripilus giganteus*) all described in [13,14]. These four enzymes were dosed at 0.25 mg enzyme protein (EP)/g DM each as mono component preparations and the blend was further supplemented by acetyl xylan esterase from *Flavolaschia* sp. [11] and feruloyl esterase type A from *Aspergillus niger* [15] each dosed at 0.5 mg EP/g DM. Finally, a commercially available cellulase preparation was added, named CellicTM CTec which is based on the well-known *T. reesei* cellulase enzyme complex, supplemented by β -glucosidase and hydrolysis-boosting proteins GH61 [16,17]. The cellulase preparation was dosed at 4 mg EP/g DM. Samples were incubated at 50°C for 24 hours and enzymatic reactions stopped by heating to 100°C for 10 min. All enzymatic hydrolysis experiments were performed in triplicate. Hereafter, all samples were analyzed for free monosaccharides arabinose, glucose and xylose by HPAEC-PAD on a BioLC Dionex system, equipped with a CarboPac PA1 column (analytical 4 mm \times 250 mm) also from Dionex, according to the method given in [18]. Before pretreatment experiments the basic monosaccharide composition of raw destarched corn bran was established by methods described in [11]. This composition

was generally applied for calculating yields after enzymatic hydrolysis.

Diffuse reflectance near infrared spectroscopy (NIR)

Diffuse reflectance NIR spectra were obtained using a monochromator-based XDS instrument connected to a Rapid Content Analyzer (FOSS ANALYTICAL, Hillerød). The instrument scans the full-visible and near infrared range spanning from 400 nm to 2500 nm using a Dual detector (silicon 400–1100 nm; lead sulfite 1100–2500 nm) and at a resolution of 2 nm⁻¹. Wavelength accuracy and acquisition rate were specified as <0.05 nm and 2 scans per second, respectively.

The destarched corn bran samples were measured directly as powders with no further physical pretreatment using a round quartz cuvette (Hellma, Sussex, UK) specified by a 1.8 cm diameter. The sample amount was 150 mg and the measurement window (spot-size) for spectra acquisition reduced to 10 mm. Thirty-two scans were performed for each spectrum acquisition.

Four Polysaccharide Standards were used for comparison of spectral features. Cellulose (Avicel) and Xylan (Birchwood) were purchased from Sigma–Aldrich (Steinheim, Germany) and water soluble and insoluble wheat arabinoxylan from Megazyme (Bray, Ireland), respectively. Each polysaccharide standard was simply measured directly by NIR as the powder.

Chemometrics

Spectra were acquired by the Software VISION (Version 3.4.0.0, Foss Analytical, Hillerød, Denmark) and thereafter exported for examination using Matlab 2010 (MathWorks, Natick, MA, USA) concerning the following chemometric methods.

All spectra were pretreated using Multiplicative Signal Correction (MSC) and mean-centering. MSC [19,20] removes scatter components in spectra which can arise when measuring directly on powders. The technique calculates an average spectrum of all samples followed by a univariate regression between the average spectrum and all others. The estimated offset and slope were finally subtracted from and divided by each individual spectrum, thereby yielding a corrected spectrum. Individual scatter effects are therefore corrected.

Further on principal component analysis (PCA) [21], partial least squares [22] and backwards interval partial least squares models (biPLS) [23] were performed using PLS TOOLBOX (Version 6.0.1, Eigenvector Research). Dendrograms [24,25] were calculated and drawn by a MATLAB program using average linkage algorithm [26]. Short descriptions of these methods are given in the following.

Principal component analysis

PCA is a mathematical method which decomposes a set of observations (set of spectra) into a set of principal components (PCs) whose number is less or equal to the number of original variables. Concerning spectroscopy a set of spectra can therefore be expressed by several orthogonal ‘factor spectra’ called *loadings*. Furthermore each spectrum can be recombined by a linear combination of those *loadings* using individual *scores*. Usually the application of PCA implies a remarkable reduction of necessary variables (PCs) to describe a spectral system. If only few PCs are necessary to characterize a spectral system it is therefore possible to visualize differences between spectra by plotting the scores of

those PCs against each other. Thus the position of a sample in a score plot describes the characteristics of the measured sample.

Dendrogram

Dendrograms are used for hierarchal clustering analysis to achieve classification. Usually it is very difficult to visualize clusters along PCA scores using more than three PCs. Dendrogram plotting overcomes that disadvantage because it can handle visualization of samples with more than three independent variables using a hierarchal family tree like structure.

The algorithm performs $n - 1$ steps where n is the sample size. In all the steps it calculates the Euclidean distance of the two closest samples. Thereafter it combines those two closest samples to a new one with, for example, average coordinates (average linkage algorithm). After $n - 1$ steps a dendrogram can be drawn to visualize hierarchically in which samples are most probably comparable even considering many independent variables. This tool is very useful if classification using PCA along a sample set needs to consider more than three factors.

Partial least squares regression

Partial least squares regression functions are comparable to PCA. In contrast to PCA such a regression calculates loadings which maximize variance using supervising information concerning, for example, analyte concentration. It finds projections of scores which maximize spreading of the *scores* due to different underlying analyte concentrations and therefore enables multivariate calibration modeling.

Interval partial least squares and backwards interval partial least squares regression

In comparison to usual PLS interval partial least squares regression (iPLS) optimizes model performance by selecting spectral regions which contain highly significant information about the analyte. Spectral intervals with high degree of interference from the matrix will be ignored because they deteriorate the model performance. Backwards interval partial least squares regression is working in the same manner as iPLS, but instead of adding spectral intervals with significant information it subtracts regions which deteriorate the model performance. Both iPLS and biPLS are usually combined with interval selection algorithms to optimize computing time. Genetic algorithms have been reported [27,28] to be very efficient in that sense. Besides improvement of modeling performance interval selection methods as iPLS and biPLS lead to highlight

spectral regions and identification of specific bands in the spectra concerning the modeled analyte.

Results and discussions

NIR spectra

The sample set consisted of 43 powder samples, which were organized in groups according to the different pretreatment conditions they had undergone (Table 1). All spectra showed different absorbencies concerning different NIR vibration regions (Fig. 2). In addition the figure also shows the spectra of four polysaccharide standards namely, for example, water-insoluble arabinoxylan (blue), water-soluble arabinoxylan (red), xylan (cyan) and cellulose (green). Several polysaccharide NIR absorbencies could be identified due to either combination vibration bands at 2000–2500 nm, first overtone at 1600–2000 nm or second overtone vibration bands at 1400–1600 nm (Fig. 2). Strong absorbencies due to moisture appeared at 1950 nm, 2450 nm and less intensely around 1450 nm (Fig. 2). Carbohydrate bands could be identified for CH_2 and CH between 2300 and 2400 nm due to combination bands, at 1700–1800 nm due to first overtone and around 1400–1500 nm due to second overtone bands, respectively [10]. ROH combination bands appeared around 2100 nm (Fig. 2). The spectral features of the carbohydrates and the subtle differences among the samples could be recognized in all the 43 spectra. Noticeably all 43 corn bran spectra were most comparable to the insoluble arabinoxylan standard spectrum (blue spectrum in Fig. 2). Additionally some samples showed stronger absorbencies related to cellulose features (green spectrum in Fig. 2) than others. Regardless that there are spectral features of the carbohydrates being visible in the spectra many were strongly overlapped with spectroscopic responses from the sample matrix. This emphasizes the need of chemometric tools for multivariate analysis to prove the underlying carbohydrate chemistry.

PCA and hierarchal cluster analysis

The spectra were MSC pretreated followed by mean-centering. All 43 samples were considered for PCA Analysis and hierarchal classification based on the PCA scores.

The PCA decomposition showed that 11 PCs were necessary to explain the significant variance in the spectra. The number of components could be reduced to 5 by excluding the spectra of group D (harsh acidic, Table 1) thus indicating that those samples were least comparable to the rest of the groups. This finding was

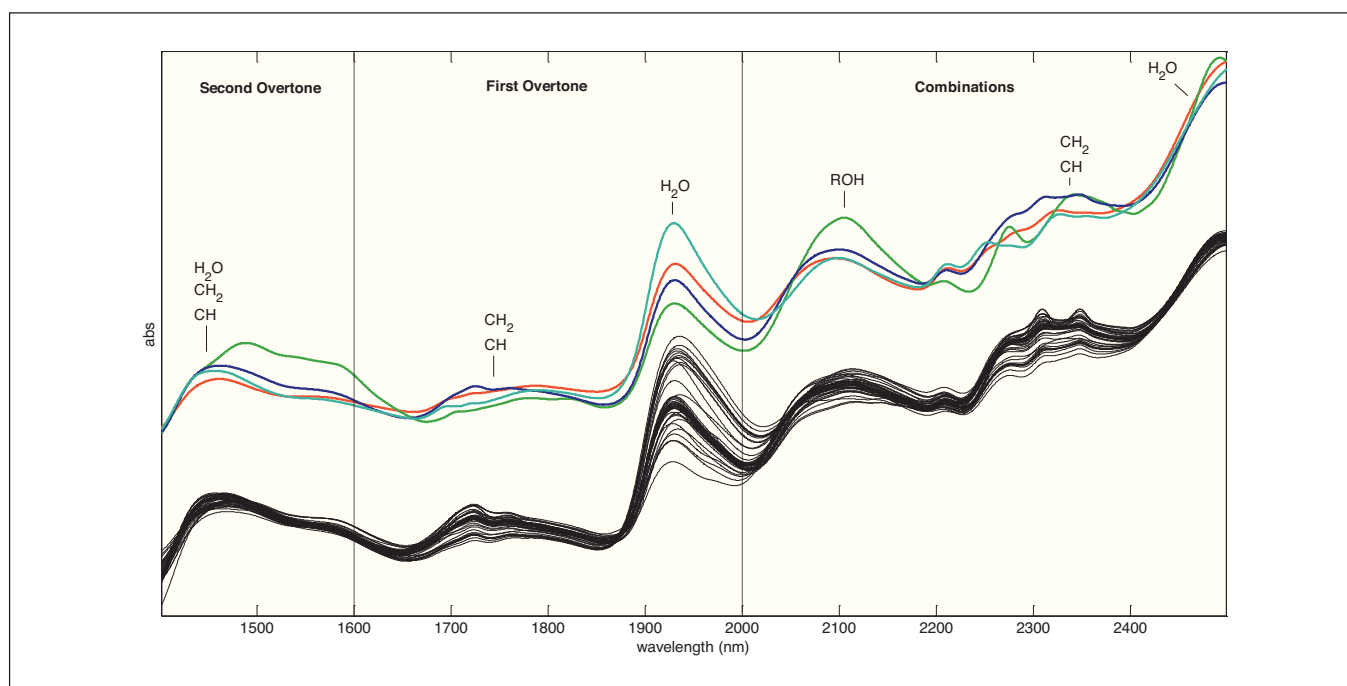
TABLE 1

Group assignment of samples according to pretreatment conditions

	pH		Temperature (°C)		Time (min)	
	Average	Interval	Average	Interval	Average	Interval
A ^{a,b}	–	–	–	–	–	–
B ^a	4.2	4.2–4.2	190.0	190–190	10.0	10–10
C	5.8	1.7–9.8	125.0	100–150	65.0	10–120
D	1.5	0.9–2.1	150.0	150–150	32.0	5–65
E	3.8	1.8–9.1	143.8	125–150	187.5	120–240

^a Powders with different particle size due to milling.

^b No pretreatment.

**FIGURE 2**

NIR spectra. Information from the visible range has been cut off and each spectrum has been MSC pretreated (black). Standard spectra for arabinoxylan soluble (red), arabinoxylan insoluble (blue), cellulose (green) and xylan (cyan) have been included with an offset for comparison. Various combination, first overtone and second overtone vibration bands have been highlighted for chemical interpretation.

also evident from the scores plot in which group D clustered with the highest distance to the others (Fig. 3).

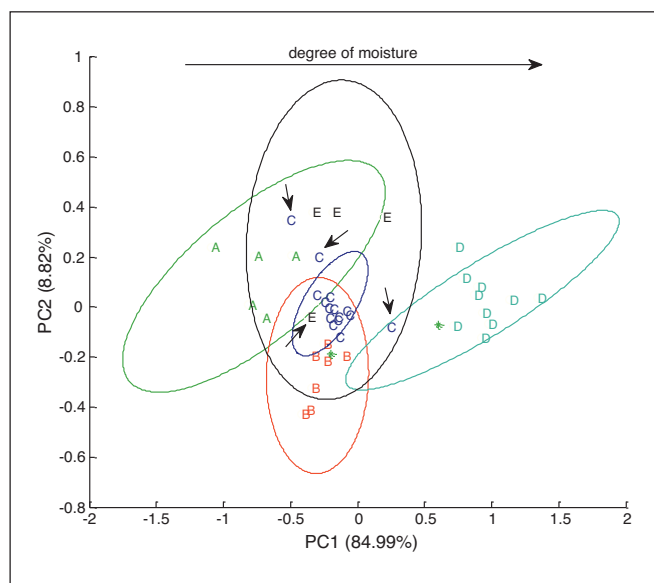
It was also evident from the PCA scores plot that noticeably the scores of group A, which belonged to samples with different particle sizes, spread more than scores from other groups. This is in contrast to scores from group B, which were also measured on different particle sizes (Fig. 3). Particle size influences NIR spectra

remarkably due to scattering effects which have been proven in several studies for determination of particle size itself [20]. However, in this particular case the application of MSC pretreatment removed those effects and the spread of group A scores was therefore not explained by different particle sizes, but rather by other features, notably the water content of the samples.

As water has a high impact on NIR spectra the scores especially along PC1 were highly affected by the water content of the samples as illustrated in Fig. 3. The interpretation of the first PCA loading showed that different amounts of moisture in the powder samples imply high leverage on the PC1 scores. Thus the spreading of group A samples, which occurred mainly along PC1, can be explained by large variations of moisture content of the samples. This can also be directly observed from the MSC pretreated spectra regarding different intensities of the first overtone vibration band due to H₂O at 1950 nm. Group B showed no significant shift on PC1 scores thus indicating comparable amounts of moisture in these samples. Differences between the two groups probably arose from different storage conditions which might have affected the uptake of moisture from the surroundings.

Two samples from group A (marked with stars in Fig. 3) were excluded from the cluster as outliers. Group D samples revealed the highest amount of moisture. In group C there was also one sample which indicated a very high amount of moisture, but this sample has not been identified as an outlier due to the following interpretation using a dendrogram.

When including the discriminative power of the second PC a classification not only due to moisture was possible. Nonetheless 11 PCs were needed in the PCA decomposition to describe the 43 spectra completely in relation to their full underlying chemical composition as mentioned before.

**FIGURE 3**

PCA score plot colored due to pretreatment grouping. All groups are presented with 95% confidence regions. Two samples have been identified as outliers (*). Four samples (marked with arrows) rather clustered to a foreign group.

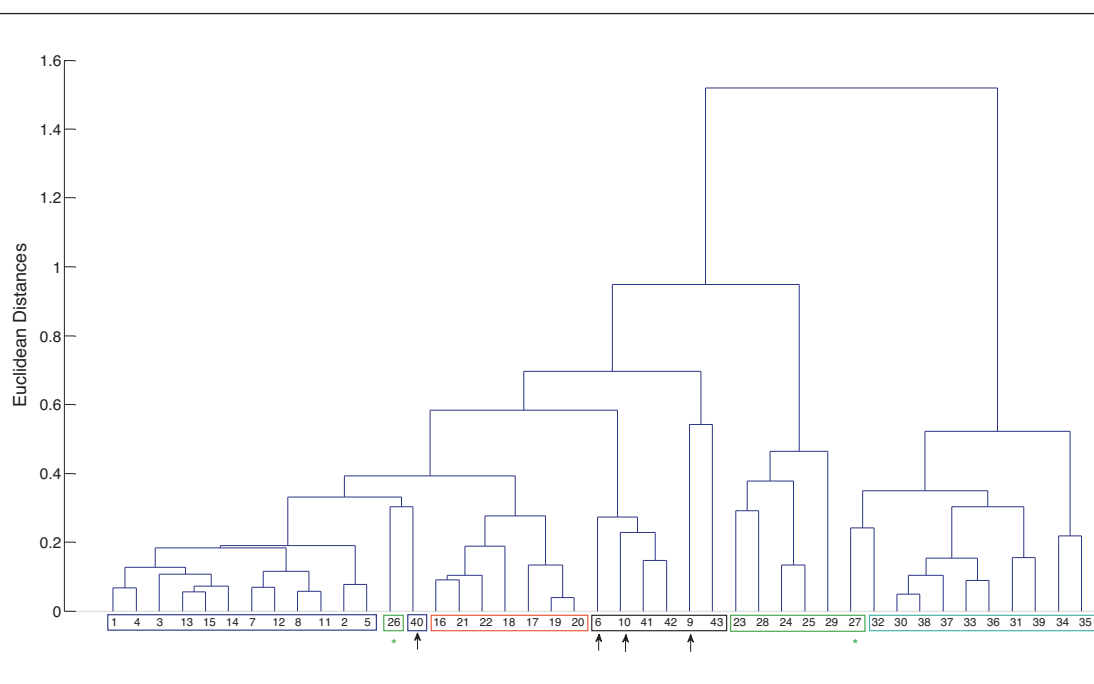


FIGURE 4

Hierarchal clustering using a dendrogram. Samples 26 and 27 are identified as outliers by the dendrogram (*). Samples marked with arrows rather clustered to foreign groups. In this particular case they have been colored according to the group they actually clustered to green – group A, red – group B, blue – group C, light blue – group D and black – group E.

As it was difficult to visualize the data in more than two dimensions all 43 scores from five PCs were used to draw a dendrogram (Fig. 4). The Euclidean distances were calculated by average linkage algorithm as described in [26]. The outlier samples 26 and 27 due to much higher amount of moisture could be clearly detected and were therefore removed (marked with * in Figs 3,4). Additionally it can be concluded that group D was the most outlying group because its averaged Euclidean distance to the rest of the samples was very high. That result was already indicated above by the remarkably increasing number of necessary PCs to describe the system including that group. Thus, the result of the Euclidean distance calculation justified the decision to not consider those samples for the calibration of multivariate prediction models for the enzymatic hydrolysis.

Finally it was noted that hierarchal clustering enabled strong classification power along the different sample sets. Especially by interpreting the dendrogram in Fig. 4 it was found that four

samples were rather clustered to a foreign group. Those samples 6, 9, 10 and 40 have been highlighted with arrows in Fig. 3 and recolored according to the groups which they rather belong to in Fig. 4. Those reassignments could be confirmed by examining the pretreatment parameters of those samples. Sample 40 was pretreated at a pH comparable to that from group C. Thus it can be seen that pH has a strong influence on the NIR spectra and therefore on the carbohydrate composition as described previously in [4].

The findings show that PCA in combination with hierarchal analysis is perfectly suitable for discriminating between the different kinds of pretreatment. Further discriminant analysis methods such as linear discriminant analysis (LDA) [29] or partial least squares discriminant analysis (PLSDA) [30] can be used to establish intelligent instrumentation devices with implemented pattern recognition. At this point such kind of study was beyond the scope of the present work.

TABLE 2

Relative monosaccharide release after enzymatic hydrolysis for the groups representing different pretreatment conditions

	Arabinose (%)		Xylose (%)		Glucose (%)	
	Average	Interval	Average	Interval	Average	Interval
A ^{a,b,c}	9.2	2.5–16.5	3.7	0.9–7.5	15.2	5.5–30.2
B ^a	25.1	22.1–28.4	17.3	14.7–21.2	65.5	51.0–86.4
C	16.7	1.4–26.3	9.5	4.2–25.3	24.7	10.4–63.7
D	13.2	5.7–18.3	n.d.	n.d.	57.1	41.3–66.7
E	6.3	0.0–25.4	11.7	10.2–14.4	59.5	51.5–69.1

^a Powders with different particle size due to milling.

^b No pretreatment.

^c Outliers have been excluded from multivariate calibration.

TABLE 3

Modeling parameters as leave-one-out cross-validated RMSECV, RMSEC, regression coefficients R^2 and number of latent variables (LV) are presented for the PLS/biPLS prediction models using the full spectra and the selected spectral intervals, respectively

	RMSECV (%)		RMSEC (%)		R^2		Number of LVs	
	Full spectra	biPLS	Full spectra	biPLS	Full spectra	biPLS	Full spectra	biPLS
Arabinose	3.12	1.39	1.07	0.99	0.82	0.96	7	6
Xylose	1.16	0.77	0.90	0.62	0.95	0.98	3	3
Glucose	8.68	5.69	6.37	3.83	0.84	0.93	5	5

Multivariate prediction models for enzymatic hydrolysis using biPLS

PLS models were calibrated for prediction of enzymatic release of glucose, xylose and arabinose using HPLC data from the hydrolyzed samples on the MSC and mean centered NIR spectra of the pretreated material before hydrolysis (as illustrated in Fig. 1). Only samples from groups A, B, C and E were considered for the PLS calibration (Table 2) due to outlying characteristics of groups D as discussed earlier. To highlight the most informative spectral regions iPLS and biPLS regression were applied, respectively. BiPLS generated better model performances and is therefore the only

method considered in the following. The associated modeling performance of biPLS in comparison to PLS on the whole NIR spectra is illustrated in Table 3, where additional important parameters such as RMSEC, RMSECV and regression coefficients R^2 are also stated. To emphasize the actual underlying chemistry due to the monosaccharide content selected spectral regions have been cut out and pretreated separately.

