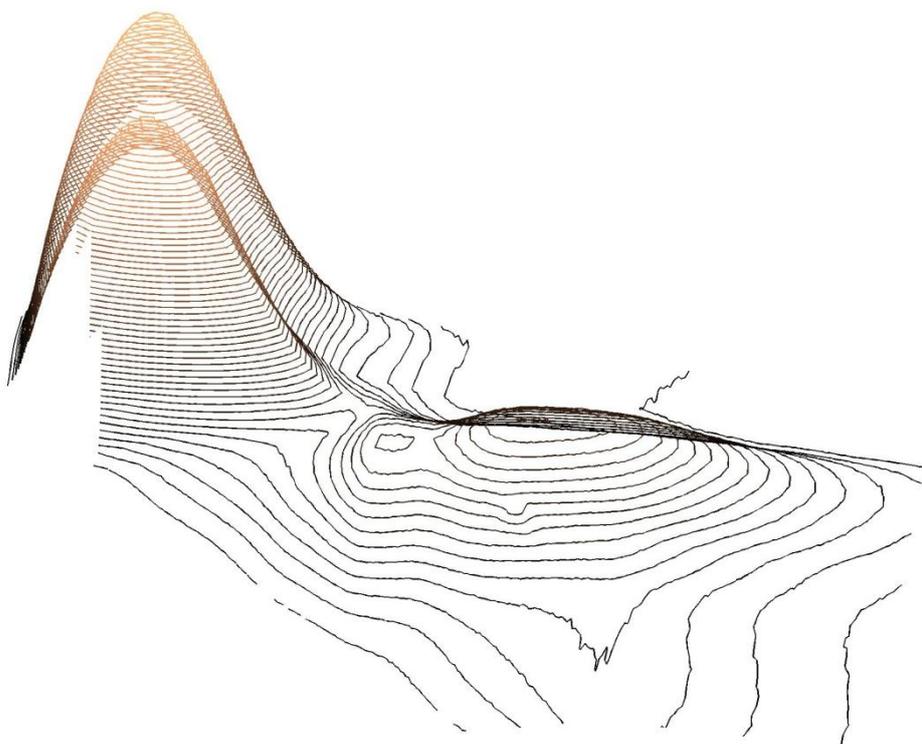


DEPARTMENT OF FOOD SCIENCE
FACULTY OF LIFE SCIENCES
UNIVERSITY OF COPENHAGEN

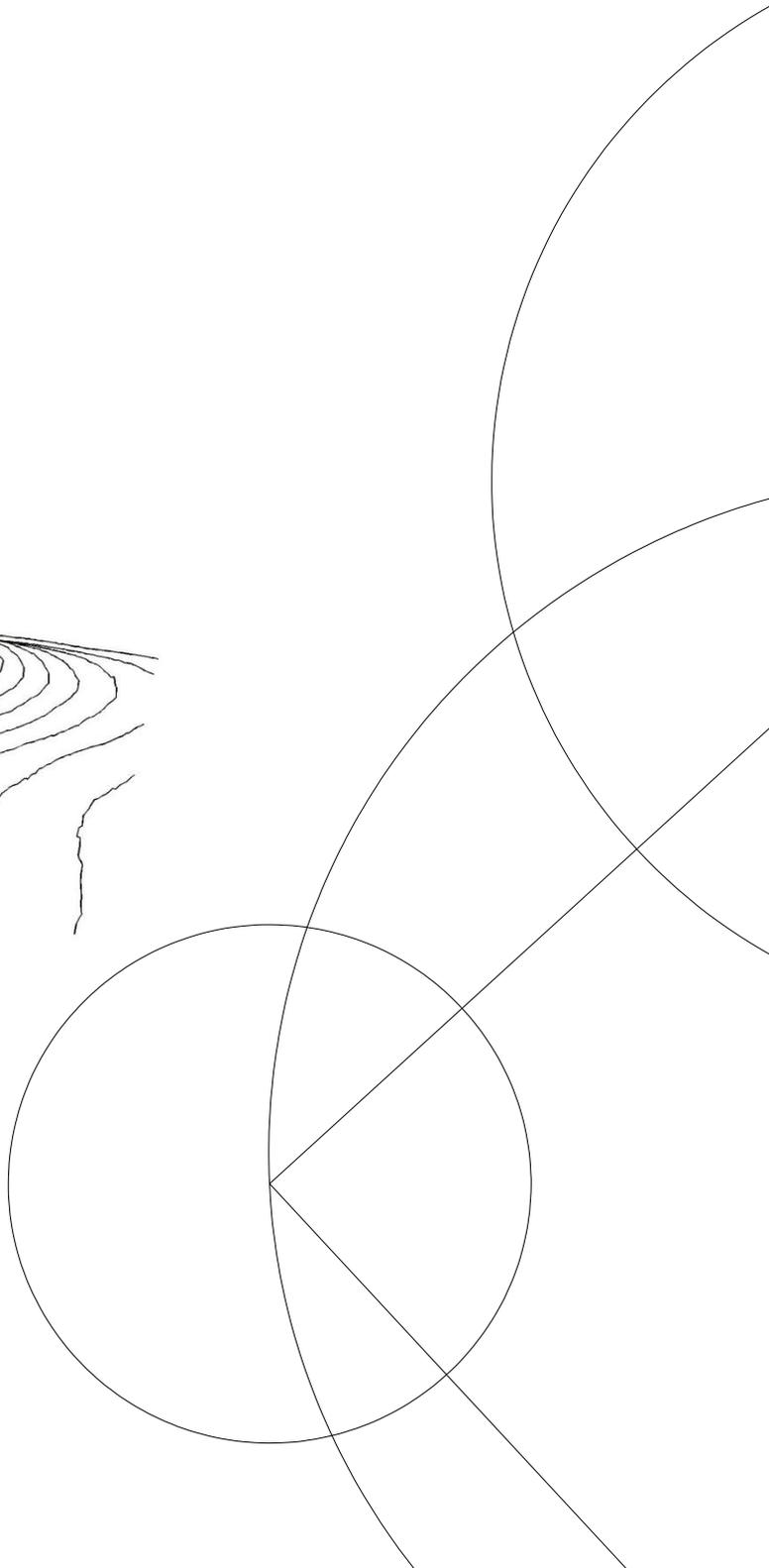


Fluorescence Spectroscopy and Chemometrics -Applied in Cancer Diagnostics and Metabonomics



PhD thesis - 2011

Anders Juul Lawaetz



**Fluorescence Spectroscopy and
Chemometrics
-Applied in Cancer Diagnostics and
Metabonomics**

PhD thesis by
Anders Juul Lawaetz 2011

Supervisor:
Rasmus Bro

Preface

This PhD thesis has been conducted in the Quality and Technology research group at the Department of Food Science, Faculty of Life Sciences, University of Copenhagen, under supervision of Professor Rasmus Bro. The project has been financed by the VILLUM KANN RASMUSEN FOUNDATION (now VILLUM FONDEN) as part of the project; Metabonomic Cancer Diagnostics.

I am very grateful to Rasmus for his supervision and support during this project. His great expertise in chemometrics has been inspiring, and with his always positive attitude he has given me motivation to look forward and continue when various obstacles have come in my way.
“Life is beautiful!”

Special thanks go to Colin Stedmon for our discussions and collaboration about standardization of fluorescence spectroscopy, and for his comments to a part of this thesis. Also great thanks to Maja Kamstrup-Nielsen for good collaboration on the colorectal cancer paper, and for comments on a part of this thesis. Thanks to Bonnie Schmidt for collaboration on the rotation paper, to Hans Jørgen Nielsen and Ib Jarle Christens for providing the plasma samples and collaboration on the colorectal cancer paper and to Abdelrhani Mourhib for his assistance in preparing the plasma samples.

The Q&T group is a fantastic place to work and I owe a great thanks to you all for making a great working environment academically as well as socially. Special thanks go to Karin Kjeldahl and Åsmund Rinnan for all their Matlab assistance, and to Karin also for comments on a part of this thesis. Thanks to Marta my office partner for listening to all my complaints and laughing at my “jokes”. Christian Lyndgaard is thanked for table tennis matches, music discussions and a lot of good laughs.

Finally a great thanks to friends and family for all your support, especially to my loved ones Tobias, Alberte and Agnes and last but not least to Stine, without you all this wouldn't have been possible.

Anders Juul Lawaetz
February 2011

Additional preface 2nd edition May 2012

In this 2nd edition of my thesis some errors have been corrected; wrong axis on figure 2 and some mix up of words in the part about sensitivity and specificity, along with some typos. Since the original thesis paper III has been accepted in a slightly altered version than the original. The accepted paper is included in this edition.

Anders Juul Lawaetz
May 2012

Abstract

The objective of this PhD project is to use fluorescence spectroscopy as a tool to discriminate between cancer patients and healthy controls by measuring on a sample of blood serum or plasma. Further by using PARAFAC to extract relevant chemical information from the fluorescence landscapes, fluorescence spectroscopy might be a potential metabonomic tool. The fluorescence excitation emission matrixes are decomposed by PARAFAC yielding estimates of the underlying chemical compounds in the blood sample. By using the PARAFAC components as a basis for discrimination between cancer and non-cancer, it is possible to achieve understanding of the chemical changes causing the discrimination. Since fluorescence spectroscopy is sensitive and specific towards even small chemical changes, the combination of fluorescence spectroscopy and PARAFAC can be seen as an interesting alternative metabonomic tool. Examples are presented with colorectal cancer and breast cancer. The results show that fluorescence spectroscopy can be used to discriminate between cancer and control samples at a level comparable with known biomarkers, and by using PARAFAC, chemical knowledge about the discrimination is achieved.

This thesis will go through some of the basic theory of the methods applied in both fluorescence spectroscopy and chemometrics. One important aspect in fluorescence spectroscopy is the instrument dependent bias in the measured data. For identical samples different instruments will give slightly different solutions, and in order to be able to compare fluorescence data across instruments, spectral correction is necessary. An example of some of the challenges in applying spectral correction by use of a commercial solution is given. Besides the chemometric methods

applied in this thesis some other aspects of chemometrics in connection to metabonomics are also discussed

The thesis will briefly go through the basics of cancer and cancer detection and screening, focusing on colorectal cancer. A literature study on how fluorescence spectroscopy on blood has been used in connection to oncology has been conducted.

Three scientific papers have been prepared in connection with this PhD project. **Paper I** is presenting an example of an improved method for intensity calibration of fluorescence spectroscopy by use of the integrated area under the water Raman peak. **Paper II** is an application of rotation of a PCA model to facilitate interpretation of the solution of a metabonomic application of St. Johns Worth **Paper III** is an example of fluorescence spectroscopy and chemometrics applied on blood plasma samples to discriminate between patients with colorectal cancer and various control groups. PARAFAC is applied and the potential for using fluorescence spectroscopy along with PARAFAC as a metabonomic tool is presented.

Resumé

Formålet med dette PhD projekt har været at anvende fluorescens-spektroskopi til at skelne mellem kræftpatienter og raske ved at måle på enten en blodplasma- eller serumprøve. Samtidig er formålet at anvende PARAFAC til at ekstrahere relevant kemisk information fra fluorescensmålingerne for at undersøge om fluorescensspektroskopi er et muligt værktøj til brug inden for metabonomics. Fluorescenslandskaber kan med PARAFAC blive nedbrudt til kvalitative og kvantitative estimater af de underliggende kemiske komponenter i blodprøven. PARAFAC-komponenterne kan så danne udgangspunkt for en diskrimination mellem kræftprøver og kontroller, og dermed er der bedre mulighed for at opnå en øget forståelse for de kemiske ændringer, som er årsag til forskellen mellem kræftpatienter og raske. Denne mulighed gør fluorescensspektroskopi sammen med PARAFAC til et interessant alternativt værktøj inden for metabonomics. Der vil blive gennemgået eksempler indenfor tyktarmskræft og brystkræft. Resultaterne viser, at det er muligt at anvende fluorescensspektroskopi til at skelne mellem kræftpatienter og kontroller på niveau med kendte biomarkører. Ved at anvende PARAFAC er det samtidig muligt at opnå kemisk viden omkring diskriminationen.

Denne afhandling vil gennemgå teorier, der ligger bag de metoder indenfor fluorescensspektroskopi og kemometri, der er anvendt i afhandlingen. Data fra fluorescensspektroskopi kan være meget afhængige af det spektrofouorometer, som er anvendt til at optage spektrene. Identiske prøver målt på forskellige spektrofouorometre kan give forskellige resultater. Hvis en metode, der er baseret på fluorescens-spektroskopi, skal kunne anvendes globalt, kræver det en spektral korrektion. Et eksempel på udfordringerne omkring implementering af spektral korrektion fra en kommercielt tilgængelig løsning er beskrevet.

Ud over de kemometriske metoder, som er anvendt i afhandlingen, vil andre metoder til kemometri i forbindelse med metabonomics blive diskuteret.

Der vil blive givet en kort gennemgang af kræft og metoder til at detektere og screene for kræft, med særligt fokus på tyktarmskræft. Et litteraturstudie er gennemført omkring anvendelsen af fluorescensspektroskopi på blodprøver i forbindelse med kræft.

I forbindelse med dette PhDprojekt er der blevet forberedt tre videnskabelige artikler. **Paper I** gennemgår et eksempel på en forbedret metode til at udføre intensitets kalibrering af fluorescensspektre ved at benytte arealet under vands raman spektrum. **Paper II** viser, hvordan anvendelsen af rotation af scorer og loadings i en PCA model kan gøre fortolkningen mere entydig. Metoden er anvendt i et metabonomic studie af Hyperikum-præperater. **Paper III** gennemgår et eksempel på anvendelse af fluorescensspektroskopi og kemometri på blodplasma til diskrimination af patienter med tyktarmskræft og forskellige kontrolgrupper. PARAFAC er anvendt, og potentialet for at kombinationen af fluorescensspektroskopi og PARAFAC er en mulig metode til metabonomics bliver præsenteret.

List of Publications

Paper I

Lawaetz A J, Stedmon CA. *Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water*. *Appl.Spectrosc.* 63 (2009); 936-940

Paper II

Lawaetz A J, Schmidt B, Staerk D, Jaroszewski J W, Bro R. *Application of Rotated PCA Models to Facilitate Interpretation of Metabolite Profiles: Commercial Preparations of St. John's Wort*. *Planta Med.* 75 (2009); 271-279

Paper III

Lawaetz A J, Bro R, Kamstrup-Nielsen M, Christensen I J, Jørgensen L N, Nielsen H J. *Fluorescence Spectroscopy as a Potential Metabonomic Tool for Early Detection of Colorectal Cancer*. *Metabolomics*, DOI 10.1007/s11306-011-0310-7

Additional Publications by the Author

I

Hougard A B, **Lawaetz A J**, Ipsen R, *Front face Fluorescence Spectroscopy and Multi-way Data Analysis for Characterization of Milk Pasteurized Using Instant Infusion*. *J. Food Chemistry*, submitted

II

Holm J, Schou C, Babol L N, **Lawaetz A J**, Bruun S W, Hansen M Z, Hansen S I, *The Interrelationship Between Ligand Binding and Self-Association of the Folate Binding Protein. The Role of Detergent-Tryptophan Interaction*. *Biochimica et Biophysica Acta, General Subjects* 1810 (2011); 1330-1339

Table of Contents

Preface	2
Abstract	4
Resumé	6
List of Publications	8
List of Abbreviations	11
Chapter 1: Introduction	12
Chapter 2: Fluorescence spectroscopy:.....	16
External Conditions Affecting Fluorescence Emission	20
Concentration Effects	22
Chapter 3: Standardization and Quality Assurance of Fluorescence Spectra	26
Test of BAM Standards.....	28
Intensity Calibration. (Paper I)	33
Chapter 4: Chemometrics/Data Analysis.....	36
PCA.....	36
Multi-Way Data Analysis – PARAFAC	37
PARAFAC and Fluorescence Spectroscopy	38
PLS	40
Classification	41
Sensitivity and Specificity	44
Rotation of PCA Scores (Paper II)	46
Chapter 5: Cancer, Definitions, Detection, Screening	50
Chapter 6: Fluorescence Spectroscopy on Blood Samples in a Diagnostic Context	56
Porphyrins in Acetone Extracts.....	57
Chapter 7: Examples of Fluorescence Spectroscopy and Multiway Analysis Applied in Detection of Cancer.....	62
The Multivariate Approach	62
Multivariate to Multiway! PARAFAC Opens for the Metabonomic Approach.....	63

Example 1 - Colorectal Cancer (Paper III)	63
Example 2 - PARAFAC on Breast Cancer Data	70
Porphyrins in Plasma – and PARAFAC	75
Chapter 8: Conclusion	78
Reference List.....	81

List of Abbreviations

AUC	Area under the ROC Curve
BAM	Bundesanstalt für Materialforschung und –prüfung (Federal Institute for Materials Research and Testing)
CEA	CarcinoeEmbryonic Antigen
CRC	Colo-Rectal Cancer
ECVA	Extended canonical Variate Analysis
EEM	Excitation Emission Matrix
FOBT	Fecal Occult Blood Test
LDA	Linear Discriminant Analysis
MS	Mass spectrometry
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
PLS	Partial Least Squares (Regression)
PLS-DA	PLS – Discriminant Analysis
PMMA	Polymethyl methacrylate
QDA	Quadratic Discriminant Analysis
ROC	Receiver Operating Characteristics
RU	Raman Units
trp	Tryptophan

Notation:

Fluorescence excitation/emission pairs are expressed as the excitation wavelength/emission wavelength designated nm. For example the tryptophan emission maximum at 350 nm after excitation at 280 nm is written 280/350 nm

Three-way arrays are denoted as underlined bold capitals, two-way matrices are denoted as bold capitals, vectors are denoted as a lower case letters in bold and scalars are denoted as a lower case letters in italic.

Chapter 1: Introduction

New ways for easy and/or early detection of lethal diseases especially cancer is the main focus of a lot of research. The “omics” are huge topics in these areas, and the number of publications in metabonomics and metabolomics has increased exponentially within the last ten to fifteen years [78]. Metabonomics/Metabolomics deals with quantitative and qualitative measurements of metabolites/small molecules in tissue or body fluids in humans, or in plants in various extracts. The aim is to make profiles or fingerprints of specific cellular processes, and thereby hopefully gain enhanced understanding of the process, or change in the process upon stress (e.g. due to disease) [75;78]. The original definition of metabolomics is “*the comprehensive and quantitative analysis of all metabolites in a biological system*” [21], and for metabonomics it is “*the quantitative measurements of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification*” [73]. Despite the differences in the two definitions, the terms are often used indiscriminately. In this thesis the focus is on the metabolic response to cancer, and following the above definitions the term *metabonomics* will be used.

The major parts of all metabonomic studies are based on data from mass spectrometry (MS) measurements, coupled to chromatographic techniques, or data from nuclear magnetic resonance (NMR) spectroscopy. These techniques can measure a great number of metabolites qualitative as well as quantitatively, and are thus excellent for the purpose. The MS-based methods are the most applied methods, as these are more sensitive, whereas NMR spectroscopy is more specific and with a higher reproducibility. It requires no pre separation of the samples and is thus faster and non-destructive [37;99]. The increased focus on the area has opened for new ways or new methods in performing

“omics” studies. The traditional ways have been through MS coupled to a chromatographic pre separation step, or NMR, but other methods such as Capillary Electrophoreses or even IR spectroscopy has been suggested [59]. In this thesis the possibilities of using fluorescence spectroscopy as a metabonomic tool will be discussed and evaluated.

Fluorescence spectroscopy and metabonomics

Compared to the traditional analytical methods in metabonomics, fluorescence spectroscopy has a much higher sensitivity, and can thus detect compounds in much lower concentrations. A drawback though is that the number of specific compounds measurable by fluorescence spectroscopy is low compared to the two others. In reference to the definition of metabolomics [21], fluorescence spectroscopy cannot be used to measure the total profile of metabolites. The high specificity of fluorescence spectroscopy allows for fluorescence spectroscopy to measure the important aromatic amino acids, and to differentiate between amino acids in different proteins or in the same protein but at different locations. This makes fluorescence spectroscopy a powerful tool for measuring specific and potentially important metabolites, and to detect even small changes in for example the micro environment of a blood sample [58]. Based on that, fluorescence spectroscopy can be seen as a potential metabonomic tool to measure the change in metabolites upon for example pathophysiological stimuli. Further; compared to MS or NMR, fluorescence is a fast and easy to use method and relatively cheap.

In this thesis examples and discussions are shown for the use of fluorescence spectroscopy on blood as a tool in cancer diagnostics. The possibilities of introducing fluorescence spectroscopy along with PARAFAC as an alternative method for performing metabonomic research is presented. Autofluorescence on blood samples to detect cancer has been suggested before, and it is this work that this thesis will try to bring one step further. This is pursued in a study using a larger number of samples than the previous studies and further by applying PARAFAC analysis of

the fluorescence landscapes. By applying PARAFAC there is a chance for providing better options for understanding of the metabolic changes behind a possible discrimination between cancer and non-cancer (Paper III).

Compared to targeted metabolite analysis, which is a focused analysis of specific compounds/metabolites, metabonomic studies provide a fairly unbiased measure of the metabolites and changes upon metabolic response. Instead of only measuring a few specific metabolites, the analytical methods often applied can measure several hundred metabolites. One of the major challenges in all “omics” studies is to extract the important/relevant biological information out of this sometimes complex data output, and for example discover new biomarkers. Multivariate data analysis/ chemometrics is part of the answer to this problem, and has thus in the recent years become an indispensable part of metabonomics [96]. This thesis will briefly go through some of the standard methods applied in metabonomics such as PCA and PLS. Similar to the research in the analytical part of metabonomics, research is conducted in finding new dedicated chemometric solutions to extract information from metabonomic data in the best possible way [88;96]. In Paper II made as part of this PhD, an application of rotations of a PCA model in a metabonomic study is presented. Rotations are applied to facilitate better conditions for interpretation of the result, and are well known methods within psychometrics, but more infrequent in chemometrics and natural sciences. In the study in Paper II rotations are applied on metabonomic data for the first time. The result shows that there is a potential for more frequent use within this field [53;88;96].

There are some challenges in applying fluorescence spectroscopy as a clinical method/diagnostic tool. Fluorescence spectroscopy outcome can vary depending on the spectrometer applied, both in terms of spectral characteristics and intensity of the signal. Consequently there is a need for calibration of the data before a method based on fluorescence spectroscopy can be globally

applied. An improved method on how to perform intensity normalization using the water Raman signal of a pure water sample is presented (Paper II), and in this thesis there is a discussion on the topic of spectral calibration.

Chapter 2: Fluorescence spectroscopy

A major part of the work done during this PhD has been related to fluorescence spectroscopy. This chapter will briefly go through some of the basic principles in fluorescence spectroscopy.

Fluorescence spectroscopy deals with excitation and emission in molecules. Any molecule can go into an electronically excited state when exposed to light of a wavelength (energy level) equal to the energy gap between the ground state and excited state. This is known as molecular absorbance of light. The amount of light absorbed is proportional to the concentration of the absorbing molecule. This connection is described in Lambert-Beers law, where the wavelength dependent absorbance A is described

$$A = \log_{10} \left(\frac{I_0}{I} \right) = \epsilon cl$$

Where A is the absorbance I_0 and I the intensity of incoming and transmitted light, ϵ the molar absorptivity expressed in $\text{L} \times \text{mol}^{-1} \times \text{cm}^{-1}$, c the concentration in $\text{mol} \times \text{L}^{-1}$ and l the effective pathway of the sample in cm [61]. Measurement of the absorbance of a sample over a wavelength range results in an absorbance spectrum.

Absorbance only deals with the transition from ground state to excited state. Fluorescence involves the relaxation from excited to ground state. In most molecules this occurs as rapid non-radiative decay. For a limited number of molecules with certain characteristics (see below) the relaxation is through emission of light. This phenomenon is called fluorescence.

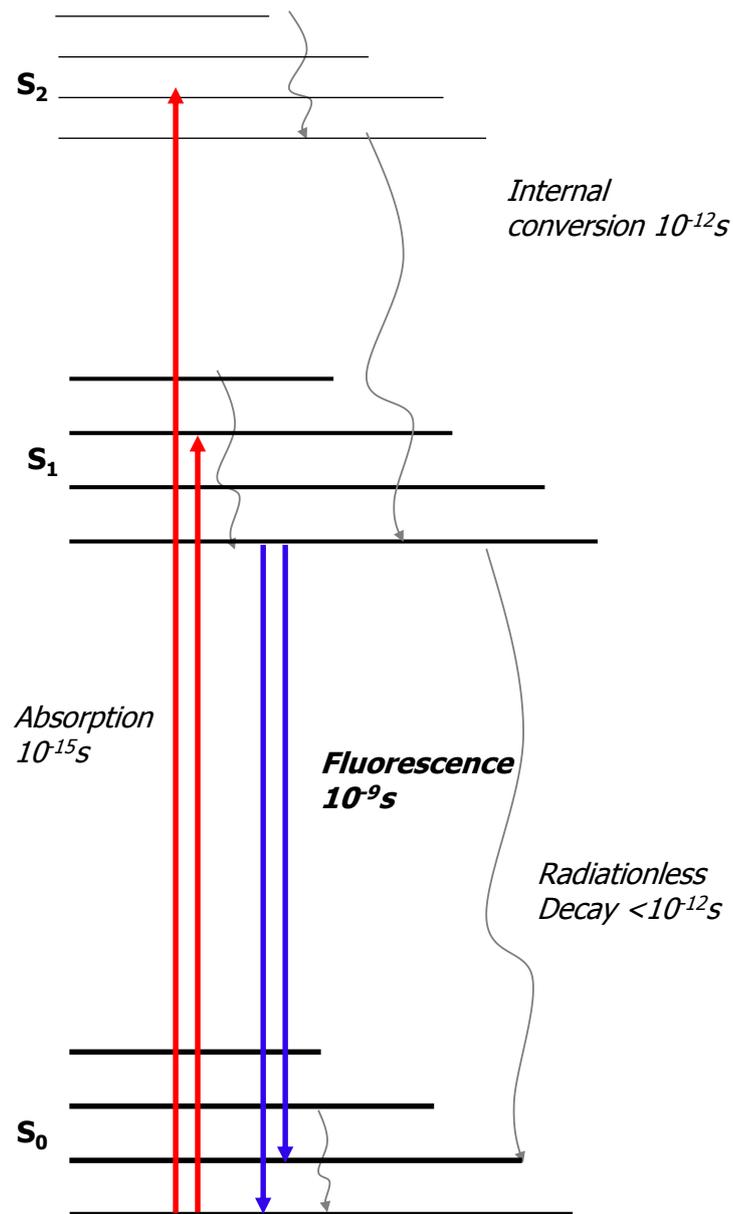


Figure 1: Example of a Jablonski diagram.

