

FACULTY OF SCIENCE

UNIVERSITY OF COPENHAGEN



# Applications of Raman Spectroscopy and Multivariate Data Analysis in Food and Pharmaceutical Sciences

LOTTE BØGE LYNDGAARD

PHD THESIS

2013



# Applications of Raman Spectroscopy and Multivariate Data Analysis in Food and Pharmaceutical Sciences



PHD THESIS

BY

Lotte Bøge Lyndgaard

**Title:**

Applications of Raman Spectroscopy and Multivariate Data Analysis in Food and Pharmaceutical Sciences

**Submission:****Defence:****Supervisors:**

Associate Professor Frans van den Berg  
&  
Professor Rasmus Bro  
Quality and Technology  
Department of Food Science  
Faculty of Science  
University of Copenhagen

**Opponents:**

Professor Thomas de Beer  
Laboratory of Pharmaceutical Process Analytical Technology  
Faculty of Pharmaceutical Sciences  
Ghent University Belgium

Associate Professor Søren Hassing  
Department of Chemical Engineering,  
Biotechnology and Environmental Technology  
Faculty of Engineering  
University of Southern Denmark

Associate Professor José Manuel Amigo  
Quality and Technology  
Department of Food Science  
Faculty of Science  
University of Copenhagen

**Cover illustration:**

Image by Dick Wieboldt

# Preface

This PhD project was carried out in the research group Quality and Technology (Q&T) at the Department of Food Science, Faculty of Science, University of Copenhagen. The work was sponsored by the Technology and Innovation foundation through the consortium QbD – ‘Quality by Design in the Food and Pharmaceutical Industries’. The project was supervised by Associate Professor Frans van den Berg and Professor Rasmus Bro from the University of Copenhagen.

I am particular grateful to my supervisors for giving me the opportunity to do a PhD in the Quality and Technology group and for sharing their knowledge and ideas. They have both inspired me and supported me in different ways and levels. A special thanks to Frans for always being available and helping me through all phases of this project.

Gratitude is shown to all the people I worked with during my three month stay at the Department of Analytical Chemistry, Universitat de Barcelona. I am very grateful to my supervisor during the stay Associate Professor Anna de Juan for introducing me to Multivariate Curve Resolution and for her very warm and kind personality. Furthermore, Elisa, Silvia, Sara, and Messi are also thanked for making my stay a good experience both scientifically and socially. I also wish to thank Rolf Spångberg at Qpharma and Christopher Gilmour from International partnership of microbicides for good collaboration in the vaginal ring project. Klavs Martin Sørensen and people at Carometec are greatly acknowledged for letting me be part of the NitFom project.

I am very thankful to my colleagues at Q&T for contributing to a perfect working environment both socially and academically. Acknowledge to Professor Søren Balling Engelsen for inspiring and valuable discussions about spectroscopy. The company of my office mate Signe Hoff has been greatly appreciated at all times – celebrating victories and supporting through rough times – and lovely coffee breaks at our office balcony. A special thanks to Maja Kamstrup-Nielsen for your support and great friendship, and for proofreading the thesis.

I am grateful for the sympathy and support from my family. The greatest thanks to my wonderful husband Christian - your love, care, and incredible support has been so important to me. Finally, thanks to my lovely daughter Luna for taking my mind elsewhere and making me remember what is most important in life.

Lotte Bøge Lyndgaard,  
Frederiksberg, November 2013

## Abstract

Raman spectroscopy is a light scattering technique which measures molecular vibrations. It can provide unique information of chemical composition and has advantageous properties such as being non-invasive, rapid, and requiring no or minimal sample preparation. These properties, along with a huge development in instrumentation in recent years, have given Raman spectroscopy a strong position among the leading analytical techniques in a wide range of scientific areas.

The aim of this thesis is to investigate the practical utility of non-resonance, vibrational Raman spectroscopy and multivariate data analysis for analyzing the quality of food and pharmaceutical products through feasibility studies. Furthermore, the thesis aims at exemplifying and investigating the use of multivariate data analysis for handling artefacts and extracting information from Raman spectroscopic data. Finally, the thesis seeks to evaluate the advantages and limitations of Raman spectroscopy for the studied applications. The work is organized in two main parts. Part I contains an introduction to Raman spectroscopy including a brief description of the Raman effect, Raman instrumentation, spectral interpretation, and data analysis. Literature reviews related to the feasibility studies are presented including Raman studies of food lipids, quantitative Raman studies of active pharmaceutical ingredient (API) in common solid dosage formulations (tablet/capsule) and more special dosage formulations (ointment/patch/elastomers). Finally, Raman analysis through packaging is reviewed. Part II consists of three research papers with quite diverse applications of Raman spectroscopy, which are summarized below.

The quality of pork carcasses greatly depends on the composition and thickness of backfat which is organized in two layers. Pork adipose tissue consists primarily of triglycerides of saturated, unsaturated, and polyunsaturated fatty acids. The composition of these fatty acids varies with the depth of the fat tissue. In the outer layer (just below the skin) there is a large fraction of unsaturated fatty acids. Consequently, if this layer becomes too thick, and with a large fraction of unsaturated fat, it can lead to an oily appearance, rancidity development and reduced slicing yield. Thus, an analytical tool to determine the thickness and composition of the outer layer is valuable for both the meat processing industry and consumer acceptability. In **Paper I**, depth profiles (from skin surface to meat) of adipose tissue of 16 pigs were studied by Raman spectroscopy. Principal component analysis showed that it is possible to discriminate between fat layers by Raman spectroscopy. The separation of layers was mainly based on variations in the C=C stretching and =C-H twisting bands of the Raman spectrum attributed to variation in fraction of unsaturation. The results suggest that Raman spectroscopy could be applied as a potential rapid measurement technique for grading of pork carcasses.

Raman spectroscopy has potential for non-invasive detection and quantification of chemical solids in concealing packaging e.g. plastic bottles, bags, or blister packs. Evaluation of pharmaceutical product composition through packaging material has a multitude of advantages, but successful quantification of paracetamol through a blister package has never been proved. **Paper II** shows that the concentration of the API paracetamol in a ternary mixture can be predicted through a blister package using a combination of Raman spectroscopy and multivariate curve resolution-alternating least squares (MCR-ALS). In a concentration range of 0-20% w/w, paracetamol was predicted with a root-mean-square-error-of-prediction (RMSEP) of 0.8% w/w. Quantification inside a blister package can be challenging, if the blister package material varies between production batches, as it will interfere with the Raman signal. To overcome this problem, a novel modification of the correlation constraint in MCR-ALS is proposed for handling multiple group (multiset) matrix effects. The results showed that by this new multivariate approach, quantification of paracetamol was possible despite blister package variations. The method may also be suitable for reducing signal interferences in other sample matrices.

Human immunodeficiency virus (HIV) is the world leading cause of death for women in the age range from 15-44. Women in the sub-Saharan countries are especially vulnerable due a mix of biology and social factors, and there are urgent needs for a tool, which can help these women protect themselves. For this purpose, an HIV preventive polymeric drug delivery device (an intra-vaginal ring) for women was developed by the International partnership of microbicides. In **Paper III**, Raman spectroscopy was applied to quantify the API dapivirine in the intra-vaginal rings with the purpose of testing the technique as a rapid, cheap, and non-invasive alternative to the current high pressure liquid chromatography (HPLC) standard method. Customized reference rings of known dapivirine concentrations were used to build calibration models, and the models were verified by rings from the production using the HPLC as reference analysis. A data analysis approach using the ratios of specific Raman bands proved to work similarly well as the multivariate approach (partial least squares regression) resulting in low prediction errors (RMSEP). The accuracy of the results indicated that Raman spectroscopy can be used as a faster and cheaper alternative to HPLC. Additionally, a fast and non-destructive measurement opens up for testing all production samples automatically through continuously monitoring of the production process performance (process capability). On top of this it was shown that Raman, in combination with a ring spinning module, could provide an estimate of how homogenous the API was distributed within the ring.

In conclusion, the research of this thesis has shown the practical utility of Raman spectroscopy for analyzing the quality of food and pharmaceutical products. Furthermore, in this thesis new ways of using multivariate methods for handling artefacts and extracting more information from Raman spectra have been developed and demonstrated.

# Sammendrag

Ramanspektroskopi er en lysspredningsteknik, der måler molekylære vibrationer. Den kan give unik information om kemisk sammensætning og har fordelagtige egenskaber såsom at være ikke-invasiv, hurtig og med ingen eller minimal prøveforberedelse. Disse egenskaber, sammenholdt med en stor udvikling i Ramaninstrumenter i de seneste år, har givet Ramanspektroskopi en stærk position blandt de førende analytiske teknikker indenfor en lang række videnskabelige områder.

Formålet med denne afhandling er, at undersøge den praktiske anvendelse af ikke-resonans, vibrationel Ramanspektroskopi og multivariat dataanalyse til kvalitetsanalyse af fødevare- og farmaceutiske produkter gennem feasibility studier. Endvidere, forsøger afhandlingen at eksemplificere og undersøge brugen af multivariat dataanalyse til håndtering af artefakter og til udtrækning af information fra Raman spektroskopisk data. Endelig evalueres fordele og begrænsninger ved Ramanspektroskopi for de undersøgte anvendelser. Afhandlingen er organiseret i to hoveddele. Del I indeholder en introduktion til Ramanspektroskopi som inkluderer en kort beskrivelse af Raman effekten, instrumentbeskrivelse, spektralfortolkning, og dataanalyse. Litteratur-gennemgange relateret til feasibility studierne er præsenteret inkluderende Ramanstudier af fødevarelipider, kvantitative Ramanstudier af aktive farmaceutiske ingredienser (API) i almindelige/typiske faste stoffers doseringsformer (tabletter/kapsler) og i mere specielle doseringsformer (salver/plastre/elastomere). Endelig er Raman analyse gennem emballage undersøgt. Del II består af tre videnskabelige artikler med forskellige applikationer af Ramanspektroskopi, som er sammenfattet i nedenstående.

Kvaliteten af slagtesvin afhænger i høj grad af sammensætningen og tykkelsen af rygfedtet, som er organiseret i to lag. Svinefedtvæv består primært af triglycerider af mættede, umættede og polyumættede fedtsyrer. Sammensætningen af disse fedtsyrer varierer med dybden af fedtvævet. I det ydre lag (lige under flæskesværen) er der en stor mængde af umættede fedtsyrer. Hvis dette lag bliver for tykt og med en for stor mængde umættet fedt, kan det føre til et olieagtigt udseende, udvikling af harskhed samt et reduceret skæringsudbytte. Derfor er et analyseværktøj til at bestemme tykkelse og sammensætning af det ydre fedtlag værdifuldt for både slagteindustrien og forbrugeraccepten. I **Artikel I** er dybdeprofilen (fra hudoverflade til kød) af fedtvæv fra 16 svin undersøgt med Ramanspektroskopi. Principal komponent analyse viste at det er muligt at skelne mellem fedtlagene med Ramanspektroskopi. Separationen mellem lagene var hovedsagelig baseret på variationerne i C=C stræknings- og =C-H twisting-båndene i Ramanspektret, hvilket kan tillægges variation i graden af umættethed. Disse resultater tyder på, at Ramanspektroskopi kan anvendes som en potentiel hurtig måleteknik til sortering af slagtesvin.

Ramanspektroskopi har potentiale for ikke-invasiv detektion og kvantificering af faste kemiske stoffer/tabletter pakket i emballage som f.eks. plastikbeholdere, poser eller blisterpakker. Undersøgelse af sammensætningen af farmaceutiske produkter gennem emballage har en lang række fordele, men vellykket kvantificering af paracetamol igennem en blisterpakke har aldrig været vist. **Artikel II** viser, at koncentrationen af API'en paracetamol i en ternær pulverblanding kan prædikteres gennem en blisterpakning ved at bruge en kombination af Ramanspektroskopi og multivariate curve resolution-alternating least squares (MCR-ALS). I et koncentrationsområde på 0-20% w/w var paracetamol prædikteret med en prædiktionsfejl (RMSEP) på 0,8% w/w. Kvantificering i en blisterpakke kan være udfordrende, hvis blisterpakkingsmaterialet varierer mellem produktionsbatcher, idet det vil interferere forskelligt med Ramansignalet. For at overvinde dette problem blev en helt ny modificering af korrelationsrestriktionen i MCR-ALS udviklet til at håndtere flergruppematricoeffekter (multisæt). Resultaterne viser, at med den nye multivariate fremgangsmåde var kvantificering mulig, på trods af variationer i blisterpakkemateriale. Metoden kan også tænkes at være passende til at reducere signalinterferenser i andre prøvematrixer.

Human immunodeficiency virus (HIV) er den hyppigste dødsårsag blandt kvinder i alderen 15-44 år på verdensplan. Kvinder i lande syd for Sahara er specielt udsatte på grund af et mix af biologiske og sociale faktorer, og der er akut behov for et redskab, som kan hjælpe disse kvinder med at kunne beskytte sig selv. Til dette formål er en HIV-beskyttende polymer til medicinering (en intra-vaginalring) udviklet til kvinder af International partnership of microbicides. I **Artikel III** blev Ramanspektroskopi anvendt til at kvantificere API'en dapivirine i intra-vaginalringe med det formål at teste teknikken som et hurtigt, billigt og ikke-invasivt alternativ til den nuværende HPLC standardmetode. Specialfremstillede referenceringsring med kendte dapivirinekoncentrationer blev brugt til at bygge en kalibreringsmodel, og modellen blev verificeret med ringe fra produktionen ved brug af HPLC som referenceanalyse. En dataanalytisk fremgangsmåde hvor forholdet mellem specifikke Ramanbånd var brugt, viste sig at virke ligeså godt som den multivariate fremgangsmåde (PLS) hvor begge resulterede i lave prædiktionsfejl. Nøjagtigheden af resultaterne indikerede, at Ramanspektroskopi kan anvendes som et hurtigt og billigt alternativ til HPLC. Endvidere giver en hurtig og ikke-destruktiv måling mulighed for at teste alle produktionsprøver automatisk gennem en kontinuerlig monitorering af produktionens procesdydeevne. Ydermere blev det vist, at Raman, i kombination med et ring-spindingsmodul, kunne give et estimat for, hvor homogent API var fordelt rundt i ringen.

Forskningen i denne afhandling har vist den praktiske anvendelighed af Ramanspektroskopi til analyse af kvaliteten af fødevarer- og farmaceutiske produkter. I denne afhandling er der endvidere udviklet og demonstreret nye måder, hvormed man ved hjælp af multivariate metoder kan håndtere artefakter og udtrække kemisk information fra Ramanspektret.



# List of Publications

## PAPER I

Lyndgaard, L.B., Sørensen, K.M., van den Berg F. & Engelsen, S.B. (2012). Depth profiling of porcine adipose tissue by Raman spectroscopy. *Journal of Raman Spectroscopy*, 43, 482–489

## PAPER II

Lyndgaard, L.B., van den Berg, F. & de Juan, A. (2013). Quantification of paracetamol through tablet blister packages by Raman spectroscopy and MCR-ALS, *Chemometrics and Intelligent Laboratory Systems*, 25, 58-66

## PAPER III

Lyndgaard, L.B., Spångberg, R., Gilmour, C., Lyndgaard, C.B. & van den Berg, F. (2013). A process analytical approach for quality control of dapivirine in HIV preventive vaginal rings by Raman spectroscopy, Accepted to *Journal of Raman Spectroscopy*.

## Additional work by the author

## POSTER I

Lyndgaard, L.B., Andersen, K.K., Christensen, M., Acar, E., Egebo, M. & Bro, R. (2011). Predicting coagulation properties of milk with seasonal variations by mid-infrared spectroscopy. Awarded best poster at 12<sup>th</sup> Scandinavian Symposium on Chemometrics, Billund.

# Table of contents

<b>PREFACE</b> .....	<b>I</b>
<b>ABSTRACT</b> .....	<b>II</b>
<b>SAMMENDRAG</b> .....	<b>IV</b>
<b>LIST OF PUBLICATIONS</b> .....	<b>VII</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 BACKGROUND .....	1
1.2 AIMS AND SCOPE OF THE THESIS .....	2
1.3 OUTLINE OF THESIS .....	4
<b>2. THE RAMAN EFFECT</b> .....	<b>5</b>
2.1 LIGHT SCATTERING .....	5
2.2 POLARIZABILITY .....	8
2.3 POLARIZABILITY TENSOR .....	10
2.4 MOLECULAR VIBRATIONS .....	11
2.5 RAMAN SPECTROSCOPY COMPARED TO OTHER SPECTROSCOPIC TECHNIQUES .....	13
<b>3. RAMAN INSTRUMENTATION</b> .....	<b>17</b>
3.1 OVERVIEW OF INSTRUMENT COMPONENTS AND THE LIGHT PATH .....	17
3.2 LASERS .....	17
3.4 INSTRUMENTAL DESIGN .....	20
3.5 LASER-LINE REJECTION .....	21
3.6 SPECTROGRAPH .....	22
3.7 SAMPLING INTERFACES (PROBES) .....	24
3.8 ACQUISITION PARAMETERS .....	28
<b>4. THE RAMAN SPECTRUM</b> .....	<b>31</b>
4.1 DISPLAY .....	31
4.2 SPECTRAL INTERPRETATION .....	31
4.3 SOURCES OF NOISE .....	35
4.4 TEMPERATURE AND PRESSURE EFFECTS .....	38
4.5 BAND FITTING .....	39
<b>5. RAMAN DATA ANALYSIS</b> .....	<b>43</b>
5.1 UNIVARIATE VS. MULTIVARIATE .....	43
5.2 DATA PREPROCESSING .....	44
5.2.1 <i>Removal of background</i> .....	45
5.2.2 <i>Removal of noise</i> .....	47
5.2.3 <i>Normalization (object-wise)</i> .....	48

5.3 QUALITATIVE ANALYSIS.....	51
5.3.1. <i>Principal component analysis (PCA)</i> .....	51
5.4 QUANTITATIVE RAMAN SPECTROSCOPY .....	54
5.4.1 <i>Partial Least Squares (PLS) regression</i> .....	55
5.4.2 <i>Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS)</i> .....	56
<b>6. RAMAN ANALYSIS OF FOOD AND PHARMACEUTICAL PRODUCTS .....</b>	<b>61</b>
6.1 RAMAN SPECTROSCOPY OF FOODS .....	62
6.1.1 <i>Food macro components</i> .....	62
6.1.2 <i>Food lipids</i> .....	63
6.1.3 <i>Raman scattering properties of food lipids</i> .....	65
6.1.4 <i>Overview of Raman studies of edible fats and oils</i> .....	69
6.2 RAMAN ANALYSIS OF ACTIVE PHARMACEUTICAL INGREDIENTS (API) IN DIFFERENT MATRICES .....	70
6.2.1 <i>Raman scattering properties of API and excipients</i> .....	70
6.2.2 <i>Overview of quantitative studies of API in solid dosage formulations</i> .....	71
6.2.3 <i>Raman analysis through packaging</i> .....	73
6.2.4 <i>Raman analysis of API in other solid and semisolid dosage formulations</i> .....	75
<b>7. CONCLUSIONS AND PERSPECTIVES.....</b>	<b>77</b>
<b>REFERENCES.....</b>	<b>83</b>

**PAPERS I-III**

**POSTER I**



# 1. INTRODUCTION

## 1.1 Background

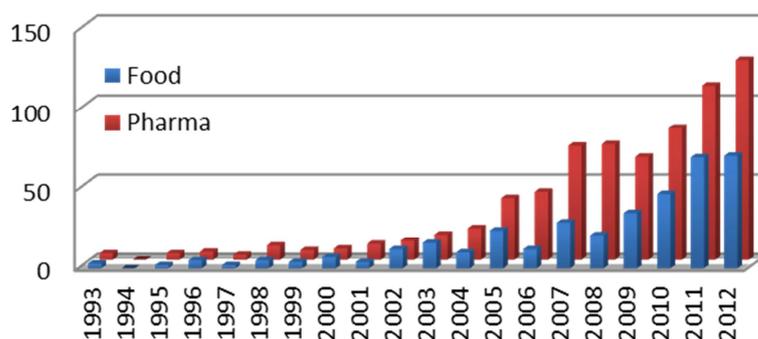
Have you ever taken a bite of a bad piece of fruit? Would it not be nice to be able to determine whether this fruit was good or bad before you experience the taste? Most food product testing is done destructively, which means that the product e.g. has to be consumed to determine whether it is good or bad. One of the problems is that you can only show whether or not that particular piece/sample was good, but you cannot tell whether the next piece will be good. This thesis will introduce you to a technique called Raman spectroscopy that has the potential to optimize quality testing for a broad range of products in the food and pharmaceutical industry.

Consider what happens to laser light hitting a sample. Some of the light is transmitted, some is absorbed, and some is scattered. Most of the scattered light will have the same wavelength as the incident light (albeit travelling in a different direction), but a small fraction of the scattered light is shifted in wavelength because the vibrations and rotations of the molecules in the sample have interacted with a portion of the light. The spectrum of this wavelength-shifted light is called the Raman spectrum. The Raman spectrum contains bands that are characteristic and proportional to concentrations of specific molecules in the sample, thus Raman provides a good basis for qualitative and quantitative analysis.

Raman spectroscopy originates from 1928 where Krishna and Raman discovered scattering at the molecular level. They found out that sunlight that interacts with matters creates a unique pattern. For many years, Raman spectroscopy has been a recognized research technique largely confined to experts in dedicated academic or industrial research laboratories. The instruments were large and complicated, and the experiments could be quite complex. However, in the past decades new developments in the Raman instrumentation (e.g. improved lasers, introduction of charge-coupled device (CCD) detectors and fiber optics) have expanded the possibilities for the use of Raman spectroscopy. With these advances, fundamental problems of a weak Raman signal and fluorescence interference can be overcome.

Raman spectrometers have become smaller, sometimes even portable, and easy-to-use, with an analysis time reduced to seconds. Furthermore, non-destructive analysis can be provided without the necessity of sample preparation. Chemometrics has since the 1970s added value to multivariate analytical chemical techniques in many ways. Raman spectroscopy is no exception, as chemometric tools have made it possible to analyze Raman spectral data of increasing complexity. The combination of these developments has made Raman spectroscopy applicable to a wide range of scientific areas. This is exemplified by the steep increase in the number of peer-reviewed

publications of Raman spectroscopy in food science and in pharmaceutical science through the last decades as shown in Figure 1.



**Figure 1.** Numbers of published peer-reviewed papers for the last two decades within Raman spectroscopy and food or pharmaceutical sciences (Web of Science).

The pharmaceutical industry has been encouraged to use process analytical technology (PAT) in order to increase process understanding, optimize manufacturing efficiency, and boost reproducibility of product quality by the PAT guidance issued in 2004 by US food and drug administration (FDA, 2004). One of the key elements in implementing PAT (in a production) is to apply process analyzers such as rapid and non-destructive spectroscopic methods like Raman spectroscopy. Despite the many advantages as an analytical tool for quality control in both food and pharmaceutical productions, Raman spectroscopy is a less recognized method than comparable optical methods such as ultraviolet-visual (UV-VIS), near-infrared (NIR), and infrared (IR) spectroscopy. This thesis seeks to elucidate some of the possibilities of modern Raman spectroscopy within food and pharmaceutical sciences.

## 1.2 Aims and scope of the thesis

The following aims are identified; all pertaining to non-resonance, vibrational Raman spectroscopy. In this thesis I will:

- *Exemplify and investigate the usefulness of multivariate data analysis for handling artefacts and extracting information from Raman spectroscopic data.*
- *Investigate the practical utility of Raman spectroscopy for analyzing the quality of food and pharmaceutical products through feasibility studies.*
- *Evaluate the advantages and limitations of Raman spectroscopy for the applications investigated in the feasibility studies.*

On the basis of three feasibility studies, this thesis seeks to introduce the reader to some fundamental aspects to consider when applying Raman spectroscopy for specific

food and pharmaceutical applications, focusing on quantitative and instrumental properties along with the challenges in spectral interpretation, preprocessing and multivariate data analysis. The motivation, objective and methodology for each of the three feasibility studies are presented in the following sections.

Raman spectroscopy measures how easily an electron cloud of a molecule can be distorted (change in polarizability) during a vibration. The polarizability of molecules with triple or double bonds (e.g. carbon-carbon double bonds) can easily change during vibration, because they have a high electron density between the atoms; their electron cloud can thus easily be distorted. As a result, the Raman signal for the vibrations of such molecules is strong. This strong signal is found for the double bonds between carbons present in unsaturated fatty acids (UFA) making Raman spectroscopy an excellent method for compositional analysis of fat and edible oils. In **Paper I**, Raman spectroscopy is applied to analyze pork backfat, whose thickness and composition affect a number of quality parameters. The meat processing industry is highly competitive, and grading meat in different qualities is extremely important. The backfat of pork consists of two different layers: an outer layer with a high degree of unsaturation and the inner layer with a lower degree of unsaturation. Raman spectroscopy combined with pre-processing and classification methods was proposed as a fast method to discriminate between the fat layers of carcasses. To further characterize fatty acid composition Raman band fitting and multivariate calibration was investigated.

The ability to evaluate the pharmaceutical product composition through sealed packing has a multitude of advantages: revealing counterfeit products, controlling deterioration due to damaged packing material, testing of products subjected to extreme storage conditions, or testing the quality close to the product release. Raman spectroscopy has the potential for identifying materials through their packaging. This has been shown in several qualitative studies, but the number of studies where products are not only identified but quantified is very limited. In **Paper II**, the use of Raman spectroscopy and multivariate curve resolution-alternating least squares (MCR-ALS) was investigated with the purpose of quantifying the active pharmaceutical ingredient (API) paracetamol in pharmaceutical mixtures inside blister packages.

It is projected that 34 million people live with the human immunodeficiency virus (HIV) worldwide. The problem is particular widespread in sub-Saharan Africa, where young women are especially vulnerable, and at least twice as likely to become infected as young men. As a practical and discreet tool to help women protect themselves from HIV infection, the International partnership of microbicides invented an intra-vaginal ring containing the API dapivirine. In **Paper III**, the objective was to test Raman spectroscopy as a fast and non-invasive alternative to the current standard reference method (high-performance liquid chromatography) in

quantifying dapivirine in intra-vaginal rings. In the quantification, the use of a novel sampling system of wide-area illumination Raman and sample rotation in a special built ring spinner was investigated. An additional aim was to explore the multiple possibilities of using Raman spectroscopy in the process and quality control of intra-vaginal rings.

### **1.3 Outline of thesis**

The thesis consists of an introductory part (**Chapters 1-7**), which serves as introduction to three scientific papers (**Papers I, II and III**). In the introductory part, the reader is presented with major aspects of Raman spectroscopy important for successful application. This includes an introduction to the Raman effect (**Chapter 2**), instrumentation (**Chapter 3**), spectral interpretation (**Chapter 4**), and data analysis (**Chapter 5**). A literature review of Raman studies of food lipids is provided along with a review of the quantitative Raman studies of active pharmaceutical ingredient (API) in common solid dosage formulations (tablet/capsule) and more special dosage formulations (ointment/patch/elastomers). In addition, Raman analysis through packaging is reviewed (**Chapter 6**). Finally, the conclusions and perspectives of this PhD study are summarized (**Chapter 7**).

## 2. The Raman Effect

The following chapter provides a very brief introduction to Raman spectroscopy as utilized in this thesis. This chapter is not intended as comprehensive description, but rather as an introduction for people that are not familiar with Raman spectroscopy. For a more in-depth description, the reader is referred to the literature (Long, 2002).

Raman spectroscopy is increasingly being used for quantitative purposes such as industrial quality control in particular in the pharmaceutical industry. The continuous developments in Raman instrumentation have made it simple to use, and today Raman spectroscopy can be applied for analyzing many types of samples without having a deep understanding of the nature of the Raman effect. However, some basic knowledge of the Raman effect can be advantageous in assessing how suitable Raman spectroscopy is when analyzing a specific type of sample, what measurement conditions should be considered, and how spectra can be interpreted and analyzed. In addition, an increased understanding might facilitate that much more information about a molecule and its surroundings can be obtained, enhance the interpretation, recognize and avoid possible pitfalls, and the background required to understand some more modern developments such as surface enhanced Raman spectroscopy (SERS) will become clearer. Therefore, this chapter will provide a basic description of Raman spectroscopy.

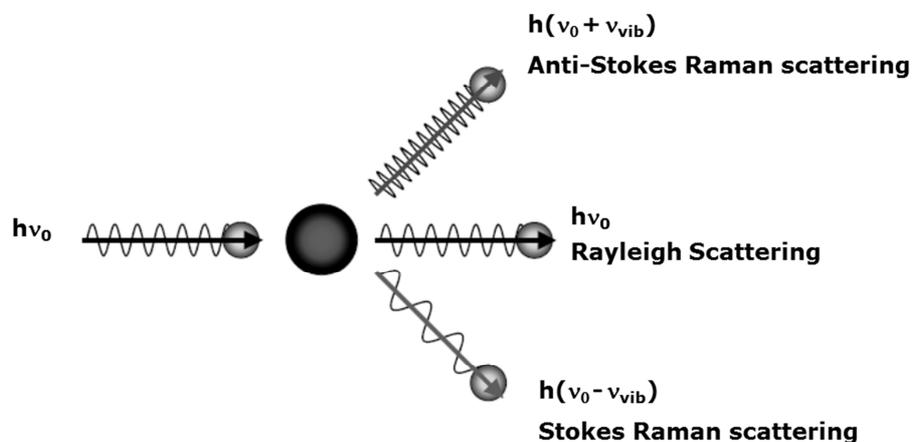
Theoretically, there are two different ways to describe the phenomenon of Raman scattering: “the classical theory” or “the quantum theory”. The classical theory is based on wave theory of light, and can be very useful in understanding one of the essential concepts in Raman spectroscopy – the polarizability of a molecule. However, the relationship between molecular properties and Raman scattering is better explained by quantum theory (Sasic & Ozaki, 2007; Smith & Dent, 2005). Therefore, the following chapter is based on a combination of the two.

### 2.1 Light scattering

When light interacts with matter, it can be transmitted, absorbed or scattered. When light is scattered from a molecule, most of the photons are elastically scattered, which means that the scattered light has the same energy as the incident light (Rayleigh scattering). However, a tiny fraction of the light is inelastically scattered i.e. at frequencies different from the frequency of the incident light. This process is known as the Raman effect. Raman scattering can occur due to changes in the electronic, rotational, or vibrational energy of a molecule (McCreery, 2000). The present thesis will consider the vibrational Raman scattering only.

Light can be considered as a stream of particles called photons. Raman spectroscopy is often thought of in terms of a collision model, where photons can be imagined to

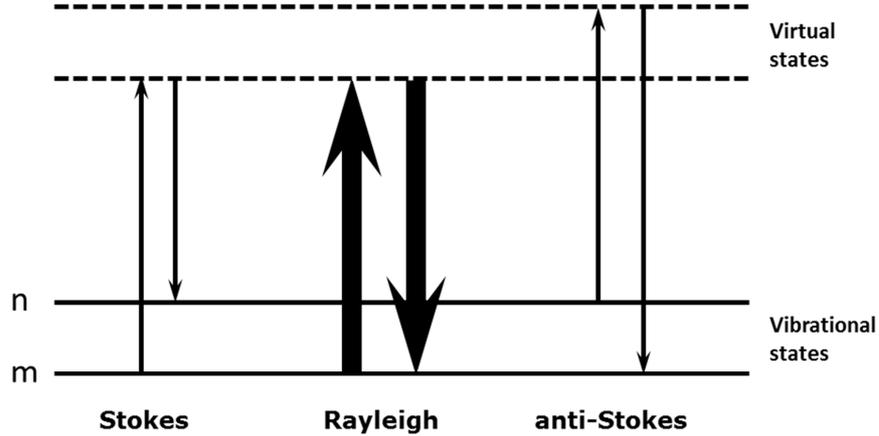
undergo a collision with molecules. A photon-molecule collision is depicted in Figure 2. If an incident photon delivers an  $h\nu$  quantum of energy to the molecule, the energy of the scattered photon is reduced to  $h(\nu_0-\nu)$ , and the frequency of the scattering photon becomes  $\nu_0-\nu$ .



**Figure 2.** Collision between incident light and a molecule. After collision, most of the light is scattered with unchanged energy as Rayleigh scattering. A small fraction is scattered with a different energy (frequency) as Stokes and anti-Stokes Raman scattering (modified from Sasic & Ozaki, 2007).

On the contrary, when an incident photon gains  $h\nu$  energy from the molecule, the energy of the scattering photon rises to  $h(\nu_0+\nu)$ , and the frequency of the scattering photon becomes  $\nu_0+\nu$ . Scattering, in which an incident photon exchanges energy with a molecule, is known as Raman Scattering. Scattered light having a lower frequency ( $\nu_0-\nu$ ) than the incident light is called Stokes Raman scattering, while light scattered with higher frequency ( $\nu_0+\nu$ ) is called anti-Stokes Raman scattering (Figure 2) (Sasic & Ozaki, 2007; Banwell, 1983).

The energy exchange between incident light photons and illuminated target molecule arises because the photons interact and distort the electron cloud surrounding the nuclei of the molecule. The photons act as an electric field transferring energy to the molecule, causing the electrons to polarize (be distorted in an opposite direction compared to the nuclei) and go to a higher “virtual” energy level. This interaction can be considered a very short-lived (usually less than  $10^{-14}$  seconds) complex between light photons and electrons in the molecule, a timeframe in which the nuclei do not have time to move. This results in a high energy form of the molecule with altered electron geometry, but without any large nuclei movement. The complex is, however, not stable and the photon is re-emitted as scattered light with a small change in energy. The energy difference between the incident and the scattered photon corresponds to the energy needed for vibrational excitation of a given vibrational mode of a functional group (Smith & Dent, 2005).



**Figure 3.** A simplified Jablonski energy diagram of the Rayleigh and the Raman scattering processes of a given vibrational mode. The lowest energy vibrational state  $m$  (so-called ground state of the molecule) is shown at the foot with states of increasing energy above it. Upwards arrows denote the energy of incident photons, whereas downwards arrows denote the energy of scattered photons.

Numerically, the energy difference between the initial and the final vibrational states,  $\bar{\nu}$ , or Raman shift in wavenumbers ( $\text{cm}^{-1}$ ), is calculated through Equation 1.

$$\bar{\nu} = \left( \frac{1}{\lambda_{\text{incident}}} - \frac{1}{\lambda_{\text{scattered}}} \right) \times 10^7 \quad (1)$$

in which  $\lambda_{\text{incident}}$  and  $\lambda_{\text{scattered}}$  are the wavelengths (nm) of the incident and the Raman scattered photons, respectively. It is important to note that the Raman shift is independent of the incident laser frequency. The vibrational energy is ultimately dissipated as heat. However, the heat dissipation does not cause a measureable temperature change in a material due to the very low vibrational energy levels (Pelletier, 1999a).

The Raman effect is weak. It has been stated that the challenge of Raman spectroscopy is to detect the Raman “needle” in the Rayleigh “haystack”. This becomes evident by comparing the relative intensity of the incident light and the scattering as shown in Table 1.

**Table 1.** The relative intensities of incident light and scattering

Incident light	:	Rayleigh scattering	:	Raman scattering
1		$10^{-6}$		$10^{-10}$

Only approximately one out of  $10^8$  incident photons is shifted. Even fewer are anti-Stokes scattered because the majority of molecules are in the ground vibrational energy state at room temperature, and not in the excited state as required for generating anti-Stokes scattering (Figure 3). The Boltzmann distribution (Equation 2)

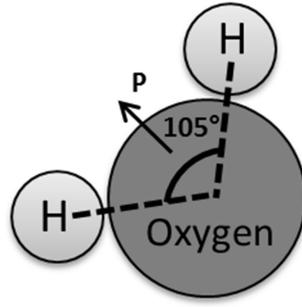
describes the relationship between temperature and the fraction of molecules in an excited state:

$$\frac{N_n}{N_m} = \frac{g_n}{g_m} \exp \frac{-\Delta E}{kT} \quad (2)$$

where  $N_n$  and  $N_m$  are the number of molecules in the excited and ground energy state, respectively,  $g_n$  and  $g_m$  are the degeneracies of the excited and ground vibrational state, respectively,  $k$  is Boltzmann constant,  $T$  is the temperature in kelvin, and  $\Delta E$  is the energy difference between the vibrational energy states. It can be deduced from the Boltzmann distribution that as temperature increases, the relative proportion of molecules in ground and excited states changes and the relative intensity between the Stokes and anti-Stokes changes accordingly. However, at room temperature most molecules are likely to be in the ground vibrational state, and therefore, the intensity of the Stokes Raman will always be much larger than the anti-Stokes Raman intensity. Even though the Stokes and the anti-Stokes spectra contain the same frequency information, the intensity of the Stokes Raman is usually much higher, and therefore typically only the Stokes Raman spectra are recorded (Pelletier, 1999a; Bakeev, 2010). As a curiosity, it can be mentioned that the anti-stokes Raman band are not prone to fluorescence interference, which turns out to be a major obstacle in Raman spectroscopy.

## 2.2 Polarizability

In order to obtain a better understanding of the concept of polarizability, some basic facts about the dipole moment of a molecule will be explained. To absorb or scatter light the so-called dipole moment in the molecule is important. The dipole moment is the center of the positive and the negative charges. Even though the total charge on a molecule might be zero, the nature of chemical bonds is such that the positive and negative charges do not completely overlap in most molecules. Such molecules are said to be polar because they possess a permanent dipole moment. A good example is the dipole moment of the water molecule (Figure 4). Molecules with mirror symmetry like oxygen, nitrogen, carbon dioxide, and carbon tetrachloride have no permanent dipole moments. Even if there is no permanent dipole moment, it is possible to induce one by the application of an external electric field. The dipole moment is induced because the electric field can change the distribution of the electrons slightly within the molecule, by pushing the negative electrons and the positive nuclei in opposite directions. This is called polarization, and the change of the dipole moment induced is a measure of the polarizability of the molecular species. A change in the dipole moment can both be in length and direction, since it is a vector (Keresztury, 2002).



**Figure 4.** The asymmetry of the water molecule leads to a permanent dipole moment,  $P$ , in the symmetry plane pointed toward the more positive hydrogen atoms.

In the classical theory of Raman spectroscopy, a dipole moment can be induced in the molecule by the incident radiation, since it presents an electric field. The magnitude of the induced dipole moment,  $P$ , depends both on the magnitude of the applied field  $E$  (the intensity of laser light), and on the ease with which the electron cloud in the molecule can be distorted. This can be expressed by Equation (3):

$$P = \alpha E \quad (3)$$

where  $\alpha$  is the polarizability of a molecule (Banwell, 1983). The efficiency of Raman scattering depends on the change in polarizability of the molecule when it vibrates. Molecular vibrations cause the position of the atoms held together by the chemical bond to change, which may lead to a change in the polarizability of the molecular electron cloud. As the strength of the scattered light is proportional to the electron cloud displacement, molecular vibrations can cause a periodic change in the intensity of the scattered light. The position of the displacement is equal to the energy of the chemical bond that has been polarized. The energy can be added to or subtracted from the original frequency of the light. If we define (virtual) normal coordinates  $Q$  to fix the position of a molecule in the three dimensional space,  $Q$  will change during molecular vibration and will thus be given by the derivative  $\partial Q$ . In Raman spectroscopy, the change in polarizability,  $\alpha$ , during a molecular vibration is measured. The change is described by the polarizability derivative,  $\partial\alpha/\partial Q$ . This leads to the so-called *selection rule* for a Raman active vibration, namely that there must be a change in polarizability during the vibration, formally stated as  $\frac{\partial\alpha}{\partial Q} \neq 0$ . Scattering intensity is proportional to the square of the induced dipole moment, that is, to the square of the polarizability derivative  $\left(\frac{\partial\alpha}{\partial Q}\right)^2$ .

If a molecular vibration does not change the molecular polarizability, then the derivative will be zero, and the intensity of the Raman band will be absent. The vibrations of highly polar bonds are usually weak, since an electric field cannot induce a large change in the dipole moment, and stretching and bending of the bond does not change this. Strong Raman scatters are typically double bonds, such as carbon-carbon bonds, where the electron cloud can easily be distorted in an external electric field.

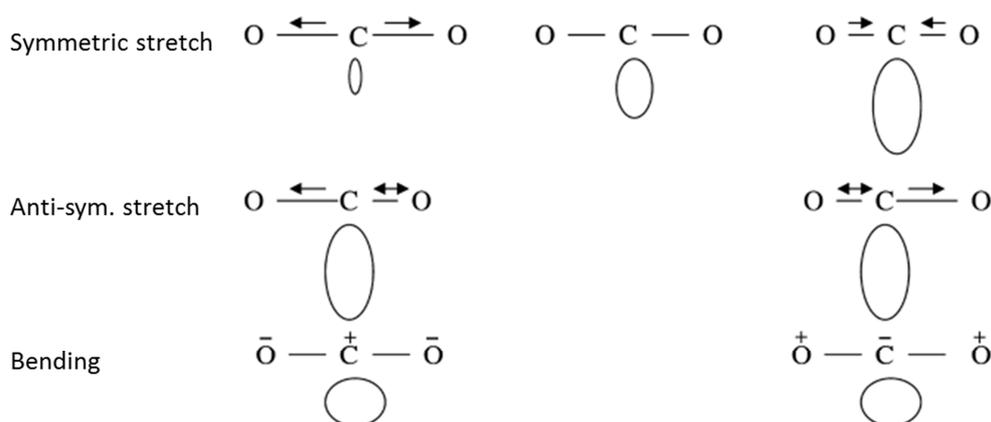
Stretching and bending of the bond will change the distribution of the electron density substantially and cause a large change in the induced dipole moment. Last but not least, the observed strength of the Raman band is proportional to the concentration of the species (further details in section 5.4), as well as to the intensity of the excitation laser (further described in section 3.2).

### 2.3 Polarizability tensor

The incident (laser) light causing the Raman effect is polarized in one plane, but the effect on the electron cloud is in all three directions x, y, and z. Therefore, three dipole moments are considered. The polarization angle effect upon the polarizability components is usually labeled,  $\alpha_{xx}$ , where the first subscript refers to the direction of the polarizability of the molecule, and the second x refers to the polarization of the incident light. Thus,  $P_x = \alpha_{xx}E_x + \alpha_{xy}E_y + \alpha_{xz}E_z$ . Similar expressions can be formulated for  $P_y$  and  $P_z$  and thus the overall polarizability of the molecule is a tensor as shown in Equation 4 (McCreery, 2000):

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \quad (4)$$

Often, the polarizability in various directions is represented by ellipsoids as shown for the symmetric stretching of a CO<sub>2</sub> molecule in Figure 5 (Pelletier, 1999a).



**Figure 5.** Molecular vibrations of the carbon dioxide molecule. The ellipsoids represent the polarizability,  $\alpha$ , of the molecule. The volume of the ellipsoids is proportional to  $1/\sqrt{\alpha}$  and the figure shows how the three vibrational modes affect the polarizability. The symmetric stretching is Raman active because of the change in the polarizability ellipsoid for the two extreme displacements. For the anti-symmetric stretching and the bending, the shapes of the ellipsoids for two extreme displacements are the same, thus, these two vibrations are not Raman active (modified from Wong, 1994).

The figure shows an example of a molecule (CO<sub>2</sub>) with a center of symmetry. A molecule has a center of symmetry when, for any atom in the molecule, an identical

atom exist diametrically opposite this center in an equal distance from the center. There may or may not be an atom at the center. For molecules with a center of symmetry the *rule of mutual exclusion* exists, which states that no normal vibrational modes can be both IR and Raman active.

## 2.4 Molecular vibrations

Raman spectroscopy is used to obtain information about structure and properties of molecules from their vibrational transitions. Even at the absolute zero all molecules are in a constant state of vibration. In vibrational spectroscopy this molecular property is utilized by studying the interaction between light and these molecular vibrations. The energy of most molecular vibrations corresponds to that of the infrared region of the electromagnetic spectrum, which falls between  $4000\text{ cm}^{-1}$  and  $200\text{ cm}^{-1}$ . Vibrational energy for a molecular bond mainly depends on the mass of the atoms involved and the strength of the covalent bond. In the classical description, molecular vibrations are modeled with a ball-and-spring concept, where the atoms in a molecule are imagined as balls, and the chemical bond strengths between them as a spring. If the masses of two atoms in a bond are  $m_1$  and  $m_2$ , respectively, and the restoring force of the spring,  $F$ , is proportional to the displacement  $x$  of the atoms from their equilibrium position, then:

$$F = kx \quad (5)$$

where  $k$  is the force constant of the spring, in  $\text{N m}^{-1}$ , which is a measure of the strength of the bond between the two atoms. Equation 5 is Hooke's law, and the resulting motion is a simple harmonic. Based on these assumptions, the ball-and-spring model is a harmonic oscillator. The vibrational frequency of the harmonic oscillator is given by:

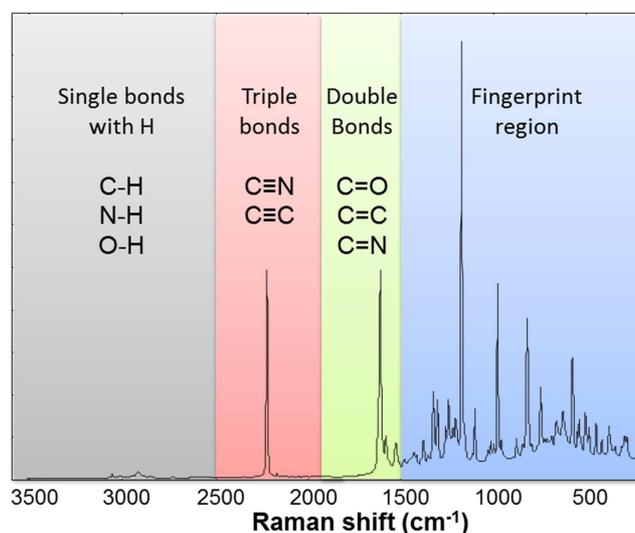
$$\nu = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}} \quad (6)$$

where  $c$  is the speed of light and  $\mu$  is the so-called reduced mass:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (7)$$

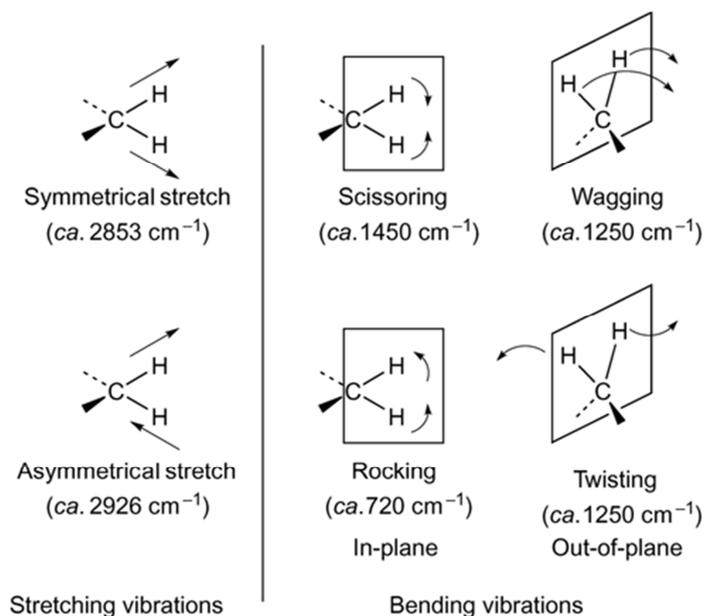
The stronger the bonding between two atoms is, the higher the force constant and the higher the vibrational frequency. For triple bonds the bond strength is higher than for double bonds resulting in a higher vibrational frequency for triple bonds than double bonds. Conversely, the heavier the vibrating masses, the lower the vibrational frequency. Thus, the vibrational frequency for bonds involving hydrogen has the highest frequency because hydrogen has a small mass. The concept of bond vibration is not very exact. For each vibration, all atoms in a molecule vibrate, but they can be more or less localized. The X-H vibrations are very localized, so are the triple bonds,

but for double bonds in for example peptides, they become less localized and so forth. An overview of how the Raman spectrum can be divided into regions is shown in Figure 6.



**Figure 6.** Fundamental regions of the Raman spectrum. The underlying spectrum is Dapivirine, which is an antiretroviral drug used in treatment and prevention of HIV infection.

The main types of vibrations are so-called stretching and bending modes. The vibrations of a CH<sub>2</sub> group in a poly-methylene chain are shown in Figure 7. Stretching vibrations can be symmetrical and anti-symmetrical with slightly different frequencies. The bending vibration can occur in different patterns recognized as scissoring, rocking, wagging, and twisting, all occurring at different frequencies (Miller, 2001; Wartewig, 2003).



**Figure 7.** The six normal modes of molecular vibrations shown for a methylene group in a poly-methylene chain (modified from Miller, 2001).

A complex molecule has a large number of vibrational modes ( $3N-6$ , where  $N$  is the number of atoms). Some of these molecular vibrations are localized to individual bonds or functional groups, while others are more delocalized and must be considered as vibrations of the entire molecule. When the molecule is irradiated with light, the vibrating bond will only absorb energy if the frequencies of the light and the vibration are compatible.

A group frequency is a vibrational frequency that is a localized vibration characteristic for a particular chemical functional group. Some group frequencies fall within a restricted range, regardless of the compound in which the group is found, while other group frequencies are highly affected by the matrix of which the group is a part. Functional groups of special interest in Raman spectroscopy are C-C, C=C, C≡N, and aromatic groups, mostly originating from the skeleton of molecules.

## 2.5 Raman spectroscopy compared to other spectroscopic techniques

When choosing the most appropriate analytical method for a given problem, a number of practical and theoretical properties must be considered. In order to provide a platform for choosing the most suitable methodology, this section outlines a number of characteristics of commonly used spectroscopic techniques in process analytical applications. The emphasis is foremost on describing these properties with regards to their applicability for rapid quality control and/or process monitoring. Table 2 provides an overview of properties of five frequently used spectroscopic techniques.

**Table 2.** Comparison of properties of five spectroscopic methods. Note that it is the multivariate use which is considered here.