As an example, Figure 5d, presents the spectral region from 2172 to 2272 nm which was highly significant for total arabinose content. Looking at this selected interval it was clearly visible that there were trends in spectra in relation to the total amount of

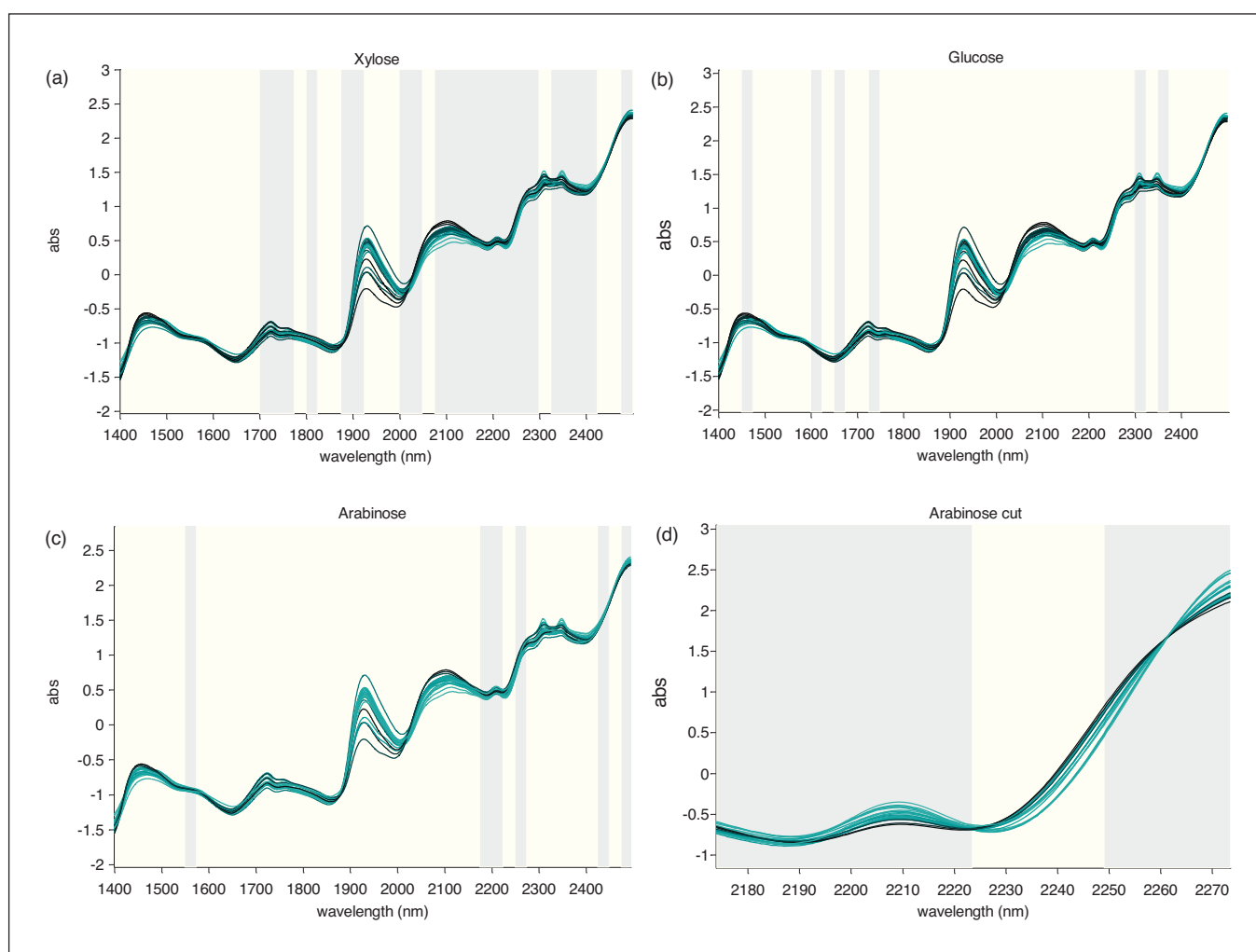
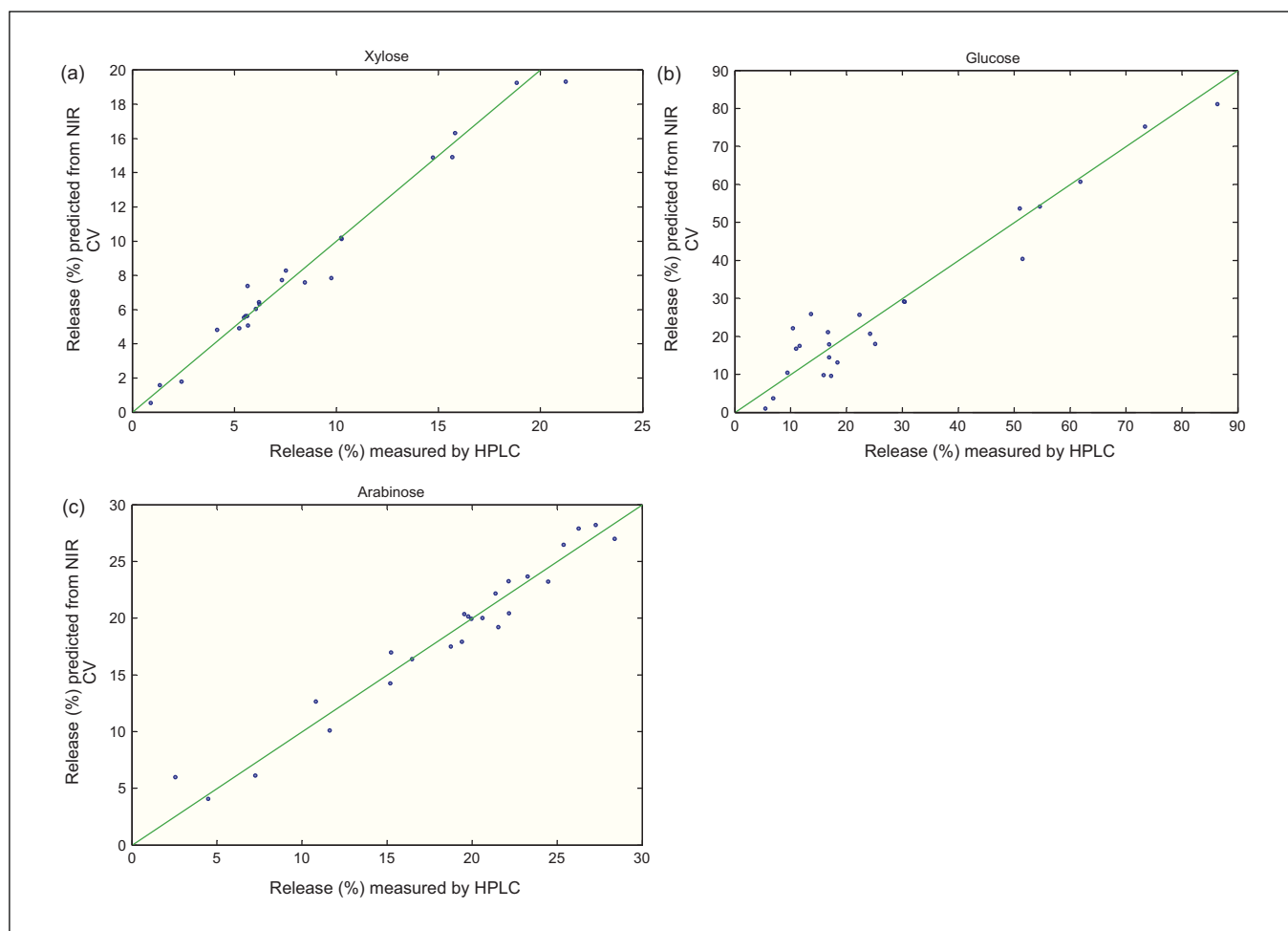


FIGURE 5

Selected spectral intervals using biPLS highlighted in grey. Xylose (a), glucose (b) and arabinose (c). Spectra are colored according to analyte concentration (black – low conc.; light blue – high conc.) Spectral range from 2172 to 2272 nm is cut out to emphasize underlying arabinose bands (d).

**FIGURE 6**

Cross-validated prediction of enzymatic release of xylose **(a)**, glucose **(b)** and arabinose **(c)** from NIR versus measured Monomer Release by HPLC after hydrolysis.

arabinose being present in the pretreated DCB samples. Additionally it is remarkable that samples with high glucose amounts showed exceptionally strong absorbencies in the second overtone regions around 1600 nm, which are highly influenced by cellulose bands, as it can be seen from the cellulose standard spectrum in Fig. 2. The spectral regions have been selected by biPLS. Finally one can state that the spectral regions selected for multivariate calibration can be interpreted in agreement with earlier discussed spectral features of the measured commercial polysaccharide standards. This result emphasizes that interval selection in combination with PLS is suitable for highlighting chemical information about arabinose, glucose and xylose in the NIR spectra. Accordingly strong support for multivariate modeling of monosaccharide content can be ascertained.

All spectral intervals which were selected for the calibration of PLS prediction models are shown in Fig. 5. Spectrum color indicated the extent of enzymatic release concerning each monomer (black indicates low releases, light blue, high releases) measured by HPLC. Out of the three monomers in question xylose was modeled with the best performance because it had the highest abundance in DCB as reported in [11]. Meanwhile biPLS also showed that xylose modeling performance could not

be improved by leaving rather many segments out as it could for arabinose and glucose. All three models could be improved by leaving out spectral segments with strong water absorbencies due to moisture.

The leave-one-out cross-validated result of the biPLS calibration models for arabinose, xylose and glucose are shown in Fig. 6. Results have also been validated by leaving out subgroups while calibrating the models which did not decrease model performance significantly. Poorest performance was reported for the model of glucose release after hydrolysis. This might be influenced by two factors. First, the content of total glucose monomers is lower compared to that of arabinose and xylose [11]. As mentioned earlier the spectra were generally best described by the standard spectrum of insoluble arabinoxylan and therefore the crucial information about the glucose monomers in the biomass may be hidden behind the higher content of arabinoxylan constituents. Secondly, differences in glucose origin from either crystalline or amorphous cellulose might also be speculated to interfere with the modeling as crystalline cellulose is more difficult to hydrolyze than amorphous. However previous studies on crystalline lactose [31] have shown that NIR is capable of differentiating between both species.

Conclusion

Two major hypotheses were tested and confirmed in this study. First it was possible to enable classification between the different pretreatments of the DCB samples. It was shown that NIR spectroscopy is a powerful non-destructive tool for rapid evaluation of the pretreatment process. Hence, both of the hypotheses set up for the study were found to be valid, and this offers remarkable potentials for fast screening and evaluation of the effects of biomass pretreatment. As NIR can also be used to measure suspensions further studies can be carried out measuring NIR spectra even during the pretreatment process and therefore examine the effect of pretreatment time-resolved. This will give valuable information about optimization parameters and thereby enhance monosaccharide release and reduce the process cost. In addition the method is highly suitable for detection of outliers, can be utilized using discriminant analysis and implemented within intelligent instrumentation.

Furthermore, by means of sophisticated chemometric methods it has been demonstrated that the monosaccharide release of the pretreated material can be predicted quantitatively from the solid

pretreated DCB samples directly without need for the enzymatic hydrolysis itself. Thus once having a quantitative calibration model established there is no demand for the reducing sugar and HPLC analysis to determine the monosaccharide release or composition of the biomass polymer. The explicit effects of the different pretreatments can be assessed rapidly and reliably.

Industry has not recognized the great potential of NIR spectroscopy yet as a fast and reliable method eminently suitable for high-throughput screening and on-line monitoring tasks in carbohydrate analytics. The exploitation of this generic method could be extended to a range of biomass feedstocks, using both high-throughput and/or on-line analyses of the pretreatments as well as the enzymatic hydrolysis data. With proper multivariate calibration NIR methodology can be developed further to directly predict enzymatic susceptibility of pretreated biomasses.

Acknowledgement

This work was partially financed by the 7th Framework Programme via the Marie Curie Initial Training Network, LeanGreenFood.


References

- Mansfield, S.D. *et al.* (1999) Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnol. Progr.* 15, 804–816
- Saha, B.C. and Bothast, R.J. (1999) Pretreatment and enzymatic saccharification of corn fiber. *Appl. Biochem. Biotechnol.* 76, 65–77
- Pedersen, M. and Meyer, A.S. (2010) Lignocellulose pretreatment severity – relating pH to biomatrix opening. *New Biotechnol.* 27, 739–750
- Agger, J. *et al.* (2011) pH catalyzed pretreatment of corn bran for enhanced enzymatic arabinoxylan degradation. *New Biotechnol.* 28, 125–135
- Cotta, M.A. and Saha, B.C. (2009) *Fuel Ethanol Production from Barley Straw* (Research NCfAU ed), U.S. Department of Agriculture
- Liu, L. *et al.* (2010) Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. *Carbohydr. Polym.* 81, 820–829
- Bruun, S. *et al.* (2010) Prediction of the degradability and ash content of wheat straw from different cultivars using near infrared spectroscopy. *Ind. Crops Prod.* 31, 321–326
- Pohl, F. and Senn, T. (2011) A rapid and sensitive method for the evaluation of cereal grains in bioethanol production using near infrared reflectance spectroscopy. *Bioresour. Technol.* 102, 2834–2841
- Delwiche, S.R. *et al.* (2011) Starch waxiness in hexaploid wheat (*Triticum aestivum* L.) by NIR reflectance spectroscopy. *J. Agric. Food Chem.* 59, 4002–4008
- Zou, X.B. *et al.* (2010) Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667, 14–32
- Agger, J. *et al.* (2010) Enzymatic xylose release from pretreated corn bran arabinoxylan: differential effects of deacetylation and deferuloylation on insoluble and soluble substrate fractions. *J. Agric. Food Chem.* 58, 6141–6148
- Bjerre, A.B. *et al.* (1996) Pretreatment of wheat straw using combined wet oxidation and alkaline hydrolysis resulting in convertible cellulose and hemicellulose. *Biotechnol. Bioeng.* 49, 568–577
- Sorensen, H.R. *et al.* (2006) A novel GH43 alpha-L-arabinofuranosidase from *Humicola insolens*: mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan. *Appl. Microbiol. Biotechnol.* 73, 850–861
- Sorensen, H.R. *et al.* (2007) Enzymatic hydrolysis of wheat arabinoxylan by a recombinant ‘minimal’ enzyme cocktail containing beta-xylosidase and novel endo-1,4-beta-xylanase and alpha-(L)-arabinofuranosidase activities. *Biotechnol. Progr.* 23, 100–107
- Faulds, C.B. and Williamson, G. (1994) Purification and characterization of a ferulic acid esterase (Fae-III) from *Aspergillus niger* – specificity for the phenolic moiety and binding to microcrystalline cellulose. *Microbiology-UK* 140, 779–787
- Rosgaard, L. *et al.* (2007) Evaluation of minimal *Trichoderma reesei* cellulase mixtures on differently pretreated barley straw substrates. *Biotechnol. Progr.* 23, 1270–1276
- Harris, P.V. *et al.* (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49, 3305–3316
- Sorensen, H.R. *et al.* (2003) Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between alpha-L-arabinofuranosidases, endo-1,4-beta-xylanases, and beta-xylosidase activities. *Biotechnol. Bioeng.* 81, 726–731
- Geladi, P. *et al.* (1985) Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39, 491–500
- Martens, H. *et al.* (2003) Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Anal. Chem.* 75, 394–404
- Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441
- Wold, S. *et al.* (1983) The multivariate calibration-problem in chemistry solved by the PLS method. *Lect. Notes Math.* 973, 286–293
- Leardi, R. and Norgaard, L. (2004) Sequential application of backward interval partial least squares and genetic of relevant spectral regions. *J. Chemom.* 18, 486–497
- Aldecoa, R. and Marin, I. (2010) Jerarca: Efficient analysis of complex networks using hierarchical clustering. *Plos One* 5
- Stanberry, L. *et al.* (2003) Cluster analysis of fMRI data using dendrogram sharpening. *Hum. Brain Mapp.* 20, 201–219
- Massart, D.L. *et al.* (1997) *Handbook of Chemometrics and Qualimetrics*. Elsevier Science B.V.
- Leardi, R. and Gonzalez, A.L. (1998) Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemom. Intell. Lab. Syst.* 41, 195–207
- Leardi, R. *et al.* (2002) Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Anal. Chim. Acta* 461, 189–200
- Baum, A. *et al.* (2010) Differentiation between origins of extra virgin olive oils by GC-C-IRMS using principal component analysis, linear discriminant analysis, and hierarchical cluster analysis. *Spectroscopy* 25, 40–+
- Yang, Z. *et al.* (2008) Discrimination of wood biological decay by NIR and partial least squares discriminant analysis (PLS-DA). *Spectrosc. Spect. Anal.* 28, 793–796
- Norgaard, L. *et al.* (2005) Multivariate near-infrared and Raman spectroscopic quantifications of the crystallinity of lactose in whey permeate powder. *Int. Dairy J.* 15, 1261–1270

PAPER 2

Andreas Baum, Malgorzata Dominiak, Silvia V. Melgosa, William G.T. Willats, Karen M. Søndergaard, Per W. Hansen, Jørn D. Mikkelsen; *FTIR and carbohydrate microarray analysis for prediction and characterization of enzymatic versus acidic extracted pectin*; Carbohydrate Polymers, submitted

Prediction of Pectin Yield and Quality by FTIR and Carbohydrate Microarray Analysis

Andreas Baum^{1,2}  · Malgorzata Dominiak^{1,3} · Silvia Vidal-Melgosa^{4,5} · William G. T. Willats⁴ · Karen M. Søndergaard³ · Per W. Hansen² · Anne S. Meyer¹ · Jørn D. Mikkelsen¹

Received: 4 May 2016 / Accepted: 5 September 2016
© Springer Science+Business Media New York 2016

Abstract Pectin production is complex, and final product quality assessment is generally accomplished at the end of the process using time-consuming off-line laboratory analysis. In this study, pectin was extracted from lime peel either by acid or by enzymes. Fourier transform infrared spectroscopy and carbohydrate microarray analysis were performed directly on the crude lime peel extracts during the time course of the extractions. Multivariate analysis of the data was carried out to predict final pectin yields. Fourier transform infrared spectroscopy (FTIR) was found applicable for determining the optimal extraction time for the enzymatic and acidic extraction processes, respectively. The combined results of FTIR and carbohydrate microarray analysis suggested major differences in the crude pectin extracts obtained by enzymatic and acid

extraction, respectively. Enzymatically extracted pectin, thus, showed a higher degree of esterification (DE 82 %) than pectin extracted by acid (DE 67 %) and was moreover found to be more heterogeneously esterified when probed with the monoclonal antibodies JIM5, JIM7, and LM20. The data infer that enzymatic pectin extraction allows for extraction of complex, high DE pectin, and that FTIR and carbohydrate microarray analysis have potential to be developed into online process analysis tools for prediction of pectin extraction yields and pectin features from measurements on crude pectin extracts.

Keywords Enzymatic extraction · Multivariate analysis · PLS · Antibodies · Online monitoring · Chemometrics

Andreas Baum and Malgorzata Dominiak contributed equally to the manuscript.

Electronic supplementary material The online version of this article (doi:10.1007/s11947-016-1802-2) contains supplementary material, which is available to authorized users.

✉ Andreas Baum
aba@kt.dtu.dk

¹ Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Lyngby, Denmark

² Foss Analytical, Foss Allé 1, 3400 Hillerød, Denmark

³ Department of Hydrocolloid Science, DuPont Nutrition Biosciences ApS, Edwin Rahrs Vej 38, 8220 Brabrand, Denmark

⁴ Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

⁵ Present address: Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany

Introduction

Comprehensive monitoring of pectin yield and quality during extraction, as opposed to quality check after extraction, is needed in the industry to achieve optimal extraction conditions and to enhance process control. Pectin designates a family of plant cell wall polysaccharides, which is principally made up of four structural units: homogalacturonan (HG), xylogalacturonan, rhamnogalacturonan type I (RGI), and rhamnogalacturonan type II (RGII). HG consists of an unbranched chain of α -1,4-linked galacturonic acid residues, which may be methyl esterified and/or acetylated and which may be intervened by rhamnose residues (Voragen et al. 2008). Quantitatively, HG (of different degrees of esterification) is the most dominant component in commercial “pectin,” which in fact designates a series of hydrocolloids used as thickening and gelation agents in foods and in other products such as pharmaceuticals, personal care, and cosmetics (Ciriminna et al. 2015). Pectin is approved for food use in the European Union (E440 (i) and E440 (ii)) and is

usually produced by extraction from apple pomace or citrus peels (lemon, lime, and orange). Recently, significant efforts have been made to develop pectin extraction methods that are more sustainable than the classical acid-assisted extraction, e.g., using enzymes instead of acid (Lim et al. 2012; Wikiera et al. 2015; Zykwiniska et al. 2008), by ultrasound and microwave assisted methods (Boukroufa et al. 2015; Grassino et al. 2016) or by employing new heating strategies in order to lower the energy expenditure (Rudolph and Petersen 2011). Nevertheless, current industrial practice for large-scale pectin production involves extended acid extraction of the raw material, usually involving HNO₃ treatment for 3–12 h at 50–90 °C (Rolin et al. 1998). The final product quality assessment of the pectin is based on data obtained from a number of classical laboratory methods, such as size exclusion chromatography for determination of molecular weight, high-performance liquid chromatography analysis for determination of the monosaccharide composition, and intrinsic viscosity and calcium sensitivity measurements. These methods are carried out in an off-line fashion, are time consuming, and therefore unsuitable for process control. In this paper, two advanced analytical techniques, namely FTIR and carbohydrate microarrays, are assessed with regard to their ability to monitor extractions of pectin by either acid or enzymes.

Fourier transform infrared spectroscopy (FTIR) is based on measurement of absorption in the mid-infrared frequency range. Since organic molecules have intramolecular and intermolecular bonds that absorb in this frequency range, each spectrum obtained will reflect the chemical (structural) composition of the sample being analyzed. Hence, FTIR, in combination with appropriate chemometric methodology required to extract patterns in the spectra obtained, is a rapid, non-destructive analytical method (Kuligowski et al. 2012) that can be employed for a number of complex, analytical applications including quality assessment. FTIR has proven useful for detecting, e.g., the authenticity and adulteration of milk (Nicolaou et al. 2010) and virgin olive oil samples (Rohman and Man 2010) and has more recently shown applicability within more advanced applications to determine enzyme activity by spectral evolution profiling and multiway analysis (Baum et al. 2013a, b).

Carbohydrate microarrays are chip-based tools to study interactions of biomolecules with carbohydrates and thus represent an alternative methodology for carbohydrate analyses (de Paz and Seeberger 2012). Thousands of binding events can be assessed in parallel on a single slide with very small amounts of samples, making this analysis high-throughput capable (Fangel et al. 2012). Carbohydrate microarrays are not only applied in medical, animal, and prokaryote research (de Paz and Seeberger 2012; Fangel et al. 2012) but have also been found to be a valuable tool in pectin research. HG-directed monoclonal antibodies, differing in relation to pattern and degree of esterification, have been developed and applied for analysis of plant cell walls (Clausen et al. 2003; Sørensen and Willats 2011).

The hypothesis of this study is that FTIR and carbohydrate microarrays represent advanced tools for quantitative prediction of pectin yield from measurements on crude extracts during the extraction. The two methods either are high-throughput capable or can enable rapid measurements, therefore having the potential to move pectin quality control closer to the extraction process. In addition, the crude pectin extracts, obtained by acidic and enzymatic treatment of lime peels, can be structurally characterized using the two analytical techniques.

Materials and Methods

Three pectin extractions were carried out. Crude pectin extracts were taken at different time points during the time course of the extractions (Table 1). Pectin extractions I and II were conducted enzymatically using Laminex C2K, while extraction III was conducted using HNO₃. Aliquots of all crude extracts were measured using FTIR spectroscopy and carbohydrate microarray analysis. Pectin yields, used as

Table 1 Pectin yields (%), obtained by propan-2-ol precipitation, at specific time points during enzymatic (extractions I and II) and acidic (extraction III) extraction

Extraction time (min)	Pectin yield (%)		
	Extraction I	Extraction II	Extraction III
0	6.67	6.67	8.15
2.5	9.63		
5	10.4	8.89	14.1
7.5	11.1		
10	11.1	11.1	14.8
12.5	11.9		
15	12.6		
17.5	13.3		
20	13.3		
35	14.8		
30	15.6	16.3	19.3
35	17.0		
40	17.0		
45	17.0		
50	17.0		
55	17.0		
60	18.5	18.5	21.5
70	19.3		
80	20.0		
90	20.7		
100	21.5		
110	20.7		
120	24.5	22.2	23.0
240		23.0	24.4

reference values for the multivariate regression models, were determined by propan-2-ol precipitation. In addition, the monomeric composition of the final products was determined using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Detailed protocols of the applied methods are given below.

Materials

Dry lime peel was obtained from DuPont's pectin plant in Mexico (Dupont Nutrition and Biosciences Mexicana S.R.L, Tecmán, Colima, Mexico). The pre-treatment involved milling of the peel to pass a 35-mesh size screen (centrifugal mill Retsch ZM 200, Haan, Germany). Several standard polysaccharides were used as benchmarks in the final analytical evaluation, namely pectin E81 from lime peel, with a degree of esterification (DE) of 81 %; pectin F31 (DE 31 %) derived from E81 treated with *Aspergillus niger*'s pectin methyl esterase (both from DuPont Nutrition Biosciences, Brabrand, Denmark), prepared as described in Limberg et al. (2000); and RGI from soy bean (Megazyme, Bray, Ireland). The enzyme preparation, Laminex C2K, derived from *Penicillium funiculosum* (obtained from DuPont Industrial Biosciences, Leiden, the Netherlands), was a commercial preparation containing cellulase, xylanase and arabinoxylanase activities and a low level of pectinase activity. The enzyme preparation was chosen as a suitable candidate for pectin extraction in a previous study (Dominiak et al. 2014).

Enzymatic Extraction

For each sample, 1.35 g of milled lime peel was mixed with 40 mL of 50 mM citric acid buffer pH 3.5. The mixture was heated to 50 °C before addition of 75 µL of the enzyme Laminex C2K, equivalent to 235 carboxymethylcellulose-dinitrosalicylic acid activity units. After addition, each sample was incubated in a water bath at 190 rpm for 0–240 min (exact incubation times for extractions I and II are presented in Table 1). The reaction was terminated by heating at 100 °C for 5 min to inactivate the enzyme. The crude extract was separated from the residual biomass by centrifugation for 10 min at 3500 rpm and stored at –20 °C until further analyses.