The mechanism of the excitation/relaxation in the molecule can be illustrated through the Jablonski diagram seen in figure 1. Dependent on the energy of the light, the molecule is excited to different electronic singlet states S_1 , S_2 , etc. Relaxation through emission of light though will in principle always occur from the lowest energy excited electronic state of a molecule (S_1) (Kasha's rule) [46], thus when excited to a higher energy excited electronic singlet state (S_2 or S_3), the molecule will undergo internal

conversion prior to emission. Emission almost always occur at the lowest excited state, and thus at a specific energy level (wavelength), independent of the energy (wavelength) of the excitation light [52;97]. There is a loss of energy through light emission, and the emitted light is always red shifted (towards longer wavelength) relative to the excitation light [52;97]. The difference between excitation and emission wavelength is called the Stokes shift which relates to the energy loss.

An emission spectrum is measured as the light emitted (fluorescence) across a broad wavelength range upon excitation at a fixed wavelength. Similarly, an excitation spectrum can be measured by measuring the emission at one fixed wavelength while exciting the molecule over a wavelength range. When measuring several emission spectra over a range of shifting excitation wavelengths a fluorescence landscape or an excitation emission matrix (EEM) will occur.

As stated above, excitation to different singlet states and to their different vibrational levels occurs at specific excitation wavelengths. This is reflected in the excitation spectrum by “spikes” or “fingers” on top the overall absorbance spectrum reflecting the transitions to the different singlet states (e.g. $S_0 \rightarrow S_1$, $S_0 \rightarrow S_2$) and to different vibrational levels of the singlet states as seen in the Jablonski diagram. Theoretically, when measuring a single fluorescing molecule, the excitation and absorption spectrum will be identical. Emission almost always occurs from the lowest singlet state S_1 to the ground state i.e. $S_1 \rightarrow S_0$ transition, and the emission spectrum therefore will most often have only a single peak (Gaussian) shape. If there are “spikes” or “fingers” in the emission spectrum it is due to transition to a higher vibrational level of S_0 . The shape of excitation and emission spectra is often described in the mirror image rule, which says that the emission spectra, the $S_1 \rightarrow S_0$ transition, is a mirror image of the excitation/absorbance spectrum of the $S_0 \rightarrow S_1$ transition [52;97].

As a consequence of the above described properties, the emission spectrum from a given fluorophore measured upon different

excitation wavelengths will only vary in intensity not shape, and the shape of the emission spectrum is thus independent of the excitation wavelength. The opposite is also true; the excitation spectrum is independent of emission. The fact that the fluorescence spectrum is measured as a function of two factors; excitation and emission wavelength, makes fluorescence spectroscopy a specific method that allows the scientist to assign spectra to specific chemical compounds.

Fluorescence spectroscopy is a measure of photons. In modern instruments this is often done as single photon counting, but traditionally this has been done as an average conversion of light pulses into an analog electrical signal [52]. Both methods are capable of detecting few photons accurately which makes fluorescence spectroscopy a highly sensitive method. It is reported 100-1000 times more sensitive compared to other spectroscopic methods [52;93].

Fluorophores are typically compounds with aromatic rings, conjugated double bonds or similar rigid structures that prevent relaxation through torsional energy. Common examples of fluorophores are the aromatic amino acids, tyrosine, phenylalanine and tryptophan, where especially the latter is widely used in protein science. Other important fluorophores found in biological samples are coenzymes NAD(P)H and FAD and a suite of vitamins (A, B, D and E) [12]. Important to this thesis, the fluorescence properties of human serum were studied by Wolfbeis and Leiner (1985) [104]. They concluded that there are two dominant parts in the fluorescence landscape from blood; one area in the ultraviolet spectral area which is due to the aforementioned amino acids, and one area in the near ultraviolet and visible area which is characteristic of e.g. NAD(P)H, riboflavin and bilirubin (example of an EEM of a blood sample in the figure below). Naturally occurring fluorophores in a sample are called intrinsic fluorophores and emission from those are called autofluorescence, as opposed to extrinsic fluorophores, which are designed fluorophores that bind to a specific molecule

and are added to a sample before measuring fluorescence. Globally, the field of extrinsic fluorophores (fluorescent probes) is a much larger field than the field of autofluorescence, and it is widely applied in molecular biology and the search for new biomarkers [27]. Only the field of autofluorescence is addressed in this thesis.

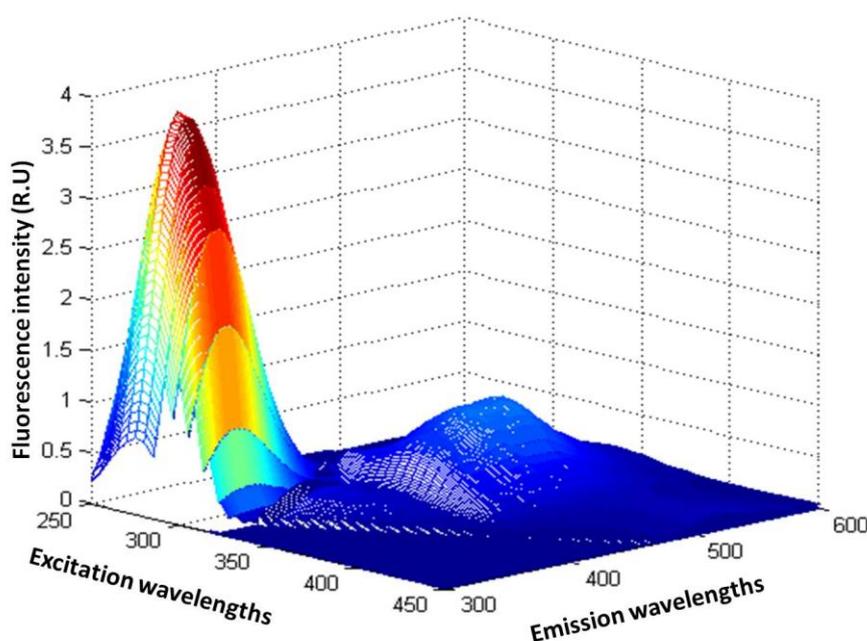


Figure 2: Measured fluorescence landscape (EEM) of a diluted blood plasma sample. The visual wavelength area (app 380:600 nm) is plotted on a different scale to see the spectral shape. It is clear that there are two areas of fluorescence the UV area (below 380 nm) which is dominated by the “amino acid” fluorescence, and the visual area (above 380 nm) which is different cofactors and vitamins.

External Conditions Affecting Fluorescence Emission

The local environment surrounding a fluorophore can affect the fluorescence signal. Factors such as pH, temperature, concentration and polarity can in one way or another affect the emission from a given fluorophore. The polarity of the solvent is an especially important factor as it causes a shift in the emission. When the molecule is excited, the dipole moment is higher than in the ground state. In a highly polar environment a “solvent”

relaxation will occur, making the dipole moment between ground state and excited state smaller, and thus a lower energy difference between the two states. This will lead to a shorter emission wavelength (a blue shift) compared to the same molecule in a non polar environment. This is relevant when measuring a fluorophore in different solvents, but also when measuring macro molecules such as proteins that can contain several fluorescing groups e.g. tryptophan, at different positions, or two different proteins where the tryptophan group is located at different sites. A tryptophan molecule located on the outside of a protein can have a rather different local environment compared to a tryptophan molecule located central in the protein. The two tryptophan groups will then have different emission maximum, and can be discriminated from one another. Below is an example with fluorescence measurements of a folate-binding protein in suspension. Tryptophan is expected to be the dominant fluorophore in this protein, and it has different tryptophan groups at different locations. The emission spectrum following excitation at 280 nm (expected tryptophan excitation) has maximum at 353 nm and a significant shoulder at 330 nm (figure 3 right plot). The excitation spectra at these two emission wavelengths (left plot), are identical with maximum in 280 nm, this is an example of difference emission in an internal and external sited tryptophan group [36].

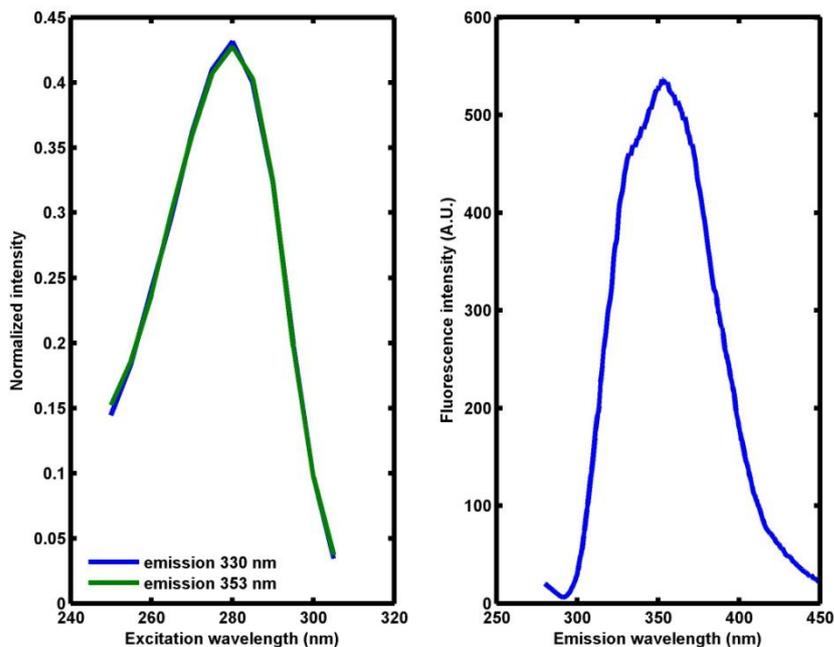


Figure 3: Excitation and emission spectra of folate binding protein in pH 7.4 suspension. Left; normalized excitation spectra, emission at 330 and 353 nm. Right; emission spectrum following the excitation at 280 nm. Data from [36]

Concentration Effects

The fluorescence intensity is dependent on the overall absorbance of the sample, and hence of the concentration of the fluorophore, but also from other absorbing substances in the sample. At low concentrations the relation between concentration and intensity known from Lambert-Beers law is also valid for fluorescence intensity. At high concentrations, the intensity can be affected by concentration quenching (sometimes described as inner filter effect). Part of the excitation and/or emitted light is reabsorbed by the sample, and the measured intensity of the fluorescence is thus decreased (quenched). In high concentration samples, the linear relation between concentration and fluorescence intensity is no longer valid (i.e. cannot be described by Lambert-Beers law). Depending on instrument and measuring conditions, the linear dependence is only present at absorbance below approximately 0.05-0.1 [52]. The concentration quenching can be reduced or removed by reducing the absorbance in the sample by either diluting the sample or by reducing the pathway. For solid

samples or other samples that cannot be diluted, a way of reducing the pathway in fluorescence measurements is to change the measuring geometry from a right angle setup to a front face setup, where fluorescence is measured on the surface of the sample, and the pathway is reduced to the penetration depth of the light into the sample. It is also possible to correct for inner filter effect by normalizing the intensity to the absorbance at excitation and emission wavelengths [52]

Concentration quenching in blood plasma

Blood plasma is a highly absorbant and thus dilution or other precautions must be taken against concentration quenching. Wolfbeis and Leiner, some of the pioneers in fluorescence measurements of blood (see later) suggested a 500 fold dilution in the Ultra Violet (UV) area (200 – 400nm) and a 20 fold dilution in the visual area (400-750nm) [104], others have suggested diluting to Optical density of 0.5 [64]. In an important study by Nørgaard et al. (see later), they adapted the Wolfbeis and Leiner dilutions but then added undiluted serum samples, which they measured in a front face setup to reduce concentration quenching. If concentration quenching can be reduced, there is a good rationale for measuring on the undiluted samples. Dilution is laborious and introduces an extra operational step where errors can be made. There are some other risks connected with dilution. One is that some compounds are diluted to a concentration below the detection limit, and thus potential discriminators are removed from the matrix. Another risk is the change in pH and/or polarity of the samples that dilution can cause, which can affect the emission profile of the sample. See example of how tryptophan emission is red shifted in the diluted sample in figure 4.

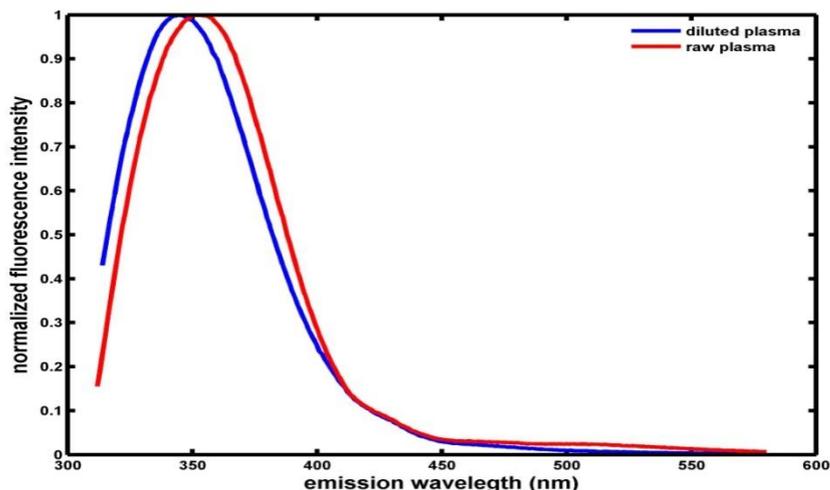


Figure 4: Emission spectra (excitation at 295 nm) of a plasma sample undiluted and diluted 100 times in PBS buffer (pH 7.4). Figure adapted from paper III.

In the study with colorectal cancer conducted in connection to this PhD project a 100 fold dilution was chosen for the whole spectral area and like Nørgaard et al. the undiluted samples, in this case blood plasma, were included. For practical reasons it was not possible to measure the undiluted samples in front face geometry, instead in a standard right angle setup, but in a special cell with a shorter pathway in the excitation direction to reduce the absorbance, and hence the amount of concentration quenching. To minimize the laboratory work, the 100 fold dilution was chosen as a compromise for both the ultra violet and the visual spectral area. It might not be sufficient to give a linear dependence between intensity and concentration in the ultra violet spectral area, and there is a risk that the measured spectra to some extent are affected of concentration quenching, which can influence the achieved results.

Chapter 3: Standardization and Quality Assurance of Fluorescence Spectra

Fluorescence spectroscopy has many advantages as explained in the previous chapter, but there are also some drawbacks. Fluorescence is generally dependent on the instrument used for data acquisition. Instruments differ in spectral resolution, and in wavelength accuracy in either the emission or excitation channel. The same sample measured on different instruments can give different results in both intensity and spectral characteristics. In order to compare results, and/or to pool data from different instruments to use in a joint data analysis, the data needs to be corrected. In Figure 5 is an example where spectral correction is needed. The same solution of the fluorophore DCM is measured on two different instruments, a difference in both spectral shape and maximum position is observed.

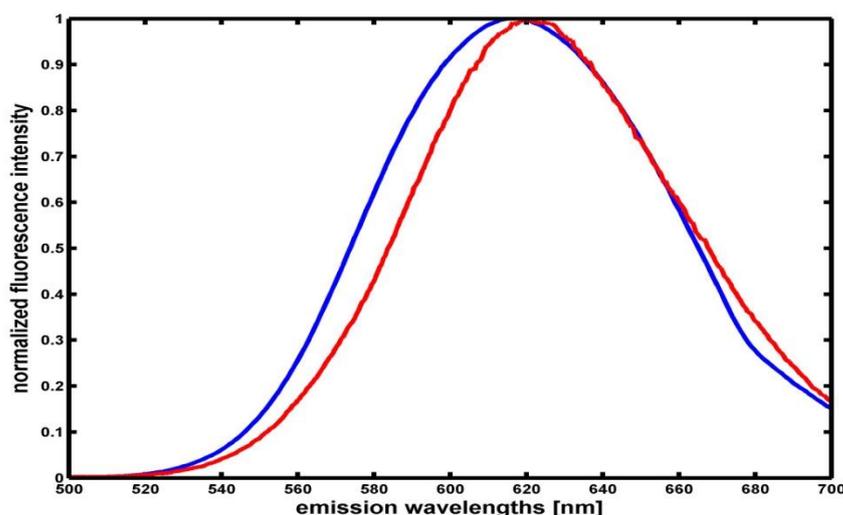


Figure 5: A solution of the fluorophore DCM measured on two different instruments. Spectra are normalized to maximum intensity. Blue is measured on a FS-920 (Edinburgh Instruments) and red is measured on a LS-55 (Perkin Elmer) same settings are used.

Depending on the purpose of the measurements and the aim of the subsequent data analysis, there are different needs for spectral correction and different methods to apply. Smaller independent measurements for feasibility studies or internal evaluation measured over a short interval of time needs none or minor corrections. On the other hand, a much more thorough correction is necessary if measurements are part of a larger study with a global perspective, or if the samples are measured over a longer period of time or on several different fluorescence instruments.

For spectral correction, different options/standards exist depending on the purpose. One class of standards are standards that are used for determination and correction of instrument dependent spectral bias. These can be divided into two classes; physical standards and spectral fluorescence standards. Physical standards are standardized light sources and/or detectors that can be mounted within the sample compartment in the instrument. These standards are expensive in use and require expert skills to use, and are thus not convenient for the broad community of fluorescence instrument operators [35;83]. They are typically used by national metrological institutes such as the National Institute of Standards and Technology (NIST) and the German Federal Institute for Materials Research and Testing (BAM), or the larger instrument suppliers to correct new instruments. Most new instruments are thus “born” with a correction file made for this specific instrument [14;35;83]

For the average fluorescence spectroscopy user, the spectral fluorescence standards are the obvious choice. A spectral fluorescence standard is a chemical compound with a known and stable spectral profile. Though finding suitable fluorescence standards has not been easy, the optimal fluorophore has a broad and unstructured emission spectrum, with little overlap between emission and excitation spectrum, small temperature dependence and of course a very stable emission profile[84]. Thus until recently only one certified reference material was available; Quinine sulphate from NIST [98] and only covering the spectral

area from 395 to 565 nm. Within the last few years more focus has come to this area, and two sets of standards have been made commercially available. A kit of five standard solutions from the BAM [80;84] covering the spectral area from 300 to 800 nm and a set of two cuvette shaped dyed glass standards from NIST that cover the spectral area from 395 to 780 nm when combined with quinine sulphate [14-17]. The BAM kit is even accompanied by a software (linkcorr®) that easily makes the correction file, with an attached estimate of the uncertainty. The corrections performed within the software are based on an algorithm made by Gardecki & Maroncelli (1998) [24], that fits a common, smooth correction factor for the whole spectral area covered by the standards.

These new initiatives have made spectral corrections of instrument specific spectral bias more accessible for the average user, though as the following will show, there is still some work to be done.

Test of BAM Standards

One of the aims of this PhD project was to explore the possibilities of a model for early cancer detection based on fluorescence measurements. To apply such a model clinically there is indeed a need for quality assurance of the data, and spectral correction is thus vital. Therefore a test of the BAM kit has been made, doing a small inter laboratory calibration of three instruments.

For the test we used two different BAM kits, one was used on two instruments on University of Copenhagen and one kit on an instrument at Danish National Environmental Research Institute. The instruments on University of Copenhagen were an LS-55 from Perkin Elmer, and an FS-920 from Edinburgh Instruments, the instrument at Danish National Environmental Research Institute was a Varian Cary Eclipse. By applying different kits at the two sites we have a situation similar to what we will find when data from two different laboratories should be compared. The achieved correction files from the BAM software linkcorr®

were then later used to correct emission spectra from a set of three fluorescence reference materials. (see below in figure 7). Results from the BAM solutions are shown in figure 6. The left column of plots in figure 6 shows the measured and the technical BAM spectra from each instrument. The technical spectra are the “true” emission profile of the standards as reported by the supplier, and it is clear that there is a difference between technical and measured spectra, and also a difference at spectra measured at the different instruments. Hence, there is a need for a calibration of the instruments. The correction factors in the middle plots have similar shapes for the all three instruments. The huge difference in the scale on the y axis is due to the different intensity scales in the instruments. The right columns of plots show how there is a reasonable agreement between the technical BAM spectra and the measured BAM spectra after correction. As the technical spectra are the ones used as target spectra for the correction, a good agreement is expected between the corrected spectra and the technical BAM spectra. Some deviations are seen between the technical and corrected spectra in the overlap between some of the emission standards. The deviations occur at the beginning or the tail of a spectrum, in the area, where the overlapping spectra have a higher intensity. The reason for this is found in the algorithm used for combining the correction factors for each of the standards. The standards have overlapping spectra, and for two overlapping spectra the spectrum with higher intensity is used to obtain the correction factor.

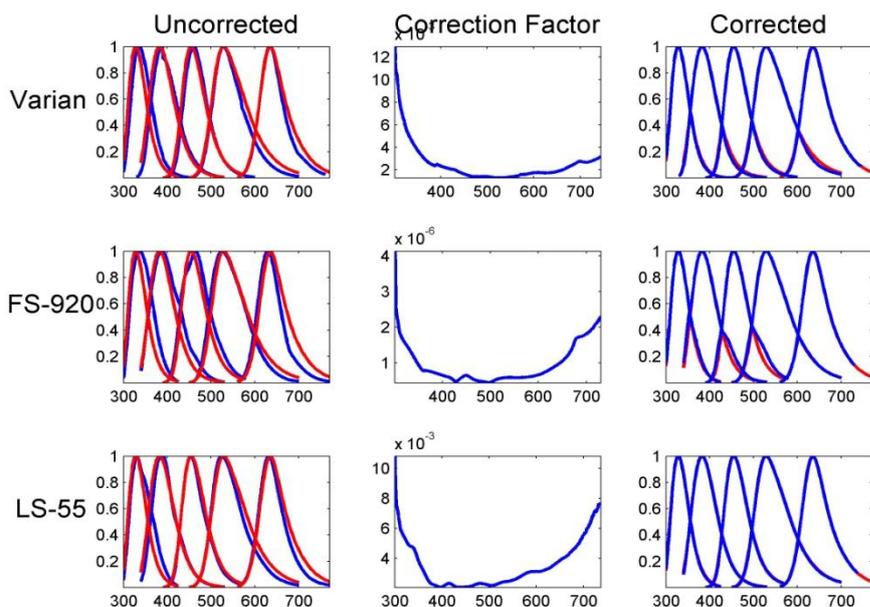


Figure 6: BAM correction applied on three different instruments. The red spectra are the technical BAM spectra, the blue spectra are the measured spectra. The correction factors are derived from the linkcorr® software

In order to test the derived correction factors, we applied them on the set of three fluorescence standards in polymethyl methacrylate (PMMA) matrix from Starna covering the spectral area from approximately 300-600 nm, all equally measured on the same three instruments. If a proper emission correction is applied, we expect spectral agreement between the corrected spectra of the standard blocks from the three instruments. However this is not the case. As can be seen in figure 7, the correction does improve the agreement of some of the spectra slightly, but for others the agreement after correction is worse than before.

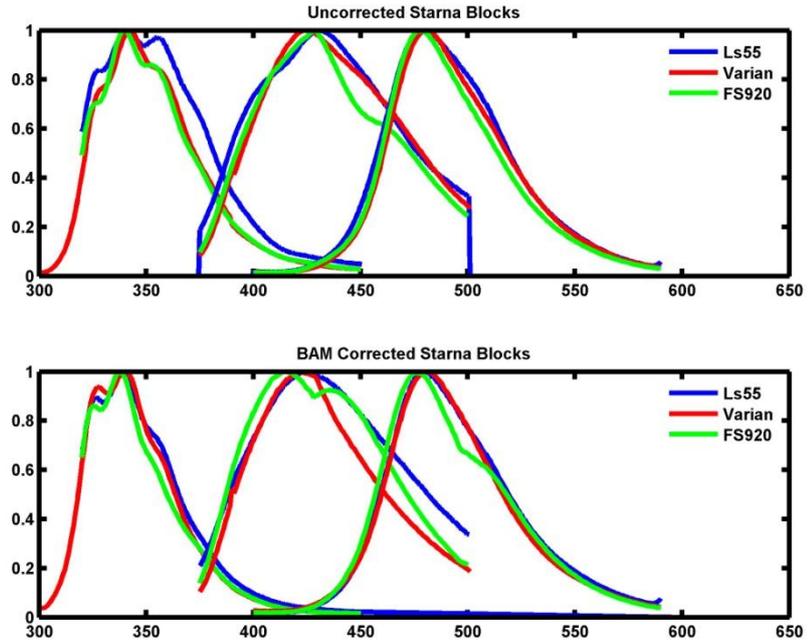


Figure 7: Starna reference blocks 3-5 measured on three different instruments; before (upper plot) and after (lower plot) BAM emission correction.