	<b>NIR absorption</b>	<b>Raman scattering</b>	<b>IR absorption</b>	<b>UV-VIS absorption</b>	<b>Fluorescence emission</b>
Resolution	Low	Medium	Medium	Low	Medium
Sensitivity	Low	Medium	Medium	High	High
Selectivity	Medium	High	High	Low	Very high
'Main Interferences'	Water	Fluorescence	Water	Scattering	Quenching
Sample preparation	None	None	None	None	None
Sampling flexibility	High	High	Low	Medium	Medium
Typical analysis time	Seconds	Seconds	Minutes	Seconds	Minutes
Prize of a robust instrument	++	++++	++	+	+++
Molecular phenomena	Combination- and overtones of molecular vibrations	Fundamental molecular vibrations	Fundamental molecular vibrations	Valence electron transitions	Electron transitions
Other characteristics	-Nearly all molecules absorb -Quartz transparent	-Nearly all molecules scatter -Quartz transparent	-Nearly all molecules absorb -Quartz absorb	-Few molecules absorb -Quartz transparent	-Few all molecules fluoresce -Quartz transparent

The spectroscopic methods listed in Table 2 are based on different regions of the electromagnetic spectrum, on different physical principles and have different sensing capabilities. The measurements properties can be categorized in terms of three main aspects:

*1. Information content.* The method must be sensitive enough to measure the analyte or quality parameter of interest and this depends on the molecular properties of the analyte. In non-resonance Raman spectroscopy the change in polarizability during a molecular vibration is measured. Thus, the Raman signal is large for vibrations that do not change the symmetry of a molecule, vibrations of C-C double and triple bonds, and for molecules containing S, Cl or Br. IR spectroscopy is based on absorbance of IR radiation by fundamental molecular vibrations with a permanent dipole moment or an induced dipole moment during vibration. The C=O stretch gives particular large absorbance in IR. NIR is especially sensitive to strong anharmonic vibrations which are typical for X-H bond vibrations where X is a heavier atom. This NIR selection criterion gives preference to C-H, N-H, and O-H vibrations which are abundant in e.g. biological systems. Absorption of UV-VIS radiation causes valence electron transitions in molecules enabling measurement of analyte down to very low concentrations (ppm range). For fluorophore molecules the analytes can be measured to even lower concentrations (ppb).

The spectral resolution can have an influence on the ability to distinguish between analytes. NIR, UV-VIS, and fluorescence typically have low spectral resolution whereas IR and Raman have high spectral resolution. Chemometrics can, however, often be used to overcome problems of low spectral resolution by resolving the chemical information mathematically.

The amount of interference is also important, as it may cover/interfere with the signal of the analyte of interest, and thereby give a biased result. For Raman spectroscopy the major interference is fluorescence, whereas for NIR and especially IR water can interfere and cover the signal of interest.

The selectivity plays a major role in the ease of calibration. As shown in this thesis, quantification by Raman spectroscopy can be achieved with a small to modest set of reference values by selection of analyte specific spectral bands or regions; the same can be anticipated for IR and fluorescence spectroscopy applications (assuming the analyte of interest is a fluorophore for the latter). This is not the case for NIR or UV-VIS where the low selectivity often translates into larger investments for quantification. Large calibration sets, incorporating natural variation such as different/uncorrelated levels of chemical species other than the analyte of interest, are required to extract only the analyte-relevant information during chemometric modelling.

2. *Ease of sampling.* The choice of spectroscopic method also depends on the requirements regarding analysis time, sample preparation and whether measurements must be done e.g. in-line in a process or off-line in a laboratory. NIR has been the primary choice in many industrial applications as it is fast, requires no sample preparation, and can measure remotely and non-invasively. Raman possesses many of the same properties and can sometimes measure analytes which NIR cannot determine. However, Raman probes must be shielded from external lights (i.e. room light) during measurement. For IR the possibilities of using optical fibers are very restricted, thereby decreasing the sampling flexibility (i.e. making on-line and in-line installations limited). For both UV-VIS and fluorescence optical fiber probes are available, making *in-situ* process sampling possible, and at the same time allowing the spectrometer to be positioned safely away from the process.

3. *Prize and stability.* From a purchasing cost perspective the prize of robust spectroscopic instruments ranges from \$50k to \$200k (Bakeev, 2010) where Raman installations are found on the high-end of the scale. For an in-line application the installation cost are between two to five times the purchasing costs (utility lines for nitrogen purchasing, IT infrastructure, etc.) and this ratio can be much higher for Raman installation due to laser-related safety precautions. In long-term stability and maintenance cost the life-span of the laser in Raman spectroscopy is also a disadvantage compared to light sources used in e.g. NIR or IR. However, these cost-disadvantages might be offset by the before mentioned ease of calibration (smaller data sets and no advanced data skills required by the user) and the ‘assurance’ or ‘confidence’ Raman measurement can bring, e.g. the clearly recognizable spectral band(s) uniquely associate with the analyte of interest e.g. the active pharmaceutical ingredient as exemplified in Paper III.



## 3. RAMAN INSTRUMENTATION

Relatively recent, there has been a large development in Raman instrumentation, which has expanded the possible instrument configurations. A new user considering a Raman application must decide which of many available configurations is most suited for either the specific problem, the sample type, or ideally a combination of both (McCreery, 2000). This chapter will describe commonly used components in commercial Raman instruments and their influence on instrumental design and performance. Since the majority of the research in the present thesis has been carried out using a dispersive instrument, the main focus will be on this instrument design and non-dispersive (Fourier transform, FT) instruments will only be touched upon briefly.

### 3.1 Overview of instrument components and the light path

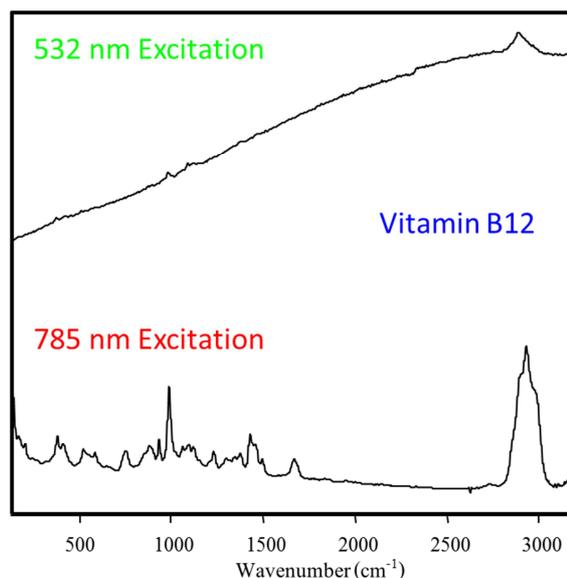
All Raman instruments consist of a laser as a monochromatic light source, an interface between sample and instrument, a filter to remove Rayleigh scatter, a spectrograph to separate the Raman scattered light by wavelength, a detector, and a communication system to report analysis results (Bakeev, 2010).

The light path in dispersive fiber optical instruments can be described as follows: monochromatic light from the laser is coupled into an optical fiber bundle. The fiber bundle carries the laser excitation to the probe. Silica Raman scattering and fluorescence generated in the collection fiber is filtered out at the probe head, and the laser light is focused onto the sample. Backscatter from the sample is filtered to remove the light at the laser wavelength (Rayleigh line), collected, focused, and sent to the base unit of the Raman instrument through another fiber bundle. In the base unit, a holographic notch filter removes any residual laser (Rayleigh) light and transmits the Raman scattered light to a holographic transmission grating. The grating disperses the transmitted light and projects it onto a charge-coupled device (CCD) camera via lenses. Subsequently, the accumulated charge on the pixels of the CCD camera is read out and converted into a Raman spectrum (Bakeev, 2010; Swierenga *et al.*, 1999).

### 3.2 Lasers

A highly monochromatic light source is essential for Raman measurements. The light source should provide radiation with an appropriate wavelength and sufficient power. Due to properties such as brightness and high power density, lasers are almost exclusively used for excitation. The most commonly used lasers are neodymium-doped yttrium aluminum garnet (Nd:Yag) lasers with possible excitation wavelengths of 1064 nm and 532 nm, and diode lasers which can excite at 630 or 785 nm.

The choice of laser is critical to the success of Raman for a given application and has major effects on both the observed spectrum and the instrument design. When selecting the right laser wavelength it often involves a trade-off between three factors: Raman cross section (defined as the ratio between the power of the scattered radiation and the intensity of the incident beam), detector sensitivity/signal, and background scattering from the sample. Usually, shorter wavelengths can be detected with higher quantum efficiency (% photons detected) and less noise resulting in an improved sensitivity. However, shorter wavelengths are more likely to excite fluorescence since more electronic transitions occur in the ultraviolet (UV) and visible region than in the infrared. So, if a fluorescent sample is to be measured, the applied laser wavelength should be as long as permitted by the sensitivity requirements. Finding the maximum signal-to-noise-ratio (SNR) is therefore a compromise between higher sensitivity at shorter wavelengths and lower fluorescence background at longer wavelengths (McCreery *et al.*, 2000). In Figure 8, vitamin B12 was measured with two different lasers, a green (532 nm) and a red (785 nm) one. The figure illustrates how an increase in the excitation wavelength lowers the fluorescence background of the Raman spectrum.



**Figure 8.** The effect of excitation laser wavelength. Vitamin B12 measured with green laser at 532 nm and red laser at 785 nm. At 532 nm, fluorescence is sufficiently intense that the Raman signal is unrecoverable from background, but at 785 nm the background is greatly reduced permitting observation of the Raman signal of vitamin B12.

The intensity of the Raman signal,  $I_{\text{Raman}}$ , is proportional to the fourth power of the inverse of the incident wavelength,  $\lambda$ , as shown in Equation 8.

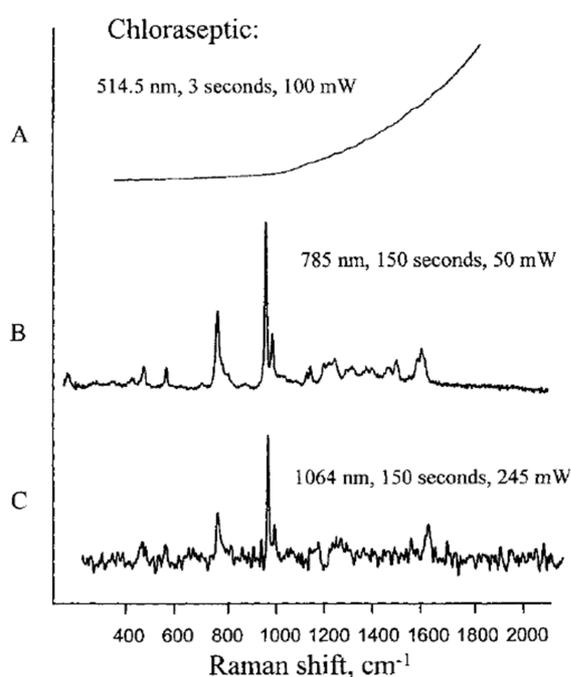
$$I_{\text{Raman}} = (1/\lambda)^4 \quad (8)$$

As illustrated in Table 3, there is a large intensity difference between using a UV (480 nm) laser and an NIR (1064 nm) source.

**Table 3.** Decrease in Raman intensity as a function of laser wavelength (Bakeev, 2010).

Wavelength $\lambda$ (nm)	480	532	633	785	840	1064
Intensity $(1/\lambda)^4$ (relative %)	100%	66.3%	33.1%	14.0%	10.7%	4.1%

Using a light source with a short wavelength, the number of Raman photons available for detection is much higher than if a longer wavelength laser is applied. It is therefore desirable to use the shortest laser wavelength possible to maximize the signal. When the laser wavelength increases, the SNR improves because the background typically decreases much more than the signal. As longer wavelengths are applied, the background continues to decrease. However, the SNR will eventually degrade again as the detector noise and dark noise become factors. The effect of detector noise is shown in Figure 9 for the common throat medicine Chloraseptic.

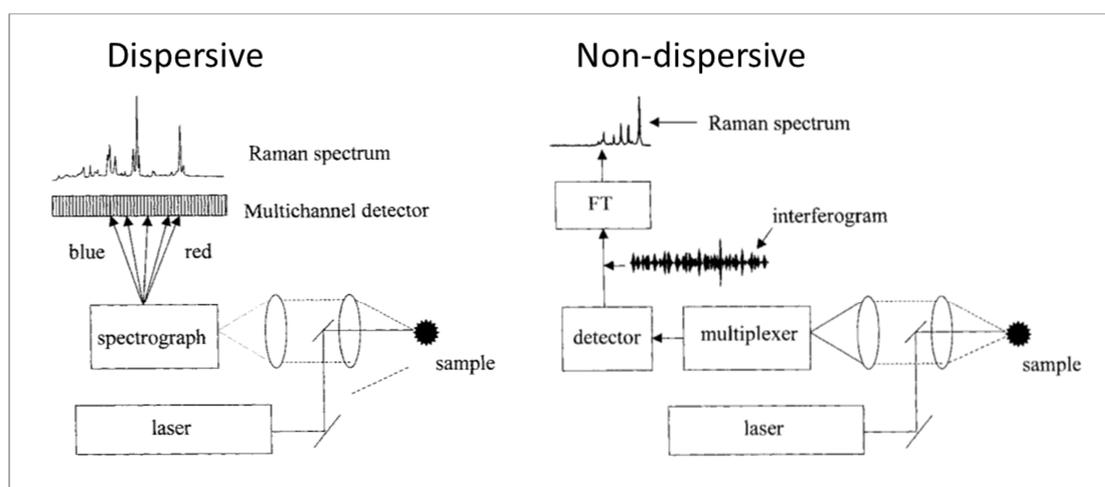


**Figure 9.** Raman spectra of the throat medicine Chloraseptic obtained with three different spectrometers. (A) 514.5 nm 3 sec. exposure, (B) 785 nm 150 sec. exposure, CCD detector, and (C) FT-Raman 150 sec. exposure (McCreery, 2000).

At 514.5 nm, the fluorescence from red dye obscures the Raman scattering. At 785 nm, the fluorescence is much reduced and the Raman scattering of phenol, glycerin, and saccharin can be observed. At 1064 nm, the fluorescence is still low, but the detector noise has become a significant factor. In this case, the 1064 nm laser dictates the choice of a noisier detector, since the CCD used at 785 nm is insensitive above about 1050 nm. This example illustrates an important rule of thumb: the laser wavelength should be as long as required to avoid fluorescence, but not so long that the signal strength and detector noise become problematic (McCreery, 2000).

### 3.4 Instrumental design

There are two fundamentally different ways a Raman spectrometer can be designed: dispersive or non-dispersive (Figure 10).



**Figure 10.** Overview of the operation of a dispersive multichannel Raman spectrometer and a non-dispersive FT-Raman spectrometer (McCreery, 2000).

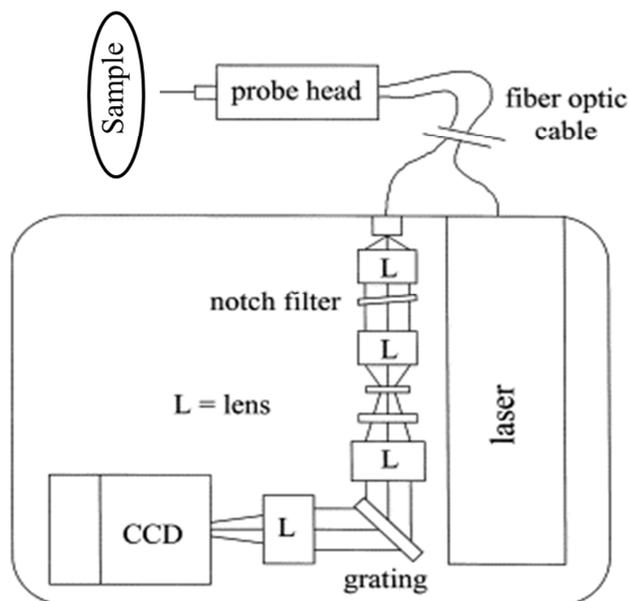
#### 3.4.1 Non-dispersive Raman

Non-dispersive Raman spectrometers are nearly always associated with FT modulation of the signal and involve no physical separation of the wavenumbers. Usually wavelengths are modulated so that each wavelength has a characteristic modulation frequency. The composite modulated signal is then monitored by a single detector and demodulated by a Fourier transform. Non-dispersive spectrometers are used in FT-Raman instruments with the excitation at 1064 nm.

As stated previously, CCD detectors are insensitive for laser wavelength excitation above about 1050 nm, which means that for non-dispersive instruments where a 1064 nm laser is used, a noisier detector has to be applied. The current detectors of choice for Raman measurements at wavelengths longer than 1000 nm are high purity p-type germanium or indium gallium arsenide (InGaAs) detectors. These detectors are noisier than CCDs, but tend to have high quantum efficiencies (Pelletier, 1999a). The quantum efficiency is a measure of the sensitivity of a detector, and it is defined as the probability that a photon striking the detector generates an electronically measurable signal, usually a photoelectron.

#### 3.4.2 Dispersive Raman

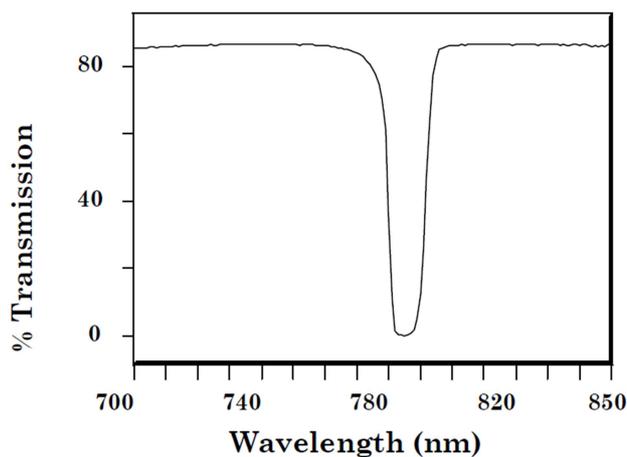
The dispersive principle is based on diffracting the scattered radiation by a grating. These types of spectrometers are used in systems equipped with laser excitation of wavelength from visible up to 1050 nm. An overview of a typical dispersive Raman instrument is shown in Figure 11.



**Figure 11.** An overview of the elements of a dispersive CCD Raman instrument (modified from Swierenga *et al.*, 1999).

### 3.5 Laser-line rejection

Raman instruments require a laser-line rejection device to filter out the reflected laser radiation and Rayleigh scatter from the much weaker Raman scattering. There are two types of filters used for Raman spectroscopy: the edge filter and the notch filter. The edge filter usually blocks all wavelengths above or below a certain wavelength, whereas only a smaller region of wavelengths is blocked by the notch filter. This is illustrated in Figure 12.



**Figure 12.** Transmission curve for a 785 nm Holographic SuperNotchPlus filter. At the laser wavelength 785 nm the transmission is less than 0.0001%, thus a very effective suppression of the Rayleigh light is provided (Kaiser Optical Systems).

Currently the most popular choice for laser light rejection is the holographic notch filter. The filter acts as a “wavelength-selective mirror”; a narrow band of wavelengths is reflected, while all other wavelengths are transmitted with high efficiency. As illustrated in Figure 12, the bandwidth of a notch filter can be very narrow and the reflectivity very high with a transmission at the center of the notch of less than 0.0001%. For Raman applications where low wavenumber bands are of interest this type of notch filter is ideal. Only drawbacks of the holographic filter are their sensitivity to water and temperatures above 80°C (Lewis & Edwards, 2001).

## 3.6 Spectrograph

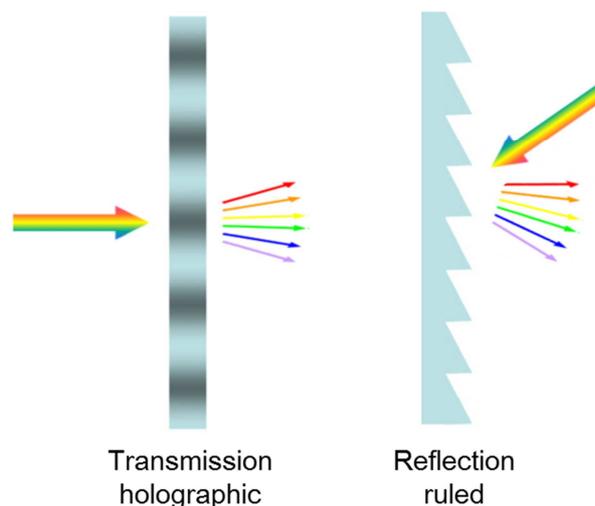
### 3.6.1 Design

The components in a dispersive spectrograph are a slit or input aperture, an input collimator, a grating, and an output collimator. The light enters the spectrograph through the slit or input aperture, and is collimated and directed to the grating by the input collimator. The grating disperses the light into its constituent wavelengths, and the output collimator images the light from the grating onto the detector.

The spectrograph is sometimes referred to as the heart of the Raman instrument, and in the development of Raman instrumentation much focus has been on spectrograph design and components. A thorough description of all possible spectrograph designs and available components will not be presented here. Instead, it has been selected to describe one of the newest spectrograph design concepts, *the axial transmissive spectrograph*. This type of spectrograph contains a volume holographic transmission grating in combination with lenses, in order to achieve almost aberration-free performance. The lenses are used as collimating and focusing elements, and they are placed on an axis with a holographic transmission grating in the collimated region between them. To get the highest diffraction efficiency, the grating is placed in an angle close to 45°. In the overview of the dispersive Raman instrument in Figure 11, an axial transmissive spectrograph design is shown in combination with a pre-filter stage (Lewis & Edwards, 2001).

### 3.6.2 Gratings

A grating is an optical component with a periodic structure, whose fundamental purpose is to bend different wavelengths of light from a common input path into different angular output paths. This is done through diffraction and dispersion. Diffraction means bending light through the use of a periodic phase change, and dispersion is defined as an angular separation of different wavelengths. There are two types of diffraction gratings: reflection and transmission. Figure 13 shows example of a ruled reflection grating, and holographic transmission grating.



**Figure 13.** Diffraction of light with two types of gratings. Left: a volume phase holographic transmission grating and right: a ruled reflection grating (modified from Bertarelli *et al.*, 2011).

The ruled gratings are created by etching a large number of parallel grooves onto the surface of a substrate, followed by a coating with a highly reflective material. Holographic gratings, on the other hand, are created by interfering two UV beams to create a sinusoidal index of refraction variation in a piece of optical glass. This process results in a much more uniform spectral response, but a much lower overall efficiency. Ruled gratings are simpler and cheaper to produce, but exhibit much more stray light. Furthermore, the holographic grating can function as both the dispersive element and the focusing optic at the same time. The wavelength range and the optical resolution of a spectrometer are determined by the grating.

### 3.6.3 Detector

The detector of choice for most Raman measurements at wavelengths below 1050 nm are charge coupled device (CCD). A CCD detector is made of silicon (or other photosensitive semiconductors) and registers the Raman signal via electron-hole pairs produced by photons of sufficient energy. The CCD camera consists of two-dimensional arrays of pixels that each can be considered as an independent detector. The horizontal pixels are calibrated to correspond to the wavenumber axis, and the vertical pixels measure the intensity of the Raman signal; i.e. the image on the CCD camera is an electronic picture of the Raman signal, which is then converted into a spectrum. For light to be detected by a CCD detector, the following must occur: 1) light must reach the photosensitive silicon, 2) light must be absorbed by the photosensitive silicon, 3) the absorbed photon must produce a photoelectron, and 4) the photoelectron must be transported to the output amplifier. The most significant source of noise in the CCD is shot noise (further explained in section 4.3.1). The sensitivity of a detector is measured by the quantum efficiency, which is defined as the probability that a photon striking the detector generates an electronically measurable signal, usually a photoelectron. In most Raman instruments, the CCD is

cooled well below room temperature (e.g. stabilized to  $-40^{\circ}\text{C}$  using multiple-stage Peltier elements). This greatly improves the sensitivity of the detector system, and makes it insensitive to temperature fluctuations (Sasic & Ozaki, 2007; Pelletier, 1999a).

### 3.7 Sampling interfaces (probes)

The optical interface in the Raman instrument illuminates the sample with laser light, collects the Raman scatter from the sample material, and directs it to the spectrograph. This can be done in two ways by direct imaging or by optical fibers. Direct imaging is mainly used in non-dispersive Raman instruments, whereas optic fibers are quite commonly used in dispersive instruments, because mid-IR light is difficult to transmit through most optical fibers (McCreery, 2000, Lewis & Edwards, 2001). Due to remote sensing and flexibility, the fiber-coupled sampling probes have been preferred in the recent years, thus the rest of this section will focus on fiber optical probes.

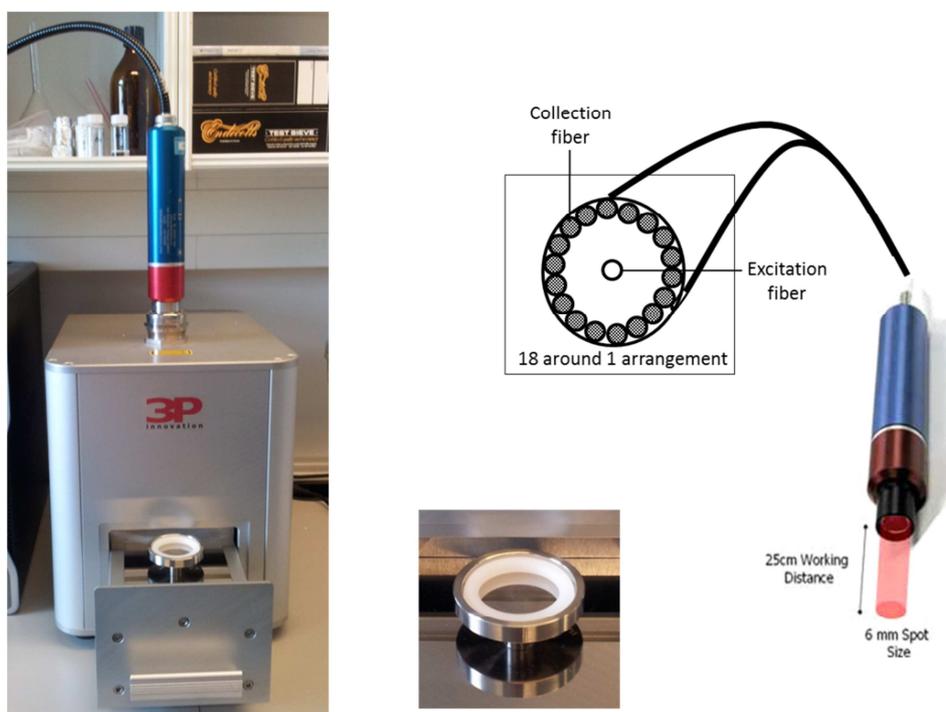
For optical fiber probes there is a possible working distance between the measurement point and the instrument up to 100m. When the laser light travels through the fiber silica, Raman scatter is generated and will overpower the analyte signal if it is not removed; thus a notch filter (as described in section 3.5) is applied prior to collection, so the laser light is blocked from the return fiber. The size, number, and arrangement of the fibers differ between interfaces. Most Raman fiber optics probes use  $180^{\circ}$  backscatter geometry, which refers to the angle between the sample and the collection optics. The light is delivered to the sample by a single fiber, and the  $180^{\circ}$  scattered Raman signal is collected through another fiber or bundle of fibers. The use of transmission mode ( $0^{\circ}$  geometry) is common for many spectroscopic techniques e.g. NIR. However, it has not been widely used in Raman spectroscopy. In transmission Raman, the excitation and the collection fibers would be on opposite sides of the sample. This technique is beneficial when the bulk concentration of a sample is to be found, but not optimal when only part of the sample is to be inspected (Bakeev, 2010; Lewis & Edwards, 2001).

In the selection of the Raman probe, it is crucial to consider the nature of the sample being measured e.g. phase, density, and transparency/opacity. Furthermore, probe functionalities such as the confocality and spot size are also very important. The confocality is a measure of how localized/focused the signal collection is in relation to the sample. The confocality for some probes is designed with a long focus making it possible to measure a sample through a process window, and only obtain data from the sample within. The laser spot size is determining for the sampling volume and can range from a few  $\mu\text{m}$  up to 6 mm. Small volume sampling approaches often include optical microscopes, whereas probes ideal for larger volume samples include wide-area-illumination (WAI) and spatially offset Raman spectroscopy (SORS, Matousek *et*

*al.*, 2005) probes, and transmission set-ups. For fragile samples that are easily affected by temperature changes, it would be appropriate to use large spot sizes, where the relative energy densities are low. Probes with small spot sizes work very well for samples that are homogeneous on a size scale equal to or smaller than the volume probed. Whereas large spot size probes, providing a representative average of a large area, are suited for bulk samples. However, a major drawback with many WAI probes is that they are limited only to measure a frequency range from 1800-200  $\text{cm}^{-1}$ . To elucidate how different optical fiber probes can be applied, four different examples of experimental set-ups are illustrated. All examples use the same instrument, but three different probes: WAI probe, immersion probe, and imaging probe.

### 3.7.1 Wide-area illumination (WAI) probe

WAI probes use a bundle of fibers for illumination and collection in a so-called  $N$ -around-1 configuration, where a single excitation fiber is surrounded by  $N$  collection fibers (Figure 14, right). Using this probe, it is possible to measure with a spot size diameter of 3-6 mm. The signal from the collection fibers is added to give one spectrum. In Papers II and III the WAI was applied. The wide spot size allows a much greater portion of a sample to be interrogated in a single measurement than with a small sample volume probe. The WAI approach reduces the need for multiple measurement points and can therefore reduce the analysis time.

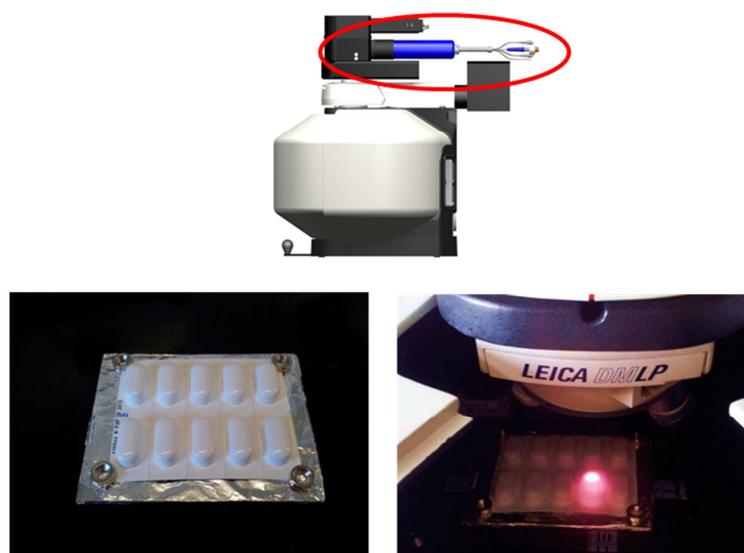


**Figure 14.** On the left: experimental set-up for measuring API in vaginal rings with wide-area illumination (WAI) probe and a ring spinner. On the right: WAI probe with a cross section view of the optical fiber showing a so-called  $N$ -around-1 configuration, in this case an 18-round-1.

The left side of Figure 14 shows an experimental set-up using a WAI probe. In this set-up, the sample (in this case a vaginal ring) was placed in a custom-built ring

spinner (3P Innovation Limited, UK), and a WAI probe (Kaiser optical Pharmaceutical-Area-Testing probe) was placed on top of the ring spinner. Measurements were carried out using an exposure time of 30 sec while spinning the ring at 17 rpm. The WAI probe was chosen because of its large spot size diameter (6 mm) allowing a representative sampling of almost the entire ring. The measurements were used to successfully predict the active ingredient in vaginal rings (Paper III).

In some situations, the WAI probe can be mounted on a measurement chamber as shown in Figure 15. The motorized stage makes sampling easier, as the samples can be moved automatically. In this example, the WAI probe was used for measuring different locations on an object over relatively wide distances, in this case different mixtures inside the same blister package. The experimental set-up is shown in Figure 15.



**Figure 15.** Experimental set-up measuring through a sealed blister package by WAI probe. Top: the WAI probe connected to the Raman microprobe. Bottom left: PVC blister package containing sample mixtures. Bottom right: the laser spot illuminating the blister package.

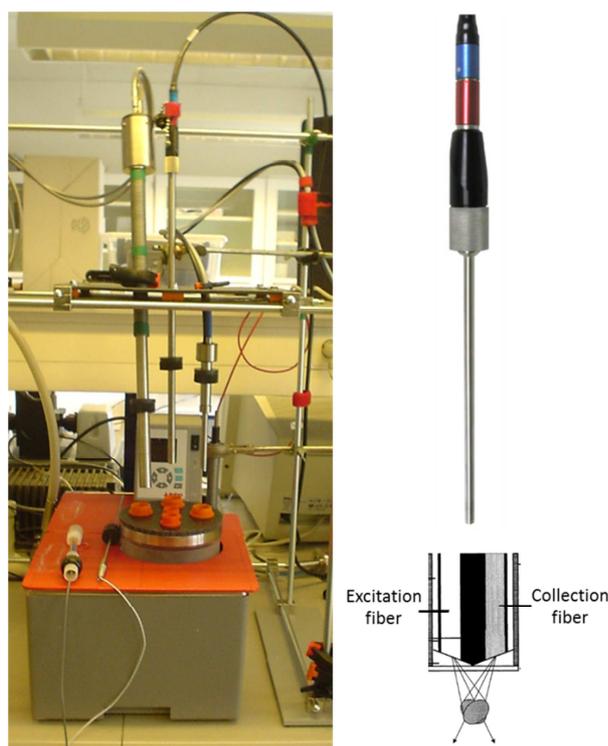
The blister package containing sample mixtures was placed on the motorized stage, and the use of a 6 mm diameter spot size allowed a very large portion of the sample mixtures to be probed. As earlier described, the use of a large sampling area will even out occurrence of sample heterogeneity (Paper II).

### ***3.7.2 Immersion probe***

Immersion probes are very robust devices that can be mounted directly into a process stream, reaction vessel, etc. It uses only a single fiber for excitation and another for collection. It is equipped with a filtered probe head, which removes scattered light before and the laser line after it strikes the sample, which reduces the background signal. The probe can be fitted with either an optic providing a long working distance

(fixed focus 3 mm from the tip of the optics), or with a short working distance (fixed focus 0 mm from the tip of the optics).

A Raman immersion probe with short focus length was applied to monitor a fermentation process, in this case lactic acid bacteria fermentation in milk. The experimental set-up with an immersion probe and a small volume fermenter is shown in Figure 16.



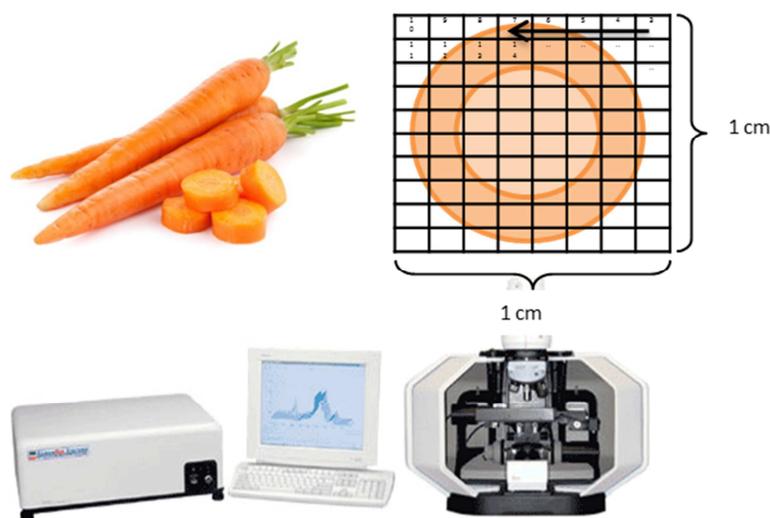
**Figure 16.** Raman immersion probe. On the left: experimental set-up using a Raman immersion probe to measure a fermentation. On the right: the fiber configuration providing focus at the tip of the optics (modified from Lewis & Edwards, 2001).

The probe can be immersed directly into the fermenter, and continues measurements can be carried out during the entire fermentation. The fermenter should be fully covered to avoid spectral interference from the room light. By means of this set-up, fermentations may be monitored in real time. However, if color or phase changes occur during the fermentation, the Raman measurements may be highly affected. In addition, sampling issues such as slurry homogeneity should be carefully considered in order to reduce pre-analytical biased results.

### ***3.7.3 Optical microscope (mapping of cross-sections)***

Alternatively, probes can be used in conjunction with a Raman microscope, where the laser can be launched into the optical path of the microscope. In this set-up, measurements of high spatial resolution can be obtained. Furthermore, with this set-up Raman images can be generated through rapid scanning of a sample stage. With a

Raman microscope it is possible to look at the spatial distribution of a sample, in this case a cross section of a carrot. The cross section was placed on a motorized stage, and a mapping of an area 1x1 cm was recorded (Figure 17).



**Figure 17.** Raman microprobe used for mapping of cross sections of carrots.

For samples where it is interesting to know the distribution of multiple compounds over a small area, this set-up is ideal (sometimes it may be enough to do multiple point scans or a line scan instead of a full mapping). The microscope can be equipped with objectives to provide appropriate magnification of the sample.

### 3.8 Acquisition parameters

When the right sample interface is selected for a measurement, there are different parameters that can be tuned or selected prior to acquisition. In this section, some of these parameters will be briefly described.

Exposure: The exposure is the time where the spectrally dispersed light emerging from the spectrometer is allowed to strike the pixels of the CCD array and accumulate charge. In most dispersive Raman systems, a mechanical shutter is used. When not exposing the shutter prevent light from striking the CCD detector. Between exposures the detector may be set to do a cleaning, where any charge that is accumulated in the pixels is flushed from the CCD. The mechanical shutter helps reducing the amount of charge that must be flushed, making sure that CCD is empty before initiating an exposure.

Accumulation (co-addition): Signal accumulation (co-addition of successive spectra) is a technique for very effective reduction of random noise inherent in the Raman spectrum. It is a straightforward procedure, where the number of spectra,  $N$ , to be accumulated are first specified and collected successively. Each spectrum is added to an accumulating buffer on a pixel-by-pixel basis. When all  $N$  spectra are acquired and

summed to data, collection is completed. By this approach, random noise is, in general, reduced by the square root of the number of spectra that are summed (Lewis & Edwards, 2001). Spectra are typically (but not always) summed rather than averaged (as is done e.g. in NIR absorption measurement), because the Raman technique is a scatter counting principle, which will follow Poisson statistics.

Dark subtraction: Dark subtraction is another important data preprocessing step that Raman software may need to perform. Basically, a CCD detector will accumulate charge in its pixels even when there is no light striking the CCD. The amount of charge accumulated is proportional to the exposure time, but will vary considerably from pixel to pixel. The dark-charge rate will be affected by temperature and design of the CCD chip. A significant dark-charge contribution may induce false bands in the Raman spectrum, thus it has to be subtracted. Typically, it is carried out by collecting a dark spectrum (i.e. an exposure with the camera shutter closed) and saving that spectrum for subtraction from all subsequently spectra to be recorded under similar conditions. A new dark spectrum must be acquired whenever the acquisition parameters are changed.



## 4. The Raman spectrum

This chapter focuses on the interpretation of Raman spectra. It will include an approach to spectral assignment, a description of the major sources of noise and spectral interferences. Finally, the concept of band fitting is described.

### 4.1 Display

As described previously (section 2.1) the presentation of Raman spectra usually only contain the Stokes part with the anti-Stokes absent. Even though a graphical scale should be displayed as increasing, it is common use to display the wavenumber scale from the highest to the lowest value in the Raman spectrum. This originates from the fact that Raman spectroscopy measures a shift and not an absolute value. In addition, the high to low format makes comparisons to the IR spectrum easier as spectra can be overlaid and band positions compared. The intensity scale in the Raman spectrum varies from instrument to instrument, and will often be displayed in arbitrary units, thus direct quantification is not easy.

### 4.2 Spectral interpretation

Interpretation of a Raman spectrum can be carried out at different levels depending on how detailed information is desired. The spectrum of a molecule can be the subject of a full mathematical interpretation in which every band is carefully assigned e.g. by density functional theory (DFT) calculations (Fairchild *et al.*, 2009). In other cases the spectrum is just used to confirm whether or not the sample is or contains a particular molecule. For identifying an unknown compound most spectroscopic laboratories use computer aided spectrum interpretation e.g. spectral libraries containing a large range of pre-recorded reference spectra. A library search is carried out using a number of matrices to match a target spectrum with either the identical library spectrum or the most similar library spectrum (Ferraro *et al.*, 2003). In vibrational spectroscopy the spectral bands are typically assigned to molecular vibrations by means of assignment tables.

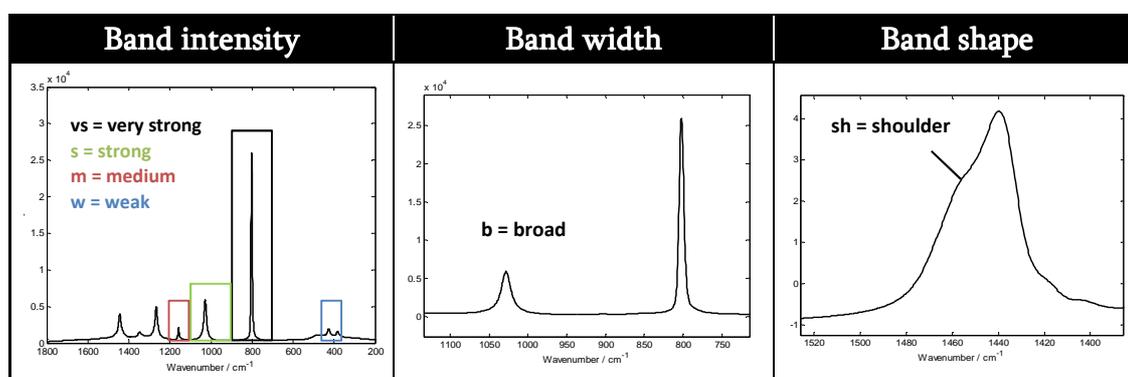


Figure 18. Visual illustrations of abbreviations commonly used in spectral interpretation.

In order to obtain a successful spectral assignment a systematic and knowledge-based approach is required. It involves some practice to learn. One important thing to keep in mind when interpreting a spectrum is that a spectral band may not only be assigned based on its band position, but other band parameters such as width and shape may also be of great use. A way to extract band parameters through band fitting will be explained in section 4.5. In spectral interpretation characterization of the spectral bands are normally included, and common terms and abbreviations to do this are listed in Figure 18.

As briefly described in section 2.4, the frequencies of vibrational bands can be divided into two broad categories: group frequencies and fingerprint frequencies. In the spectral interpretation the group frequencies usually give rise to sharp and strong intensity Raman bands above  $1500\text{ cm}^{-1}$ , while fingerprint bands are often weak and found below  $1500\text{ cm}^{-1}$  (as was shown in Figure 6). However, it may occur that a group frequency is present below  $1500\text{ cm}^{-1}$ . A powerful way of using the fingerprint region (below  $1500\text{ cm}^{-1}$ ) in the interpretation, is to raise a specific question based on the observations in the high-frequency region, and go back to this region to seek answers - e.g. there is a saturated C-H band (at  $2970\text{-}2850\text{ cm}^{-1}$ ). Is there evidence for C-CH<sub>3</sub> at  $1375\pm 5\text{ cm}^{-1}$ ? Box 1 gives an approach to assignment of a Raman spectrum. The approach is not comprehensive, and focuses mainly on the bands particular strong in Raman. It is intended to be used in combination with assignment tables as shown in Table 4 (high frequency range  $3600\text{-}1200\text{ cm}^{-1}$ ) and Table 5 (low frequency range  $1200\text{-}0\text{ cm}^{-1}$ ).

**Box 1.** Approach to interpretation of a Raman spectrum (based on notes from IR and Raman interpretation course by Griffiths and Hannah).

Start with high frequency end of the spectrum, and confirm with assignment tables

3600-3100  $\text{cm}^{-1}$

Are there any X-H bonds

- **3600  $\text{cm}^{-1}$**  OH stretching band is generally weak
- **3400  $\text{cm}^{-1}$**  NH stretching band is medium strong

3200-2700  $\text{cm}^{-1}$

Look for CH stretching bands

- **>3000  $\text{cm}^{-1}$**  olefinic (a hydrogen on an unsaturation C-atom) or aromatic
- **<3000  $\text{cm}^{-1}$**  aliphatic (a hydrogen on a saturated C-atom)

2700-1900  $\text{cm}^{-1}$

Are there triple bonds or cumulative double bonds (2300-1900  $\text{cm}^{-1}$ )

- **2250-2100  $\text{cm}^{-1}$**   $\text{C}\equiv\text{C}$  stretching bond
- **2590-2540  $\text{cm}^{-1}$**  SH stretching band
- **2250-2100  $\text{cm}^{-1}$**  SiH stretching band

1800-1550  $\text{cm}^{-1}$

Look for bands of double bonds

- **1750  $\text{cm}^{-1}$**  C=O carbonyl band is weak and broad
- **1680-1630  $\text{cm}^{-1}$**  C=C stretching band is strong and sharp

<1500  $\text{cm}^{-1}$

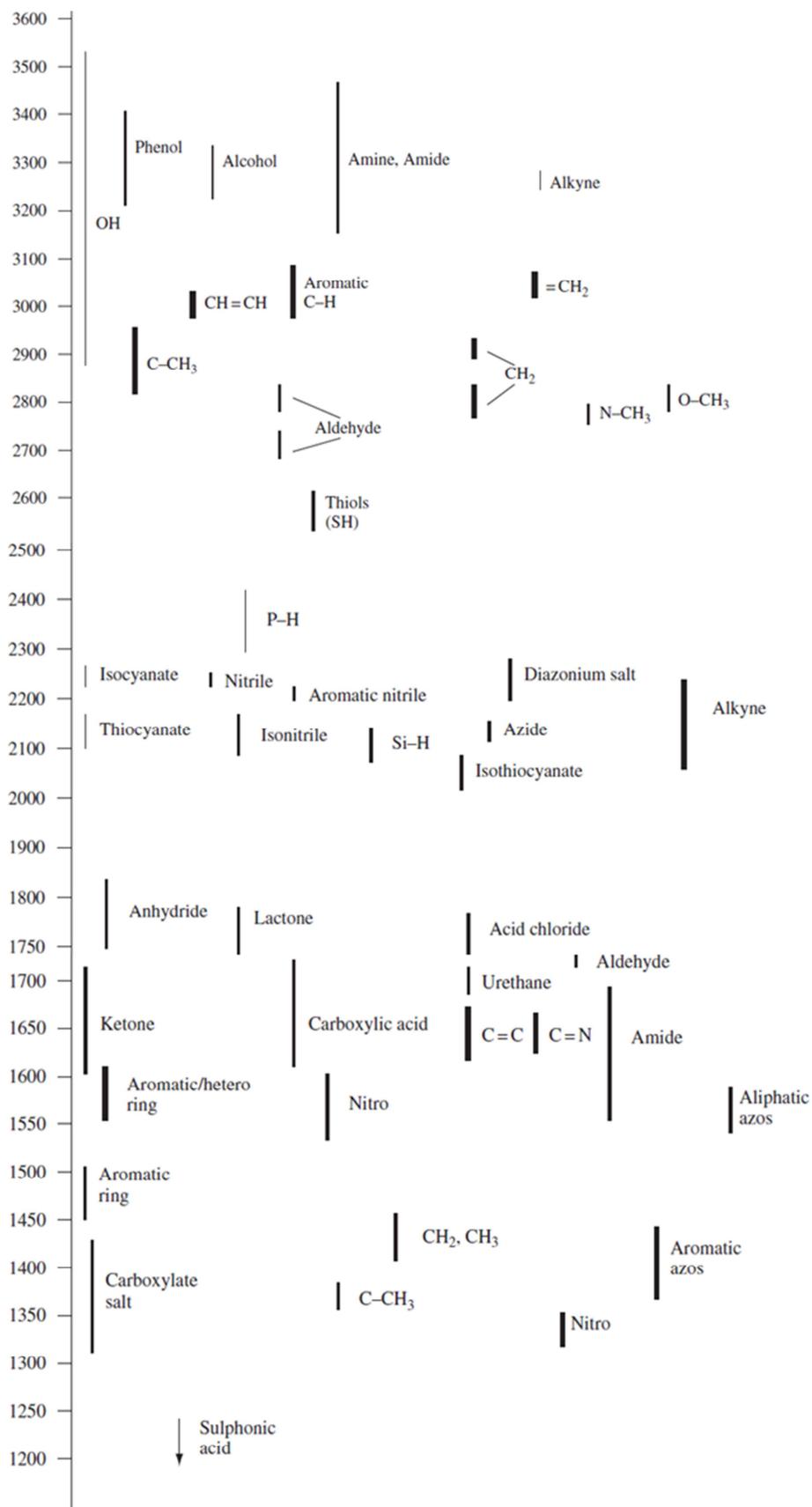
Mainly fingerprint frequencies but may contain some group frequency bands.

Care must be taken when assigning bands in this region.