Acidic Extraction

For each sample, 1.35 g lime peel was mixed with 40 mL deionized water and the mixture was heated to 70 °C. The pH was adjusted to 1.7 with 42 % nitric acid, and the mixture was then incubated in a water bath at 190 rpm (more details about the incubation times in Table 1). The crude extract was separated from the residual biomass by centrifugation for 10 min at 3500 rpm. The pH of the crude extracts was increased to 3.5, and the extracts were then stored at –20 °C until further analyses.

Determination of Pectin Yield

Pectin crude extracts (two replicates) were precipitated by mixing with a double volume of propan-2-ol, followed by agitation for 1 h. The precipitated pectin samples were separated by filtration through a cotton canvas and dried overnight in a 40 °C ventilated oven. Dried pectin samples were weighed and ground to <0.5 mm in an Ultra Centrifugal Mill ZM 200 (Retsch, Haan, Germany).

High-Performance Liquid Chromatography and Colorimetry

To verify and compare the structural information obtained by FTIR and carbohydrate microarray analyses, the chemical composition of the propan-2-ol precipitated pectin samples was determined by HPAEC-PAD and colorimetric measurements. Each pectin sample was washed three times in 80 % ethanol followed by washing one time in 96 % ethanol to remove sugars. After hydrolysis in 1 M sulfuric acid (6 h, 100 °C), the samples were diluted; filtered through a 0.45-µm filter; and the quantitative content of D-fucose, L-rhamnose, D-arabinose, D-galactose, D-glucose, and D-xylose was determined. The separation and quantification were performed using an ICS-3000 system (Dionex Corp., Sunnyvale, CA) utilizing a CarboPac™ PA1 (4 mm × 250 mm) analytical column and a CarboPac™ PA1 guard column. Eluent A consisted of deionized water and eluent B of 150 mM NaOH. Twenty-minute isocratic elution with 15 mM NaOH was followed by 10-min washing with 150 mM NaOH. The flow was kept at 1 mL min^{–1}. Before injection of each sample (50 µL), the column was re-equilibrated with 15 mM NaOH for 8 min. Data analysis was carried out using Chromeleon 6.80 (Dionex Corp., Sunnyvale, CA).

The galacturonic acid content was determined colorimetrically by the *m*-hydroxybiphenyl method in a microtiterplate format reaction (van den Hoogen et al. 1998) after hydrolysis for 1 h at 80 °C in concentrated sulfuric acid with 0.0125 M tetraborate (Blumenkrantz and Asboe-Hansen 1973).

The degree of esterification was determined by alkaline saponification followed by quantification of methanol by HPLC as described in Voragen et al. (1986).

Fourier Transform Infrared Spectroscopy

An aliquot of each crude extract was used to acquire the mid-infrared spectrum between 950 and 1550 cm^{–1}. All spectra were measured at 42 °C using a MilkoScan™ FT2 (FOSS Analytical, Hillerød, Denmark). The instrument consisted of an interferometer which scanned the IR spectrum using a cuvette with a path length of 50 µm. Acquisition was carried out according to the method described by Andersen et al. (2002) with an optical resolution of 14 cm^{–1}. All spectra were

measured against an aqueous blank using an automatic flow-through system (FOSS Analytical, Hillerød, Denmark).

Carbohydrate Microarray Analysis

Extracted Pectin Microarray Printing

The crude extracts were diluted twofold with printing buffer (55.2 % glycerol, 44 % water, 0.8 % Triton X-100) and then two further fivefold dilutions were made. Each of these dilutions was printed in triplicate employing six drops per spot. A piezoelectric microarray robot, Arrayjet Sprint (Arrayjet, Roslin, UK), equipped with a 12-sample high-capacity JetSpyder sample pickup device was used to deposit the crude extracts onto a nitrocellulose membrane with a pore size of 0.45 μm (Whatman, Maidstone, UK) at 22.5 °C and 50 % humidity. As positive controls, the three pectin standard samples E81, F31, and RGI were also deposited on the nitrocellulose membrane.

Extracted Pectin Microarray Probing

Microarrays were blocked for 1 h in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , 1.7 mM KH_2PO_4 , pH 7.5) containing 5 % w/v low fat milk powder (5%M-PBS). Thereafter, microarrays were probed for 2 h with a panel of 30 anti-rat antibodies, anti-mouse antibodies, and carbohydrate-binding modules (CBMs) diluted in 5%M-PBS to 1/10, 1/1000, and 10 $\mu\text{g mL}^{-1}$, respectively. The antibodies and CBMs were obtained from PlantProbes (Leeds, UK), INRA (Nantes, France), BioSupplies (Bundoora, Australia), and NZYTech (Lisboa, Portugal), respectively. Afterward, arrays were washed with PBS and incubated for 2 h in either anti-rat, anti-mouse, or anti-His tag secondary antibodies conjugated to alkaline phosphatase (Sigma, Poole, UK) diluted 1/5000 (for anti-rat and anti-mouse) or 1/1500 (for anti-His tag) in 5 % M-PBS. After the incubation, arrays were washed with PBS and developed with a solution of 5-bromo-4-chloro-3-indolylphosphate and nitro blue tetrazolium in alkaline phosphatase buffer (100 mM NaCl, 5 mM MgCl_2 , 100 mM diethanolamine, pH 9.5). Development was stopped by washing the arrays in deionized water.

Quantification of the Data

Developed microarrays were scanned at 2400 dpi (CanoScan 8800F, Canon, Søborg, Denmark), converted to TIFF files, and antibody and CBM signals were quantified using Array-Pro Analyzer 6.3 (Media Cybernetics, Rockville, USA). Results were presented in a heatmap, where the maximal mean spot signal was set to 100 and the rest of values

normalized accordingly. A cutoff of 5 was implemented, and color intensity was correlated to mean spot signals.

Chemometrics

Chemometric modeling was performed using MATLAB (version 2012b, The Mathworks Inc., MA, USA) and PLS Toolbox (version 6.0.1, Eigenvector Research Inc., WA, USA).

Principal Component Analysis

Principal component analysis (PCA) is a descriptive chemometric method which decomposes multivariate datasets into principal components (PCs) using loadings and scores (Hotelling 1933; Wold et al. 1987). Principal components represent orthogonal projections in the original multidimensional variable space which ideally characterize the main features of a data set. Normally, a low number of principal components is sufficient to express the measured variance of a (chemical) system, e.g., the IR spectral changes recorded during a time-resolved chemical extraction. The established PCA models were used for exploratory purposes only and were not validated (no test set).

Partial Least Squares Regression

In contrast to PCA, Partial Least Squares regression is a supervised decomposition method wherein a *Y*-block, typically containing reference analysis data, is used to find projections in the original multidimensional variable space which is suitable for prediction of desired parameters (Beebe and Kowalski 1987). Illustratively spoken, it rotates principal components, e.g., retrieved from PCA, in space to find maximal covariance between the *X*-block (multivariate dataset containing independent variables) and the *Y*-block (containing known reference values) while minimizing the residuals in a least squares sense. The resulting projections lead to latent variables which allow prediction of one or more parameters from a newly measured sample. PLS calibrations can be employed to reveal data patterns from complex data, such as FTIR spectra or carbohydrate microarray heat maps.

To obtain robust models using PLS, a number of decision criteria was required to find out about how many latent variables were necessary to describe the dataset sufficiently (Geladi et al. 1985). Leave-one-out cross validation was used to choose the appropriate amount of latent variables for PLS. In addition, extraction II samples were used as a prediction test set during the study. Thus, validation of the PLS models was achieved by calculating the root mean square error of prediction (RMSEP) for the test

set samples. The RMSEP was calculated using Eq. 1, where y_i and \hat{y}_i denote actual and model predicted pectin yield, respectively.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

Pre-Processing

Pre-processing of the data was necessary to remove irrelevant systematic variation to gain better model prediction performance and to decrease the required number of components necessary for the analysis (due to removal of undesirable offsets) (Barnes et al. 1989; Berg et al. 2006; Bro and Smilde 2003). The FTIR data were normalized using standard normal variate (SNV) and mean centered. The carbohydrate microarray data were auto-scaled and mean centered. While SNV is typically used for spectroscopic techniques to remove light scattering effects, auto-scaling is used when variables indicate different units.

Results

Results for FTIR and carbohydrate microarrays are reported in separate sections. The results are thereafter compared in relation to the HPAEC results in the “Discussion” section.

Pectin Yield, High-Performance Liquid Chromatography, and Colorimetric Results

Pectin yields (%) determined by propan-2-ol precipitation are presented in Table 1. As expected, the pectin yield increased with extraction time in all three extraction cases. The HPAEC-PAD and colorimetric results of the final pectins (Table 2) showed that both extraction methods delivered a pectin

product very rich in galacturonic acid—and since both methodologies gave a higher than 65 % by weight of galacturonic acid in the precipitate, both methods were in essence valid pectin extraction methods. The results will be utilized for comparison in the following sections.

Fourier transform infrared spectroscopy

All spectra of extractions I, II, and III are shown in Fig. 1 as difference spectra. To calculate difference spectra, the initial spectrum (first sample) was subtracted from the entire series to remove constant background signals. Therefore, the spectrum of sample 1 resulted in a zero line, while spectra at later stages evolved according to the continuous release of soluble pectin and oligosaccharides during extraction. The elapsing time was illustrated by gray scale shifting from dark to light. For comparison, benchmark spectra of arabinose, xylose, and galactose are included in Fig. 1b, c.

When investigating the spectra of the three extractions, it appeared that the spectral changes of extractions I and II were similar (Fig. 1a, b). These two extractions were performed using enzymatic hydrolysis, while the spectra of extraction III (acid) evolved differently (Fig. 1c). Hence, different molecular properties of the extracted pectin/oligosaccharides in the two types of crude extract were expected.

Individual spectral bands for pectin, including the backbone and side group vibration bands, are identified in Fig. 1a. The strong evolving peaks at 1022, 1106, and less abundant at 1149 cm^{-1} (a) were due to the high HG content (specifically due to C-C, C-O, CCH, OCH vibrations) as reported previously (Coimbra et al. 1999; Fellah et al. 2009; Kacurakova et al. 2000; Wellner et al. 1998). The band at 1149 cm^{-1} is characteristic for C-O-C vibrations of glycosidic linkages (+glycosidic ring) (Wellner et al. 1998) and is therefore also characteristic for pectin backbone structures. Bands at 1045 and 1076 cm^{-1} (b) were interpreted to be mainly caused by neutral sugars, such as arabinose, xylose, and

Table 2 Chemical composition of enzyme (extraction II) and acid (extraction III) extracted pectin samples after 240 min incubation

	Galacturonic acid (%DM)	Arabinose (%DM)	Fucose (%DM)	Galactose (%DM)	Glucose (%DM)	Rhamnose (%DM)	Xylose (%DM)	DE (%)
Enzyme extracted pectin	81.3	9.3	0.1	1.7	1.6	0.8	0.7	82
Acid extracted pectin	90.1	3.3	0.1	3.9	0.4	1.4	0.2	67
F31	74.5 ^a	0.2	0.1	3.3	0.1	1.0	0.1	31
E81	78.7 ^a	0.2	0.1	4.0	0.2	1.3	0.1	81
RGI from soybean	51 ^b	3.4 ^b	10.3 ^b	12.3 ^b	n.d.	6.4 ^b	13.7 ^b	n.d.

Samples E81, F31, and RGI were added for comparison

^a Data from Limberg et al. (2000)

^b Data from the manufacturer specification sheet

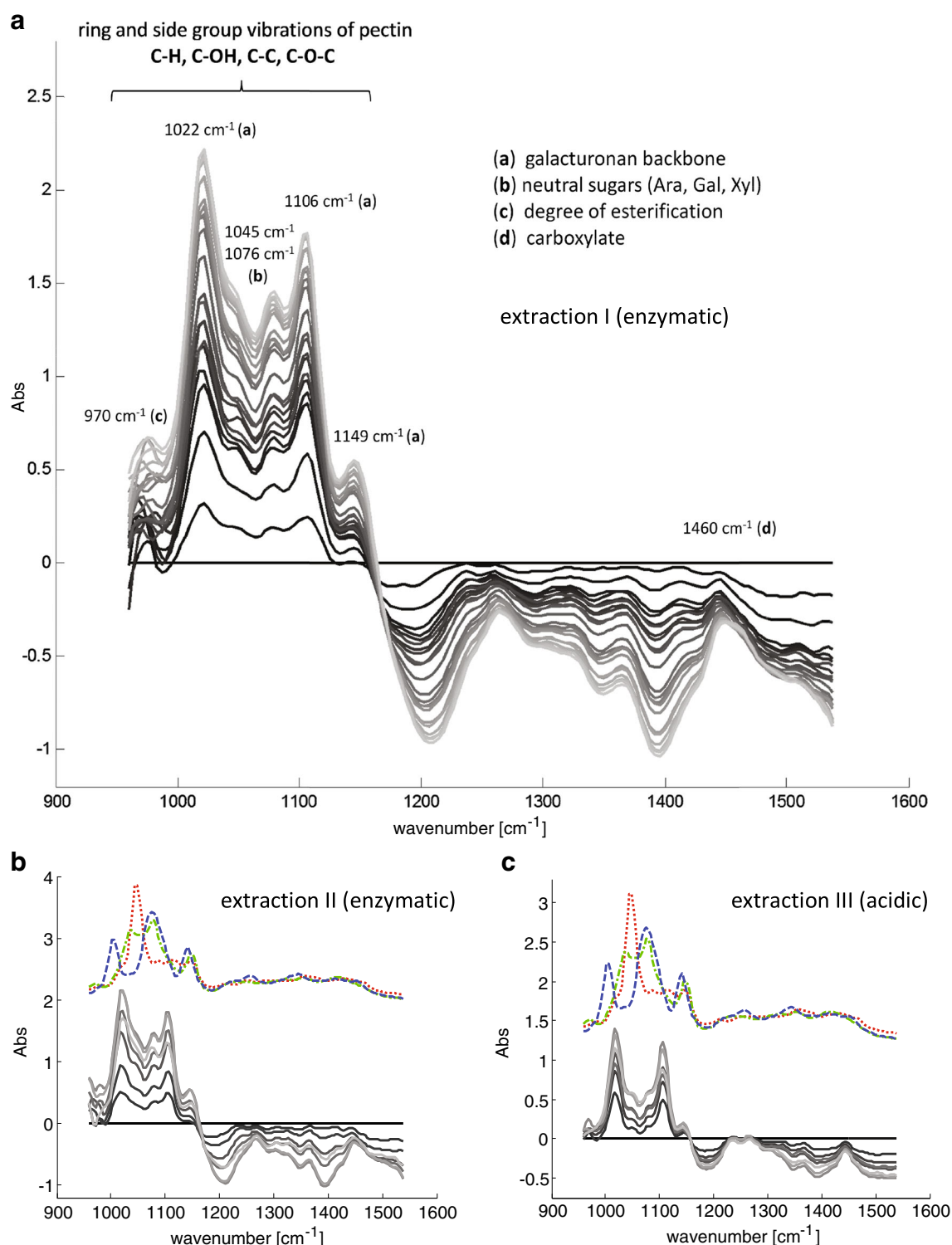


Fig. 1 FTIR spectra of crude pectin extracts: **a** enzymatic extraction I, **b** enzymatic extraction II, and **c** acidic extraction III. Changing gray scale from dark to light indicates elapsing extraction time. Spectra are displayed as difference spectra to highlight changes during extraction

and to eliminate background signals. Monosaccharide standard spectra are included in **b** and **c** for comparison (dotted—xylose; dashed—arabinose; dashed dot—galactose)

galactose (Coimbra et al. 1998). Higher amounts of neutral sugars were detected in the enzymatic crude extracts. This was especially evident when looking at the intensity ratios of

the two band groups (a) vs. (b) as shown in Fig. 1a–c. The intensities of the bands (b) were relatively high for enzymatic extracted samples, which suggested higher amounts of neutral

sugars in the extracts (Fig. 1a, b), as compared to acidic extracts (Fig. 1c). As neutral sugars are mainly found in side chains of pectin rhamnogalacturonan I (RGI) structural domains, these bands indicated “hairy regions” in the pectin molecules. Alternatively, the bands could have been a response caused by free neutral sugars or shorter oligomers, which could have independently coexisted in the crude extract.

An additional band at 970 cm^{-1} (c) indicated the degree of esterification. Accordingly, at 1460 cm^{-1} (d), one can observe a carboxylate band which tends to be weaker, when highly esterified pectin is present (Coimbra et al. 1998, 1999; Fellah et al. 2009; Kačuráková et al. 1999; Wellner et al. 1998). It could be concluded that the degree of esterification (DE) of the enzymatic extracted pectin was higher than that of acidic extracted pectin.

Principal Component Analysis

PCA was used to analyze the spectral changes of the crude pectin samples, extractions I and II in one common model (Fig. 2a) and extraction III in a separate model (Fig. 2b).

Both PCA models required only one principal component to describe 97.3 and 98.6 % of the variance in the spectra, respectively. This was not surprising since the extraction pattern was expected to stay constant over the entire time period of the extraction. The spectral loadings for both PCA models

are given in Fig. 2c, d and appear as expected for neutral sugar rich and less neutral sugar rich extracted material.

As shown in Fig. 1, the spectra converged for all three extractions with elapsing time. This trend was also described by the PC 1 scores which also converged as shown in Fig. 2a, b. In addition, the PC 1 scores of both PCA models correlated to a high extend ($R^2 > 0.92$) with the estimated pectin yields determined by propan-2-ol precipitation (Table 2), therefore indicating potential of PCA analysis to assist the optimization of the pectin extraction time. As it can be seen from Fig. 2a, b, the three extractions indicated no additional release of pectin after 120 min.

Partial Least Squares Modeling

Multivariate calibrations were established to predict the pectin yield when measuring the crude extract at different time points during the extraction. PLS models were prepared for extraction I (model 1) and extractions I and III (model 2). In both cases, the extraction II samples were used as a prediction test set for validation of the model acknowledging the fact that they did not contain extraction III samples due to the limited sample size. The calibrations of models 1 and 2 are presented in Fig. 3a, b. Calibration and test set samples are indicated as circles and triangles, respectively. Leave-one-out cross validation indicated the necessity of two and four latent variables for model 1 and model 2, respectively. In the previous section, it

Fig. 2 PC1 scores derived from PCA models plotted against extraction time for **a** enzymatic extractions I (circles) and II (triangles) and for **b**, acidic extraction III. PC1 loadings describing the main spectral features of the enzymatic and acidic crude extracts are presented in **c** and **d**. Characteristic band regions for neutral sugars are marked in both loadings with arrows

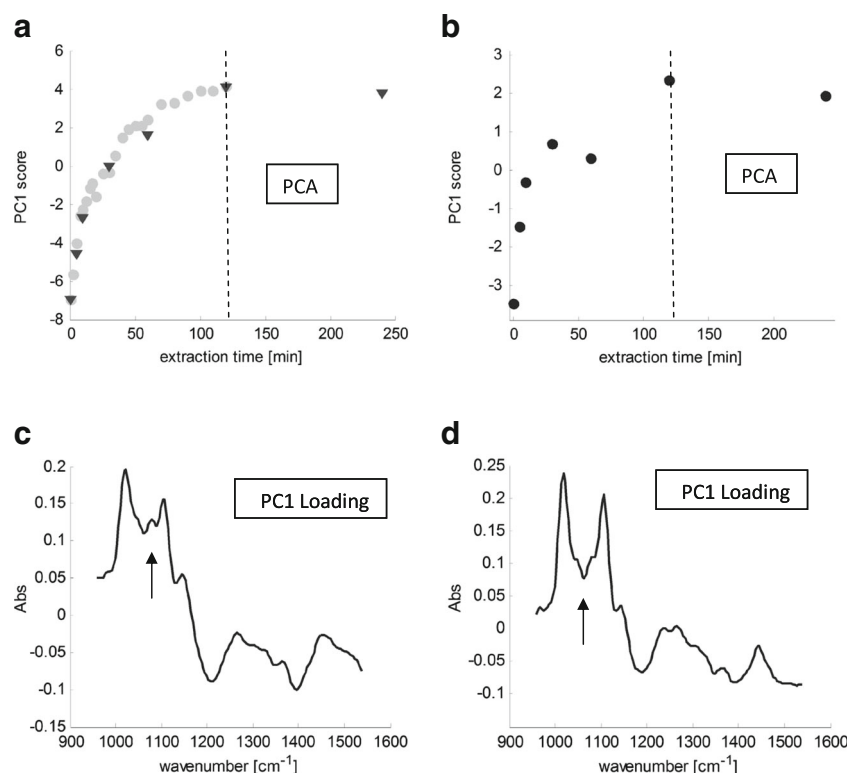
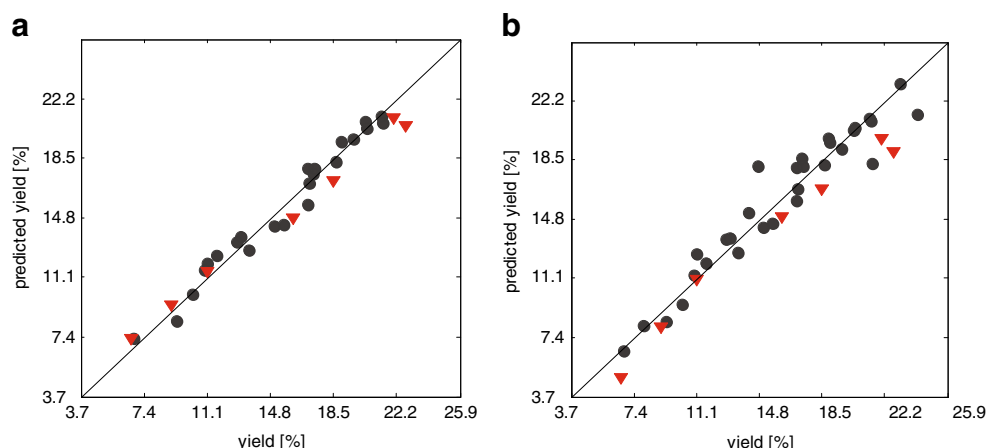


Fig. 3 FTIR pectin yield models: **a** PLS model 1 calibrated on enzymatic extraction I (circles) and tested/validated on enzymatic extraction II (triangles). **b** PLS model 2 calibrated on extractions I and III (enzymatic and acidic extractions, circles) and tested/validated on enzymatic extraction II (triangles)



was discussed that PCA models could be established using only one principal component to represent the spectral variances well. However, minor spectral shifts were present, i.e., for bands around 1160 and 1450 cm^{-1} (Fig. 1). A second latent variable of model 1 compensated for these shifts and resulted in a lower RMSECV. Model 2 required four latent variables as additional spectral variance related to extraction III samples was present. Both models could predict the yield of the extracted pectin with high confidence. Furthermore, the test set samples (validation) correlated with the reference values to a high extend. All figures of merit, including R^2 and RMSEP, are given in Table 3.

Carbohydrate Microarray Analysis

In carbohydrate microarray analysis, antibodies with well-defined sugar-binding epitopes were utilized to obtain information about the sample deposited on the array. The specific binding of the antibody is detected by the use of a fluorescently tagged or enzyme-linked secondary antibody. Within a certain range, the concentration of the deposited epitope covaries with the spot signal produced by the antibody or CBM binding. Such arrays therefore provide direct information about the relative abundance of specific epitopes across the sample set being tested. From the heatmap data displaying the binding responses of 30 different antibodies and CBMs, it was evident that predominantly pectin-related epitopes were detected in

the crude pectin samples, and many of the non-pectin related probes thus produced no signal in the microarray analysis (entire heatmap not shown). A size reduced heatmap of the carbohydrate microarray analysis of the three pectin extractions is shown in Fig. 4. To evaluate the performance of the microarray, three pectin samples were included as positive controls, namely E81 (a), F31 (b), and RGI (c). Further information on structure and composition of these samples is given in Table 2.