For Starna block three and five (first and last spectrum), the correction seems to have some positive effect. Especially between the LS-55 and the Varian, the spectra are well aligned after the correction. For Starna block four (middle spectrum) the agreement between spectra from the three spectrometers is worse after than before correction. Both the spectra from the FS920 and the Varian have strange curvatures. This is illustrated in figure 8, where we see that the BAM corrected spectrum from block four measured on the FS-920 has a strange curvature exactly in 428 nm which is the point of intersection between the two BAM spectra, and there is a similar dip in the correction factor from the FS920 in 428 nm (inserted in the figure).

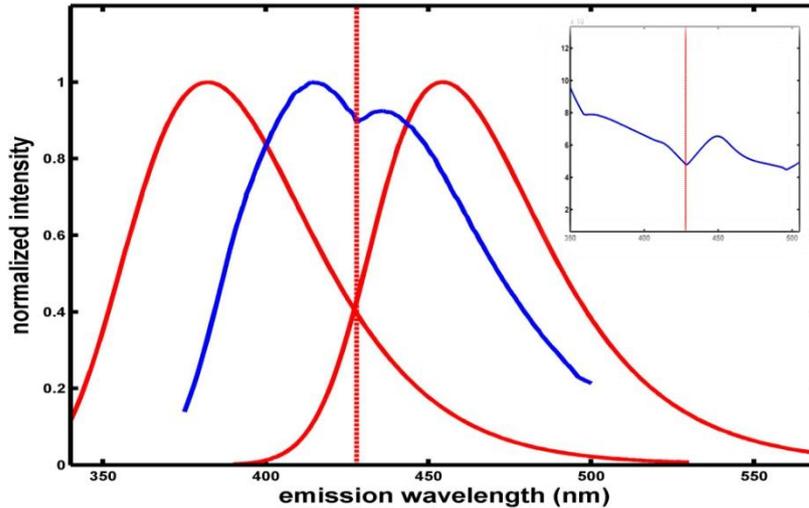


Figure 8: BAM standards 3 and 4 (red) and BAM corrected Starna Block 4 (blue) measured on the FS-920 instrument. Inserted is a section of the BAM derived correction factor from the FS-920 instrument. The vertical red line is at the intersection of the BAM standards at 428 nm in both main figure and inset.

For the Varian the deviation also occurs just after the point of the overlap between two BAM standards. It is unclear if the reason is a problem with the algorithm aligning the correction factors derived from the different standards. The algorithm calculates the correction factor for each BAM standard as the relation between the technical spectrum and measured spectrum. The individual correction factors are then normalized and intersections are smoothed by a weighted average (weighted by the measured intensity) of a range of 8 nm on each side of the intersection of the overlapping spectra [24]. Changing the spectral range of averaging did not solve the problem. It could be interesting to try a different algorithm to see if that could derive a more stable correction factor. Another reason for the problem could be that the signals of the standards are insufficient in that area, indicating that an extra standard covering that particular spectral area is needed to obtain a better correction. No suitable standard was found to test this.

The above result is by no means a proof that the commercial BAM correction kit does not work, but it is clear that even with the ready to use kit as the BAM, spectral correction is not

straightforward to perform, and more work needs to be done in order to make proper and simple spectral correction.

Excitation correction

The spectral output of the light source is different over the wavelength spectrum. In addition, the intensity at each wavelength can change with time, and finally fluctuations in the lamp can appear. Correction of the excitation channel is thus also necessary. Most spectrofluorometers are equipped with a reference detector located between the excitation monochromator and the sample compartment. A beam splitter leads a fraction of the excitation light to the reference detector, which will then correct the final result for the wavelength dependent output, by normalizing to the reference signal [52]. Otherwise, correction of the excitation channel is typically done using a quantum counter, which is a compound with a constant emission rate independent of the excitation wavelength. Thus the emission output is proportional to the output of the light source, and the corrected quantum counter excitation spectrum should be a flat line. Quantum counters applied are often Rhodamine B or Rhodamine 101. The latter is less temperature dependent and has a broader wavelength range. [45;68;69]

Intensity Calibration. (Paper I)

The last step in the calibration/data correction before data analysis is the intensity correction. This is necessary for making quantitative comparisons between fluorescence data from different instruments, or data from same instrument made with different instrumental settings. Intensity correction is fairly simple, as it is done by relating the intensity of the acquired spectra to the spectra of a known standard measured at the same settings. The challenge is to find a suitable standard. Typically quinine sulphate or recently the standards from NIST [16;17] has been used, as they have a stable intensity profile. An alternative and less widespread method is to use the Raman scatter peak of water [18;98]. Raman scatter is a physical property of water, and

the intensity of the Raman peak is theoretically connected to the excitation wavelength. The Raman peak is therefore an excellent stable standard for intensity. The Water Raman approach has mostly been applied by use of only the maximum intensity (Peak height) of the emission of Raman spectrum [13]. In Paper I a slight alternative approach where the whole integral of the water Raman peak is used to correct for intensity is described [54]. The integral of the peak, or the area under the peak (A_{rp}), is defined as the area calculated by the trapezoidal rule covering the spectral area in an interval of peak maximum $\pm 1800 \text{ cm}^{-1}$ at the Raman peak following the 350 nm excitation. This equals the area spanning the spectral area from 371 nm to 428 nm. The relatively broad area is defined in order to ensure that the intensity on either side of the peak should be low and within the area of instrumental noise. The spectrum to correct is simply normalized to the integrated area of the Raman peak, and the intensity of the fluorescence spectrum becomes relative to A_{rp} on a new scale of Raman units (R.U.).

The shape/broadening of an emission spectrum depends of the instrument settings, especially the slit widths. The width of the Raman peak is accordingly dependent on the settings, and by using the whole integral of the Raman peak, instead of only the intensity at peak maximum is also possible to intensity correct data from different instrument settings to the same scale of Raman units (Figure 9). The correction is dependent on a suitable signal to noise level of both the spectrum to correct and the Raman peak; the example in the figure below with ex/em slit of 1.5/5 nm illustrates this.

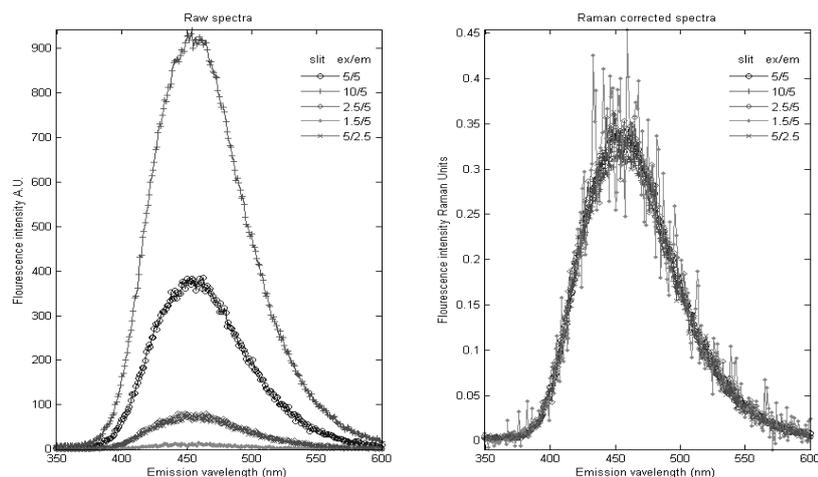


Figure 9: Spectra of quinine sulfate solution obtained at different slit settings (left) before and (right) after Raman correction. In the left plot all spectra are on the same scale of Raman Units. The spectrum with slit settings of 1.5/5 nm has a low intensity in the raw spectrum (as has the Raman peak at the same settings) and hence a low signal to noise value, this is the reason for the very noisy corrected spectrum. Figure adapted from Paper I

By presenting fluorescence intensities on a relative scale of Raman Units, fluorescence results become inter comparable between instruments, independent of the original scale of the instruments, provide that the same excitation wavelength is used for the Raman peak. There is a strong wavelength dependence of the area under the Raman peak of λ^{-4} [7], thus it is important to report the excitation wavelength used for correction. The scale of R.U. is becoming standard within the field of aquamarine science and the area of fluorescence measurements of dissolved organic matter [91].

Chapter 4: Chemometrics/Data Analysis

The general topics for this thesis are within the fields of metabonomics and spectroscopy, and the analysis of data from these disciplines. Very often metabonomic data are synonymous with spectroscopic data, and thus the areas are in many cases coinciding. Common for both, coinciding or not, are the often large number of variables compared to samples, a situation that exclude us from using traditional statistic methods when analysing the data. Instead multivariate statistics/chemometric methods are applied. The chemometric methods applied in this thesis are mostly standard methods that are all thoroughly described in the literature. Thus, in most cases only a brief summary is given here.

Chemometrics is the discipline of extracting relevant chemical information from often complex data structures acquired by measuring on any chemical/biological matrix. A huge advantage of chemometric tools is that they can often be used when classical statistical tools have problems. For example spectroscopic data, which often consist of highly correlated data points (e.g. absorbance of neighbouring wavelength points), is a challenge for traditional statistical methods. In chemometrics, linearly independent latent variables, which reflect the major variations/trends in data, are extracted. Chemometrics can thus reduce complex data structures to more simple systems with few latent variables that describe the important chemical variations in the samples.

PCA

One of the most fundamental and most applied chemometric methods is Principal Component Analysis (PCA). PCA is a useful

tool to get an overview of the data, to see initial clustering or to detect outliers [38;100;101].

Given a data matrix \mathbf{X} of size $i \times j$ (objects \times variables); PCA will reduce \mathbf{X} into a systematic part and a residual (noise) part. The systematic part consists of possibly few latent variables, principal components, which summarize the most important variance in the data. The residuals are the part of \mathbf{X} not explained by the PCA model. The projection of I objects in \mathbf{X} onto the first loading vector provides the score values of the first component, \mathbf{t}_1 . The direction of maximum numerical variation in the J dimensional variable space, is then described by the first loading vector \mathbf{p}_1 . The PCA decomposition can be described by the following equation:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

\mathbf{T} is the score matrix, \mathbf{P}^T is the transposed loading matrix, and \mathbf{E} represents the residuals. The scores and loadings are determined so as to minimize the residuals in the least squares sense [101].

Multi-Way Data Analysis – PARAFAC

PCA is a method for data in matrices (two-way data). When extending to three-way data as for example in fluorescence Excitation Emission Matrices (EEMs) (samples \times emission \times excitation), PCA cannot work directly on such a data array. It is then possible to unfold the data to a two-way matrix or it is possible to apply multi-way techniques such as PARAFAC or Tucker 3 directly on the data array, and thus exploit the second-order advantages of the three way structure [89].

PARAFAC is the multi-way analysis of choice in this thesis due to its great advantages when applied on fluorescence EEMs where it can give estimates of the underlying emission and excitation spectra (se later). PARAFAC is based on the work of Cattell (1944) [11], and originally presented by Harshman (1970) [30] and

Carroll and Chang (1970) [10]. PARAFAC can be seen as a generalization of PCA to higher order data [8]. The data array is decomposed into trilinear components of three loading vectors, often described as one score vector and two loading vectors (in case of higher order data the decomposition is extended to quadrilinearity, quintilinearity,... etc.) . The part of $\underline{\mathbf{X}}$ which is not described by the components is the residuals. In the perfect model, the sum of the components (the model) explains all the systematic variations in $\underline{\mathbf{X}}$ and leaves all noise in the residuals (see graphical example in figure 10 below). The parameters of model are estimated as to minimize the sum of squares of the residuals in the equation

$$\underline{\mathbf{X}}_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk}$$

where a_{if} , b_{jf} and c_{kf} are the i^{th} elements of the loading vectors for the f^{th} PARAFAC component.

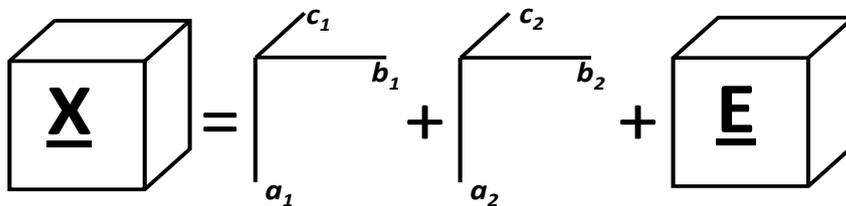


Figure 10: The decomposition of $\underline{\mathbf{X}}$ in a two component PARAFAC model into scores (a), loadings ($b+c$) and residuals (E) presented graphically.

PARAFAC and Fluorescence Spectroscopy

One important feature of the PARAFAC model is the uniqueness of the solution, as opposed to a bilinear model which has rotational freedom. Thus if the correct number of PARAFAC components is used on data with an approximately true trilinear structure and an appropriate signal to noise value, the solution from the PARAFAC model will give estimates of the true

underlying profiles of the variables [4;8]. This makes PARAFAC perfect for fluorescence spectroscopy when applied on EEMs. The loadings and scores can be treated as estimates of the excitation and emission spectra, and relative concentrations of the fluorophores in the samples respectively [4;8]. Below (Figure 11) is illustrated a decomposition of fluorescence EEMs into estimates of the true underlying excitation and emission spectra of the present chemical fluorophores. The sample depicted (contour plot lower right) is an EEM in the UV-area of a sample from an experiment with effect of detergent (reduced Triton X) on a folate binding protein (FBP); the experiment is fully described in [36]. It is clear from the contour plot that there are two defined peaks which both have excitation maximum around 280 nm, and emission maximum at approximately 320 and 350 nm respectively. The best PARAFAC model on the data though, gives three components in this case. Thus, we can extract information on three chemical compounds from the sample. From the excitation loadings (lower left figure) it is seen that all three have excitation maximum around 280 nm, which is the typical excitation maximum of the amino acid tryptophan, the expected dominating peak in protein emission. The reduced Triton X has excitation maximum at the same wavelength. The emission loadings though (upper right figure) all have different maximum values. The reduced triton X is known to have maximum at approximately 300 nm (the green loading), and the two other can be assigned to differently located tryptophans. In figure 12 the score values representing the relative concentrations of the three components are plotted as a bar plot. The PARAFAC solution thus allows us to describe the set of complex EEMs from the samples, as a matrix of concentrations of three defined chemical compounds. This nice relationship has made the combination of PARAFAC and fluorescence spectroscopy to a well established tool [4;12;91].

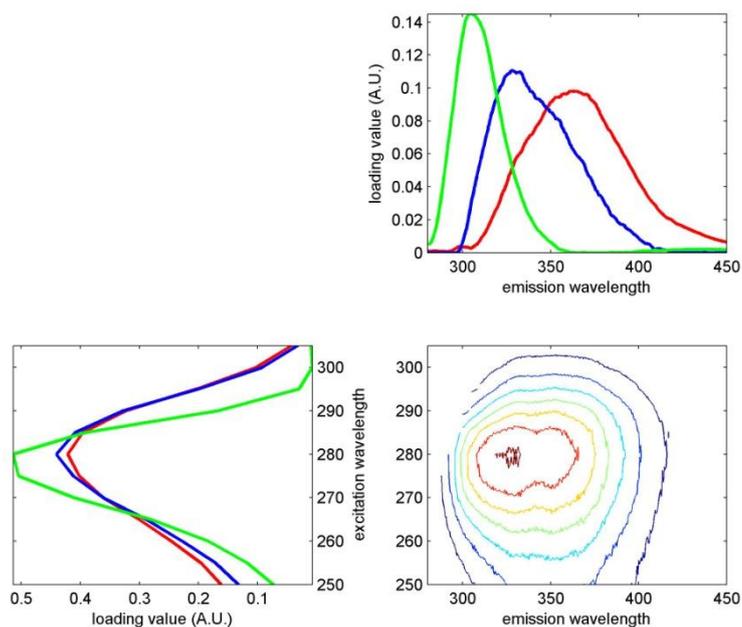


Figure 11: An example of a PARAFAC decomposition of an EEM; lower right plot is an example of a contour plot of an EEM of a mixture of protein and detergent. Upper right and lower left plot are PARAFAC emission and excitation loadings respectively.

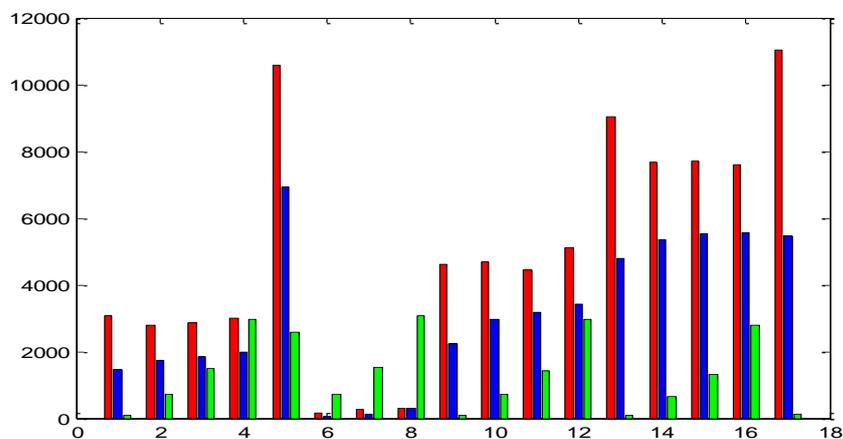


Figure 12: Score values from the PARAFAC model, this represents the relative concentrations of the three components in the samples.

PLS

PCA and PARAFAC are both unsupervised methods used for exploring data or data mining. Partial Least Squares (PLS) regression is a regression method for establishing a mathematical

relation between \mathbf{X} and \mathbf{Y} . As opposed to PCA and PARAFAC, PLS is a supervised multivariate method for two-way data where a set of dependent variables, held in a matrix \mathbf{Y} (or a vector \mathbf{y}), are introduced. PLS regression will then find the variation in \mathbf{X} that best describes the covariance between \mathbf{X} and \mathbf{Y} [26;102;103]. PLS regression can be described by the equation

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

Where the regression coefficients are found by maximizing the covariance of the scores in a “PCA-like” decomposition of \mathbf{X} and \mathbf{Y} described by

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}$$

\mathbf{T} and \mathbf{P}^T are the scores and transposed loadings in \mathbf{X} , and \mathbf{U} and \mathbf{Q}^T are the scores and the transposed loadings of the \mathbf{Y} space, \mathbf{E} and \mathbf{F} are the residuals in \mathbf{X} and \mathbf{Y} respectively [26; 103]. If \mathbf{Y} is a matrix with several dependent variables, PLS is denoted PLS 2, when \mathbf{y} is a vector of only one variable it is called PLS1 [103].

Classification

In chemometrics and statistics one is often trying to solve the problem of classification of samples into classes based on measurements of various parameters (quality measurements, chemical profiles, spectral profiles, etc.). Several methods exist to perform classification; which is the better depends on the nature of the data and the purpose of the analysis. A classical method is Fishers Linear Discriminant Analysis (LDA) from 1936 [22], or its closely related Canonical Discriminant/Variate Analysis (CDA or CVA) [39;40]. Discriminant analysis seeks the direction in the data that maximizes the distance between the groups, as opposed to PCA which will find the direction with maximum variance in the data. This is illustrated in the figure below (Figure 13); PCA would find the direction of the blue arrow as the direction of the

largest variance, whereas a discriminant analysis would find the magenta direction as the direction separating the groups [6].

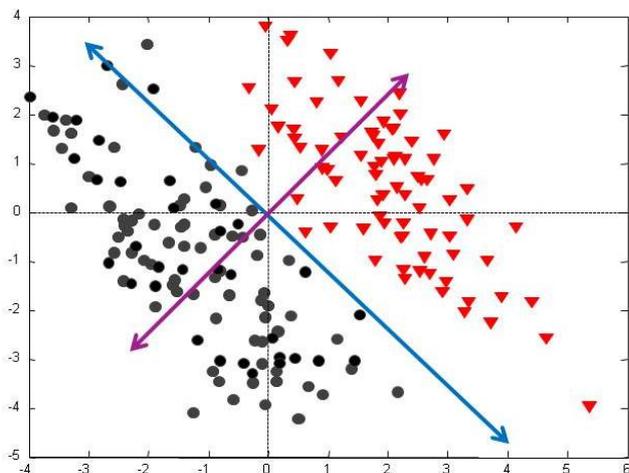


Figure 13: Major directions found in a dataset by either PCA (blue line) or LDA (magenta line)

LDA and CDA are still widely applied methods, and in case of full rank linear data they might still be the best methods in terms of misclassification rate [31]. For nonlinear data, a quadratic version of LDA exists (QDA) and likewise more advanced techniques that will be able to handle such data [87]. The problem with LDA arises when data does not have full rank (rank deficiency) due to either more variables than samples, or highly correlated variables. LDA cannot be applied directly on such data due to a noninvertible covariance matrix. In that case, the rank of the data must be reduced by for example PCA prior to using LDA. Another possibility is to use methods like Extended canonical variate analysis (ECVA) that solves the eigenvector problem in CVA by finding the direction of maximum distance between groups by applying PLS, and hereby overcome the rank deficiency problem [76].

PLS-DA is a popular classification method applied in chemometrics and often applied in metabonomic studies. PLS-DA classification is a discriminant analysis like LDA, and as illustrated in figure 13, PLS-DA will also search for the direction

that best separates the groups. Basically, PLS-DA is a PLS regression, but instead of using a continuous \mathbf{Y} , \mathbf{Y} is a binary dummy matrix representing class membership. For a dataset with i samples and k classes, \mathbf{Y} is a matrix of size $(i \times k)$ where each row contains $(k-1)$ zeros and the value of one in the column representing its class. Modelled class membership is calculated from the predicted \hat{Y} value according to a given threshold value. Given a threshold value of say 0.5, a predicted value $\hat{y} > 0.5$ means that the sample is assigned to the class, and a $\hat{y} \leq 0.5$ means that the sample is not assigned to the class [103]. A threshold value of 0.5 is not necessarily the best solution. Any threshold value between 0 and 1 can be used dependent on the problem at hand.

PLS-DA is widely applied in metabonomic studies and some criticism has been made [50]. A common problem is for example lack of or improper validation. With the use of a dummy y consisting of only zeros and ones over fit is a huge risk if proper validation is not applied. In the often very large number of variables used in e.g. "omics" or spectroscopic applications there is a great chance, that even if there is no correlation of interest between \mathbf{X} and \mathbf{y} , it is possible to find an arbitrary direction in the \mathbf{X} space that correlates nicely to the zero-one direction in \mathbf{y} . A proper validation of the model would show that this correlation is not valid.

Another important thing to remember in PLS-DA is how to choose the number of Latent Variables (LV). In PLS the right number of LV's to use is based on evaluation of root mean squared error of prediction or cross validation (RMSEP / RMSECV), or predicted residual sum of squares (PRESS) which is a measure of the total prediction error. In PLS-DA the \mathbf{Y} values of 1 and 0 have been set to the samples to state class membership, and are not really related to the nature of the data, and thus we cannot expect perfect prediction. From a classification perspective it is of no interest if the sample is predicted as 0.95 or 0.85, where the latter will result in a higher RMSECV, in both cases the sample would be classified as belonging to the class. Instead of

prediction error the classification error is a much better criterion to evaluate upon, as it is more important for model performance [50].

Sensitivity and Specificity

The performance of a binary classification like LDA or PLS-DA is often given in terms of sensitivity and specificity. Sensitivity is the measure of positives that are correctly classified as positives (true positives) as a fraction of all positives, and specificity is the measure of negatives that are correctly classified as negative (true negatives) as a fraction of all negative samples [3;70]. A classification with perfect discrimination (no overlap) between two classes will thus result in a sensitivity and specificity of 100%. The concept of sensitivity and specificity is closely related to the concept of type I and type II errors. A low sensitivity, i.e. a high rate of false positives is synonymous with a high rate of type I error, and in parallel, a low specificity is synonymous with a high type II error meaning a high rate of false negatives.

Classification models for discrimination between two groups, e.g. diseased and healthy, are based on some defined threshold value of a certain classifier, for example the concentration of a biomarker or a predicted value from a multivariate projection. In a perfect (100%) classification the threshold value is naturally given by the class separation though this is seldom the case. Often there will be an overlap between the groups, and the threshold value must be determined by the analyst. When choosing a threshold value in a non-perfect classification it is a trade-off between the specificity and sensitivity values [25]. In for example diagnostic tests, a false positive result can be expensive due to unnecessary and sometimes high risk follow up tests and it can cause unnecessary anxiety for the patient [9]. A high specificity is thus preferred in these kinds of tests and a stricter threshold can be set to obtain this, the trade-off is then a higher amount of false negative and a lower sensitivity [108;109]. The threshold value, where the sum of sensitivity and specificity is maximized, can be found as the intersection between the probability distributions for

the outcome of positives and negatives from the test. The relation between “threshold”, sensitivity and specificity can be illustrated in a Receiver Operator Characteristic (ROC) curve [70] (see example in Figure 14).