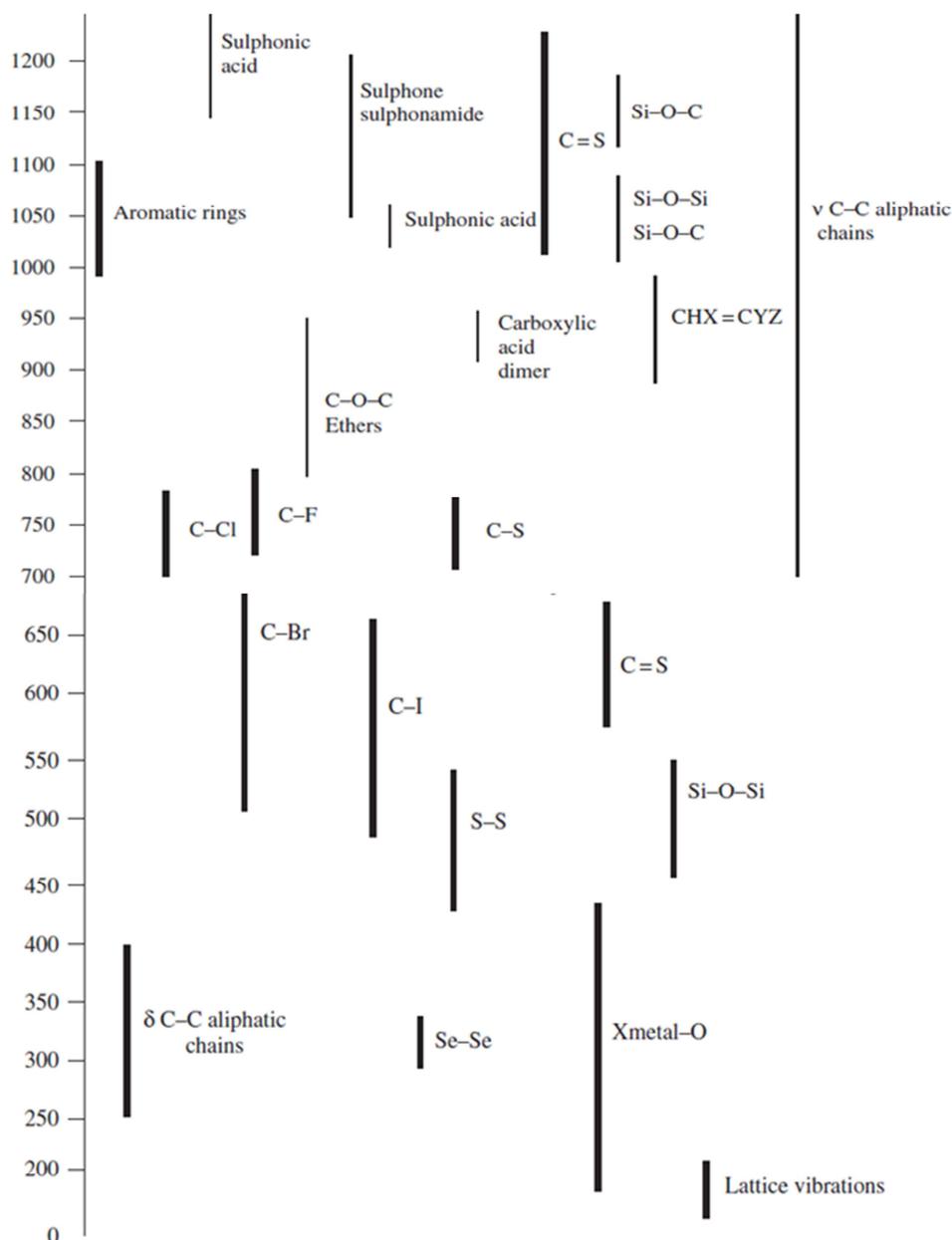
- **1450  $\text{cm}^{-1}$**   $\text{CH}_2$  scissoring band
- **1375  $\text{cm}^{-1}$**   $\text{CH}_3$  "umbrella" sym. stretching band
- **1300  $\text{cm}^{-1}$**  All-in-phase twisting band for 4 or more  $\text{CH}_2$  in a row
- **1000  $\text{cm}^{-1}$**  Phenyl mono 1,3- and 1,3,5-substituted is strong
- **<650  $\text{cm}^{-1}$**  usually inorganic groups, metal-organic groups or lattice vibrations

It is important to note that it might be impossible to assign all of the bands found in a spectrum, so instead of trying to assign every weak band the focus should be on the very intense bands. If the physical state of the analyzed sample is known, it should also be taken into account when interpreting the spectra, as it can strongly influence the spectra by shifting band positions and/or changing the band shape. This will be further described in section 4.4.

**Table 4.** Raman assignment table 3600-1200  $\text{cm}^{-1}$ . Thickness of lines indicates the band intensity (Smith & Dent, 2005)



**Table 5.** Raman assignment table 1200-0  $\text{cm}^{-1}$ . Thickness of lines indicates the band intensity (Smith & Dent, 2005).



### 4.3 Sources of noise

Raman spectra consist of two parts, the Raman signal and the noise part. The definition of noise in this context is the part of the spectra containing unwanted information not related to the chemical composition of the sample. The quality of a Raman spectrum is often determined by the signal-to-noise (SNR) ratio, and it is thus important to increase the signal and reduce the noise as much as possible. To be able to do so, it is critical to understand the noise, and how it affects the spectra. The main sources of noise will be described.

### 4.3.1 Shot noise

Shot noise is usually the dominant source of noise from the CCD. It is a direct consequence of the statistical nature of light. Shot noise arises from the random timing and energy of scattered photons. Two consecutive spectral measurements of the same sample will always show some variation in number of photons collected at any pixel. Shot noise level is equal to the square root of the number of detected photons. The signal-to-noise ratio of shot-noise-limited Raman spectra will increase linearly with the square root of accumulation time. In other words, for a CCD Raman more detected Raman photons (longer acquisition time) leads to a decreased relative standard deviation and increased SNR in the Raman spectra, this is illustrated in Table 6 (McCreery, 2000; Pelletier, 1999a).

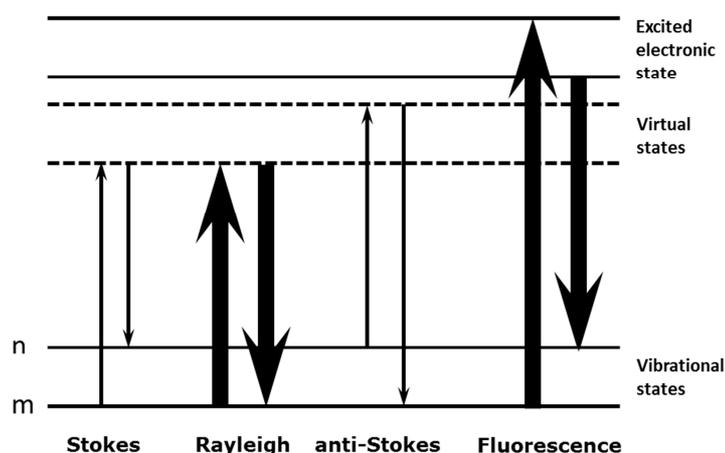
**Table 6.** The impact of shot noise on the relative standard deviation and uncertainty. The table illustrates how an increased accumulation time can reduce the uncertainty (modified from McCreery, 2000).

Shot noise = $\sqrt{\# \text{ detected photons}}$ = relative standard deviation			
Accum time:	# detected photons	STD	Uncert.
1 s	10000	100	$\pm 1 \%$
4 s	40000	200	$\pm 0.5 \%$
Uncertainty decreases with measurement time			

Another type of noise related to the detector is so-called dark noise, which is defined as the rate of spontaneous generated electrons in the detector. The amount of dark noise is strongly dependent on the temperature, and it can thus be reduced by sufficient detector cooling, and by dark subtraction (described in section 3.8).

### 4.3.2. Fluorescence

Laser-induced fluorescence is perhaps the most common background noise observed in the Raman spectra. The fluorescence process is illustrated by the energy state diagram in Figure 19. The sample is raised into the first excited electronic singlet state by absorption of a photon. The sample rapidly relaxes to the lowest vibrational level of the first excited singlet state and, after a period of time (1-10 nanoseconds), the sample relaxes back to the ground electronic state by emitting a photon of fluorescence.



**Figure 19.** Jablonski diagram compares Rayleigh, Raman Stokes and anti-Stokes with fluorescence. Fluorescence occurs when electrons are excited to another electronic state. The fluorescence signal is much larger than the Raman signal.

Fluorescence in a Raman spectrum can be challenging, because the fluorescence signal is often very high compared to the Raman signal, and therefore the Raman signal can be hidden under strong fluorescence. Fluorescence spectral features are usually much broader than Raman bands, and often look like a slowly changing baseline in the Raman spectrum.

Fluorescence problems are not always predictable and often occur in samples not typically described as fluorescent. That may be due to the fact that even very low concentrations of a fluorescent material in a bulk material can produce a fluorescence background that is much stronger than the Raman signal. The identities and concentrations of impurities in a sample are often unknown, and the magnitude of fluorescence must therefore be determined by experiment. Nevertheless, when considering analyzing a sample with Raman spectroscopy, it is useful to understand what conditions lead to fluorescence background, and how it can be minimized.

Common fluorescent impurities in biological systems include flavins, porphyrins and tryptophan degradation products. Colorful samples often contain much fluorescence; however, color is no guide as to whether a sample will fluoresce. Clear, water white crystals have been observed to cause fluorescence at all illuminating wavelengths. The following three approaches have been used to reduce the fluorescent background:

Longer excitation wavelength: by applying a longer excitation wavelength the fraction of common material that can absorb light and create an excited electronic state will decrease.

Photobleaching: illuminating the sample with laser light a period of time before data acquisition begins can sometimes reduce the fluorescence background, due to the

destruction of fluorescent material. This process is called photobleaching. However, this type of fluorescence reduction is often not convenient in on-line and in-line Raman applications, where the sample moves quickly before a significant fraction of the sample can be photobleached (Pelletier, 1999a).

Spectral preprocessing: by applying different mathematical methods the fluorescence background can be removed or reduced from a spectrum. The most commonly preprocessing techniques applied for Raman spectra will be described in section 5.2.

### ***4.3.3 Other background light***

There are many sources of externally generated noise that might affect the Raman spectrum. One external source of noise in the spectrum is cosmic rays. They can occur during exposure, and will affect the detector, by creating exceptionally high charges on few pixels. In the spectrum they will appear as very sharp and strong intensity bands (Lewis & Edwards, 2001). Cosmic ray filtering can usually be applied to eliminate the effect from cosmic rays. In the preprocessing section 5.2.2 this will be further described. Additionally, sunlight and a range of other light sources (room lighting) add similar sharp bands in the spectrum. Therefore, Raman measurements should be performed in darkness e.g. by using a sample chamber as shown in Figure 15 in the previous chapter. Finally, any Raman signal is masked by strong blackbody radiation. Blackbody radiation is defined as the radiation all objects emit because their temperature is above the absolute zero.

## **4.4 Temperature and pressure effects**

The Raman spectrum contains bonding information on chemical species, and will therefore be sensitive to anything that can change the chemical bonding such as temperature and pressure. The spectrum will change as a function of physical or chemical changes of the measured sample. Thus the Raman spectrum will be sensitive to all factors which may induce such sample changes. However, other - less obvious - factors such as changes in the refractive index can also affect the Raman spectrum.

The effects of temperature and pressure on the Raman spectra have been widely studied (Pelletier, 1999b; Wang Wei-Wei *et al.*, 2010). A change in temperature or pressure of a sample can change a Raman band height, width, and/or integrated area. Causes for these spectral changes can be divided into changes that are truly altering the sample composition or structure, and changes that do not. Both types of changes are listed in Table 7.

**Table 7.** Overview of temperature and pressure related changes.

Temperature and Pressure effects	
<b>Sample composition or structure</b> <ul style="list-style-type: none"><li>• chemical equilibrium</li><li>• density</li><li>• phase</li><li>• crystal lattice distortion</li></ul>	<b>Other changes</b> <ul style="list-style-type: none"><li>• refractive index</li><li>• vibrational excited state distribution</li><li>• vibrational anharmonicity</li><li>• vibrational lifetime changes</li></ul>

#### Density and phase changes

Increase in sample density is equal to concentrating the sample analytes in a smaller volume. A higher sample concentration results in a stronger Raman signal. Different phases of the same material often have very different Raman spectra. For example, the OH stretching of liquid water appears at  $3400\text{ cm}^{-1}$  and is about  $400\text{ cm}^{-1}$  wide. The same vibration in water vapor is found at  $3652\text{ cm}^{-1}$  and is less than  $1\text{ cm}^{-1}$  wide. However, in situations where it is desired to look at phase changes, this spectral influence may be seen as a property.

#### Changes in vibrational excited state distribution

A change in temperature will change the distribution of molecules between the vibrational energy states according to Boltzmann distribution (Equation 2, section 2.1). Increases in temperature will result in a higher concentration of molecules in the excited state. This will lead to an increase in the intensity of the Raman scatter, as the Raman cross section is larger for transitions that occur from higher energy states. The high and low frequency bands are affected unequally. The lower frequency vibrations will be more affected, because their vibrational energy is smaller compared to the thermal energy in the sample (Pelletier, 1999a).

In section 5.2 preprocessing methods to reduce or eliminate unwanted spectral changes due to temperature or pressure will be described.

### **4.5 Band fitting**

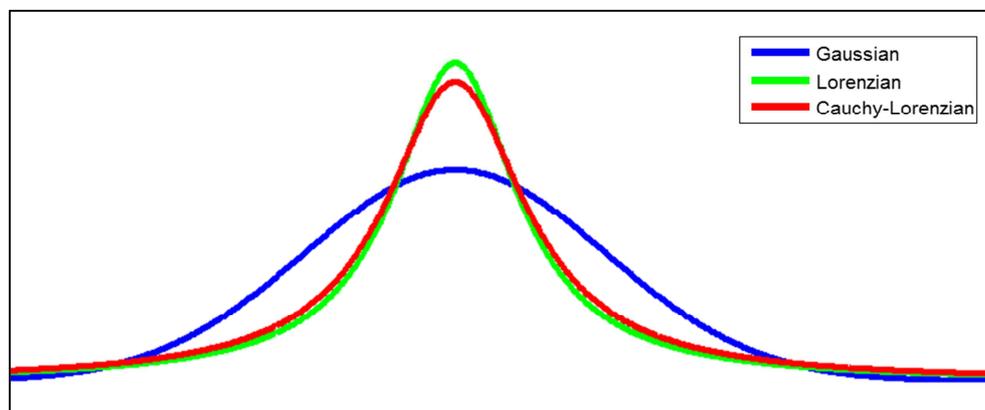
Physical and chemical relevant information can be found in the bands of the Raman spectra. Often only the band intensity is considered, however, the band width and the exact band position (frequency) may also contain important information regarding

sample composition. A way to extract parameters such as band width and position is band fitting. Band fitting is the process of least-squares fitting of mathematical functions to describe individual bands. Beside band parameter extraction band fitting can also be used for resolving overlapping bands. One of the critical parameter in using band fitting for Raman spectra is to choose the correct mathematical function that fits the band shape.

Most vibrating molecules interact with their surrounding molecules (the environment). The interactions between each molecule and its environment are slightly different, thus the vibrational frequencies are slightly different. The observed band shape is the sum of these individual molecules light scattering. At equilibrium, the population of vibrational states is given by the Boltzmann distribution (Equation 2), and at room temperature the majority is in the ground vibrational state. When illuminated by laser light some molecules are excited to a higher “virtual” state, but fall quickly back to the ground state, causing a change in the light scattered (the Raman shift). The rapid return of a molecule from excited to ground state is called the relaxation lifetime,  $\tau_a$ . All of the excited molecules vibrate together (coherently) at first, but motion and slight differences in vibrational frequencies randomizes this over time, breaking coherence. The time the molecules vibrate coherently is called the coherence lifetime,  $\tau_c$ . The coherence fades because components start to interfere, and cancel one another (so-called dephasing). The total lifetime  $\tau$  of a relaxation is a combination of the two factors  $\tau_a$  and  $\tau_c$ . Dependent on the molecular environment the magnitudes of the two factors change. In solids where the environment is not in motion, the excited molecule relaxes before incoherence becomes dominant,  $\tau_c \gg \tau_a$ . In gasses,  $\tau_c \ll \tau_a$  as collisions and rotations happen quickly, and the incoherence sets in rapidly, thus dephasing is dominating. In between these two extremes are liquids. In liquid interactions prevent extremely rapid motion, but the molecules are not locked in place. Therefore  $\tau_a$  and  $\tau_c$  are of similar magnitude (Adar, 2012). This is just one argument why band shapes in Raman spectra can differ greatly from application to application.

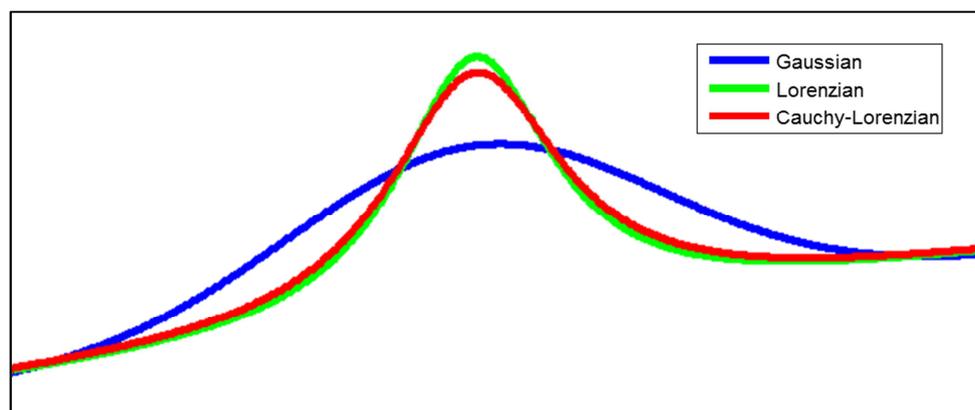
In order to describe the shape of Raman bands most accurately different mathematical functions have been proposed. As the shape of Raman bands change with the molecular environment, the mathematical function that describe the shape most appropriate will also change dependent on the molecular environment. This means that in order to perform optimal band fitting the sample type (solid, liquid, gas, or an intermediate/transition) has to be taken into account. For solids a Gaussian function fits well, as the band shape is a bell curve. For gasses the band is more sharp in the center and has longer wings, which fits well with a Lorentzian profile. The shape of a Raman band of a liquid has features of both Gaussian and Lorentzian character. Two functions fit this shape: a simple combination of Gaussian and Lorentzian (Cauchy-Lorentzian) or the more complex combination Voigt profile

(where the two characters are combined in a Fourier Transform integral). The Voigt profile allows the portions of Gaussian and Lorentzian to have different band widths (Di Rocco & Cruzado, 2012). In Figure 20, band profiles of Gaussian, Lorentzian, and Cauchy-Lorentzian are compared.



**Figure 20.** Examples of Gaussian, Lorentzian and Cauchy-Lorentzian profiles. The Raman bands of solid sample often have a gaussian shape, whereas the bands of gasses fits well with a Lorentzian profile. In between are the Cauchy-Lorentzian profile which often fits well to the Raman bands of liquids.

The Raman spectrum may contain signals due to fluorescence (as was explained in section 4.3.2). Removal of the fluorescence is often carried out using background removal routines (see section 5.2.1), but band fitting can also be used. Typically, fluorescence tends to exhibit considerable asymmetry, and therefore the Log-normal profile or a polynomial with a slope are often used to model it. Figure 21 shows the Gaussian, Lorentzian, and Cauchy-Lorentzian profiles with a slope.



**Figure 21.** Taking asymmetric fluorescence contributions into account in the band fitting of Raman bands. Gaussian, Lorentzian, and Cauchy-Lorentzian profiles with a polynomial slope.

There are many different purposes for applying band fitting: resolving overlapping bands, or for extraction of band parameters like band width and position. In Paper I Cauchy-Lorentzian band fitting was applied to Raman spectra of pork backfat to extract information on unsaturated fatty acids.



## 5. Raman data analysis

A Raman spectrum often contains a large amount of information that describes the chemical composition and the physical properties of the sample, but also unwanted non-sample information originating from noise and external light. In an exploratory study, it can be of interest to look at all this simultaneously, but often not all information is relevant for a given study. Therefore, the key in the analysis of spectral data, such as Raman, is how to find and extract the appropriate information.

Historically, the quantitative analysis of Raman spectra has been limited to the univariate linking of single Raman band intensities or Raman band ratios to some reference values. Only in recent years, the uses of multivariate analysis and chemometrics have been introduced.

This chapter will focus on the preprocessing methods appropriate for Raman spectroscopic data. Methods for noise reduction, background removal, and normalization are described. Furthermore, multivariate techniques for qualitative and quantitative analysis will be explained.

### 5.1 Univariate vs. multivariate

Univariate analyses use a single value for calculation of the property of interest - e.g. using a single intensity to calculate concentration. The observed intensity for a species is usually assumed to be a linear function of the concentration of that species:

$$I_R = \sigma L C I_0 k \quad (9)$$

where  $I_R$  is the observed Raman intensity,  $\sigma$  is the apparent Raman scattering efficiency (depending primarily on species, environment, and excitation wavelength),  $L$  is the interrogated volume,  $C$  is the species' concentration,  $I_0$  is the incident intensity, and  $k$  is the instrumental throughput (a measure of the light-gathering power of the instrument). When  $\sigma$ ,  $L$  and  $I_0$  are assumed to be constant, Equation 10 can be simplified to:

$$I_R = mC \quad (10)$$

where  $m$  is a parameter empirically fit through a calibration using known concentrations and observed intensities. When  $m$  is determined, observed band intensity can be used to predict an unknown concentration. If the property of interest is not a concentration, but for instance a physical state as crystallinity, the concentration can be kept constant, and the equation can be rewritten using  $I_R$  and  $\sigma$  (Shaver, 2001).

Univariate analysis is very intuitive and straightforward to apply; however, it can be critical to base results on only one measured band intensity. If many chemical components are present in a sample, the band selected for univariate analysis may be a function of several of the components concentrations, thus individual contributions cannot be discriminated. A way to overcome this problem is to use the intensities of several bands or regions of the spectra simultaneously to determine various concentrations:

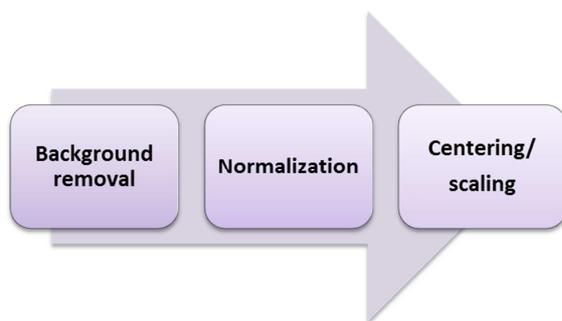
$$I_{R,i} = \sum m_{j,i} C_j \quad (11)$$

where  $I_{R,i}$  is the intensity observed for region or band  $i$ ,  $m_{j,i}$  is the empirically determined weighting factor for species  $j$  and band  $i$ , and  $C_j$  is the concentration of species  $j$ .

Compared to univariate data analysis, multivariate data analysis is more complex and misinterpretation or over-fitting is more likely to occur. However, multivariate analysis is advantageous in aspects like handling of interference, noise reduction, outlier detection, and for exploratory analysis. Most often, very small spectral changes can be detected and several compounds in a sample can be analyzed simultaneously. In multivariate analysis, the relationship between variables both within a spectrum and between individual spectra is utilized to obtain more accurate and precise determinations (Shaver, 2001, Bro, 2003).

## 5.2 Data Preprocessing

There is no substitute for optimal collection of Raman data, but, after proper data collection, preprocessing of spectral data is the most important step before multivariate modeling (e.g., Principal Component Analysis (PCA) and Partial Least Squares (PLS)). In Raman spectroscopy, the spectra are often affected by several factors other than the chemical components of interest, and if the effects from these unwanted factors are not eliminated, they can create challenges in the subsequent qualitative or quantitative analysis. In Raman spectroscopy, the typically unwanted effects originate from measurement noise or systematic errors related to interferences from optical and physical variations - e.g. changes in sample density, opacity, surface topography, or position. As previously described, chemical sample treatment and clever instrumental tuning can minimize spectral interferences, but another approach is to use mathematical preprocessing. Several studies have investigated the preprocessing methods for Raman data (Afseth *et al.*, 2006a; Rinnan *et al.*, 2009). This section describes preprocessing methods suitable for Raman spectra. The methods are ordered in groups according to which unwanted effect (noise, density, fluorescence, or other background phenomena) they are targeted in minimizing. An overview of the preprocessing described in this section and a general approach to preprocessing of Raman data is suggested in Figure 22.



**Figure 22.** A common order of Raman preprocessing steps.

Typically, normalization should be performed before any centering or scaling or other column-wise preprocessing steps and after background or offset removal.

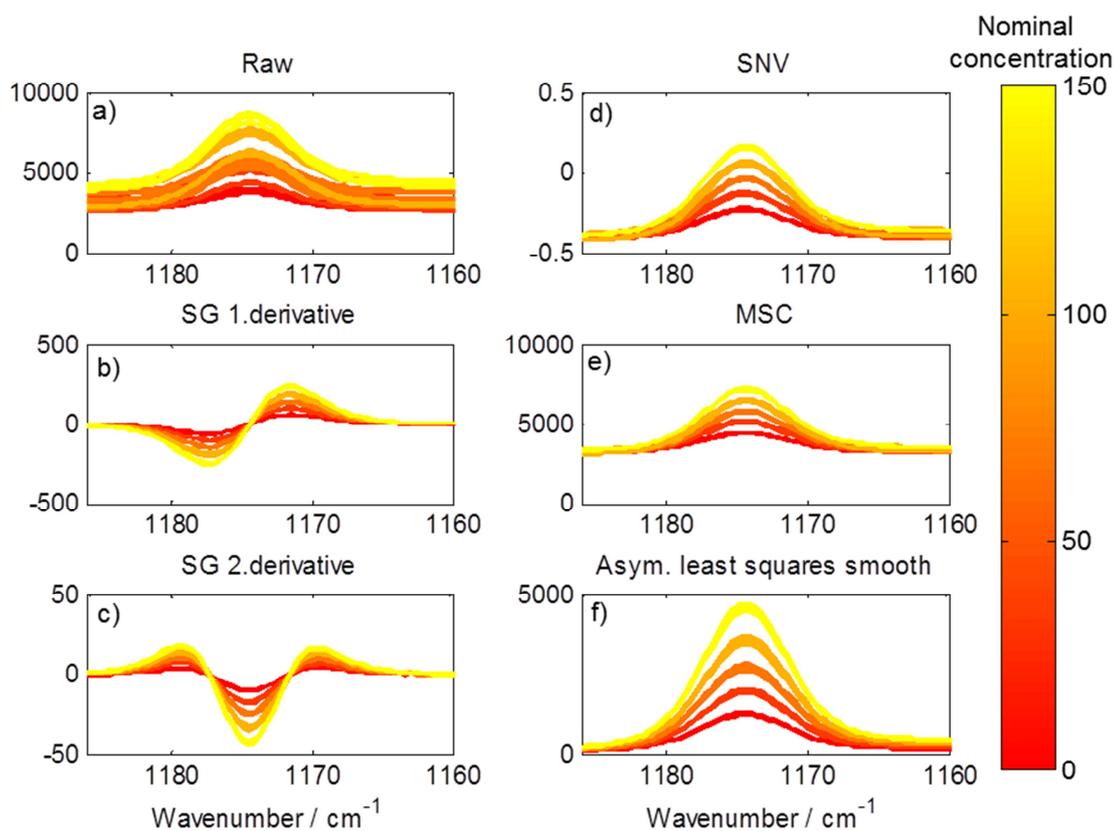
### ***5.2.1 Removal of background***

The background contribution in a Raman measurement can cause the baseline of the spectrum to vary from zero, in the literature techniques used to correct for this variation are referred to as “baseline-correction” or “background-correction” techniques. As previously described, a background in the Raman spectrum arise from laser-induced fluorescence, non-laser-induced emission processes (e.g. room light, sunlight, blackbody radiation), or Raman scattering of something beside the analyte (solvent, substrate, optics, etc.). Sometimes the source of background can be removed by experimental modifications (excitation wavelength or sample preparation). Eventually, many Raman spectra contain some unavoidable background, although it may not be strong enough to interfere with quantification.

The background causes a reduction of the signal to noise ratio of the Raman bands related to shot noise. Shot noise is always present, also in the fluorescence signal. If the fluorescence emission is orders of magnitude stronger than the Raman signal, the shot noise due to this fluorescence can be larger or of similar size as the Raman signal. Background correction technique can in this case remove broad structure of the fluorescence while still failing to remove shot noise imparted by the background (Shaver, 2001).

A common method used to remove unwanted background from signal of interest is taking the derivatives of the measured response. By taking the derivatives a filter is created between high-frequency (sharp and narrow) features like the Raman bands and low-frequency (broad and smooth) features such as the baseline. Point difference and Savitsky-Golay (SG, Savitzky & Golay, 1964) derivatives are the most common used approaches to calculating derivatives. Point difference, which is the simplest form, is done by simple subtracting each spectral point from one of the adjacent/neighbor points (variables). This subtraction removes the signal which is the same between the two variables and leaves only the part of the signal which is different. When performed on the entire spectrum the lower frequency signals will be

de-emphasized. A second derivative would be calculated by repeating the process, which will further accentuate higher-frequency features. In SG derivatives of a polynomial function are used to fit some local window of points around each spectral variable and the derivative is computed from this polynomial estimates instead of the original data. Each time a derivative is performed on the spectrum, the offset of the spectrum is removed and subsequently higher order shapes get simplified: linear trends become offsets, quadratic become linear, and so forth. Figure 23 b) and c) shows examples of how SG works on Raman spectra.

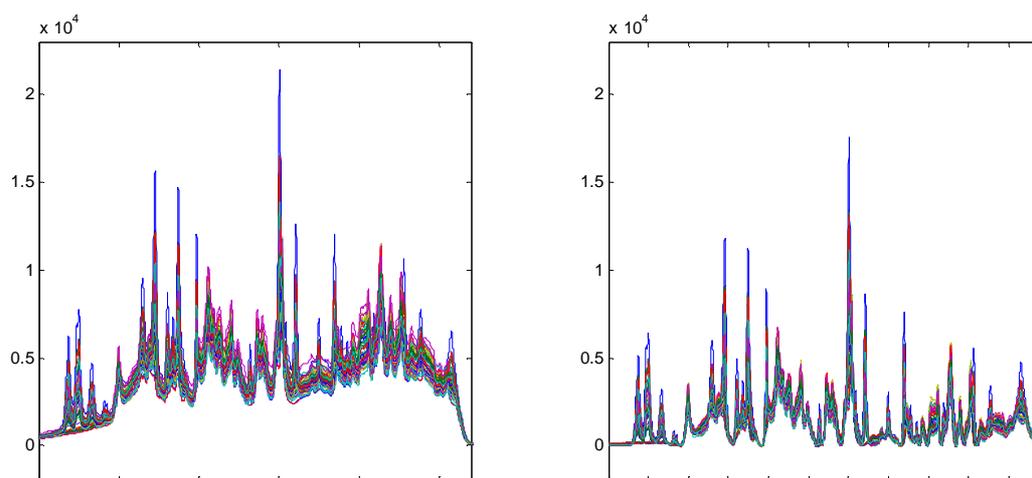


**Figure 23.** The effect of preprocessing of Raman spectra of HIV-preventive vaginal rings colored according to nominal concentration of API (data from Paper III) a) raw, b) Savitsky-Golay first derivative, c) Savitsky-Golay second derivative, d) standard normal variate, e) multiplicative signal correction, and f) asymmetric least squares smoothing.

The SG deritivation algorithm requires selection of the number of local points (window size) to fit the function to, the polynomial order, and the order of the derivatives. In order not to smoothen out non-noise features it is important to select a window size similar or smaller than the spectral size of these non-noise features. The SG algorithm can also be applied for simple smoothing by changing the derivative order (see section 5.2.2).

Weighted least squares (or asymmetric weighted least squares) is another preprocessing method for removing baseline offset from data, and the method is potentially very suited for Raman data where the signal in some variables is due only

to background (e.g. fluorescence). The variables containing only background provide a good estimate on how much background should be removed in the nearby variables. Points most likely due to baseline alone are determined by iteratively fitting a polynomial to each spectrum and up weighting variables below the polynomial in the next iteration. This iterative procedure continues until a user-defined number of variables are left. The overall effect is an automatic removal of the background while avoiding that highly negative bands are created. The baseline is usually approximated by low-order polynomials, but a specific baseline reference can be used. If a polynomial is used for determining the baseline, the order should be selected carefully. Additional unwanted variance might be induced if a too high order is applied. In general, subtraction of baseline is not as numerically safe as using the derivatives. However, interpretation of the spectra and loadings is easier (Eilers, 2004). Figure 24 shows an example of asymmetric least squares smoothing applied to correct the baseline fluctuation in the Raman spectra of ternary pharmaceutical mixtures, and in Figure 23f the technique is compared to other preprocessing methods in the preprocessing of Raman spectra of HIV-preventive vaginal rings.



**Figure 24.** Asymmetric least squares smoothing of Raman spectra of ternary pharmaceutical mixtures (data from Paper II).

### 5.2.2 Removal of noise

#### Random noise (shot noise)

The most significant source of noise is shot noise (or statistical noise) as described in section 4.3.1. There is a range of techniques available to remove random-noise from the Raman spectra (Beebe *et al.*, 1998). They vary from single-spectrum smoothing algorithms to multivariate analysis, where the correlation between an entire set of spectra is used to isolate signal from noise in each spectrum.

In single-spectrum techniques the relationship between adjacent spectral points are used to isolate signal from noise. Whereas noise usually appears as high-frequency

random fluctuations from one variable to the next, the signal is assumed to be lower frequency and more gradually changing from variable to variable. A simple boxcar averaging averages each point in the spectrum with a number of its neighbors. Effectively, the number of counts for each point is increased and, therefore, the relative error is decreased because of the square root relationship described earlier. The assumption is that adjacent points are essentially equal in the signal they report. In practice, bands are changing over small pixel ranges and the band shape will be distorted by the smoothing process. Weighting functions which decrease the influence of further-away points can be used in smoothing to reduce the band shape changes.

Savitsky-Golay smoothing (Savitsky & Golay, 1964) is a related approach which permits some lower frequency changes to occur over each small window of points. The algorithm fits individual polynomials to windows around each point in the spectrum, and these polynomials are then used to smooth the data. The input to the algorithm is the order of the polynomial and the size of the window (filter width). Increasing the window size and decreasing the polynomial order will increase the smoothing. A rule of thumb is that the window size should be on the order of, or smaller than, the nominal width of non-noise features (Rinnan *et al.*, 2009). Other approaches such as Fourier and wavelet filtering have also proven to be successful for removal of noise.

### Cosmic Ray Filter

Intense random noise spikes are common in data of CCD devices. These spikes are often caused by high-energy particles, such as cosmic rays, striking the CCD detector. In the spectrum cosmic rays appear as very narrow bands of high intensity. Because of the high intensities these bands can have a large influence on data analysis if not removed. A filtering algorithm provides a straightforward way of removing these effects. In the filtering algorithm, two separate data collections are performed and the resulting spectra are compared on a pixel-by-pixel basis. If the count difference for a given pixel exceeds twice the value of the square root of the minimum value, the smaller of the two counts is used. Otherwise, the two counts are averaged. This eliminates the sharp spikes produced by random cosmic background radiation, without distorting spectral band shapes (Shaver, 2001).

### **5.2.3 Normalization (object-wise)**

A large cause of systematic variation in the Raman spectral measurement arises from total intensity variations i.e. uniform intensity changes throughout the spectrum. As shown earlier (Equation 9) the observed Raman intensity can be related to either the concentration or the scattering efficiency of a species, under the assumption that all other parameters (incident intensity ( $I_0$ ), interrogated volume ( $L$ ), and instrumental throughput ( $k$ )) are constant. However, in practice this is very unlikely, since all three

parameters are subject to changes. The incident intensity will change with laser-intensity output and laser throughput to the sample. The interrogated volume can be influenced by changes in physical sample properties such as density, opacity, position, absorptivity and refractive index. Changes in the instrumental throughput may also occur over long time periods despite the stability of modern instruments. Often instruments can provide information on the changes in the throughput and incident intensity, but changes in the interrogated volume are harder to determine and correct (Shaver, 2001).

Quantitative Raman models are in most cases based on determination of the intensity, and will therefore be influenced by total intensity variations. A way to avoid this is to base models on other band parameters such as band width or position, which are unaffected by the changes in total intensity. However, quantitative models based on such parameters will often be subjects to significant interferences.

In most situations, a normalization of the entire spectrum is necessary. Normalization is performed by dividing each variable of a spectrum with a constant, and different normalizations are distinguished from each other by the choice of this constant. A common approach to normalization is to use the height of a single reference band. The height is the measured range between the baseline and the maximum intensity. This band can be an internal or external reference. A typical choice for reference bands in Raman spectra are solvent bands such as water bending ( $1650 \text{ cm}^{-1}$ ), C=C stretching ( $1660 \text{ cm}^{-1}$ ) or methylene scissoring ( $1450 \text{ cm}^{-1}$ ) (Pelletier, 1999a). Instead of using a single reference band, so-called vector normalization can be applied. In this type of normalization different vector-norms can be used as scaling factor.

**Table 8.** Overview of vector normalization methods

Normalization method	Description	$X_{i,\text{norm}} = \frac{x_i - a_i}{w_i}$
1-Norm	Normalization to the sum of the absolute value of all variables for a given sample	$a_i = 0$ $w_i = \sum_{j=1}^n  x_{i,j} $
2-Norm	Normalization to the sum of the squared value of all variables for a given sample	$a_i = 0$ $w_i = \sum_{j=1}^n x_{i,j}^2$
Euclidian/vector	Normalization to the square root of the sum of the squared value of all variables for a given sample	$a_i = 0$ $w_i = \sqrt{\sum_{j=1}^n x_{i,j}^2}$
SNV	Normalization to the standard deviation of all pooled variables. Prior to normalization the mean value of the variables are subtracted.	$a_i = \bar{x}$ $w_i = \sqrt{\frac{\sum_{j=1}^n (x_{i,j} - \bar{x})^2}{n - 1}}$

The most common to use is total sum of absolute values of all variables (1-Norm), or to use the square root of the sum of squared elements (Euclidian). Alternatively, the vector-norm called 2-Norm can be applied, where variables with larger values (higher intensities) will be more heavily weighted in the scaling (Rinnan *et al.*, 2009).

Another approach to normalization is standard normal variate (SNV). This method, proposed by Barnes *et al.* (1989), has proved to be very suitable for Raman data, and it has been applied as preprocessing method for all data obtained in this thesis (Papers I, II, and III). SNV is a weighted normalization method i.e. not all spectral points contribute equally to the normalization. Mathematically, SNV correction is identical to autoscaling of the rows instead of the columns of the matrix. Applying SNV, the individual mean value is subtracted from each sample spectrum and afterwards divided by the standard deviation of all pooled variables for the given sample. This normalization will give the spectrum a unit standard deviation of 1.

SNV is applied to every spectrum individually, and not on an entire set of spectra. The method works very well when much of the signal in a sample is the same in all samples, thus the number of bands are approximately the same, so that neither the mean spectral value nor the standard deviation will differ significantly between spectra. However, in situations where the overall signal differs significantly from sample to sample, problems may occur. In fact, it is quite possible that this normalization can lead to non-linear responses to what were originally linear relationships. When the method is used prior to quantification it should be carefully compared to other normalization methods. In Figure 23 d) SNV is compared to other preprocessing techniques.

#### Multiplicative signal correction (MSC)

For spectroscopic data one of the most widely used preprocessing techniques is MSC (Geladi *et al.*, 1985). Like SNV, MSC is used to compensate for additive and multiplicative effects. This is done by calculating the slope ( $b_{ref}$ ) and intercept ( $b_0$ ) of the regression between each individual sample spectrum ( $\mathbf{X}_{org}$ ) and a reference spectrum ( $\mathbf{X}_{ref}$ ), which is typically the mean spectrum of the calibration set:

$$\mathbf{X}_{org} = b_0 + b_{ref} \cdot \mathbf{X}_{ref} + \mathbf{e} \quad (12)$$

The corrected spectra are obtained by subtracting intercept ( $b_0$ ) from each variable in all the sample spectra followed by a division with the slope ( $b_{ref}$ ). The technique has been further developed to extended multiplicative signal correction (EMSC) (Martens & Stark, 1991). In addition to additive and multiplicative effects, EMSC can also account for e.g. wavelength dependent scattering effects. One disadvantage with MSC/EMSC is that the mean spectrum is used in the calculations and therefore the correction can be affected by outliers and extreme samples. Thus, it can sometimes be

advantageously to use SNV or derivatives, as they work independently of other spectra.

### 5.3 Qualitative analysis

In qualitative analysis the goal is to find and utilize similarities and differences between samples to group them without prior knowledge. Because of the very well resolved bands found in the spectra, Raman spectroscopy provides a very powerful tool for distinguishing between samples with only minimal differences. However, the variations between spectra can be very small and difficult to identify visually, thus chemometric methods can be useful for exploring spectral data. Approaches used for finding patterns and clusters within a data set are called unsupervised techniques because no prior information needs to be presented to the model prior to analysis.

#### 5.3.1. Principal component analysis (PCA)

The most commonly used unsupervised multivariate technique is principal component analysis (PCA). The method is fundamental and can be used for many different purposes: data reduction, pattern recognition, outlier detection, or classification.

Equation 13 expresses the PCA model. A data matrix,  $\mathbf{X}$ , can be decomposed into a score matrix ( $\mathbf{T}$ ) and a loading matrix ( $\mathbf{P}$ ), which capture the main variation in data, leaving un-modeled the unsystematic variation in the residuals,  $\mathbf{E}$ .

$$\mathbf{X}_c = \mathbf{TP}^T + \mathbf{E} \quad (13)$$

where  $\mathbf{X}_c$  is mean centred (subtracting the mean of each variable from the original measurement) or autoscaled (subtracting the mean of each variable from the original measurement and dividing with the standard deviation of each variable). Systematic variations in samples and variables are found in the score matrix ( $\mathbf{T}$ ) and loading matrix ( $\mathbf{P}$ ), respectively. By PCA, the main variation in a multidimensional data set is found by creating new linear combinations of the original variables. These linear combinations are called principal components (PCs) and they are found from consecutive orthogonal subtraction of the largest variation in the data. This means that the first principal component is the direction through the data that explains the largest variation in the data. The second and subsequent PCs must be orthogonal to the previous PCs and describe the maximum amount of the remaining variation. Each PC is the outer product of a score vector ( $\mathbf{t}_i$ ) and a loading vector ( $\mathbf{p}_i$ ), where  $i$  denote the component number.

$$\mathbf{X}_c = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_i\mathbf{p}_i^T + \mathbf{E} \quad (14)$$

The number of PCs expresses the number of observed variations in the data i.e. the number of independent phenomena. If the original variables in a data set co-vary, the

number of PCs is expected to be much smaller than the number of original variables. In case of spectral data such as Raman spectra, where the variables are highly collinear, the data matrix is likely to be decomposed into a considerably lower number of PCs than the number of variables (Wold *et al.*, 1987).

The result of a PCA model can be graphically illustrated as a score plot and a loading plot. From the score plot, sample similarities (trends and groupings) can be identified as well as outliers. The loading plot shows the importance of the original variables for each PC, e.g. which variables are responsible for the separation of samples or responsible for a particular sample to be an outlier.

Different algorithms can be applied to find the components in PCA. Typically, either the non-linear iterative partial least squares (NIPALS) algorithm or singular value decomposition (SVD) is used. NIPALS is an iterative procedure involving sequential computation of the eigenvectors (loadings) by order of explained variance (see Box 3). SVD uses eigenvalue decomposition in a non-sequential way. Both methods have their benefits: NIPALS is much faster for large data sets, because not all components are necessarily calculated. Furthermore, NIPALS can handle missing values, which SVD cannot directly. SVD is numerically more stable and more efficient than NIPALS if not only the first few components are required. Thus, SVD is better for discriminating between components that only differ slightly, which becomes relevant for very large matrices (Wold *et al.*, 2001; Wu *et al.*, 1997; Geladi & Kowwalski, 1986).

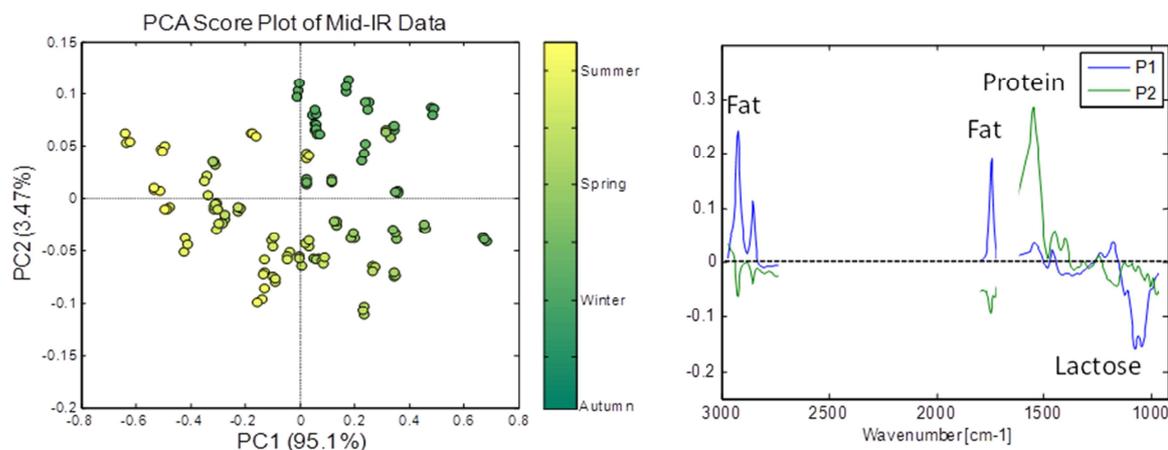
In PCA, there is an orthogonality constraint on the PCs, forcing them to be completely uncorrelated. This means that the interpretation is simplified in the sense that each component describes a variance structure (pattern) in the data. However, it also means that if chemical compounds in a sample co-vary, the information about the individual compounds will not be resolved, and the loadings will thus not directly represent the pure spectra of the chemical compounds. Such co-variations are much better handled by methods such as multivariate curve resolution-alternating least squares (MCR-ALS, section 5.4.2), which does not impose the orthogonality constraint.

### Examples of how PCA can be applied

#### *Exploring seasonal variations in milk (Poster I)*

Cow milk samples collected at dairies from 2009/2010 were analyzed by mid-IR spectroscopy. The aim of the study was to explore the seasonal variation in milk using mid-IR spectroscopy. In order to identify seasonal trends a visual inspection of the IR spectra was first initiated. However, such inspection can often be very difficult since spectral sample differences may be very small, and the underlying pattern of interest (this case seasonal variation) can be complex combinations of small spectral bands. Moreover, this can be especially challenging if knowledge on spectral interpretation is

limited, thus no guide to where to look in the spectrum. As a tool to help explore the IR-spectra PCA was applied. The results are shown in Figure 25.

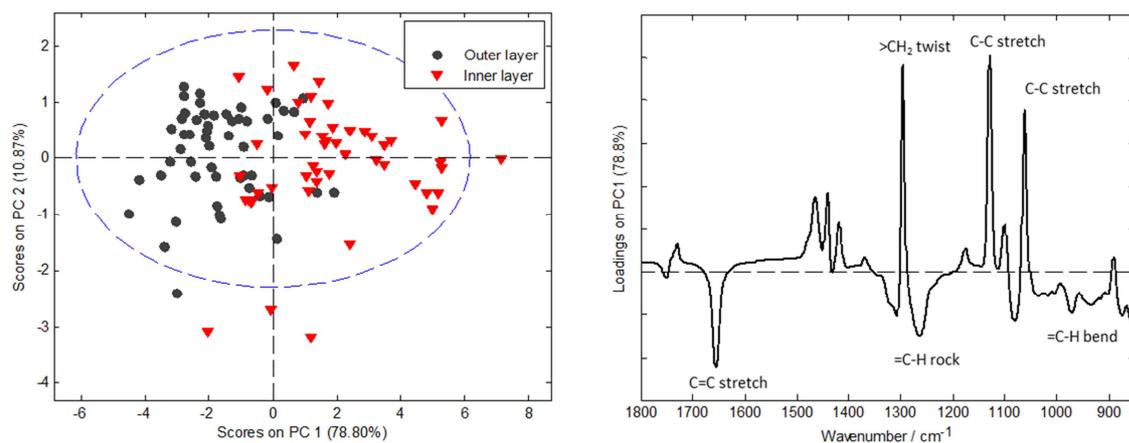


**Figure 25.** PCA score plot and loadings of mid-IR spectra of cow milk samples with seasonal variations.

In the left side Figure 25 a PCA score plot of the first and second principal component is shown. All milk samples are colored according to which time of the year they originate from, and the plot shows a clear seasonal variation in the milk. In the right side of the figure the corresponding loadings show that the seasonal variation is correlated to the molecular vibrations of fat, lactose and protein, indicating that in raw milk these constituents vary gradually during the season. Careful inspection of the seasonal variation shows that autumn milk samples are high in both fat and protein. From autumn to winter the protein in the milk decreases, and from winter to spring the milk decreases in fat. From spring to summer an increase in the protein is seen, and from summer to autumn the milk increases in fat. The score plot thus gives a nice graphical representation of the seasonal variation in milk.

#### *Discriminating between fat layers (Paper I)*

Knowledge about the depth of the individual fat layers in pork backfat may help improve the quality, because each layer has different sensorial and technological properties. In Paper I Raman spectroscopy was investigated and applied to discriminate between fat layers in samples of porcine backfat. The PCA score plot in Figure 26 (left) shows an almost complete separation of fat layers along the first principal component which describe 79% of the total spectral variation.



**Figure 26.** Principal component analysis of Raman spectra from porcine backfat with varying distance to skin surface. Left: PCA score plot of the first and second principal component shows the separation between samples from the inner and outer fat layer of porcine backfat. Right: PCA loadings on the first principal component.

When examining the corresponding loading vector for the first component (Figure 26, right) it is observed that samples from the inner fat layer are positively correlated to high intensity of the bands  $1296\text{ cm}^{-1}$  ( $>\text{CH}_2$  twist),  $1125\text{ cm}^{-1}$  (C-C in-phase stretch) and  $1060\text{ cm}^{-1}$  (C-C out-of-phase stretch). Furthermore, the samples from the inner layer also contain higher amounts of the  $>\text{CH}_2$  scissoring deformation ( $1438\text{ cm}^{-1}$ ). These findings indicate that the inner layer contains a higher amount of  $\text{CH}_2$  presumably due to a higher amount of saturated fatty acids. The samples from the outer layer have negative PC1 score values which is correlated to high intensity of the three bands related to double bonds  $1655\text{ cm}^{-1}$  (C=C stretch),  $1266\text{ cm}^{-1}$  (=C-H rock) and  $970\text{ cm}^{-1}$  (=C-H band). This is in accordance with the fact that saturation will increase as a function of penetration depth (from the skin surface).

#### 5.4 Quantitative Raman spectroscopy

Under constant experimental conditions, the number of Raman scattered photons is proportional to analyte concentration, and as previously shown (Equation 9) the intensity of Raman scattering can be expressed by an equation analogous to Lambert-Beer's law. This fundamental relationship between intensity and concentration is the basis of using Raman for quantitative analysis. There are different approaches to quantitative analysis of Raman spectra. Sometimes, the intensity of a single band or the ratio of bands can be used for quantification, but in other situations more advanced methods like PLS need to be applied. In Paper III, simple and more advanced approaches to quantification are studied and compared. MCR-ALS is usually considered to be an unsupervised technique such as PCA. However, when applying certain constraints (e.g. correspondence among species or correlation constraints) MCR-ALS becomes a supervised technique that similarly to PLS can be applied for quantitative purposes. Paper II presents a more complex Raman quantification case comparing the use of PLS with a novel extension of MCR-ALS. In the following

sections, PLS and MCR-ALS will be explained. For PLS, it was decided to focus on the model, whereas for MCR-ALS both the model and the algorithm is described, to provide a better background to Paper II, where an extension to MCR-ALS has been developed.