Assessment of the responses on the antibodies for pectin in the microarray revealed clear differences in the profiles of the enzymatic extractions I and II and the acidic extraction III (Fig. 4). All crude extracts obtained during the acidic extraction III thus showed interactions with monoclonal antibody (mAb) JIM5 (Clausen et al. 2003). JIM5 binds to unesterified and partially methyl-esterified HG having preference in binding low methylated pectins. In addition, JIM5 also recognized the epitopes of the control pectin (b, F31) with high affinity, confirming the presence of a HG structure with lower degree of esterification in the acidic extracted pectin. In contrast, no JIM5 signal was detected for any of the enzymatically extracted samples. Instead, relatively strong signals from mAb JIM7 (Clausen et al. 2003) and LM20 (Verherbruggen et al. 2009) were found. JIM7 and LM20 are binders of methyl-esterified HG binding preferably to high methylated pectins. This suggests that the enzymatic crude extracts contained pectin with a relatively high DE. Data also indicated that the acid extracted pectin contained both low and high DE pectin and that it was much more heterogeneous in terms of DE than that of the enzymatic extracted pectin. The high DE (82 %) of enzymatically extracted pectin and the low DE (67 %) of the acidic extracted pectin (Table 2) corroborated the microarray data. The heatmap results also confirmed that acidic extraction and low pH are not a preferable combination for achieving homogeneous pectin with high DE values.

Moreover, the carbohydrate microarray results suggest that acidic extracted pectin contained additional regions with lower degree of esterification, which could not be detected in the

Table 3 Figures of merit for PLS models

	FTIR		Microarray	
	Model 1	Model 2	Model 3	Model 4
R^2_{cal}	0.97	0.93	0.89	0.88
R^2_{pred}	0.99	0.98	0.91	0.73
RMSEP [%]	1.19	1.70	4.30	4.74

RMSEP root mean square error of prediction

	Extraction I										
	0	56	0	0	41	0	15	7	7	0	
	0	63	0	0	40	0	9	5	5	2.5	
	0	63	0	0	47	0	11	6	8	5	
	0	62	0	0	46	0	10	0	8	7.5	
	0	71	0	0	47	0	11	5	8	10	
	0	67	0	0	46	0	9	6	8	12.5	
	0	63	0	0	49	0	9	5	8	15	
	0	61	0	0	47	0	7	0	8	17.5	
	0	69	0	0	45	0	8	0	6	20	
	0	71	0	0	47	0	8	0	7	25	
	0	66	0	0	46	0	5	0	6	30	
	0	68	0	0	49	0	6	0	8	35	
	0	70	0	0	48	0	9	0	8	40	
	0	75	0	0	49	0	7	0	10	45	
	0	76	0	0	51	0	6	0	9	50	
	0	76	0	0	52	0	6	0	9	55	
	0	81	0	0	49	0	8	0	8	60	
	0	73	0	0	47	0	0	0	7	70	
	0	82	0	0	51	0	5	0	9	80	
	0	73	0	0	55	0	0	0	8	90	
	0	74	0	0	42	0	5	0	6	100	
	0	81	0	0	51	0	0	0	8	110	
	0	77	0	0	49	0	0	0	8	120	
	Extraction II										
	0	53	0	0	37	0	14	6	7	0	
	0	63	0	0	45	0	17	10	9	5	
	0	64	0	0	46	0	13	7	9	10	
	0	75	0	0	53	0	12	7	11	30	
	0	70	0	0	53	0	10	0	11	60	
	0	60	0	0	45	0	6	0	7	120	
	0	73	0	0	48	0	0	0	7	240	
	Extraction III										
	18	39	0	0	34	0	7	6	0	0	
	29	81	0	0	72	18	39	33	26	5	
	22	66	0	0	58	12	28	22	19	10	
	28	78	0	0	67	26	41	35	22	30	
	25	76	0	0	66	29	40	37	23	60	
	23	55	0	0	48	26	30	26	12	120	
	24	69	0	0	63	35	43	39	19	240	
Controls											
	50	65	18	0	50	0	10	39	0	a	
	100	48	77	48	25	0	19	34	0	b	
	14	30	0	46	5	56	50	34	23	c	

Fig. 4 Crude pectin extracts from extractions I to III were printed as microarrays and probed with a panel of cell wall glycan directed monoclonal antibodies (mAbs) and carbohydrate binding modules (CBMs). Binding results of a selection of pectin-related monoclonal antibodies are presented in this *heatmap* in which color intensity was correlated to mean spot signal values. Antibody names and their corresponding epitopes (recognized by a particular antibody) are depicted at the *top* of the heatmap. Each sample extraction time is shown on the *right side* of the heatmap. Controls include **a** a commercially produced pectin with degree of esterification (DE) 81 % extracted from lime peel (E81), **b** a pectin with DE 31 % derived from E81 by treatment with *Aspergillus niger*'s pectin methylesterase, and **c** RGI from soy bean

enzymatically extracted pectin. The lower degree of esterification of acid extracted pectin was not surprising as low pH enhances the hydrolysis of ester bonds and can, therefore, reduce the degree of esterification in acidically extracted pectin (Rolin et al. 1998). Another explanation might be that a low level of polygalacturonase activity, present in the Laminex C2K preparation (unpublished data), induced degradation of the non-esterified HG stretches, thereby leaving highly esterified HG available for binding to the nitrocellulose membrane.

Partial Least Squares Modeling

When analyzing the obtained heatmap signals column-wise for each of the 30 antibodies/CBMs (Supplementary Fig. S1), none could be directly identified to correlate with the extracted amount of pectin. However, when combining all 30 antibody and CBM responses using multivariate PLS regression, a correlation did occur. As performed for FTIR, two models were established. Model 3 (Fig. 5a) was calibrated on extraction I samples, while model 4 (Fig. 5b) was calibrated on extractions I and III samples. In both cases, the extraction II samples served as the prediction test set. High R^2 values of both models indicated strong correlations in the data sets. All figures of merit are summarized in Table 3. While only two latent variables were used for modeling the enzymatic extraction (model 3), five were necessary to establish a combined model for acidic and enzymatic extraction (model 4). This was not surprising when looking at the microarray heatmap (Fig. 4), showing that the signal pattern was very different for the enzymatic and the acidic extractions. Especially since JIM5 and antibodies recognizing RG I regions gave strong responses mainly for samples from the acidic extraction, which led to a higher rank in the data.

Discussion

Multivariate Analysis

FTIR and carbohydrate microarray analysis could predict the pectin yield well using PLS modeling. PLS models 2 and 4, which included the crude acidic extraction III samples, had

lower prediction performance for both FTIR and microarray analysis. In particular, the yield prediction ability for carbohydrate microarray data decreased, when including the crude acidic extraction III. The higher RMSEP did result not only from a worse precision but also from a bias induced by the prediction samples itself. This bias was evident by the fact that both carbohydrate microarray models (3 and 4) predicted all test set samples (extraction II) with an intercept offset (Fig. 5). This bias was not recognizable for predictions using the FTIR models 1 and 2 (Fig. 3). It should be pointed out at this point that the sample size was limited and further efforts will, therefore, be necessary to validate and consolidate the results. Nonetheless, the study showed in principle that both techniques could be used to establish quantitative calibration models to predict the final pectin yield from crude samples measured during the extractions.

Structural Characterization

FTIR and carbohydrate microarrays provided comprehensive understanding of structural properties of the crude pectin extracts even prior to precipitation. However, certain contemplations in the results were present. While FTIR detected high amounts of neutral sugars in the enzymatic extracted samples, which was in agreement with compositional analysis of precipitated pectins (Table 2) and literature data (Nielsen 1996; Zykwinska et al. 2008), carbohydrate microarray analysis suggested lower amounts of RGI and neutral sugars in those extracts. In fact, carbohydrate microarray analysis detected a low level of arabinan by LM6 (Willats et al. 1998) and neither RGI backbone by INRA-RU1 and INRA-RU2 (Ralet et al. 2010) nor galactan by LM5 (Jones et al. 1997) in the final enzymatically extracted pectin sample (240 min extraction). However, the crude extracts investigated in this study probably contained oligomers with a sufficient degree of polymerization (DP) to be precipitated by propan-2-ol. On the other hand, the oligomers must have been too short to bind to the nitrocellulose membrane and, hence, were not detected by

microarray analysis. The results, furthermore, suggested that the enzymatic cocktail decreased the amount of neutral sugars in the enzymatically extracted pectin over time, as the antibody responses of the monoclonal antibodies INRA-RU1 (RGI backbone) and LM5 (galactan) decreased with increasing extraction time. The presence of pectinase activity in the used Laminex C2K preparation would explain the hydrolysis of the pectin, yielding more oligomers as extraction time elapsed.

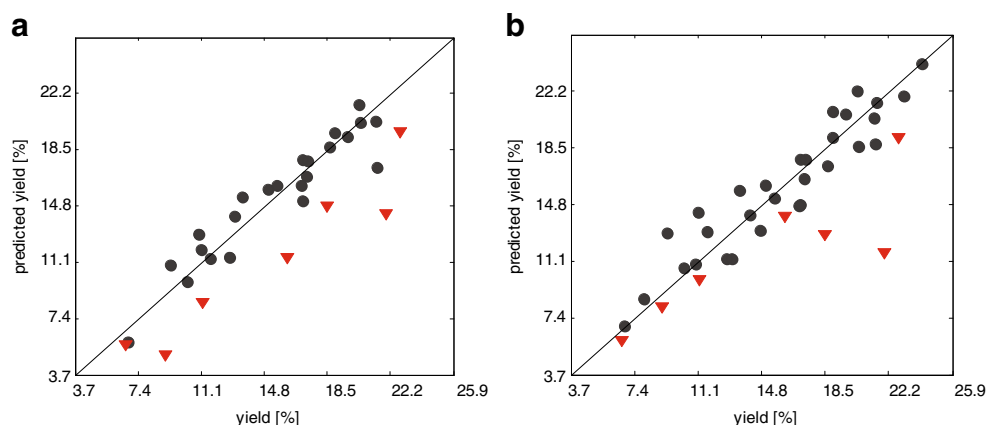
Conclusion

The present study illustrated the potential analytical power of FTIR spectroscopy and carbohydrate microarray analysis in a new application namely for direct assessment of crude pectin extracts during enzymatic and acidic extraction. Two aspects have been investigated and compared with respect to the two techniques, namely the ability to predict final pectin yield (and optimal extraction time) from crude samples taken during the extraction and the ability to predict the structural quality and complexity of final pectin product.

While both techniques were able to predict the respective pectin yield during the extractions, especially FTIR showed high accuracy and precision. Furthermore, simple PCA on FTIR data indicated possibilities to determine the optimal extraction time, even without the need of interrupting the process. Rapid and non-invasive FTIR spectroscopy, therefore, enables possibilities for online monitoring of pectin extraction processes and has great potential to replace cumbersome and time consuming quantitative chemical analysis.

Secondly, both techniques gave insight into structural features of the extracted pectins. The results confirmed major differences between the enzymatically and the acidically extracted products with respect to degree of esterification and abundance of RGI pectin regions. Due to its high-throughput capability, carbohydrate microarray analysis has proven to be

Fig. 5 Carbohydrate microarray pectin yield models: **a** PLS model 3 calibrated on enzymatic extraction I (circles) and tested/validated on enzymatic extraction II (triangles). **b** PLS model 4 calibrated on extractions I and III (enzymatic and acidic extractions, circles) and tested/validated on enzymatic extraction II (triangles)



a remarkable tool to assess structural features of the extracts that in turn can be used for pectin quality prediction.

Hence, both FTIR and carbohydrate microarray analysis have potential to move pectin analysis closer to pectin production, which is particularly important in industrial process and quality control.

Acknowledgments This work was financed by the Seventh Framework Program via the Marie Curie Initial Training Network, LeanGreenFood (EU-ITN 238084).

References

- Andersen, S. K., Hansen, P. W., & Andersen, H. V. (2006). Vibrational Spectroscopy in the Analysis of Dairy Products and Wine. Handbook of Vibrational Spectroscopy. doi:10.1002/0470027320.s6602.
- Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772–777.
- Baum, A., Hansen, P. W., Meyer, A. S., & Mikkelsen, J. D. (2013a). Simultaneous measurement of two enzyme activities using infrared spectroscopy: a comparative evaluation of PARAFAC, TUCKER and N-PLS modeling. *Analytica Chimica Acta*, 790, 14–23.
- Baum, A., Meyer, A. S., Garcia, J. L., Egebo, M., Hansen, P. W., & Mikkelsen, J. D. (2013b). Enzyme activity measurement via spectral evolution profiling and PARAFAC. *Analytica Chimica Acta*, 778, 1–8.
- Beebe, K. R., & Kowalski, B. R. (1987). An introduction to multivariate calibration and analysis. *Analytical Chemistry*, 59(17), 1007A–1017A.
- Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7(1), 1.
- Blumenkrantz, N., & Asboe-Hansen, G. (1973). New method for quantitative determination of uronic acids. *Analytical Biochemistry*, 54(2), 484–489.
- Boukroufa, M., Boutekdjiret, C., Petigny, L., Rakotomanomana, N., & Chemat, F. (2015). Bio-refinery of orange peels waste: a new concept based on integrated green and solvent free extraction processes using ultrasound and microwave techniques to obtain essential oil, polyphenols and pectin. *Ultrasonics Sonochemistry*, 24, 72–79.
- Bro, R., & Smilde, A. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, 17(1), 16–33.
- Ciriminna, R., Chavarria-Hernández, N., Inés Rodríguez Hernández, A., & Pagliaro, M. (2015). Pectin: a new perspective from the biorefinery standpoint. *Biofuels, Bioproducts and Biorefining*, 9(4), 368–377.
- Clausen, M. H., Willats, W. G. T., & Knox, J. P. (2003). Synthetic methyl hexagalacturonate hapten inhibitors of anti-homogalacturonan monoclonal antibodies LM7, JIM5 and JIM7. *Carbohydrate Research*, 338(17), 1797–1800.
- Coimbra, M. A., Barros, A., Barros, M., Rutledge, D. N., & Delgadillo, I. (1998). Multivariate analysis of uronic acid and neutral sugars in whole pectic samples by FT-IR spectroscopy. *Carbohydrate Polymers*, 37(3), 241–248.
- Coimbra, M. A., Barros, A., Rutledge, D. N., & Delgadillo, I. (1999). FTIR spectroscopy as a tool for the analysis of olive pulp cell-wall polysaccharide extracts. *Carbohydrate Research*, 317(1), 145–154.
- de Paz, J. L., & Seeberger P. H. (2012). Recent advances and future challenges in glycan microarray technology. In Y. Chevotot (Ed.), Carbohydrate Microarrays: Methods and Protocols. *Methods in Molecular Biology* (pp 1–12). Humana Press.
- Dominiak, M., Wichmann, J., Vidal Melgosa, S., Willats, W., Meyer, A., Søndergaard, K., & Mikkelsen, J. (2014). Application of enzymes for efficient extraction, modification, and development of functional properties of lime pectin. *Food Hydrocolloids*, 40, 273–282.
- Fangel J. U., Pedersen H. L., Vidal-Melgosa S., Ahl L. I., Salmean A. A., Egelund J., et al. (2012) Carbohydrate microarrays in plant science. In J. Normanly (Ed.), High-Throughput Phenotyping in Plants: Methods and Protocols. *Methods in Molecular Biology* (pp 351–62). Humana Press
- Fellah, A., Anjukandi, P., Waterland, M. R., & Williams, M. A. K. (2009). Determining the degree of methylesterification of pectin by ATR/FT-IR: methodology optimisation and comparison with theoretical calculations. *Carbohydrate Polymers*, 78(4), 847–853.
- Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, 39(3), 491–500.
- Grassino, A. N., Brnčić, M., Vikić-Topić, D., Roca, S., Dent, M., & Brnčić, S. R. (2016). Ultrasound assisted extraction and characterization of pectin from tomato waste. *Food Chemistry*, 198, 93–100.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417.
- Jones, L., Seymour, G. B., & Knox, J. P. (1997). Localization of pectic galactan in tomato cell walls using a monoclonal antibody specific to (1 → 4)-[β]-D-galactan. *Plant Physiology*, 113(4), 1405–1412.
- Kacurakova, M., Capek, P., Sasinkova, V., Wellner, N., & Ebringerova, A. (2000). FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. *Carbohydrate Polymers*, 43(2), 195–203.
- Kačuráková, M., Wellner, N., Ebringerová, A., Hromádková, Z., Wilson, R., & Belton, P. (1999). Characterisation of xylan-type polysaccharides and associated cell wall components by FT-IR and FT-Raman spectroscopies. *Food Hydrocolloids*, 13(1), 35–41.
- Kuligowski, J., Cascant, M., Garrigues, S., & de la Guardia, M. (2012). An infrared spectroscopic tool for process monitoring: sugar contents during the production of a depilatory formulation. *Talanta*, 99, 660–667.
- Lim, J., Yoo, J., Ko, S., & Lee, S. (2012). Extraction and characterization of pectin from Yuza (Citrus junos) pomace: a comparison of conventional-chemical and combined physical-enzymatic extractions. *Food Hydrocolloids*, 29(1), 160–165.
- Limberg, G., Körner, R., Buchholt, H. C., Christensen, T. M. I. E., Roepstorff, P., & Mikkelsen, J. D. (2000). Analysis of different de-esterification mechanisms for pectin by enzymatic fingerprinting using endopectin lyase and endopolygalacturonase II from *A. niger*. *Carbohydrate Research*, 327(3), 293–307.
- Nicolaou, N., Xu, Y., & Goodacre, R. (2010). Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. *Journal of Dairy Science*, 93(12), 5651–5660.
- Nielsen, B. U. (1996). Fiber-based fat mimetics: pectin. In S. Roller & S. A. Jones (Eds.), *Handbook of fat replacers* (pp. 161–173). Boca Raton: CRC Press.
- Ralet, M.-C., Tranquet, O., Poulain, D., Moïse, A., & Guillon, F. (2010). Monoclonal antibodies to rhamnogalacturonan I backbone. *Planta*, 231(6), 1373–1383.
- Rohman, A., & Man, Y. B. C. (2010). Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. *Food Research International*, 43(3), 886–892.
- Rolin, C., Nielsen, B. U., & Glahn, P. E. (1998). In S. Dumitriu (Ed.), *Polysaccharides—structural diversity and functional versatility* (pp. 377–431). New York: Marcel Dekker Inc.
- Rudolph B., & Petersen S. A. (2011). US Patent App. 13/013,709 Pub: US20120190831 A1. Methods for Steam Flash Extraction of Pectin. Cp Kelco Aps, assignee. United States

- Sørensen, I., & Willats, W. G. T. (2011). Screening and characterization of plant cell walls using carbohydrate microarrays. In Z. Popper (Ed.), *The plant cell wall: methods and protocols. Methods in Molecular Biology* (pp 115–121). India: Humana Press.
- van den Hoogen, B. M., van Weeren, P. R., Lopes-Cardozo, M., van Golde, L. M. G., Barneveld, A., & van de Lest, C. H. A. (1998). A microtiter plate assay for the determination of uronic acids. *Analytical Biochemistry*, 257(2), 107–111.
- Verhertbruggen, Y., Marcus, S. E., Haeger, A., Ordaz-Ortiz, J. J., & Knox, J. P. (2009). An extended set of monoclonal antibodies to pectic homogalacturonan. *Carbohydrate Research*, 344(14), 1858–1862.
- Voragen, F., Beldman, G., & Schols, H. (2008). Chemistry and enzymology of pectins. *Advanced dietary fibre technology*, 379–397. doi:10.1002/9780470999615.ch33.
- Voragen, A. G. J., Schols, H. A., & Pilnik, W. (1986). Determination of the degree of methylation and acetylation of pectins by h.p.l.c. *Food Hydrocolloids*, 1(1), 65–70.
- Wellner, N., Kačuráková, M., Malovíková, A., Wilson, R. H., & Belton, P. S. (1998). FT-IR study of pectate and pectinate gels formed by divalent cations. *Carbohydrate Research*, 308(1), 123–131.
- Wikiera, A., Mika, M., & Grabacka, M. (2015). Multicatalytic enzyme preparations as effective alternative to acid in pectin extraction. *Food Hydrocolloids*, 44, 151–161.
- Willats, W. G. T., Marcus, S. E., & Knox, J. P. (1998). Generation of a monoclonal antibody specific to (1→5)- α -L-arabinan. *Carbohydrate Research*, 308(1), 149–152.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52.
- Zykwinska, A., Boiffard, M.-H., Kontkanen, H., Buchert, J., Thibault, J.-F., & Bonnin, E. (2008). Extraction of green labeled pectins and pectic oligosaccharides from plant byproducts. *Journal of Agricultural and Food Chemistry*, 56(19), 8926–8935.

PAPER 3

Andreas Baum, Anne S. Meyer, Javier L. Garcia, Max Egebo, Per W. Hansen, Jørn D. Mikkelsen; Enzyme activity measurement via spectral evolution profiling and PARAFAC; *Analytica Chimica Acta*, Volume 778, 17 May 2013, Pages 1-8



Enzyme activity measurement via spectral evolution profiling and PARAFAC



Andreas Baum^{a,b}, Anne S. Meyer^a, Javier Lopez Garcia^a, Max Egebo^b,
Per Waaben Hansen^b, Jørn Dalgaard Mikkelsen^{a,*}

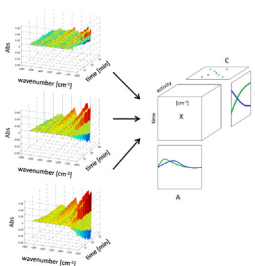
^a Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

^b Foss Analytical, Foss Allé 1, DK-3400 Hillerød, Denmark

HIGHLIGHTS

- Enzyme activity determination using FTIR and PARAFAC.
- Directly and universally applicable without the need of any external standards.
- The method provides alternative solutions to the cumbersome reducing sugar analysis.
- Enzymatic hydrolysis of plant polymers and sugar release can be determined.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 18 January 2013

Received in revised form 5 March 2013

Accepted 12 March 2013

Available online 21 March 2013

Keywords:

Chemometrics

Fourier transform infrared spectroscopy

Enzyme kinetics

Substrate evolution

Product evolution

Multiway

ABSTRACT

The recent advances in multi-way analysis provide new solutions to traditional enzyme activity assessment. In the present study enzyme activity has been determined by monitoring spectral changes of substrates and products in real time. The method relies on measurement of distinct spectral fingerprints of the reaction mixture at specific time points during the course of the whole enzyme catalyzed reaction and employs multi-way analysis to detect the spectral changes. The methodology is demonstrated by spectral evolution profiling of Fourier Transform Infrared (FTIR) spectral fingerprints using parallel factor analysis (PARAFAC) for pectin lyase, glucose oxidase, and a cellulase preparation.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Quantitative assessment of enzyme activity is a fundamental measurement required for determining the amount of an enzyme. The enzyme activity is defined as the amount of active enzyme which will catalyze the transformation of a particular molar amount of substrate per time unit under standard conditions. Its measurement is decisive for defining enzyme dosage, reaction

time, substrate use and product yields in practical enzyme catalyzed reactions. A quantitative enzyme activity assay – optimally measured directly as the reaction rate defining the initial substrate consumption rate – is also the fundamental measurement used for studying enzyme action, specificity, kinetics, possible inhibition, and for identifying a specific enzyme.

Existing large scale industrial enzyme based processes range from specific conversions of crude penicillins to defined β -lactam antibiotics [1] to multi-enzymatic degradation of cellulose and other biomass materials for novel bio-based fuel production and biorefinery products [2]. Current activity assay methods for biomass and polysaccharide degrading enzymes are particularly

* Corresponding author. Tel.: +45 60856300.

E-mail address: jdm@kt.dtu.dk (J.D. Mikkelsen).

cumbersome because of the complexity of the substrates. The currently used enzyme assays therefore rely on the use of artificial substrates and assessment of their conversion by laborious spectrophotometric techniques, which may only measure the reaction indirectly, or are based on crude measurements such as reducing end analyses, or employ advanced chromatography-based analyses of changes in the substrate or product levels [3–5]. A main disadvantage of all existing enzyme assays is that they rely on measurement of either the product formation or the substrate consumption, not both at the same time.