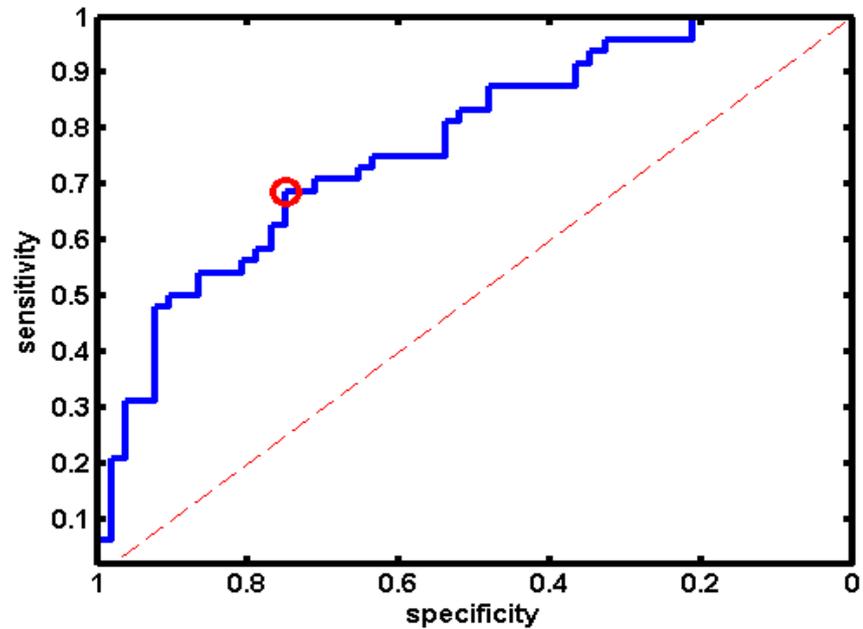


Figure 14: Example of a ROC-curve. The red marker on the curve represents the point of maximized sum of specificity and sensitivity. The dashed diagonal represents the random outcome line. The abscissa is reversed to represent specificity values; alternatively it can be (1-specificity).

In the ROC-curve the sensitivity is along the ordinate and 1-specificity (i.e. false positive rate) along the abscissa (some time it is depicted with the specificity along a reverse abscissa). The upper left corner then represents the perfect classification with a sensitivity and specificity of 100%, and the diagonal line represents a random outcome. The closer the outcome is to the top left corner the better [70]. In a perfect two-class classification model the ROC-curve will follow the ordinate and the upper limit of the scheme and thus give an area under the curve (AUC) of 1. For non-perfect models the AUC will be smaller than one. The extent of overlap between the groups is decisive for the AUC; a large overlap will give small sensitivity and specificity values independent of the threshold value. A total overlap and thus no discrimination will give a ROC curve following the diagonal, and an AUC of 0.5. AUC's lower than 0.5 could indicate a wrong

hypothesis, and by inverting the test a good classification could be obtained [108;109]. The ROC curve and the AUC allows us to compare different diagnostic tests at any given specificity or sensitivity value independent of the prevalence of the disease [70]

Rotation of PCA Scores (Paper II)

PCA scores and loading plots are good reviews of the major trends in data but sometimes the result can be difficult to interpret. Imagine that we are specifically interested in how specific variables influence the variation in data, and we then have a complex solution where these variables have similar loading values in all components, then it would be difficult to draw any conclusions regarding their influence in the samples. In that case it might help to look at the system from a different perspective. This is possible by rotating the solution towards a more simple solution. The solution of a PCA model is not unique, meaning that there is rotational freedom in the model. The scores or loadings in the model can be rotated if their associated loadings or scores counterparts are similarly counter-rotated [14]. The rotation principle can be described as follows:

If we define an orthogonal $m \times m$ rotation matrix \mathbf{Q} ($\mathbf{Q} \times \mathbf{Q}^T = \mathbf{I}$), we can rotate a PCA model by \mathbf{Q} , simply by multiplying the original score and loading matrices \mathbf{T} and \mathbf{P}^T by \mathbf{Q} and \mathbf{Q}^T , whereby the rotated scores, \mathbf{S} , and the new rotated basis, \mathbf{M} , are obtained:

$$\mathbf{TQ} = \mathbf{S} \text{ and } \mathbf{Q}^T \mathbf{P}^T = \mathbf{M}^T$$

This means that

$$\mathbf{TP}^T = \mathbf{TQQ}^T \mathbf{P}^T = \mathbf{SM}^T$$

The original PCA model is then converted into the rotated model, $\mathbf{X} = \mathbf{SM}^T + \mathbf{E}$ with scores \mathbf{S} and loadings \mathbf{M} that are rotated versions of \mathbf{T} and \mathbf{P}^T . The new model explains exactly the same

variation, though with different components. Notice that the components are no longer principal components as they no longer represents the original directions found in the least squares fit.

The idea behind rotation of a PCA model is to establish better conditions for interpretation of the model. This is typically done by rotating towards a more simple structure. In the example below (Figure 15), orthogonal rotation of the loadings has been applied on the loadings from a PCA model on some chromatographic data. In the left plot which shows a section of the original loadings, the information about the compound in the chromatogram is spread over five components, whereas in the right plot with the rotated loadings, almost all the information is concentrated in only one component. Thus we can find samples containing this compound in the direction of this rotated component only.

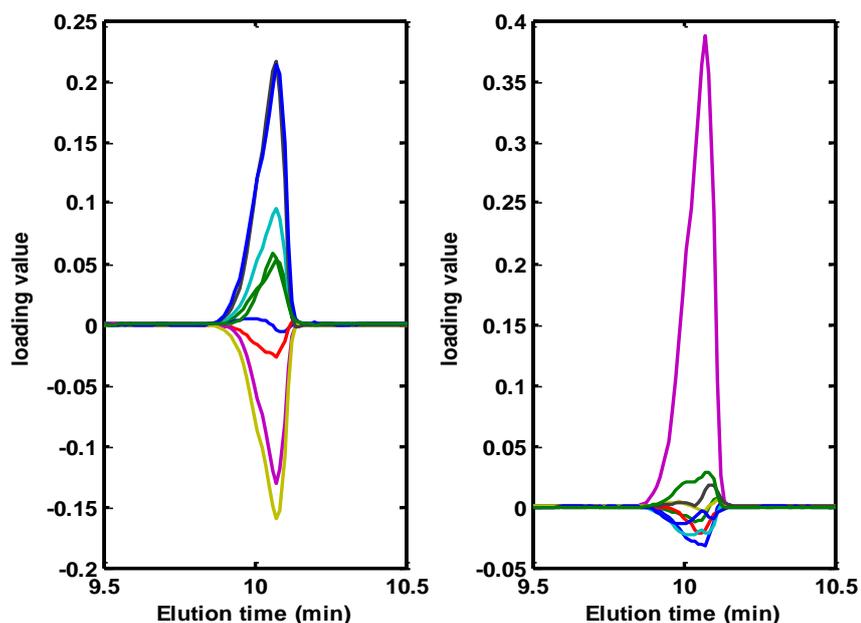


Figure 15: loadings from a PCA model on chromatographic data. Left: original loadings, right: rotated loadings.

There are many different principles on how to determine the rotation matrix \mathbf{Q} . In the above example the varimax criterion has been applied. The varimax criterion suggested by Kaiser [43] is the most often applied criterion for orthogonal rotation of any

coordinate system. It is often described together with the quartimax rotation principle under the common name orthomax rotation [28] maximizing the objective function:

$$V = \sum_{j=1}^J \sum_{f=1}^F p_{jf}^4 - \frac{\gamma}{J} \sum_{f=1}^F \left(\sum_{j=1}^J p_{jf}^2 \right)^2$$

p_{jf} is the rotated loading value for the j^{th} variable on component f , $j = 1, \dots, J$ are the variables, and $f = 1, \dots, F$ the components; γ ($0 \leq \gamma \leq 1$) determines the rotation, for $\gamma = 0$ the function becomes the quartimax criterion and if $\gamma=1$ it is the varimax criterion. The typical differentiation between the two extreme methods of the orthomax criterion is that the varimax criterion is said to simplify the columns of the loading matrix, whereas, the quartimax criterion simplifies the rows [47]. Thus maximizing the varimax criterion provides a solution where the loading values in one specific component are either high (in absolute value) or close to zero to the extent possible. Maximizing the quartimax criterion provides a solution where one variable will have high loading values in only one component, and low in the others. A drawback of quartimax can be that it often leads to one loading vector representing a general offset in the data whereas this is not typically the case for varimax.

Theory of rotations is primarily described in connection with psychometrics, and only few applications are published within chemometrics and natural sciences. There are some differences in how PCA are performed in psychometrics and chemometrics that has an influence on the rotation criterion. In chemometrics the loadings are typically normalized columnwise to a unit length of one whereas in psychometrics normalization of loadings is uncommon. Normalization of the loadings will cause that the two “extreme” solutions of the orthomax criterion; varimax and quartimax, or in fact all solutions to orthomax, will provide the same solution [29;48]. The reason for this is illustrated below. Since rotations primarily are described in connection to

psychometrics this fact is sometimes ignored, and thus not often recognized in the chemometric world.

Looking at the second term of the orthomax equation above

$$\frac{\gamma}{J} \sum_{f=1}^F \left(\sum_{j=1}^J p_{jf}^2 \right)^2$$

The squared loading elements of each column are summed, and thus for the normalized loadings the sum $\sum_{j=1}^J p_{jf}^2$ is equal to one regardless of rotation. Hence, the second term of the equation will be constant, and maximizing the whole criterion will only be a matter of maximizing the first term. This will reduce any orthomax criterion, when applied to normalized loadings, to the quartimax criterion.

The problem is of course only relevant when rotation is applied to loadings. Following the 'symmetry' of the PCA model it is equally possible to rotate the scores in the PCA model towards simple structure. The scores are not usually subject to normalization and orthomax rotation is then dependent on the value of γ .

In Paper II rotations of both scores and loadings are implemented in an application of metabolic profiling of St. Johns Worth. Rotations enhanced interpretation of the metabolic background of sample clustering, on both a level of individual variables and on the total profile of clustered samples. It is hereby shown how rotations with advantage can be applied to complex "omics" data to facilitate the visual interpretation, and we believe that the method has general applicability in metabonomic, metabolomic, and metabolite profiling studies.

Chapter 5: Cancer, Definitions, Detection, Screening

A large part of this thesis is concerned with measurement of autofluorescence in blood samples to detect cancer. The following will briefly describe some of the methods usually applied to detect, and screen for cancer. Focus will be on colorectal cancer.

Cancer is a disease caused by malignant cells that display a significant growth dysregulation, often resulting in tumours that have the ability to invade adjacent tissue and in some cases spread to other parts of the body (metastasis). A tumour is the result of mutations in cells that allow them to proliferate abnormally. Thus, it is a combination of “gain of function” mutations in the genes that induce growth of the cell, and a “loss of function” mutations in the genes that normally restrain growth (tumour suppressors). Cancer tumours differ from benign tumours by their ability to invade and destroy surrounding or distant tissue, whereas benign tumours are circumscribed and therefore often can be removed [49;72;79]. The reason why malignant tumours can spread to adjacent tissue lies in a change in the mechanisms on the surface of cells that control how cells interact with the extracellular matrix. The cell-cell contact is weaker and it thus allows the cancer cell to leave the tumour and grow in non-homologous tissue [51].

In this PhD project a set of blood plasma samples from colorectal cancer patients have been measured. Colorectal cancer is caused by malignant tumours appearing in the colon, rectum or appendix. The tumours stem from adenomatous polyps that develop into malignant tumours. This progress is in many cases very slow, and in the vast majority of colorectal cancers, a change in the polyps occurs up to 15 years before the malignant change

begins. The frequency of adenoma polyps in the colon increases with age, and can be identified in up to 20% of the population, but only 10% of the adenomas will develop into cancer [5;27;79].

The progression of colorectal cancer is typically staged in the Dukes classification system from Dukes A to Dukes D, depending on how deep the tumour penetrates the layers of the bowel wall and whether it has spread to the lymph and finally other organs. Other general staging systems exist for all types of cancer; for example the “Roman Numeral Staging” which classifies cancers in stages I – IV, where I is “local cancers that have not spread”, and IV is “cancers that have metastasized to other parts of the body, with II and III in between [86].

In Denmark colorectal cancer is the second and third most frequent malignant disease among women and men respectively. In all of Europe it was the most frequent cancer type in 2008, and the third leading cause of all cancer related deaths in the industrialized part of the world [20;95].

Detection and diagnostics

Traditional detection and diagnostics of cancers is done by various techniques depending on the type of cancer. Often a combination of tumour biopsy and methods like ultrasound images, Xray, CT or MR scanning is applied. Common about these methods is that they are either laborious, and/or requires expensive equipment. Especially the CT and MR scanners are extremely expensive, and require educated personnel to operate [1]. These methods are typically not applied unless there is a suspicion of cancer, i.e. the patient has experienced symptoms (for example a lump in the breast or testicles, or blood in the stool), or there is a “positive” test result from a screening program (see below). Hence these methods are best suited to diagnose and stage cancer in a progressed stage. The standard method for examination of the colon and hence for colorectal cancer is colonoscopy, but this is also an expensive and laborious method,

and there is a small risk of colonic perforation. Thus colonoscopy is not suitable for population screening.

Screening programs for cancer

By early detection of cancer, the chances for cure and survival are better, as the disease has not yet developed much, and the risk of metastases is smaller [90]. In a test of a screening program, early detection of colorectal cancer has reduced mortality by 33% in an 18 years follow up study [42]. Early detection of cancers can thus be crucial. To induce early detection of cancers huge effort is put into finding methods to do systematic screening for cancer. For some types of cancers, screening programs are established, for example mammography for breast cancer and smear test for cervix cancer. Both are in many countries offered systematically to women of a certain group of age.

Due to the slow evaluation of adenomas to malignant tumours in colorectal cancer, and the improved survival rates with early detection, a lot of effort has been put into developing screening programs for colorectal cancer. Following these efforts, large screening programs have been tested and in some countries launched as permanent programs [32;33]. The most applied method in these screening programs is the fecal occult blood test (FOBT), detecting blood in the stool that can indicate cancer. The FOBT is most often followed by a colonoscopy in positive cases [9;42].

Two different approaches of FOBT exist; one is a guaiac based method, sensitive to peroxidase activity, which reacts to the haem group in blood. The risk is that there can be false positives from other peroxidases from e.g. fruit or from haem from red meat. The other type is an immunochemical test specifically sensitive towards human haemoglobin, and this avoids diet induced false positives [42]. There are large deviations on reported sensitivity and specificity values for FOBT in the literature. Generally specificity values are higher than sensitivity values, and the immunochemical method show slightly better accuracy than the guaiac [9]. In most of the executed screening programs for

colorectal cancer the guaiac method was preferred, though recent studies recommend the immunochemical method [32-34]. Tests have proved that screening with FOBT can reduce mortality with 13-33%. A problem with the FOBT test for population screening is a low compliance; in an American screening program less than 60% of the target group were up to date with the screenings [55]. Even though the test is easy to perform, and in most cases can be performed at the patient's own home, the unpleasantness of handling in stool samples is likely to be the reason for the low compliance [55;92]. A screening method requiring only a blood or urine sample would probably give a better compliance and huge effort is put into finding a screening method based on body fluid samples, most likely as a biomarker [71;82].

Biomarkers

Since the 1960's scientists have realized that cancer or tumour cell activity can be reflected in the blood stream. Cancer cells or mechanisms connected to the cancer cells can result in reduced or elevated concentrations of some molecules. More recent results have shown that changes in DNA, of for example the growth and tumour suppressor regulations, can be seen as a result of cancer activity [90]. These molecules or DNA sequences found in the blood are called serum tumour markers or biomarkers. Due to the easiness of sampling blood, huge efforts have been put into discover and use suitable markers to monitor cancer treatment, and further to detect and diagnose cancer [81;105]. Serum tumour markers or biomarkers can be molecules such as enzymes, isoenzymes, serum proteins and hormones. The concentration of a certain biomarker can give information on the stage of cancer and can thus be used for an individual targeted treatment, and subsequently monitoring of the effect of treatment and progression of the cancer. Today more than 50 named markers are known and applied clinically. Some of the most applied markers are the prostate-specific antigen (PSA) used for prostate screening, and the carcinoma-associated glycoprotein antigen (CA-125) used to diagnose ovarian cancer [27]. Despite the heavy research in biomarkers for early detection, even the best known

and most applied markers have low specificity values leading to a high number of false positives. Elevated levels of these markers do not manifest until an advanced stage of the malignancy, and the clinical use is thus “limited” to applications of prognosis, selection, and monitoring of cancer treatment [63].

Biomarkers in colorectal cancers

For colorectal cancer a number of biomarkers have been suggested.

Carcinoembryonic antigen (CEA) is a well-known biomarker that has elevated levels in colorectal cancer patients. Elevated levels of CEA are also found in other groups; for example liver patients and smokers, and the sensitivity and specificity for detecting colorectal cancer are too low (0.34/0.93) to recommend stand-alone diagnostic use. CEA is thus primarily used to detect recurrence in patients with previous colorectal cancer, or as a supplement to other examinations in the detection of the disease [23;94].

Other biomarkers have been suggested for diagnosis or early detection of colorectal cancer, e.g. carbohydrate antigen 19-9 (CA 19-9), the Plasma Lysophosphatidylcholine Levels, free DNA, or urokinase receptor (Upar), some with more promising results than others, though none have yet been accepted for clinical use [19;23;60;71;107].

Chapter 6: Fluorescence Spectroscopy on Blood Samples in a Diagnostic Context

In the following a brief review is given of some of the work applied in the field of using fluorescence spectroscopy on human blood to diagnose or detect cancer or other diseases. An overview of some of the important publications is given in table 1.

The idea of using autofluorescence measurements of blood to discriminate cancer from non-cancer was first presented by Leiner, Wolbeis and co-workers in the 1980's [55-58;104]. They considered the fluorescence EEM of a blood sample as a "fingerprint" that can be used to monitor the health status of a person. The hypothesis was that it would be possible to observe deviations in the fluorescence spectrum from "normal" healthy subjects to diseased subjects [58]. This theory fits well into the present theories of metabonomics, and without knowing it Leiner, Wolfbeis and co-workers in fact introduced the theory of fluorescence based metabonomics. In their series of studies on fluorescence spectroscopy (EEMs) on blood (sera) from both rats and humans they discovered deviations in the autofluorescence from tumour-bearing subjects compared to healthy subjects. In rats with a hepatoma they reported a decrease in tryptophan fluorescence and an altered fluorescence from NAD(P)H. They used rats which they killed at different time of cancer progress, and observed the differences in fluorescence even at an early stage of the disease. In sera from human patients with gynecological tumours compared to age and weight matched healthy women they experienced a blue shift in the tryptophan fluorescence. Based on that study they suggested that the relative ratio between the fluorescence at 287/365 nm and 287/337 could be used to detect cancer. In a follow up study by Hubmann et al. (1990), it was suggested that an increased level of alpha-2

globulins and a decreased level of albumin in malignancies could explain the observed tryptophan shift [41;58]. Based on the findings by Leiner and co-workers, Madhuri et al. (1997) did a study where they used the ratio between tryptophan and NADH fluorescence to detect and prognose oral cancer. The results were promising, though they only used a limited number of samples. Nørgaard et al. (2007) did a study where they adapted the setup from Leiner and measured fluorescence EEMs on human serum from breast cancer patients in different stages and healthy controls. Their study though did not involve spectral assignment, but instead they used a multivariate approach (see more later), which gave good classification between cancer and control. The result was compared with three traditional tumour markers used for breast cancer, which also were substance for a multivariate analysis. Their study showed that the spectral analysis gave better results than the tumour markers. Common to all the studies above is the relatively low number of samples. A larger number of samples are required to validate the results.

Porphyrins in Acetone Extracts

The studies above all use raw or diluted plasma or serum, but other approaches have been made to use autofluorescence of blood in a diagnostic context. Instead of using whole (or diluted) serum or plasma for analysis, several publications suggest an acetone extract of blood for the fluorescence measurements [44;61;64-66]. The acetone extracts is from either plasma, serum or the solid cellular elements left when the plasma or serum is separated from the blood. These studies are all about measuring porphyrin levels in the blood samples. The background for this approach is some earlier studies which have shown elevated porphyrin levels in cancer patients [85].

Porphyrins are present in blood only in small concentrations, and the area of porphyrin emission is dominated by other fluorophores in higher concentrations. Hence, it can be hard to detect porphyrins directly in the raw serum or plasma by

fluorescence. Using an acetone extraction of blood leads to removal of many of the interfering fluorophores, and porphyrins can then easily be measured by fluorescence. Studies on acetone extract of blood serum or plasma have all found that porphyrin emission around 630 nm in cancer patients is elevated compared to healthy patients [66]. The reason for the elevated levels of porphyrin in the cancer patients is however not totally clear. One theory is that the elevated levels found in cancer is due to an over expression of porphyrins in the cancer cells. In a study made by Aiken and Hue (1994) on pure sera (no acetone extract), they had similar observations of elevated levels of porphyrin emission, though they found no difference in the levels of total porphyrin measured by HPLC in the same samples [2]. The latter suggests that the reason is rather a change in the relative protein composition in cancer patients. The effect of that is a decrease in the amount of protein-bound bilirubin in sera and consequently a loss in the intensity of the background emission signal. Another explanation could be a change in serum lipoprotein composition [2]. Despite the rather positive results of the acetone extract method, this method has not yet been clinically accepted.

Table 1: Overview of some important publications in fluorescence spectroscopy on human blood and cancer.

Work	Blood source	Application/Aim of study	N samples	Discrimination method/conclusion
Wolfbeis and Leiner (1985)	Human serum	To map and assign total amount of fluorophores in human serum		A list of expected important fluorophores found in blood serum.
Leiner et al. (1983)	Rat serum Diluted 500 times in PBS	Yoshida ascites hepatoma-bearing rats in different stages (fluorescence measured in the UV area)	8 cancer 8 controls	Decrease in tryptophan fluorescence in tumour bearing animals.
Leiner et al. (1983)	Rat serum Diluted 20 times in PBS	Yoshida ascites hepatoma-bearing rats in different stages (fluorescence measured in the Vis area)	8 cancer 8 controls	Clustering method based on two wavelengths (trp and NAD(P)H
Leiner et al. (1986)	Human Serum Diluted 20 and 500 times in PBS	gynecological tumours	31 cancer 19 controls	Fluorescence intensity of 287/365 expressed as percent of 287/337
Hubmann et al. (1990)	"Synthetic" and human sera diluted to $Abs_{280} < 0.1$ A	To study the reasons for the observed differences found by Leiner et al.	30 human serum samples	
Madhuri et al. (1997)	Human Plasma diluted to O.D 0.5	Oral cancer	3 controls 13 cancer (3 in stage I, 4 in stage II and 3 in stage III and IV respectively)	Ratio between intensity of trp and NAD(P)H

Work	Blood source	Application	N samples	Discrimination method
Aiken and Hue (1994)	Synthetic, bovine and human sera	Different types of cancer	26 control 120 cancer	Difference in porphyrin fluorescence 505/620 and 505/660
Madhuri et al. (1999)	Acetone extract of human plasma	Liver disease	13 controls 19 liver patients	Discriminant analysis using the ratios between emission 465/620 and emission 465/520 after excitation at 405 nm
Madhuri et al. (2003)	Acetone extract of human plasma	Oral cancer	36 healthy controls 20 cancer in stage I and II 23 cancer in stage III and IV	Discriminant analysis using different ratios of porphyrin emission.
Masilamani et al. (2004)	Acetone extract of cellular elements of blood	Various cancer types	25 healthy control 70 non cancer patients 5 benign tumour patients 77 cancer	Relative intensity between 400/590 and 400/630
Lualdi et al. (2007)	Acetone extract of human plasma	Colorectal cancer	169 healthy controls 172 patients	Difference in intensity at 405/623
Nørgaard et al. (2007)	Human Serum Undiluted, 20 and 500 times diluted	Breast cancer	13 controls 11 solitary metastases 15 multiple metastases	ECVA on unfolded EEM (best result with ECVA on emission spectra from ex 230 nm
Kalaivani et al. (2008)	Acetone extract of cellular elements of blood	Breast cancer	35 controls 28 early stage cancer 18 advanced stage cancer	LDA on six ratios between porphyrin and flavin peaks

Chapter 7: Examples of Fluorescence Spectroscopy and Multiway Analysis Applied in Detection of Cancer

This chapter will discuss the multivariate and multiway advantages of applying fluorescence spectroscopy in a diagnostic context. Some examples are shown where fluorescence spectroscopy and multivariate and multiway methods have been used in an attempt to detect cancer from human blood samples.

The Multivariate Approach

In the literature examples discussed in the previous chapter, only the paper from Nørgaard et al. (2008) [77] uses a multivariate approach to analyse the spectra. The others use either intensities at single wavelength pairs, or ratios between wavelengths. The conclusion from the studies of Leiner and co-workers was that their method/approach had potential for pattern recognition. The operational output from their study was a simple ratio between two data points in the EEM consisting of, in their case, 300 data points, thus they did most likely not fully exploit the possibilities in their data. The reason is that they did not possess the necessary tools to do this. From a metabonomic point of view this exclusive focus on single wavelength pairs can be a limiting factor compared to the approach where the whole spectral area is measured. The latter gives better options for getting a broader insight to changes in the metabolic state. The same criticism can be applied to the acetone extract approach where the method is related only on porphyrins.

In the Nørgaard et al. study [77], the full EEMs were measured, and in order to fully exploit the potential of the full data matrix they applied chemometrics on the unfolded EEMs. The result was a good classification between breast cancer patients and controls by using multivariate classification techniques on the total

spectral area, and perfect classification using only emission spectra following excitation at 230 nm. By using chemometrics on the full data matrix instead of only a limited number of data points, Nørgaard et al. were able to, not only, classify cancer from healthy, but in addition they found interesting subgroups among the diseased that could give further information on the patients. Hereby they utilized the collected data better.