### 5.4.1 Partial Least Squares (PLS) regression

PLS is a commonly used and well described method for multivariate regression (Wold *et al.*, 2001, Geladi & Kowalski, 1986, Sjöström *et al.*, 1983). In PLS regression, a set of dependent variables  $\mathbf{y}$  or  $\mathbf{Y}$  is modeled from a set of independent variables  $\mathbf{X}$ . The purpose of this is to estimate  $\mathbf{Y}$  from future measurements of  $\mathbf{X}$  only, requiring a representative calibration set with known values of  $\mathbf{X}$  and  $\mathbf{Y}$ .

PLS is well suited for spectroscopic data sets for two reasons: 1) they often contain more variables than samples and 2) many spectral variables co-vary, which means that the number of variables is much higher than the number of independent phenomena in the data. Mathematically, this leads to rank deficiency, which means that the simpler regression method, multivariate linear regression (MLR) does not work.

In PLS, the individual variations in  $\mathbf{X}$  and  $\mathbf{Y}$  and their mutual correlation are simultaneously maximized. Explaining PLS is often done by introducing the concept of an outer relation (PCA on  $\mathbf{X}$  and  $\mathbf{Y}$  individually) and an inner relation where the association between  $\mathbf{X}$  and  $\mathbf{Y}$  scores is found by linear regression. However, PLS does not actually work that simple, because when PCA is calculated on the  $\mathbf{X}$  and  $\mathbf{Y}$  separately, their mutual relation will be weak. A more optimal way is to give  $\mathbf{X}$  and  $\mathbf{Y}$  information about each other by exchanging scores, which is the process done in the NIPALS algorithm (Box 2, Wold *et al.*, 2001, Geladi & Kowalski, 1986).

<b>Box 2.</b> Non-linear iterative partial least squares (NIPALS, Wold <i>et al.</i> , 2001; Geladi & Kowalski, 1986)	
Model: $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$	7. Iterate step 3 through 6 until convergence. For a single response of $\mathbf{y}$ the algorithm will converge in one run
1. Pre-process $\mathbf{X}$ and $\mathbf{Y}$	8. Solve $\mathbf{X} = \mathbf{t}\mathbf{p}^T$ wrt $\mathbf{p} \rightarrow \mathbf{p}^T = (\mathbf{t}^T\mathbf{t})^{-1}\mathbf{t}^T\mathbf{X}$
2. Select a random column of $\mathbf{Y} \rightarrow \mathbf{u}$	9. Deflate $\mathbf{X}$ and $\mathbf{Y}$ : $\mathbf{X}_{\text{new}} = \mathbf{X} - \mathbf{t}\mathbf{p}^T$
3. Solve $\mathbf{X} = \mathbf{u}\mathbf{w}^T$ wrt $\mathbf{w}$ and normalize $\mathbf{w}$	$\mathbf{Y}_{\text{new}} = \mathbf{Y} - \mathbf{u}\mathbf{q}^T$
4. Solve $\mathbf{X} = \mathbf{t}\mathbf{w}^T$ wrt $\mathbf{t} \rightarrow \mathbf{t} = \mathbf{X}\mathbf{w}$	10. Continue with steps 2 through 9 until wanted number of factors are extracted
5. Solve $\mathbf{Y} = \mathbf{t}\mathbf{q}^T$ wrt $\mathbf{q} \rightarrow \mathbf{q}^T = (\mathbf{t}^T\mathbf{t})^{-1}\mathbf{t}^T\mathbf{Y}$	11. $\mathbf{b} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}$ (regression vector)
6. Solve $\mathbf{Y} = \mathbf{u}\mathbf{q}^T$ wrt $\mathbf{u} \rightarrow \mathbf{u} = \mathbf{Y}\mathbf{q}(\mathbf{q}^T\mathbf{q})^{-1}$	

#### ***5.4.2 Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS)***

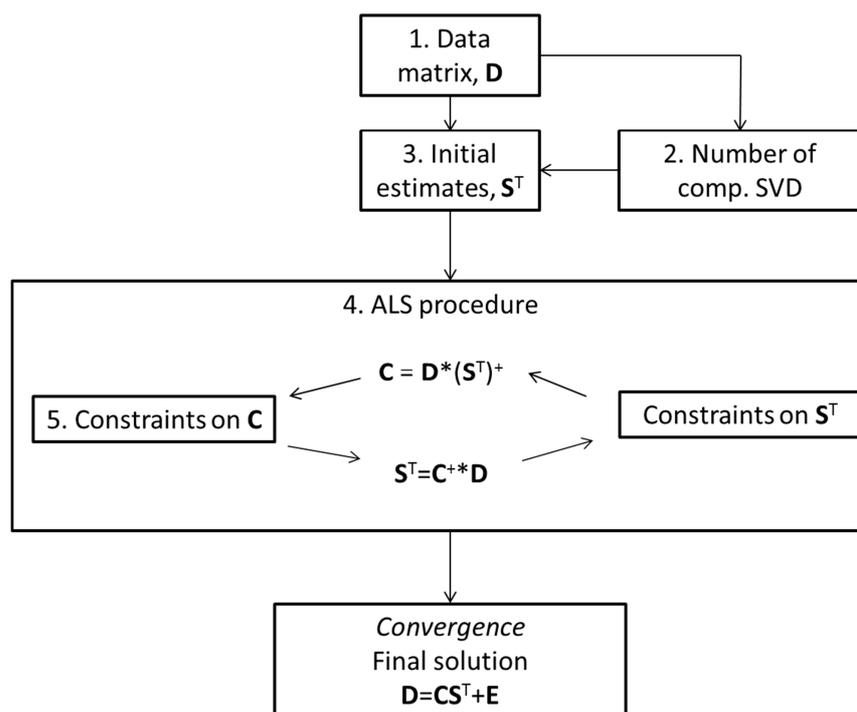
The iterative MCR-ALS method is a bilinear decomposition of data matrix (**D**), into a chemically meaningful concentration matrix (**C**) and a spectral matrix (**S**), and a residual matrix (**E**) (Tauler, 1995). The MCR-ALS model is given by:

$$\mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad (15)$$

As previous described, MCR does not impose the orthogonality constraint (opposed to PCA) when extracting components, thus the components can be direct spectral estimates of pure chemical compounds. This is nice because the components scores are then proportional to the concentrations of actual chemical compounds. A drawback of MCR-ALS (compared to PLS) is that it provides a solution that is not unique, which means that a number of solutions may be obtained with different combinations of **C** and **S** matrices. This is called rotational ambiguity. A way to overcome rotational ambiguity is by applying different constraints on both **S** and **C**, which will help to minimize the number of possible solutions, and maybe lead to a desired uniqueness.

In the following the iterative process of building the MCR-ALS model will be described. Figure 27 gives an overview of the iterative process in MCR-ALS. It consists of the following main steps:

1. Building the data matrix
2. Estimation of the number of components (the rank) present in data matrix
3. Finding initial estimates of spectral (or concentration) profiles
4. Iterative ALS calculation of spectral and concentration profiles under constraints
5. Checking for convergence



**Figure 27.** Overview of the steps in the multivariate curve resolution-alternating least squares (MCR-ALS) procedure (Paper II).

1. A data set consists of  $n$  samples containing  $nc$  compounds. This can be arranged in a data matrix  $\mathbf{D}$ . Under the assumption of the natural spectroscopic bilinear model,  $\mathbf{D}$  can be expressed as shown in Equation 15.  $\mathbf{D}$  consists of  $n$  rows of different individual spectra measured for different samples and  $m$  columns of spectral wavenumbers. The columns in  $\mathbf{C}$  represent the concentrations of  $nc$  components in these  $n$  samples and the rows in  $\mathbf{S}^T$  are their related  $nc$  spectra of length  $m$ .  $\mathbf{E}$  is the residual matrix with the variance not explained by the bilinear model. The goal of curve resolution methods is to perform the matrix decomposition according to Equation 15, i.e. estimate  $\mathbf{C}$  and  $\mathbf{S}^T$  matrices from the analysis of the data matrix  $\mathbf{D}$ . An infinite number of possible solutions for the mathematical decomposition of  $\mathbf{C}$  and  $\mathbf{S}^T$  exists unless additional information is provided for the bilinear decomposition.

2. If the number of chemical compounds present is known, this is often a good estimate of how many components to use. If it is more difficult to determine number of components, SVD can be applied to help this determination. From the unique orthogonal decomposition, the number of significant components directly related to chemical compounds can often be estimated (Malinowski, 1992).

3. The initial estimates used should be a set of either  $nc$  concentration or spectral profiles. In some cases SIMPLISMA (Windig & Guilment, 1991) initial estimates from the purest spectral profiles can be determined.

4. If spectral profiles are the initial estimates, concentration profiles, and after that improved spectra, can be calculated by least squares with the following expressions:

$$\mathbf{C} = \mathbf{D}(\mathbf{S}^T)^+ \quad (16)$$

$$\mathbf{S}^T = \mathbf{C}^+\mathbf{D} \quad (17)$$

where  $(\mathbf{S}^T)^+$  and  $\mathbf{C}^+$  are the Moore-Penrose pseudo-inverse matrices of the spectra matrix  $\mathbf{S}^T$  and  $\mathbf{C}$ , respectively (which are equal to  $\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}$  and to  $(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$  if  $\mathbf{S}^T$  and  $\mathbf{C}$  are full rank matrices). This is implemented in an alternating least squares cycle where in each iteration new  $\mathbf{C}$  and  $\mathbf{S}^T$  estimates are obtained. Within each ALS step, constraints can be applied to modify the initial profiles towards chemically meaningful ones.

5. Convergence of the algorithm is checked by evaluating the relative change in the lack of fit of the reconstructed matrix  $\mathbf{D} = \mathbf{C}\mathbf{S}^T$  between two consecutive iterations. When this difference is below a preset threshold value the iterative optimization stops.

As previous described constraints can be applied on both concentrations and spectral profiles, and these may help minimize problems of rotational ambiguity and provide a unique solution. In MCR-ALS step 4 (Figure 27) it is possible to apply some of the following described constraints:

*a. Non-negativity constraint* - Concentrations of chemical species are always values equal to or larger than zero. Therefore non-negativity can be applied to the set of concentration profiles (Manne, 1995; Bro & DeJong, 1997). Non-negativity can also be applied to spectral profiles. However, care must be taken as preprocessing like derivative and SNV-correction results in spectra containing negative values.

*b. Selectivity and local rank constraint* - This constraint is applied when some species are absent in a particular sample (e.g. samples with known “0% concentration”). This constraint can be used within a matrix or subset in a multiset structure and increases the accuracy in the definition of the profiles (Manne, 1995; Tauler *et al.*, 1995).

*c. Correspondence among species* - This constraint is only used if the data are structured as a multiset with  $ns$  subsets. Correspondence among species defines the number and identity of components in each single matrix along the augmented data set. This information is usually introduced in the algorithm through a binary coded matrix of size  $ns$  by  $nc$ , which by 1 and 0 indicates if a species is present or absent in the related subset (Tauler & Barcelo, 1993).

*d. Correlation constraint* - In quantitative analysis a correlation constraint can be applied to the concentration profiles (Antunes *et al.*, 2002). The data is divided into a calibration set and a test set. The initial step when applying a correlation constraint is

to create a mask matrix  $\mathbf{C}^*$ , of size  $n$  by  $nc$  (similar to the  $\mathbf{C}$  matrix), with real numbers equal to known concentration values in the positions to be constrained (i.e. concentration values of the calibration set) and missing values in the positions left unconstrained (the test set samples). In each MCR-ALS iteration, the concentrations estimated by ALS for the calibration set are regressed against the known values in  $\mathbf{C}^*$  (reference concentrations). For a particular chemical component, a local linear regression model between  $\mathbf{c}$  and  $\mathbf{c}^*$  is estimated by least squares:

$$\mathbf{c} = \mathbf{b} \mathbf{c}^* + \mathbf{b}_0 + \mathbf{e} \quad (18)$$

Since only the calibration set has known values in  $\mathbf{c}^*$ , only those entries (indicated by  $\mathbf{c}^*_{\text{cal}}$ ) can be used in the local model:

$$\mathbf{c}_{\text{cal}} = \mathbf{b} \mathbf{c}^*_{\text{cal}} + \mathbf{b}_0 \quad (19)$$

Next, the model parameters  $\mathbf{b}$  and  $\mathbf{b}_0$  are used to predict the concentrations of the unknown and unconstrained values in the test set:

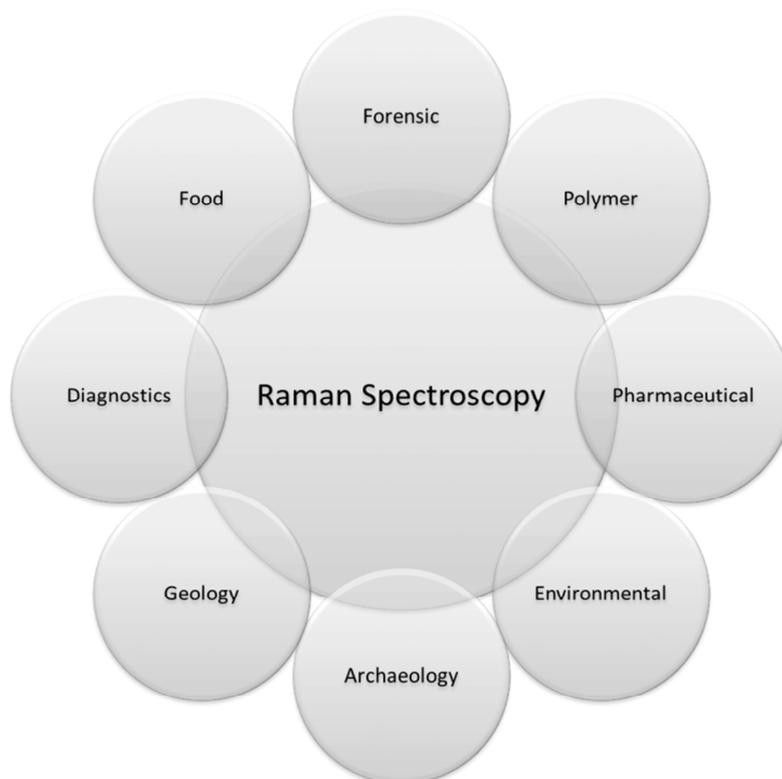
$$\mathbf{c}^*_{\text{test}} = (\mathbf{c}_{\text{test}} - \mathbf{b}_0) / \mathbf{b} \quad (20)$$

Each ALS iteration is completed after updating the  $\mathbf{C}$  matrix with the predictions ( $\mathbf{C}_{\text{ALS}}$  is substituted by  $\mathbf{C}^*_{\text{cal}}$  and  $\mathbf{C}^*_{\text{test}}$ ) and new spectral profiles  $\mathbf{S}^T$  are found. This constraint, as all others, can be applied to one or more chemical components (concentration profiles) in the data set. For each one of them, an individual calibration model (Equation 19) would be built. Note that the fact of using individual calibration models per compound instead of a global one is of utmost importance, since this allows selecting which compounds must be calibrated (not necessarily all the ones contributing to the measured signal). This one-at-a-time calibration strategy allows keeping the structure of the real bilinear spectroscopic model (equal to the one used in the direct calibration method classical least squares, which builds a global calibration model for all compounds) but enables calibration in the presence of interferences, which are modeled using other, softer constraints.



## 6. Raman analysis of food and pharmaceutical products

Raman spectroscopy has during the last years gained increased interest as a (potentially) quantitative, fast measurement method in the food and pharmaceutical industries. Moreover, it is becoming increasingly used in a broader range of scientific research areas. As shown in Figure 28, applications in various fields of science such as forensic, pharmaceutical, environmental, petrochemistry/polymer, archaeology, geology, biology, diagnostics, and food can be found.



**Figure 28.** Overview of the broad range of scientific research areas where Raman spectroscopy has been applied.

The technique was initially applied to examine inorganic compounds, but it rapidly grew due to an extensive use in polymer science. More recently the authorities within the pharmaceutical science recognized the possibilities of using the technique for rapid on-line quality control during production, leading to a growth in application to pharmaceutical process monitoring. Other successful applications have been established within e.g. paint production, semiconductors, art, archaeology, and biotechnology. There have also been major advances in forensic investigation and process analysis. Compared to other analytical methods Raman spectroscopy is viewed by many as a niche technique. However, due to the recent year's instrumental developments, the technique has become applicable to more complex sample systems and the applications are fast growing. Major advantages of the technique, such as minimal sample preparation, compatibility with aqueous solutions, remote sampling possibilities (by the use of optical fiber), and the ability to analyze materials

in/through glass containers, water or inside sealed packaging makes it a versatile tool for at-line, on-line or in-line analysis in a broad range of fields. In line with the papers presented in this thesis the main focus of this chapter will be on three specific Raman application fields: 1) the analysis of food lipids, 2) the analysis of API through packaging, and 3) the analysis of API in different solid dosage formulations.

## **6.1 Raman Spectroscopy of foods**

### ***6.1.1 Food macro components***

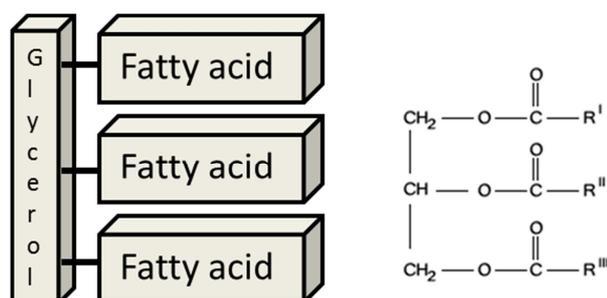
There has been a steep increase in the number of peer-reviewed publications of Raman spectroscopy in food science through the last decade as was shown in Figure 1. Review papers reporting applications of Raman spectroscopy in food science (LiChan, 1996; Yang & Ying, 2011) describe the wide range of different purposes (both qualitative and quantitative) the technique can be used for. The list of different food products Raman spectroscopy has been applied to is long and very diverse with products such as almonds (Micklander *et al.*, 2002), carrots (Killeen *et al.*, 2013), salmon (Afseth *et al.*, 2006b), oils (Muik *et al.*, 2005), chocolate (Dahlenborg *et al.*, 2012) and whisky (Boyaci *et al.*, 2012).

Raman spectroscopy is suitable for analyzing the entire range of macro components in food: Protein, carbohydrates, lipids, water, plus minor compounds such as carotenoids. From the Raman spectra of proteins information on the microenvironment and chemistry of side chains can be derived, as well as information on the conformation of the polypeptide backbone (LiChan, 1996). Stretching modes of amide I, III and skeletal provide information on relative amounts of different secondary structures in polypeptides or proteins. Furthermore, the amino acid side chains, such as cystine and cysteine, are easy to detect since the stretching modes of S-S or S-H are very Raman active. In the analysis of carbohydrates Raman spectroscopy has been used to elucidate the structural composition of di-, oligo- and polysaccharides, and to distinguish between closely related structures and isomers of glucose and fructose (Soderholm *et al.*, 1999). In addition, Raman has been applied for rapid analysis of carbohydrates in carrageenan powders (Dyrby *et al.*, 2004), beverages (Delfino *et al.*, 2011) and honey (Ozbalci *et al.*, 2013). Raman analysis has been applied in determining contents of minor food components such as carotenoids in Japanese tea (Ozaki *et al.*, 1992), amygdalin in bitter almonds (Micklander *et al.*, 2002), astaxanthin in salmon (Wold *et al.*, 2004), and melamine in milk powder (Okazaki *et al.*, 2009). In Paper I of this thesis, Raman spectroscopy was applied to analyze lipids; therefore the following sections will describe the analysis of food lipids by Raman spectroscopy in details.

### 6.1.2 Food lipids

Lipids have some primary functions in foods. First of all, they are of high nutritional value since they act as fuel molecules and source of essential fatty acids and vitamins. Furthermore, in the handling and processing of food, they are important for both flavor (giving a pleasant creamy or oily mouth feel) and texture (can act as food emulsifiers) (Belitz *et al.*, 2009).

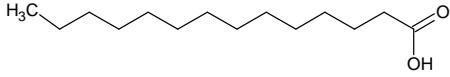
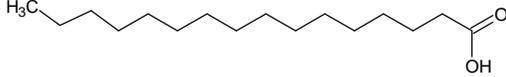
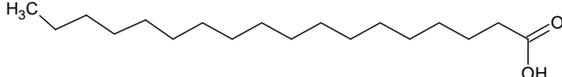
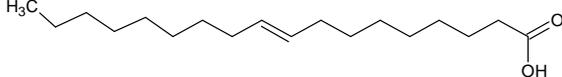
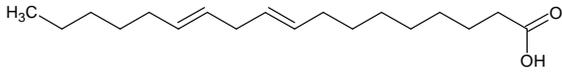
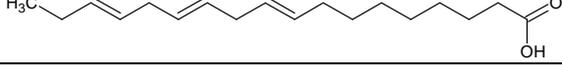
The lipids of plant and animal origin have traditionally been called fats and oils, depending on their melting point. Edible oils or fats consist nearly completely of triglycerides (also called triacylglycerols). Triglycerides (TGs) are glycerol esters of fatty acids as depicted in Figure 29 (Ouellette, 1998).



**Figure 29.** Triglycerides contain full substitution of all three hydroxyl groups of the glycerol molecule with ester linked acyl residues (modified from Quellette, 1998).

Fatty acids (FAs) are characterized by a hydrocarbon chain ending in a carboxyl functionality. They can be classified according to chain length, number, position, and configuration of their double bonds, and the occurrence of additional functional groups along the chains. When classified according to the number of double bonds they are defined as saturated or unsaturated. Saturated FAs (SFA) contain only single bonds between carbon atoms and, hence, the molecule is saturated or contains the maximum possible number of hydrogen atoms per carbon. The unsaturated FAs are divided into monounsaturated fatty acids (MUFA) in which the carbon chain contains only one double bond, and polyunsaturated fatty acids (PUFA) in which the carbon chain contains two or more double bonds (Ouellette, 1998). Unsaturated fatty acids (UFA) can exist in different geometric forms, or isomers. In most naturally occurring UFAs have *cis* formation where the hydrogens next to double bonds are on the same side of the carbon chain. If the double bond is altered, moving the hydrogens across from each other, the formation is called *trans*. *Cis* carbon chains are more bended than *trans*. *Trans* fatty acids are a health concern since they have shown to raise LDL cholesterol levels and, therefore, increase the risk of heart diseases. Some PUFAs have conjugated chains, i.e. long chains of alternating single and double bonds. The positional isomer called conjugated linoleic acid is being studied for positive health effects (Nagpal *et al.*, 2007). The major fatty acids which occur in foods are listed in Table 9.

**Table 9.** Structure of major food fatty acids (modified from Belitz et al., 2004).

Abbreviated designation	Structure	Common name
14:0		Myristic acid
16:0		Palmitic acid
18:0		Stearic acid
18:1 (9)		Oleic acid
18:2 (9,12)		Linoleic acid
18:3 (9,12,15)		Linolenic acid

Fats are solids or semisolids at room temperature and are usually from animals. They have a higher percentage of saturated FAs, whereas oils have a high percentage of unsaturated FAs. Oils are obtained from vegetable sources such as olives, sunflower, corn, rapeseed and soybeans. The unsaturated FA residues in the molecules of oils lower their melting point causing them to be liquid at room temperature. The short-chain fatty acids (<C14) are only found in fat and oils of milk, coconut and palmseed. Animals fats are dominated by C16:0 and C18:0 fatty acids, including a large portion of fully saturated TGs. Fish oil is rich in long chain unsaturated fatty acids (>C18) with a high content of conjugated double bonds (Ouellette, 1998). An overview of the positions and distribution of fatty acids in the triglycerides of different lipids sources are found in Table 10.

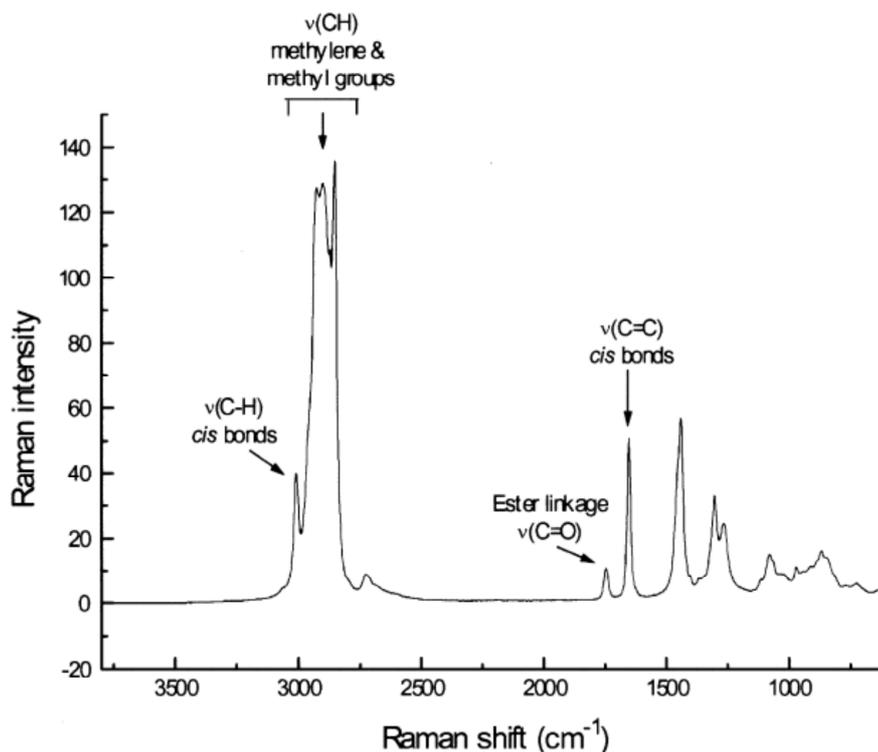
**Table 10.** FA Positional Distributions in TG from Various Sources (Compiled from Belitz *et al.*, 2009).

Lipid source	sn-Position	Fatty Acid (%)								
		8:0	10:0	12:0	14:0	16:0	18:0	18:1	18:2	18:3
Coconut	1	4	4	39	29	16	3	4	-	-
	2	2	5	78	8	1	1	3	2	-
	3	32	13	38	8	1	1	3	2	-
Cocoa butter	1	-	-	-	-	34	50	12	1	-
	2	-	-	-	-	2	2	87	9	-
	3	-	-	-	-	37	53	9	-	-
Corn	1	-	-	-	-	18	3	28	50	-
	2	-	-	-	-	2	-	27	70	-
	3	-	-	-	-	14	31	52	1	-
Sunflower	1	-	-	-	-	11	3	17	70	-
	2	-	-	-	-	1	1	22	76	-
	3	-	-	-	-	10	9	28	54	-
Olive	1	-	-	-	-	13	3	72	10	1
	2	-	-	-	-	1	-	83	14	1
	3	-	-	-	-	17	4	74	5	1
Peanut	1	-	-	-	-	14	5	59	19	-
	2	-	-	-	-	2	-	59	39	-
	3	-	-	-	-	11	5	57	10	-
Beef fat	1	-	-	-	4	41	17	20	4	1
	2	-	-	-	9	17	9	41	5	1
	3	-	-	-	1	22	24	37	5	1
Pig (outer back)	1	-	-	-	1	10	30	51	6	-
	2	-	-	-	4	72	2	13	3	-
	3	-	-	-	-	-	7	73	18	-

### 6.1.3 Raman scattering properties of food lipids

The potential of Raman spectroscopy for analysis of fats and oils has been recognized for some time (Horvat & Bailey, 1971). Raman spectroscopy determines the change in polarizability during a molecular vibration (as described in section 2.2). For molecules with triple or double bonds (e.g. carbon-carbon double bonds) there is a high electron density between the atoms. Because of this organization the electron cloud can be easily distorted by the incident photons of the laser light, and thereby induce a dipole moment. Therefore the Raman signal for the stretching vibration of such functional groups becomes strong, like the double bonds between carbons present in unsaturated fatty acids.

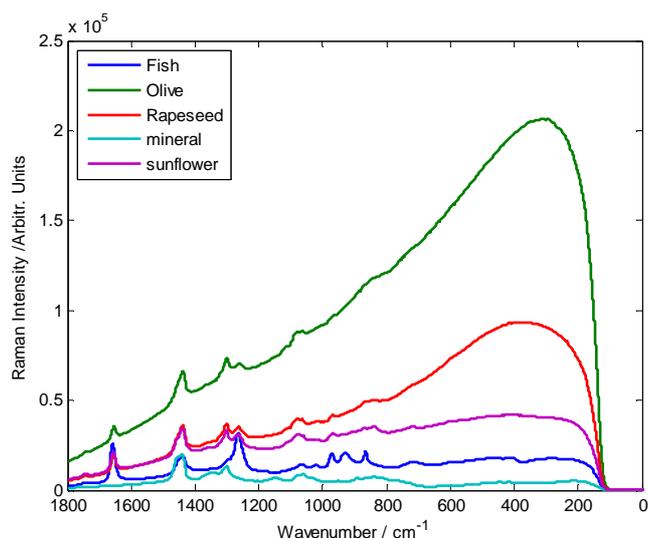
Figure 30 shows typical Raman spectra of oil, where the spectral bands are assigned to different functional groups. In relation to the measurement of the degree of unsaturation, the region from 1700 to 1600  $\text{cm}^{-1}$  is particularly interesting due to the appearance of the strong band originating from the C=C stretching vibration. Another interesting feature in the Raman spectra of fats and oils are the fact that the C=C stretching of *cis* and *trans* double bonds are observed at different wavenumbers (1656 and 1670  $\text{cm}^{-1}$ , respectively). This makes it possible to determine the type of unsaturation from the Raman spectra. The presence of conjugated double bonds C=C will give rise to a band around 1630  $\text{cm}^{-1}$ .



**Figure 30.** Raman spectrum of edible oil (Dobson, 2001).

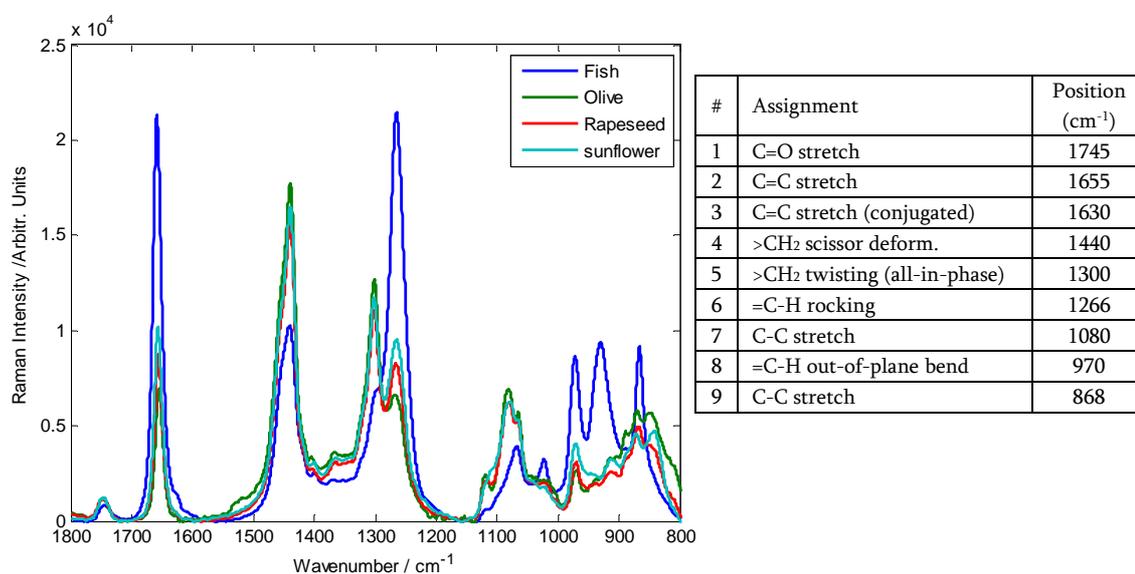
Although the spectra of most pure oils and fats appear fairly similar visually, they reflect well the differences in fatty acid composition that exist among different types of fats and oils. Detailed studies of the spectra of a large variety of fats and oils have resulted in the identification of specific bands or band ratios that may be useful in the characterization and classification of fats and oils. These include many of the bands described previously, in particular those indicating the extent and type of unsaturation. However, chemometric techniques have to be applied in order to take full advantage of the spectral information.

Fluorescence interferences may occur in the Raman spectra of fats and oils. Fluorescence will arise from compounds such as chlorophyll and tocopherols often found in plant oils (Sikorska *et al.*, 2004). Furthermore, the presence of highly conjugated double-bond structures will also cause fluorescence. Figure 31 shows the difference in the amount of fluorescence occurring in the Raman spectra of different edible oils. Clearly, the oils of plant origin contain the largest fluorescence background due to the amount of chlorophyll and other colorants, whereas the fish and mineral oils have much less fluorescence background. Fish oil contains some fluorescence due to large amount of conjugated fatty acids.



**Figure 31.** The difference in background fluorescence of different edible oils.

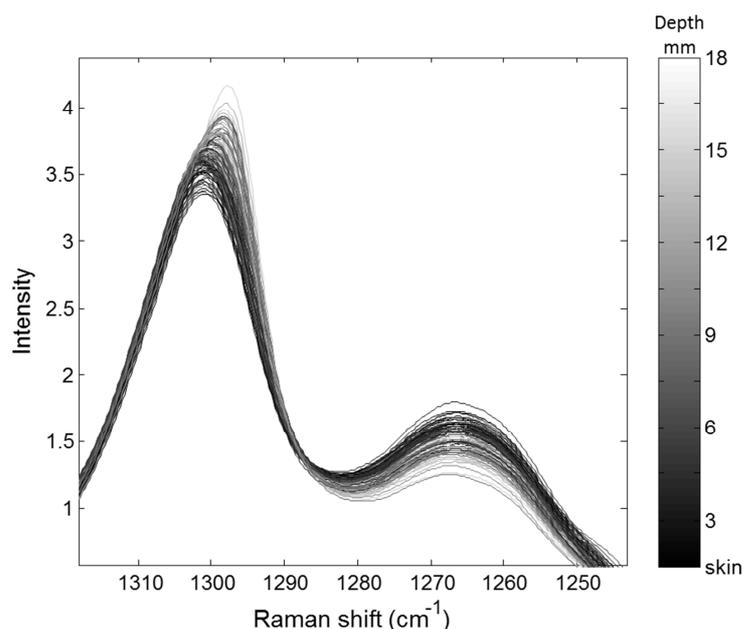
When background fluorescence is removed from the Raman spectra by asymmetric least squares smoothing the differences in distinct Raman bands become clear (Figure 32). The compositional differences between fish oil and vegetable oils are clearly reflected in the spectra. Fish oil has high intensities of bands containing molecular vibrations including double bonds such as C=C stretch  $1655\text{ cm}^{-1}$  and =C-H in-plane rocking  $1266\text{ cm}^{-1}$ . Furthermore, the all-in-phase twisting mode for 4 or more  $-\text{CH}_2-$  in a row at  $1300\text{ cm}^{-1}$  is very weak in the spectrum of fish oil due to only low amounts of saturated fatty acids.



**Figure 32.** Left) Raman spectra  $1800\text{--}800\text{ cm}^{-1}$  of edible oils, and right) table with assignments of major spectral bands.

Looking carefully there is actually a very small band at  $1630\text{ cm}^{-1}$  observable in the fish oil spectrum. This band originates from conjugated double bonds (Chmielarz *et al.*, 1995; Beattie *et al.*, 2004).

The Raman spectra of fats and oils can be affected by the temperature. As previously described (section 4.4), temperature changes can influence the spectral band shape and position. In the measurement of pork fat (Paper I) a change in the spectral band shape and position was clearly observed (Figure 33). The all-in-phase twisting motion of methylene is also affected by the fatty acid chain conformation. The band occurs at  $1305\text{ cm}^{-1}$  but shifts to a lower wavenumber of about  $1296\text{ cm}^{-1}$ .



**Figure 33.** Raman spectra of pork backfat colored according to distance from skin surface. The band at  $1300\text{ cm}^{-1}$  (assigned to all-in-phase twisting mode for 4 or more  $-\text{CH}_2-$  in a row) are systematically changing shape and position as a function of the depth (Paper I).

The pork fat samples analyzed in this study were taken with different distances from the skin to meat and as shown in Figure 33 the band gets wider and shifts to lower wavenumber as a function of depth. This spectral phenomenon has been reported for the first time in Paper I. A possible reason for the shift and shape change of the methylene twisting band is a change in the physical state of the samples. The fat samples taken just below the skin surface contains more unsaturated fatty acid, and therefore have a lower melting point than the more saturated samples taken deeper into the backfat. The temperature during measurement may have reached close to the melting point of the fat from the outer layer (just below the skin); as a consequence the samples may have melted a little and changed from a total solid to a more liquid-like state. Deeper into the backfat the amount of saturated fat is higher and therefore the melting point is also higher (Beattie *et al.*, 2006).

### 6.1.4 Overview of Raman studies of edible fats and oils

In the past decades, there have been numerous studies concerning the characterization, classification, quantification and authentication of edible fats and oils by Raman spectroscopy. Table 11 gives an overview of these studies.

**Table 11.** Overview of application of Raman spectroscopy in analysis of fats and oils

#### Qualitative

Objective	Sample type	Data analysis	Reference
Classification	Oils and fat	PCA	(Baeten <i>et al.</i> , 1998)
	Adipose tissue: beef, lamb, pork and chicken	PLSDA, PCLDA and ANN	(Beattie <i>et al.</i> , 2006)
	Brazilian vegetable oils	PCA, PLS	(Samyn <i>et al.</i> , 2012)
	Animal fats	Single band	(Motoyama <i>et al.</i> , 2010)
Discriminating fat layers	Pork backfat	PCA	(Lyndgaard <i>et al.</i> , 2012/Paper I)
Cis/trans isomer	Oils and fats	Band ratios	(Sadeghi-Jorabchi <i>et al.</i> , 1991)
	Vegetable oils	Band ratios	(Horvat & Bailey, 1971)
Adulteration	Oils	Band ratios	(Weng <i>et al.</i> , 2003)
	Extra virgin olive oil	Bayesian least squares SVM	(Dong <i>et al.</i> , 2012)
Carotenoids	Oils	Visual appearance	(Ferreira <i>et al.</i> , 2013)
Crystal structure	MUFA	Visual appearance	(Sprunt <i>et al.</i> , 2000)
Content of conjugation	Oils	Visual appearance	(Chmielarz <i>et al.</i> , 1995)
	Milk fat	MLR	(Meurens <i>et al.</i> , 2005)

#### Quantitative

Objective	Sample type	Data analysis	Reference
Determination of IV	Oils and margarines	Band ratio	(Sadeghi-Jorabchi <i>et al.</i> , 1990)
	Pork fat	PLS	(Olsen <i>et al.</i> , 2010)
	Pork fat	PLS, band fitting	(Lyndgaard <i>et al.</i> , 2012/Paper I)
	Milk fat	PLS	(El-Abassy <i>et al.</i> , 2012)
	Salmon	PLS	(Afseth <i>et al.</i> , 2006b)
Degree of SFA, MUFA and PUFA	Pork fat	PLS	(Olsen <i>et al.</i> , 2007)
Amount of $\Omega$ 6 and 3 fatty acids	Pork fat	PLS	(Olsen <i>et al.</i> , 2008)
Lipid oxidation	Vegetable oils	Band ratio	(Muik <i>et al.</i> , 2005)

In the determination of total degree of unsaturated fatty acids Raman was early proposed as an efficient method. In 1990 Sadeghi-Jorabchi *et al.* studied oils and margarines by Raman spectroscopy, and found a nice correlation between IV and the band ratios between the total area of C=C stretching band and the CH<sub>2</sub> scissoring band. Since then the band ratio method has been substituted by chemometric methods. Recently, Olsen *et al.* 2010 and Paper I proved that IV in pork adipose tissue can be predicted by Raman spectroscopy with low prediction errors using PLS regression. In further studies on pork adipose tissue Olsen *et al.* showed how the

degree of MUFA and PUFA (2007), and the amount of omega 6 and 3 fatty acids (2008) can be determined. In the determination of total fatty acid composition Beattie *et al.* (2006) found a good correlation between Raman spectra and GC for sample of adipose tissue originating from beef, lamb, pork and chicken. Fatty acid unsaturation of salmons has been studied by Afseth *et al.* (2006). The same authors also quantified fat composition in complex food model systems by Raman and NIR spectroscopy. The texture and nutritional value of hydrogenated vegetable oil and margarines are affected by the *trans* isomer content of fatty acids. *Cis/trans* geometrical isomer ratios have been studied by Raman spectroscopy in vegetable oils (Horvat & Bailey, 1971; Johnson *et al.*, 2002) and margarines (Sadeghi-Jorabchi, 1991). Muik *et al.* (2005) used Raman spectroscopy for direct monitoring of lipid oxidation in edible oils. The sensitivity of Raman spectroscopy in detecting unsaturated related changes have also been used to classify commercial and essential oils and fats according to their content of SFA, MUFA and PUFA (Beaten *et al.*, 1998; Yang & Ying, 2011). In the control of adulteration of extra virgin olive oils Raman spectroscopy combined with chemometrics have also proved promising results (Dong *et al.* 2013).

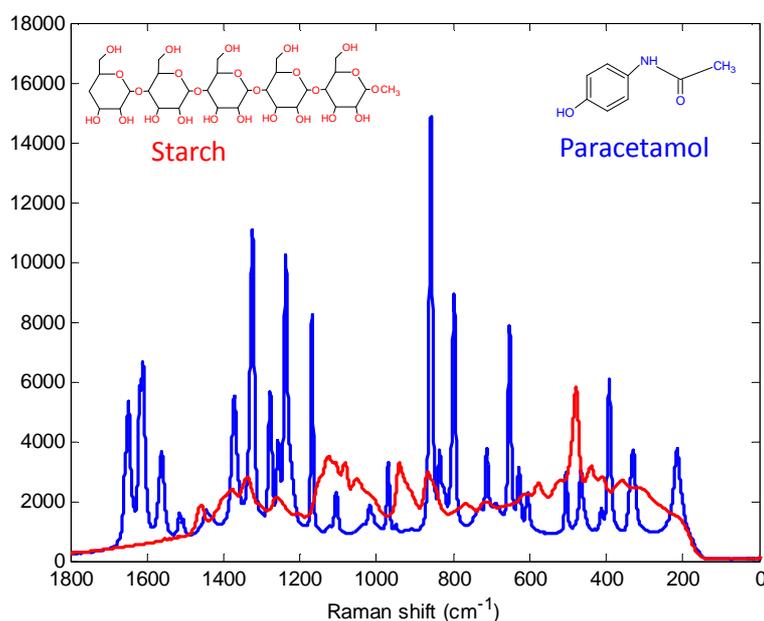
For individual fatty acids the chain length (Beattie *et al.*, 2004) and content of conjugated double bonds (Chmielarz *et al.*, 1995; Meurens *et al.*, 2005) have also been studied by Raman spectroscopy. Raman spectroscopy can also be used for investigating changes in crystal structures. Sprunt *et al.*, (2000) studied the polymorphic behavior of monounsaturated triglycerides found in many fats and oils.

## **6.2 Raman analysis of active pharmaceutical ingredients (API) in different matrices**

### ***6.2.1 Raman scattering properties of API and excipients***

FDA defines API as “any substance or mixture of substances intended to be used in the manufacture of a drug product and that, when used in the production of a drug, becomes an active ingredient in the drug product. Such substances are intended to furnish pharmacological activity or other direct effect in the diagnosis, cure, mitigation, treatment or prevention of disease or to affect the structure and function of the body”. APIs are typically small, organic and crystalline molecules that usually contain multiple nitrogen and oxygen atoms as well as aromatic and heteroaromatic rings. Furthermore, they often hold conjugated domains. The presence of many double bonds, aromatics and conjugation makes APIs strong Raman scatters. Excipients are the pharmacologically inactive compounds formulated together with the API in a medication. They act as fillers, extenders, diluents, wetting agents, solvents, emulsifiers, preservatives, flavors, absorption enhancers, sustained-release matrices, and/or coloring agents. In contrast to APIs many excipients are aliphatic and thus appear only weak in the Raman spectra (Strachan *et al.*, 2007). To serve as a reference in the spectral interpretation of pharmaceutical formulations Raman spectra of forty-three of the most common excipients were provided in a study by de (de Veij

*et al.*, 2007). Figure 34 shows the difference between the Raman spectra of the excipient starch and the API paracetamol.



**Figure 34.** Spectral differences in the Raman spectra of a common pharmaceutical excipient, starch, and an active pharmaceutical ingredient, paracetamol (data from Paper II).

### **6.2.2 Overview of quantitative studies of API in solid dosage formulations**

Quantitative analysis of the API in pharmaceutical products is the most essential part of pharmaceutical analysis. Compared to other spectroscopic methods such as mid- and near-IR, Raman has some clear advantages in the study of APIs in formulations, especially when the concentration is low. First of all, the good scattering properties of APIs resulting in clear distinct bands and low influence from excipients. In addition, Raman is less sensitive to polar bonds, which means that water (a prevalent excipient) can be present in a sample without it having any impact on the Raman spectra. This is not the case for either mid- or near-IR analysis of aqueous environments. The spectra of both of these techniques will exhibit a large water band making the analysis challenging. Furthermore, Raman has also many beneficial practical features which are similar to near-IR such as sampling of very small amounts, no sample preparation, being rapid and non-destructive, and the possibility of bringing the analysis close to production by the use of optical fiber probes. These theoretical and practical advantages of Raman spectroscopy have already been exploited in a range of quantitative studies of different APIs in solid dosage forms.

Table 12 provides an overview of quantitative Raman studies during the last decade.

**Table 12.** Overview of quantitative Raman studies on active pharmaceutical compounds 2002-2013.

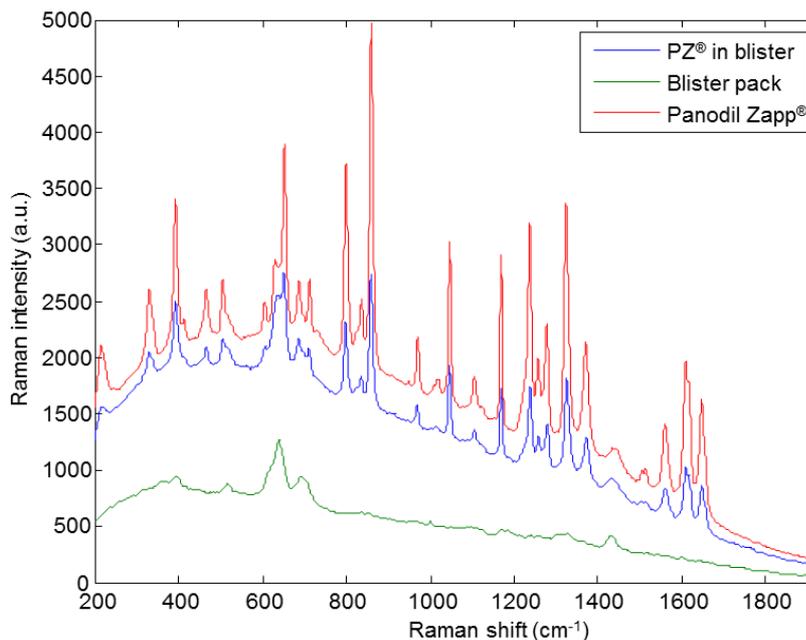
Active substance	Physical form	Laser frequency	Analysis	Reference
Acetaminophen	Tablets	1064 nm	Multivariate	(Komsta <i>et al.</i> , 2011)
Acetylsalicylic acid, acetaminophen	Tablets	1064 nm	Multivariate	(Szostak & Mazurek, 2002)
Acyclovir	Tablets in blister	1064 nm	Uni- and multivariate	(Skoulika & Georgiou, 2003)
Ambroxol	Pellets	785 nm	Multivariate	(Hwang <i>et al.</i> , 2005)
	Tablets	1064 nm	Multivariate	(Szostak & Mazurek, 2004)
	Capsules	785 nm	Multivariate	(Kim <i>et al.</i> , 2007a)
Atorvastatin calcium	Tablets	1064 nm	Multivariate	(Mazurek & Szostak, 2009)
Amphetamine	Powders	830 nm	Multivariate	(Katainen <i>et al.</i> , 2007)
Benzocaine, lidocaine, isoxsuprine, norephedrine	Powders	700 nm	Multivariate	(Fenton <i>et al.</i> , 2011)
Calcitonin	Powder	670 nm	Univariate	(Vehring, 2005)
Captopril and prednisolone	Tablets	1064 nm	Multivariate	(Mazurek & Szostak, 2006)
Carbamazepine	Powders	1064 nm	Uni- and Multivariate	(Strachan <i>et al.</i> , 2004)
chloramphenicol palmitate	Powders	1064 nm and 632.8 nm	Univariate (ratios)	(Gamberini <i>et al.</i> , 2006)
Ciprofloxacin, ciproxin	Powders	785 (handheld)	Multivariate	(Assi <i>et al.</i> , 2012)
Diclofenac sodium	Tablets and capsules	1064 nm	Multivariate	(Mazurek & Szostak, 2008)
Dipyroen	Tablets	1064 nm	Univariate	(Izolani <i>et al.</i> , 2003)
Ecstasy	Tablets	785 nm	Multivariate	(Bell <i>et al.</i> , 2004)
Naproxen	Tablets	785 nm	Multivariate	(Kim <i>et al.</i> , 2006)
Nitrofurantoin	Powder	785 nm	Multivariate	(Rantanen <i>et al.</i> , 2005)
Paracetamol	Tablets	785 nm	Multivariate	(Fransson <i>et al.</i> , 2010)
	Powders in blister	785 nm	Multivariate	(Lyndgaard <i>et al.</i> , 2013/Paper II)
Piracetam	Powders	785 nm	Multivariate	(Croker <i>et al.</i> , 2012)
Prednisone	Tablets	1064 nm	Multivariate	(Mazurek & Szostak, 2012a)
Propranolol HCl	Tablets	785 nm	Multivariate	(Johansson <i>et al.</i> , 2005)
Ranitidine	Tablets	457.9 nm	Multivariate	(Pratiwi <i>et al.</i> , 2002)
	Powders	1064 nm	Multivariate	(Chieng <i>et al.</i> , 2009)
Thiamine hydrochloride	Tablets	1064 nm	Multivariate	(Mazurek & Szostak, 2012b)

The table shows that a large number of different APIs have been quantified in powders, tablets and capsules during the past decade. The group of Mazurek and Szostak has quantified several APIs using FT-Raman in combination with PLS modeling. Dispersive Raman using 785 nm laser excitation has been applied to quantify ambroxol in pellets (Hwang *et al.*, 2005) and capsules (Kim *et al.*, 2007a). Bell *et al.* (2004) studied ecstasy tablets by Raman with a four point sampling approach and found that the concentration of API could be predicted with low errors. In a recent study by Assi *et al.* (2012), a handheld dispersive Raman instrument was tested for rapid and easy identification of ciprofloxacin hydrochloride in street marked samples.