FTIR (Fourier transform infrared) spectroscopy is based on measurement of absorption in the mid-infrared frequency range. All organic molecules have intra- and intermolecular bonds that absorb in this range, and the spectrum obtained therefore reflects the chemical (structural) composition of the sample being analyzed. The absorption pattern of the chemical bonds having an electric dipole moment that changes during vibration will in turn produce a unique fingerprint signature for the biomolecules present in the sample [6]. Although FTIR is a rapid and easy to use method, two major challenges are (a) it is difficult to identify distinct peaks and (b) interference of an inherent strong water background at the mid infrared wavelength range.

FTIR spectroscopy is used commercially to quantify constituents in, e.g. milk and wine utilizing multivariate and advanced statistical approaches to help identify the relevant information in the spectra [7]. FTIR spectroscopy has also been employed previously to determine enzyme activity [8]. In this particular case deuterated water – which exhibits a much lower absorption in the infrared range – was used to eliminate the water interference in the enzyme activity measurements. Alternatively, the enzyme activity for each individual enzyme was determined by monitoring the rate of the decrease in intensity of a specific band: for amidase (EC 3.5.1.4.) and urease (EC 3.5.1.5) the drop in the amide band at 1635 cm⁻¹ or 1605 cm⁻¹, respectively, was thus measured as a function of time [8,9]. FTIR spectroscopy has also been used to assess the influence of enzymatic treatment of a complex substrate, in particular the enzymatic degradation of sea buckthorn berries [10]. However, in order to monitor the enzymatic reaction, only the intensity of single, specific peaks were used to quantify a certain product or substrate [10]. Hence, the available methods have not exploited the full potential of monitoring the full spectral fingerprint. When relying on assessment of only one band, each method becomes highly specific in a classical univariate sense and moreover not applicable to multi-component enzymatic systems with more complex distributions of various substrates and products.

The work presented in this paper was undertaken to test the hypothesis that FTIR and chemometric multiway analysis can be used as a universally applicable approach for rapid assessment of enzyme activity without using any external standards, even on genuine, complex substrates as biopolymers (which may not be chromogenic). A more comprehensive introduction to FTIR in microbiology and examples of chemical applications of multi way chemometrics are given in [11,12], respectively.

2. Theory

Based on the assumption that different substrate(s) and product(s) have distinguishable spectral fingerprints (when measured by FTIR spectroscopy), enzyme kinetics can principally be observed directly in aqueous solutions by monitoring the evolution of the full spectral fingerprint, i.e. by assessing the time resolved changes over the bandwidth from 1000 to 1600 cm⁻¹. During this continuous measurement, as the substrate is consumed, the FTIR spectrum will change simultaneously with the development of the product molecules and as the conversion of substrate to product progresses the spectral profile will thus shape a “landscape” in relation to

enzyme activity. Those spectral landscapes, each identifying the full enzyme catalyzed reaction accomplished at one particular enzyme level, i.e. one particular enzyme activity, are referred to as *Evolution Profiles* and are illustrated for three different enzyme catalyzed reactions with glucose oxidase (EC 1.1.3.4), pectin lyase (EC 4.2.2.1), and a mixed cellulase preparation, respectively (Fig. 1B, D, F and Supplementary Fig. A).

This spectral evolution profiling in real time, combined with chemometric multi-way methodology, i.e. parallel factor analysis (PARAFAC), allows for the quantitative detection of spectral changes of both the substrate and the product during the enzyme catalyzed reaction.

2.1. PARAFAC

PARAFAC [13–15] is a numerical decomposition method suitable for decomposing multilinear datasets into loadings or scores using an alternating least squares algorithm [12]. By doing so PARAFAC finds common profiles/patterns which are present in all samples and therefore enables the advantage of fitting those common profiles to all samples simultaneously. In order to generate a calibration which correlates enzymatic activity with the spectral evolution data a multilinear data structure (tensor) must be formed:

Each *Evolution Profile* is represented by a data matrix where vectors are spectra at consecutive time points. Since all *Evolution Profiles* are recorded at the same consecutive time points those matrices can be stacked to form a tensor which is defined by its three dimensions, namely wave number (K), time (J) and activity (I). The tensor connecting those three measurement dimensions fulfills trilinearity and is a starting point for multi-way analysis (Fig. 2).

The tensor X was mathematically decomposed into three matrices A , B and C (Eq. (1) and Fig. 3) using the PARAFAC algorithm [14]:

$$x_{i,j,k} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{i,j,k} \quad (1)$$

2.2. Data pre-processing

The spectra have been scaled using standard normal variate (SNV). To do so the tensor was unfolded and scaled slab wise across the sample mode as described below [14,16].

$$\bar{x}_{ik} = \frac{\sum_{j=1}^J x_{ijk}}{J} \quad (2)$$

$$w_i = \sqrt{\frac{\sum_{k=1}^K \sum_{j=1}^J (x_{ijk} - \bar{x}_{ik})^2}{KJ - 1}} \quad (3)$$

$$x_{ijk}^{SNV} = \frac{x_{ijk}}{w_i} \quad (4)$$

Additionally the three-way array was consecutively double centered across mode 1 and mode 2. Multi-way centering, e.g. for mode 1 can be understood as unfolding the tensor to a $I \times JK$ matrix and proceeding as in usual two-way analysis [16]:

$$x_{ijk}^{cent} = x_{ijk} - \frac{\sum_{i=1}^I x_{ijk}}{I} \quad (5)$$

2.3. Evolution Profiles

All *Evolution Profiles* from FTIR data have been visualized using surf plots. To display the time-resolved spectra of an enzymatic reaction in this paper all spectra have been SNV scaled slabwise

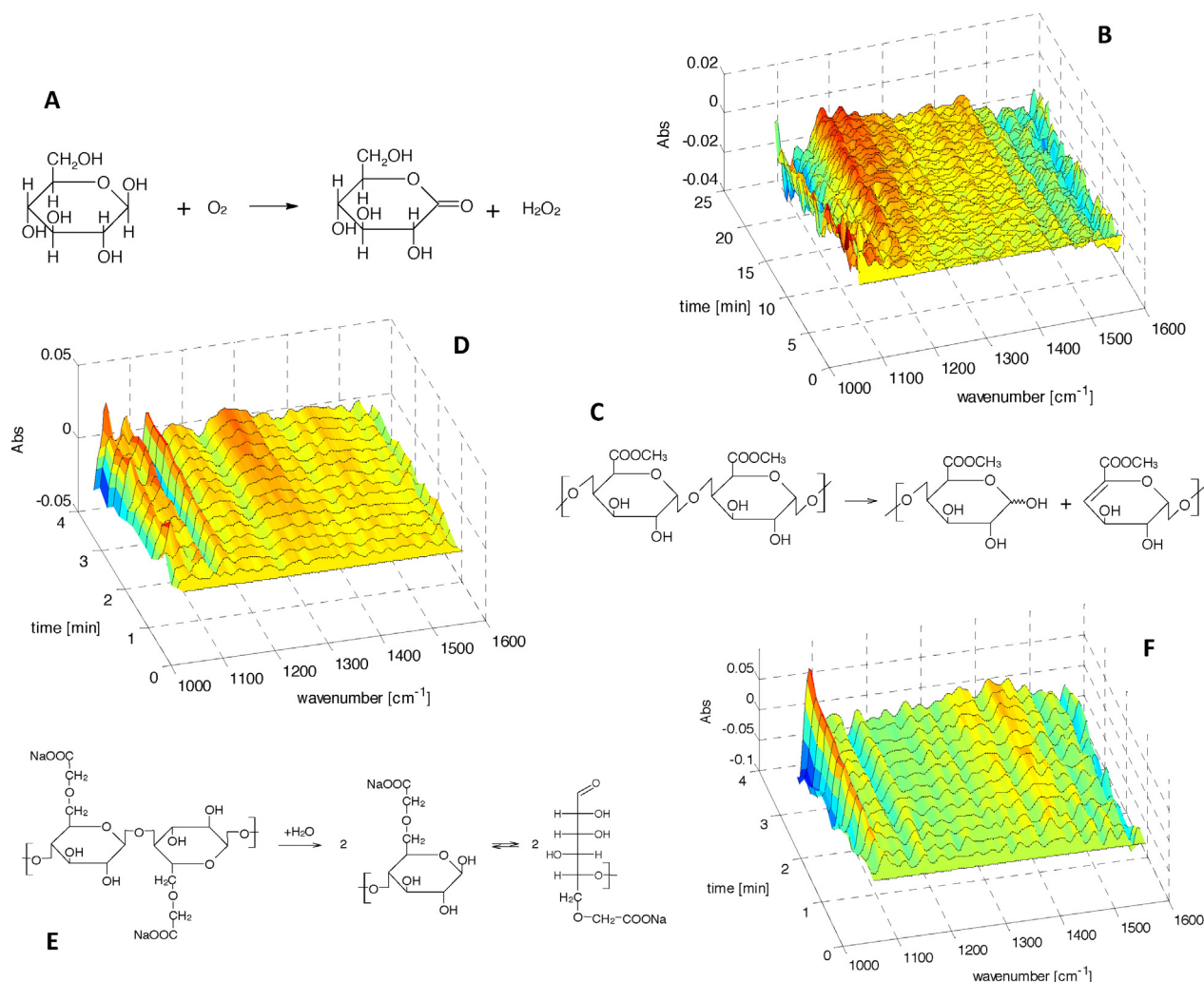


Fig. 1. (A) Enzymatic reaction of glucose oxidase; (B) Spectral Evolution Profile of glucose oxidase visualized using a surf plot. Spectra are displayed as difference spectra to emphasize temporal changes and to disable the visualization of background signals. The observation time was 20.7 min as the signal to noise ratio was low. Substrate consumption (decreasing bands) did not exhibit efficient spectral change; (C) enzymatic reaction of pectin lyase; (D) Spectral Evolution Profile of pectin lyase using 1% apple pectin apple as substrate. Observation time was 4.2 min; (E) enzymatic reaction of a cellulolytic enzyme blend (Celluclast 1.5L); (F) spectral Evolution Profile of Celluclast 1.5L using 1% CMC as substrate. The observation time was 4.2 min.

as described above and difference spectra have been calculated by subtracting the first spectrum from all remaining spectra including the first one. Constant background signals were therefore eliminated as the first difference spectrum results in a vector of zeros (zero absorption over the whole spectral range). Concluding, the following spectra contain signal changes due to the enzymatic reaction.

3. Materials and methods

3.1. Materials

Glucose oxidase from *Aspergillus niger* (TypeX-S), apple pectin and sodium carboxymethylcellulose were purchased from Sigma Aldrich (St. Louis, USA). The pectin lyase gene from *Emericella nidulans* was expressed in *Pichia pastoris* and produced in-house by Center for Bioengineering, DTU, Lyngby, Denmark. The Celluclast 1.5L was donated by Novozymes, Bagsværd, Denmark.

3.2. Experimental

All spectra were obtained using a MilkoScan™ FT2 (FOSS ANALYTICAL, Hillerød, Denmark). The instrument consisted of a FTIR

which scanned the IR spectrum in the range of 1000–1600 cm⁻¹. Acquisition was carried out according to the method described in [17] with an optical resolution of 14 cm⁻¹. The instrument was equipped with an automatic flow-through system apparatus and worked semi-automated. The optical system was hermetically sealed and pressure and humidity controlled. The path length of the cuvette was 50 μm. The temperature was equilibrated at 42 °C.

For measurements the sample was placed underneath the automatic sample intake which was supported by peristaltic pumps. All spectra have been measured against an aqueous blank (FOSS Analytical, Hillerød, Denmark). After each measurement the flow-through system was cleaned automatically. Furthermore the samples possibly were recycled back to the main reaction container in-between time-resolved measurements to avoid oxygen depletion in the cuvette.

3.3. Calibration procedure

To obtain enzyme activity calibrations several steps were performed.

- (1) *Evolution Profiles* of two control mixtures have been acquired. One control contained only the substrate without enzyme.

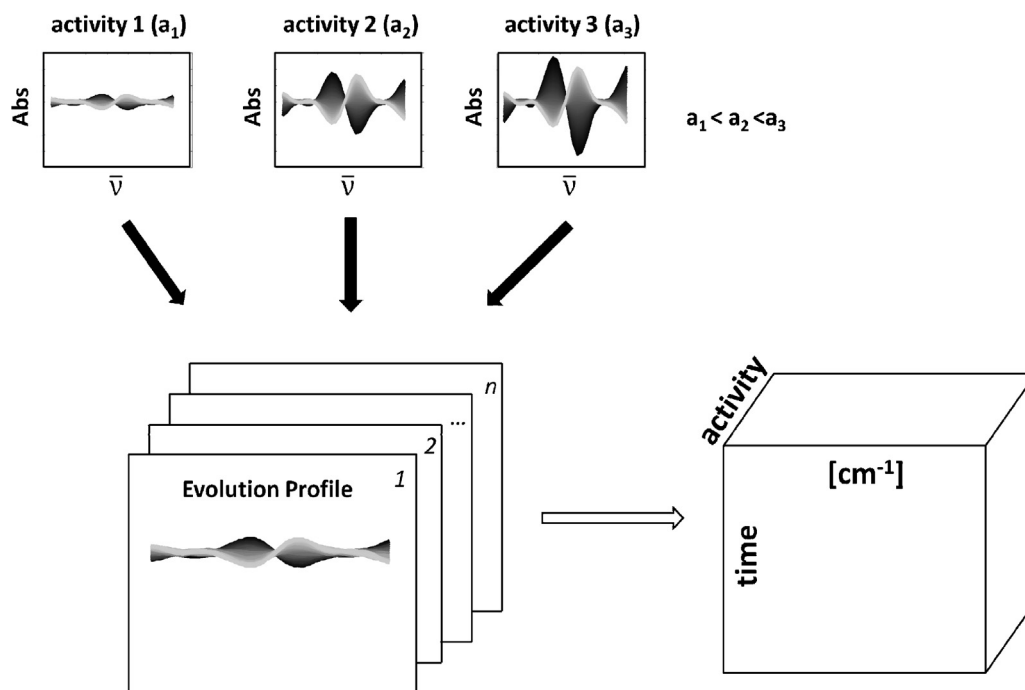


Fig. 2. Conversion from two-way to a three-way data structure. All data matrices containing time-resolved spectra for a certain enzyme activity are stacked behind each other to form a tensor. The cubic data structure fulfills trilinearity since all Evolution Profiles are recorded in the same consecutive time steps.

Buffer was added instead of the enzyme solution. The second control contained only enzyme and no substrate. Instead of substrate solution only buffer was added. The *Evolution Profiles* of both controls showed no significant change in time. That step is especially important to assure that the observed spectral evolution in spectra was due to enzyme activity and not

due to altering circumstances like temperature, precipitation, denaturation, mutarotation etc.

- (2) Various *Evolution Profiles* were acquired using equal substrate concentrations and different dosages of enzyme (activities). Depending on the nature of the enzymatic reaction, measurements have been carried out in flow-back mode where the reaction mixture was continuously led back to the reaction container to ensure access of gases as oxygen which was necessary for the reaction of glucose oxidase. The other enzymatic reactions have been pumped into the cuvette only once for continuous measurements (Pectin Lyase and Celluclast 1.5L). In this case the reaction mixture stayed inside the cuvette during the whole acquisition period. Three replicates have been measured for each calibration point. Spectra for each Evolution Profile were measured consecutively using time steps of 16.6 s for Pectin Lyase and Celluclast 1.5L and 31.0 s for glucose oxidase, respectively. Acquisition of the Evolution Profiles was carried out in random order to prevent systematic biases.
- (3) The acquired spectral data was exported to MATLAB (The Mathworks Inc., MA, USA) where it was treated using designated scripts. PARAFAC analysis was performed using PLS Toolbox 6.0.1 (Eigenvector Research Inc., WA, USA).
- (4) The obtained scores from the PARAFAC decomposition were plotted against the added enzyme activities, respectively. The absolute activities have been determined by conventional colorimetric assays as described below.

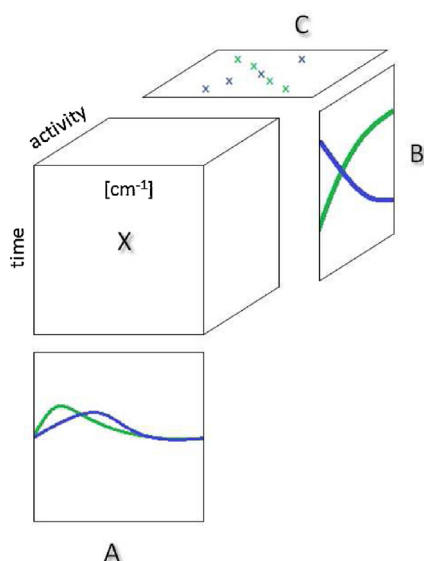


Fig. 3. Schematic illustration of a PARAFAC decomposition of time-resolved enzymatic FTIR data. The matrices contain the following information: (A) this matrix contains the true spectra (fingerprints) of the enzymatic system. Here spectral fingerprints representing both the substrate and the product are illustrated by green and blue color. (B) This matrix contains the kinetic (i.e. time resolved) behavior of substrate(s) and product(s). (C) This matrix contains the information about how abundant the change in spectra in (B) is considering the pure spectra from (A). Those values of that matrix therefore correlate with the used enzyme activity (having opposite slopes for substrate(s) and product(s) if non-negativity constraints are applied to the analysis).

3.4. Colorimetric assays

3.4.1. Pectin lyase

Pectin lyase activity was measured on 1 g L⁻¹ pectin from apple by incubating in 50 mM sodium phosphate Buffer at pH 7.0 and 42 °C. The increase in absorbance was determined at 235 nm during 4 min in an Infinite200 microplate reader (Tecan, Salzburg, Austria); the data collection was controlled by the program Tecan i-control version 1.5.14.0. The extinction coefficient used was 5.5 mM⁻¹ L⁻¹ [18,19].

3.4.2. Glucose oxidase

Incubation conditions were set to pH 6.9 and 42 °C. Samples were taken at relevant time points during a 30 min reaction and the extent of the reaction was determined by the described assay below.

20 μ L of each sample were mixed with 10 μ L o-dianisidine solution (30 mg mL⁻¹), 160 μ L 0.1 M glucose solution and 10 μ L peroxidase solution (1 mg mL⁻¹) to form a yellow-orange dye. The samples were measured photometrically at 420 nm using an Infinite200 microplate reader (Tecan, Salzburg, Austria). 1 unit was defined as the amount of enzyme which catalyzed the formation of 1 μ mol H₂O₂ per minute in 50 mM sodium phosphate buffer at pH 6.9 and 42 °C [20].

3.4.3. Celluclast 1.5L

Incubation conditions were set to pH 5.0 and 42 °C. Samples were taken at relevant time points during a 30 min reaction and the extent of the reaction was determined by reducing sugars.

The procedure for the determination of reducing sugars based on alkaline *p*-hydroxybenzoic acid hydrazide (PAHBAH) is a potentially valuable assay because of its apparent selectivity for reducing sugars, and the limited number of reported interfering substances. The method utilizes the instability of sugars in hot alkaline solution to produce the yellow anionic species of the bis(4-hydroxybenzoylhydrazones) of glyoxal and methylglyoxal in the presence of HBAH [21–23]. The reaction was catalyzed and therefore amplified for small analyte concentrations using bismuth [24].

1 M bismuth, 1 M potassium sodium tartrate, and 3 M sodium hydroxide were mixed with 0.5 M sodium hydroxide and 5% (w/v) 4-hydroxybenzoic acid hydrazide in 0.5 M hydrochloric acid in the ratio 1:899:100. The reagent was mixed with the samples and incubated at 70 °C for 10 min, cooled down to 42 °C, and the absorbance was measured at 410 nm in an Infinite200 microplate reader (Tecan, Salzburg, Austria).

1 U was defined as the amount of enzyme which catalyzed the conversion of 1 μ mol glucose per minute in 50 mM sodium acetate buffer at pH 5.0 and 42 °C.

4. Results and discussion

4.1. Enzyme activity calibrations

For each enzyme calibration several spectral *Evolution Profiles* were acquired using three replicate measurements. The total amounts of *Evolution Profiles* used for the calibrations of glucose oxidase, pectin lyase and Celluclast 1.5L were 29, 32 and 33, respectively.

The *Evolution Profiles*, each measuring a certain enzyme activity (i.e. the specific amount of an enzyme), have been visualized using difference spectra, implying that the first spectrum, and therefore all constant background signals including the water signals, were subtracted from the spectra series. The absorption of certain bands decreased over time due to substrate consumption whereas other bands grew over time due to product formation (Fig. 1B, D, F). Except for glucose oxidase, the products being formed cannot be explicitly identified as a single product, but rather a distribution of products. Nonetheless the various products lead to a combined fingerprint which was measurable by FTIR.

When examining the *Evolution Profiles* it is noticeable that the spectral change depends on the used enzyme dosage (enzyme activity) and the differences are evident for pectin lyase *Evolution Profiles* (Fig. 4). This enzyme dosage dependence was quantified using PARAFAC.

The multi way scores were obtained by decomposing the tensors of each of the three enzymatic systems (Fig. 3) and the

Table 1

Calibration parameters for the three modeled enzymatic systems.

	Pectin lyase	Glucose oxidase	Celluclast 1.5L
Time of spectral evolution	4.2 min	20.7 min	4.2 min
Number of spectra in each Evolution Profile	15	40	15
Calibration range (per mL substrate)	0–200 mU mL ⁻¹	0–6 U mL ⁻¹	0–80 mU mL ⁻¹
LOD ^a	9 mU mL ⁻¹	277 mU mL ⁻¹	3.79 mU mL ⁻¹

^a Limit of detection (LOD) only valid for used time of spectral evolution. LOD decreased for extended observation time.

corresponding enzyme activities were determined by conventional colorimetric and photometric assays.

The plotting against classical units determined by conventional assays (Fig. 5) gives the calibrations an absolute character and at the same time described goodness of fit. Various calibration parameters for the three enzymatic assays are stated in Table 1. The sign of the calibration slope of each individual calibration is not linked to substrate or product character as it can result as positive or negative when repeating the PARAFAC analysis. This results from an existing sign ambiguity in Component analysis and is further described in [25].

The unique solution derived from PARAFAC highly correlated with the enzyme activity in each case (Fig. 5). Since the PARAFAC scores were retrieved to encompass one score result for the spectral evolution of the substrate(s) and another one for the product(s), per definition two calibrations should result from the PARAFAC analysis for each enzyme (Fig. 3). Overlap of strong fingerprint abundances from the PARAFAC derived solution with substrate depletion bands, e.g. gave evidence to identify this PARAFAC Component as a substrate depletion calibration and vice versa. Hence, for pectin lyase and Celluclast 1.5L a set of two calibrations, one for the substrate consumption and one for the product formation could be established, respectively (Fig. 5A–D). However, glucose oxidase showed poorer performance than the other enzyme reactions due to lower signal/noise ratio and resulted in only one calibration (product calibration) (Fig. 5E). Even increasing observation time to 20.7 min for glucose oxidase did not improve substrate calibration performance as the spectral evolution of the substrate consumption (bands with decreasing abundance) did not exhibit efficient spectral change in the observed time period as can be seen from Fig. 1B. Additionally the product calibration indicated poorer precision and a lower linear correlation coefficient $R^2 = 0.97$ (Fig. 5E). Underlying linear correlation was validated by modeling several PARAFAC calibrations using different observation times. Assuming each calibration to be linear the correlation coefficient R^2 increased with observation time as shown in Fig. 6. Further details are given in Section 4.2.

4.2. Calibration robustness

In general calibration robustness depends on each enzymatic system itself since the *Spectral Evolution Profiles* appear to look different for various enzymatic reactions. The nature of each *Evolution Profile* is therefore specific for each investigated enzyme/substrate system. Concluding, the spectral bands of substrates and products can overlap each other to a different extent which can hide valuable information concerning the reaction kinetics.