Multivariate to Multiway! PARAFAC Opens for the Metabonomic Approach

The advantages in analysing EEMs with PARAFAC are described in chapter 3. By applying PARAFAC on the whole spectral area, the rather complex fluorescence EEMs can be reduced to scores and loadings, where the loadings are (provided a good PARAFAC model) good estimates of the fluorophores in the sample, and the score matrix contains estimates of the concentrations of the fluorophores. The result of the PARAFAC analysis can be seen as a chemical profile of the samples along with a concentration profile of the chemical compounds, thus we have a qualitative and quantitative analysis of a number of metabolites. The solution of the PARAFAC analysis of the fluorescence EEMs is thus in concurrence with the definition of metabonomics, and this opens for fluorescence spectroscopy as a metabonomic tool.

Example 1 - Colorectal Cancer (Paper III)

A larger experiment with autofluorescence on blood samples (citrate plasma) from a study on colorectal cancer was performed as part of this PhD project. The scope of this study was to apply PARAFAC on EEMs in an attempt to extract relevant background information for a potential difference between cancer and controls. This could then be an example of a fluorescence based metabonomic study. A larger set of samples (304 subjects, compared to Leiner et al. 50 samples and Nørgaard et al. 39 samples) was used in the study.

The samples in this study are a part of a larger sample set from a multi-centre cross sectional study conducted at six Danish hospitals of patients undergoing large bowel endoscopy due to symptoms associated with colorectal cancer [60;74]. The present sample set is designed as a case control study with one case group and three different control groups. These three control groups are divided into 1) healthy subjects with no findings at endoscopy, 2) subjects with other, non malignant findings and 3) subjects with pathologically verified benign adenomas. Sample handling, data acquisition and data treatments are described in paper III.

Evaluating the mean spectra of the different groups confirms some of the expected findings based on the literature. Figure 16 show mean spectra from two selected spectral areas. Tryptophan emission (290/350) in the cancer group is slightly blue shifted compared to the other groups (Figure 16 upper plot). The emission spectra following excitation at 345 nm is deviating in the cancer group (figure 16 lower plot). The lower spectra could be assigned to NAD(P)H. Generally, the cancer group is deviating in a great part of the measured spectral area when only looking at the mean spectra.

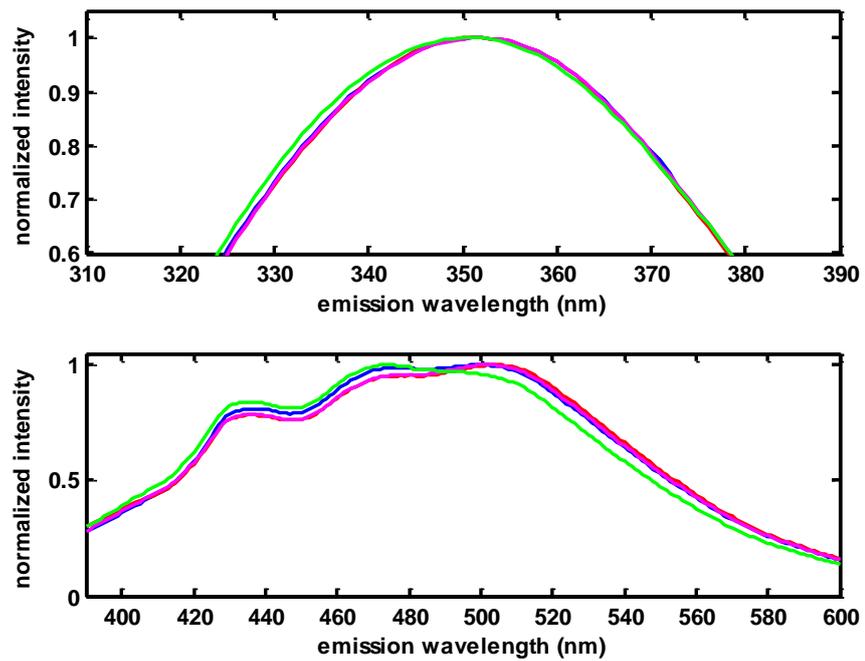


Figure 16: Mean emission spectra of the four groups; blue: no findings, red: other nonmalignant findings, magenta: adenomas, green: cancer. Upper plot; emission spectra after ex 290 nm, lower plot; emission after ex 345 nm.

The results are positive and confirm the findings from previous works. Unfortunately it is not that simple when we look at individual persons rather than the population estimates. There is a huge biological variation in the samples, and hereof following a large standard deviation on the mean values. This is illustrated in figure 17. The figure shows the mean intensity at 290/337 (resembles the 287/337 wavelength used by Leiner et al.); the cancer patients clearly has a higher mean value than the other groups, but the standard deviations in all groups are large and overlapping. A classification built exclusively on this single intensity would thus not be good. Similar exercises can be performed on the ratio between 287/337 and 287/365, which was suggested by Leiner and co-workers. Again a large variation within the groups gives heavily overlapping standard deviations, which makes classification troublesome (not shown).

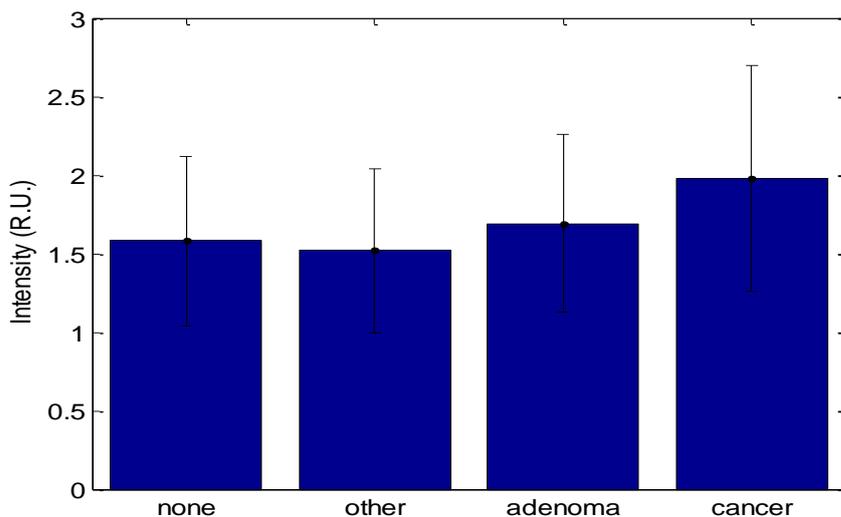


Figure 17: Mean intensity at 290/337 nm in the different groups, with the standard deviation marked as an error bar.

Multivariate/Multiway approach

Instead of using single wavelength pairs in an attempt to discriminate between the groups, we applied multivariate and multiway data analysis methods on the whole measured spectral area.

Emission and excitation loadings from PARAFAC models on the three measured setups (undiluted in two spectral areas and diluted; see chapter 1 or paper III for more information) are seen in the figure below. All together 19 components were extracted, though some of the components in the models on the diluted and the undiluted samples must reflect the same compounds, so the actual number of chemical compounds reflected is less than 19. The pooled score matrices from the three models with all 19 components, holding the relative concentrations of the compounds, were used in a discriminant analysis. The result was classification models with sensitivity and specificity values around 0.75 and with area under the ROC curve of 0.7-0.8 for discrimination between cancer and one or all of the control groups. A high specificity value is often preferred in diagnostic tests (see chapter 3), and by changing the threshold it is possible to get models with specificity value of 0.9, and a sensitivity of 0.5. (see figure 19). These results are at level with the performance of

known immunochemical tests used to classify colorectal cancer (e.g. carcinoembryonic antigen (CEA)) see chapter 5. This result confirms the result from Nørgaard that a simple fluorescence measurement can perform as good as biomarker tests.

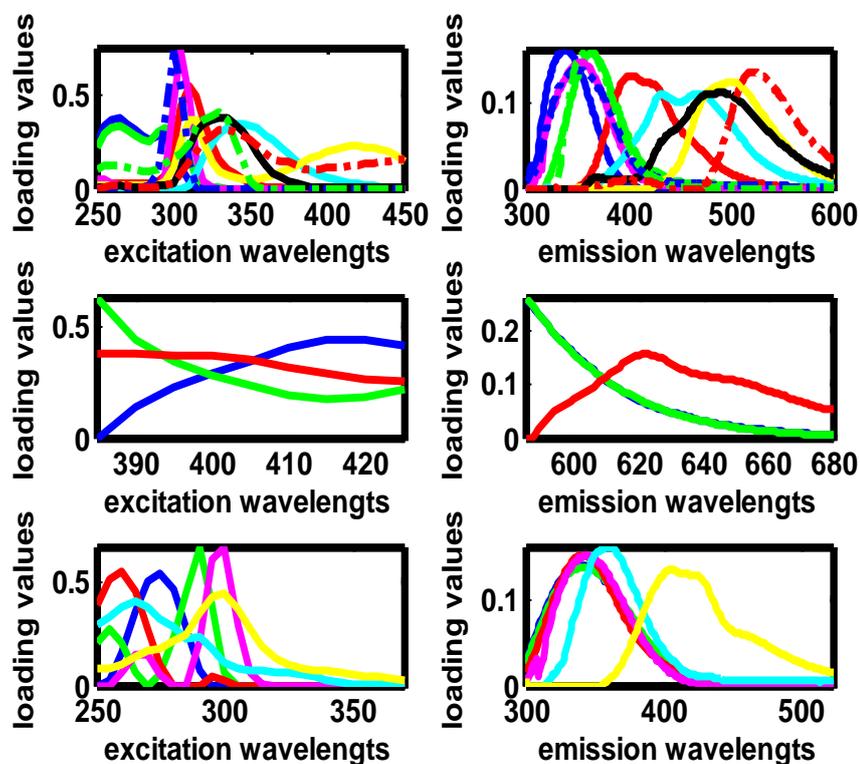


Figure 18: PARAFAC excitation and emission loadings from models on the three measuring setups on the plasma samples. Upper: Undiluted large spectral area, Middle: Undiluted Selected spectral area for Porphyrins, Lower: $100 \times$ diluted large spectral area. Figure adapted from paper III

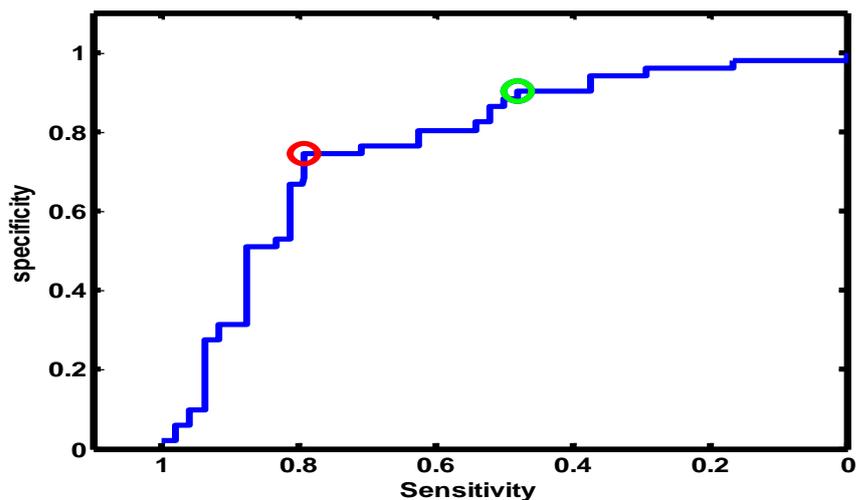


Figure 19: ROC curve from PLS-DA model based on all PARAFAC loadings, for discrimination between cancer patients and control samples from the group “other non malignant findings”. Red circle marks the “optimal” sensitivity and specificity value, green circle marks specificity and sensitivity for specificity at 0.9.

Further analysis of the results revealed that some of the PARAFAC components were more correlated to the cancer patients and some more to the controls. In both groups there were compounds which could be tentatively assigned to tryptophan, but those correlating more to the cancer patients were blue shifted compared to those correlating more to the controls (see figure 20). These results are in concurrence with the findings from Leiner, Wolfbeis and co-workers, who also reported a blue shift in tryptophan in cancer samples [56-58]. This result indicates that beside the classification, the PARAFAC based solution provides the researcher with additional information that can give an understanding of the mechanisms behind the classification in concurrence with the ideas of metabonomics.

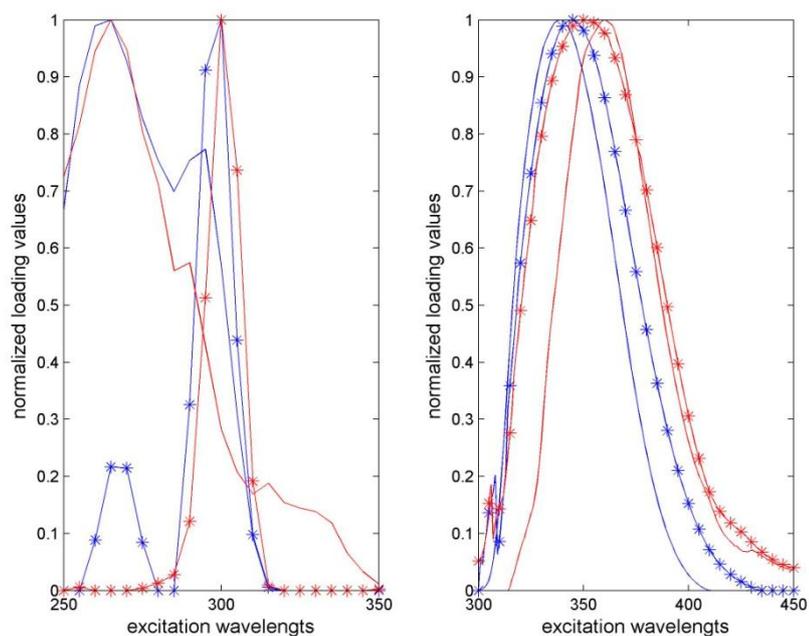


Figure 20: Four PARAFAC excitation and emission loadings: blue and blue with asterisk are correlated with cancer, red and red with asterisk are correlated with control samples. The loadings correlated with cancer are blue shifted in the emission compared to the loadings correlated to the control samples. Figure adapted from paper III

Besides using PARAFAC on the EEMs, PLS-DA was also applied directly on the unfolded spectra. This was the approach taken by Nørgaard et al. Discriminations achieved by PLS-DA was in the same level as those obtained by use of the PARAFAC scores (see paper III). In mere classification there is thus not much to gain by applying PARAFAC, but for a chemical analysis of the result, the PARAFAC loadings are much easier to interpret than the PLS-DA loadings.

Some of the PARAFAC models in the above example gave loadings that were not easy to assign to specific chemical compounds. In some cases, the shape of loadings indicates that they are not only reflecting one single chemical compound, but rather a number of compounds or maybe the interaction between several compounds. The consequence is that the qualitative aspect of the PARAFAC model is limited. The PARAFAC loadings will still reflect variations in the fluorescence spectra, and the peak

positions in the loadings will still have a chemical connection, but it is no longer possible to associate a PARAFAC component to only one single fluorophore. The scores no longer reflect only one concentration, and a change in the loading/score value cannot be traced back to the change in only one compound. From an analytical point of view this is not optimal, as it limits the possibilities to draw solid chemical conclusions from the analysis. If fluorescence spectroscopy should be used as a metabonomic tool it is necessary to solve this problem. Blood plasma is a very complex matrix with potential interference from various compounds and a huge biological variation. In order to make more optimal PARAFAC models, it is possible that many more samples are required. More samples will cover more of the biological variation found in the samples and there are thus better options for separating some of the confounding loadings. Describing more of the biological variability with more samples, more PARAFAC components are also expected and thus maybe a better chance for finding chemical compounds relevant for cancer separation. Another option that might provide better PARAFAC models is to ensure optimal measuring conditions; if for example the spectra suffer from inner filter effect, finding the optimal dilution could potentially help.

Example 2 - PARAFAC on Breast Cancer Data

In another example, PARAFAC has been applied on some of the breast cancer data from the paper of Nørgaard et al. The data used are from serum diluted 20 times and measured in the UV area. The absorbance is thus expected to be within the area of linear dependence between concentration and fluorescence, hence maybe better options for PARAFAC modelling. A five component PARAFAC model was fitted to a selected spectral area. This area was chosen because it displays some significant sub-grouping among the cancer patients with progressive cancer (see reference [77] for details)

This PARAFAC model gave a clear clustering of the cancer samples and the controls in the first vs. the fourth component (highlighted green and blue in the loading plot below). From the spectra it was expected to see discrimination in the fifth component (dashed purple) which is reflecting the peak found in the patients with progressive cancer.

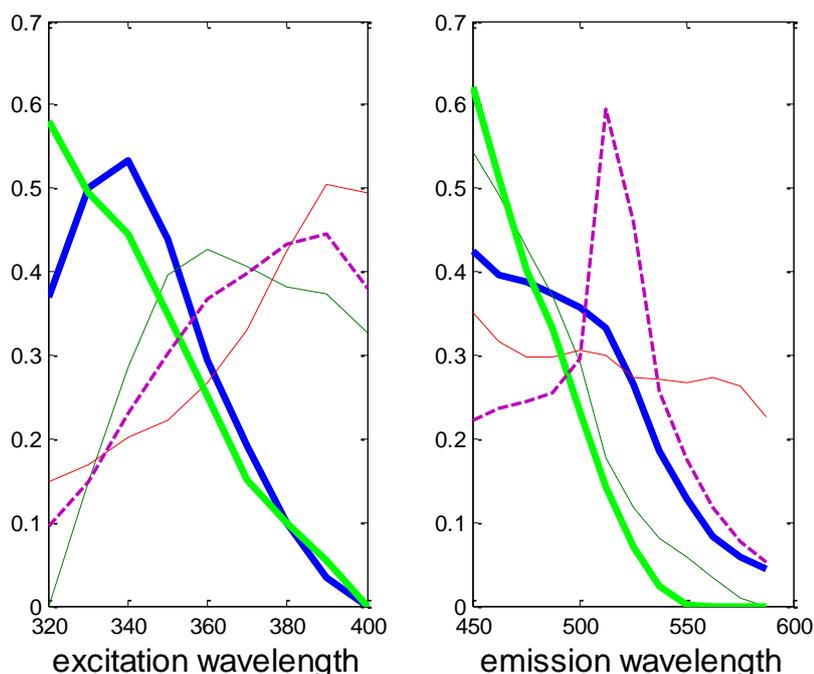


Figure 21: Excitation and emission loadings from a five component PARAFAC model on a spectral selection of the breast cancer data. The highlighted blue and green loadings (component one and four respectively) are used for separation of cancer and non-cancer, see below. The dashed purple (component five) is the loading separating the metastasis patients.

Component one (blue) which has excitation maximum at 340 nm and emission maximum just outside the modelled area can tentatively be assigned to NAD(P)H, whereas component four (green) has both excitation and emission maximum outside the modelled area, and assignment is not possible. Expanding the spectral area to cover the area where these two compounds expectably have emission maximum did not give good PARAFAC models, and did not improve the classification.

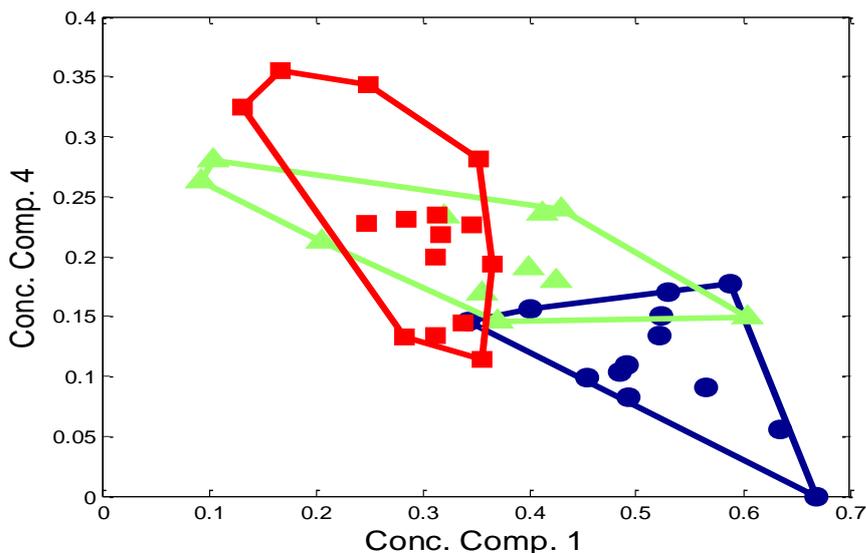


Figure 22: PARAFAC scores of component one vs. component four. The coloured classes are; Blue circles: healthy controls; green triangles: cancer patients with solitary metastases; red squares: cancer patients with multiple metastases.

The PARAFAC score plot in figure 22 shows the clustering of the samples. No separation is experienced between solitary and multiple metastasis, whereas there is good, though not perfect, separation between cancer and controls. One sample from the group of solitary metastasis is located far from the other cancer samples, and within the group of controls. From a modelling point of view there is no outlying behaviour of that sample, and removing it is hence not feasible on that account. Still, it seems plausible that the sample may be misclassified in the original sample set.

The PARAFAC model is based on a relatively small number of samples, and including more samples in the model would give a better idea whether this sample is off, or if it is within the normal variation in patients with solitary metastases.

In a direct classification based on ECVA on the unfolded spectra Nørgaard et al. achieved good results with few classification errors dependent on which spectral areas and dilutions they used. In discriminating controls from cancer (solitary and multiple in one common group), a total classification error of 10 out of 38 was

obtained, whereas when discriminating controls from solitary metastases, as a sort of early detection, they only had two errors out of 24 samples. In this study, PLS-DA models were calculated based on the scores from the PARAFAC model. Classifications were performed both on the total score matrix, and on a score matrix of the scores from component one and four only. The results are seen in table 2. Both in terms of sensitivity, specificity and area under the ROC curve the results are promising. The classification error in discriminating the cancer groups from controls, pooled or individual is at level with the results achieved from Nørgaard et al.

Table 2: Results from PLS-DA models for classification of breast cancer samples from control samples based on PARAFAC scores.

<i>Model</i>	<i>Groups</i>	<i>Sens</i>	<i>Spec</i>	<i>AUC</i>	<i>Classification error</i>	
					<i>False positive</i>	<i>False negative</i>
Full score matrix	Control vs. cancer (pooled)	0.96	0.85	0.86	2	1
	Control vs. Solitary	0.91	0.77	0.79	3	1
	Control vs. Multiple	1	0.92	0.91	1	0
Score vectors 1+4	Control vs. cancer (pooled)	0.92	0.92	0.90	1	2
	Control vs solitary	0.91	0.85	0.91	2	1
	Control vs. Multiple	0.93	0.92	0.88	1	1

There is not much difference between classifying on the whole score matrix or on the matrix of only the two discriminating components. Most of the discriminating information is thus found in those two components. Comparing the results with Nørgaard et al. there is a slightly better classification error for discrimination between cancer and controls (only cross validated results are compared), whereas in discrimination between solitary and controls the result is the same or slightly worse. The latter is possible due to the one deviating sample.

As in the previous example with colorectal cancer patients, PARAFAC analysis of the fluorescence EEMs compared to “standard” chemometric analysis of the unfolded spectra did not give improved classification. The decomposition of the complex spectra into PARAFAC components allows the analyst to focus on components one and four when searching for a biochemical understanding of the mechanisms behind the discrimination. Looking at the mean concentrations for these two components (figure 23) it is clear that there is a difference between the concentrations in controls and cancer patients. The standard deviations are rather high, which must be expected in biological samples. Especially for component one in the solitary samples the standard deviation is high, the deviating sample in the solitary sample has a high score value for this component, and can possibly explain some of this deviation. Component one, which probably can be assigned to NAD(P)H, seems to have higher concentration in the control samples than in the cancer samples, and component four (unassigned) has higher concentration in the cancer samples. These findings are in conflict with a recently published work by Yu and Heikal (2009) who reported elevated concentrations of NADH in breast cancer cells compared to healthy breast cells [106]. The findings are of course in different media, blood serum versus living breast cells, but it would be interesting to look further into. The number of samples applied in this study (39) is too small to draw solid conclusions. A larger study would allow a more thorough analysis of the mechanisms behind the discrimination, with a more accurate spectral assignment.

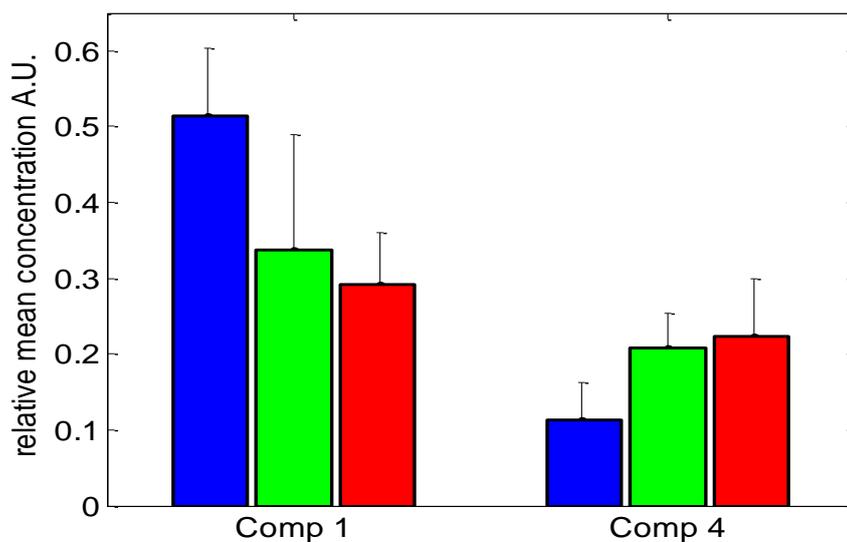


Figure 23: Relative mean concentrations of PARAFAC component one and component four in the different groups of individuals. Blue: control, green: solitary metastasis, red: multiple metastasis. Standard deviations are shown in the error bars.