The study showed that even though the instrument provided a satisfactory identity of the constituent in the samples, the quantification ability was limited.

### ***6.2.3 Raman analysis through packaging***

Paper II takes advantage of an additional property of Raman spectroscopy – the ability to analyze materials through their packaging. For safety reasons (i.e. avoid contamination and microbial growth) pharmaceutical products are always stored in some kind of packaging such as plastic or glass containers, blister packs, paper or plastic bags. The ability to evaluate pharmaceutical product composition through concealed packing has a multitude of advantages. In the pharmaceutical industry there is a growing concern regarding the counterfeiting of tablets and medications. Counterfeit drugs can be substandard drug containing too much or too little of an API, or it could be a fake drug containing a wrong API and/or a toxic ingredient. Counterfeiting is found both in so-called “lifestyle” drugs as hormones, steroids, but also in medication for treatment of life-threatening illness such as malaria and HIV/AIDS (Ricci *et al.*, 2007; Newton *et al.*, 2006). Other situations where through-packaging analysis can be advantageous is in controlling deterioration due to damaged packing material, or testing the quality as close to the product release as possible to evaluate potential harm done during product storage. Furthermore, it can be useful for verification of incoming raw materials. Raman and NIR spectroscopy are both methods with high potential in analyzing materials through packaging since non-contact measurements with up to several centimeters distance between the sample and measurement probe are possible. However, whereas NIR is limited in chemical specificity Raman has an advantage of providing clear spectral features that are in general easily assigned to chemical compounds in the pharmaceutical product. This is due to the fact that Raman spectroscopy measures the fundamental vibrations, whereas NIR measures the overtones and combinations of the fundamentals (Pelletier, 1999a). An example of detection of commercial tablets through a blister package by Raman spectroscopy is shown in Figure 35.



**Figure 35.** Through-packaging Raman analysis of the commercial pain-relieving tablet Panodil Zapp®. The green line is Raman spectrum of a blister package, the red line is the Raman spectrum of the pure tablet, and the blue line shows the Raman spectrum of the tablet inside the blister package.

Studies investigating the use of Raman to do non-invasive analysis through packaging are limited, and have mainly been on opened packaging or transparent or semi-transparent packaging. This is primarily because the sampling depth in conventional Raman is often very small (only a few microns into the sample). Thus, conventional Raman will have difficulties in collecting photons scattered from depths beneath the sample surface. However, there have been different approaches to overcome these problems and in the following sections two of these approaches - wide area illumination (WAI) and spatially offset Raman scattering (SORS) - will be described in connection to through-packaging analysis.

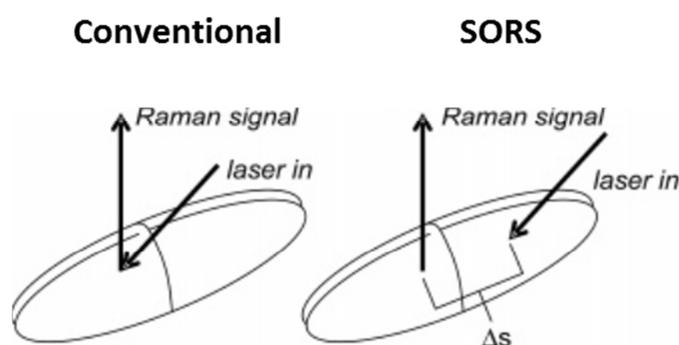
#### Wide-area illumination (WAI)

Wide-area illumination Raman is very suitable for analyzing materials through packaging, as a WAI probe often has a long focal length, which makes it possible to sample materials behind packaging barriers. In WAI probes multiple optical fibers illuminates the samples creating a large illumination laser spot up to 6 mm in diameter. This results in a much more representative sampling compared to single fiber illuminating. WAI Raman have been used for quantification of API in capsules with (Lee *et al.*, 2012) and without colors (Kim *et al.*, 2007a). Furthermore, Kim *et al.* (2007b) measured the concentration of the API, povidone, in commercial eyewash solution through a plastic (low density polyethylene) bottle with WAI. In the study a comparison showed that relative standard deviations of the ratio between the bands originating from the plastic bottle and the bands originating from the API was much lower for WAI compared to conventional results. In Paper II, WAI Raman in

combination with MCR-ALS was applied to quantify paracetamol as part of ternary powder mixtures inside white polyvinyl-chloride (PVC) blister packages.

#### Spatially offset Raman scattering (SORS)

In 2005 Matousek and colleagues discovered SORS by using offset spot illumination. The SORS innovation allow the collection of pure, subsurface Raman spectra from much deeper layers than conventional Raman can analyze (Matousek *et al.*, 2005). In brief, the SORS approach is based on the collection of Raman spectra from regions away from the point of illumination on the sample surface as demonstrated in Figure 36.



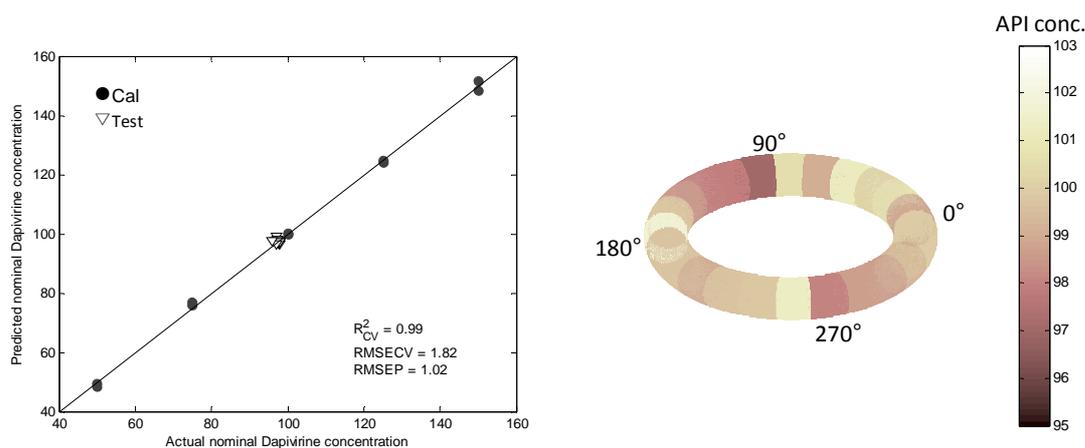
**Figure 36.** Schematic illustration showing the laser illumination onto the sample and the Raman signal collection direction for conventional Raman (left) and spatially offset Raman scattering (right) (modified from Eliasson & Matousek, 2007).

Two Raman spectra are acquired, followed by a scaled subtraction (or multivariate data analysis) to separate the signals of individual layers within the interrogated sample, resulting in the subsurface Raman spectrum. Since the first demonstration of the SORS concept (Matousek *et al.*, 2005), it has been used in different applications. In a study by Eliasson and Matousek (2007) SORS was applied for identifying a range of counterfeit pharmaceutical tablets and capsule through different packaging. More recently, Olds *et al.*, (2011) proved the power of SORS for the detection and identification of concealed substances and drugs such as APIs inside a plastic container and in a capsule in a blister pack. Earlier this year, Bloomfield *et al.*, (2013) demonstrated the use of SORS in raw material verification. Common pharmaceutical raw materials were rapidly identified through a number of frequently used packaging.

#### ***6.2.4 Raman analysis of API in other solid and semisolid dosage formulations***

Beside commonly used solid dosage forms such as tablets, capsules or powders Raman spectroscopy has also been applied to quantify APIs in other solid or semisolid drug delivery systems. Armstrong *et al.* (1996) used FT-Raman to investigate the homogeneity of API distribution in a hormone replacement therapy patch. The mapping profile revealed inhomogeneity which could affect the drug delivery rate and thereby decrease in the overall effectiveness. Dispersive Raman mapping of the

API in the cross section of HIV preventive intra-vaginal rings was investigated by Bell *et al.*, (2007), and in Paper III, a Raman method to determine the API distribution in entire rings was invented. This method may be an important tool to detect inhomogeneity in the API distribution, which may affect the release characteristics and lead to a decreased safety/efficiency, unwanted release-bursts or other side effects. Figure 37 (left) shows a calibration model for the API in vaginal rings showing good correlation between the results from Raman and reference analysis HPLC. The right side of the figure shows a determination of the API distribution for multiple spot on the entire ring circumference allowing for homogeneity testing.



**Figure 37.** The API dapivirine in HIV preventive vaginal rings can successfully be quantified using a novel Raman sampling system and PLS regression, providing a fast and non-destructive alternative to the current standard method HPLC. Dapivirine distribution over the entire ring circumference can be determined for quality assessment. Left) the PLS calibration model, and right) the distribution of API in a vaginal ring (results from Paper III).

Raman spectroscopy has also been successfully applied for determining API in gel-based pharmaceutical preparations. The API in local anaesthetic formulations was investigated before and after exposure to skin (Dennis *et al.*, 2004), and a possible phase-changes in the API was discovered from Raman spectral band broadening. De Beer *et al.* (2007) studied API in ointment for the treatment of warts and hyperkeratosis, and found that quantitative results of FT-Raman corresponded to results based on HPLC.

## 7. Conclusions and perspectives

This thesis has demonstrated the usefulness of multivariate data analysis for handling artefacts and extracting information from non-resonance, vibrational Raman spectroscopic data. This was demonstrated through several practical examples showing the use of suitable preprocessing methods and development of novel curve resolution methods.

Through three feasibility studies, the thesis has shown the practical utility of Raman spectroscopy moving closer towards the use in process analytical applications (on-line measurements directly on complex samples and similar measurement situations). In general, the studies utilize the attractive analytical properties of Raman spectroscopy i.e. high spectral resolution with rich chemical interpretability, no sample preparation, non-invasive, and rapid analysis. Additionally, each study takes advantage of individual benefits of Raman spectroscopy (strong intensity of double bonds; analysis through packaging material; wide-area illumination). For each application, Raman spectroscopy was faced with challenges that were handled in different ways through the use of multivariate data analysis. The following sections contain the conclusions and perspectives, and describe some of the limitations of Raman spectroscopy for each of the three feasibility studies.

### **Depth profiling of porcine adipose tissue**

In lipids, the double bonds between carbons present in unsaturated fatty acids (UFA) are very Raman active. The individual fat layers in the depth profile (from skin to meat) of porcine adipose tissue differ in their amount of UFA. Distinguishing between the fat layers may help improve the quality, because each layer possesses different sensorial and physical properties. The use of Raman spectroscopy for fat layer characterization has so far been unexploited. This study proves that Raman spectroscopy can be applied successfully to discriminate between fat layers in porcine adipose tissue. Based on spectral interpretation, variations in the intensities of the vibrational bands of C=C and =C-H were found to be the main differences between the fat layers. Additionally, the study found correlations between the results of Raman spectra and gas-chromatography for the iodine value (IV), the percentages of saturated, monounsaturated, and polyunsaturated fatty acids. Different spectral phenomena occurred as a function of fat depth. The three band parameters height, location, and width were extracted by Cauchy–Lorentz band fitting. Results showed large differences between the three parameters when correlating them to IV, SFA, MUFA, and PUFA, leading to the conclusion that important information can be found in not only band intensity, but also in band location and width. This research proves that Raman spectroscopy has potential for application in the pork industry.

***Practical utility and limitations of Raman.*** The meat industry is highly competitive, and in order to ensure as good a product quality as possible, the requirements for quality control are increasing. Therefore, rapid ways to measure quality parameters, as shown in this paper, can potentially be of great value providing e.g. the possibility of more precise meat grading. However, there are certain challenges to overcome before this measurement system can be applied in a real production environment (the slaughterhouse). Foremost, a way of measuring the depth profile without having to cut out small slabs must be identified. A solution to this could be to equip the Raman probe with knives, so it can be directly inserted in the carcass. A similar solution has already been invented for NIR probes.

The price of robust Raman instruments is still much higher compared to e.g. robust NIR instruments, which may often influence the choice of NIR over Raman for an application. For prediction of IV in porcine adipose tissue, which was one of the aims in Paper I, another study showed that NIR instrument proved to work similarly well for this purpose. So, if the meat industry is mainly focusing on the IV, and not on more specific information regarding fatty acid composition, they would likely prefer NIR over Raman because of economics. However, if additional information regarding the fatty acid composition is needed to determine and optimize quality, then NIR will not be sufficient and Raman could be a possible solution. In general, to increase the application rate of Raman spectroscopy, the availability of the instrument should be improved by focusing on the developing of reliable and in-expensive instruments. However, low cost Raman instruments are still lacking in one or several factors such as sensitivity, stability, and resolution.

### **Quantification of API through blister package**

In Raman spectroscopy, non-contact measurements with up to several centimeters distance between the sample and the measurement probe are possible. Evaluation of pharmaceutical product composition through packaging material has a multitude of advantages. To the author's knowledge successful quantification of paracetamol through a blister package has never been proven. This study shows that the concentration of the API paracetamol in a ternary mixture can be predicted through a blister package by Raman spectroscopy. Calibration models from two multivariate methods - PLS and MCR-ALS - performed similarly well, when only samples from the same blister batch were taken into account. However, quantification of mixtures was affected when different blister packages materials were used. To overcome this problem, a novel modification of the correlation constraint in MCR-ALS for handling multiple group (multiset) matrix effects was proposed. The results showed that by this new multivariate approach, quantification of paracetamol was possible despite blister package variations. The method may also be suitable for reducing signal interferences in other sample matrices.

***Practical utility and limitations of Raman.*** In the pharmaceutical industry quality control is rigorous, and thorough registration at all stages of production and storage is required. Often the standard procedures for quality control of the end-products involve time-consuming and invasive analysis of grab samples. The application shown in Paper II may provide a rapid and non-invasive alternative to end-product quality testing, showing the possibility of testing the product inside the packaging material. The feasibility study showed some limitations of Raman spectroscopy. The technique is very sensitive to the local environment, and even small changes in blister packaging material affected the signal. This unwanted information regarding variations in the blister package material affected the Raman signal, and made it more challenging to build calibration models for the API. However, in Paper II this problem was solved by developing new ways to apply multivariate modelling methods. Another way to overcome this problem could be to apply other spectroscopic techniques e.g. NIR or IR. Even though NIR has many of the same advantages as Raman (can – potentially – measure through packaging, remote sampling, no requirements for sample preparation) it does not provide the same high resolution and distinct spectral bands for the API in complex mixtures. IR would have a similarly high resolution as Raman; however, it would not have the same possibility for measurement through packaging material and not the same potential for on-line/in-line application in a production, due to limited availability of optical fibers.

### **Quality control of API in vaginal rings**

Recent developments in the Raman instrumentation brought the possibility of using wide-area illumination probes, where large portions of a sample can be analyzed in one measurement. Raman spectroscopy was investigated as an alternative to the current standard method HPLC in quantifying the API dapivirine in HIV preventive silicone vaginal rings. A novel sampling system using wide-area illumination and sample rotation provided rapid and representative measurements for a number of customized reference rings and rings taken from a real production. The results show that PLS regression can predict the dapivirine in vaginal rings with low errors (RMSECV and RMSEP).

***Practical utility and limitations of Raman.*** This feasibility study implies that HPLC as a time and resource demanding standard quality control method can be partly substituted by Raman spectroscopy which is rapid, non-destructive, and requires no sample preparation. Due to the short analysis time (high-throughput), a larger fraction of products can be quality assessed prior to commercial release. This opens the possibilities of intervening faster when irregularities occur, thereby reducing scrap. This study has focused on predicting dapivirine in the final product by Raman spectroscopy. However, as a multivariate spectral method, Raman can also provide

information on other sample quality attributes – a priori known or unknown – such as impurities and physical properties. Furthermore, the flexible sampling of Raman expands the measurement possibilities, so quality can be monitored multiple places during the ring production e.g. raw material verification, the mixing process, and during or right after extrusion. The results of this paper have thus demonstrated the feasibility of using Raman spectroscopy combined with multivariate data analysis as process analytical technology for rapid quality control.

### **General perspectives on Raman spectroscopic applications**

Raman spectroscopy may still suffer a little from its previous reputation as a complex technique only applicable by experts. The currency of Raman spectroscopy is much behind other spectroscopic methods such as UV, NIR, and IR. This lack of knowledge about the technique makes it less preferred, even though it may provide better results/outcomes. A way to promote and draw more attention to the technique could be by stronger communication of successful applications.

#### ***Raman data analysis***

In this thesis, the importance and benefits of using chemometrics in the analysis of Raman data have been illustrated. However, the thesis also emphasizes the importance of the spectral interpretation. It is often found that the traditional approach of using intensities of single spectral bands fails when interferences, overlapping bands, and outliers are present. On the other hand, a strictly multivariate analysis approach purely based on regression, where spectral interpretation is not taken into account, may result in quantification based on chemically meaningless variables. To the author's opinion finding the right balance between multivariate analysis and chemical interpretation is the key to successful Raman data analysis and quantitative modelling.

A second important aspect of Raman data analysis is the use of other spectral band parameters besides the band height. Frequently specific information is found in the width and position of the band, and PLS or other bilinear multivariate methods may not capture this information; band fitting methods should be more widely applied to extract these band parameters.

#### ***Future instrumental trends and options***

One of the major requirements for the successful increase in the number of Raman applications is further developments in the instrumentation. The development of wide-area illumination probes with large laser spot sizes has reduced the sampling uncertainty and opened up for a wide range of new applications particularly in the

pharmaceutical industry, where special solid dosage formulations such as vaginal rings can now be analyzed reliable and representable.

Handheld Raman instruments have been developed to provide on the spot analysis, which can be of great importance for e.g. first responders such as airport security. These small portable instruments may also be used in the food and pharmaceutical industries for raw material verification or in other situations, where unknown powders and liquids need to be non-destructively tested. Even though these instruments may provide rapid measurements with suitable resolutions and SNR for a specific purpose, their measurement quality is not yet as good as the stationary instruments.

Like the trend towards hyphenated techniques in chromatography, there is also a development in combining Raman with other sampling and measurement techniques such as AFM-Raman, DSC-Raman, and SEM-Raman. This reveals the possibility of gaining more selective measurement signals by employing the second (complementary) direction and possibly second-order chemometric methods. Also, combinations of Raman and complementary techniques such as laser-induced breakdown spectroscopy (LIBS) and laser-induced fluorescence (LIF) have increased in recent years.

Instruments that compromise the fluorescence rejection of a FT-Raman spectrometer and detector sensitivity of a dispersive instrument by using an excitation wavelength of 1000 nm have been developed, keeping the advantage of fiber-optical interfaces.



## References

- Adar, F. (2012) Band-Fitting Raman Spectra to Extract Chemical and Physical Information. *Spectroscopy*, 27(11), 12-17.
- Afseth, N.K., V.H. Segtnan and J.P. Wold. (2006a) Raman spectra of biological samples: A study of preprocessing methods. *Applied Spectroscopy*, 60(12), 1358-1367.
- Afseth, N.K., J.P. Wold and V.H. Segtnan. (2006b) The potential of Raman spectroscopy for characterisation of the fatty acid unsaturation of salmon. *Analytica Chimica Acta*, 572(1), 85-92.
- Antunes, M.C., J.E.J. Simao, A.C. Duarte and R. Tauler. (2002) Multivariate curve resolution of overlapping voltammetric peaks: quantitative analysis of binary and quaternary metal mixtures. *Analyst*, 127(6), 809-817.
- Armstrong, G., H. Edwards, D. Farwell and A. Williams. (1996) Fourier transform Raman microscopic study of drug distribution in a transdermal drug delivery device. *Vibrational Spectroscopy*, 11(2), 105-113.
- Assi, S., R.A. Watt and A.C. Moffat. (2012) On the quantification of ciprofloxacin in proprietary Ciproxin tablets and generic ciprofloxacin tablets using handheld Raman spectroscopy. *Journal of Raman Spectroscopy*, 43(8), 1049-1057.
- Baeten, V., P. Hourant, M. Morales and R. Aparicio. (1998) Oil and fat classification by FT-Raman spectroscopy. *Journal of Agricultural and Food Chemistry*, 46(7), 2638-2646.
- Bakeev, K.A. (2010). *Process analytical technology, spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. Chichester, West Sussex: Chichester, West Sussex : Wiley.
- Banwell, C.N. (1983). *Fundamentals of molecular spectroscopy*. London : McGraw-Hill.
- Barnes, R.J., M.S. Dhanoa and S.J. Lister. (1989) Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772-777.
- Beattie, J., S. Bell, C. Borggaard, A. Fearon and B. Moss. (2006) Prediction of adipose tissue composition using Raman spectroscopy: Average properties and individual fatty acids. *Lipids*, 41(3), 287-294.
- Beattie, J., S. Bell and B. Moss. (2004) A critical evaluation of Raman spectroscopy for the analysis of lipids: Fatty acid methyl esters. *Lipids*, 39(5), 407-419.
- Beebe, K.R., R.J. Pell and M.B. Seasholtz. (1998). *Chemometrics, a practical guide*. New York: Wiley.
- Belitz, H.-., W. Grosch and P. Schieberle. (2009). *Food Chemistry*. Berlin: Springer-Verlag.
- Bell, S., J. Beattie, J. McGarvey, K. Peters, N. Sirimuthu and S. Speers. (2004) Development of sampling methods for Raman analysis of solid dosage forms of therapeutic and illicit drugs. *Journal of Raman Spectroscopy*, 35(5), 409-417.
- Bell, S.E.J., A.C. Dennis, L.A. Fido, R.K. Malcolm, N.M.S. Sirimuthu, C.F. Toner and A.D. Woolfson. (2007) Characterization of silicone elastomer vaginal rings containing HIV microbicide TMC120 by Raman spectroscopy. *Journal of Pharmacy and Pharmacology*, 59(2), 203-207.
- Bertarelli, C., A. Bianco, R. Castagna and G. Pariani. (2011) Photochromism into optics: Opportunities to develop light-triggered optical elements. *Journal of Photochemistry and Photobiology C-Photochemistry Reviews*, 12(2), 106-125.

- Bloomfield, M., D. Andrews, P. Loeffen, C. Tombling, T. York and P. Matousek. (2013) Non-invasive identification of incoming raw pharmaceutical materials using Spatially Offset Raman Spectroscopy. *Journal of pharmaceutical and biomedical analysis*, 76, 65-69.
- Boyaci, I.H., H.E. Genis, B. Guven, U. Tamer and N. Alper. (2012) A novel method for quantification of ethanol and methanol in distilled alcoholic beverages using Raman spectroscopy. *Journal of Raman Spectroscopy*, 43(8), 1171-1176.
- Bro, R. (2003) Multivariate calibration - What is in chemometrics for the analytical chemist? *Analytica Chimica Acta*, 500(1-2), 185-194.
- Bro, R. and S. DeJong. (1997) A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, 11(5), 393-401.
- Chieng, N., S. Rehder, D. Saville, T. Rades and J. Aaltonen. (2009) Quantitative solid-state analysis of three solid forms of ranitidine hydrochloride in ternary mixtures using Raman spectroscopy and X-ray powder diffraction. *Journal of pharmaceutical and biomedical analysis*, 49(1), 18-25.
- Chmielarz, B., K. Bajdor, A. Labudzinska and Z. Klukowskamajewska. (1995) Studies on the Double-Bond Positional Isomerization Process in Linseed Oil by Uv, Ir and Raman-Spectroscopy. *Journal of Molecular Structure*, 348, 313-316.
- Crocker, D.M., M.C. Hennigan, A. Maher, Y. Hu, A.G. Ryder and B.K. Hodnett. (2012) A comparative study of the use of powder X-ray diffraction, Raman and near infrared spectroscopy for quantification of binary polymorphic mixtures of piracetam. *Journal of pharmaceutical and biomedical analysis*, 63, 80-86.
- Dahlenborg, H., A. Millqvist-Fureby, B.D. Brandner and B. Bergenstahl. (2012) Study of the porous structure of white chocolate by confocal Raman microscopy. *European Journal of Lipid Science and Technology*, 114(8), 919-926.
- De Beer, T.R.M., W.G. Baeyens, Y. Vander Heyden, J.P. Remon, C. Vervaet and F. Verpoort. (2007) Influence of particle size on the quantitative determination of salicylic acid in a pharmaceutical ointment using FT-Raman spectroscopy. *European Journal of Pharmaceutical Sciences*, 30(3-4), 229-235.
- de Veij, M., P. Vandenabeele, K.A. Hall, F.M. Fernandez, M.D. Green, N.J. White, A.M. Dondorp, P.N. Newton and L. Moens. (2007) Fast detection and identification of counterfeit antimalarial tablets by Raman spectroscopy RID B-7015-2008. *Journal of Raman Spectroscopy*, 38(2), 181-187.
- Delfino, I., C. Camerlingo, M. Portaccio, B. Della Ventura, L. Mita, D.G. Mita and M. Lepore. (2011) Visible micro-Raman spectroscopy for determining glucose content in beverage industry. *Food Chemistry*, 127(2), 735-742.
- Dennis, A., J. McGarvey, A. Woolfson, D. McCafferty and G. Moss. (2004) A Raman spectroscopic investigation of bioadhesive tetracaine local anaesthetic formulations. *International journal of pharmaceuticals*, 279(1-2), 43-50.
- Di Rocco, H.O. and A. Cruzado. (2012) The Voigt Profile as a Sum of a Gaussian and a Lorentzian Functions, when the Weight Coefficient Depends Only on the Widths Ratio. *Acta Physica Polonica a*, 122(4), 666-669.
- Dobson, G. (2001) Spectroscopy and spectrometry of lipids - Part 1 - Lipid analysis by vibrational spectroscopy. *European Journal of Lipid Science and Technology*, 103(12), 815-826.
- Dong, W., Y. Zhang, B. Zhang and X. Wang. (2012) Quantitative analysis of adulteration of extra virgin olive oil using Raman spectroscopy improved by Bayesian framework least squares support vector machines. *Analytical Methods*, 4(9), 2772-2777.
- Dyrby, M., R. Petersen, J. Larsen, B. Rudolf, L. Norgaard and S. Engelsen. (2004) Towards on-line monitoring of the composition of commercial carrageenan powders. *Carbohydrate Polymers*, 57(3), 337-348.

- Eilers, P. (2004) Parametric time warping. *Analytical Chemistry*, 76(2), 404-411.
- El-Abassy, R.M., P.J. Eravuchira, P. Donfack, B. von der Kammer and A. Materny. (2012) Direct Determination of Unsaturation Level of Milk Fat Using Raman Spectroscopy. *Applied Spectroscopy*, 66(5), 538-544.
- Eliasson, C. and P. Matousek. (2007) Noninvasive authentication of pharmaceutical products through packaging using spatially offset Raman spectroscopy. *Analytical Chemistry*, 79(4), 1696-1701.
- Fairchild, S.Z., C.F. Bradshaw, W. Su and S.K. Guharay. (2009) Predicting Raman Spectra Using Density Functional Theory. *Applied Spectroscopy*, 63(7), 733-741.
- FDA. (2004) "Guidance for industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance". U.S. Food and Drug Administration, Rockville, USA
- Fenton, O.S., L.A. Tonge, T.H. Moot and K.A. Frederick. (2011) Quantitative Analysis of Simulated Illicit Street-Drug Samples Using Raman Spectroscopy and Partial Least Squares Regression. *Spectroscopy Letters*, 44(4), 229-234.
- Ferraro, J., K. Nakamoto and C.W. Brown. (2003). *Introductory Raman spectroscopy*. Amsterdam Boston: Amsterdam : Boston : Academic Press.
- Ferreira, B.S., C.G. de Almeida, M. Le Hyaric, V.E. de Oliveira, H.G.M. Edwards and L.F.C. de Oliveira. (2013) Raman Spectroscopic Investigation of Carotenoids in Oils from Amazonian Products. *Spectroscopy Letters*, 46(2), 122-127.
- Fransson, M., J. Johansson, A. Sparen and O. Svensson. (2010) Comparison of multivariate methods for quantitative determination with transmission Raman spectroscopy in pharmaceutical formulations. *Journal of Chemometrics*, 24(11-12), 674-680.
- Gamberini, M., C. Baraldi, A. Tinti, C. Rustichelli, V. Ferioli and G. Gamberini. (2006) Solid state characterization of chloramphenicol palmitate. Raman spectroscopy applied to pharmaceutical polymorphs. *Journal of Molecular Structure*, 785(1-3), 216-224.
- Geladi, P. and B. Kowalski. (1986) Partial Least-Squares Regression - a Tutorial. *Analytica Chimica Acta*, 185, 1-17.
- Geladi, P., D. Macdougall and H. Martens. (1985) Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, 39(3), 491-500.
- Horvat, R. and G. Bailey. (1971) Estimation of Cis/trans Isomer Content of Edible Vegetable Oils by Raman Spectroscopy. *Journal of the American Oil Chemists Society*, 48(2), A103-&.
- Hwang, M., S. Cho, H. Chung and Y. Woo. (2005) Nondestructive determination of the ambroxol content in tablets by Raman spectroscopy. *Journal of pharmaceutical and biomedical analysis*, 38(2), 210-215.
- Izolani, A., M. de Moraes and C. Tellez. (2003) Fourier transform Raman spectroscopy of drugs: quantitative analysis of 1-phenyl-2,3-dimethyl-5-pyrazolone-4-methylaminomethane sodium sulfonate: (dipyron). *Journal of Raman Spectroscopy*, 34(10), 837-843.
- Johansson, J., S. Pettersson and S. Folestad. (2005) Characterization of different laser irradiation methods for quantitative Raman tablet assessment. *Journal of pharmaceutical and biomedical analysis*, 39(3-4), 510-516.
- Johnson, G., R. Machado, K. Freidl, M. Achenbach, P. Clark and S. Reidy. (2002) Evaluation of Raman spectroscopy for determining cis and trans isomers in partially hydrogenated soybean oil. *Organic Process Research & Development*, 6(5), 637-644.

Katainen, E., M. Elomaa, U. Laakkonen, E. Sippola, P. Niemela, J. Suhonen and K. Jarvinen. (2007) Quantification of the amphetamine content in seized street samples by Raman spectroscopy. *Journal of forensic sciences*, 52(1), 88-92.

Keresztury, G. (2002). Raman Spectroscopy: Theory. In J.M. Chalmers and P.R. Griffiths (Eds.), *Handbook of vibrational spectroscopy*, Chichester: Chichester : Wiley, 71-87.

Killeen, D.P., C.E. Sansom, R.E. Lill, J.R. Eason, K.C. Gordon and N.B. Perry. (2013) Quantitative Raman Spectroscopy for the Analysis of Carrot Bioactives. *Journal of Agricultural and Food Chemistry*, 61(11), 2701-2708.

Kim, J., J. Noh, H. Chung, Y. Woo, M.S. Kemper and Y. Lee. (2007a) Direct, non-destructive quantitative measurement of an active pharmaceutical ingredient in an intact capsule formulation using Raman spectroscopy. *Analytica Chimica Acta*, 598(2), 280-285.

Kim, M., H. Chung, Y. Woo and M. Kemper. (2006) New reliable Raman collection system using the wide area illumination (WAI) scheme combined with the synchronous intensity correction standard for the analysis of pharmaceutical tablets. *Analytica Chimica Acta*, 579(2), 209-216.

Kim, M., H. Chung, Y. Woo and M.S. Kemper. (2007b) A new non-invasive, quantitative Raman technique for the determination of an active ingredient in pharmaceutical liquids by direct measurement through a plastic bottle. *Analytica Chimica Acta*, 587(2), 200-207.

Komsta, L., H. Czarnik-Matusewicz, R. Szostak, A. Gumieniczek, R. Pietras, R. Skibinski and T. Inglot. (2011) Chemometric Detection of Acetaminophen in Pharmaceuticals by Infrared Spectroscopy Combined with Pattern Recognition Techniques: Comparison of Attenuated Total Reflectance-FTIR and Raman Spectroscopy. *Journal of AOAC International*, 94(3), 743-749.

Lee, Y., J. Kim, S. Lee, Y. Woo and H. Chung. (2012) Simple transmission Raman measurements using a single multivariate model for analysis of pharmaceutical samples contained in capsules of different colors. *Talanta*, 89, 109-116.

Lewis, I.R. and H. Edwards. (2001). *Handbook of Raman spectroscopy, from the research laboratory to the process line*. New York: New York : Marcel Dekker.

LiChan, E. (1996) The applications of Raman spectroscopy in food science. *Trends in Food Science & Technology*, 7(11), 361-370.

Long, A.D. (2002). *The Raman effect, a unified treatment of the theory of Raman scattering by molecules*. Chichester New York; Chichester: Wiley.

Lyndgaard, L.B., F. van den Berg and A. de Juan. (2013) Quantification of paracetamol through tablet blister packages by Raman spectroscopy and multivariate curve resolution-alternating least squares. *Chemometrics and Intelligent Laboratory Systems*, 125, 58-66.

Lyndgaard, L.B., K.M. Sorensen, F. van den Berg and S.B. Engelsen. (2012) Depth profiling of porcine adipose tissue by Raman spectroscopy. *Journal of Raman Spectroscopy*, 43(4), 482-489.

Malinowski, E.R. (1992) Window Factor-Analysis - Theoretical Derivation and Application to Flow-Injection Analysis Data. *Journal of Chemometrics*, 6(1), 29-40.

Manne, R. (1995) On the Resolution Problem in Hyphenated Chromatography. *Chemometrics and Intelligent Laboratory Systems*, 27(1), 89-94.

Martens, H. and E. Stark. (1991) Extended Multiplicative Signal Correction and Spectral Interference Subtraction - New Preprocessing Methods for Near-Infrared Spectroscopy. *Journal of pharmaceutical and biomedical analysis*, 9(8), 625-635.

- Matousek, P., M. Morris, N. Everall, I. Clark, M. Towrie, E. Draper, A. Goodship and A. Parker. (2005) Numerical simulations of subsurface probing in diffusely scattering media using spatially offset Raman spectroscopy. *Applied Spectroscopy*, 59(12), 1485-1492.
- Mazurek, S. and R. Szostak. (2006) Quantitative determination of captopril and prednisolone in tablets by FT-Raman spectroscopy. *Journal of pharmaceutical and biomedical analysis*, 40(5), 1225-1230.
- Mazurek, S. and R. Szostak. (2009) Quantification of atorvastatin calcium in tablets by FT-Raman spectroscopy. *Journal of pharmaceutical and biomedical analysis*, 49(1), 168-172.
- Mazurek, S. and R. Szostak. (2012b) Quantitative analysis of thiamine hydrochloride in tablets-Comparison of infrared attenuated total reflection, diffuse reflectance infrared and Raman spectroscopy. *Vibrational Spectroscopy*, 62, 10-16.
- Mazurek, S. and R. Szostak. (2008) Quantitative determination of diclofenac sodium in solid dosage forms by FT-Raman spectroscopy. *Journal of pharmaceutical and biomedical analysis*, 48(3), 814-821.
- Mazurek, S. and R. Szostak. (2012a) Quantitative Determination of Prednisone in Tablets by Infrared Attenuated Total Reflection and Raman Spectroscopy. *Journal of AOAC International*, 95(3), 744-750.
- McCreery, R.L. (2000). *Raman spectroscopy for chemical analysis*. New York : John Wiley & Sons.
- Meurens, M., V. Baeten, S. Yan, E. Mignolet and Y. Larondelle. (2005) Determination of the conjugated linoleic acids in cow's milk fat by Fourier transform Raman spectroscopy. *Journal of Agricultural and Food Chemistry*, 53(15), 5831-5835.
- Micklander, E., L. Brimer and S. Engelsen. (2002) Noninvasive assay for cyanogenic constituents in plants by Raman spectroscopy: Content and distribution of amygdalin in bitter almond (*Prunus amygdalas*). *Applied Spectroscopy*, 56(9), 1139-1146.
- Miller, C.E. (2001). Chemical principles of Near-infrared technology. In P. Williams and K. Norris (Eds.), *Near-infrared technology, in the agricultural and food industries*, St. Paul, Minn: St. Paul, Minn : American Association of Cereal Chemists.
- Motoyama, M., M. Ando, K. Sasaki and H. Hamaguchi. (2010) Differentiation of Animal Fats from Different Origins: Use of Polymorphic Features Detected by Raman Spectroscopy. *Applied Spectroscopy*, 64(11), 1244-1250.
- Muik, B., B. Lendl, A. Molina-Diaz and M. Ayora-Canada. (2005) Direct monitoring of lipid oxidation in edible oils by Fourier transform Raman spectroscopy. *Chemistry and physics of lipids*, 134(2), 173-182.
- Nagpal, R., H. Yadav, A.K. Puniya, K. Singh, S. Jain and F. Marotta. (2007) Conjugated linoleic acid: Sources, synthesis and potential health benefits - An overview. *Current Topics in Nutraceutical Research*, 5(2-3), 55-65.
- Newton, P.N., M.D. Green, F.M. Fernandez, N.P.J. Day and N.J. White. (2006) Counterfeit anti-infective drugs. *Lancet Infectious Diseases*, 6(9), 602-613.
- Okazaki, S., M. Hiramatsu, K. Gonmori, O. Suzuki and A.T. Tu. (2009) Rapid nondestructive screening for melamine in dried milk by Raman spectroscopy. *Forensic Toxicology*, 27(2), 94-97.
- Olds, W.J., E. Jaatinen, P. Fredericks, B. Cletus, H. Panayiotou and E.L. Izake. (2011) Spatially offset Raman spectroscopy (SORS) for the analysis and detection of packaged pharmaceuticals and concealed drugs. *Forensic science international*, 212(1-3), 69-77.
- Olsen, E.F., C. Baustad, B. Egelanddal, E. Rukke and T. Isaksson. (2010) Long-term stability of a Raman instrument determining iodine value in pork adipose tissue. *Meat Science*, 85(1), 1-6.

- Olsen, E.F., E. Rukke, B. Egeland and T. Isaksson. (2008) Determination of omega-6 and omega-3 fatty acids in pork adipose tissue with nondestructive Raman and Fourier transform infrared spectroscopy. *Applied Spectroscopy*, 62(9), 968-974.
- Olsen, E.F., E. Rukke, A. Flatten and T. Isaksson. (2007) Quantitative determination of saturated-, monounsaturated- and polyunsaturated fatty acids in pork adipose tissue with non-destructive Raman spectroscopy. *Meat Science*, 76(4), 628-634.
- Ouellette, R.J. (1998). *Organic chemistry, a brief introduction*. Upper Saddle River, N.J: Upper Saddle River, N.J. : Prentice Hall.
- Ozaki, Y., R. Cho, K. Ikegaya, S. Muraishi and K. Kawauchi. (1992) Potential of Near-Infrared Fourier-Transform Raman-Spectroscopy in Food Analysis. *Applied Spectroscopy*, 46(10), 1503-1507.
- Ozbalci, B., I.H. Boyaci, A. Topcu, C. Kadilar and U. Tamer. (2013) Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chemistry*, 136(3-4), 1444-1452.
- Pelletier, M.J. (1999a). *Analytical applications of Raman spectroscopy*. Oxford : Blackwell Science.
- Pelletier, M. (1999b) Effects of temperature on cyclohexane Raman bands. *Applied Spectroscopy*, 53(9), 1087-1096.
- Pratiwi, D., J. Fawcett, K. Gordon and T. Rades. (2002) Quantitative analysis of polymorphic mixtures of ranitidine hydrochloride by Raman spectroscopy and principal components analysis. *European Journal of Pharmaceutics and Biopharmaceutics*, 54(3), 337-341.
- Rantanen, J., H. Wikstrom, F. Rhea and L. Taylor. (2005) Improved understanding of factors contributing to quantification of anhydrate/hydrate powder mixtures. *Applied Spectroscopy*, 59(7), 942-951.
- Ricci, C., C. Eliasson, N.A. Macleod, P.N. Newton, P. Matousek and S.G. Kazarian. (2007) Characterization of genuine and fake artesunate anti-malarial tablets using Fourier transform infrared imaging and spatially offset Raman spectroscopy through blister packs. *Analytical and Bioanalytical Chemistry*, 389(5), 1525-1532.
- Rinnan, A., F. van den Berg and S.B. Engelsen. (2009) Review of the most common pre-processing techniques for near-infrared spectra. *Trac-Trends in Analytical Chemistry*, 28(10), 1201-1222.
- Sadeghi-Jorabchi, H., P. Hendra, R. Wilson and P. Belton. (1990) Determination of the Total Unsaturation in Oils and Margarines by Fourier-Transform Raman-Spectroscopy. *Journal of the American Oil Chemists Society*, 67(8), 483-486.
- Sadeghi-Jorabchi, H., R. Wilson, P. Belton, J. Edwardswebb and D. Coxon. (1991) Quantitative-Analysis of Oils and Fats by Fourier-Transform Raman-Spectroscopy. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, 47(9-10), 1449-1458.
- Samyn, P., D. Van Nieuwkerke, G. Schoukens, L. Vonck, D. Stanssens and H. Van den Abbeele. (2012) Quality and Statistical Classification of Brazilian Vegetable Oils Using Mid-Infrared and Raman Spectroscopy. *Applied Spectroscopy*, 66(5), 552-565.
- Sasic, S. and Y. Ozaki. (2007). Introduction to Raman Spectroscopy. In S. Sasic (Ed.), *Pharmaceutical applications of raman spectroscopy*, New Jersey: New Jersey : Wiley-Interscience.
- Savitsky, A. and M. Golay. (1964) Smoothing + Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627-1637.
- Shaver, J.M. (2001). Chemometrics for Raman Spectroscopy. In: I.R. Lewis and H. Edwards (Eds.), *Handbook of Raman spectroscopy, from the research laboratory to the process line*, New York: Marcel Dekker.

Sikorska, E., A. Romaniuk, I. Khmelinskii, R. Herance, J. Bourdelande, M. Sikorski and J. Koziol. (2004) Characterization of edible oils using total luminescence spectroscopy. *Journal of Fluorescence*, 14(1), 25-35.

Sjöström, M., S. Wold, W. Lindberg, J. Persson and H. Martens. (1983) A Multivariate Calibration-Problem in Analytical-Chemistry Solved by Partial Least-Squares Models in Latent-Variables. *Analytica Chimica Acta*, 150(1), 61-70.

Skoulika, S. and C. Georgiou. (2003) Rapid, noninvasive quantitative determination of acyclovir in pharmaceutical solid dosage forms through their poly(vinyl chloride) blister package by solid-state Fourier transform Raman spectroscopy. *Applied Spectroscopy*, 57(4), 407-412.

Smith, E. and G. Dent. (2005). *Modern Raman spectroscopy, a practical approach*. New York; Hoboken, NJ: Wiley.

Soderholm, S., Y. Roos, N. Meinander and M. Hotokka. (1999) Raman spectra of fructose and glucose in the amorphous and crystalline states. *Journal of Raman Spectroscopy*, 30(11), 1009-1018.

Sørensen, K.M., H. Petersen and S.B. Engelsen. (2012) An On-Line Near-Infrared (NIR) Transmission Method for Determining Depth Profiles of Fatty Acid Composition and Iodine Value in Porcine Adipose Fat Tissue. *Applied Spectroscopy*, 66(2), 218-226.

Sprunt, J., U. Jayasooriya and R. Wilson. (2000) A simultaneous FT-Raman-DSC (SRD) study of polymorphism in sn-1,3-distearoyl-2-oleoylglycerol (SOS). *Physical Chemistry Chemical Physics*, 2(19), 4299-4305.

Strachan, C.J., T. Rades, K.C. Gordon and J. Rantanen. (2007) Raman spectroscopy for quantitative analysis of pharmaceutical solids. *Journal of Pharmacy and Pharmacology*, 59(2), 179-192.

Swierenga, H., A. De Weijer and L. Buydens. (1999) Robust calibration model for on-line and off-line prediction of poly(ethylene terephthalate) yarn shrinkage by Raman spectroscopy. *Journal of Chemometrics*, 13(3-4), 237-249.

Szostak, R. and S. Mazurek. (2004) FT-Raman quantitative determination of ambroxol in tablets. *Journal of Molecular Structure*, 704(1-3), 229-233.

Szostak, R. and S. Mazurek. (2002) Quantitative determination of acetylsalicylic acid and acetaminophen in tablets by FT-Raman spectroscopy. *Analyst*, 127(1), 144-148.

Tauler, R. (1995) Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30(1), 133-146.

Tauler, R. and D. Barcelo. (1993) Multivariate Curve Resolution Applied to Liquid-Chromatography Diode-Array Detection. *Trac-Trends in Analytical Chemistry*, 12(8), 319-327.

Tauler, R., A. Smilde and B. Kowalski. (1995) Selectivity, Local Rank, 3-Way Data-Analysis and Ambiguity in Multivariate Curve Resolution. *Journal of Chemometrics*, 9(1), 31-58.

Vehring, R. (2005) Red-excitation dispersive Raman spectroscopy is a suitable technique for solid-state analysis of respirable pharmaceutical powders. *Applied Spectroscopy*, 59(3), 286-292.

Wang Wei-Wei, Li Liang, Li Zhan-Long, Sun Cheng-Lin, Ouyang Shun-Li, Xu Da-Peng, Qu Guan-Nan, Li Zuo-Wei, Gao Shu-Qin, Cao Jun-Sheng, Gao Feng-Li and Zhou Qiang. (2010) Effect of Temperature on the Raman Spectrum of Short-chain Polyenes Biological Molecules beta-Carotene. *Chemical Journal of Chinese Universities-Chinese*, 31(9), 1864-1867.

Wartewig, S. (2003). *IR and Raman spectroscopy, fundamental processing*. Weinheim Cambridge: Weinheim : Wiley-VCH.

- Weng, Y., R. Weng, C. Tzeng and W. Chen. (2003) Structural analysis of triacylglycerols and edible oils by near-infrared Fourier transform Raman spectroscopy. *Applied Spectroscopy*, 57(4), 413-418.
- Windig, W. and J. Guilment. (1991) Interactive Self-Modeling Mixture Analysis. *Analytical Chemistry*, 63(14), 1425-1432.
- Wold, J., B. Marquardt, B. Dable, D. Robb and B. Hatlen. (2004) Rapid quantification of carotenoids and fat in Atlantic salmon (*Salmo salar* L.) by Raman spectroscopy and chemometrics. *Applied Spectroscopy*, 58(4), 395-403.
- Wold, S., K. Esbensen and P. Geladi. (1987) Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37-52.
- Wold, S., M. Sjostrom and L. Eriksson. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109-130.
- Wong, M. (1994) Polarizability and Tensor Ellipsoid in the Raman Effect. *Vibrational Spectroscopy*, 7(2), 197-199.
- Wu, W., D. Massart and S. deJong. (1997) The kernel PCA algorithms for wide data .1. Theory and algorithms. *Chemometrics and Intelligent Laboratory Systems*, 36(2), 165-172.
- Yang, D. and Y. Ying. (2011) Applications of Raman Spectroscopy in Agricultural Products and Food Analysis: A Review. *Applied Spectroscopy Reviews*, 46(7), 539-560.

# PAPER I

---

Lyndgaard, L.B., Sørensen, K.M., van den Berg, F. & Engelsen, S.B.

Depth profiling of porcine adipose tissue by Raman spectroscopy

*Journal of Raman Spectroscopy*, 43, 482–489, 2012

# Depth profiling of porcine adipose tissue by Raman spectroscopy

Lotte Bøge Lyndgaard,<sup>a,\*</sup> Klavs Martin Sørensen,<sup>a,b</sup> Frans van den Berg<sup>a</sup> and Søren Balling Engelsen<sup>a</sup>

Raman spectroscopy was applied on a depth profile of porcine adipose tissue (from skin to meat) with the purpose of (1) discriminating between fat layers and (2) estimating the variation in fatty acid composition as a function of fat depth and fat layer: total degree of unsaturation (iodine value), fractions of saturated, and monounsaturated and polyunsaturated fatty acids. The thickness and composition of the outer layer of porcine adipose tissue influences the final quality of backfat. A too-thick outer layer is associated with problems such as oily appearance, rancidity development, and difficulties in separating muscle and adipose tissue when cutting. From principal component analysis on standard normal variate preprocessed Raman spectra (1800–800 cm<sup>-1</sup>), it was possible to discriminate between the outer and the inner backfat layer. Principal component analysis loadings showed that the separation of layer was mainly explained by variation in the bands originating from vibration of double bond C=C stretching plus C–H twisting and rocking. In the prediction of iodine value a three-component partial least squares regression model based on full range Raman spectra showed a root mean square error of cross validation of 2.00 and  $R^2 = 0.69$ . Applying Cauchy–Lorentz band fitting proved that information regarding fat unsaturation was found not only in band intensity, but also in band parameters such as location and width. The results suggest Raman spectroscopy as a potential measurement technique for rapid grading of pork carcasses. Copyright © 2011 John Wiley & Sons, Ltd.

**Keywords:** Raman spectroscopy; fat layers; iodine value (IV); depth profiling; chemometrics

## Introduction

Pork is one of the most important meat products worldwide. The backfat of porcine is very well studied because fat tissue is known to be an important aspect of pork carcass quality. Backfat consists primarily of triglycerides containing saturated, unsaturated, and polyunsaturated fatty acid (SFA, UFA, and PUFA), and the composition of these fatty acids is important in both meat processing and consumer acceptability.<sup>[1–3]</sup> From a health perspective it is preferred to have SFAs replaced with PUFAs as the evidence from epidemiologic, clinical, and mechanistic studies is consistent in finding that this will reduce the risk of coronary heart diseases.<sup>[4]</sup> However, from a production perspective fat containing mostly UFA causes problems with slicing yield and results in an unwanted oily product appearance when packaged. Furthermore, tendency to develop rancidity is increased with UFA content.<sup>[5–7]</sup> Consequently, Hugo and Roodt<sup>[3]</sup> concluded that saturated fat is necessary to manufacture high quality products, despite the fact that it is associated with inferior health properties.

Porcine backfat becomes separated into two different layers (outer and inner) with increased age and fat content, and it has been shown that structure, chemical composition, and texture are different in the two layers.<sup>[8–12]</sup> The outer layer (next to the skin) has a high degree of UFA whereas the inner layer (muscle side) consists of more SFA.<sup>[13]</sup> The degree of unsaturation is very important for the firmness of fat, as the double bonds, only present in UFA, give bends in the methylene chain leading to a less structured and more fluid fat with lower melting point.<sup>[14]</sup> SFAs have straight chains and can pack closer together, which has a positive influence on fat firmness and consistency.