As mentioned previously calibration robustness secondly depends on the observation time of the *Evolution Profile*. For a commercially useful method it is valuable to optimize the acquisition time for an *Evolution Profile*. Considering glucose oxidase Fig. 6 shows how the calibration performance declines as the observation time of the enzymatic reaction is shortened. Yet, following

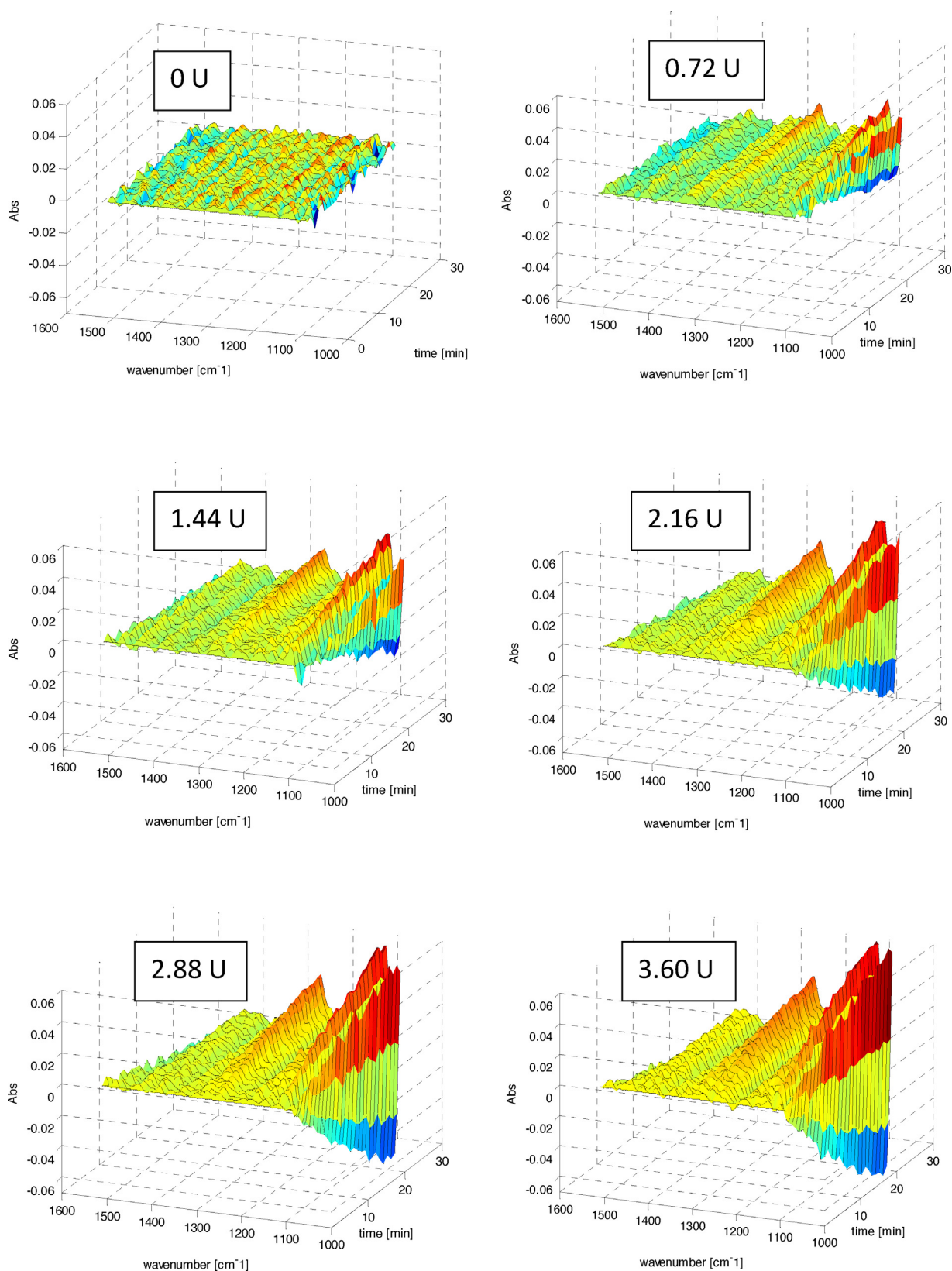


Fig. 4. Spectral Evolution Profiles for the enzymatic reaction of pectin lyase using different activities. Spectral evolution clearly depends on added enzyme activity. The observation time was extended for illustrational purposes.

the reaction for 10 min resulted in a calibration with a regression coefficient $R^2 > 0.90$. Calibration performances of pectin lyase and Celluclast 1.5L have been much more robust as observation times were much shorter (Table 1).

4.3. PCA versus PARAFAC

The PARAFAC approach has been compared to chemometric two-way methods as Principal Component Analysis (PCA).

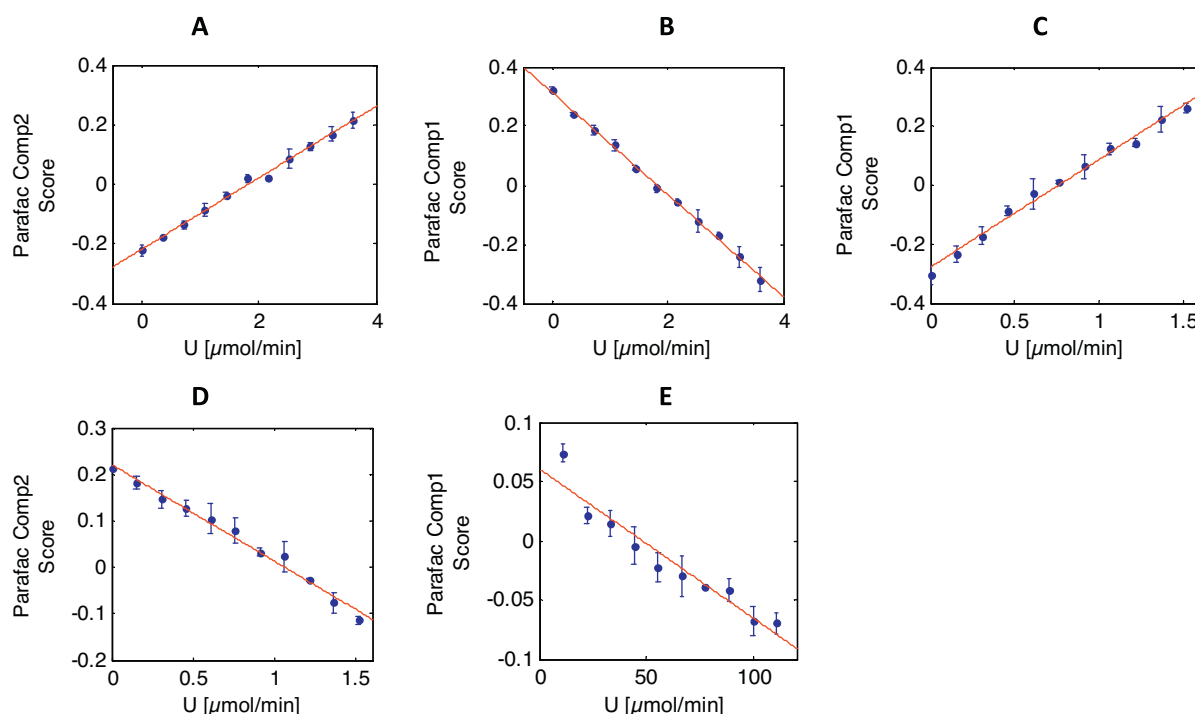


Fig. 5. PARAFAC scores from both Components were plotted against determined enzyme activities (derived from conventional assays) to build calibrations. PARAFAC is an unsupervised multi-way decomposition method and identified the scores without any knowledge about the added enzyme activity. (A) Pectin lyase product calibration ($R^2 = 0.995$); (B) pectin lyase substrate calibration ($R^2 = 0.998$); (C) Celluclast 1.5L product calibration ($R^2 = 0.98$); (D) Celluclast 1.5L substrate calibration ($R^2 = 0.99$); E: glucose oxidase product calibration ($R^2 = 0.97$). The sign of the slope of each individual calibration is not linked to substrate or product character (sign ambiguity in Component analysis [25]).

Chemometric algorithms as Principal Component Analysis (PCA) [26] are commonly used to visualize the characteristics of different spectra in a feature reduced subspace defined by its Principal Components.

To analyze the kinetics of a certain enzymatic reaction (for one *Evolution Profile*) the scores of each spectrum, derived from PCA, need to be plotted against the time. However, since PCA finds Principal Components only due to maximal variance in the spectra the outcome strongly depends on the nature of the *Evolution Profile* itself (e.g. signal-to-noise ratio). It was therefore impossible to use PCA to find reproducible kinetic profiles for the same enzymatic reaction using different amounts of enzyme, simply because the variance in spectra was distributed in a different manner, e.g. in low

activity cases noise was so abundant that the first and/or second Principal Component were highly biased. PCA was therefore not useful for enzyme activity calibrations since it did not necessarily identify the right variance in spectra being related to the enzymatic reaction.

Due to many interfering background signals (mainly water and buffer as shown in Supplementary Fig. A1) the sensitivity of FTIR is somewhat lower in comparison to other techniques as it can be seen from Fig. 5E (considerably small calibration slope). In that regard the signal-to-noise-ratios of the *Evolution Profiles* for glucose oxidase were especially low as shown in Fig. 1B. This reaction requires oxygen as a second substrate. Its measurement was therefore cumbersome since the reaction had to be pumped out of the instrument

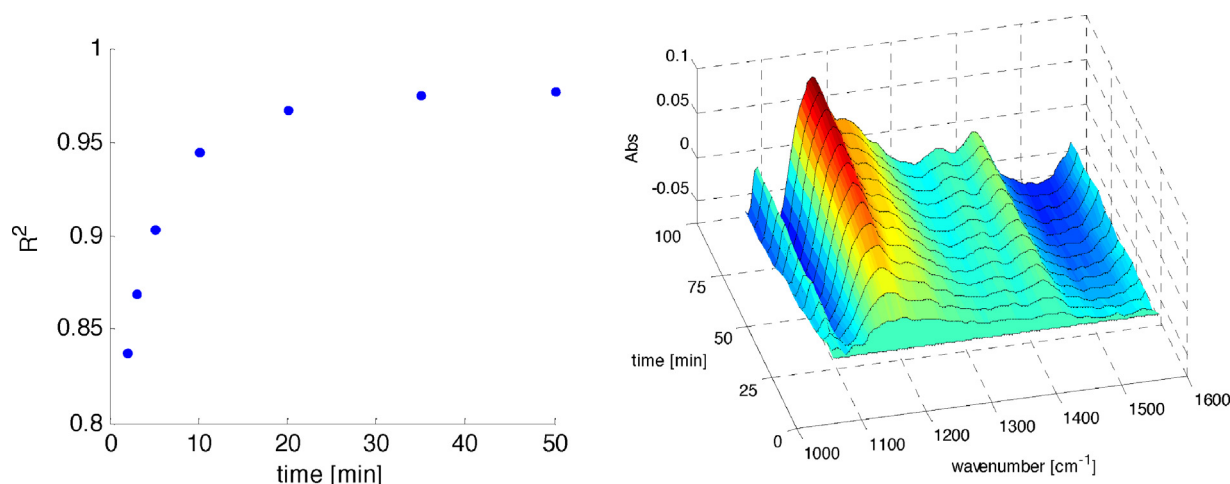


Fig. 6. Illustration of how the product calibration performance increases with observation time. The *Evolution Profile* shows the Difference spectra of a glucose oxidase reaction for an extended observation time.

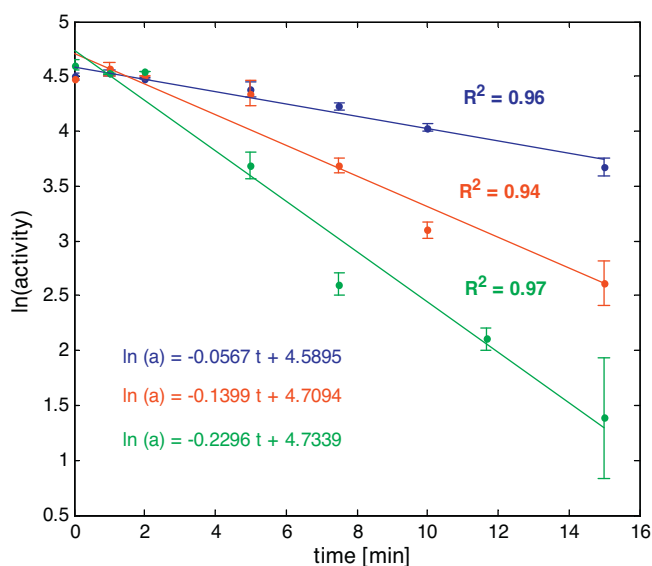


Fig. 7. Thermal inactivation of pectin lyase for 70 °C (blue), 75 °C (red) and 80 °C (green). The study was carried out as an application example of “Measuring enzyme activity using spectral evolution profiling”. All three inactivation experiments have been fitted to Eq. (7). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

after each measurement to ensure constant oxygen supply. However, the PARAFAC analysis overcomes possible constraints which may arise from low signal-to-noise ratio since the multi-linearity of the dataset interconnects all the *Evolution Profiles* (Fig. 2). A common kinetic pattern can therefore be found by PARAFAC and fitted to all *Evolution Profiles* simultaneously (Fig. 3), unlike fitting separate parameters to the unfolded data using PCA. PARAFAC therefore even recognizes the kinetics for enzymatic reactions with very small changes in spectra over time and high amounts of noise. Noisy *Evolution Profiles* due to low activity therefore benefit from strong spectral evolutions due to high enzyme activity.

4.4. Application example: thermal inactivation of pectin lyase

The proposed method was used to study the thermal inactivation of pectin lyase at three different temperatures, namely 70, 75 and 80 °C. To do so, enzymatic samples have been incubated accordingly for different time intervals using three replicates. Thereafter the precipitated protein was separated using centrifugation and the supernatant was used for the acquisition of a spectral *Evolution Profile*, respectively. The obtained scores from using a pectin lyase PARAFAC model (as in Fig. 5) were used to determine the activity of the sample after thermal treatment. All three thermal inactivation experiments at 70, 75 and 80 °C could be fitted to Eq. (6). The regression constants R^2 were 0.96, 0.93 and 0.95, respectively.

$$[E_t] = [E_0]e^{-k_D t} \quad (6)$$

$$\ln([E_t]) = \ln([E_0]) - k_D t \quad (7)$$

A linearized form (Eq. (7)) was used to determine the inactivation rate constants expecting a first order denaturation: $N \xrightarrow{k_D} D$ where N was the enzymes native and D the enzymes denaturated form. The determined inactivation rate constants were 0.0567 min⁻¹, 0.1399 min⁻¹ and 0.2296 min⁻¹, respectively (Fig. 7). With reference to the Arrhenius equation (Eq. (8)) temperature correlated with inactivation rate. The regression constant R^2 was

calculated to be 0.98 when using the linearized plot in Eq. (8).

$$k_D = k_0 \cdot e^{-(E_A/RT)} \rightarrow \ln k_D = \ln k_0 - \frac{E_A}{R} \frac{1}{T} \quad (8)$$

5. Conclusion

The presented study demonstrates a new approach to monitor enzymatic reactions continuously by taking all substrate(s) and product(s) into account. By observing temporal spectral evolution using FTIR the method inaugurates possibilities to also characterize multi component enzymatic systems. The approach does not require any chemical standards or biomarkers, but instead utilizes multivariate analysis to continuously monitor the enzymatic reaction in real time. Thus the method is directly applicable to a wide range of enzymes, even considering complex sample matrices, and may furthermore be highly suitable for automation implying intelligent instrumentation. Replacing cumbersome univariate enzyme activity assays with universal and direct spectroscopic assays promise a better and direct observation of the big picture regarding ongoing enzymatic reactions. The recent advances in multivariate analysis allow substantial alternatives to question the reasonability of traditional enzyme activity assessment.

Acknowledgements

This work was partially financed by the 7th Framework Program via the Marie Curie Initial Training Network, LeanGreenFood (EU-ITN 238084). The Fungal Generic Stock Center (FGSC) is acknowledged for the *Pichia pastoris* clone expressing the pectin lyase gene AN2569.2.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.aca.2013.03.029>.

References

- [1] S.M. Drawz, R.A. Bonomo, Clin. Microbiol. Rev. 23 (2010) 160–201.
- [2] M.W. Bevan, M.C.R. Franssen, Nat. Biotechnol. 24 (2006) 765–767.
- [3] A. Hennig, H. Bakirci, W.M. Nau, Nat. Methods 4 (2007) 629–632.
- [4] R.K. Bera, A. Anoop, C.R. Raj, Chem. Commun. 47 (2011) 11498–11500.
- [5] J.L. Reymond, Food Technol. Biotechnol. 42 (2004) 265–269.
- [6] F.L. Martin, Nat. Methods 8 (2011) 385–387.
- [7] F.A. Iñón, S. Garrigues, M. de la Guardia, Anal. Chim. Acta 513 (2004) 401–412.
- [8] K. Karmali, A. Karmali, A. Teixeira, M.J.M. Curto, Anal. Biochem. 333 (2004) 320–327.
- [9] R. Pacheco, A. Karmali, M. Serralheiro, P.I. Haris, Anal. Biochem. 346 (2005) 49–58.
- [10] C. Adina, F. Florinela, T. Abdelmoumen, S. Carmen, Rom. Biotechnol. Lett. 15 (2010) 5739–5744.
- [11] M. Beekes, P. Lasch, D. Naumann, Vet. Microbiol. 123 (2007) 305–319.
- [12] R. Bro, Crit. Rev. Anal. Chem. 36 (2006) 279–293.
- [13] R.A. Harshman, M.E. Lundy, Comp. Stat. Data Anal. 18 (1994) 39–72.
- [14] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
- [15] R.A. Harshman, UCLA Work. Pap. Phon. 16 (1970) 1–84.
- [16] R. Bro, A.K. Smilde, J. Chemometr. 17 (2003) 16–33.
- [17] S.K. Andersen, P.W. Hansen, H.V. Andersen, Handbook of Vibrational Spectroscopy, 1st edition, 2002.
- [18] L.A.M. van den Broek, E.D. den Aantrekker, A.G.J. Voragen, G. Beldman, J.P. Vincken, J. Sci. Food Agric. 75 (1997) 167–172.
- [19] R. Edstrom, H. Phaff, J. Biol. Chem. 239 (1964) 2403–2408.
- [20] S.B. Bankar, M.V. Bule, R.S. Singhal, L. Ananthanarayan, Biotechnol. Adv. 27 (2009) 489–501.
- [21] M. Koziol, Anal. Chim. Acta 128 (1981) 195–205.
- [22] A.B. Blakeney, L.L. Mutton, J. Sci. Food Agric. 31 (1980) 889–897.
- [23] M. Lever, Anal. Biochem. 47 (1972) 273–279.
- [24] P. Hartmann, S.J. Haswell, M. Grasserbauer, Anal. Chim. Acta 285 (1994) 1–8.
- [25] R. Bro, E. Acar, T.G. Kolda, J. Chemometr. 22 (2008) 135–140.
- [26] S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.

PAPER 4

Andreas Baum, Per W. Hansen, Anne S. Meyer, Jørn D. Mikkelsen; *Simultaneous measurement of two enzyme activities using infrared spectroscopy: comparative evaluation of PARAFAC, TUCKER and N-PLS modeling*; Analytica Chimica Acta, submitted



Simultaneous measurement of two enzyme activities using infrared spectroscopy: A comparative evaluation of PARAFAC, TUCKER and N-PLS modeling

Andreas Baum^{a,b}, Per Waaben Hansen^b, Anne S. Meyer^a, Jørn Dalgaard Mikkelsen^{a,*}

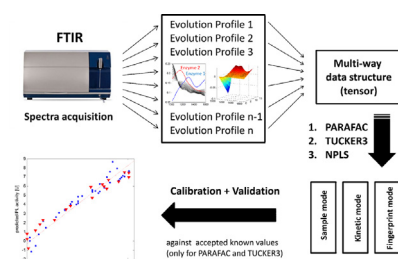
^a Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

^b Foss Analytical, Foss Allé 1, DK-3400 Hillerød, Denmark

HIGHLIGHTS

- Simultaneous enzyme activity determination of two co-acting enzymes.
- Activity calibration and validation for pectin lyase and pectin methyl esterase.
- Chemometric multiway modeling comparing PARAFAC, TUCKER3 and N-PLS.
- Kinetic investigations on genuine substrates as pectin.
- Universally applicable without using any external standards or (bio-)markers.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 17 April 2013

Received in revised form 19 June 2013

Accepted 24 June 2013

Available online 28 June 2013

Keywords:

Fourier transform infrared spectroscopy

Enzyme activity

Spectral evolution profiles

Multiway

Second-order calibration

ABSTRACT

Enzymes are used in many processes to release fermentable sugars for green production of biofuel, or the refinery of biomass for extraction of functional food ingredients such as pectin or prebiotic oligosaccharides. The complex biomasses may, however, require a multitude of specific enzymes which are active on specific substrates generating a multitude of products. In this paper we use the plant polymer, pectin, to present a method to quantify enzyme activity of two pectolytic enzymes by monitoring their superimposed spectral evolutions simultaneously. The data is analyzed by three chemometric multiway methods, namely PARAFAC, TUCKER3 and N-PLS, to establish simultaneous enzyme activity assays for pectin lyase and pectin methyl esterase. Correlation coefficients R^2_{pred} for prediction test sets are 0.48, 0.96 and 0.96 for pectin lyase and 0.70, 0.89 and 0.89 for pectin methyl esterase, respectively. The retrieved models are compared and prediction test sets show that especially TUCKER3 performs well, even in comparison to the supervised regression method N-PLS.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Comprehensive understanding of enzymatic reactions has become increasingly important as trends in biotechnological research have led to a higher industrial demand concerning diverse production of enzymes. To understand the mechanism of an

enzymatic reaction enzyme activity, as defined of how much substrate is converted into product per time unit, is of major interest.

Previously reported enzyme activity assays rely on univariate measurement principles being applied to data discontinuously obtained from Colorimetry [1] and HPLC [2–4]. Alternatively, spectroscopic measurements are used utilizing univariate peak integration methods or comparable analytical methods as described in [5–7]. Hence, each enzyme activity assay is designed to be specifically bound to a certain enzymatic system. Besides being rather cumbersome, time-consuming and not suitable for

* Corresponding author. Tel.: +45 60856300.

E-mail address: jdm@kt.dtu.dk (J.D. Mikkelsen).

high-throughput measurements these assays rely on measuring the enzyme activity by either following the substrate depletion or the product formation, but not both simultaneously. Additionally, no enzymatic assay has been reported which can continuously quantify enzyme activities of two co-acting enzymes in one sample simultaneously. Such measurements are particularly relevant in relation to processing of complex substrates, e.g., plant biomass, since in this case efficient degradation requires the concerted action of several enzymes.

Thus, it would be beneficial to develop a feasible and universal methodology to directly observe the kinetics of an enzymatic reaction, without any need to apply chemical markers. Additionally, such a method should be universally applicable to all kinds of enzymes and independent from its surrounding sample matrix.

One solution is to use a multivariate measurement technique which is able to capture information considering all reaction parameters, as to observe all products and all substrates of an enzymatic system. Such great amount of information usually overlaps within a certain measurement (e.g., spectral) range and therefore leads to fingerprints which are interrelated with the underlying chemical nature. Fourier infrared spectroscopy (FTIR) is such a fingerprinting technique. It analyses the absorption pattern of infrared light due to vibrational modes of covalent bonds and enables the possibility to monitor the overall pattern of an ongoing enzymatic reaction continuously if spectra are acquired consecutively.

However, applying such multivariate measurement principles to, e.g., time-resolved FTIR data requires the use of (multiway) chemometric tools. A method utilizing chemometric multiway methodologies for enzyme activity determination has been reported recently [8]. In this particular study spectral evolution profiling in connection with FTIR has been used to quantify the enzyme activity of one enzyme, e.g., glucose oxidase, pectin lyase, or a composition of enzymes, e.g., a cellulolytic enzyme blend. The study also discussed critical detection limits and major drawbacks of FTIR, which are related to its comparable low sensitivity, and furthermore described how chemometric multiway analysis can help to overcome these challenges.