Porphyrins in Plasma – and PARAFAC

Many of the applications with autofluorescence on blood found in the literature applied the method of acetone extract of blood serum or plasma. The target of those studies was to measure porphyrins, and the acetone extract removes some of the interference when measuring porphyrin emission. In the present colorectal cancer study, the area of porphyrin excitation and emission was also included in the measurements. The measurements were applied directly on the undiluted plasma, no acetone extract was applied. The reason for not using the acetone extract was for one thing the laborious and hazardous work connected to the extraction. The other reason was that by applying PARAFAC it should be possible to extract both estimates of the porphyrin emission spectrum along with estimates of the interfering background emission. The measured spectrum of this area was dominated by “background” signal with only a small “bump” indicating porphyrin emission. A PARAFAC model fitted to this specific area with three components gave two components of the interfering background but also one component which could be assigned to porphyrin

with maximum at 400/620 nm (see figure 17 middle plots). However, opposed to the acetone extract application described, the level of porphyrins found in the cancer patients was not elevated compared to the controls, and hence no classification of colorectal cancer vs. non cancer could be based exclusively on the porphyrin levels. This was further verified in a formal multivariate classification model based on that particular spectral range. Elevated porphyrin levels has previously been found in colorectal cancer [62;85], and it would be interesting to go further into why there was not found any differences in the porphyrin levels in this study. A direct comparison between porphyrin found by PARAFAC and porphyrin found by the acetone extract method could be interesting. An expansion of the measured spectral area could also be interesting as it would give better options for modelling the background signal. Aiken and Hue (1994) suggested that the difference between cancer and controls was a change in the interfering background emission.

Chapter 8: Conclusion

The examples in the previous chapter have shown that fluorescence spectroscopy on human blood samples, (serum or plasma) has a clear potential when it comes to discriminating between cancer and control samples. Previous results from for example Leiner and co-workers and Nørgaard et al have thus been confirmed. In the present studies PARAFAC was further applied to decompose the fluorescence data into estimates of the excitation and emission spectra of the underlying chemical compounds. It was shown how PARAFAC applied on “clinical” fluorescence data might provide better options for understanding some of the biochemical changes behind the discrimination between cancer and control samples. At the same time it was shown, that classification models based on the raw unfolded spectra performed equally well as a model based on PARAFAC scores, in terms of sensitivity and specificity values. If the purpose is only classification a two-way method like ECVA or PLS-DA might be a better choice as these methods can be more operational in terms of for example variable selection. If the purpose is to gain more understanding of the chemical mechanisms behind the classification, and thus use fluorescence spectroscopy as a metabonomic tool, PARAFAC must be applied on the spectra. To obtain conditions suitable for a chemical interpretation of the classification it requires of course chemically meaningful loadings. The PARAFAC approach thus requires optimal conditions for measuring fluorescence with limited concentration quenching and suitable concentrations of the fluorophore. In the colorectal example shown in this thesis, some of the PARAFAC loadings are ambiguous which might be due to non-optimal measuring conditions. Interpretation of the important PARAFAC components in the classification indicated a blue shift in tryptophan emission in the cancer samples, confirming findings

from Leiner and co-workers, but this effect alone could not explain the classification. In the example with breast cancer data, two PARAFAC components in a five component model are important for the classification. The measuring conditions for these samples are assumably better, and one component can be most likely be assigned to NAD(P)H. This example thus left the analyst with the knowledge that the NAD(P)H concentration has impact on cancer classification.

The above results show the potential for fluorescence spectroscopy combined with PARAFAC as a metabonomic tool, though they also reveal that much work yet has to be done. Bigger or additional sample sets should be measured, and work should be done in finding optimal measuring conditions for doing perfect PARAFAC models on blood plasma/serum which is a very complex biological medium.

The conclusions based on the results found here in these and in previous studies is that fluorescence spectroscopy seems to be a potential method within cancer diagnostics, monitoring and maybe screening. Compared to traditional methods used for detection, where some requires very expensive equipment (MR, CT, X-ray), fluorescence spectroscopy is a cheap method, though at present state not a direct alternative. The sensitivity and specific values found by fluorescence spectroscopy are comparable to any known biomarker used for colorectal cancer. An operational comparisons to biomarker analysis, will reveal that the onetime investments in fluorescence spectroscopy is higher, but the running costs of the ELISA kits detecting the biomarkers are high, and at the same time fluorescence spectroscopy is faster and much easier to use. There is a potential to use fluorescence spectroscopy as a screening method, either as a standalone method, or combined with other methods. Compared to the FOBT method mentioned in a previous chapter, fluorescence spectroscopy on a blood sample might have a better chance of satisfying compliance as it avoids unpleasantness of handling stool.

The basic concept of metabonomics is to measure all relevant metabolites affected by for example pathophysiological stimuli. The different analytical methods typically applied in metabonomics can measure different metabolites, but no methods are capable of measuring the total amount of metabolites. The individual methods have so to say an “analytical filter” which limits the number of molecules that can be detected. Compared to traditional metabonomic tools such as LC-MS and NMR, the filter of fluorescence spectroscopy might allow fewer molecules to be measured, but the sensitivity and specificity is very high. Operational fluorescence spectroscopy is cheaper and much simpler than MS and NMR. As a metabonomic tool, fluorescence spectroscopy can find some unique results, which potentially can give new insight to the metabonomic state. For a total measure of all metabolites, the different methods complement each other, and even more insight to the metabonomic state can be achieved.

To use a fluorescence based method globally it requires quality insurance of data. One aspect of this is the need for proper spectral calibration of fluorescence instruments. The test of the commercial solution from BAM applied on three instruments illustrated the need for a calibration. More work needs to be done before an easy to use solution is easy to use for the average spectroscopy user.

Reference List

- [1] Essential Medical Imaging, Cambridge University press, New York, 2009.
- [2] J.H. Aiken, C.W. Huie, J.A. Terzian, Characteristic Visible Fluorescence Emission-Spectra of Sera from Cancer-Patients. *Analytical Letters* 27 (1994); 511-21.
- [3] D.G. Altman, J.M. Bland, Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *BMJ* 308 (1994); 1552.
- [4] C.M. Andersen, R. Bro, Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J Chemometrics* 17 (2003); 200-15.
- [5] C.N. Arnold, A. Goel, H.E. Blum, C.R. Boland, Molecular pathogenesis of colorectal cancer - Implications for molecular diagnosis. *Cancer* 104 (2005); 2035-47.
- [6] M. Barker, W. Rayens, Partial least squares for discrimination. *J Chemometrics* 17 (2003); 166-73.
- [7] J.S. Bartlett, K.J. Voss, S. Sathyendranath, A. Vodacek, Raman scattering by pure water and seawater. *Applied Optics* 37 (1998); 3324-32.
- [8] R. Bro, PARAFAC. Tutorial and applications. *Chemometr Intell Lab Syst* 38 (1997); 149-71.
- [9] J.A. Burch, K. Soares-Weiser, D.J.B. St John, S. Duffy, S. Smith, J. Kleijnen, et al., Diagnostic accuracy of faecal occult blood tests used in screening for colorectal cancer: a systematic review. *J Med Screen* 14 (2007); 132-7.
- [10] J.D. Carroll, J.J. Chang, Analysis of Individual Differences in Multidimensional Scaling Via An N-Way Generalization of Eckart-Young Decomposition. *Psychometrika* 35 (1970); 283-&.
- [11] R.B. Cattell, "Parallel Proportional Profiles" And Other Principles for determining the choice of factors by rotation. *Psychometrika* 9 (1944); 267-83.
- [12] J. Christensen, L. Norgaard, R. Bro, S.B. Engelsen, Multivariate autofluorescence of intact food systems. *Chemical Reviews* 106 (2006); 1979-94.
- [13] P.G. Coble, C.A. Schultz, K. Mopper, Fluorescence contouring analysis of DOC intercalibration experiment samples: a comparison of techniques. *Marine Chemistry* 41 (1993); 173-8.

- [14] P.C. DeRose, E.A. Early, G.W. Kramer, Qualification of a fluorescence spectrometer for measuring true fluorescence spectra. *Rev Sci Instrum* 78 (2007); 033107.
- [15] P.C. DeRose, U. Resch-Genger, Recommendations for Fluorescence Instrument Qualification: The New ASTM Standard Guide. *Analytical Chemistry* 82 (2010); 2129-33.
- [16] P.C. DeRose, M.V. Smith, K.D. Mielenz, D.H. Blackburn, G.W. Kramer, Characterization of Standard Reference Material 2941, uranyl-ion-doped glass, spectral correction standard for fluorescence. *Journal of Luminescence* 128 (2008); 257-66.
- [17] P.C. DeRose, M.V. Smith, K.D. Mielenz, D.H. Blackburn, G.W. Kramer, Characterization of Standard Reference Material 2940, Mn-ion-doped glass, spectral correction standard for fluorescence. *Journal of Luminescence* 129 (2009); 349-55.
- [18] S. Determann, R. Reuter, P. Wagner, R. Willkomm, Fluorescent Matter in the Early Atlantic-Ocean .1. Method of Measurement and Near-Surface Distribution. *Deep-Sea Research Part I-Oceanographic Research Papers* 41 (1994); 659-75.
- [19] M.J. Duffy, A. van Dalen, C. Haglund, L. Hansson, E. Holinski-Feder, R. Klapdor, et al., Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *European Journal of Cancer* 43 (2007); 1348-60.
- [20] J. Ferlay, D.M. Parkin, E. Steliarova-Foucher, Estimates of cancer incidence and mortality in Europe in 2008. *European Journal of Cancer* 46 (2010); 765-81.
- [21] O. Fiehn, Metabolomics - the link between genotypes and phenotypes. *Plant Molecular Biology* 48 (2002); 155-71.
- [22] R.A. Fisher, The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7 (1936); 179-88.
- [23] E. Flamini, L. Mercatali, O. Nanni, D. Calistri, R. Nunziatini, W. Zoli, et al., Free DNA and Carcinoembryonic Antigen Serum Levels: An Important Combination for Diagnosis of Colorectal Cancer. *Clinical Cancer Research* 12 (2006); 6985-8.
- [24] J.A. Gardecki, M. Maroncelli, Set of secondary emission standards for calibration of the spectral responsivity in emission spectroscopy. *Applied Spectroscopy* 52 (1998); 1179-89.
- [25] C.A. Gatsonis, Receiver Operating Characteristic Analysis for the Evaluation of Diagnosis and Prediction. *Radiology* 253 (2009); 593-6.

- [26] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial. *Anal Chim Acta* 185 (1986); 1-17.
- [27] M.H. Hamdan, *Cancer Biomarkers*, John Wiley and sons, New Jersey, 2007.
- [28] H.H. Harman, *Modern Factor Analysis*, University of Chicago press, Chicago, 1976.
- [29] C.W. Harris, H.F. Kaiser, Oblique Factor Analytic Solutions by Orthogonal Transformations. *Psychometrika* 29 (1964); 347-62.
- [30] R.A. Harshman, Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics* 16 (1970); 84.
- [31] T. Hastie, R. Tibshirani, J. Friedman, *Linear Methods for Classifications. Elements of Statistical Learning, data mining, inference and prediction*. Springer, 2009.
- [32] P. Hewitson, P. Glasziou, L. Irwig, B. Towler, E. Watson, Screening for colorectal cancer using the faecal occult blood test, Hemocult. *Cochrane Database of Systematic Reviews* (2007).
- [33] P. Hewitson, P. Glasziou, E. Watson, B. Towler, L. Irwig, Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (Hemocult): An update. *American Journal of Gastroenterology* 103 (2008); 1541-9.
- [34] L. Hol, M.E. van Leerdam, B.M. van, A.J. van Vuuren, D.H. van, J.C. Reijerink, et al., Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. *Gut* 59 (2010); 62-8.
- [35] J. Hollandt, R.D. Taubert, J. Seidel, U. Resch-Genger, A. Gugg-Helminger, D. Pfeifer, et al., Traceability in fluorometry - Part I: Physical standards. *Journal of Fluorescence* 15 (2005); 301-13.
- [36] J. Holm, C. Schou, L.N. Babol, A.J. Lawaetz, S.W. Bruun, M.Z. Hansen, S.I. Hansen, The interrelationship between ligand binding and self-association of the folate binding protein. The Role of detergent-tryptophan interaction. *Biochimica et Biophysica Acta, General Subjects* 1810 (2011); 1330-1339
- [37] U. Holzgrabe, R. Deubner, C. Schollmayer, B. Waibel, Quantitative NMR spectroscopy - Applications in drug analysis. *Journal of Pharmaceutical and Biomedical Analysis* 38 (2005); 806-12.
- [38] H. Hotelling, Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24 (1933); 417-41.

- [39] H. Hotelling, The most predictable criterion. *Journal of Educational Psychology* 26 (1935); 139-42.
- [40] H. Hotelling, Relations between two sets of variates. *Biometrika* 28 (1936); 321-77.
- [41] M.R. Hubmann, M.J.P. Leiner, R.J. Schaur, Ultraviolet Fluorescence of Human Sera .1. Sources of Characteristic Differences in the Ultraviolet Fluorescence-Spectra of Sera from Normal and Cancer-Bearing Humans. *Clin Chem* 36 (1990); 1880-3.
- [42] F. Jenkinson, R.J.C. Steele, Colorectal cancer screening - Methodology. *Surgeon-Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* 8 (2010); 164-71.
- [43] H.F. Kaiser, The Varimax Criterion for Analytic Rotation in Factor-Analysis 2. *Psychometrika* 23 (1958); 187-200.
- [44] R. Kalaivani, V. Masilamani, K. Sivaji, M. Elangovan, V. Selvaraj, S.G. Balamurugan, et al., Fluorescence spectra of blood components for breast cancer diagnosis. *Photomedicine and Laser Surgery* 26 (2008); 251-6.
- [45] T. Karstens, K. Kobs, Rhodamine-B and Rhodamine-101 As Reference Substances for Fluorescence Quantum Yield Measurements. *Journal of Physical Chemistry* 84 (1980); 1871-2.
- [46] M. Kasha, Characterization of Electronic Transitions in Complex Molecules. *Discussions of the Faraday Society* (1950); 14-9.
- [47] H.A.L. Kiers, A comparison of techniques for finding components with simple structure. in: C.M.Cuadras, C.R.Rao, editors. *Multivariate analysis: future directions 2* (European Materials Research Society Symposia Proceedings). *Elsvier Science Publishers*, 1993.
- [48] H.A.L. Kiers, J.M.F. Tenberge, The Harris-Kaiser Independent Cluster Rotation As A Method for Rotation to Simple Component Weights. *Psychometrika* 59 (1994); 81-90.
- [49] K.W. Kinzler, B. Vogelstein, Introduction. in: K.W. Kinzler, K.J. Voss, editors. *The Genetic basis of Cancer*. *McRgaw-Hill Companies*, 2002.
- [50] K. Kjeldahl, R. Bro, Some common misunderstandings in chemometrics. *J Chemometrics* 24 (2010).
- [51] W.S. Klug, M.R. Cummings, The genetic basis of Cancer. in: W.S. Klug, M.R. Cummings, editors. *Essentials of Genetics*. 2005.
- [52] J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer, 2006.

- [53] A.J. Lawaetz, B. Schmidt, D. Staerk, J.W. Jaroszewski, R. Bro, Application of rotated PCA models to facilitate interpretation of metabolite profiles: commercial preparations of St. John's Wort. *Planta Med* 75 (2009); 271-9.
- [54] A.J. Lawaetz, C.A. Stedmon, Fluorescence intensity calibration using the Raman scatter peak of water. *Appl Spectrosc* 63 (2009); 936-40.
- [55] J. Lee, V. Reis, S. Liu, L. Conn, E. Groessl, T. Ganiats, et al., Improving Fecal Occult Blood Testing Compliance Using a Mailed Educational Reminder. *Journal of General Internal Medicine* 24 (2009); 1192-7.
- [56] M. Leiner, R.J. Schaur, O.S. Wolfbeis, H.M. Tillian, Fluorescence Topography in Biology .2. Visible Fluorescence Topograms of Rat Sera and Cluster-Analysis of Fluorescence Parameters of Sera of Yoshida Ascites Hepatoma-Bearing Rats. *Ircs Medical Science-Biochemistry* 11 (1983); 841-2.
- [57] M. Leiner, O.S. Wolfbeis, R.J. Schaur, H.M. Tillian, Fluorescence Topography in Biology .1. Ultraviolet Fluorescence Topograms of Rat Sera and Decrease of Tryptophan Fluorescence in Yoshida Ascites Hepatoma-Bearing Rats. *Ircs Medical Science-Biochemistry* 11 (1983); 675-6.
- [58] M.J. Leiner, R.J. Schaur, G. Desoye, O.S. Wolfbeis, Fluorescence topography in biology. III: Characteristic deviations of tryptophan fluorescence in sera of patients with gynecological tumors. *Clin Chem* 32 (1986); 1974-8.
- [59] J. Lindon, E. Holmes, J. Nicholson, *Metabonomics Techniques and Applications to Pharmaceutical Research & Development*. *Pharmaceutical Research* 23 (2006); 1075-88.
- [60] A.F. Lomholt, G. Hoyer-Hansen, H.J. Nielsen, I.J. Christensen, Intact and cleaved forms of the urokinase receptor enhance discrimination of cancer from non-malignant conditions in patients presenting with symptoms related to colorectal cancer. *British Journal of Cancer* 101 (2009); 992-7.
- [61] G.F. Lothian, Beer's law and its use in analysis. A review. *Analyst* 88 (1963); 678-85.
- [62] M. Lualdi, A. Colombo, E. Leo, D. Morelli, A. Vannelli, L. Battaglia, et al., Natural fluorescence spectroscopy of human blood plasma in the diagnosis of colorectal cancer: Feasibility study and preliminary results. *Tumori* 93 (2007); 567-71.
- [63] J.A. Ludwig, J.N. Weinstein, Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 5 (2005); 845-56.

- [64] S. Madhuri, P. Aruna, M.I. Summiya Bibi, V.S. Gowri, D. Koteeswaran, R.J. Schaur, et al., Ultraviolet fluorescence spectroscopy of blood plasma in the discrimination of cancer from normal. *Proc SPIE* 2982 (1997); 41-5.
- [65] S. Madhuri, S. Suchitra, P. Aruna, T.G. Srinivasan, S. Ganesan, Native fluorescence characteristics of blood plasma of normal and liver diseased subjects. *Medical Science Research* 27 (1999); 635-9.
- [66] S. Madhuri, N. Vengadesan, P. Aruna, D. Koteeswaran, P. Venkatesan, S. Ganesan, Native fluorescence spectroscopy of blood plasma in the characterization of oral malignancy. *Photochemistry and Photobiology* 78 (2003); 197-204.
- [67] V. Masilamani, K. Al-Zhrani, M. Al-Salhi, A. Al-Diab, M. Al-Ageily, Cancer diagnosis by autofluorescence of blood components. *Journal of Luminescence* 109 (2004); 143-54.
- [68] W.H. Melhuish, Calibration of Spectrofluorimeters for Measuring Corrected Emission Spectra. *Journal of the Optical Society of America* 52 (1962); 1256-&.
- [69] W.H. Melhuish, Modified Technique for Determining Wavelength-Sensitivity Curve of A Spectrofluorimeter. *Applied Optics* 14 (1975); 26-7.
- [70] C.E. Metz, Basic Principles of Roc Analysis. *Seminars in Nuclear Medicine* 8 (1978); 283-98.
- [71] A. Mishra, M. Verma, Cancer Biomarkers: Are We Ready for the Prime Time? *Cancers* 2 (2010); 190-208.
- [72] R.M. Nakamura, W.W. Grody, General Considerations in the use and application of Laboratory tests for the evaluation of cancer. in: R.M. Nakamura, W.W. Grody, J.T. Wu, R.B. Nagle, editors. *Cancer Diagnostics current and future trends*. Human Press, Totowa, New Jersey 2010.
- [73] J.K. Nicholson, J.C. Lindon, E. Holmes, 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29 (1999); 1181-9.
- [74] H.J. Nielsen, N. Brunner, C. Frederiksen, A.F. Lomholt, D. King, L.N. Jorgensen, et al., Plasma tissue inhibitor of metalloproteinases-1 (TIMP-1): a novel biological marker in the detection of primary colorectal cancer. Protocol outlines of the Danish-Australian endoscopy study group on colorectal cancer detection. *Scand J Gastroenterol* 43 (2008); 242-8.
- [75] A. Nordstrom, R. Lewensohn, Metabolomics: Moving to the Clinic. *Journal of Neuroimmune Pharmacology* 5 (2010); 4-17.

- [76] L. Nørgaard, R. Bro, F. Westad, S.B. Engelsen, A modification of canonical variates analysis to handle highly collinear multivariate data. *J Chemometrics* 20 (2006); 425-35.
- [77] L. Nørgaard, G. Soletormos, N. Harrit, M. Albrechtsen, O. Olsen, D. Nielsen, et al., Fluorescence spectroscopy and chemometrics for classification of breast cancer samples - a feasibility study using extended canonical variates analysis. *J Chemometrics* 21 (2007); 451-8.
- [78] M. Oldiges, S. Lutz, S. Pflug, K. Schroer, N. Stein, C. Wiendahl, Metabolomics: current state and evolving methodologies and tools. *Appl Microbiol Biotechnol* 76 (2007); 495-511.
- [79] I.M. Oving, H.C. Clevers, Molecular causes of colon cancer. *European Journal of Clinical Investigation* 32 (2002); 448-57.
- [80] D. Pfeifer, K. Hoffmann, A. Hoffmann, C. Monte, U. Resch-Genger, The calibration kit spectral fluorescence standards - A simple and certified tool for the standardization of the spectral characteristics of fluorescence instruments. *Journal of Fluorescence* 16 (2006); 581-7.
- [81] D.F. Ransohoff, Developing molecular biomarkers for cancer. *Science* 299 (2003); 1679-80.
- [82] D.F. Ransohoff, How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. *Journal of Clinical Epidemiology* 60 (2007); 1205-19.
- [83] U. Resch-Genger, D. Pfeifer, K. Hoffmann, G. Flachenecker, A. Hoffmann, C. Monte, Linking Fluorometry to Radiometry with Physical and Chemical Transfer Standards: Instrument Characterization and traceable Fluorescence Measurements. in: U. Resch-Genger, editor. *Standardization and Quality Assurance in Fluorescence Measurements I*. vol. 5, Springer, 2008.
- [84] U. Resch-Genger, D. Pfeifer, C. Monte, W. Pilz, A. Hoffmann, M. Spieles, et al., Traceability in fluorometry: Part II. Spectral fluorescence standards. *Journal of Fluorescence* 15 (2005); 315-36.
- [85] E. Rossi, P. Garcia-Webb, Red cell zinc protoporphyrin and protoporphyrin by HPLC with fluorescence detection. *Biomed Chromatogr* 1 (1986); 163-8.
- [86] R.W. Ruddon, *Characteristics of Human Cancer*. Cancer Biology. Oxford University Press, New York 2007.
- [87] R.S. Ryback, M.J. Eckardt, R.R. Rawlings, L.S. Rosenthal, Quadratic Discriminant-Analysis As An Aid to Interpretive Reporting of Clinical Laboratory Tests. *Jama-*

- Journal of the American Medical Association 248 (1982); 2342-5.
- [88] F. Savorani, G. Tomasi, S.B. Engelsen, icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance* 202 (2010); 190-202.
- [89] A. Smilde, R. Bro, P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*, Wiley, Chichester, 2004.
- [90] S. Srivastava, M. Verma, D.E. Henson, Biomarkers for Early Detection of Colon Cancer. *Clinical Cancer Research* 7 (2001); 1118-26.
- [91] C.A. Stedmon, S. Markager, R. Bro, Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry* 82 (2003); 239-54.
- [92] C.L. Stokamer, C.T. Tenner, J. Chaudhuri, E. Vazquez, E.J. Bini, Randomized controlled trial of the impact of intensive patient education on compliance with fecal occult blood testing. *Journal of General Internal Medicine* 20 (2005); 278-82.
- [93] G.M. Strasburg, R.D. Ludescher, *Theory and Applications of Fluorescence Spectroscopy in Food Research*. *Trends in Food Science & Technology* 6 (1995); 69-75.
- [94] E. Tan, N. Gouvas, R.J. Nicholls, P. Ziprin, E. Xynos, P.P. Tekkis, Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surgical Oncology* 18 (2009); 15-24.
- [95] The Danish Cancer Society, <http://www.cancer.dk> 2010
- [96] J. Trygg, E. Holmes, T. Lundstedt, Chemometrics in metabonomics. *Journal of Proteome Research* 6 (2007); 469-79.
- [97] B. Valeur, *Molecular Fluorescence: Principles and Applications*, Wiley-VCH, 2007.
- [98] R.A. Velapoldi, M.S. Epstein, A Fluorescence Standard Reference Material -Quinine Sulfate Dihydrat. *NBS Spec Publ* 260-64 (1980).
- [99] D.S. Wishart, M.J. Lewis, J.A. Morrissey, M.D. Flegel, K. Jeroncic, Y.P. Xiong, et al., The human cerebrospinal fluid metabolome. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 871 (2008); 164-73.
- [100] H. Wold, E.J. Mosbaek, Estimation Vs Operative Use of Parameters in Models for Non-Experimental Situations. *Econometrica* 34 (1966); 46-&.