The thickness of the backfat layers is an important quality parameter in porcine meat production. Measurements of backfat thickness can be a decision tool for optimizing breeding and conditions for growth, longevity in gilts, and for quality control and for carcass grading.<sup>[15]</sup> Currently, it is general practice to measure the total backfat thickness, and not the thickness of the individual layers. However, studies have shown that knowledge about the depth of the individual fat layers may help improve the quality.<sup>[16]</sup>

Measurements of fat layers have been carried out with ultrasound and ultrasound imaging.<sup>[15,17]</sup> However, no chemical information about the fat layers can be withdrawn from ultrasound measurements. The chemical composition of each layer is potentially critical as it affects sliceability, visual appearance, and the oxidative stability.<sup>[7]</sup> Thus, although ultrasound can discriminate between fat layers, it does not reveal variations in composition of the individual layers. To the authors knowledge a depth profiling of the composition of fatty acids in backfat has not yet been proposed.

The standard method for determining the fat quality (firmness) is iodine value (IV), which measures the degree of unsaturation through the total number of double bonds in the fatty acids.

\* Correspondence to: Lotte Bøge Lyndgaard, University of Copenhagen, Faculty of Life Sciences, Department of Food Science, Quality and Technology, Frederiksberg C, Denmark. E-mail: Lottebs@life.ku.dk

<sup>a</sup> University of Copenhagen, Faculty of Life Sciences, Department of Food Science, Quality and Technology, Frederiksberg C, Denmark

<sup>b</sup> Carometec A/S, Transformervej 9, Herlev, Denmark

Previous studies have found higher IV, i.e., higher levels of UFA and lower levels of SFA, for the outer layer compared with the inner layer.<sup>[13]</sup> Monziols *et al.*<sup>[18]</sup> found that IV decreases with a negative gradient from outside inwards, and feed and race have been shown to affect the degree of unsaturation in pork fat. IV can be found by the direct chemical method, but usually it is calculated from gas chromatography (GC). GC fatty acid determination is a relatively time-consuming analysis and thus, fast analytical methods such as near infrared transmission (NIT) spectroscopy have been investigated as fast and uncomplicated alternatives, and recently a new NIT based instrument with hand-held probe (NITFOM) has showed very promising results for online carcass grading (*results not published*).

This study investigates Raman spectroscopy for determining fatty acid composition of pork backfat as a function of the depth. Raman spectroscopy has during the last years gained increased interest as a potential quantitative fast measurement method in the food and pharmaceutical industries. Raman spectroscopy determines the change in polarizability during a molecular vibration. For molecules with multiple bonds (e.g. double bonds) there is a high electron density between the atoms. Therefore, the Raman signal for the stretching vibration of such functional groups becomes strong. In lipids the double bonds between carbons present in UFA are very Raman active and for this reason Raman was earlier proposed as an efficient method for determination of IV in oils and margarines,<sup>[19]</sup> and recently, Olsen *et al.*<sup>[20]</sup> proved that IV in pork fat can be predicted by Raman spectroscopy with low prediction errors.

In this study, we investigate the performance of Raman spectroscopy to characterize fatty acid composition of pork backfat with emphasis on estimating the IV as a function of fat depth and fat layer.

## Materials and Methods

### Sample preparation

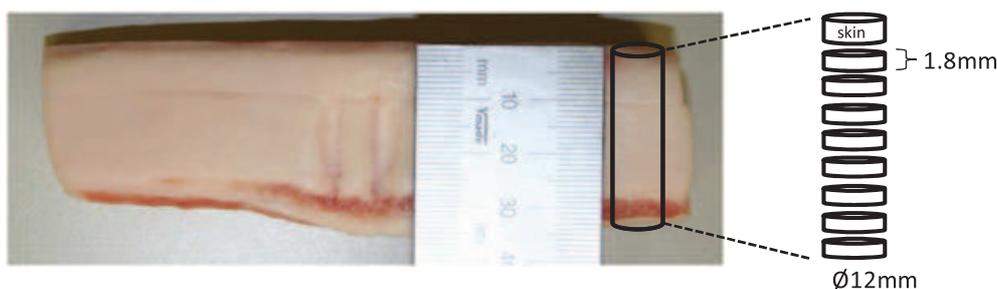
The samples for this study were 16 pork carcasses taken from the daily production stock of a slaughterhouse (Tönnies Fleischwerk, Rheda, Germany). The carcasses were preselected based on the feeding regimes. Variability in the fatty acid composition was ensured because the carcasses were selected from two extreme feeding methods and a normal center group.<sup>[21,22]</sup> Carcasses were selected to have a minimum loin fat thickness of 16 mm (to facilitate proper depth description), but no further selection restrictions were set on lean meat content or carcass weight.

Carcasses were slaughtered within a time span of 30 min. Following the normal processing regime, the carcasses were

slaughtered and left in the slaughterhouse buffer cooler. After 24 h the carcasses were marked for easy identification, and after primary separation, the loin of each carcass was taken off the cut floor leaving the muscle, fat and skin intact on each loin cut to reduce auto-oxidation of the fat. From each of the loins, an approx. 30 cm long piece was cut from the neck end. The 16 loins were kept in dry-ice cooled containers, and 10 h after being taken off the slaughter line the processing began; the complete experiment, including Raman measurements, spanned less than 48 h from the slaughter. Cylinders were drilled out using an Ø12 mm rotary biopsy punch drill and subsequently sliced in discs (height 1.8 mm) to create a depth profile (as shown in Fig. 1). The slicing was done with a slicing guide and a scalpel. The first disc, containing skin, was cut to a thickness of 2.4 mm and discarded, as was any disc containing meat. Care was taken to keep the samples at maximum +5 °C. In total 134 discs were acquired from the 16 loins. Each disc was divided in two, and one half was analyzed by Raman spectroscopy and the other half was used for reference analysis.

### Reference analysis (gas chromatography)

Gas chromatography was applied to determine fatty acid composition. An Agilent Technologies 6890A GC system was used with a capillary column (DB-WAX 30.0 m × 250 µm, and 0.25 µm film thickness; Agilent Technologies) coupled to a flame ionization detector. Prior to injection the fat samples were methylated: fat was saponified by addition of sodium hydroxide and heated to approx. 100 °C for 5 min. After cooling to room temperature, methylation was achieved by adding borate-fluoride in methanol, and hydroquinone in methanol-solution acting as an antioxidant. Extraction of the fatty acids for GC analysis was accomplished by further heating the sample at approx. 100 °C for 5 min, and after cooling to room temperature adding sodium chloride solution and *n*-heptane, then shaking the mixture for 30 s. Allowing the hydrophobic heptane phase (containing the esters) to separate from the hydrophilic salt phase, the mixture was set to rest for 10 min, after which the top phase was transferred to a vial for injection into the GC instrument. One microliter of methylated sample was injected with an autosampler with a split ratio of 175:1, and run for 10 min. Helium and hydrogen were used as carrier gasses with pressure of 22.5 psi and flow rate of 184 ml/min, respectively. In the injector and the detector the temperatures were 250 °C, whereas the oven temperature was maintained at 210 °C. *n*-heptane was used as cleaning agent for the column. A few of the samples were run on a GC-MS instrument under similar instrumental conditions to allow a clear identification of fatty acids, and the assignment of reference retention peaks was



**Figure 1.** Sampling of porcine backfat. Depth profiles created by drilling out a cylinder and subsequently slicing it into discs (Ø12 mm and height 1.8 mm). The number of discs for each carcass depends on the thickness of the fat layer.

transferred to the sample retentions by dynamic time warping.<sup>[23]</sup> Sixteen single fatty acids (C14:0, C15:0, C15:1, C16:0, C16:1, C17:0, C17:1, C18:0, C18:1, C18:2, C18:3, C18:4, C20:0, C20:1, C20:2, C20:3) were identified. The results were reported as the percentage area of each individual fatty acid expressed in percent of the total amount of fatty acids. IVs were calculated on the basis of these results, by applying the American Oil Chemists' Society recommended practice 1c-8522, 'Calculated Iodine Value' with the extensions by Petursson.<sup>[24]</sup>

### Raman spectroscopic measurements

Raman spectra were recorded on a RamanRxn1 instrument (Kaiser Optical Systems Inc, MI, USA) equipped with a 785 nm near-infrared diode laser (Invictus, Kaiser Optical Systems Inc., MI, USA). For these experiments a single holographic grating and a thermoelectric cooled charge-coupled device detector, operated at  $-40^{\circ}\text{C}$ , were used. The probe was an Mk II filtered probe attached to a microscope with a  $10\times$  objective. The spectra were acquired using a total of 16 accumulations of 1 s exposure and they were stored as Raman shifts in the range  $1800\text{--}200\text{ cm}^{-1}$ . The samples were randomized and measured on a metal plate. Measurements were carried out at four different positions on each sample disc to increase sample representation. The average of the four spectra was used to even out the occurrence of heterogeneity within the sample.

### Data processing

#### Preprocessing method

The presence of fluorescence in biological samples is one of the main challenges of Raman spectroscopy. Fluorescence often gives raise to varying baseline, masking of the Raman bands. Different preprocessing techniques have been proposed to minimize the spectral fluctuations caused by fluorescence and different optical properties of the sample surface being measured.<sup>[25,26]</sup> In the present work, standard normal variate (SNV) scaling<sup>[27]</sup> proved to be the most suited because partial least squares (PLS) models based on SNV corrected spectra provided the best model (giving parsimonious models with the lowest prediction error). Applying SNV, the individual mean value is subtracted from each spectrum and afterwards divided by the standard deviation of all pooled variables for the given sample (Eqn 1). Thereby, both additive and multiplicative effects are eliminated.

$$X_{\text{SNV}} = \frac{X_i - \bar{x}}{SD} \quad (1)$$

Two multivariate methods – principal component analysis (PCA)<sup>[28]</sup> and PLS regression<sup>[29]</sup> – were applied. All data analyses were carried out using MATLAB R2010 (Mathworks, MA, USA).

#### Band fitting

In Raman spectroscopy both wavenumber shifts and bandwidth can contain relevant chemical and structural information. Band fitting was used to locate wavenumber shifts and changes in band shapes. The so-called Cauchy–Lorentz mathematical functions was applied to describe the bands of the Raman spectrum.

$$x(\sigma) = p_3 \times \frac{1}{\pi} \times \frac{p_2}{(\sigma - p_1)^2 + p_2^2} + p_4 + p_5 \times \sigma \quad (2)$$

In this analysis three different band parameters – band location ( $p_1$ ), bandwidth ( $p_2$ ), and intensity (band height,  $p_3$ ) – for each spectrum  $x$  were investigated. Equation 2 was fitted by non-linear least squares regression on a spectral segment window  $\sigma$  of  $100$  or  $150\text{ cm}^{-1}$  width ( $334$  or  $501$  data points) surrounding the spectral band of interest. The first-order polynomial in Eqn 2 was added to compensate for baseline offset ( $p_4$ ) and slope ( $p_5$ ) difference and was fitted simultaneously with the other model parameters. Parameters in Eqn 2 are estimated using the Levenberg–Marquardt algorithm<sup>[30,31]</sup> and implemented in MATLAB employing in-house routines.

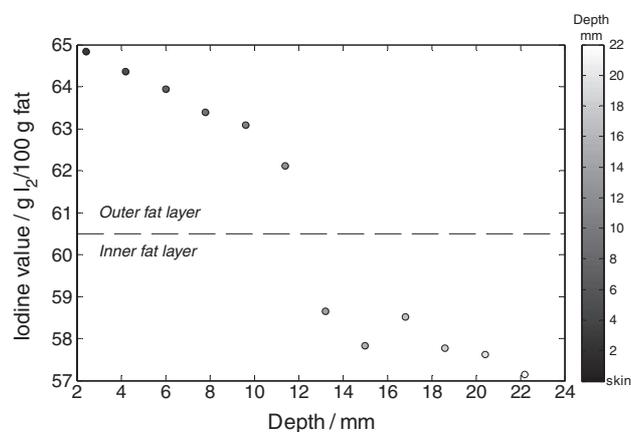
## Results and Discussion

### Iodine depth profile

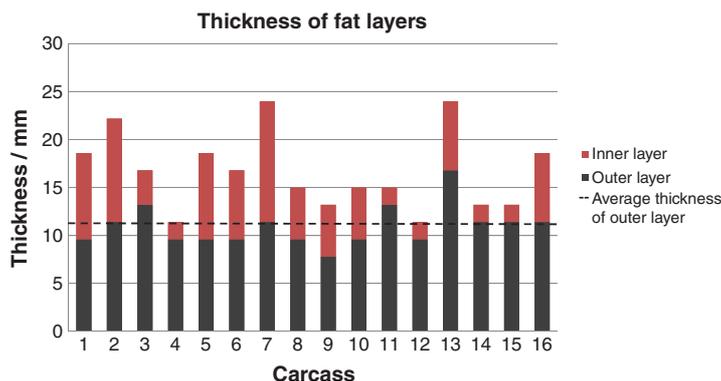
Figure 2 shows the degree of unsaturation (expressed by iodine values) as a function of fat layer depth (measured from skin) for one carcass. This trend is representative for the 16 carcasses. The profile clearly show that overall the degree of unsaturation decreased with depth. The fat layer just below the skin had the highest iodine value, which slowly decreased with depth until a substantial sudden decrease in the iodine value occurred at depths between  $10.5$  and  $12$  mm, at which stage the iodine value levels out. The sudden decrease indicates a change in fat layer from the outer to the inner. The average height of the outer fat layer of the samples included in this study is  $11.6$  mm (as indicated in Fig. 3). The observed change in the iodine value as a function of depth is in accordance with previous observations.<sup>[18,32–34]</sup>

### Raman spectral interpretation

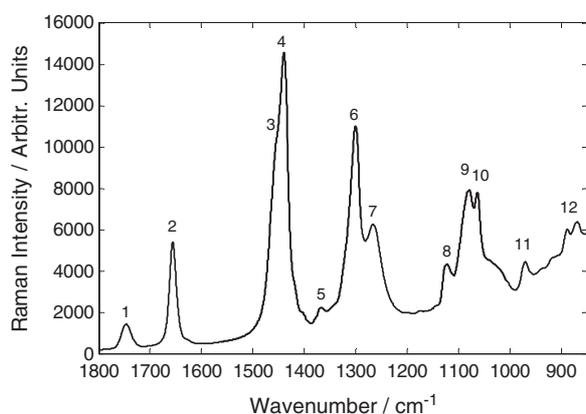
The bands in the Raman spectrum of pork adipose tissue can be assigned to different molecular vibrations (see Fig. 4 and Table 1). The table includes structural and group assignments for the spectral bands based on literature.<sup>[19,35–38]</sup> Furthermore, the table



**Figure 2.** Iodine value (GC) of porcine backfat as a function of depth location (distance from skin surface) in one pig carcass (no. 7). Between  $11.4$  and  $13.2$  mm from the skin there is a sudden decrease in iodine value indicating a change in fat layer. The broken line marks the fat layer change.



**Figure 3.** Thickness of the outer and inner fat layers of backfat for the 16 porcine carcasses.



**Figure 4.** Average Raman spectrum of pork adipose tissue with assignment numbers corresponding to Table 1.

shows coefficients of determination ( $R^2$ ) for the maximum heights of each band and IV. The spectrum shows that the vibrations from C=C double bonds and polymethylene  $\text{CH}_2$  dominate the spectra. Furthermore, carbonyl stretching C=O vibrations can be observed although not nearly as strongly as in the infrared region.<sup>[39]</sup> The following section will present more detailed descriptions of the most interesting bands found in different regions of the spectra going from high to low wavenumbers.

#### The region 1800–1600 $\text{cm}^{-1}$

In the region 1800–1600  $\text{cm}^{-1}$  two bands are present (1 and 2 in Fig. 4). The carbonyl band at 1745  $\text{cm}^{-1}$  with weak intensity gives an indirect indication of the average fatty acid chain length because short chains will give a higher concentration of carbonyl stretching and thereby a higher signal intensity.<sup>[40]</sup> The second band appearing at 1655  $\text{cm}^{-1}$  corresponds to the stretching vibration of C=C and should therefore be a good measure of the degree of unsaturation in fat. A linear regression between the intensity of the C=C stretching band and IV showed an  $R^2 = 0.60$  for the present study. The exact location of the band is important as it indicates the conformation of the C=C bond. Bailey and Horvat<sup>[41]</sup> showed that the wavenumber of the band will shift and the width of the band will vary with respect to the relation between *cis* (1650  $\text{cm}^{-1}$ ) and *trans* (1668  $\text{cm}^{-1}$ ) isomers. In addition, the presence of conjugated double bonds will shift the band position approx.  $-20 \text{ cm}^{-1}$ .<sup>[36]</sup>

#### The region 1600–1400 $\text{cm}^{-1}$

In the scissoring region 1600–1400  $\text{cm}^{-1}$  there is one dominating band around 1438  $\text{cm}^{-1}$  (4 in Fig. 4) originating from the methylene scissoring deformation. The band has a shoulder at 1460  $\text{cm}^{-1}$  (3 in Fig. 4), which corresponds to the deformation of methyl. Literature describes that the relative intensities of these bands to each other do not increase with the chain length, but the

**Table 1.** Assignments of bands in the Raman spectra of adipose tissue. The band numbers correspond to bands in Fig. 4. For each assigned band the position, functional group, mode of vibration, intensity, and coefficient of determination ( $R^2$ ) between maximum height and IV is given

Band	Band position ( $\text{cm}^{-1}$ )	Functional group	Mode of vibration	Intensity	$R^2$ to IV
1	1745	C=O	Stretch	w	0.08 <sup>b</sup>
2	1655	C=C	Stretch	m	0.60
3	1460	$\text{CH}_3$	Antisymmetric deformation	a	—
4	1438	$>\text{CH}_2$	Symmetric deformation (scissor)	s	0.40 <sup>b</sup>
5	1368	$\text{CH}_3$	Symmetric deformation (umbrella)	w	0.60 <sup>b</sup>
6	1301	$>\text{CH}_2$	Twisting (all-in-phase)	s	0.52 <sup>b</sup>
7	1266	=C-H	Symmetric rock <i>cis</i>	m	0.71
8	1125	C-C	Aliphatic in-phase stretch	w	0.49 <sup>b</sup>
9	1080	C-C	Aliphatic stretch	w	0.24
10	1068	C-C	Aliphatic out-of-phase stretch	w	0.49 <sup>b</sup>
11	970	=C-H	Out-of-plane bend <i>cis</i>	w	0.70
12	868	C-C	Stretch	w	0.41 <sup>b</sup>

<sup>a</sup>Shoulder.

<sup>b</sup>Negative correlation.

locations can change up to approx.  $15\text{ cm}^{-1}$ .<sup>[36]</sup> However, the shift in location mainly occurs in short-chain fatty acids ( $<10:0$ ), while they become negligible for longer chains such as the ones investigated in this study. In a study of edible oils Bailey and Horvat<sup>[41]</sup> showed that the ratio between the intensity of the C=C band,  $I_{1660\text{cm}^{-1}}$ , and the  $\text{CH}_2$  scissoring band,  $I_{1440\text{cm}^{-1}}$ , showed a fairly good correlation with the iodine value determined from GC ( $R^2$  was not reported). In this study, an  $R^2 = 0.64$  was found between the ratio ( $I_{1655\text{cm}^{-1}}/I_{1438\text{cm}^{-1}}$ ) and the IV.

#### Region $1400\text{--}1200\text{ cm}^{-1}$

In this region, a very weak band from the methyl umbrella deformation is found at approx.  $1368\text{ cm}^{-1}$ . Furthermore, this spectral region contains bands from methylene,  $\text{CH}_2$ , in-phase twisting and symmetric rocking of unconjugated *cis* double bonds,  $=\text{C-H}$ . The twisting band is affected by the conformation of the polymethylene backbone chain and will shift in location from  $1301\text{--}1295\text{ cm}^{-1}$ . In addition the bandwidth can change. The rocking band near  $1266\text{ cm}^{-1}$  occurs when the polymethylene chain becomes longer than four. Previous results<sup>[42]</sup> have shown that the level of *cis* unsaturation can be estimated from the rocking band. In addition, in a study of different oils and fats, Baeten et al.<sup>[40]</sup> showed a very high correlation between total amount of unsaturation and the intensity of the rocking band ( $R^2 = 0.93$ ). In the present study the rocking band also shows a high correlation with IV ( $R^2 = 0.71$ ). Likewise the ratio between intensities of the rocking and twisting bands is well correlated to the degree of unsaturation ( $R^2 = 0.69$ ). Li-Chan<sup>[38]</sup> showed that the total content of *cis* isomer can be determined from the ratio between the band intensities at  $1265$  and  $1303\text{ cm}^{-1}$ . However, the correlation to *cis* isomer was not investigated for the present study.

#### Region $1200\text{--}800\text{ cm}^{-1}$

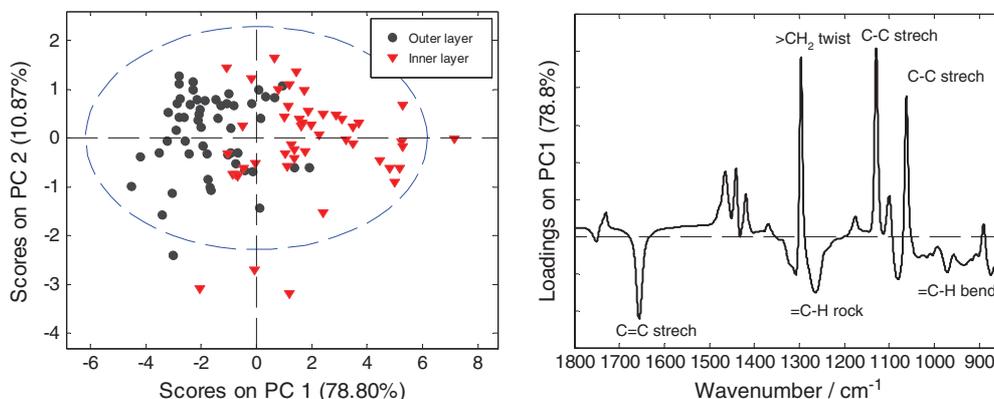
In the so-called fingerprint region from  $1200\text{--}800\text{ cm}^{-1}$  we find a series of bands from the C-C skeleton vibrations. Three of the main bands in this region originate from vibrations of the aliphatic C-C stretching:  $1125\text{ cm}^{-1}$  (in-phase),  $1068\text{ cm}^{-1}$  (out-of-phase), and  $1080\text{ cm}^{-1}$  (with a phase angle between  $0$  and  $\pi$ ). These bands vary according to the conformation of the fatty acid chains and are affected by intermolecular interactions.<sup>[36]</sup> The band at  $970\text{ cm}^{-1}$  corresponds to the bending of C-H attached to double bonds with *cis* conformation.<sup>[43]</sup> In the present study the intensity of this band has an  $R^2 = 0.70$  to IV.

#### Fast discrimination between fat layers

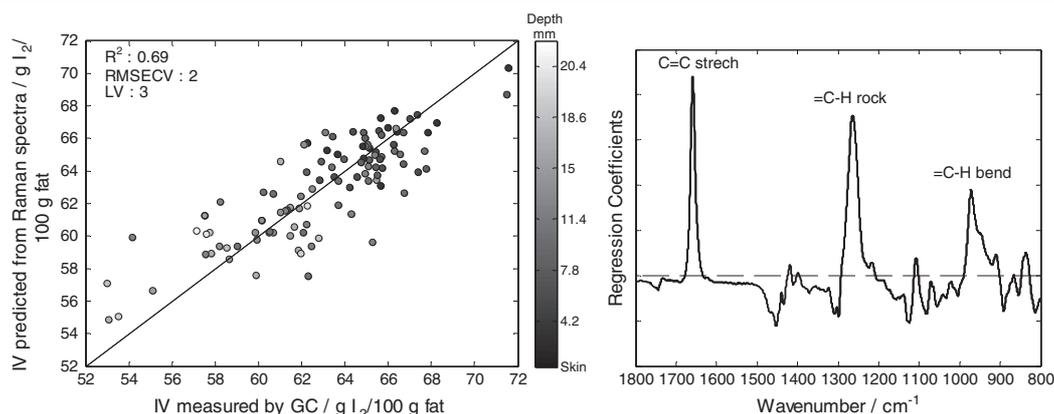
It is known that adipose tissue is a mixture of different fatty acids, saturated and unsaturated, and that the amount of unsaturated fatty acids decreases as a function of backfat depth.<sup>[13]</sup> Because of the Raman spectra change with respect to the degree of saturation, it should be possible to discriminate between adipose tissue from different depths, e.g., adipose tissue  $2\text{ mm}$  from the skin compared with adipose tissue  $10\text{ mm}$  from the skin. One objective of the present study was to investigate if Raman spectra can discriminate between the outer and inner fat layer from adipose tissue. On the basis of the IV found from GC analysis, we classified the samples as part of the inner or outer layer. The boundary between inner and outer layer was defined as the point with the largest difference in IV between consecutive discs (Fig. 2). Coloring the Raman spectra according to location of the sample (inner/outer fat layer) revealed that layers could be distinguished for a number of spectral bands (results not shown). Subsequently, a PCA was conducted on the SNV processed spectra. The PCA score plot in Fig. 5 (left) shows an almost complete separation of fat layers along the first principal component, which describe 79% of the total spectral variation. When examining the corresponding loading vector for the first component (Fig. 5, right) it is observed that samples from the inner fat layer are positively correlated to high intensity of the bands  $1296\text{ cm}^{-1}$  ( $>\text{CH}_2$  twist),  $1125\text{ cm}^{-1}$  (C-C in-phase stretch), and  $1060\text{ cm}^{-1}$  (C-C out-of-phase stretch). Furthermore, the samples from the inner layer also contain higher amounts of the  $\text{CH}_2$  scissoring deformation ( $1438\text{ cm}^{-1}$ ). These findings indicate that the inner layer contains a higher amount of  $\text{CH}_2$  presumably because of a higher amount of saturated fatty acids, which is consistent with the observation of low IV values in the inner layer (Fig. 2). The samples from the outer layer have negative first score values, which is correlated to high intensity of the three bands related to double bonds  $1655\text{ cm}^{-1}$  (C=C stretch),  $1266\text{ cm}^{-1}$  ( $=\text{C-H}$  rock), and  $970\text{ cm}^{-1}$  ( $=\text{C-H}$  bend). This is in accordance with the fact that saturation will increase as a function of penetration depth (from the skin surface).

#### Prediction of iodine value, SFA, MUFA and PUFA

The results from the PLS analysis (Fig. 6, left) showed that by using three components the IV could be predicted with a relatively high correlation ( $R^2 = 0.69$ ) and a fairly low error of cross



**Figure 5.** Principal component analysis of Raman spectra from porcine backfat with varying distance to skin surface. Left: PCA score plot of the first and second principal component shows the separation between samples from the inner and outer fat layer of porcine backfat, respectively. Right: PCA loadings on the first principal component.



**Figure 6.** Left: predicted versus reference iodine value by PLS and Raman spectroscopy in pork fat sample from 16 carcasses (134 samples) taken at various depths from the skin surface. Right: regression vector for a PLS model (3 latent variables) for prediction of iodine value from Raman spectra.

validation (random subset with 18 splits and 5 iterations, root mean square error of cross validation (RMSECV) = 2.00). The result is in agreement with the results of Olsen *et al.*,<sup>[20]</sup> who reported a test-set validation with a prediction error (root mean square error of prediction) of 2.04 for six latent variables. The regression vector for the PLS model (Fig. 6, right) showed as expected a high contribution from the C=C band at 1655  $\text{cm}^{-1}$  and =C-H bands at 1266  $\text{cm}^{-1}$  and 970  $\text{cm}^{-1}$ . Table 2 summarizes the results of prediction models for IV, SFA, MUFA, PUFA. The table shows that a five-component PLS model predicts % of PUFA of the total fatty acids with a low error (RMSECV = 1.17) and an  $R^2 = 0.72$  suggesting Raman spectroscopy is a suitable method for fast determining the content of PUFA in a product. For the food industry this may be a beneficial and rapid tool to determine if the product has the right balance between high nutritional value (high PUFA) and good oxidative stability (low PUFA).

### Spectral phenomena occurring as a function of fat depth

Our results showed some characteristic spectral changes occurring as functions of distance from the skin. Most notable are the occurrence of distinct, sharp bands at 1060 (Fig. 7(A)) and 1129  $\text{cm}^{-1}$  (Fig. 7(B)), a decrease in the 1266  $\text{cm}^{-1}$  band (Fig. 7(C)), a shift and narrowing of the band at 1301  $\text{cm}^{-1}$  (Fig. 7(C)), and a decrease in the 1655  $\text{cm}^{-1}$  band (Fig. 7(D)). The same spectral features were found in a study of lipoproteins by Chan *et al.*<sup>[44]</sup> Figures 7(A) and (B) show bands developing systematically with increased depth suggesting that some C-C stretching vibrations do not exist in the fat from the most outer layer, but gradually start to increase as the measurement move inwards. This may be explained by the fact that the hydrocarbon chains in the most outer fat are highly disordered (unsaturated), but become more and more ordered (saturated) deeper into the

fat. The all-in-phase twisting motion of methylene (Fig. 7(C)) is also affected by the fatty acid chain conformation. The band occurs at 1305  $\text{cm}^{-1}$  but shifts to a lower wavenumber of about 1296  $\text{cm}^{-1}$ . When the fatty acids chains are straight molecules as is the case for SFAs, the twist motion occurs at a lower wavenumber than when the fatty acids chains become more bent or crooked as is the case for UFAs. Thus, as the fat becomes more and more saturated, the methylene twisting band shifts to lower frequencies. Another possible reason for the shift and shape change of the methylene twisting band is a change in the physical state of the samples. The adipose tissue in the outer layer (just below the skin) contains more unsaturated fatty acid, and therefore it has a lower melting point than the more saturated fat in the inner layer. The temperature during measurement may have reached close to the melting point of the fat from the outer layer; as a consequence the samples may have melted a little and changed from a total solid to a more liquid-like state. Deeper into the backfat the amount of saturated fat is higher and therefore the melting point is also higher.<sup>[44,45]</sup> In this study the shifting and shape change were only observed for the methylene twisting band. The intensity of C=C stretching band at 1655  $\text{cm}^{-1}$  (Fig. 7(D)) decreases along with the fat depth and degree of unsaturation.

### Band fitting

Usually, band intensity is the important parameter considered in spectral analysis, and often spectral shifts are considered unwanted and neglected such as in PLS modeling. Advanced data pre-processing methods are developed to correct for shifts prior to the analysis, e.g., co-shift, icoshift,<sup>[46]</sup> and warping.<sup>[23]</sup> However, like in nuclear magnetic resonance spectra, wavenumber shifts and bandwidth can also contain relevant chemical and structural information. Therefore, we investigated the importance of bandwidth and location compared with the band intensity for three of the major bands: C=C stretching (1655  $\text{cm}^{-1}$ ), methylene twisting (1301  $\text{cm}^{-1}$ ), and rocking (1265  $\text{cm}^{-1}$ ). This was done by fitting mathematical Cauchy-Lorentz distributions (Eqn 2) that could describe the band with regard to height, width, and shape, and afterwards correlating each of them to the total amount of SFA, MUFA, PUFA, and to the IV; the coefficients of determination ( $R^2$ ) are shown in Table 3.

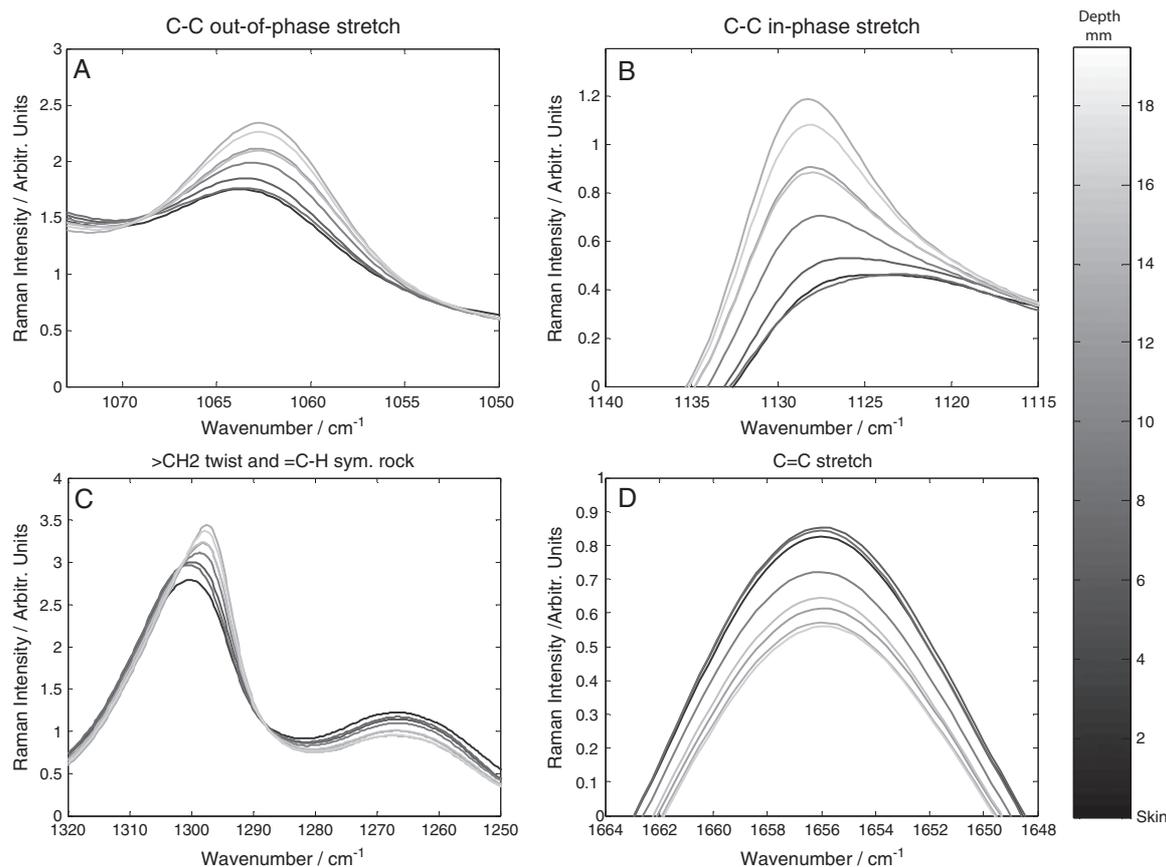
The results show that for the =C-H rocking band at 1266  $\text{cm}^{-1}$  the highest correlation was found between band height and IV.

**Table 2.** PLS predictions models for IV, SFA, MUFA and PUFA

	RMSECV	$R^2$	No. latent variables	Range
IV <sup>a</sup>	2.00	0.69	3	52–69
SFA <sup>b</sup>	2.24	0.50	4	36–51
MUFA <sup>b</sup>	2.28	0.57	7	34–50
PUFA <sup>b</sup>	1.17	0.72	5	11–23

<sup>a</sup>gram  $I_2$ /100 gram sample.

<sup>b</sup>% of total fatty acids.



**Figure 7.** Raman spectral features occurring as functions of backfat depth (dark → light, skin → meat). The shown spectra are from carcass no. 7. (A) out-of-phase C–C stretching; (B) in-phase C–C stretching; (C) >CH<sub>2</sub> twisting and =C–H sym. rocking; and (D) C=C stretching.

Fairly high correlations were also found between both width and IV, and location and IV. However, when examining the band at 1301–1296 cm<sup>-1</sup> both location and width show much higher correlations with IV, SFA, and MUFA than band height. This is in agreement with the fact that the bandwidth and location of methylene twisting change depending on the order of hydrocarbon chains. For the C=C band at 1656 cm<sup>-1</sup> results show that the width of the band is not very important, as it only shows low correlations with the reference values. However, both location and intensity/height seem to be important. The location of the band has high correlations to MUFA (negatively) and PUFA (positively), while height has the highest correlation to IV. This

indicates that this particular band changes the position according to conformation, and that the presence of conjugated double bonds will shift the band position.

## Conclusions

This study has proved that Raman spectroscopy can be applied successfully to discriminate between fat layers in porcine adipose tissue. On the basis of spectral interpretation, it was found that differences in intensities of the vibrational bands of C=C and =C–H were the main variations between the layers.

Another objective of this study was to estimate the IV, SFA, MUFA, and PUFA on the basis of the Raman spectra. The study found that a three-component PLS regression model could predict the IV with an  $R^2=0.69$  and RMSECV=2.00. The best PLS regression model was found for prediction of PUFA ( $R^2=0.72$  and RMSECV=1.17).

Different spectral phenomena occurred as a function of fat depth. The three band parameters height, location, and width were extracted by Cauchy–Lorentz band fitting. Results showed large differences between the three parameters when correlating them to IV, SFA, MUFA, and PUFA, leading to the conclusion that important information can be found in not only band intensity, but also band location and width.

This research has proved that Raman spectroscopy has potential value for the pork industry. However, the methodology used in this study was excision based, and to implement it in a real production setting it would need some development. A possible way to develop the method and make it more suitable for online

**Table 3.** The coefficient of determination ( $R^2$ ) between IV, SFA, MUFA, PUFA and the three parameters height, location, and width for bands at 1266, 1301 and 1656 cm<sup>-1</sup>

		IV	%SFA	%MUFA	%PUFA
1266 cm <sup>-1</sup>	=C–H sym. rock				
	height	0.67	0.45 <sup>a</sup>	0.05	0.36
	location	0.46	0.44 <sup>a</sup>	0.15	0.11
1301 cm <sup>-1</sup>	>CH <sub>2</sub> twist				
	height	0.01	0.01	0.01	0.02 <sup>a</sup>
	location	0.45	0.49 <sup>a</sup>	0.21	0.08
1656 cm <sup>-1</sup>	C=C stretch				
	height	0.67	0.45 <sup>a</sup>	0.05	0.34
	location	0.06	0.03	0.40 <sup>a</sup>	0.54
	width	0.09	0.01 <sup>a</sup>	0.03 <sup>a</sup>	0.18

<sup>a</sup>Negative correlation.

analysis could be to combine an optical probe with knives or needles, so the probe can be inserted into an intact pork carcass while measuring. This has already been investigated for near-infrared spectroscopy, but thus not for Raman spectroscopy to our knowledge.

### Acknowledgements

This work was carried out in collaboration between Quality and Technology (Department of Food Science, University of Copenhagen) and Carometec A/S (Denmark). The authors are grateful for generous financial support from the Danish National Advanced Technology Foundation through the NITFOM project and from Danish Agency for Science, Technology and Innovation through the Consortium 'QbD - Quality by Design in the Food and Pharmaceutical Industries' ([www.qbd.dk](http://www.qbd.dk)). Tönnies Fleischwerk (Germany) is thanked for providing the raw material.

### References

- [1] K. R. Glaser, C. Wenk, M. R. L. Scheeder, *Arch. Anim. Nutr.* **2002**, *56*, 51.
- [2] F. M. Whittington, N. J. Prescott, J. D. Wood, M. Enser, *J. Sci. Food Agric.* **1986**, *37*, 753.
- [3] A. Hugo, E. Roodt, *Food Rev. Int.* **2007**, *23*, 175.
- [4] A. Astrup, J. Dyerberg, P. Elwood, K. Hermansen, F. B. Hu, M. U. Jakobsen, F. J. Kok, R. M. Krauss, J. M. Lecerf, P. LeGrand, P. Nestel, U. Riserus, T. Sanders, A. Sinclair, S. Stender, T. Tholstrup, W. C. Willett, *Am. J. Clin. Nutr.* **2011**, *93*, 684.
- [5] S. J. Maw, V. R. Fowler, M. Hamilton, A. M. Petchey, *Meat Sci.* **2003**, *63*, 185.
- [6] A. J. Kempster, A. W. Dilworth, D. G. Evans, K. D. Fisher, *Anim. Prod.* **1986**, *43*, 517.
- [7] J. D. Wood, M. Enser, F. M. Whittington, C. B. Moncrieff, A. J. Kempster, *Livest. Prod. Sci.* **1989**, *22*, 351.
- [8] G. Sturm, I. Karl, B. Schwarz, G. Siebert, *Z. Ernährungswiss.* **1982**, *21*, 2.
- [9] J. M. Eggert, A. P. Schinckel, S. E. Mills, J. C. Forrest, D. E. Gerrard, E. J. Farrand, B. C. Bowker, E. J. Wynveen, *Swine Day Report*, Purdue University, West Lafayette, IN, USA, **1998**.
- [10] W. W. Christie, D. M. Jankinson, J. H. Moore, *J. Sci. Food Agric.* **1972**, *23*, 1125.
- [11] M. Camara, J. Mourot, C. Fevrier, *Ann. Nutr. Metab.* **1996**, *40*, 287.
- [12] B. Leuret, J. Mourot, *Prod. Anim.* **1998**, *11*, 131.
- [13] A. Daza, J. Ruiz-Carrascal, A. Olivares, D. Menoyo, C. J. Lopez-Bote, *Food Sci. Technol. Int.* **2007**, *13*, 135.
- [14] R. J. Ouellette, *Organic Chemistry: A Brief Introduction*, Prentice Hall, Upper Saddle River, N.J. **1998**.
- [15] F. J. Mcevoy, A. B. Strathe, M. T. Madsen, E. Svalastoga, *Acta Vet. Scand.* **2007**, *49*, 32.
- [16] W. G. Moody, S. E. Zobrisky, *J. Anim. Sci.* **1966**, *25*, 809.
- [17] T. Koch, S. Lakshmanan, S. Brand, M. Wicke, K. Raum, D. Morlein, *Meat Sci.* **2011**, *88*, 67.
- [18] M. Monziols, M. Bonneau, A. Davenel, M. Kouba, *Meat Sci.* **2007**, *76*, 54.
- [19] H. Sadeghi-Jorabchi, P. J. Hendra, R. H. Wilson, P. S. Belton, *J. Am. Oil Chem. Soc.* **1990**, *67*, 483.
- [20] E. F. Olsen, C. Baustad, B. Egelanddal, E. O. Rukke, T. Isaksson, *Meat Sci.* **2010**, *85*, 1.
- [21] W. S. Kwak, J. S. Kang, *Bioresour. Technol.* **2006**, *97*, 243.
- [22] I. Moreira, F. L. Mourinho, P. L. D. Carvalho, D. Paiano, L. M. Piano, I. S. Kuroda, *Brazil. J. Anim. Sci.* **2009**, *38*, 2408.
- [23] T. Skov, F. van den Berg, G. Tomasi, R. Bro, *J. Chemom.* **2006**, *20*, 484.
- [24] S. Petursson, *J. Am. Oil Chem. Soc.* **2002**, *79*, 737.
- [25] A. Rinnan, F. van den Berg, S. B. Engelsen, *TrAC, Trends Anal. Chem.* **2009**, *28*, 1201.
- [26] N. K. Afseth, V. H. Segtnan, J. P. Wold, *Appl. Spectrosc.* **2006**, *60*, 1358.
- [27] I. Barnes, M. S. Dhanoa, S. J. Lister, *Appl. Spectrosc.* **1989**, *43*, 772.
- [28] H. Hotelling, *J. Educ. Psychol.* **1933**, *24*, 417.
- [29] S. Wold, H. Martens, H. Wold, *Lect. Notes Math.* **1983**, *973*, 286.
- [30] D. W. Marquardt, *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431.
- [31] R. H. Barham, W. Drane, *Technometrics* **1972**, *14*, 757.
- [32] H. K. Dean, T. P. Hilditch, *Biochem. J.* **1933**, *27*, 1950.
- [33] E. H. Thompson, E. Allen, *J. Anim. Sci.* **1969**, *29*, 127.
- [34] D. P. Lo Fiego, P. Santoro, P. Macchioni, E. De Leonibus, *Meat Sci.* **2005**, *69*, 107.
- [35] N. K. Afseth, J. P. Wold, V. H. Segtnan, *Anal. Chim. Acta* **2006**, *572*, 85.
- [36] J. R. Beattie, S. E. J. Bell, B. W. Moss, *Lipids* **2004**, *39*, 407.
- [37] Y. M. Weng, R. H. Weng, C. Y. Tzeng, W. L. Chen, *Appl. Spectrosc.* **2003**, *57*, 413.
- [38] E. C. Y. Li-Chan, *Trends Food Sci. Tech.* **2002**, *7*, 341.
- [39] D. K. Pedersen, S. Morel, H. J. Andersen, S. B. Engelsen, *Meat Sci.* **2003**, *65*, 581.
- [40] V. Baeten, P. Hourant, M. T. Morales, R. Aparicio, *J. Agric. Food Chem.* **1998**, *46*, 2638.
- [41] G. F. Bailey, R. J. Horvat, *J. Am. Oil Chem. Soc.* **1972**, *49*, 494.
- [42] H. Sadeghi-Jorabchi, R. H. Wilson, P. S. Belton, J. D. Edwardswebb, D. T. Coxon, *Spectrochim. Acta, Part A* **1991**, *47*, 1449.
- [43] M. D. Guillen, N. Cabo, *J. Sci. Food Agric.* **1997**, *75*, 1.
- [44] J. W. Chan, D. Motton, J. C. Rutledge, N. L. Keim, T. Huser, *Anal. Chem.* **2005**, *77*, 5870.
- [45] J. R. Beattie, S. E. J. Bell, C. Borggaard, A. Fearon, B. W. Moss, *Lipids* **2006**, *41*, 287.
- [46] F. Savorani, G. Tomasi, S. B. Engelsen, *J. Magn. Reson.* **2010**, *202*, 190.

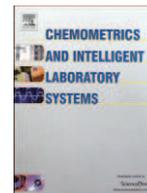
# PAPER II

---

Lyndgaard, L.B., van den Berg, F. & de Juan, A.

Quantification of paracetamol through tablet blister packages by  
Raman spectroscopy and MCR-ALS

*Chemometrics and Intelligent Laboratory Systems, 25, 58-66, 2013*



# Quantification of paracetamol through tablet blister packages by Raman spectroscopy and multivariate curve resolution–alternating least squares



Lotte B. Lyndgaard<sup>a,\*</sup>, Frans van den Berg<sup>a</sup>, Anna de Juan<sup>b</sup>

<sup>a</sup> University of Copenhagen, Faculty of Science, Department of Food Science, Quality and Technology Research Group, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

<sup>b</sup> Universitat de Barcelona, Department of Analytical Chemistry, Chemometrics Group, Diagonal 645, 08028 Barcelona, Spain

## ARTICLE INFO

### Article history:

Received 4 January 2013

Received in revised form 20 March 2013

Accepted 23 March 2013

Available online 6 April 2013

### Keywords:

Raman spectroscopy

MCR-ALS

Paracetamol

Blister package

Quantitative determination

Matrix effects

## ABSTRACT

Raman spectroscopy has potential for non-invasive detection and quantification of chemical solids concealed within packages e.g. plastic bottles, bags or blister packs. In this study we demonstrate how paracetamol as part of ternary mixtures can be quantified inside blister packages using the combination of Raman spectroscopy and multivariate curve resolution–alternating least squares (MCR-ALS). An MCR-ALS model with correlation constraint could predict paracetamol concentrations in a test set with an error of 0.8% w/w in the concentration range 0–20% w/w. Additionally, a new multivariate approach to overcome sample matrix effects is proposed to handle variation in the blister packing material between different production batches. These variations interfere with the Raman signal and therefore make quantification challenging with conventional regression methods. By a novel modification of the correlation constraint in MCR-ALS, quantification despite matrix effects is made possible. Calibration models with low prediction errors of paracetamol could be obtained if a simple calibration set of unshielded samples (samples measured directly, hence without blister package) was combined with a very small set of samples measured through blister. Models with prediction errors of 1.3, 1.4 and 1.9% w/w were obtained for samples measured through three distinct blister packages. The novelty in the modification allows working with a multiset structure, making local regression models that can handle both signal interferences and matrix effects (e.g. packing material, temperature, instrumental variations).

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The ability to evaluate pharmaceutical product composition through concealed packing has a multitude of advantages. Just to name some: revealing counterfeit products, controlling deterioration due to damaged packing material, or testing the quality as close to the product release as possible to evaluate potential harm done during product storage. To achieve this goal, Raman and near-infrared (NIR) spectroscopy are methods with high potential since non-contact measurements with up to several centimeters distance between the sample and measurement probe are possible. Samples can be measured through clear glass or plastic bottles, bags, or blister packages [1]. Contrary to NIR, Raman has an attractive spectral advantage: clear spectral features that are in general easily assigned to chemical compounds in the pharmaceutical product. This is due to the fact that Raman spectroscopy measures the fundamental vibrations, whereas NIR measures the overtones and combinations of the fundamentals [2].

Even though collection of Raman spectra through packing material is possible, it still presents some challenges when quantitative predictions of product parameters are to be made. The biggest challenge is that the packing material itself acts as interference in the measured

Raman spectra in two ways: adding new spectral features to the spectra measured and causing a decrease in the signal intensity linked to the product drug compounds. Traditionally, spectral knowledge has been utilized in the analysis of the Raman spectrum and often height or other features of a single specific band have been used for quantification [2]. This type of univariate calibration cannot account for interfering signals, such as those from packing material or partially overlapping bands from other sample compounds. Multivariate calibration methods like partial least squares (PLS) regression and multivariate curve resolution–alternating least squares (MCR-ALS) are the appropriate choices when dealing with interference and overlapping bands. In PLS regression, interferences like package material can be partly handled when they are properly represented in a calibration data set. If this precaution is not met samples containing interferences will be predicted as outliers with a large systematic error [3,4]. A second problem may arise when different interferences have to be taken into account in the model; they may cause deviations from the linear behavior of the calibration model, e.g. due to different matrix effects. The principle behind MCR is to resolve mixture spectra into the concentration-weighted contributions of the pure spectra of each individual compound. MCR-ALS works differently from PLS since the unexpected interferences, such as packing material contributions, are modeled explicitly and separately from the contributions of relevant compounds. There are different ways to instruct the MCR method to model interferences and one of

\* Corresponding author. Tel.: +45 35 33 36 00; fax: +45 35 33 32 45.  
E-mail address: [lottebs@life.ku.dk](mailto:lottebs@life.ku.dk) (L.B. Lyndgaard).