On the other hand, this paper extends the concept consequently to distinguish and quantify the enzyme activity of two enzymes, i.e. pectin lyase and pectin methyl esterase, simultaneously [9,10]. Such a multi-enzyme analysis could be of high interest to determine side activities in enzyme production and to provide a tool to control the quality of fermentation processes. Monitoring of several activities could therefore be of potential interest in both, industrial processes and scientific research.

The present study is carried out by continuously monitoring the combined spectral evolution of the enzymatic reactions using FTIR. Thereafter the underlying spectral patterns related to the two enzyme activities are deconvoluted to build individual enzyme activity calibrations using three different chemometric multiway methods on the same data. The retrieved models, derived from Parallel Factor Analysis (PARAFAC), TUCKER3 and multilinear Partial Least Squares regression (N-PLS), are compared toward feasibility and calibration performance.

Further comprehensive introduction and overview on different applications of multiway chemometrics is given by Bro [11,12].

2. Theory

2.1. Spectral evolution profiling

In this chapter a simulated dataset is used for illustrational purpose.

If we consider an enzymatic reaction which converts a substrate *S* into a product *P*1, following first order or pseudo first order

kinetics, the hypothesis can be stated that (taking mid-infrared spectroscopic measurements into account) substrate and product have distinguishable fingerprints which should evolve from one to another. This spectral evolution is linked to the enzymatic reaction and its kinetics and is presented in Fig. 1A.

A second enzymatic reaction from substrate *S* to product *P*2 is exemplified as in Fig. 1B. If we combine both enzymatic reactions and let them occur simultaneously we can imagine the time-resolved spectra to appear as illustrated in Fig. 1C, assuming that both enzymes are present in equal amounts. In all scenarios the reaction is stopped before reaching a total conversion, where the final product fingerprints are illustrated as red and blue dashed lines (Fig. 1A–C).

All enzymatic reactions are monitored using very high initial substrate concentrations to assure that there is no significant (competing) interaction between the enzymes. This is of particular importance because interaction of the enzymes would disturb independent assessment of individual enzyme activities. Fig. 1A–C also shows the spectral changes using difference spectra visualized utilizing surf plots. These plots have advantages visualizing small changes in spectra and furthermore enhance visibility of spectral convergence since the dimension of time is resolved in space [8]. When no significant interaction between the enzymes during the observation time can be assured a perfect superimposition of the individual “enzymatic landscapes” as illustrated in Fig. 1C can be expected.

In the present study pectin lyase (PL) and pectin methyl esterase (PME) have been used simultaneously to degrade the same substrate, namely pectin from citrus peel. Biopolymers as pectin can serve as substrates for several enzymatic reactions at the same time. While pectin lyase catalyzes the de-polymerization of homogalacturonan regions in pectin, pectin methyl esterase catalyzes the de-esterification of the methyl ester groups of the pectin. The chemical equations of both reactions are shown in Fig. 2.

2.2. Chemometrics

Each acquired spectrum can be represented by a vector of mid infrared absorption values. Hence, a time-resolved evolution profile forms a matrix with *n* vectors at *n* consecutive time points.

In order to quantify the enzyme activity of both enzymes simultaneously all evolution profiles, represented by data matrices, were stacked on top of each other to form a tri-linear data structure as illustrated in Fig. 3A and B. Since all evolution profiles have been acquired using the same consecutive time steps this data structure transformation adds up one more measurement dimension, namely activity. Instead of working with unfolded matrices of the single evolution profiles one three-way array, mathematically spoken tensor, is therefore obtained. This tensor is thereafter decomposed by multiway methods as PARAFAC, TUCKER3 (Fig. 3A) or NPLS (Fig. 3B) into three mode loadings *A*, *B* and *C* which in this particular paper refer to sample mode, kinetic mode and fingerprint mode. The loadings of the sample mode are also referred to as scores.

2.2.1. PARAFAC

PARAFAC [13] is a mathematical tensor decomposition method which results in a unique solution meaning that the solution has no rotational freedom as in ambiguous models derived from e.g., PCA [14]. The three-way array *X* is decomposed into three matrices *A*, *B* and *C* by minimizing the sum of squares of the residuals in *E* according to Eq. (1), where *a*, *b* and *c* are elements of those matrices while *R* represents the number of components. A schematic illustration is given in Fig. 3. Ideally the core tensor *G* is super-diagonal meaning that all positions along the dashed line contain ones while the remaining positions contain zeros, respectively.

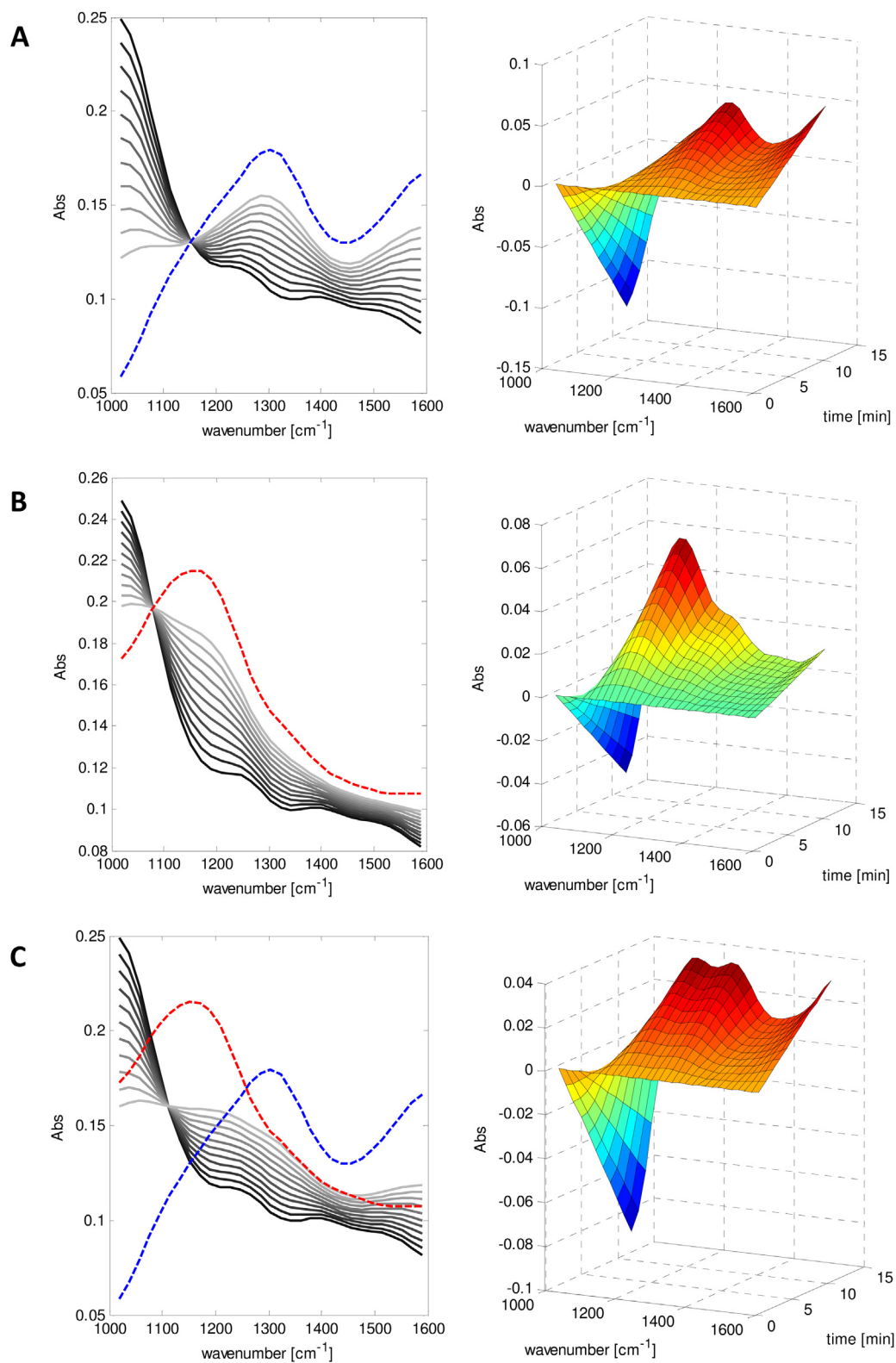


Fig. 1. Simulated data for three enzymatic reactions. (A) Enzyme 1 catalyzes the conversion of the substrate to product P1; (B) Enzyme 2 converts the substrate to product P2; (C) Enzyme 1 + Enzyme 2 (half dose enzyme 1 + half dose enzyme 2) convert the Substrate to both products (P1 and P2) at the same time. The spectral changes from Substrate fingerprint to product fingerprint are illustrated on the left side, whereas difference spectra using surf plots are shown on the right side. Individual final product fingerprints are given by red and blue dashed lines, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

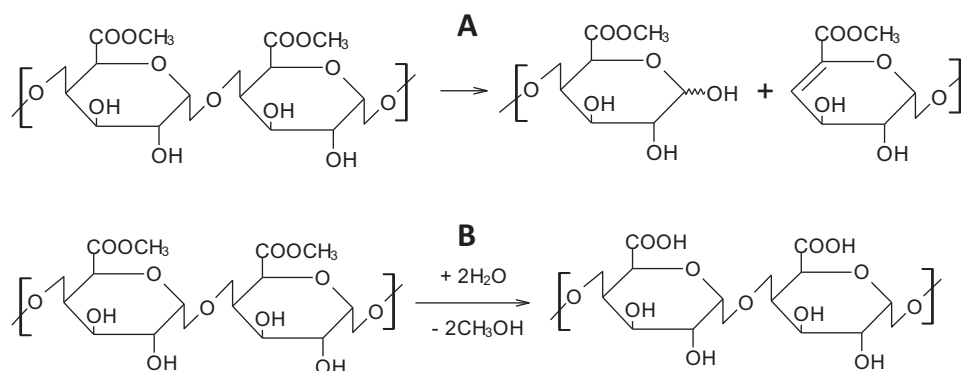


Fig. 2. Enzymatic reactions for (A) pectin lyase and (B) pectin methyl esterase.

According to the here present models the loadings in mode C (fingerprint mode) identify spectral bands which change during the course of the enzymatic reactions, while the components in B (kinetic mode) represent the time-resolved shape of those fingerprint changes. As a consequence the scores in mode A (sample mode) describe the abundance of the underlying spectral evolutions, derived from mode B and C. This was described in more detail in [8]

$$X_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (1)$$

2.2.2. TUCKER3

TUCKER3 is a multiway tensor decomposition method [15] which decomposes a three way array, e.g., $X (I \times J \times K)$ into three loading matrices, $A (I \times P)$, $B (J \times Q)$ and $C (K \times R)$ and a core tensor, G [16,17]. The model does not result in a mathematical unique solution and has therefore rotational freedom. TUCKER3 can be expressed as given in Eq. (2), where a_{ip} , b_{jq} and c_{kr} are elements of the loading matrices, g_{pqr} are elements of the core array and e_{ijk} are the individual residuals.

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \quad (2)$$

Unlike in PARAFAC analysis, the core tensor G , is not super-diagonal and enables possibilities to analyze interaction between the different modes. Decomposition of multiway arrays into loading matrices which may have different individual ranks can

therefore be analyzed. Alternatively, the TUCKER3 decomposition can be written as in Eq. (3), where \otimes denotes the Kronecker product. A, B and C are linked to sample, kinetic and fingerprint mode as described for PARAFAC.

$$X = AG(C \otimes B)^T + E \quad (3)$$

2.2.3. N-PLS [18]

Multilinear Partial Least Squares regression can be described quite similarly to bilinear PLS, where both matrices X and Y are simultaneously decomposed while finding maximal covariance of the scores of both decompositions [19]. Detailed elaboration on the mathematical description of N-PLS and its algorithms were described by Bro in [20]. Unlike in PARAFAC and TUCKER3 N-PLS uses accepted known reference values (Y) to find suitable projections for calibration modeling and therefore comprises a risk of over fitting. PLS models are usually validated by cross validation [21] and tested on prediction sets.

3. Materials and methods

3.1. Reagents

Pectin lyase EC 4.2.2.10 family PL1 was a cloned enzyme from *Aspergillus aculeatus* and expressed as a recombinant enzyme in *Saccharomyces cerevisiae*; the enzyme was a gift from Novozymes A/S (Bagsværd, Denmark) [22]. Pectin methyl esterase EC 3.1.1.11 family CE8 was also cloned from *A. aculeatus* as described in [23]. The monocomponent, recombinant enzyme was a gift from Novozymes A/S (Bagsværd, Denmark). Pectin from citrus peel with a degree of esterification around 80% was donated from Dupont, Brabrand, Denmark.

3.2. FTIR

All spectra were obtained using a MilkoScan™ FT2 (FOSS ANALYTICAL, Hillerød, Denmark). The instrument consists of a FTIR interferometer that scanned the full infrared spectrum, was equipped with an automatic flow-through system apparatus and worked semi-automated. The optical system was hermetically sealed and pressure and humidity controlled. The path length of the cuvette was 50 μm. The temperature was equilibrated at 42 °C using heat exchangers.

All spectra have been measured against an aqueous blank. After each measurement the flow-through system was cleaned automatically (with water). The time required to measure one spectrum accounted 16.6 s resulting in 11 min acquisition time for one evolution profile containing 40 spectra.

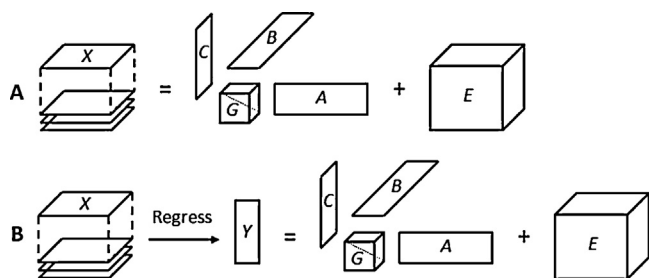


Fig. 3. The data cube X is decomposed into A (sample mode), B (kinetic mode) and C (fingerprint mode). E refers to the residuals. (A) Schematic illustration of PARAFAC and TUCKER3 decomposition. While in PARAFAC the core tensor G is superdiagonal (containing ones along the dashed diagonal and zeros in the other positions) in TUCKER3 the core tensor G can contain values deviating from zero in the positions outside the dashed diagonal. (B) Schematic illustration of N-PLS (multiway PLS) regression. The major difference to unsupervised tensor decomposition methods as PARAFAC or TUCKER3 is that N-PLS is guided by information of accepted known values (Y) to find suitable projections for calibration.

3.3. Conventional enzymatic assays

The accepted known enzyme activities of PL and PME have been determined by conventional assays as described below.

3.3.1. Pectin lyase

Pectin lyase activity was measured on 10 g L^{-1} pectin by incubating in 50 mM sodium acetate Buffer at pH 5.0 and 42°C . The increase in absorbance was determined at 235 nm during 4 min in an Infinite200 microplate reader (Tecan, Salzburg, Austria); the data collection was controlled by the program Tecan i-control version 1.5.14.0. The extinction coefficient used was $5.5 \text{ mM}^{-1} \text{ L}^{-1}$. The activity of PL was defined as the amount of enzyme releasing $1 \mu\text{mol}$ of product per minute calculated using the molar absorption coefficient $5500 \text{ M}^{-1} \text{ L}^{-1}$ [9,24,25].

3.3.2. Pectin methyl esterase

PME cleaves esters into alcohols and its corresponding acid (Fig. 2). The PME activity was measured using an autotitrator. A 10 g L^{-1} pectin solution was prepared in 50 mM sodium acetate buffer at pH 5.0 and 42°C and 0.1 M sodium hydroxide was added continuously to maintain pH 5.0 during the enzymatic reactions. The activity of PME was defined as the amount of enzyme necessary to release $1 \mu\text{mol}$ of methanol per minute as estimated from the consumed sodium hydroxide equivalents during the autotitration [10,26].

3.4. Chemometric methods

After calibration data acquisition the data were exported to Matlab (The Mathworks Inc., MA, USA) where they were treated using designated scripts. PARAFAC, TUCKER3 and N-PLS analysis were performed using PLS Toolbox 6.0.1 (Eigenvector Research Inc., WA, USA).

3.4.1. Data structure

The three-way array, X , indicated the dimensions of 56 samples \times 131 wavenumbers \times 40 time points ($I \times J \times K$). Before decomposition of the tensor the data were preprocessed.

3.4.2. Data preprocessing

Variables selected for the analysis throughout the paper have been wavenumbers ranging from 1018 to 1507 cm^{-1} . The spectra have been scaled using Standard Normal Variate (SNV) along the spectral axis. To do so the tensor was unfolded and normalized slabwise across the sample mode as described in Eqs. (4)–(6) [13,27]

$$\bar{X}_{ik} = \frac{\sum_{j=1}^J x_{ijk}}{J} \quad (4)$$

$$W_i = \sqrt{\frac{\sum_{k=1}^K \sum_{j=1}^J (x_{ijk} - \bar{X}_{ik})^2}{KJ - 1}} \quad (5)$$

$$x_{ijk}^{SNV} = \frac{x_{ijk}}{W_i} \quad (6)$$

Additionally the data cube was consecutively double centered across the sample and the kinetic mode [27]. Multiway centering, e.g., for the sample mode can be understood as unfolding the tensor to a $I \times JK$ matrix and proceeding as in usual two-way analysis (Eq. (7)):

$$x_{ijk}^{cent} = x_{ijk} - \frac{\sum_{i=1}^I x_{ijk}}{I} \quad (7)$$

3.4.3. Validation

In this paper the data were split into two sets. Random two thirds of the samples were used to generate models using leave-one-out (loo) cross validation and the remaining set of one third of the samples was used as a test set to check for goodness of prediction.

Strategies for evaluation of prediction errors are usually given by figures of merit in multivariate analysis. In this paper the root mean square error of prediction (RMSEP) was used to describe the degree of agreement between predicted values y_i and the accepted known values \hat{y}_i of the test set using a total model derived from the calibration set. Those values are calculated as shown in Eq. (8).

$$\text{RMSEP} = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{n} \quad (8)$$

3.4.4. Calibration accuracy and precision

Besides evaluation of the RMSEP calibration accuracy was determined by calculating the linear regression coefficient R^2 of model predicted enzyme activity against accepted known enzyme activity as described in Eq. (9).

$$R^2 = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad (9)$$

The variables \hat{y} and \bar{y} represent the predicted value using determined slope and intercept of the linear regression and the mean of y , respectively, while y_i represents the actual observed value.

The precision is a measure of repeatability and can be understood as of how much replicates of the same accepted known enzyme activity differ in prediction. This span defines the precision of the individual calibration and is usually expressed using standard deviation. However, if several calibration points indicate different numbers of replicates the pooled standard deviation s_p gives an improved estimate of the experimental precision as described in Eq. (10). Hereby, s_1, s_2, \dots, s_k represent the standard deviations of the individual calibration points while n_1, n_2, \dots, n_k denote the amount of replicates at every calibration point.

$$s_p = \frac{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}}{n_1 + n_2 + \dots + n_k - k} \quad (10)$$

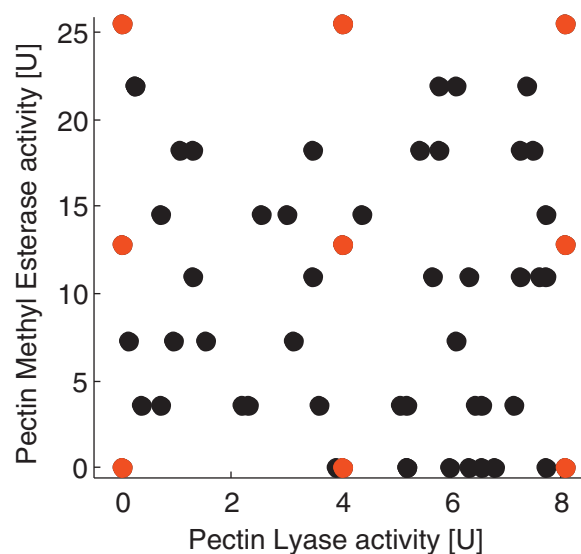


Fig. 4. Experimental design: red samples define calibration ranges, while black samples are randomly placed to create a normally distributed sampling scenario. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

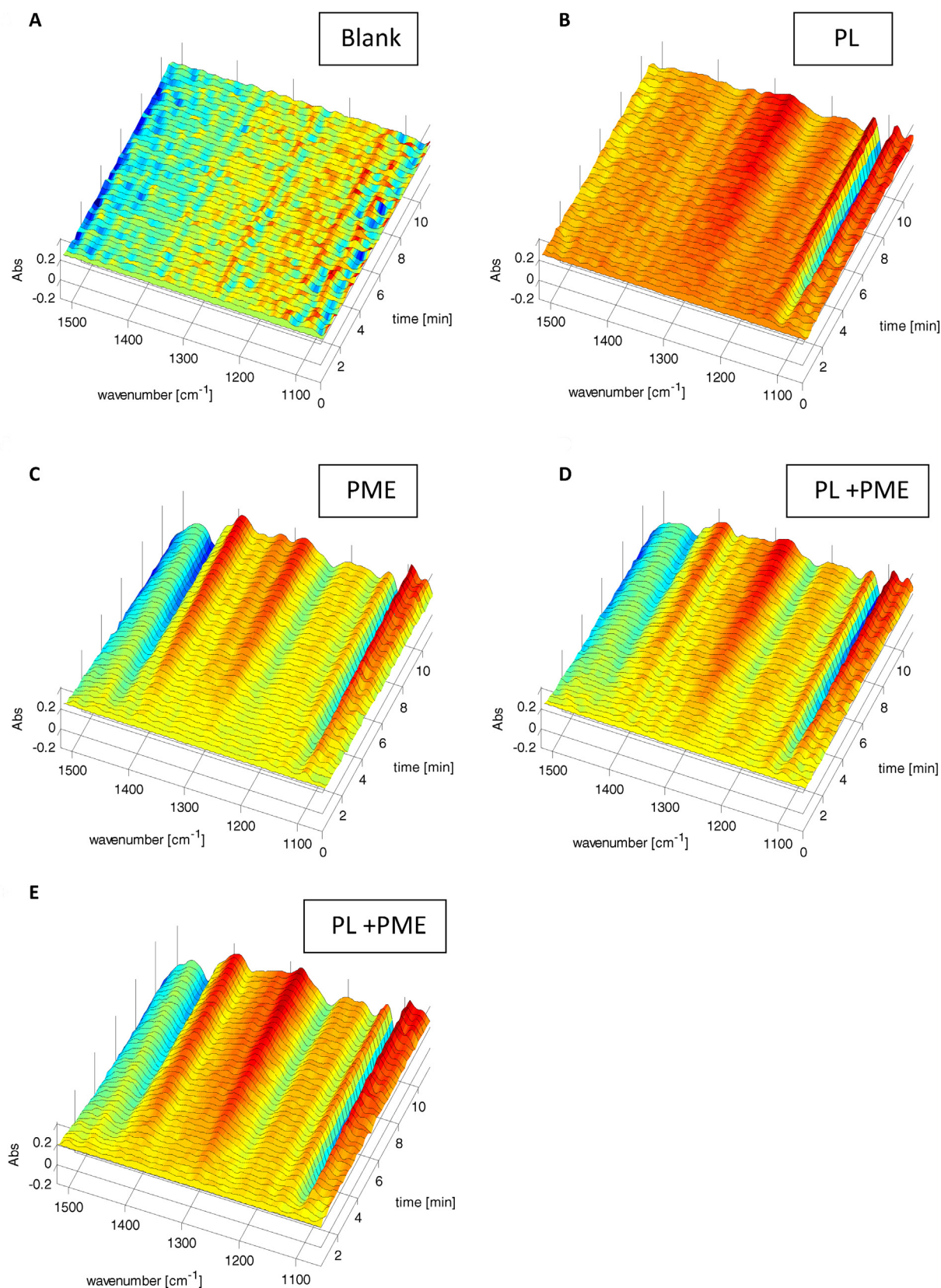


Fig. 5. Spectral evolutions (11 min each) for the two simultaneously profiled enzymatic reactions of pectin lyase and pectin methyl esterase. A shows a substrate blank (no significant spectral change) while B and C show spectral evolutions for PL (8.07 U) and PME (25.53 U), respectively. D shows a mixed spectral evolution of both enzymes (4.04 U PL and 12.77 U PME) while E shows spectral evolution for an enzyme mixture of 8.07 U PL and 25.53 U PME. Spectra are displayed as difference spectra, meaning that the initial spectrum has been subtracted from the whole series to eliminate constant background signals.