- [101] S. Wold, K. Esbensen, P. Geladi, Principal Component Analysis. *Chemometr Intell Lab Syst* 2 (1987); 37-52.
- [102] S. Wold, H. Martens, H. Wold, The Multivariate Calibration-Problem in Chemistry Solved by the Pls Method. *Lecture Notes in Mathematics* 973 (1983); 286-93.
- [103] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab Syst* 58 (2001); 109-30.
- [104] O.S. Wolfbeis, M. Leiner, Mapping of the Total Fluorescence of Human-Blood Serum As A New Method for Its Characterization. *Anal Chim Acta* 167 (1985); 203-15.
- [105] J.T. Wu, Types of circulating tumor markers and their clinical applications. in: R.M. Nakamura, W.W. Grody, J.T. Wu, R.B. Nagle, editors. *Cancer Diagnostics current and future trends*. Human Press, Totowa, New Jersey 2010.
- [106] Q.R. Yu, A.A. Heikal, Two-photon autofluorescence dynamics imaging reveals sensitivity of intracellular NADH concentration and conformation to cell physiology at the single-cell level. *Journal of Photochemistry and Photobiology B-Biology* 95 (2009); 46-57.
- [107] Z. Zhao, Y. Xiao, P. Elson, H. Tan, S.J. Plummer, M. Berk, et al., Plasma Lysophosphatidylcholine Levels: Potential Biomarkers for Colorectal Cancer. *J Clin Oncol* 25 (2007); 2696-701.
- [108] M.H. Zweig, G. Campbell, Receiver Operating Characteristic (Roc) Plots - A Fundamental Evaluation Tool in Clinical Medicine (Vol 39, Pg 561, 1993). *Clin Chem* 39 (1993); 1589.
- [109] M.H. Zweig, G. Campbell, Receiver-Operating Characteristic (Roc) Plots - A Fundamental Evaluation Tool in Clinical Medicine. *Clin Chem* 39 (1993); 561-77.

Paper I

Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water

Reprint from *Applied Spectroscopy* 63 (2009); 936-940

applied spectroscopy

Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water

A. J. LAWAETZ* and C. A. STEDMON

Dept. Food Science, Faculty of Life Sciences, University Copenhagen, Rolighedsvej 30, DK-1958, Frederiksberg, Denmark (A.J.L.); and Department of Marine Ecology, National Environmental Research Institute, University of Aarhus, Frederiksborgvej 399, Roskilde, 4000 Denmark (C.A.S.)

Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water

A. J. LAWAEZ* and C. A. STEDMON

Dept. Food Science, Faculty of Life Sciences, University Copenhagen, Rolighedsvej 30, DK-1958, Frederiksberg, Denmark (A.J.L.); and Department of Marine Ecology, National Environmental Research Institute, University of Aarhus, Frederiksborgvej 399, Roskilde, 4000 Denmark (C.A.S.)

Fluorescence data of replicate samples obtained from different fluorescence spectrometers or by the same spectrometer but with different instrument settings can have great intensity differences. In order to compare such data an intensity calibration must be applied. Here we explain a simple calibration method for fluorescence intensity using only the integrated area of a water Raman peak. By applying this method to data from three different instruments, we show that it is possible to remove instrument-dependent intensity factors, and we present results on a unified scale of Raman units. The method presented is a rapid and simple approach suitable for routine measurements with no need for hazardous chemicals.

Index Headings: Fluorescence spectroscopy; Intensity standardization; Water Raman peak.

INTRODUCTION

Fluorescence spectroscopy is a popular analytical method applied in a wide range of fields.¹⁻³ It has the advantage of being highly sensitive (down to ppb) and selective, which makes it a powerful analytical tool for both quantitative and qualitative analysis.⁴ There are, however, certain drawbacks. The fluorescence signal can be instrument dependent, and therefore there is a need for standardization if fluorescence data are to be compared between instruments.

For fluorescence data there are three data correction stages that have to be considered prior to data analysis (Fig. 1). The first stage is correction for the spectral properties of the instrument (excitation and emission correction). This removes the instrument-specific spectral biases. This correction step is important and new approaches are currently being assessed and developed by the National Institute of Standards and Technology (NIST, Gaithersburg, MD) and the Federal Institute for Materials Research and Testing (BAM, Germany).⁵⁻⁷ The second stage is correction for the absorption properties of the sample (often referred to as inner filter effects⁴). In optically thin samples with low absorbance, this is not necessary.⁸ After completing these first two stages, data should be spectrally inter-comparable between instruments and over time. However, the intensity of the fluorescence signal is not yet calibrated, and this is done in the final stage. In the same sense as the spectral position of the peak can be influenced by the instrument, the intensity of the fluorescence signal is also very dependent on the instrument. Different instruments have different detector systems and/or use different photomultipliers; hence they often use different scales for the fluorescence intensity. Additionally, the fluorescence intensity is almost always (except for photon counting systems) given in arbitrary units (A.U.). This makes quantitative fluorescence

spectroscopy difficult across different instruments. When working with well characterized and known fluorophores, this is a minor problem that can be circumnavigated using a series of concentration standards. This is, for example, done in the routine measurement of the plant pigment *chlorophyll a* in marine research.⁹ However, when working with a complex mixture of potentially unknown fluorophores a different approach is required. To date the majority of fluorescence studies attempt to avoid this problem either by carrying out all measurements on the same instrument or by using an external well-characterized standard such as quinine sulfate or, more recently, using reference standard material 2941, both supplied by NIST.¹⁰⁻¹²

There is, however, an alternative less widespread method that uses the scattering properties of pure water as a quantitative standard. This involves the properties of the water Raman peak. The technique has been applied before^{13,14} but not explained or demonstrated thoroughly or clearly in the literature. Also, applications so far are nearly exclusively found within the aquatic sciences though this method can be applied universally to a broad range of fluorescence applications. Here we explain the approach in a simple way and emphasize its utility as a straightforward and robust calibration technique. The technique is demonstrated for single excitation and emission wavelength pair fluorescence measurements, but it is equally applicable for 2d and 3d fluorescence spectra. A minimum of data are presented in order to make the presentation user friendly and to stress the simplicity of the method.

RAMAN SCATTER BAND AND THE CALIBRATION APPROACH

Pure water has two clear scatter peaks: Rayleigh and Raman (Fig. 2). The first is due to direct scattering of the incident light and therefore occurs at the same wavelength as the excitation. The water Raman peak is, however, a result of non-elastic scatter. A fraction of the incident photons lose energy to vibration in water molecules and the photon is then scattered at a higher wavelength than the incident light. The energy loss in water has a fixed frequency of approximately 3400 cm^{-1} .¹⁵ The Raman peak has a relatively low intensity and is often overshadowed by the fluorescence of even moderate concentrations of fluorophores that fluoresce at these wavelengths.

The wavelength-dependent Raman cross-section of water is a fixed property of water and the integral of the measured Raman peak (A_{rp}) (Fig. 2) is directly proportional to it.¹⁵ A_{rp} can therefore be used to calibrate measurements made on different instruments, or made with different instrumental settings as the peak height and width will vary accordingly. Raman peaks measured for each setup used to measure samples

Received 13 March 2009; accepted 19 May 2009.

* Author to whom correspondence should be sent. E-mail: ajla@life.ku.dk.

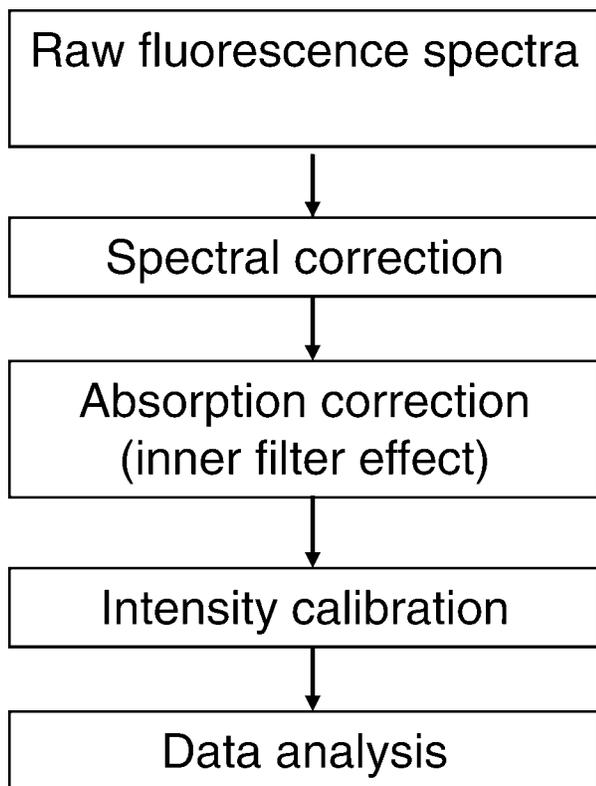


FIG. 1. Flow chart of the different correction steps for fluorescence data.

can then be used to calibrate the data accordingly. This is particularly relevant for applications in which the concentration span of the fluorophore of interest is so wide that it is necessary to change the instrument setup, or equally in situations where samples are measured at different locations. The spectral position of the Raman peak in a water-based solution for any specific excitation wavelength (λ_{ex}) can be calculated using Eq. 1 as illustrated in Fig. 2:

$$\begin{aligned} \text{Raman peak position } [\lambda_{\text{em}}(\text{nm})] \\ = 1 \times 10^7 \left(\frac{1 \times 10^7}{\lambda_{\text{ex}}} - 3400 \right)^{-1} \end{aligned} \quad (1)$$

In order to calculate the integral of the Raman peak (A_{rp}) one needs to define the wavelength band over which to integrate. Assuming pure water is used, the signal on either side of the peak should be very low and within instrumental noise. As the width of the Raman peak varies depending on instrumental setup, we suggest that a relatively broad fixed band is used. This ensures that it is valid for as broad as possible a range of excitation wavelengths and instrumental parameters. We have chosen to define the Raman peak width as peak position $\pm 1800 \text{ cm}^{-1}$. For an excitation wavelength of 350 nm, this equates to a band spanning from 371 to 428 nm. Depending on applications, alternative Raman peaks (i.e., from a different excitation wavelength) may be used, but it is important to report which excitation wavelength has been used for the calibration. We suggest the Raman peak from 350 nm excitation as this is already used for signal-to-noise determinations and it is within the range of many fluorescence spectrophotometers. A_{rp} is dependent on the excitation

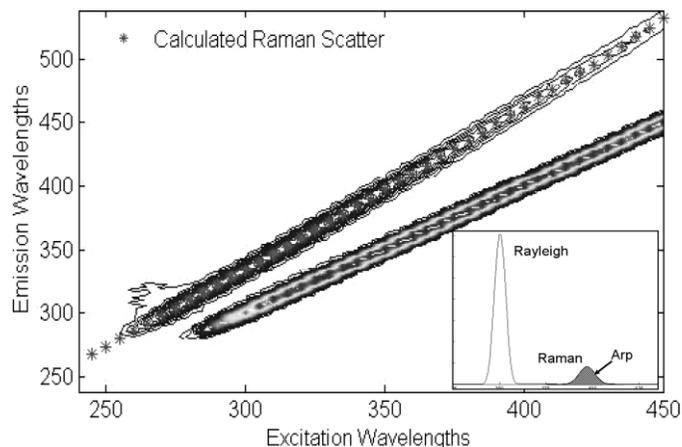


FIG. 2. Excitation–emission matrix for a MilliQ water sample showing the Rayleigh and Raman scatter bands. Also plotted are the calculated Raman peak wavelengths according to Eq. 1. Insertion is the emission spectrum at 350 nm excitation showing the Rayleigh and Raman peaks. The area under the Raman peak has been marked to illustrate the A_{rp} .

wavelength chosen and is calculated according to Eq. 2:

$$A_{\text{rp}}^{\lambda_{\text{ex}}} = \int_{\lambda_{\text{em}}^1}^{\lambda_{\text{em}}^2} I_{\lambda_{\text{em}}} d\lambda_{\text{em}} \quad (2)$$

I_{λ} is the measured spectrally corrected intensity of the Raman peak at emission wavelength λ . For practical use, A_{rp} is obtained by summing the intensity at every wavelength. It is important to note that some fluorescence spectrophotometers record emission data at intervals other than every 1 nm. This has to be taken into consideration before calculating A_{rp} .

To perform the calibration the fluorescence of a sample at any wavelength is normalized to A_{rp} determined daily for the particular instrumental setup (Eq. 3). The fluorescence signal at all measured wavelengths is now calibrated to so-called Raman Units (R.U.), which is in turn quantitatively independent of instrument specificities and therefore comparable to measurements from other instruments or from the same instrument but with different settings.

$$F_{\lambda_{\text{ex}}, \lambda_{\text{em}}} (\text{R.U.}) = \frac{I_{\lambda_{\text{ex}}, \lambda_{\text{em}}} (\text{A.U.})}{A_{\text{rp}}} \quad (3)$$

It is important to note that this approach differs from another commonly used Raman intensity normalization approach, in which the measured signal is normalized to the peak intensity alone rather than A_{rp} (e.g., Giana et al. (2003)¹⁶; Holbrook et al. (2006)⁷). Only normalizing to A_{rp} will result in a truly universal scale that should be independent of instrumental parameters, provided that spectral corrected data is used.^{5,6} This will be demonstrated with some simple examples.

EXPERIMENTAL

In order to demonstrate the approach, a simple inter-calibration was carried out on three different instruments. A series of concentration standards of quinine sulfate were made using quinine sulfate obtained from NIST in perchloric acid (HClO_4) according to the procedure from Velapoldi and Mielenz.¹⁰ The series of concentration standards consisted of concentrations of 0, 1, 2.5, 5, and 7.5 ppb. Fluorescence

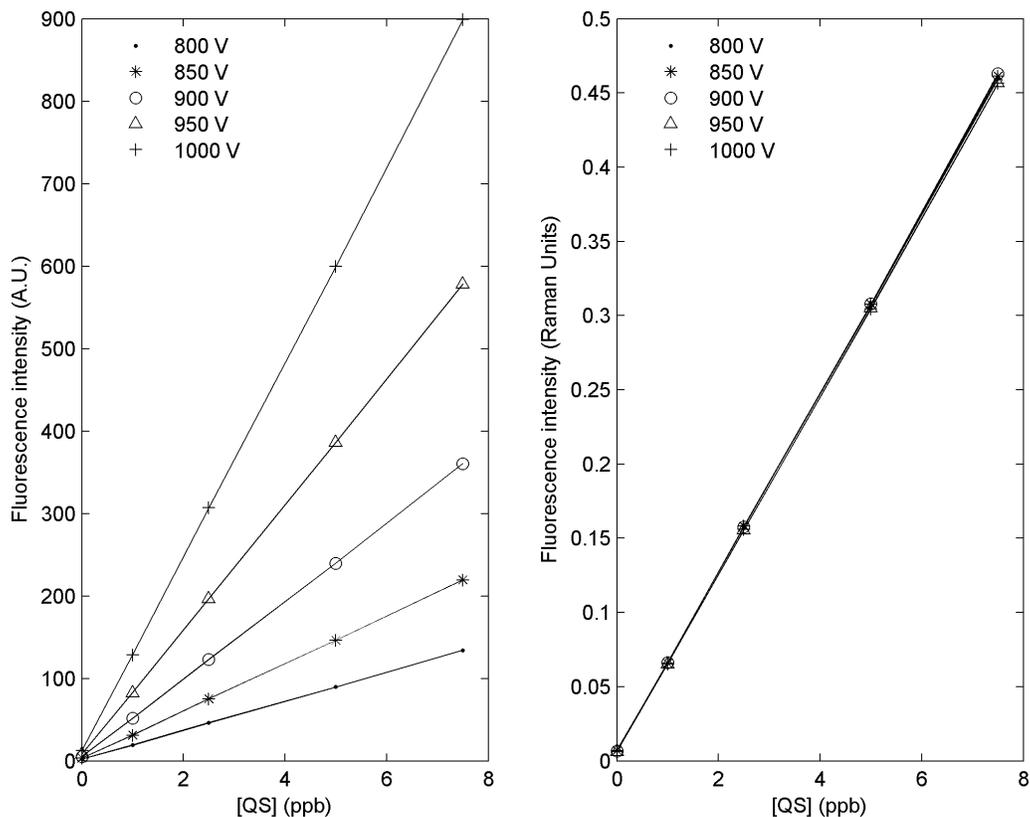


FIG. 3. Fluorescence of quinine sulfate at different concentrations measured using variable photomultiplier tube (PMT) voltages on a Varian Cary Eclipse. (Left) before and (right) after Raman calibration. Raman units can easily be converted to QS-equivalents by the equation $QS = RU/0.0767$.

measurements were carried out on a Varian Cary Eclipse, an LS 55 Perkin Elmer, and an FS920 Edinburgh Instruments fluorescence spectrophotometer. For all measurements excitation and emission slits were set to 5 nm. Water Raman spectra were recorded with an excitation wavelength of 350 nm and emission wavelengths from 365 to 430 nm. Fluorescence of quinine sulfate (QS) was measured at an excitation wavelength of 250 and emission wavelength of 450 nm. The photomultiplier tube (PMT) voltage was varied during the measurements. For the Varian, the voltages used were 800, 850, 900, 950, and 1000 V, and for the LS55 700, 750, 800, 850, and 900 V were used. The FS920 is a photon-counting instrument and hence the detector voltage could not be varied. All the samples (QS solutions and water samples) were measured using the different instrument setups. All samples were measured in replicates of at least five and the mean value was calculated and used for the data analysis. The whole experiment was repeated a month later with a new batch of QS standards. Additionally, spectra of a single sample were measured (ex. 250 nm, em. 300–600 nm) on the Varian with different excitation and emission slit width settings (5/5, 10/5, 2.5/5, 1.5/5, and 5/2.5 nm). Five parts per billion (5 ppb) QS and detector voltage of 950 V was used for these measurements. Emission spectra were corrected for the wavelength-dependent spectral bias using a correction factor derived by use of secondary emission standards provided from BAM.⁵ In addition, during measurement the source intensity in all three instruments was normalized to that of an internal reference detector.

RESULTS AND DISCUSSION

As stated above, sometimes it can be necessary to change the instrument setup in order to obtain the best spectra from a set of samples. The effects of this and the result of the subsequent Raman calibration are illustrated in Fig. 3. A greater detector voltage results in a greater fluorescence signal on the same solution (Fig. 3, left panel). This makes quantitative comparison of the measurements impossible. Applying Raman calibration to the data removes these differences and places all measurements on an equal scale of Raman units (Fig. 3, right panel). It is now straightforward to make quantitative judgments of the fluorophore of interest across measurements with different instrument setups.

The instruments applied in this study were from three different manufacturers, and they use different techniques for detecting the fluorescence signal. The FS 920 from Edinburgh Instruments is a photon-counting instrument, whereas the two other instruments use an arbitrary scale from 0 to 1000, but these are not equally calibrated. This will, of course, give three different results in terms of intensity, thereby making inter-instrumental comparison impossible without applying intensity calibration. The Raman calibration applied here removes these instrument-specific intensity factors and is thus suitable for such inter-instrument intensity calibration. The results of two series of concentration standards of QS measured on two different days on the three instruments after Raman calibration are shown in Fig. 4. Put on a Raman unit scale, the same concentration of QS gives the same intensity independent of instrument.

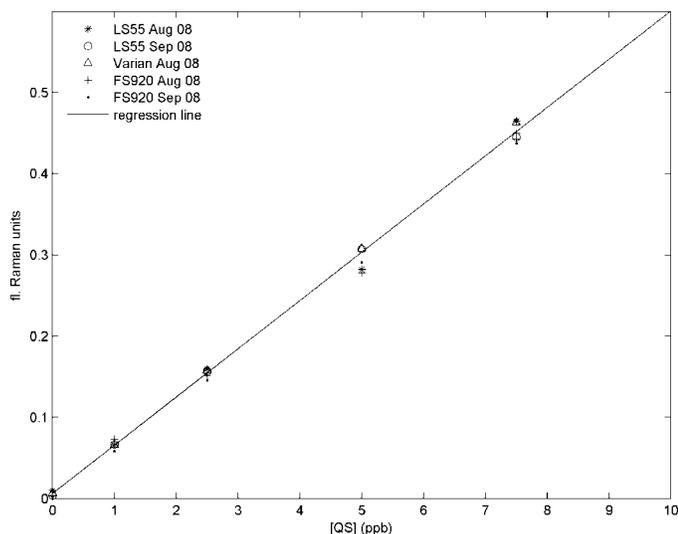


FIG. 4. Raman corrected fluorescence intensity of a standard series of quinine sulfate. Two identical series are measured on two different days, on three different instruments.

In the above example we have varied the PMT voltage of the instruments, but as illustrated in Fig. 5 similar results are obtained if excitation and emission slit widths are varied instead. Different slit widths result in different intensities of the spectrum, but after Raman correction the intensities are equal. Of course, changing the slit widths can in extreme cases change the spectral shape of the peak. Especially if there are narrow and well defined peaks, some of the resolution will be lost. This is of course not handled by the Raman correction. The Raman corrected spectrum with slit settings 1.5/5 is noisy due to the fact that both the Raman peak and the raw spectrum have low intensity and, hence, have a low signal-to-noise ratio. This is a

general limitation of the method: if the instrument setup applied does not allow measuring a Raman peak with a proper signal-to-noise ratio, too much noise will be introduced to the normalized spectra and the method will not work as well as intended. In some situations, using a Raman peak of a lower excitation wavelength can solve this problem, as it will give a better signal-to-noise ratio on the Raman peak. However, in general the 350 nm excitation Raman peak is suitable. There is a fixed relationship dependence between A_{rp} from different excitation wavelengths,¹⁵ which makes it possible to recalibrate to different Raman peaks, should it be required.

The above results reveal that Raman calibration is a suitable tool for calibrating fluorescence measurements onto a “global” scale that makes it possible to quantitatively compare measurements from different settings on one instrument or between instruments. This calibration method is not only applicable for single excitation and emission wavelength pair fluorescence measurements (Fig. 3) but is also valid for spectral measurements (Fig. 5) as well as excitation–emission matrices (EEMs). Equation 3 is valid for all excitation–emission wavelength pairs measured, hence the subscripts λ_{ex} and λ_{em} . The integral of one, fixed, Raman peak (A_{rp}) is used to normalize the whole spectrum or EEM.

A major advantage of this approach compared to other calibration methods, such as the quinine sulfate method, is that no standards are required, thus removing operational steps (weighing, dilution, etc.), and the risk of degradation of the chemicals, which all can cause errors. The Raman approach involves no hazardous chemicals and requires only pure water (preferably deionized and ultraviolet exposed), which is available in most laboratories. To simplify the approach further, sealed cuvettes (cells) with pure water are also available from most instrument manufacturers and are suitable for this approach. In Fig. 6 a plot of three water Raman peaks is shown; one is from a fresh MilliQ sample, whereas the other

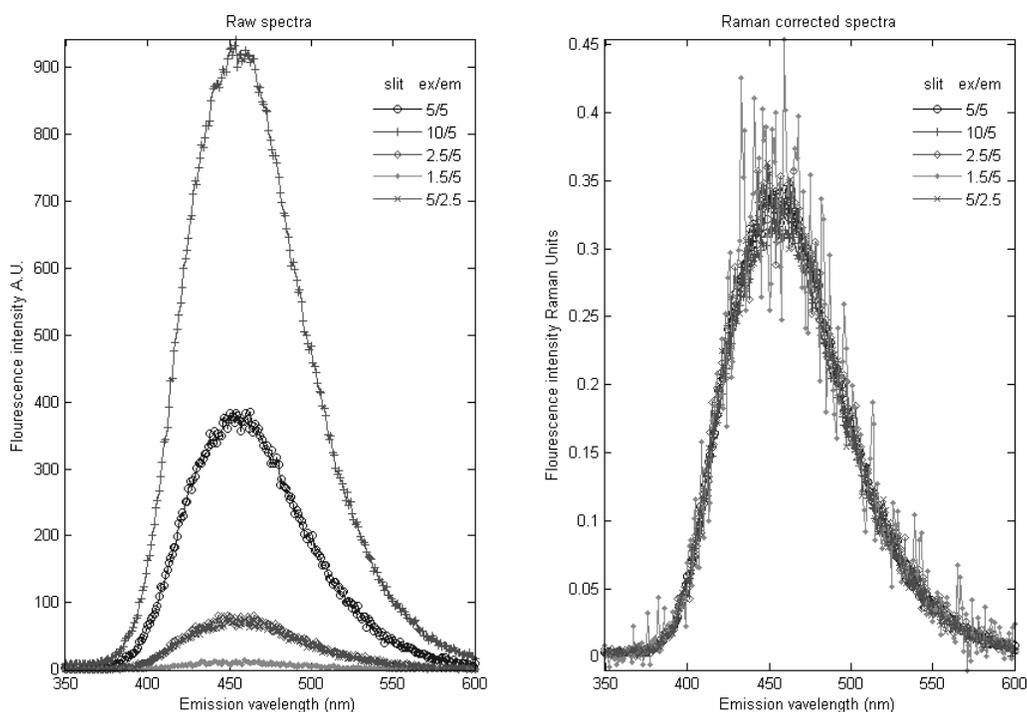


FIG. 5. Spectra of quinine sulfate solution obtained at different slit settings (left) before and (right) after Raman correction.