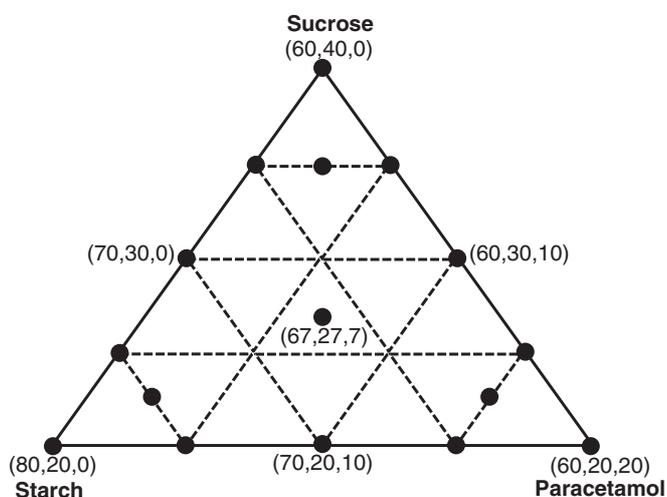
them is by applying constraints during the alternating least squares optimization. When quantification of compounds is the main goal of the analysis, a so-called correlation constraint can be applied to the concentration profiles. This constraint, first suggested by Antunes et al. [5], is designed to correlate known concentrations in a calibration set to concentrations found during the ALS optimization. In a study by Goicoechea et al. [6] it was shown how correlation constrained MCR-ALS was a feasible method for achieving analyte quantification in the presence of unexpected interferences. Unwanted interferences in spectroscopic measurements can arise from e.g. color differences [7], temperature variations [8,9] and instrumental variations [10].

In our study, MCR-ALS with a correlation constraint will be used to perform quantification of pharmaceutical compounds in a presence of spectral interferences and matrix effects rising from a blister package material. Until now the correlation constraint has always been applied in a global way to a single data matrix, where the calibration set and the test set are included. A new implementation proposed in this study allows working with a multiset structure, making local regression models that can handle diverse interferences and matrix effects. Raman spectroscopy is used to analyze ternary mixtures of pharmaceutical compounds through three different types of blister packages. The objectives are 1) to investigate if paracetamol, the active pharmaceutical ingredient (API) in the mixture, can be quantified using this measurement setup and 2) to find a solution for a potential industrial calibration challenge: differences in blister packing material. To pursue the objectives, the new implementation of MCR-ALS will be investigated and differences between this approach and PLS will be described.

## 2. Materials and methods

### 2.1. Sample preparation

18 mixtures of paracetamol (Sigma-Aldrich, Denmark), starch (wheat unmodified, Sigma-Aldrich, Denmark) and sucrose (Fluka Analytical, Denmark) were prepared according to the ternary mixture design shown in Fig. 1. Each compound was varied at six levels: paracetamol (0–20% w/w) which was considered the API of interest, sucrose (20–40% w/w) and starch (60–80% w/w). The flasks containing 5 g mixtures were placed on a shaker board at 900 rpm for 10 min. The blister packages used for this study were white polyvinyl-chloride (PVC) types from a number of different batches retrieved from a commercial product purchased in pharmacies.



**Fig. 1.** Overview of the mixture design used in the present study. Six different concentration levels for each of the three components: paracetamol (0–20% w/w), sucrose (20–40% w/w), starch (60–80% w/w). The center point is triplicated. In total 18 ternary mixtures.

### 2.2. Raman measurements

Raman measurements were carried out with a RamanRXn1 instrument (Kaiser optical systems, Inc, Michigan, USA). The laser was a 785 nm diode laser (Invictus, Kaiser Optical Systems, Inc., Michigan, USA) with a laser power of 200 mW. A single holographic grating and a Peltier-cooled CCD detector operating at  $-40^{\circ}\text{C}$  were used. For this experiment the pharmaceutical area testing (PhAT) probe was used because of its large spot size diameter (6 mm) with a focus depth of  $\pm 12$  mm; this allowed a large portion of the sample to be measured. Each mixture was measured directly (without blister) and through each of the three blister packages giving four data sets:  $D_0$ , the mixtures measured directly (unshielded) and  $D_{b(1,2,3)}$ , the mixtures measured through each of the blister packages 1, 2 or 3. For the measurements through blisters each mixture was placed in the blister and the backside (flat side) of the blister package was covered with a metal plate wrapped up in aluminum foil and held together with magnets (closely mimicking intact concealed commercial products, see Fig. 2). The samples were measured in a randomized order under conditions shielded from external light. The spectra were acquired using a total of two accumulations of 10 s exposure and stored as Raman intensities in the range  $1800\text{--}200\text{ cm}^{-1}$ . Measurements were carried out in triplicate illuminating at three slightly altered positions on the sample. Measuring at three different positions and using a wide area illumination were done in order to even out occurrence of sample heterogeneity. Although it might be statistically more sound to average the spectra of the three recordings made on one sample, we will treat them as separate measurements to take into account the natural powder heterogeneity. Our data matrix for one measurement setup thus consists of  $18 \times 3 = 54$  objects.

### 2.3. Data preprocessing

Fluorescence often gives rise to a varying baseline, which can distort the Raman signal. Different preprocessing techniques have been suggested to minimize the spectral fluctuations caused by fluorescence and different optical properties of the sample surface being measured [11]. In the present work various preprocessing techniques were investigated, and standard normal variate (SNV) scaling [12,13] proved to be the most suited technique to correct the spectra providing parsimonious models with the lowest prediction error for both MCR-ALS and PLS models. Applying SNV, the individual mean value is subtracted from each spectrum and the result is divided by the standard deviation over all pooled variables for the given sample (Eq. (1)). Thereby both additive and multiplicative effects are eliminated.

$$x_{snv} = (x_i - \bar{x}) / SD. \quad (1)$$

In this paper two different multivariate analysis methods – MCR-ALS and PLS – have been applied. A detailed description of PLS is given elsewhere [14,15]. All data analysis was carried out using Matlab 2011b (MathWorks Inc. Natick, MA, USA) with in-house written scripts, the PLS toolbox (Eigenvector Research, WA, USA) and the MCR-ALS GUI (free software available at <http://www.mcrals.info>).

## 3. Theory

### 3.1. MCR-ALS

The iterative MCR-ALS method has been widely used for bilinear decomposition of chemical data into chemically meaningful concentration and spectral profiles [16–21]. Fig. 3 gives an overview of the iterative process in MCR-ALS. It consists of the following main steps:

1. Building the data matrix
2. Estimation of the number of components present in data matrix

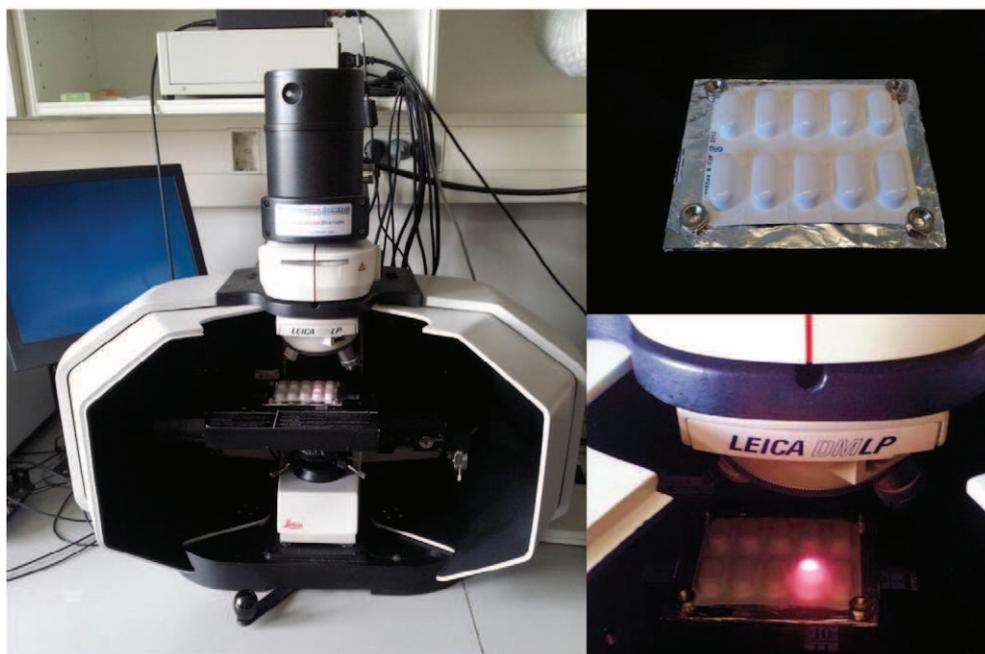


Fig. 2. Measurement set-up. Left: the blister package in the Raman microprobe. Upper right: the blister package with the backside (flat side) covered by a metal plate wrapped up in aluminum foil, held together with magnets. Lower right: a large portion of the sample measured using the pharmaceutical area testing (PHAT) probe with a 6 mm spot size.

3. Finding initial estimates of spectral (or concentration) profiles
4. Iterative ALS calculation of spectral and concentration profiles under constraints
5. Checking for convergence

Each of the listed steps will now be described in the context of the data set analyzed in this paper, especially the novel modification of the so-called correlation constraint.

1. The data set consists of Raman spectra of  $n$  mixture samples containing  $nc$  compounds in different concentrations. This can be arranged in a data matrix  $\mathbf{D}$ . Under the assumption of the natural spectroscopic bilinear model,  $\mathbf{D}$  can be expressed as

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (2)$$

where  $\mathbf{D}$  consists of  $n$  rows of different individual spectra measured for different pharmaceutical mixtures and  $m$  columns of spectral

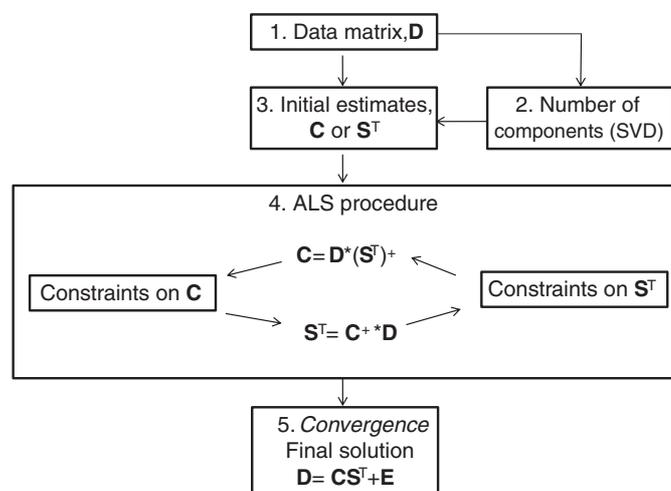


Fig. 3. Overview of the steps in the multivariate curve resolution-alternating least squares (MCR-ALS) procedure.

wavenumbers. The columns in  $\mathbf{C}$  represent the concentrations of  $nc$  components in these  $n$  mixtures and the rows in  $\mathbf{S}^T$  are their related  $nc$  spectra of length  $m$ .  $\mathbf{E}$  is the residual matrix with the variance not explained by the bilinear model. The goal of curve resolution methods is to perform the matrix decomposition according to Eq. (2), i.e. estimate  $\mathbf{C}$  and  $\mathbf{S}^T$  matrices from the analysis of the data matrix  $\mathbf{D}$ . An infinite number of possible solutions for the mathematical decomposition of  $\mathbf{C}$  and  $\mathbf{S}^T$  exist unless additional information is provided for the bilinear decomposition.

2. In order to determine the correct number of components, singular value decomposition (SVD) is typically applied to  $\mathbf{D}$  first. From the unique orthogonal decomposition, the number of significant components directly related to chemical compounds can be estimated [22].
3. The initial estimates used should be a set of either  $nc$  concentration or spectral profiles. In our case SIMPLISMA [23] initial estimates from the purest spectral profiles were determined.
4. If spectral profiles are the initial estimates, concentration profiles, and after that improved spectra, can be calculated by least squares with the following expressions:

$$\mathbf{C} = \mathbf{D}(\mathbf{S}^T)^+ \quad (3)$$

$$\mathbf{S}^T = \mathbf{C}^+\mathbf{D} \quad (4)$$

where  $(\mathbf{S}^T)^+$  and  $\mathbf{C}^+$  are the Moore-Penrose pseudo-inverse matrices of the spectra matrix  $\mathbf{S}^T$  and  $\mathbf{C}$ , respectively (which are equal to  $\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}$  and to  $(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$  if  $\mathbf{S}^T$  and  $\mathbf{C}$  are full rank matrices). This is implemented in an alternating least squares cycle where new  $\mathbf{C}$  and  $\mathbf{S}^T$  estimates are obtained in each iteration. Within each ALS step, constraints can be optionally applied to modify the initial profiles towards chemically meaningful ones.

5. Convergence of the algorithm is checked by evaluating the relative change in the lack of fit of the reconstructed matrix  $\mathbf{D} = \mathbf{C}\mathbf{S}^T$  between two consecutive iterations. When this difference is below a preset threshold value the iterative optimization stops.

For the present problem the following constraints were applied in MCR-ALS step 4 (Fig. 3):

- Non-negativity constraint** – Concentrations of chemical species are always values equal to or larger than zero. Therefore non-negativity has been applied to the set of concentration profiles [24,25]. Non-negativity was not applied to spectral profiles because SNV-corrected spectra contain negative values.
- Selectivity and local rank constraint** – This constraint is applied when some species are absent in a particular sample (e.g. samples with 0% paracetamol). This constraint can be used within a matrix or subset in a multiset structure and increases the accuracy in the definition of the profiles [17,24].
- Correspondence among species** – This constraint is only used if the data are structured as a multiset with  $ns$  subsets. Correspondence among species defines the number and identity of components in each single matrix along the augmented data set. This information is usually introduced in the algorithm through a binary coded matrix of size  $ns$  by  $nc$ , which by 1 and 0 indicates if a species is present or absent in the related subset [26]. In the present data, the multiset consists of two subsets (samples without and with blister interference), and via the coding matrix the absence and presence of blister contribution is set.
- Correlation constraint** – In quantitative analysis a correlation constraint can be applied to the concentration profiles [5,6,27,28]. The data is divided into a calibration set and a test set. The initial step when applying a correlation constraint is to create a mask matrix  $\mathbf{C}^*$ , of size  $n$  by  $nc$  (similar to the  $\mathbf{C}$  matrix), with real numbers equal to known concentration values in the positions to be constrained (i.e. concentration values of the calibration set) and missing values in the positions left unconstrained (the test set samples). In each MCR-ALS iteration, the concentrations estimated by ALS for the calibration set are regressed against the known values in  $\mathbf{C}^*$  (reference concentrations). For a particular chemical component, a local linear regression model between  $\mathbf{c}$  and  $\mathbf{c}^*$  is estimated by least squares:

$$\mathbf{c} = \mathbf{b}\mathbf{c}^* + \mathbf{b}_0 + \mathbf{e}. \quad (4)$$

Since only the calibration set has known values in  $\mathbf{c}^*$ , only those entries (indicated by  $\mathbf{c}_{\text{cal}}^*$ ) can be used in the local model:

$$\mathbf{c}_{\text{cal}} = \mathbf{b}\mathbf{c}_{\text{cal}}^* + \mathbf{b}_0. \quad (5)$$

Next, the model parameters  $\mathbf{b}$  and  $\mathbf{b}_0$  are used to predict the concentrations of the unknown and unconstrained values in the test set:

$$\mathbf{c}_{\text{test}}^* = (\mathbf{c}_{\text{test}} - \mathbf{b}_0) / \mathbf{b}. \quad (6)$$

Each ALS iteration is completed after updating the  $\mathbf{C}$  matrix with the predictions ( $\mathbf{C}_{\text{ALS}}$  is substituted by  $\mathbf{C}_{\text{cal}}^*$  and  $\mathbf{C}_{\text{test}}^*$ ) and new spectral profiles  $\mathbf{S}^T$  are found. This constraint, as all others, can be applied to one or more chemical components (concentration profiles) in the data set. For each one of them, an individual calibration model Eq. (5) would be built. Please note that the fact of using individual calibration models for each compound instead of a global one is of utmost importance, since this allows selecting which compounds must be calibrated (not necessarily all the ones contributing to the measured signal). This one-at-a-time calibration strategy allows keeping the structure of the real bilinear spectroscopic model (equal to the one used in the direct calibration method classical least squares, which builds a global calibration model for all compounds) but enables calibration in the presence of interferences, which are modeled using other softer constraints.

### 3.2. MCR-ALS modified correlation constraint

Until now, the correlation constraint has always been applied in a global context to a single data matrix, where a calibration set and a test set are included. The modified implementation suggested here allows us to work with multiset structures. The major advantage of using a multiset structure, compared to a single data matrix, is the possibility of expanding the amount of data and thereby improving the robustness of the regression results. A multiset can be the combination of subsets of data from different experiments, techniques or data collected under different conditions; the latter is the case for the data in our investigation. Depending on the individual types of data we suggest different ways of using the modified correlation constraint:

- Using a global regression model. This model would include a pooled calibration set of samples from different subsets to predict test samples (again possibly inside different subsets). This is the conventional way to use this constraint.
- Using local regression models for the different subsets. This allows working in more diverse scenarios, such as the following ones:
  - Cases where some subsets are subject to a correlation constraint and others are modeled freely.
  - Cases where different local regression models are applied to each subset of experiments. Within this scenario two different situations can be distinguished.
    - Simultaneous analysis of subsets without matrix effect.* This means that the scaling relationship between concentration and measured response (i.e. between  $\mathbf{C}_i$  and  $\mathbf{D}_i$  in a matrix context) is the same for all  $i$  subsets. In this case, the correlation constraint works as usual, with the exception that local models are applied instead of a global regression model.
    - Simultaneous analysis of subsets with matrix effect.* This means that the relationship between concentration and response (i.e. between  $\mathbf{C}_i$  and  $\mathbf{D}_i$  in a matrix context) is not the same in all subsets. This would mean that updating the concentration profiles by the real concentration values would cause a conflict for the general multiset bilinear model  $\mathbf{D} = \mathbf{C}\mathbf{S}^T$  since the common matrix  $\mathbf{S}^T$  would not have the same scale intensity for each subset. To avoid this situation and to benefit from the multiset analysis, a rescaling step of the concentration values is needed within the ALS optimization. Fig. 4 gives an overview on how the correlation constraint was applied for the case of subsets with matrix effect.

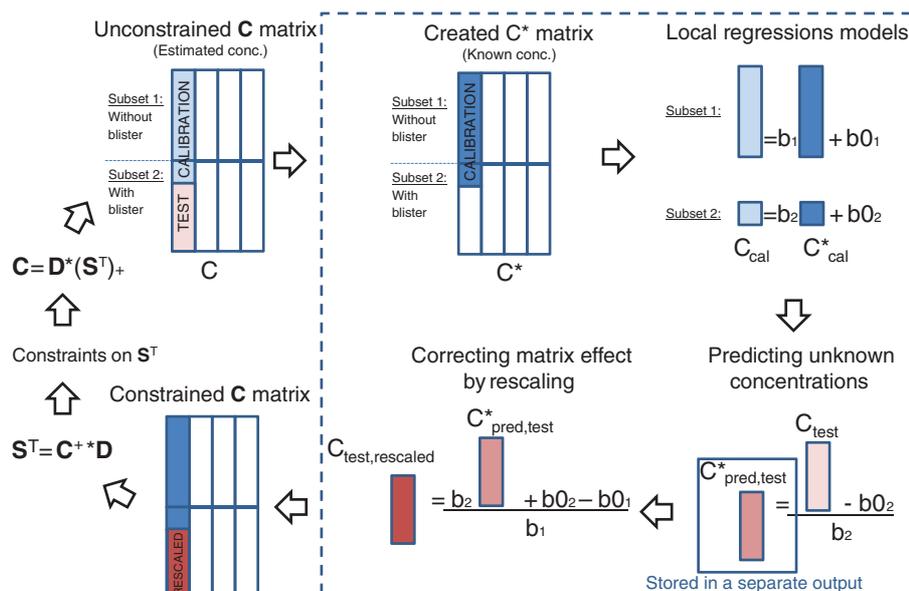
The option b2 is well suited for the data of our research because the multiset formed by two subsets (sub-matrices) – calibration samples without blister  $\mathbf{D}_0$  and samples with blister interference  $\mathbf{D}_b$  – presents a matrix effect. The correlation constraint is modified so that instead of making one single linear regression between concentrations from the ALS and known values, two local regression models are made: one for the calibration samples without blister (Eq. (7)) and one for the calibration samples with blister (Eq. (8)):

$$\mathbf{c}_{\text{cal}} = \mathbf{b}_1\mathbf{c}^* + \mathbf{b}_{0,1}. \quad (7)$$

$$\mathbf{c}_{\text{cal}} = \mathbf{b}_2\mathbf{c}^* + \mathbf{b}_{0,2}. \quad (8)$$

To predict the concentration of a test set/unknown sample with blister interference the second local model (Eq. (8)) is used.

As mentioned before, the two subsets ( $\mathbf{D}_b$  with and  $\mathbf{D}_0$  without blister interference) do not have the same relationship between concentration values and measured responses (i.e. there is a matrix effect due to the reduction of signal intensity caused by the presence of the blister). This means that the two local regression models are different in slope and offset, which would lead to  $\mathbf{S}^T$  matrices with different



**Fig. 4.** Correlation constraint applied on multiset structure. An unconstrained  $C$  matrix is entering the broken line box, where local linear regression models between known concentrations and estimated ALS concentrations are conducted for each subset. Concentrations of test set samples are predicted with one of the local models and stored in a separate output. The relationship between concentration values and measured signals are not the same for each subset. In order not to lead to  $S^T$  matrices with different signal intensities, the test set predictions are thus rescaled according to one of the local models taken as the reference model (this case the model for subset 1). After rescaling, the output of the box is a constrained  $C$  matrix, and a new  $S^T$  can be found based on all available data, and a new iteration can be initiated.

signal intensities. To correct for this and still keep the real predicted concentrations, the predicted concentrations from the local regression models are stored in a separated output variable and the values of concentration profiles are all rescaled according to one of the local models taken as the reference model. In this case, the concentration values of samples with blister would be rescaled as follows:

$$c_{\text{test, rescaled}}^* = (b_2 c_{\text{test}}^* + b_{0,2} - b_{0,1}) / b_1. \quad (9)$$

These concentrations are then used for the update of  $C$ , and finally this ALS iteration is completed. A new  $S^T$  can then be found based on all available data and a new iteration can be initiated. The real concentration values for the blister samples ( $c_{\text{test}}^*$ ) are stored in a separate output argument and are recovered at the end of the MCR-ALS optimization.

## 4. Results and discussion

### 4.1. Spectral information on mixture compounds and influence of blister package

Each sample in our investigation is a mixture of three compounds: paracetamol, starch and sucrose. Fig. 5 shows the Raman spectra of each of these compounds plus the spectrum of the mixture of the three, where spectral characteristics from each of the three compounds are identifiable. Paracetamol has some high intensity bands at approximately  $840 \text{ cm}^{-1}$  and from  $1550$  to  $1670 \text{ cm}^{-1}$  originating from out-of-plane C–H bend and Amide I and Amide II bands, respectively [29]. Starch shows a relevant band around  $480 \text{ cm}^{-1}$  that can be assigned to ring vibration [30]. In the mixture spectrum this band can also be identified. Sucrose shows molecular vibrations at  $824 \text{ cm}^{-1}$ , CH deformation of 1,3,5 tri substituted benzene,  $1119 \text{ cm}^{-1}$ , C–O stretching of C–O–H and  $1445 \text{ cm}^{-1}$ ,  $\text{CH}_2$  scissoring vibration [31]. The amount of fluorescence background differs for the pure spectra of the different compounds. Both sucrose and starch contain a considerable fluorescence originating from the functional group C–O–C (ether linkage), which is part of the six-membered

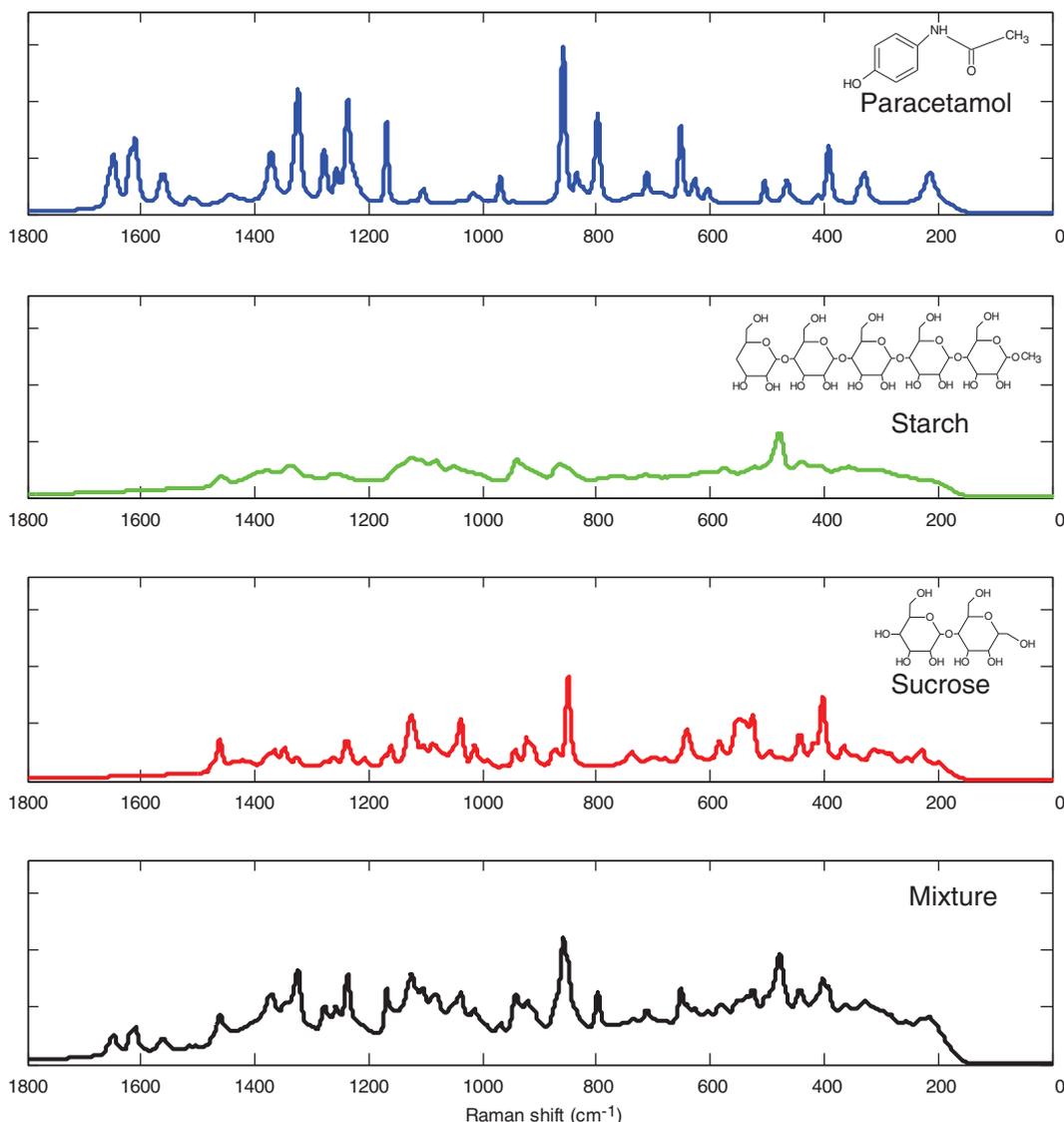
ring structure in both molecules [32]. Even though spectral features from each mixture compound can be recognized in the mixture spectrum, it should be noticed that some spectral bands are overlapping or found in more than one compound. Therefore assigning the spectral bands of this mixture spectrum to different compounds and further quantification based on single bands would lead to erroneous results. An algorithmic method to separate the contributions of the pure compounds into the mixture spectra needs to be applied. A second challenge in this study is that each mixture was measured through a blister package. The blister pack is made of white colored polyvinyl chloride (PVC) and, as shown in Fig. 6, PVC has a large fluorescence background and distinct bands at  $640$ ,  $692$  and  $1000 \text{ cm}^{-1}$ . The figure also shows how the interference from the blister lowers the signal from the mixture compounds, and makes spectral assignment even more challenging.

### 4.2. Quantitative analysis of paracetamol

The API in the ternary mixture is paracetamol. Hence, quantification of this compound was the main focus of our study, where two situations/scenarios are considered (Section 4.2.1) the case where only one type of blister packing material is involved and (Section 4.2.2) the case where variation in blister packing material is present.

#### 4.2.1. Quantification of paracetamol through a blister package

The concentrations of paracetamol in mixtures measured through a blister package were predicted applying two different multivariate methods – PLS and MCR-ALS – to the Raman spectra. In order to use MCR-ALS for quantification, a regression step during the ALS iterative optimization was introduced as defined in the theory section. The results in Fig. 7 show the actual vs. the predicted concentrations of paracetamol for models with 36 samples in the calibration set and 18 in the test set (replicates of samples were kept in either the calibration or the test). With both methods – PLS and MCR-ALS – it was possible to predict the contents of paracetamol through the blister package with acceptable model and prediction errors. MCR-ALS is shown to have a somewhat lower RMSEP of  $0.8 \text{ w/w\%}$ . Several studies [33–35] have shown the potential of measuring through blister packages by



**Fig. 5.** Raman spectra of the three mixture compounds; a) paracetamol, b) starch, c) sucrose and d) ternary mixture at center point in design (starch 67% w/w, sucrose 27% w/w and paracetamol 7% w/w).

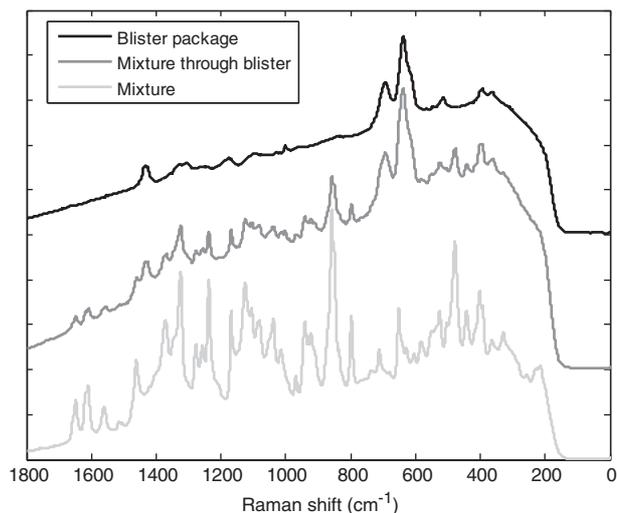
Raman spectroscopy, but these focus on the qualitative aspect, while none reported a quantitative evaluation. A study by Fransson et al. [36] has shown direct (not through blister package) quantitative determination of paracetamol in tablets by Raman spectroscopy. In this study the content of paracetamol was in the range 75–125% of the nominal value and they obtained RMSEP values in the range of 2.4–3.4%, giving a relative error of (RMSEP over range) of ~5–6%. Although not directly comparable, in the present study a relative error of ~4% was obtained (range: 0–20% w/w).

#### 4.2.2. Quantification of paracetamol through a blister, where multiple blister materials are considered

In this study each mixture in the design was measured through blister packages of three different batch numbers. When inspecting the pure spectra of the blister packages they appear different (results not shown), and this will thus induce an error when trying to quantify paracetamol in mixtures measured through three different blister packages with one classical PLS or MCR-ALS model. In Fig. 8 the concentrations of paracetamol found from classical MCR-ALS (using six components, one for each of the chemical compounds and one for each of the blisters) as a function of the actual concentrations are

given. The figure shows clearly the systematic error/matrix effect (differences in slopes corresponding to different blisters). The small batch-to-batch change in the blisters could be due to a difference in thickness or addition of a different (amount of) excipient or color agent to the PVC. In a real production setting the blister package may change from one batch to the next, and one obvious strategy is to build a calibration model on unshielded samples (thus without blister interference). But when using the model based on unshielded samples to predict samples measured through blister a systematic prediction error occurs. This systematic prediction error occurs no matter if PLS or MCR-ALS is applied. The performance of the MCR-ALS model (referred to as “classical” from now on) is summarized in Table 1.

As described in the theory section, there is a possibility to work with multiset structures in MCR-ALS, which could be beneficial in the modeling of interferences (e.g. blister material variations). In a multiset structure where the unshielded mixture samples are seen as one subset, and samples measured through blister are seen as another subset, there is a possibility to add information on the absence/presence of one particular species, the blister package. By explicitly including this information in one of the model components in the MCR-ALS, the blister interference



**Fig. 6.** Spectral effect of measuring through a tablet blister package. Raman spectra of blister package 3 (black), the center point mixture through blister package 3 (dark gray) and the center point mixture (starch 67% w/w, sucrose 27% w/w and paracetamol 7% w/w) (light gray). Individual offsets are added to the spectra for better visualization.

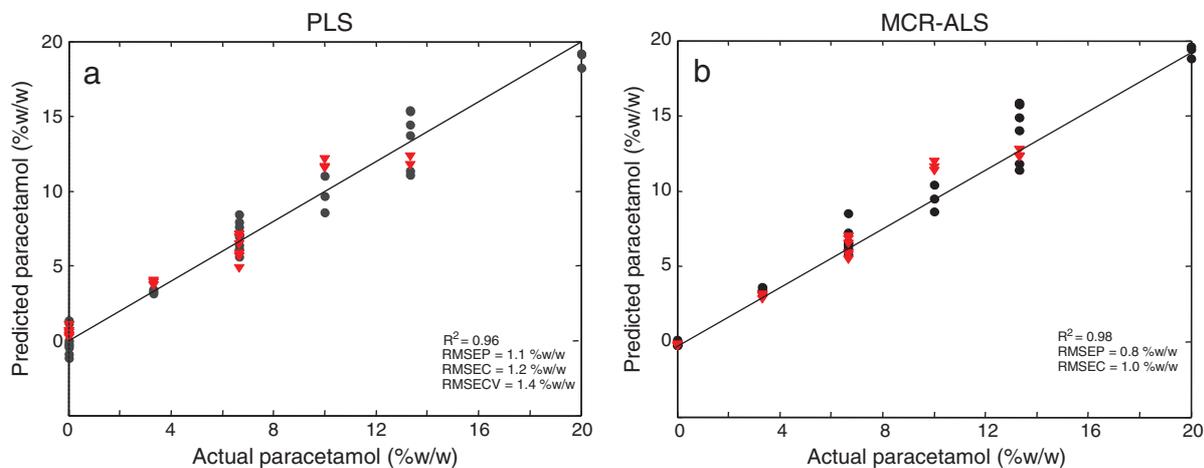
will become better represented and modeled by this component, and thereby better separated from the other true chemical MCR-ALS components. This will potentially lead to better resolution of the three mixture constituents and the blister interference. In this case, it makes sense to work with the multiset for a better separation of the blister from the rest of compounds and apply the regression model only to the blister subset (only one local model for the blister samples and let the matrix of unshielded samples unconstrained). This would correspond to the new correlation constraint option 2.a. The performance of a model using multiset approach option 2.a is shown in Table 1. In the other option of the modified correlation constraint step of the MCR-ALS cycle (option 2.b2), the linear regression was split up, where one regression was based on the calibration samples without blister interference and the second was based on the calibration samples with blister interferences. A test set of samples with blister interferences was predicted using the local regression model. In Table 1 the performance of the MCR-ALS models for three different blister packages is summarized. The table shows that using a multiset approach decreases the RMSEP considerably and if the predictions are slope- and intercept-

corrected (matrix effect corrected), the prediction errors can be decreased even further. Moreover, the table shows the effect of varying the number of concentration levels and in-blister samples in the calibration set for a model where the correlation constraint is applied to a multiset with matrix effects (option 2.b2). The result shows that for two of the three blister packages the errors are similarly low no matter if two or five concentration levels are included. The MCR-ALS calibration model based on 51 unshielded and six shielded samples in two concentration levels of paracetamol (the lowest and highest level) is shown in Fig. 9. This result proves that a calibration model that can handle considerable spectral differences due to the absence or presence of blister is established. It only requires measurements through a blister of samples at two concentration levels. In the study by Fransson et al. [36] a similar conclusion was reached where a calibration model based on only two concentration levels was sufficient.

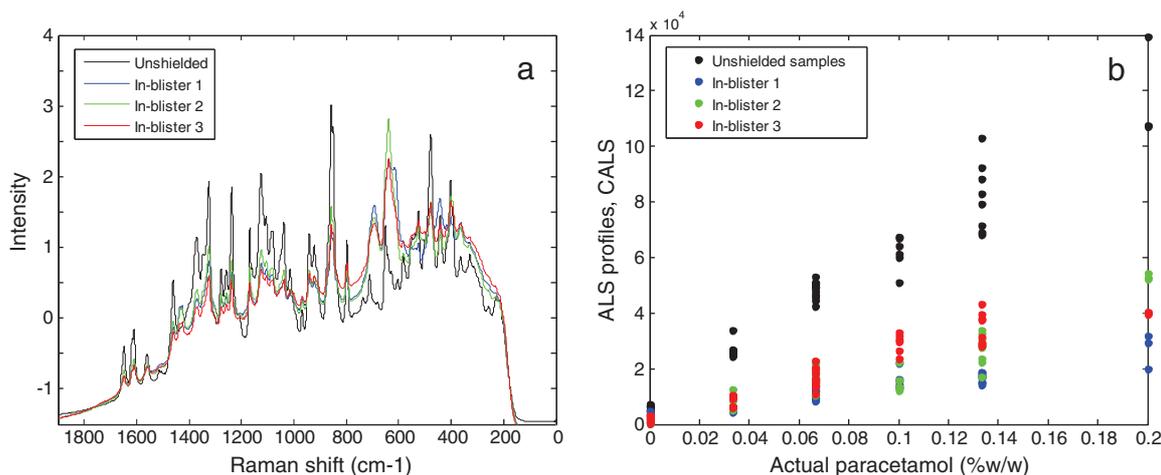
In PLS regression it is not possible to work with a multiset structure in the same way as in MCR-ALS. Blister interference may not be kept isolated in one latent variable, and will thereby be present in all latent variables and potentially influence the model predictions. Even though the MCR-ALS seems to work better than the PLS for this particular quantification case PLS is often the method used, because its background and use is more generally known. However, regarding the recovery of qualitative information (spectral) of the analyzed compound and possible interferences, MCR-ALS has an advantage over PLS as concluded in a quantitative study by Azzouz and Tauler [27].

The matrix effect was in our calibration task caused by variation in blister material, but there are many other external parameters that could cause the same type of matrix effect e.g. temperature variations, color differences and instrumental variations. In this study we have presented the modified correlation constraint in MCR-ALS as a way to overcome a matrix effect. Another approach to solve a matrix effect is external parameter orthogonalisation (EPO) PLS presented by Roger et al. [8]. EPO-PLS is designed to work for external parameters which are quantitative e.g. temperature effects, and not as in our case variation of a more qualitative character (different excipients in blister material).

Our results show that a calibration model to predict paracetamol with low prediction error could be obtained based on a simple calibration set of unshielded (or universal) samples combined with a very small set of samples measured through blister. This suggests a great potential for Raman spectroscopy as a rapid method for quality monitoring and control of pharmaceutical mixtures inside blisters with only a minor effort in measurements. The advantage of measuring



**Fig. 7.** Partial least squares (PLS) and multivariate curve resolution-alternating least squares (MCR-ALS) predictions of paracetamol through a blister package ( $D_{b3}$ ) vs. the actual concentrations (% w/w). ● Calibration set (36 samples) and ▼ test set (18 samples). a) PLS model based on one latent variable b) MCR-ALS model based on four components.



**Fig. 8.** The effect of measuring paracetamol through different blister packages. a) Mean standard normal variate (SNV) corrected Raman spectra of mixtures measured unshielded and in blisters 1–3. b) Predicted paracetamol concentration scores from unconstrained multivariate curve resolution-alternating least squares (MCR-ALS) using six components vs. the actual concentrations.

a product through packing materials makes it possible to do a rapid product verification of the composition of the product actually inside the package, without the need to physically break the seal and open the package. A wide range of studies have already showed how Raman spectroscopy can be used to detect counterfeit medicines through packaging [35,37,38]. Furthermore, storage studies of products inside a blister may be another field where this measurement setup can be applied.

This study demonstrated the possibility of quantification through a blister package. Studies have shown how Raman spectroscopy can detect through a range of other packaging materials such as plastic containers [39], bottles [37] and plastic bags [40]. Therefore it could be interesting to test if our quantification approach works for other types of packaging materials which would expand the application range even further.

## 5. Conclusions

The main conclusion of this study was that it was possible to predict the concentration of paracetamol in a ternary mixture through a

blister package by means of Raman spectroscopy. To the authors' knowledge this is the first study to quantify paracetamol through blister packaging. Calibration models from two multivariate methods PLS and MCR-ALS performed similarly well when only samples from the same blister batch are taken into account.

In addition, we propose a modification of the correlation constraint in MCR-ALS as a way to improve quantification of a compound in a dataset with varying interferences. The interference in the present study was caused by variations in blister material, and it gave rise to a difference in the relationship between concentration and spectral response. Thus, it was not possible to build a global calibration model using neither PLS nor MCR-ALS. The solution to this problem was the use of a multiset structure and the novel modification of the correlation constraint. On the one hand, using a multiset structure with samples in and out of the blister helps to model the blister contribution much better. On the other hand, the modification of the correlation constraint adapted to multiset analysis, enabled the possibility to model the samples with and without blister material interference in local regression models. Further optimization by adding a rescaling step of the concentration values was also implemented in ALS optimization to account

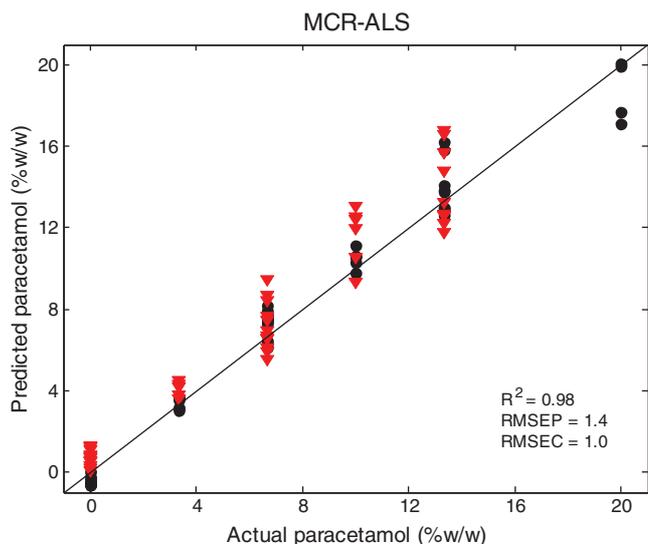
**Table 1**

Overview of MCR-ALS model statistics/performance for predicting paracetamol through tablet blister packages. The table shows MCR-ALS model description, number and type of samples in calibration and test set, number of components, blister package number, coefficient of correlation, root mean square error of calibration (RMSEC) or cross validation (RMSECV) and the root mean square error of prediction (RMSEP).

Model description	Calibration set no. of samples/ conc. levels		Test set no. of samples	No. of components	Blister package no.	R <sup>2</sup>	RMSEC/RMSECV (% w/w)	RMSEP (% w/w)
	D <sub>0</sub> <sup>a</sup>	D <sub>b</sub> <sup>b</sup>						
Classical	51/6	–	54	4	1	0.49	1.0	7.2
Multiset approach (option 2.a) Subset 1 (unshielded samples) modeled without the correlation constraint	51/6	6/2	48	4	1	0.61	5.2	3.3
Multiset approach (option 2.b2) Both subsets modeled with correlation constraints and a rescaling step is applied	51/6	6/2	48	4	1	0.96	1.4	1.3
					2	0.98	1.0	1.9
					3	0.98	1.0	1.4
		9/3	45	4	1	0.96	1.3	1.4
					2	0.97	1.1	1.9
					3	0.98	1.0	1.4
		12/4	42	4	1	0.95	1.4	1.3
					2	0.97	1.2	1.7
				3	0.97	1.0	1.4	

<sup>a</sup> D<sub>0</sub>: mixtures measured unshielded.

<sup>b</sup> D<sub>b</sub>: mixtures measured through blister.



**Fig. 9.** Multivariate curve resolution-alternating least squares (MCR-ALS) predictions of paracetamol through blister package 3 ( $D_{63}$ ) vs. the actual concentrations (% w/w). ● Calibration set was based on 51 unshielded samples (in six concentration levels) and six samples measured through blister (in two concentration levels – highest and lowest) and the predicted ▼ test set consisted of 48 samples measured through blister.

for the matrix effect caused by the presence of the blister. Using this approach paracetamol concentration could be predicted through blister package despite variation in packing material.

#### Acknowledgment

LBL would like to thank the innovation consortium “QbD – Quality by Design in the Food and Pharmaceutical Industries” ([www.qbd.dk](http://www.qbd.dk)) for financially supporting this study. AJ acknowledges funding from the Spanish government through grant CTQ2012-11572-02.

#### References

- [1] K.A. Bakeev, *Process Analytical Technology, Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Wiley, Chichester, West Sussex, 2010.
- [2] M.J. Pelletier, *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford, 1999.
- [3] P.X. Zhang, D. Littlejohn, Interference assessment and correction in the partial least squares regression method for multicomponent determination by UV spectrophotometry, *Chemometrics and Intelligent Laboratory Systems* 34 (1996) 203–215.
- [4] R. Bro, Multivariate calibration – what is in chemometrics for the analytical chemist? *Analytica Chimica Acta* 500 (2003) 185–194.
- [5] M.C. Antunes, J.E.J. Simao, A.C. Duarte, R. Tauler, Multivariate curve resolution of overlapping voltammetric peaks: quantitative analysis of binary and quaternary metal mixtures, *Analyst* 127 (2002) 809–817.
- [6] H.C. Goicoechea, A.C. Olivieri, R. Tauler, Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data, *Analyst* 135 (2010) 636–642.
- [7] S. Romero-Torres, J. Perez-Ramos, K. Morris, E. Grant, Raman spectroscopy for tablet coating thickness quantification and coating characterization in the presence of strong fluorescent interference, *Journal of Pharmaceutical and Biomedical Analysis* 41 (2006) 811–819.
- [8] J. Roger, F. Chauchard, V. Bellon-Maurel, EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits, *Chemometrics and Intelligent Laboratory Systems* 66 (2003) 191–204.
- [9] P. Eilers, B. Marx, Multivariate calibration with temperature interaction using two-dimensional penalized signal regression, *Chemometrics and Intelligent Laboratory Systems* 66 (2003) 159–174.
- [10] J.D. Rodriguez, B.J. Westenberger, L.F. Buhse, J.F. Kauffman, Standardization of Raman spectra for transfer of spectral libraries across different instruments, *Analyst* 136 (2011) 4232–4240.
- [11] A. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *Trac-Trends in Analytical Chemistry* 28 (2009) 1201–1222.
- [12] N.K. Afseth, V.H. Segtnan, J.P. Wold, Raman spectra of biological samples: a study of preprocessing methods, *Applied Spectroscopy* 60 (2006) 1358–1367.
- [13] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Applied Spectroscopy* 43 (1989) 772–777.
- [14] S. Wold, A. Ruhe, H. Wold, W.J. Dunn, The collinearity problem in linear-regression – the partial least-squares (PLS) approach to generalized inverses, *SIAM Journal on Scientific and Statistical Computing* 5 (1984) 735–743.
- [15] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems* 58 (2001) 109–130.
- [16] R. Tauler, A. Izquierdoridorsa, R. Gargallo, E. Casassas, Application of a new multivariate curve resolution procedure to the simultaneous analysis of several spectroscopic titrations of the copper(II)-polyinosinic acid system, *Chemometrics and Intelligent Laboratory Systems* 27 (1995) 163–174.
- [17] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution, *Journal of Chemometrics* 9 (1995) 31–58.
- [18] S. Navea, A. de Juan, R. Tauler, Modeling temperature-dependent protein structural transitions by combined near-IR and mid-IR spectroscopies and multivariate curve resolution, *Analytical Chemistry* 75 (2003) 5592–5601.
- [19] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometrics and Intelligent Laboratory Systems* 76 (2005) 101–110.
- [20] A. de Juan, R. Tauler, Multivariate curve resolution (MCR) from 2000: progress in concepts and applications, *Critical Reviews in Analytical Chemistry* 36 (2006) 163–176.
- [21] S. Piqueras, J. Burger, R. Tauler, A. de Juan, Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution, *Chemometrics and Intelligent Laboratory Systems* 117 (2012) 169–182.
- [22] E.R. Malinowski, Window factor-analysis – theoretical derivation and application to flow-injection analysis data, *Journal of Chemometrics* 6 (1992) 29–40.
- [23] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, *Analytical Chemistry* 63 (1991) 1425–1432.
- [24] R. Manne, On the resolution problem in hyphenated chromatography, *Chemometrics and Intelligent Laboratory Systems* 27 (1995) 89–94.
- [25] R. Bro, S. DeJong, A fast non-negativity-constrained least squares algorithm, *Journal of Chemometrics* 11 (1997) 393–401.
- [26] R. Tauler, D. Barcelo, Multivariate curve resolution applied to liquid-chromatography diode-array detection, *Trac-Trends in Analytical Chemistry* 12 (1993) 319–327.
- [27] T. Azzouz, R. Tauler, Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples, *Talanta* 74 (2008) 1201–1210.
- [28] S.E. Richards, E. Becker, R. Tauler, A.D. Walmsley, A novel approach to the quantification of industrial mixtures from the vinyl acetate monomer (VAM) process using near infrared spectroscopic data and a quantitative self modeling curve resolution (SMCR) methodology, *Chemometrics and Intelligent Laboratory Systems* 94 (2008) 9–18.
- [29] H. Moynihan, I. O'Hare, Spectroscopic characterisation of the monoclinic and orthorhombic forms of paracetamol, *International Journal of Pharmaceutics* 247 (2002) 179–185.
- [30] M.R. Almeida, R.S. Alves, L.B.L.R. Nascimbem, R. Stephani, R.J. Poppi, L.F.C. de Oliveira, Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis, *Analytical and Bioanalytical Chemistry* 397 (2010) 2693–2701.
- [31] L. Silveira Jr., L.M. Moreira, V.G.B. Conceicao, H.L. Casalechi, I.S. Munoz, F.F. da Silva, M.A.S.R. Silva, R.A. de Souza, M.T.T. Pacheco, Determination of sucrose concentration in lemon-type soft drinks by dispersive Raman spectroscopy, *Spectroscopy-an International Journal* 23 (2009) 217–226.
- [32] H.J. He, R.P. Yu, T. Zhu, Z.B. Gu, H. Xu, Study of fluorescence spectra of starch suspension, *Guang Pu Xue Yu Guang Pu Fen Xi* 26 (2006) 1636–1639.
- [33] P. Matousek, A.W. Parker, Bulk Raman analysis of pharmaceutical tablets RID G-5445-2011, *Applied Spectroscopy* 60 (2006) 1353–1357.
- [34] C. Eliasson, M. Claybourn, P. Matousek, Deep subsurface Raman spectroscopy of turbid media by a defocused collection system, *Applied Spectroscopy* 61 (2007) 1123–1127.
- [35] W.J. Olds, E. Jaatinen, P. Fredericks, B. Cletus, H. Panayiotou, E.L. Izake, Spatially offset Raman spectroscopy (SORS) for the analysis and detection of packaged pharmaceuticals and concealed drugs, *Forensic Science International* 212 (2011) 69–77.
- [36] M. Fransson, J. Johansson, A. Sparen, O. Svensson, Comparison of multivariate methods for quantitative determination with transmission Raman spectroscopy in pharmaceutical formulations, *Journal of Chemometrics* 24 (2010) 674–680.
- [37] C. Eliasson, P. Matousek, Noninvasive authentication of pharmaceutical products through packaging using spatially offset Raman spectroscopy, *Analytical Chemistry* 79 (2007) 1696–1701.
- [38] C. Ricci, C. Eliasson, N.A. Macleod, P.N. Newton, P. Matousek, S.G. Kazarian, Characterization of genuine and fake artesunate anti-malarial tablets using Fourier transform infrared imaging and spatially offset Raman spectroscopy through blister packs, *Analytical and Bioanalytical Chemistry* 389 (2007) 1525–1532.
- [39] C. Eliasson, N.A. Macleod, P. Matousek, Non-invasive detection of powders concealed within diffusely scattering plastic containers, *Vibrational Spectroscopy* 48 (2008) 8–11.
- [40] H. Tsuchihashi, M. Katagi, M. Nishikawa, M. Tatsuno, H. Nishioka, A. Nara, E. Nishio, C. Petty, Determination of methamphetamine and its related compounds using Fourier transform Raman spectroscopy, *Applied Spectroscopy* 51 (1997) 1796–1799.