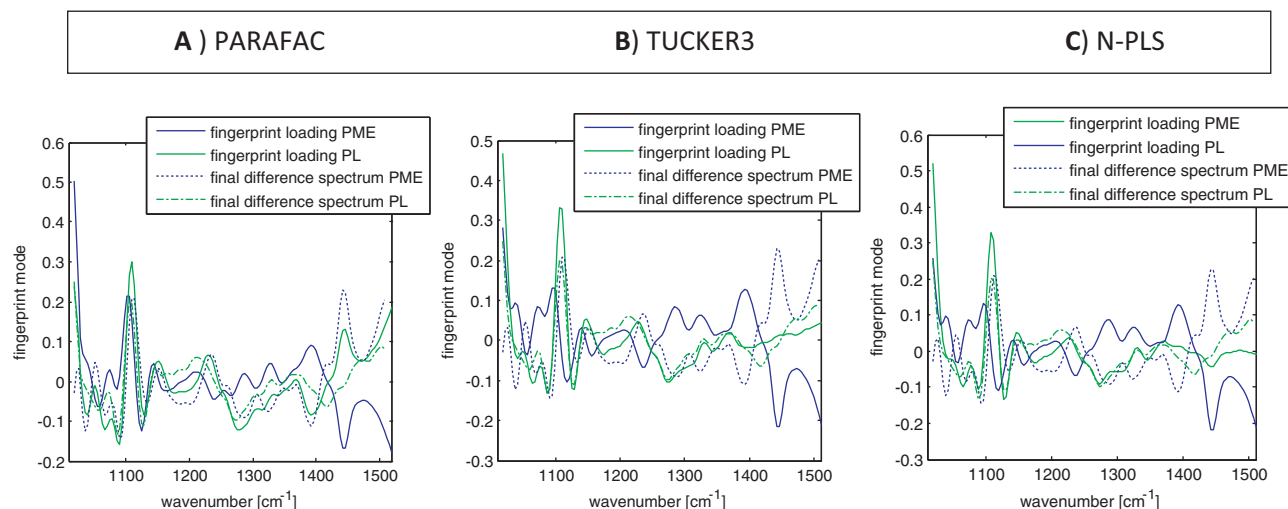


Fig. 6. Fingerprint mode loadings for (A) PARAFAC, (B) TUCKER3 and (C) N-PLS. The PARAFAC fingerprint mode loadings highly deviate from the other two model loadings in the spectral range between 1400 and 1500 cm^{-1} . Final difference spectra of PL and PME reactions are displayed in dashed style to highlight PL and PME reaction character.

In the present study replicates did not indicate constant enzyme activity ratios (Fig. 4). However, the experimental acquisition of evolution profiles toward the same enzymatic reaction has been found to be highly reproducible as it was published previously in a related study [8].

3.4.5. Experimental design

To build two calibration models, one for each enzyme, PL and PME, an appropriate experimental design was necessary [28]. Each sample represents spectral information which results from overlap of the two individual enzymatic evolution profiles. A superimposed “spectral landscape” which contains information related to both enzyme amounts can therefore be expected (as in Fig. 1C). Samples with one particular PL activity can be superimposed by PME activities which span from zero to a certain maximum value and vice versa. To guaranty robust calibration models all possible degrees of interference had to be considered.

To achieve exactly this goal nine samples have been defined by mixing enzyme activities (with 20 mL of 10 g L^{-1} substrate solution) as illustrated in Fig. 4 by red points. Those samples define the calibration range wherein additional 47 samples (black points) were randomly placed.

4. Results

Evolution profiles of two control mixtures have been acquired. One control contained only the substrate without enzyme(s) (Fig. 5A). Buffer was added instead of the enzyme solution. The second control contained only enzyme(s) and no substrate (not shown). Instead of substrate solution only buffer was added. The evolution profiles of both controls did not show significant spectral change in time. That step is especially important to ensure that the observed spectral evolution is due to enzyme activity and not due to altering circumstances like temperature changes, precipitation, denaturation, mutarotation, etc.

Various evolution profiles were acquired using equal substrate concentrations and different enzyme activities. Measurements were carried out in random order to prevent from systematic biases. Replicates have been measured according to the described experimental design (Fig. 4).

Some selected evolution profiles using different enzyme activities of PL and PME are displayed in Fig. 5. Fig. 5A shows no significant spectral change due to absent enzymes, while Fig. 5B and C shows evolution profiles concerning pure added enzyme amounts of PL

and PME, respectively. Fig. 5D and E shows combined evolution profiles using mixed activities of both enzymes. An additive character of both individual evolution profiles can be recognized, although being difficult to observe clearly. To achieve a better understanding of the underlying spectral patterns the multiway models were approached. The tensor was decomposed using three multiway methods, namely PARAFAC, TUCKER3 and N-PLS. The obtained data were then compared in order to identify the best multiway method for the purpose.

4.1. PARAFAC

A two component model using PARAFAC was obtained. The rank of the tensor was validated to be two after appropriate chemometric preprocessing. Without the used multiway preprocessing a rank three system would have been expected. This can be explained by the fact that three spectral changes occurred, namely one for the substrate consumption and one for each of the product formations for the enzymatic reactions of PL and PME, in total three. Since the substrate depletion can be expressed as linear dependent of the two spectral changes of product formation the overall rank shall be two after multiway centering the tensor across the kinetic mode (removes the offset due to constant background signals).

The fingerprint mode loadings are presented in Fig. 6A, where the blue fingerprint indicates the PL reaction and the green fingerprint the PME reaction, respectively. Noticeably both spectral fingerprints indicated signals in the spectral range between 1400 and 1550 cm^{-1} , although the PL reaction did not show any significant spectral changes in this range. This will be a reason for poorer calibration performance of PARAFAC as it will be discussed later.

The sample mode scores were correlated against the accepted known enzyme activities and the retrieved calibrations have been used to predict individual enzyme activities using leave-one-out cross validation as the result is presented in Fig. 7. Accepted known enzyme activities have been determined by conventional enzyme activity assays as described in Section 3.3.

Both PARAFAC calibrations in Fig. 7 showed poor linear correlation according to the regression coefficients $R_{\text{PL}}^2 = 0.67$ and $R_{\text{PME}}^2 = 0.52$ for calibration and $R_{\text{PL}}^2 = 0.48$ and $R_{\text{PME}}^2 = 0.70$ for the prediction test set samples, respectively. Precision of both calibrations was rather dissatisfying as it can be seen from the relatively large pooled standard deviations (Table 1) or prediction deviations in Fig. 7.

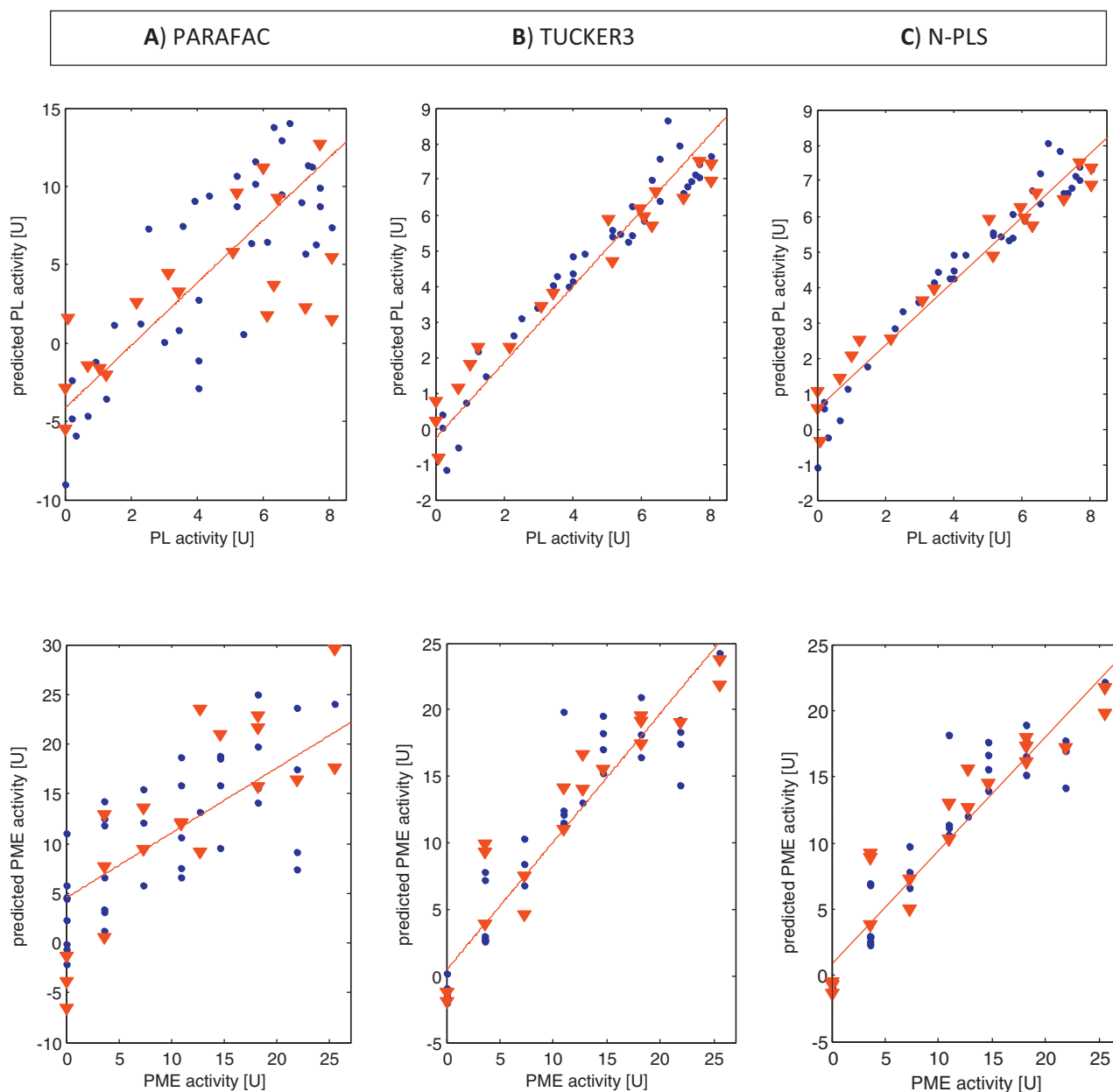


Fig. 7. PL and PME calibrations using multiway scores derived from PARAFAC, TUCKER3 and N-PLS models. Top: PME, bottom: PL; the blue points indicate leave-one-out cross validated calibration samples while red triangles indicate predictions for the test set samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Tucker

As shown in Fig. 8 the PARAFAC loadings for the kinetic mode are not very different from each other. That can be explained by the fact that both reaction rates were changing similarly (close to

initial reaction rate) and therefore had comparable shape. In fact this suggested that the rank of the tensor in the second mode could be reduced to one. The N-PLS kinetic mode indicated that both loading shapes should indeed be very similar. A tucker model with the rank [212] was therefore calculated. TUCKER3 kinetic and

Table 1

Regression coefficients for calibration and prediction set samples are given together with pooled standard deviation s_p and root mean square error of prediction RMSEP.

	PARAFAC		TUCKER3		N-PLS	
	PL	PME	PL	PME	PL	PME
R^2_{cal}	0.67	0.52	0.92	0.88	0.95	0.89
R^2_{pred}	0.48	0.70	0.96	0.89	0.96	0.89
s_p	1.10 U	1.62 U	0.26 U	1.19 U	0.24 U	1.10 U
RMSEP	3.60 U	5.31 U	0.62 U	2.72 U	0.70 U	2.85 U

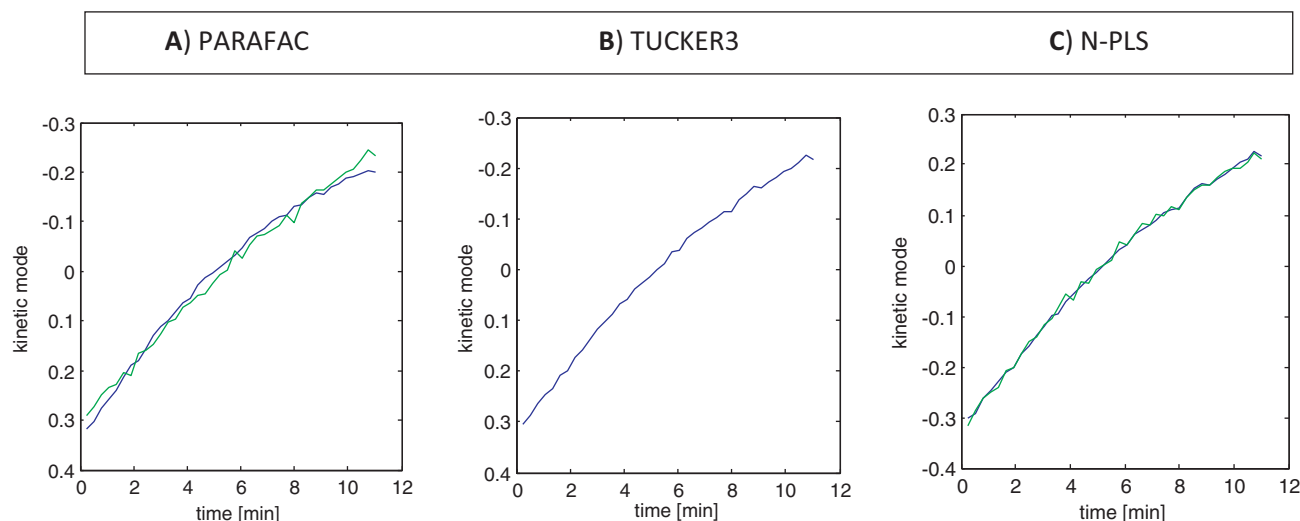


Fig. 8. (A) PARAFAC kinetic mode loadings for both, PL (blue) and PME (green), (B) TUCKER3 kinetic mode loading. The rank in the TUCKER3 kinetic mode was one as only one loading resulted from the analysis. (C) N-PLS kinetic mode loadings for both, PL (blue) and PME (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

fingerprint mode loadings are shown in Figs. 8B and 6B, respectively. The enzyme activity calibrations resulting from the sample scores are shown in Fig. 7B.

The TUCKER3 model showed much better accuracy than the PARAFAC model, as the regression coefficients were $R^2_{PL} = 0.92$ and $R^2_{PME} = 0.88$ for calibration samples and $R^2_{PL} = 0.96$ and $R^2_{PME} = 0.89$ for the prediction test set samples, respectively. Interestingly the fingerprint mode loading for the second component did not show any significant peaks in the range between 1400 and 1500 cm^{-1} as it was expected for PL.

4.3. N-PLS

N-PLS modeling has been done using leave-one-out cross validation. The results are presented in Fig. 7C. Regression coefficients of the linear correlations between the accepted known enzyme activity and the predicted values were $R^2_{PL} = 0.95$ and $R^2_{PME} = 0.89$ for calibration samples and $R^2_{PL} = 0.96$ and $R^2_{PME} = 0.89$ for the prediction test set samples, respectively. Kinetic mode and fingerprint mode loadings are presented in Figs. 8C and 6C. They are similar to the TUCKER3 loadings in both cases.

The cross validation showed that only one latent variable was needed to calibrate the system for each enzyme, as one expects only one underlying evolution profile for each enzyme. This could be verified by the RMSECV value which did not decrease for more than one latent variable. The calibration performances of the N-PLS

calibrations were comparable to the TUCKER3 model calibration performances (Fig. 7) in accuracy and precision as it can be seen from RMSEP values and pooled standard deviations in Table 1 and Fig. 9. The N-PLS fingerprint mode loadings were also more comparable to the TUCKER3 loadings (Fig. 6). PARAFAC showed quite some deviations in that case.

5. Discussion

Three multiway models have been investigated to determine enzyme activities of PL and PME simultaneously, namely PARAFAC, TUCKER3 and N-PLS. Kinetic mode loadings as well as spectral fingerprint loadings have been presented in Figs. 8 and 6. While it was possible to establish enzyme activity assays using all three multiway methods deviations between the models occurred, especially in the spectral fingerprint mode. In particular the PARAFAC fingerprint loading of PL indicated spectral changes which were not present in the actual spectral evolution. This led to decrease of calibration performance in terms of accuracy and precision as it can be seen in Table 1 and Fig. 9. In Fig. 9 the pooled standard deviation s_p of the calibration samples and the RMSEP (prediction samples) have been normalized over the activity calibration ranges of PL and PME, respectively, to make them comparable. When comparing accuracy and precision of the PARAFAC model with TUCKER3 and N-PLS performance, PL clearly performed worse than PME due to the given reasons. Overall, precision and accuracy of the TUCKER3

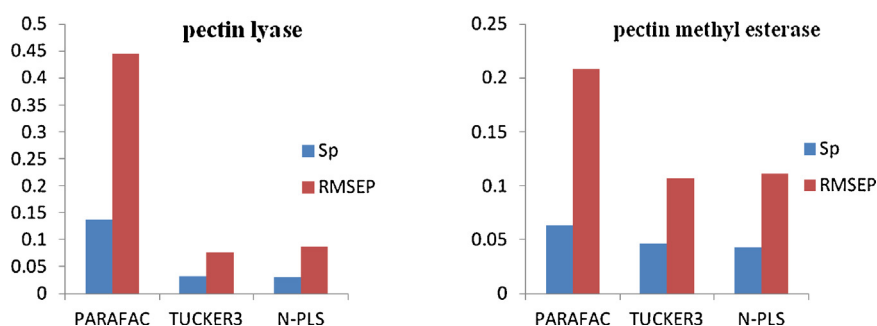


Fig. 9. Pooled standard deviation s_p (experimental precision) and root mean square error of prediction RMSEP (accuracy) are given for (a) pectin lyase and (b) pectin methyl esterase comparing the three multiway models. Both parameters have been normalized over calibration range. Accuracy and precision have been worst for PARAFAC considering both enzymes. TUCKER3 and N-PLS performance was comparable.

model were comparable to the N-PLS models which are generally more acceptable since they result from supervised regression.

While N-PLS performed slightly better for cross validated calibration set, TUCKER3 was even a bit superior when predicting the test set (RMSEP). However, the deviations are not likely to be significant since the sample size was not big enough to draw evident conclusions from it.

Nonetheless, it is surprising that TUCKER3 performed so well, even in comparison to N-PLS since it is a non-supervised regression method. This has clear advantages since PLS methods may cause over fitting under certain conditions. This cannot occur when using TUCKER3 (or PARAFAC).

6. Conclusion

The present study has shown that spectral evolution profiling using FTIR in combination with chemometric multiway analysis provides a useful and feasible approach to quantify enzyme activity of two co-acting enzymes simultaneously. When comparing the three multiway models, i.e. PARAFAC, TUCKER3 and N-PLS, mainly TUCKER3 surprises with its outstanding performance. As both, accuracy and precision were comparable to the regression results from N-PLS, although resulting from an unsupervised analysis; it shows clearly that TUCKER3 has great potential when analyzing kinetic data. This is especially valuable when PARAFAC, as an alternative unsupervised analysis, struggles to resolve the underlying phenomena as it could be seen during this study.

The methodology shows major advantages due to its universal applicability which shall be tested further in the near future. In addition, it offers the possibility to be implemented in high-throughput instrumentation which paves the road to industrial implemented solutions. Since many enzymatic reactions still rely on monitoring only one substrate or product the here presented approach enables much better observation of the whole reaction pattern. This is particularly important when enzyme kinetics is to be investigated on genuine substrates where two (or more) enzymes act concerted on the same substrate. At present those reactions can hardly be monitored by univariate measurement principles as they result in very complex reaction patterns. Time-resolved multivariate analysis as well as FTIR spectroscopy therefore inaugurates new possibilities to monitor such reactions to gain a better understanding of their complex underlying nature.

Acknowledgements

This work was partially financed by the 7th Framework Program via the Marie Curie Initial Training Network, LeanGreenFood (EU-ITN 238084). Further acknowledgement shall be given to Rasmus Bro for the continuous constructive support during the project work.

References

- [1] P. Nagaraja, H. Krishna, A. Shivakumar, A.K. Shrestha, *Clin. Biochem.* 45 (2011) 139–143.
- [2] F. Li, C.K. Lim, T.J. Peters, *Biomed. Chromatogr.* 2 (2005) 164–168.
- [3] J. Øbro, I. Sørensen, P. Derkx, C.T. Madsen, et al., *Proteomics* 9 (2009) 1861–1868.
- [4] H.A. Gavilighi, A.S. Meyer, J.D. Mikkelsen, *Biotechnol. Lett.* 35 (2013) 205–212.
- [5] C. Adina, F. Florinela, T. Abdelmoumen, S. Carmen, *Rom. Biotechnol. Lett.* 15 (2010) 5739–5744.
- [6] K. Karmali, A. Karmali, A. Teixeira, M.J.M. Curto, *Anal. Biochem.* 333 (2004) 320–327.
- [7] K. Karmali, A. Karmali, A. Teixeira, M.J.M. Curto, *Anal. Biochem.* 331 (2004) 115–121.
- [8] A. Baum, A.S. Meyer, J.L. Garcia, M. Egebo, et al., *Anal. Chim. Acta* 778 (2013) 1–8.
- [9] L.V. Thomassen, D.M. Larsen, J.D. Mikkelsen, A.S. Meyer, *Enzyme Microb. Technol.* 49 (2011) 289–297.
- [10] H.C. Buchholt, T.M.I.E. Christensen, B. Fallesen, M.-C. Ralet, et al., *Carbohydr. Polym.* 58 (2004) 149–161.
- [11] R. Bro, Ph.D. Thesis, Faculty of Science, University of Copenhagen, 1998.
- [12] R. Bro, *Crit. Rev. Anal. Chem.* 36 (2006) 279–293.
- [13] R. Bro, *Chemom. Intell. Lab. Syst.* 38 (1997) 149–171.
- [14] H. Abdollahi, S. Sajjadi, *Chemom. Intell. Lab. Syst.* 103 (2010) 144–151.
- [15] L.R. Tucker, *Contrib. Math. Psychol.* (1964) 109–127.
- [16] R. Henrion, *Chemom. Intell. Lab. Syst.* 25 (1994) 1–23.
- [17] P.M. Kroonenberg, J. De Leeuw, *Psychometrika* 45 (1980) 69–97.
- [18] C. Durante, M. Cocchi, M. Grandi, A. Marchetti, et al., *Chemom. Intell. Lab. Syst.* 83 (2006) 54–65.
- [19] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130.
- [20] R. Bro, *J. Chemometr.* (1996) 47–61.
- [21] M.W. Browne, *J. Math. Psychol.* 44 (2000) 108–132.
- [22] N.Q. Khanh, E. Ruttkowski, K. Leidinger, H. Albrecht, M. Gottschalk, *Gene* 106 (1991) 71–77.
- [23] S. Christgau, L.V. Kofod, T. Halkier, L. Andersen, et al., *Biochem. J.* 319 (1996) 705–712.
- [24] L.A. van den Broek, E.D. den Aantrekker, A.G. Voragen, G. Beldman, et al., *J. Sci. Food Agric.* 75 (1999) 167–172.
- [25] R. Edstrom, H. Phaff, *J. Biol. Chem.* 239 (1964) 2403–2408.
- [26] S.-H. Yoo, M.L. Fishman, B.J. Savary, A.T. Hotchkiss Jr., *J. Agric. Food Chem.* 51 (2003) 7410–7417.
- [27] R. Bro, A.K. Smilde, *J. Chemometr.* 17 (2003) 16–33.
- [28] W.T. Federer, *Experimental Design*, 1955.