# PAPER III

---

Lyndgaard, L.B., Spångberg, R., Gilmour, C., Lyndgaard, C.B. & van den Berg, F.

A process analytical approach for quality control of dapivirine  
in HIV preventive vaginal rings by Raman spectroscopy

*Accepted to Journal of Raman Spectroscopy, 2013*

## **A process analytical approach for quality control of dapivirine in HIV preventive vaginal rings by Raman spectroscopy**

Lotte B. Lyndgaard<sup>1\*</sup>, Rolf Spångberg<sup>2</sup>, Christopher Gilmour<sup>3</sup>, Christian B. Lyndgaard<sup>1</sup> and Frans van den Berg<sup>1</sup>

<sup>1</sup>Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark; <sup>2</sup>Qpharma, Malmö, Sweden; <sup>3</sup>International partnership of microbicides, Silver Spring, MD, USA

\*Corresponding author. Quality and Technology, Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark. Email: [lottebs@food.ku.dk](mailto:lottebs@food.ku.dk)

## **Abstract**

An estimated 34 million people are living with HIV worldwide. A polymeric drug delivery device for women has been developed to help reduce male to female vaginal transmission of HIV-1. The device is an intravaginal slow-release ring containing the antiretroviral drug dapivirine as the API. In all pharmaceutical delivery systems API content is a critical control point. Conventional, quantification of the API in vaginal rings is a stepwise, destructive and time-consuming process involving solvent extraction and high pressure liquid chromatography. This study investigates the potential of Raman spectroscopy for fast and non-destructive quantification of dapivirine in intravaginal rings. Wide-area illumination measurements were carried out on rotating rings using a custom build ring spinner. Customized reference rings of known concentrations were used to build calibration models, and the models were verified by rings from production using the established method as reference analysis. Bi- and multivariate calibration methods were applied. Models based on band ratios and PLS regression models performed similarly well resulting in low model and prediction errors. Next to an alternative reference analysis for quality assurance Raman can also be used as a production process performance analysis and optimization tool due to its non-destructive nature and speed of analysis. Using measurements on connected spots over the entire ring circumference the within ring variation in API was determined. ANOVA showed that there was a statistical difference in API distribution over the rings, information that can be of use in process optimization, for example. However, Tukey's HSD test proved that no regions of the ring were out of specifications, indicating a stable production system.

**Keywords:** Raman Spectroscopy, Vaginal rings, Medical device, Multivariate calibration, Quantification

## Introduction

It is projected that 34 million people live with the human immunodeficiency virus (HIV) worldwide; the vast majority are in low- and middle-income countries. An estimated 2.5 million people were newly infected with the virus in 2011. Globally HIV is the leading cause of death for women in the age range 15-44. The problem is especially widespread in sub-Saharan Africa, where young women are at least twice as likely to become infected as young men.<sup>[1]</sup> Women are particularly vulnerable to infection due to a mix of biology and social factors, and they urgently need tools to protect themselves. For this purpose a newly developed HIV preventive vaginal ring has shown promising results. An intravaginal ring (IVR) is a polymeric drug delivery device designed to provide controlled release of drugs to the vagina over an extended period of time. The International Partnership for Microbicides (IPM) is researching the use of the active pharmaceutical ingredient (API) dapivirine in an IVR. Several phase I and II studies have proven the ring as a safe and viable drug delivery method, with the potential to reduce female vaginal HIV infection rates. The ring is intended for 28 days consecutive use for women as a complementary prevention technology to safe sex practices. The ring formulation is intended to be used continuously by HIV-uninfected women for the prevention of male to female vaginal transmission of HIV-1.<sup>[2-5]</sup>

One of the critical control points in the production of IVR, as with many other medical delivery devices, is the concentration of API. To continuously monitor and control if the concentration of API is maintained within the product specifications is fundamental. The standard analytical method of API content testing for vaginal rings is high pressure liquid chromatography (HPLC) which is a multi-step, destructive, time-consuming and costly process involving API extraction by solvents prior to analysis. This method has been applied in numerous IVR studies in e.g. stability testing and release profiling.<sup>[6]</sup> In recent years the interest in novel analytical methods for pharmaceutical product analysis has increased. This is mainly because of the US Food and Drug Administration (FDA) initiative on process analytical technologies (PAT)<sup>[7]</sup>, advocating rapid and non-destructive analytical methods. However, some processes and drug delivery systems, like IVRs, are particularly challenging, and applied research is needed for the evaluation of novel methods. In the present study, Raman spectroscopy is assessed for characterizing IVRs and quantifying dapivirine. In line with the PAT framework, control charts are proposed in utilizing the responses provided by Raman spectroscopic monitoring. By Raman spectroscopy IVRs can be analyzed directly, without any tedious or time-consuming sample preparation, all non-contact and non-destructively. In Raman

spectroscopy the molecular vibrations in a sample can be measured, thereby providing a unique spectral fingerprint from which the sample compounds can be identified and quantified. The technique has been widely applied in pharmaceutical testing, since it carries several attractive features. The Raman signals from API's are usually high whereas the signals from excipients are typically low due to their electronic structure. This is explained by the fact that Raman spectroscopy measures the polarizability of a molecule - i.e. how easy can the charge distribution (electron cloud) of a molecule be distorted from its normal shape by an external electric field. Many APIs contain aromatic or conjugated domains, which are easily polarized, and therefore show a strong Raman signal. Conversely, aliphatic compounds, such as most excipients, are very weak Raman scatters, since they have strong dipole moments and electrons are therefore not easily polarized.<sup>[8,9]</sup> In the IVR of our study the excipient is a silicone elastomer which does show a large Raman signal. However, the major bands of silicone do not overlap with the bands of the API, and therefore do not affect the quantification of the API. The fact that the Raman signal of silicone is large gives the possibility to obtain additional valuable information from the Raman fingerprint regarding the silicone, such as potential batch-to-batch variations and impurities that might interfere with the product functionality.

It has previously been shown that Raman spectroscopy can be used in qualitatively characterizing a cross section of a silicone elastomer vaginal ring.<sup>[10]</sup> However, no studies have focused on the quantification or developed calibration models to predict API content in entire IVRs. The basis for quantitative Raman spectroscopy lies in the fact that Raman scattering is proportional to the concentration of the substance.<sup>[9]</sup> In the quantitative data analysis of Raman spectra different approaches have been applied: 1) Univariate analysis using single band intensity, 2) Bivariate using the ratio of two bands<sup>[10]</sup>, or 3) multivariate, where thousands of variables are included. Where uni- and bivariate analysis is very intuitive and straightforward to apply, multivariate analysis is more complex and misinterpretation or over-fitting is more likely to occur. However, multivariate analysis is advantageous in aspects like handling of interference, noise reduction, outlier detection and for exploratory analysis. Most often very small spectral changes can be detected and several compounds in a sample can be analyzed simultaneously.<sup>[11]</sup>

The IVR is designed to provide a controlled release of drug and the distribution of API within the ring formulation is known to influence this drug release profile. Changes in the release characteristics may lead to a decreased safety/efficiency, unwanted release-bursts or other side

effects. Determination of the API distribution over the ring is thus a valuable quality tool. So far the only analytical method showing potential for API mapping in IVRs is Raman spectroscopy.<sup>[10]</sup> Raman has been widely used for API mapping of other types of pharmaceuticals such as tables,<sup>[12,13]</sup> transdermal tape,<sup>[14]</sup> solid dispersions,<sup>[15,16]</sup> and transdermal drug delivery device.<sup>[17]</sup>

In this study the objective is to build a calibration model that can determine the dapivirine concentration in IVRs. Our approach was to base the calibration model on customized reference rings with known API concentrations, and validate the model with rings from full scale production and compare to the HPLC reference analysis. The performance of bivariate and multivariate calibration models is evaluated. We show how control charts based on model predictions can be used to evaluate if rings from different production batches stay within the API specification limits. Finally, we investigate the spatial distribution of API within a selection of rings by comparing the API concentration in multiple spots around the ring and testing for significant difference.

## **Materials and methods**

### **The intravaginal rings (IVRs)**

Production rings: Our product consists of 8 g matrix-type vaginal rings containing 25 mg of dapivirine dispersed in a two part platinum-cured silicone matrix. The rings dimensions were: Outer diameter 56mm and a cross sectional chord diameter of 7.7mm (Figure 1). 70 rings of seven different batches from a medium-scale production (Qpharma, Sweden) were used for verification of the calibration model. The rings were made by high temperature injection molding.



**Figure 1.** Intravaginal dapivirine ring. Experimental setup with the Raman pharmaceutical area testing (PhAT) probe placed on top of a ring spinner (3P innovation). Ring is placed in the small cup, which rotate when the drawer is closed.

Customized reference rings: Ten silicone elastomer rings of five different concentrations levels of dapivirine (50, 75, 100, 125 and 150% of nominal content) were custom made, kindly provided by IPM (Maryland, US). Dimensions were the same as for the rings from production. These reference rings were used for establishing a calibration model for the dapivirine concentration.

### **Reference analysis by High Pressure Liquid Chromatography (HPLC)**

Rings were cut in halves along the length using a scalpel, and transferred into a flask with a solution of norethindrone (internal standard) and acetone. The flask was shaken on a mechanical shaker for 24 hours at 180 rpm. Then the extract was evaporated to dryness and the residue re-dissolved in methanol. This was followed by mixing on a whirl mixer for 20 seconds and 2 minutes of ultrasonic treatment. Finally, the extract was transferred into HPLC sample vials. The gradient elution was performed on a reverse-phase, isocratic HPLC system (Waters, Dublin, Ireland) consisting of a Waters Alliance 2695 Separation module with pump and autosampler. The system was equipped with a Waters 2487 Dual  $\lambda$  Absorbance detector (UV) and a Waters 2996 Photodiode Array Detector. The analytical column used was a Luna C18. The injection volume was 10  $\mu$ l, and the mobile phase consisted of 65% methanol and 35% water. A flow rate of 0.75ml.min<sup>-1</sup> and a run time of 12 minutes were used.

## Raman measurements

The Raman spectra were recorded on a RamanRxn1 instrument (Kaiser Optical Systems Inc, MI, USA) equipped with a 785nm near-infrared diode laser (Invictus, Kaiser Optical Systems Inc., MI, USA). A single holographic grating and a thermoelectric cooled charge-coupled device detector, operated at  $-40^{\circ}\text{C}$ , were used. For this experiment the pharmaceutical area testing (PhAT) probe was used because of its large spot size diameter (6 mm) allowing a large portion of the sample to be measured. Figure 1 shows the measurement set-up where the ring was placed in a special build ring spinner (3P Innovation Limited, UK). The PhAT probe was placed on top of the ring spinner, and measurements were carried out spinning the ring at 17 rpm. The spectra were stored as Raman shifts in the range  $1800 - 200 \text{ cm}^{-1}$ . Two types of measurements were carried out: *1. Calibration of Dapivirine concentration*: ten reference rings measured in duplicates were used as calibration set, and ten rings from real production were used as independent test set. All rings were measured in a randomized order using 30 s of exposure time while rotating in the ring spinner. *2. Ring heterogeneity testing of dapivirine concentration*: rings from two different production batches were tested by means of stationary Raman measurements at twenty-four different spots on the ring circumference. By turning the ring in steps of  $15^{\circ}$  ( $= 360^{\circ}/24$ ) the entire ring perimeter was covered. Triplicate measurements of each spot were obtained by manually rotating the ring three times (in steps of  $15^{\circ}$ ). An exposure time of 10 s was used.

## Data analysis

In the present work various preprocessing techniques were tested, but standard normal variate (SNV) scaling proved to be the most suited technique to correct the spectra, providing parsimonious models with the lowest prediction error. Applying SNV, both additive and multiplicative effects in the spectra were corrected.

### *Calibration methods*

In this paper the following three methods are applied to quantify the API:

1. Band ratios

The ratio between the maximum intensity of a API band and a silicone matrix band can be used to quantify drug content in a ring:

$$\text{Ratio} = \frac{\text{maximum intensity of API band}}{\text{maximum intensity of matrix bands}} \quad (1)$$

This computation of ratio will compensate for possible instrumental differences such as laser fluctuations and changes in the instrument-sample interaction such as laser spot positioning on the ring surface.

## 2. Band ratios using band fitting parameters

In Raman spectroscopy both wavenumber shifts and bandwidths can contain relevant chemical and structural information. Band fitting in selected windows was used to locate wavenumber shifts and changes in band shapes. The so-called Cauchy–Lorentz mathematical function<sup>[18]</sup> was applied to describe the bands of the Raman spectrum.

$$x(\sigma) = p_3 \times \frac{1}{\pi} \times \frac{p_2}{(\sigma - p_1)^2 + p_2^2} + p_4 + p_5 \times \sigma \quad (2)$$

In this analysis three different band parameters – band location  $p_1$ , bandwidth  $p_2$ , and band intensity/height  $p_3$  – for each spectral window  $x$  were determined. The first-order polynomial in Equation 2 was added to compensate for baseline offset ( $p_4$ ) and slope ( $p_5$ ) difference and was fitted simultaneously with the other model parameters. Equation 2 was fitted by nonlinear least squares regression on a spectral window segment  $\sigma$  of 100 or 150  $\text{cm}^{-1}$  width (334 or 501 data points) surrounding the spectral band of interest using the Levenberg–Marquardt algorithm<sup>[19,20]</sup> and implemented in MATLAB employing in-house routines. The ratios between band heights determined from band fitting were found by means of Equation 1.

## 3. Partial least squares (PLS) regression

PLS is a commonly used and well described method for multivariate regression.<sup>[21-24]</sup> In PLS regression a dependent variable  $\mathbf{y}$  is regressed on a set of independent variables  $\mathbf{X}$ . The purpose of this is to estimate  $\mathbf{y}$  from future measurements of  $\mathbf{X}$  only, thus requiring a representative calibration

set with known values of  $\mathbf{X}$  and  $\mathbf{y}$  during training. PLS is well suited for spectroscopic data sets for two reasons: 1) they often contain more variables than samples, hence multivariate linear regression (MLR) does not work, and 2) many spectral variables co-vary, which means that the number of variables is much higher than the number of independent phenomena in the data.

In PLS the individual variations in  $\mathbf{X}$  and  $\mathbf{y}$  plus their mutual correlation are simultaneously maximized. This can be considered as finding an outer relation (decomposition of  $\mathbf{X}$  and  $\mathbf{y}$  individually), and an inner relation which links the two blocks together. As in the established method principal component analysis (PCA) the outer relation (decomposition) of two blocks using  $I$  components are found by:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_I\mathbf{p}_I^T + \mathbf{E} = \mathbf{TP}^T + \mathbf{E} \quad (3)$$

$$\mathbf{y} = \mathbf{u}_1\mathbf{q}_1 + \mathbf{u}_2\mathbf{q}_2 + \dots + \mathbf{u}_I\mathbf{q}_I + \mathbf{e} = \mathbf{Uq} + \mathbf{e} \quad (4)$$

The inner relation of the two decompositions is found by relating the  $\mathbf{X}$  and the  $\mathbf{y}$  scores for each component  $i$  and determining the regression coefficient:

$$\mathbf{u}_i = \mathbf{b}_i\mathbf{t}_i \quad (5)$$

$$\mathbf{b}_i = \mathbf{t}_i^T(\mathbf{t}_i^T \mathbf{t}_i)^{-1}\mathbf{u}_i \quad (6)$$

The inner relation is improved by rotating the components, which means that  $\mathbf{t}$  and  $\mathbf{u}$  iteratively changes places in the NIPALS algorithm.<sup>[23,24]</sup> To ensure that the scores are orthogonal loading weights are introduced in the final step of PLS.

#### *Model validation and statistics*

All models were evaluated using segmented cross-validation (replicates kept together in a segment) during the calibration stage. For the PLS regression model the optimal number of components was also determined by the cross-validation. All models were further validated using an independent test set. Model performance was assessed based on root mean square error of cross validation (RMSECV) and the root mean square error of prediction (RMSEP).

#### *Statistical methods*

ANOVA and Tukey's HSD (honestly significant difference) tests were applied to determine if any of the spots on the ring perimeter were significantly different.

## Process validation

The condition of a production process can be determined from calculation of the process capability index:

$$Cp = \frac{UTL-LTL}{6\sigma} \quad (7)$$

where  $\sigma$  expresses the standard deviation of a process under (statistical) control, and UTL and LTL are the upper and lower tolerance limits, respectively. The minimal acceptable value in Equation 7 is 1, while a score of 2 or more is required under the six-sigma rules.<sup>[25]</sup>

In some cases where a process offset is present, and the offset corrected capability index (Equation 8) can be used:

$$Cpk = \min\left(\frac{UTL-mean}{3\sigma}, \frac{mean-LTL}{3\sigma}\right) \quad (8)$$

where mean is the average value of the measurements.

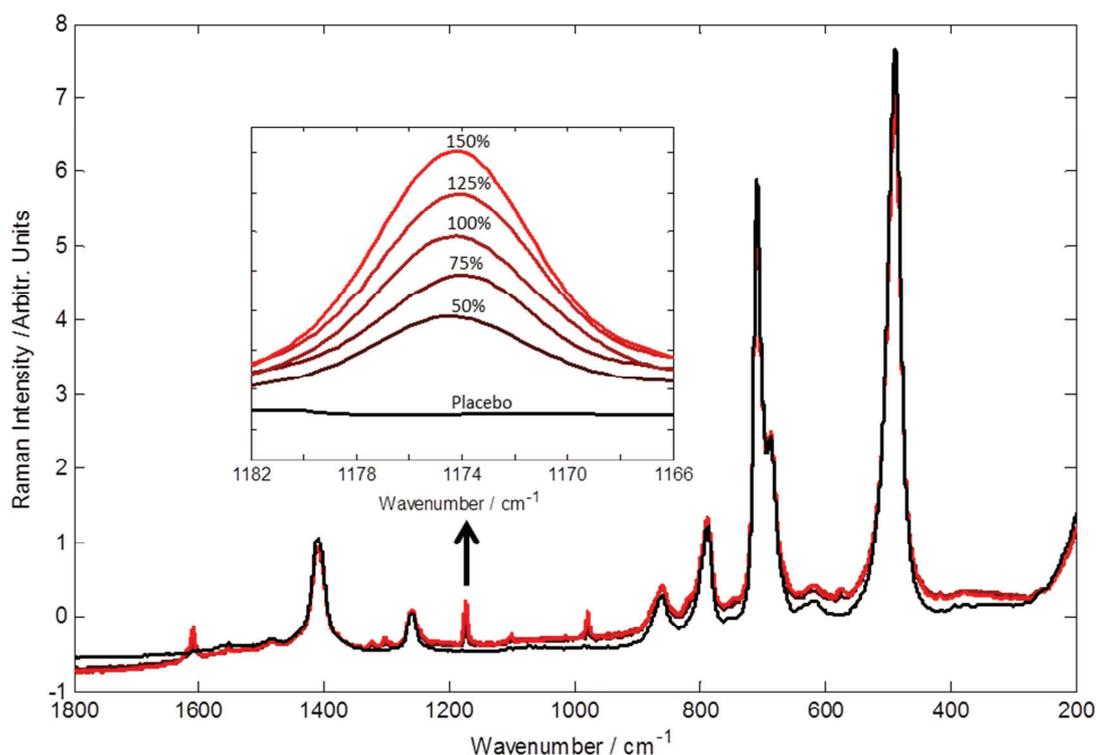
## Software

All multivariate data analyses and processing was carried out using MATLAB R2011b (Mathworks, MA, USA), the PLS toolbox (Eigenvector) and in-house routines ([www.models.life.ku.dk](http://www.models.life.ku.dk)). SAS JMP 9.0.2 (SAS Institute Inc., NC, USA) was used for statistical testing.

## Results and discussion

### Spectral interpretation

Figure 2 shows Raman spectra of intravaginal silicone rings; one placebo and five with varying API (dapivirine) concentrations (50-150 %).

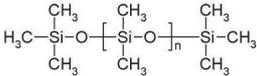
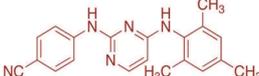


**Figure 2.** Raman spectra 1800-200 cm<sup>-1</sup> of silicone intravaginal rings. The spectra are coloured according to nominal concentration of API (dapivirine). A zoom of one of the major bands originating from API shows systematic increase in intensity with increasing concentration. Band assignments can be found in Table 1.

In the spectrum of the ring containing only silicone elastomer the most intensive band appears at 490 cm<sup>-1</sup> originating from Si-O-Si symmetric stretch. At 710 and 690 cm<sup>-1</sup> the symmetric stretch and rock vibrations of Si-C give rise to two very strong bands, and the Si-(CH<sub>3</sub>)<sub>2</sub> antisymmetric rocking of the methyl groups gives a medium strong band at 790 cm<sup>-1</sup>. At wavenumber 860 cm<sup>-1</sup> the Si-O stretching band is found, and at higher wavenumbers 1260 and 1415 cm<sup>-1</sup> symmetric and antisymmetric deformations of Si-CH<sub>3</sub> appear.<sup>[26-29]</sup> The most intense bands observed in the Raman spectrum of dapivirine appear at 1610, 1175 and 980 cm<sup>-1</sup>. Bands at 1175 and 980 cm<sup>-1</sup> can be assigned to the C-H in-plane bend and the ring breathing of the phenyl ring, respectively. The band around 1610 cm<sup>-1</sup> may be due to C=C ring stretching or C=N stretch.<sup>[30-35]</sup> In Table 1 the

assignments to functional groups and mode of vibration are summarized for both the spectral bands originating from silicone elastomer and dapivirine.

**Table 1.** Assignments of major bands occurring in the Raman spectrum of a silicone based intravaginal ring (Figure 2). The assignments are coloured according to their origin. For each assigned band the position, functional group, mode of vibration and intensity is given.

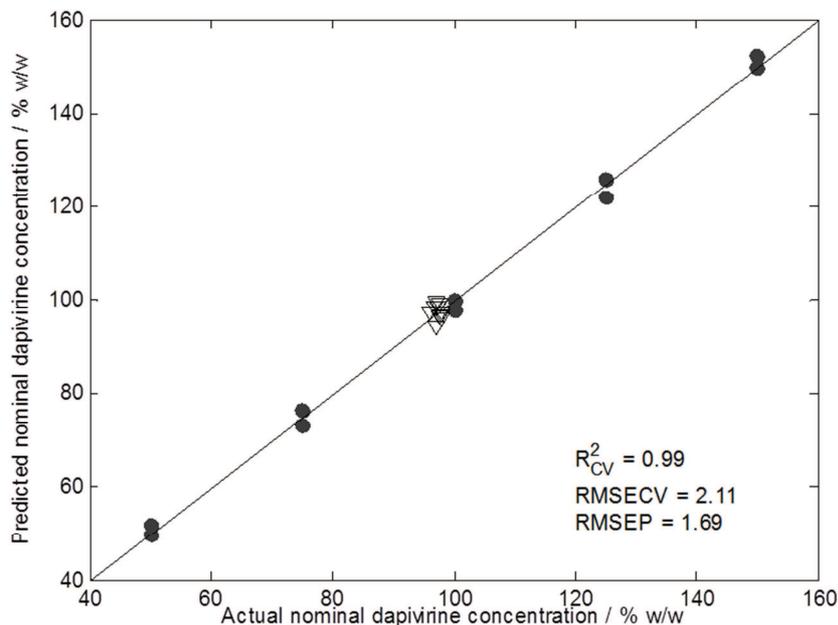
<b>Silicone elastomer:</b>		<b>Dapivirine:</b>	
			
<b>Band position (cm<sup>-1</sup>)</b>	<b>Functional group</b>	<b>Mode of vibration</b>	<b>Intensity</b>
1622	C=N	Stretch	<sup>a</sup>
1610	Phenyl ring	C=C ring stretch	w
1413	Si-CH <sub>3</sub>	Antisym. CH <sub>3</sub> deformation	m
1325	C-N/N-H	Stretch/bend	w
1305	-	-	w
1260	Si-CH <sub>3</sub>	Sym. CH <sub>3</sub> deformation	m
1175	Phenyl ring	C-H in-plane bend	m
1105	-	-	w
980	Phenyl ring	Sym. ring breathing	m
860	Si-O	Stretch	m
820	N-H	Bend	<sup>a</sup>
790	Si-(CH <sub>3</sub> ) <sub>2</sub>	Antisym. CH <sub>3</sub> rock	m
710	Si-C	Sym. Stretch	vs
690	Si-CH <sub>3</sub>	Sym. Rock	vs
620	-	-	w
575	-	-	w
489	Si-O-Si	Sym. Stretch	vs

<sup>a</sup> shoulder

### Calibration of dapivirine concentration

From the spectroscopic interpretation the major API and silicone bands were identified. Due to the well resolved bands and easy assignments of Raman spectra it is appealing to calibrate heights of individual API bands to API concentrations. However, the signal may be affected by small variations in sample properties causing e.g. changes in the depth of light penetration, and normalization of spectral bands can be useful. Therefore ratios between the intensities of the API band at 1175 cm<sup>-1</sup> and the silicone bands at 489 and 1413 cm<sup>-1</sup> (intensities summed) were used to quantify the API in the rings. Prior to the calculation of ratios the spectra were SNV corrected in order to reduce baseline variations. Figure 3 shows the actual vs. the predicted dapivirine

concentrations from a linear regression model between band ratios and actual dapivirine concentrations. The model has a low bias, RMSECV, and RMSEP (Table 2).



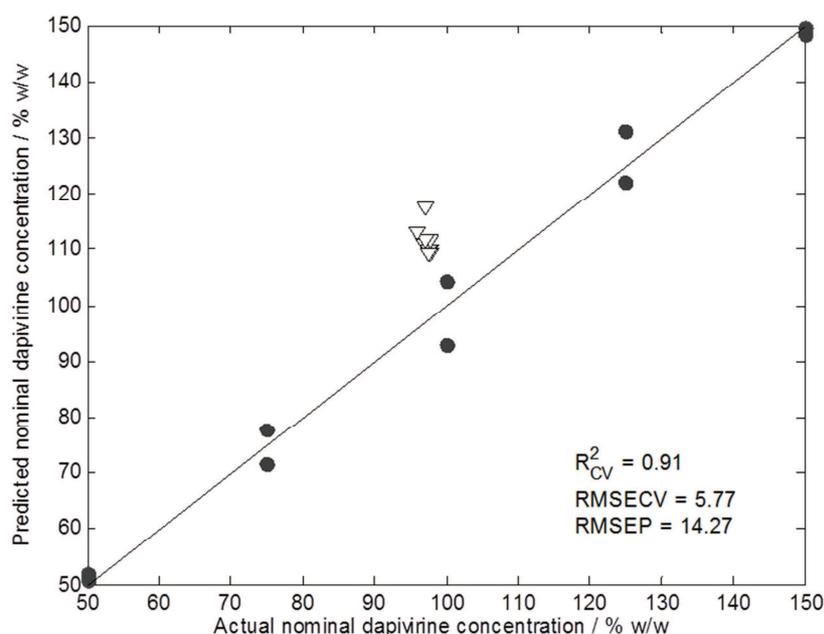
**Figure 3.** Actual vs. predicted dapivirine concentrations in intravaginal rings. Predictions found from band ratios of major API band and silicone bands in the Raman spectra. Calibration set of reference rings ●, and test set of rings from production ▽.

A second approach in establishing a calibration model was to use Cauchy-Lorentz band fitting as described in the previous section. Functions were fitted to the major API and silicone bands of the spectrum using non-linear regression (Equation 2), and thereby band heights were determined. The ratios between these heights were calculated similarly to the previous method. With this approach the baseline is estimated, and fluctuations can therefore be handled without using further spectral pre-processing. Furthermore, parameters like band width and position can also be extracted, which previously have shown to give valuable chemical information.<sup>[36]</sup> A linear regression model between band ratios found from band fitting and the actual dapivirine concentrations showed a little lower RMSECV, but a much higher RMSEP than the previous model (Table 2). The higher RMSEP might be explained by the fact that SNV was not applied prior to the band fitting (as this would distort the band shapes). In SNV correction the mean of the full spectrum is used, and it therefore works in a global way taking the full spectrum into account. This is not the case for band fitting, and the method will thus work more locally by fitting inside specific spectral regions. In the spectra of the production rings there could be some global effects not seen in the calibration, which are corrected for by SNV but not when applying band fitting.

**Table 2.** Overview of the model performances of bi- and multivariate calibration models. For each model the following parameters are given: coefficient of correlation ( $R^2$ ), root mean square error of cross validation (RMSECV) and the root mean square error of prediction (RMSEP). Furthermore the numbers of latent variables (LV) is shown for the PLS models.

Method	LV	$R^2_{CV}$	RMSECV	RMSEP
Band ratios	-	0.99	2.11	1.69
Band fitting parameters (heights) ratios	-	0.99	1.80	6.07
PLS (spectral region 200-1800 $\text{cm}^{-1}$ )	2	0.91	5.77	14.27
PLS (spectral region 800-1800 $\text{cm}^{-1}$ )	2	0.99	1.82	1.02

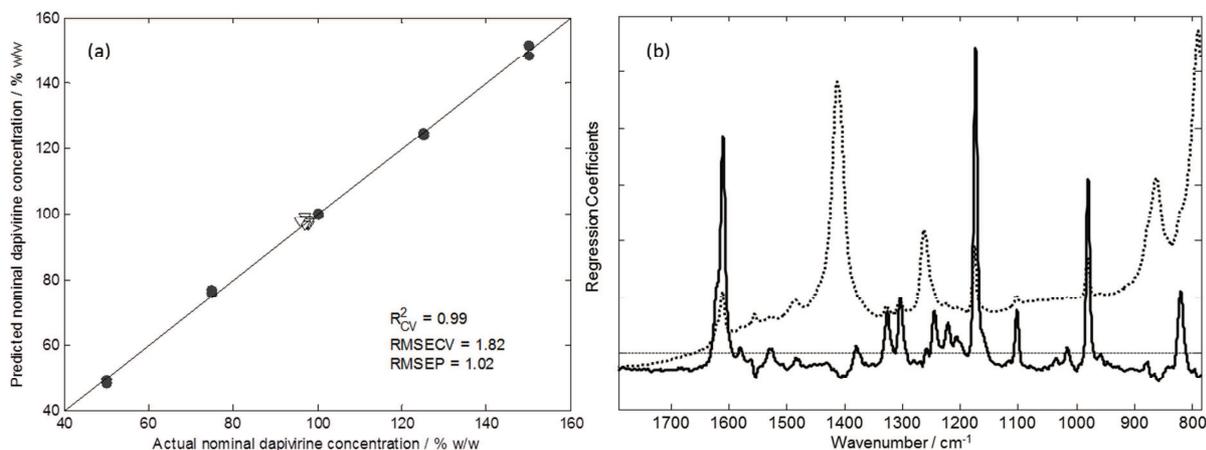
PLS regression on the SNV corrected spectra using the full spectral range resulted in systematic errors in the predictions of the rings from the production (Figure 4, Table 2).



**Figure 4.** Actual vs. predicted dapivirine concentrations in intravaginal rings. Predictions found from PLS regression on SNV corrected Raman spectra. Calibration set of reference rings●, and test set of rings from production  $\nabla$ . Systematic prediction error occurs.

This could be assigned to variations between reference rings and samples from production, deviations due to functional and/or compositional difference in the silicone elastomers. If we avoid the major spectral bands from silicone elastomer at 489 and 690  $\text{cm}^{-1}$ , and only use the spectral range from 800 to 1800  $\text{cm}^{-1}$  in the PLS regression the systematic prediction errors disappear, and the model performs well. Fig. 5a shows the actual vs. predicted dapivirine, and the regression vector, for a PLS where only the selected spectral region is included. The PLS regression vector in

Fig. 5b shows that bands originating from dapivirine are most important for the PLS model. The performance of the different calibration models are summarized in Table 2.



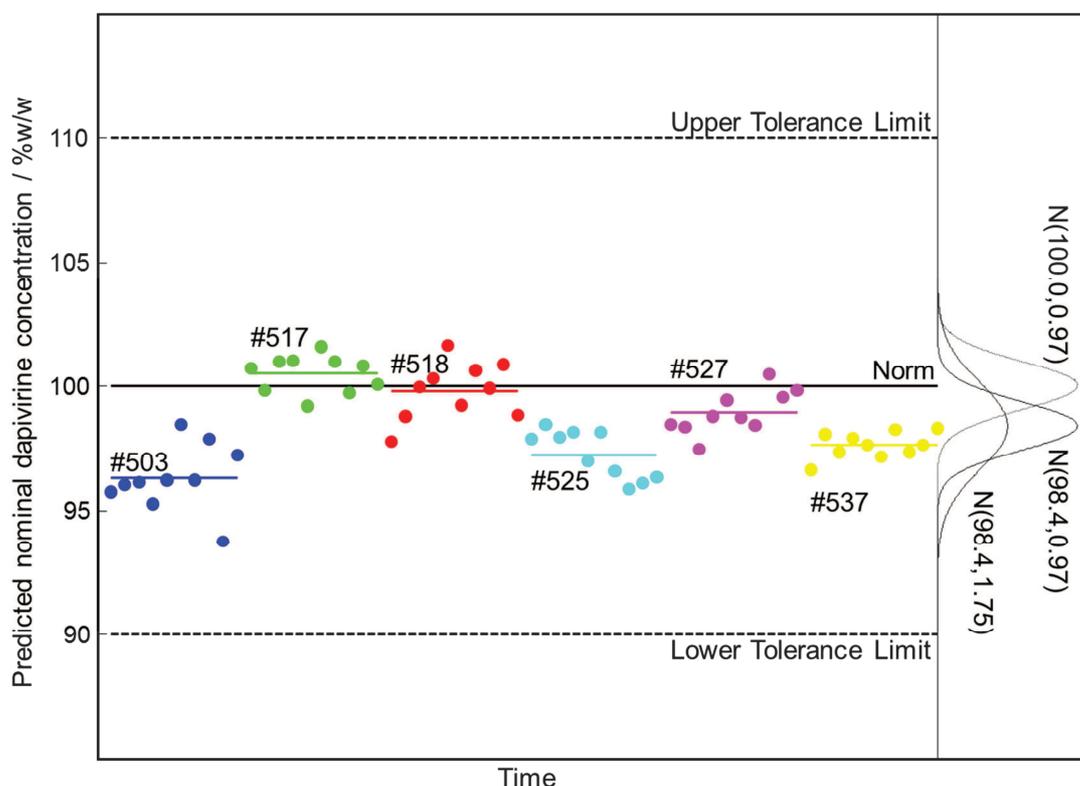
**Figure 5.** a): Actual vs. predicted dapivirine concentrations in intravaginal rings. Predictions found from PLS regression on SNV corrected Raman spectra (800-1800 cm<sup>-1</sup>). Calibration set of reference ring ●, and test set of rings from production ▽. b) dotted line shows the mean Raman spectra and the solid line shows the PLS regression vector (two latent variables) for prediction of dapivirine from Raman spectra.

The results show that the bivariate model without band fitting predicts well, but not quite as well as the multivariate PLS model on a selected spectral range. The reason for this performance difference could be that bivariate models in general are less likely to capture background phenomena that may interfere with predictions than multivariate models. A possible compositional difference in the silicone elastomer between the calibration set and the validation set may induce this type of interference that a multivariate calibration model can handle better than a bivariate model. The requirement of SNV before ratio-modelling and reduction of the PLS spectral range hint at a polymer-matrix difference between the customized reference rings and production rings, something also noticeable by visual comparison of the two. The low RMSEP value, based on independent HPLC reference analysis of production rings, proves that the matrix differences play no significant role in the final PLS model.

### Multivariate process control

To illustrate the power of rapid spectroscopy based process analysis we use some preliminary test results. Ten IVRs from six batch-runs were analyzed by Raman spectroscopy and the findings, together with the nominal or target value and upper and lower tolerance limits based on likely predetermined product specifications, are shown in Figure 6. It is important to remember that these batches (organized according to time) are part of the process optimization stage and the bias and variance are thus not truly representative for final production. However, based on the observations,

the initial process capability can be assessed. An established way to express this is via the so-called process capability index  $C_p$  (Equation 7). A value of the  $C_p$  of 1 is often accepted, while a score of more than 2 is required under the six-sigma rules.<sup>[25]</sup> Using the test batches two scenarios can be evaluated: 1) assuming no batch-to-batch differences an overall standard deviation can be computed resulting in a  $C_p$  of 1.9; a likely more realistic estimate would be the pooled standard deviation over the six batches which gives a  $C_p$  of 3.4. As can be concluded from these numbers, and observed from Figure 6, the process is capable even during the process optimization. But we also observe a modest bias in the means during this stage. There the offset corrected capability index (Equation 8) might give a more fair indication. This results in a  $C_{pk}$  of 1.6 and 2.9 for no batch-to-batch differences assumed and the pooled uncertainty, respectively.

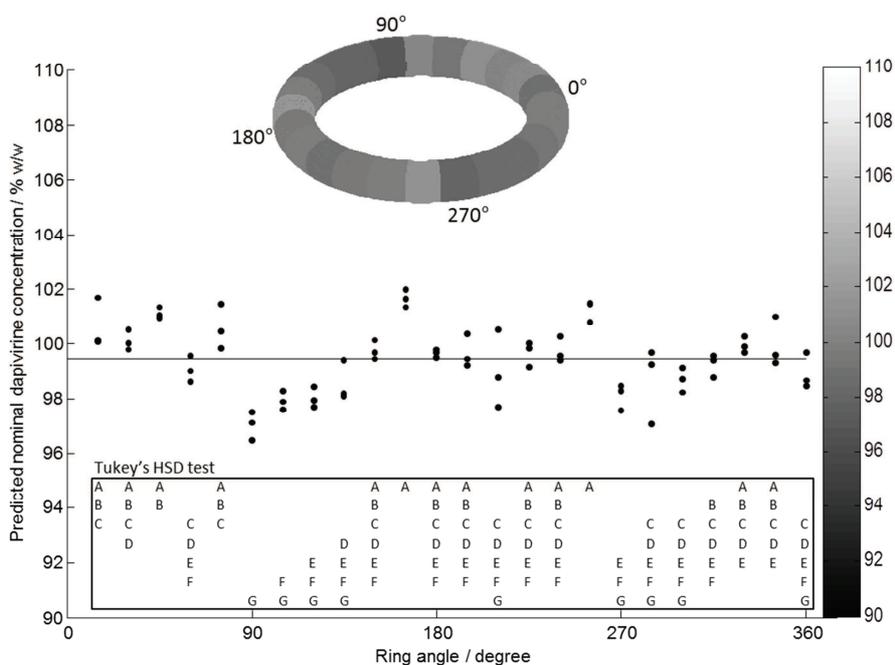


**Figure 6.** Concentration predictions of ten IVRs collected from six batch runs using Raman measurements and a PLS calibration model; (●) estimation, (–) batch mean. Nominal/Norm values and Upper and Lower Tolerance Limits (UTL, LTL) indicated. The anticipated Gaussian population distributions for ensemble/overall variance ( $n = 60$ ) and for pooled within-batch variance ( $n = 54$ ) are shown on the right.

### Ring heterogeneity testing of dapivirine concentration

The measurements for the calibrations model were carried out on rotating rings providing an average value of the API in the entire ring. Distribution of API in the ring provides additional product characterization. Therefore it was investigated if the concentration of dapivirine was the

same at different spots around the ring. From Raman spectra on 24 different spots on the entire ring circumference dapivirine concentrations were predicted using the established PLS calibration model. In Figure 7 an example of the characteristics for one ring is shown. The figure shows the dapivirine concentration in each of the spots measured in triplicates with 15° differences. The top of the figure shows a contour plot of the ring coloured according to dapivirine concentration. One way ANOVA between spots showed  $P < 0.0001$  which proves there are significant difference between spots. In conjunction to the ANOVA Tukey's HSD test was performed (Figure 7, bottom). This multi comparison test showed that none of the spots were completely different than at least one of its "neighbour" spots, indicating that the API is not highly concentrated in a few places but rather smoothly distributed. Other production rings investigated showed similar results. Such ring heterogeneity testing would be very time and resource demanding with HPLC whereas an on-line Raman installation would be rapid and could easily be automated. Furthermore, additional information on ring deviations - e.g. impurities, physical defects etc. - can be detected as they would invoke spectral abnormalities, and thereby, Raman could also function as a rapid fault screening method.



**Figure 7.** Distribution of dapivirine in an intravaginal ring from batch 527. Raman spectra of 24 spots covering the entire ring circumference were obtained in triplicates. Predictions of Dapivirine were based on PLS regression model of SNV corrected Raman spectra (800-1800  $\text{cm}^{-1}$ ). Nominal concentration of dapivirine versus ring angle. Top: Contour plot of the ring coloured according to Dapivirine concentration. Bottom: Tukey's HSD test.

## **Conclusions**

Raman spectroscopy was investigated as an alternative to the current standard method HPLC for quantifying the API dapivirine in HIV preventive silicone vaginal rings. A novel sampling system using wide-area illumination and sample rotation provided rapid and representative measurements for a number of customized reference rings and rings taken from a real production. Our results show that PLS regression can predict the dapivirine in vaginal rings with low errors (RMSECV and RMSEP). It is demonstrated how predictions using Raman can be used in building process control charts for estimating the process capability. Furthermore, ring heterogeneity was tested by measuring on a large number of spot covering the entire ring circumference. ANOVA results show significant difference in dapivirine concentration within the ring, although all spot stay well within specifications.

The results of this feasibility study implies that HPLC as a time and resource demanding standard quality control method can be partly substituted by Raman spectroscopy which is rapid, non-destructive and requires no sample preparation. Due to the short analysis time (high-throughput) a larger fraction of products can be quality assessed prior to commercial release. This opens the possibilities of intervening faster when irregularities occur, thereby reducing scrap. This study has focused on predicting dapivirine in the final product by Raman spectroscopy. However, as a multivariate spectral method, Raman can also provide information on other sample quality attributes – a priori known or unknown - such as impurities and physical properties. Furthermore, the flexible sampling of Raman expands the measurement possibilities, so quality can be monitored multiple places during the ring production e.g. raw material verification, mixing process and extrusion. The results of this paper have thus demonstrated the feasibility of using Raman spectroscopy combined with multivariate data analysis as process analytical technology for rapid quality control.

## **Acknowledgements**

LBL would like to thank Technology and Innovation foundation for financial support through the Consortium QbD – ‘Quality by Design in the Food and Pharmaceutical Industries’. The reference rings and the ring spinner were kindly provided by IPM (Maryland, US).

## References

- [1] UNAIDS. UNAIDS report on the global AIDS epidemic **2012**, [http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120\\_UNAIDS\\_Global\\_Report\\_2012\\_with\\_annexes\\_en.pdf](http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf)
- [2] J. Romano, B. Variano, P. Coplan, J. Van Roey, K. Douville, Z. Rosenberg, M. Temmerman, H. Verstraelen, L. Van Bortel, S. Weyers, M. Mitchnick, *AIDS Res. Hum. Retroviruses* **2009**, *5*, 483.
- [3] A. Nel, S. Smythe, K. Young, K. Malcolm, C. McCoy, Z. Rosenberg, J. Romano, *J. Acquir. Immune Defic. Syndr.* **2009**, *4*, 416.
- [4] E. T. Montgomery, A. van der Straten, H. Cheng, L. Wegner, G. Masenga, C. von Mollendorf, L. Bekker, S. Ganesh, K. Young, J. Romano, A. Nel, C. Woodsong, *Aids and Behavior* **2012**, *7*, 1787.
- [5] A. van der Straten, E. T. Montgomery, H. Cheng, L. Wegner, G. Masenga, C. von Mollendorf, L. Bekker, S. Ganesh, K. Young, J. Romano, A. Nel, C. Woodsong, *Aids and Behavior* **2012**, *7*, 1775.
- [6] S. M. Fetherston, P. Boyd, D. F. McCoy, M. C. McBride, K. Edwards, S. Ampofo, R. K. Malcolm, *Eur. J. Pharm. Sci.* **2013**, *3*, 406.
- [7] United States Food and Drug Administration, *Guidance for industry PAT - A framework for innovative pharmaceutical development, manufacturing and quality assurance*, U.S. Department of Health and Human Services, Rockville, **2004**.
- [8] K. C. Gordon, C. M. McGoverin, *Int. J. Pharm.* **2011**, *1-2*, 151.
- [9] C. J. Strachan, T. Rades, K. C. Gordon, J. Rantanen, *J. Pharm. Pharmacol.* **2007**, *2*, 179.
- [10] S. E. J. Bell, A. C. Dennis, L. A. Fido, R. K. Malcolm, N. M. S. Sirimuthu, C. F. Toner, A. D. Woolfson, *J. Pharm. Pharmacol.* **2007**, *2*, 203.
- [11] R. Bro, *Anal. Chim. Acta* **2003**, *1-2*, 185.
- [12] W. Lin, J. Jiang, H. Yang, Y. Ozaki, G. Shen, R. Yu, *Anal. Chem.* **2006**, *17*, 6003.
- [13] S. Sasic, *Appl. Spectrosc.* **2007**, *3*, 239.
- [14] T. Sakamoto, T. Matsubara, D. Sasakura, Y. Takada, Y. Fujimaki, K. Aida, T. Miura, T. Terahara, N. Higo, T. Kawanishi, Y. Hiyama, *Pharmazie* **2009**, *3*, 166.
- [15] A. Docoslis, K. L. Huszarik, G. Z. Papageorgiou, D. Bikiaris, A. Stergiou, E. Georganakis, *Aaps Journal* **2007**, *3*, 43.
- [16] N. Furuyama, S. Hasegawa, T. Hamaura, S. Yada, H. Nakagami, E. Yonemochi, K. Terada, *Int. J. Pharm.* **2008**, *1-2*, 12.
- [17] G. Armstrong, H. Edwards, D. Farwell, A. Williams, *Vib. Spectrosc.* **1996**, *2*, 105.

- [18] C. S. Forbes, M. Evans, *Statistical distributions*, Wiley-Blackwell, Oxford, **2010**.
- [19] D. W. Marquardt, *J. Soc. Ind. App. Math.* **1963**, 2, 431.
- [20] R. H. Barham, W. Drane, *Technometrics* **1972**, 3, 757.
- [21] M. Sjostrom, S. Wold, W. Lindberg, J. Persson , H. Martens, *Anal. Chim. Acta* **1983**, 1, 61.
- [22] R. Bro, *Håndbog i multivariabel kalibrering*, Jordbrugsforlaget, Frederiksberg, **1996**.
- [23] P. Geladi, B. Kowalski, *Anal. Chim. Acta* **1986**, , 1.
- [24] S. Wold, M. Sjostrom , L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, 2, 109.
- [25] D.C. Montgomery *Introduction to statistical quality control*, John Wiley, Hoboken, N.J, **2005**.
- [26] D. Cai, A. Neyer, R. Kuckuk , H. M. Heise, *J. Mol. Struct.* **2010**, 1-3, 274.
- [27] M. R. Aguiar, C. Verissimo, A. C. S. Ramos, S. A. Moshkalev , J. W. Swart, *J. Nanosci. Nanotechnol.* **2009**, 7, 4143.
- [28] A. Geissberger, F. Galeener, *Phys. Rev. B* **1983**, 6, 3266.
- [29] S. Hunt, G. Cash, H. Liu, G. George , D. Birtwistle, *J. Macromol. Sci., Pure Appl. Chem.* **2002**, 9, 1007.
- [30] R. Borges, A. Abras , H. Beraldo, *J. Braz. Chem. Soc.* **1997**, 1, 33.
- [31] A. Raj, K. Raju, H. T. Varghese, C. M. Granadeiro, H. I. S. Nogueira , C. Y. Panicker, *J. Braz. Chem. Soc.* **2009**, 3, 549.
- [32] J. Vazquez, J. Gozalez, F. Marquez , J. Boggs, *J. Raman Spectrosc.* **1998**, 6, 547.
- [33] J. Faniran, *J. Chem. Crystallogr.* **1975**, 3, 191.
- [34] D. Boo, K. Kim , M. Kim, *Bull. Korean Chem. Soc.* **1987**, 4, 251.
- [35] H. Okamoto, H. Inishi, Y. Nakamura, S. Kohtani , R. Nakagaki, *Chem. Phys.* **2000**, 1-2, 193.
- [36] L. B. Lyndgaard, K. M. Sorensen, F. van den Berg , S. B. Engelsens, *J. Raman Spectrosc.* **2012**, 4, 482.

# POSTER I

---

Lyndgaard, L.B., Andersen, K.K., Christensen, M., Acar, E., Egebo, M. & Bro, R.

Predicting coagulation properties of milk with seasonal variations  
by mid-infrared spectroscopy

*Awarded best poster at 12th Scandinavian Symposium on Chemometrics, 2011*