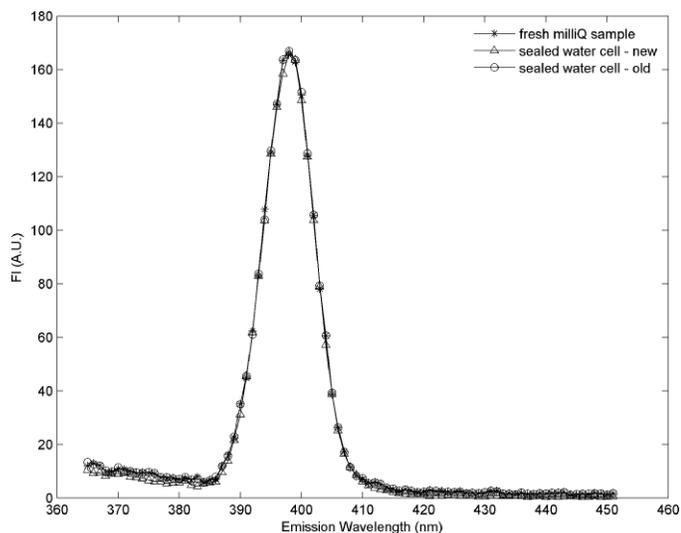


FIG. 6. Water Raman peak from a fresh MilliQ sample and two sealed water cells, a new and an old; the old one is approximately 10 years old.

two are obtained from sealed water cells. The three spectra are more or less identical (coefficient of variation for $A_{rp} < 2\%$), even though one of the sealed water cells is 10 years old. The sealed water cell can thus ensure a uniform water quality every day for a long period of time. It should also be noted that this calibration procedure is universal, and the Raman signal of water can be used irrespective of sample solvent/matrix.

Fluorescence intensity differences, inter-instrumental or due to instrument setup, can be calibrated for using the integral of the water Raman peak (A_{rp}). By Normalizing all fluorescence data to the integral of the Raman peak from excitation at 350 nm we have shown that it is possible to calibrate fluorescence data onto a global scale of Raman Units. Hereby we enable a direct comparison of fluorescence intensity from different

instruments or from the same instrument using different instrumental settings and over time. It is important to stress that this is only an intensity normalization/calibration procedure. No spectral changes occur from applying this method. The problem of spectral correction should, if necessary, be addressed as an independent operation prior to the intensity calibration.⁵⁻⁸

ACKNOWLEDGMENTS

A.J.L. is financed by the Villum Kann Rasmussen Foundation. C.A.S. was supported by the Danish Research Council (grant no. 274-05-0064 and 272-0785).

1. J. Christensen, L. Norgaard, R. Bro, and S. B. Engelsen, *Chem. Rev.* **106**, 1979 (2006).
2. O. S. Wolfbeis and M. Leiner, *Anal. Chim. Acta* **167**, 203 (1985).
3. P. G. Coble, S. A. Green, N. V. Blough, and R. B. Gagosian, *Nature (London)* **348**, 432 (1990).
4. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Springer, New York, 2006), 3rd ed.
5. U. Resch-Genger, D. Pfeifer, C. Monte, W. Pilz, A. Hoffmann, M. Spieles, K. Rurack, J. Hollandt, D. Taubert, B. Schonenberger, and P. Nording, *J. Fluorescence* **15**, 315 (2005).
6. P. C. DeRose, E. A. Early, and G. W. Kramer, *Rev. Sci. Instrum.* **78**, 033107 (2007).
7. R. D. Holbrook, P. C. DeRose, S. D. Leigh, A. L. Rukhin, and N. A. Heckert, *Appl. Spectrosc.* **60**, 791 (2006).
8. C. A. Stedmon, A. G. Ryder, N. Harrit, and R. Bro, *Pure Appl. Chem.*, paper submitted (2009).
9. C. S. Yentsch and D. W. Menzel, *Deep-Sea Res.* **10**, 221 (1963).
10. R. A. Velapoldi and K. D. Mielenz, *NBS Spec. Publ.* 260-64 (1980).
11. P. G. Coble, C. A. Schultz, and K. Mopper, *Marine Chem.* **41**, 173 (1993).
12. P. C. DeRose, M. V. Smith, K. D. Mielenz, D. H. Blackburn, and G. W. Kramer, *J. Luminescence* **128**, 257 (2008).
13. C. A. Stedmon, S. Markager, and R. Bro, *Marine Chem.* **82**, 239 (2003).
14. S. Determann, R. Reuter, P. Wagner, and R. Willkomm, *Deep-Sea Res. Part I-Ocean. Res. Papers* **41**, 659 (1994).
15. G. W. Faris and R. A. Copeland, *Appl. Opt.* **36**, 2686 (1997).
16. H. E. Giana, L. Silveira, Jr., R. A. Zângaro, and M. T. T. Pacheco, *J. Fluorescence* **13**, 489 (2003).

Paper II

Application of rotated PCA models to facilitate interpretation of metabolite profiles: commercial preparations of St. John's Wort

Reprint from *Planta Med.* 75 (2009); 271-279

Application of Rotated PCA Models to Facilitate Interpretation of Metabolite Profiles: Commercial Preparations of St. John's Wort

Author

Anders Juul Lawaetz¹, Bonnie Schmidt¹, Dan Staerk², Jerzy W. Jaroszewski³, Rasmus Bro¹

Affiliation

¹ Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Copenhagen, Denmark
² Department of Basic Sciences and Environment, Faculty of Life Sciences, University of Copenhagen, Copenhagen, Denmark
³ Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark

Key words

- *Hypericum perforatum* L.
- St. John's wort
- Clusiaceae
- principal component analysis (PCA)
- orthogonal rotation
- metabolite profiling

Abstract

▼ This paper describes the application of orthogonal rotation of models based on principal component analysis (PCA) of ¹H nuclear magnetic resonance (NMR) spectra and high-performance liquid chromatography-photo diode array detection (HPLC-PDA) profiles of natural product mixtures using extracts of antidepressive pharmaceutical preparations of St. John's wort as an example. ¹H-NMR spectroscopy of complex mixtures is often used in metabolomic, metabonomic and metabolite profiling studies for assessment of sample composition. Interpretation of the derived chemometric models may be complicated because several sample properties often contribute to each principal component and because the in-

fluence of individual metabolites may be shared by several principal components. Furthermore, extensive signal overlap in ¹H-NMR spectra poses additional challenges to the interpretation of PCA models derived from such data. Orthogonal rotation of PCA models derived from ¹H-NMR spectra and HPLC-PDA profiles of the extracts of St. John's wort preparations facilitate interpretation of the model. Using the varimax criterion, rotation of loadings provides simpler conditions for understanding the influence of individual metabolites on the observed clustering. Alternatively, rotation of scores simplifies the understanding of the influence of whole metabolite profiles on the clustering of individual samples.

Introduction

▼ ¹H-NMR spectroscopy is an attractive analytical technique for assessment of samples of biological origin, i. e., biofluids and plant extracts. The technique is non-destructive, applicable to intact biomaterial and information-rich with regard to molecular structure elucidation. Thus, the technique has been widely used as the analytical platform to generate information-dense data in metabonomic, metabolomic, and metabolite profiling studies. However, ¹H-NMR spectra of biological samples can be extremely complex as they may contain thousands of distinctive resonances. Therefore, visual inspection of a series of such spectra may only release a small percentage of the total information available.

Computer-based methods are often used to reduce the complexity of data to a suitable level. In ¹H-NMR-based metabolite profiling studies, principal component analysis (PCA) is often used [1]. Graphical outputs from PCA enable researchers across disciplines to discuss detailed facets of

conceivably complex mathematical models. A PCA model uses orthogonal and intrinsically abstract latent variables. This means that interpretation of the model in terms of finding the connection between loadings and the variables used in the analysis can be difficult. Even though a PCA bi-plot of scores and loadings provides insight into the structure of the data, it can still be difficult to interpret the many correlations occurring in NMR-based metabonomic studies.

In this study we explore a route to simplify the interpretation of complex PCA models with respect to the influence of individual compounds on the observed clustering of samples. ¹H-NMR spectra and HPLC-PDA profiles of extracts of 24 commercially available preparations of St. John's wort, a popular herbal medicine, are used as model data sets. Metabolite profiles based on ¹H-NMR spectroscopy have previously proven useful for assessment of herbal medicines or plant extracts using different two-way chemometric methods [2], [3], [4], [5], [6], [7], [8], [9].

received June 4, 2008

revised September 26, 2008

accepted October 27, 2008

Bibliography

DOI 10.1055/s-0028-1112194
 Planta Med © Georg Thieme
 Verlag KG Stuttgart · New York
 Published online 2008
 ISSN 0032-0943

Correspondence

Anders Juul Lawaetz

Department of Food Science
 Faculty of Life Sciences
 University of Copenhagen
 Rolighedsvej 30
 1958 Frederiksberg C
 Denmark
 Tel.: +45-3533-3254
 Fax: +45-3533-3245
 ajla@life.ku.dk

In an earlier study comprising commercial preparations of St. John's wort obtained from retail stores in Denmark, interpretation of the full-resolution $^1\text{H-NMR}$ data was based on separate PCA models derived from samples formulated as tablets and capsules, respectively [8]. Moreover, another data set derived from St. John's wort preparations originating from several continents and based on HPLC-PDA profiles has been analyzed by applying parallel factor (PARAFAC) analysis [10]. This provided relative concentrations of individual compounds, which were used to facilitate comparison of samples by PCA. Interpretation of the PCA model in terms of constituents responsible for the differences and similarities in composition between preparations was straightforward, because the analysis focused on well-characterized compounds. However, the influence of each compound was shared by several components, complicating the interpretation of the PCA model, because more components had to be interpreted to understand the interrelationship between individual compounds and the samples. In the present study the data set size of the originally investigated $^1\text{H-NMR}$ data set [8] has been extended to include St. John's wort preparations from several continents, previously investigated [10] by HPLC-PDA. The aim of this study is to be able to interpret the full-resolution $^1\text{H-NMR}$ data as well as HPLC-PDA data from PCA models based on the entire collection of samples.

To simplify complex PCA models of data sets based on $^1\text{H-NMR}$ spectra and HPLC-PDA profiles, rotations of loadings and scores have been performed. Such rotations can lead to model representations where individual variables are more exclusively related to distinct components rather than being shared across many. Rotations can be performed with techniques such as varimax and quartimax rotation [11]. The use of rotations in multivariate data analysis is not a new approach, and it has been used for decades in some areas, e.g., in psychometrics [12]. However, in natural sciences in general and metabolomics and metabonomics in particular the use of rotations of PCA models is far more limited, although a few recent examples can be found [13], [14], [15], [16]. Traditionally, the use of rotations in PCA modelling has not been strictly needed, because multivariate modelling has usually been performed on fairly simple data. Even though data sets with hundreds or thousands of variables have often been used in chemometrics, the real underlying complexity of the data was usually low, involving either a few independent components, or many but highly correlated variables as in analysis of profiles of electronic or vibrational spectra. Nowadays, data sets such as those arising in metabolomic, metabonomic and metabolite profiling studies have a much higher complexity. Thus, rather than analyzing, e.g., UV spectra profiles spanning the whole variable domain, it is common to study variables represented by separate narrow peaks, like those present in $^1\text{H-NMR}$ and mass spectra. This creates the need for additional mathematical tools to simplify interpretation.

In this paper, we present an application of rotated PCA models of $^1\text{H-NMR}$ and HPLC-PDA data representing complex natural mixtures.

Materials and Methods



Extracts of St. John's wort preparations

The extracts of twenty-four different commercial preparations of St. John's wort were the same as described elsewhere [10].

Thirteen preparations were formulated as tablets (preparations 1 – 4, 11, 12, 14, 16, 17, 21–24), and the remaining as capsules (preparations 5–10, 13, 15, 18–20). Preparations 1 – 10, 23, and 24 originated from Europe, preparations 11 and 17 from Asia, preparations 12, 13, 15, 16, 18, and 20–22 from North America and preparations 14 and 19 from Africa. For one of the brands, two different batches were obtained (preparations 7 and 8). For acquisition of NMR data, the extracts were lyophilized twice with 2 mL of methanol and 90 mL of water and once with 2 mL of D_2O .

NMR experiments

NMR experiments were performed on a Bruker Avance spectrometer (^1H resonance frequency of 600.13 MHz) (Bruker BioSpin) using standard Bruker library pulse sequences. 1 D $^1\text{H-NMR}$ spectra were recorded using a 5 mm TXI probe. Samples of the extracts (15 mg) were dissolved in 700 μL of $\text{DMSO-}d_6$ (99.8 atom% of deuterium) and transferred into 5 mm NMR tubes. Each sample was prepared in triplicate. For each sample 128 transients were collected using 64 k data points with a spectral width of 16 ppm, using 30° pulses and inter-pulse delay of 4.41 s in order to obtain practically fully relaxed spectra. The spectra were Fourier-transformed to 128 k data points, using line broadening of 0.1 Hz, and referenced to internal TMS.

HPLC-PDA data

The HPLC-PDA data, aligned using an extended algorithm of correlation optimized warping (COW), were the same as described previously [10].

Pre-treatment of $^1\text{H-NMR}$ data

$^1\text{H-NMR}$ data were phased and referenced in Xwin-nmr ver. 3.1 (Bruker BioSpin) and imported into MATLAB ver. 7.0.1 software (MathWorks) for further data pre-treatment and data analysis. A cubic polynomial baseline correction was applied and regions corresponding to residual solvents ($\delta=2.48 - 2.54$), water ($\delta=3.16 - 3.52$), residual extraction solvents ($\delta=7.36 - 7.40$, $7.76 - 7.82$, and $8.56 - 8.60$), and compounds not considered interesting in relation to this work (fatty acids $\delta=0.82 - 0.88$, $1.20 - 1.30$, and $2.14 - 2.22$) were excluded. This exclusion resulted in a data set containing 77,717 variables (NMR descriptors). After exclusion of the specified regions, data were autoscaled using an offset of 500,000 with the following equation:

$$x_{ij}^{\text{autoscaled with offset}} = \frac{x_{ij} - \bar{x}_j}{s_j + \text{offset}}$$

This offset was chosen based on a visual assessment of the magnitude of variables containing only noise or baseline. The use of an offset prevents these noise areas to have too much influence on the models.

Software for rotation

The rotations were performed using a general rotation tool for PCA models made in MATLAB. The tool is available from www.models.life.ku.dk (September, 2008).

Results and Discussion



Principles of rotation of PCA models

Given a data matrix \mathbf{X} of size $I \times J$ (objects \times variables), principal component analysis is a way of partitioning \mathbf{X} into a systematic part and a residual (noise) part. The systematic part consists of possibly a few latent variables, i.e., principal components that summarize the largest variance in the data. The projection of I objects in \mathbf{X} onto the first loading vector \mathbf{p}_1 provides the score values of the first component, \mathbf{t}_1 , which describes maximal variation in the data. Subsequent components are found similarly and describe as much as possible of yet unexplained variation. PCA can be described by $\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$ where \mathbf{T} is the score matrix holding the above score vectors as columns, \mathbf{P}^T is the transposed loading matrix, and \mathbf{E} is the residuals. The scores and loadings are determined so as to minimize the residuals in a least-squares sense.

For a given PCA model, it is possible to rotate the scores or loadings in the model without affecting the overall fit of the solution if either the loadings or the scores are similarly counter-rotated [17]. Hence, rotation is simply a way to represent the systematic variation differently. The actual variation described by the overall PCA model is not changed.

If \mathbf{Q} is an $m \times m$ orthogonal matrix, i.e., $\mathbf{Q} \times \mathbf{Q}^T = \mathbf{I}$ and we define $\mathbf{S} = \mathbf{TQ}$ and $\mathbf{M}^T = \mathbf{Q}^T \mathbf{P}^T$ then $\mathbf{TP}^T = \mathbf{TQQ}^T \mathbf{P}^T = \mathbf{SM}^T$. Hence, $\mathbf{X} = \mathbf{SM}^T + \mathbf{E}$ is a model with scores \mathbf{S} and loadings \mathbf{M} that are rotated versions of the original ones, and which represents exactly the same fit to the data.

The major challenge when applying a rotation to a PCA model is how to choose the rotation matrix \mathbf{Q} . From a mathematical point of view, there is an infinite number of ways to define \mathbf{Q} and different criteria for its choice have been developed [11], [12], [17], [18], [19]. The main principle of these criteria is to rotate towards a simpler structure, i.e., the rotation procedure seeks to establish a simpler relationship within the individual loadings so that these become easier to interpret [20]. An example of a simple structure could be the largest possible loading of a variable in one component, resulting in diminished loadings of the same variable for other components. Hence, samples in the direction of this loading vector can be clearly associated with distinct variables.

As an example, consider a loading matrix, which reads:

$$\mathbf{P} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{bmatrix}$$

This is a complicated structure because both original variables contribute significantly in both components (columns of \mathbf{P}). Rotating this model by a rotation matrix \mathbf{Q} (which in this case happens to be equal to \mathbf{P}) yields

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

This is a loading matrix with an obviously simple structure, because now every manifest or measured variable is only associated with one latent variable. Thus, rotations are used to obtain another view of the model in which each variable is maximally

correlated with one component and reaches a near-zero correlation with other components. The fit of the overall explained variance of the model is unchanged upon the rotation, but the scores and the contribution of explained variance of each component in the PCA model as well as the loadings will inevitably change. In a PCA model, the first component explains the largest fraction of variance and the subsequent components describe progressively smaller fractions. Upon rotation, this is no longer the case.

Rotating the PCA model towards simplicity of scores rather than simplicity of loadings is equally feasible, as follows from the symmetry of the PCA model. However, most studies published so far have used rotation of the PCA model for obtaining simpler loadings [15], [16], [17]. Rotation of scores can be particularly useful when a certain clustering is expected among the samples, as shown in the following paragraphs.

Two general categories of rotations are available, orthogonal and oblique rotations. In the first category, the angular dependence between the original set of loading vectors is preserved (as in the simple example stated above), whereas in the latter category, the angles between loading vectors are not necessarily preserved. The quartimax and varimax criteria are orthogonal rotations, whereas criteria such as oblimin, promax and simplimax represent oblique rotations [19], [21].

One advantage of the orthogonal rotations is that orthogonality makes the numerical approaches simpler and better behaved. A potential drawback could be that orthogonality between loadings is seldom the reality of the underlying features. However, the aim of rotations as presented here is not to find the 'true' profiles, but rather to find a mathematical representation that can simplify interpretation. None of the traditional rotation methods, be they orthogonal or not, can provide estimates of real profiles in normal situations. Hence, the choice of rotation method should generally not be guided by a quest for true profiles. If such estimates are sought, then the family of curve-resolution methods is useful. In this study we focus on orthogonal rotations.

Among orthogonal rotations, the quartimax criterion described by Ferguson [11], [17], [22], as well as the varimax criterion described by Kaiser [11], [23] have been commonly described under the orthomax criterion [24]. The principle of the orthomax rotations is to maximize the orthomax criterion given by:

$$V = \sum_{j=1}^J \sum_{f=1}^F p_{jf}^4 - \frac{\gamma}{J} \sum_{f=1}^F \left(\sum_{j=1}^J p_{jf}^2 \right)^2$$

where p_{jf} is the loading value for variable j on component f , $j = 1, \dots, J$ represent variables, and $f = 1, \dots, F$ represent components; $0 \leq \gamma \leq 1$. If $\gamma = 0$ the equation becomes the quartimax criterion and if $\gamma = 1$ it becomes the varimax criterion.

The varimax criterion is by far the most often applied method among the orthogonal rotations [13], [14], [15], [16]. Maximizing the varimax criterion provides a solution where the variance of the squared loading elements is maximized. For two competing solutions, the one having a higher varimax criterion value will have optimized loading values in each principal component, i.e., values that are either high (in absolute value) or close to zero. This is a solution that fulfils the definition of a simple structure [11]. On the other hand, maximizing the quartimax criterion

maximizes the variance of each (squared) variable, i.e., optimizes loadings for each variable to a high value in one component and low or zero values in other components. Hence, quartimax minimizes the number of components needed to explain each variable. Kaiser stated that there is a possible bias with the quartimax, as it tends to give one general factor [11]. The varimax rotation principle is the rotation principle applied in this study.

Interpretation of rotated PCA models based on $^1\text{H-NMR}$ spectra of St. John's wort extracts

$^1\text{H-NMR}$ spectroscopy is a non-selective technique that gives unique signals for each hydrogen-containing secondary metabolite above a certain concentration limit. Preparations of St. John's wort are complex mixtures containing many different metabolites, and $^1\text{H-NMR}$ spectra of these preparations are very complex and show hundreds of signals.

Interpretation of the derived PCA models at the individual compound level requires assignment of individual resonances of these compounds. Assignment of $^1\text{H-NMR}$ spectra of major constituents of extracts of commercial preparations of St. John's wort was performed using 2 D NMR experiments (COSY, TOCSY, *J*-resolved, HSQC and HMBC) with reference to data reported by Bilia et al. [25]. Due to the complexity of the $^1\text{H-NMR}$ spectra of the extracts, complete assignments were limited to major constituents. The 2 D NMR results guided the choice of authentic samples used for spiking, performed in order to confirm identification, especially the differentiation between closely related compounds. This led to the assignment of all resonances of chlorogenic acid, rutin, hyperoside, isoquercetin, quercetrin, and quercetin.

PCA models presented in this study are all based on data sets obtained from the full-resolution $^1\text{H-NMR}$ spectra (77,717 variables). Using the full spectral resolution rather than binned (integrated) data enhances the interpretation possibilities of derived models. $^1\text{H-NMR}$ spectra of natural product extracts often contain signals from several closely related compounds, and the use of integrated data may lead to loss of identity of individual signals and hence to loss of important information [8].

The PCA model of the preprocessed $^1\text{H-NMR}$ data used 12 components to explain 93% of the total variance in data. The number of components was chosen based on the explained variance, and on evaluation of loadings and residuals. 2 D score plots of the

first six components are shown in **Fig. 1**. An excellent separation according to supplier was achieved indicating that considerable differences between the preparations exist. This is likely due to the fact that standardizations according to procedures described in relevant pharmacopoeias [26], [27] only require standardization of a few among many constituents present. The score plots shown in **Fig. 1** clearly illustrate that it is hardly possible to find any exclusive preparation, i.e., none of the preparations is completely differentiated from the others by means of specific scores and loadings. The individual clustering of preparations shows that the content of all detected hydrogen-containing compounds is different between suppliers. Interpretation of the contributions of individual plant metabolites to the observed clustering is of utmost importance for understanding the patterns displayed in the score plots.

The loadings of the PCA model were subsequently rotated using the varimax criterion, while the scores were counter-rotated. Score plots of the first six components of the rotated PCA model are shown in **Fig. 2**. It is apparent that the first five components mainly describe features in individual preparations (preparations 14, 9, 4, 17, and 15, respectively), whereas the sixth component describes features in several preparations. Thus, rotation of loadings enabled exclusive clustering of individual preparations by means of specific loadings. This simplifies interpretation, because individual metabolites only influence a few components in the rotated PCA model as opposed to the non-rotated model, where the influences of individual metabolites are partitioned over several components. Moreover, it is also apparent from **Fig. 2** that the explained variance of each component has changed upon rotation and that the explained variance does not follow component number in a descending order. Nevertheless, the total variance explained by the rotated and the original model is exactly the same.

The loadings derived from the non-rotated as well as the rotated PCA model have been transformed using the reciprocal of the scaling factor for each variable to be able to interpret the loadings of autoscaled data. In **Fig. 3**, the back-scaled loadings corresponding to the first six components are shown for both models. It is apparent that the loadings of the rotated PCA model are more simple to interpret, e.g., the signal at $\delta=5.18$ [H-1 of glucose (Glc) in sucrose (Suc), a pharmaceutical excipient] almost exclusively influences the fourth component, and the resonan-

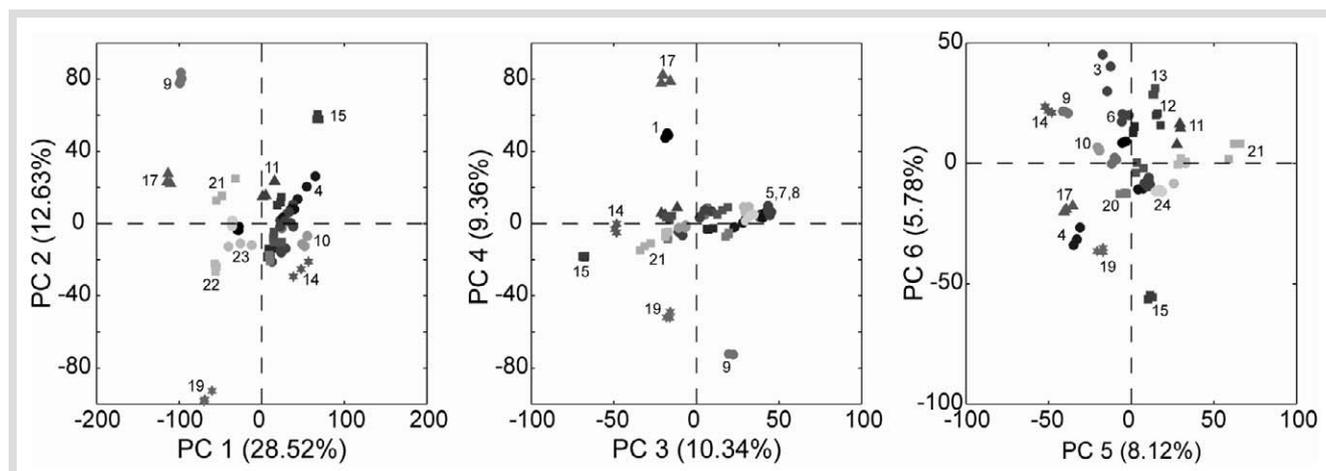


Fig. 1 Score plots of the first six components derived from a PCA model based on $^1\text{H-NMR}$ spectra of 24 preparations of St. John's wort. All samples were measured in triplicate.

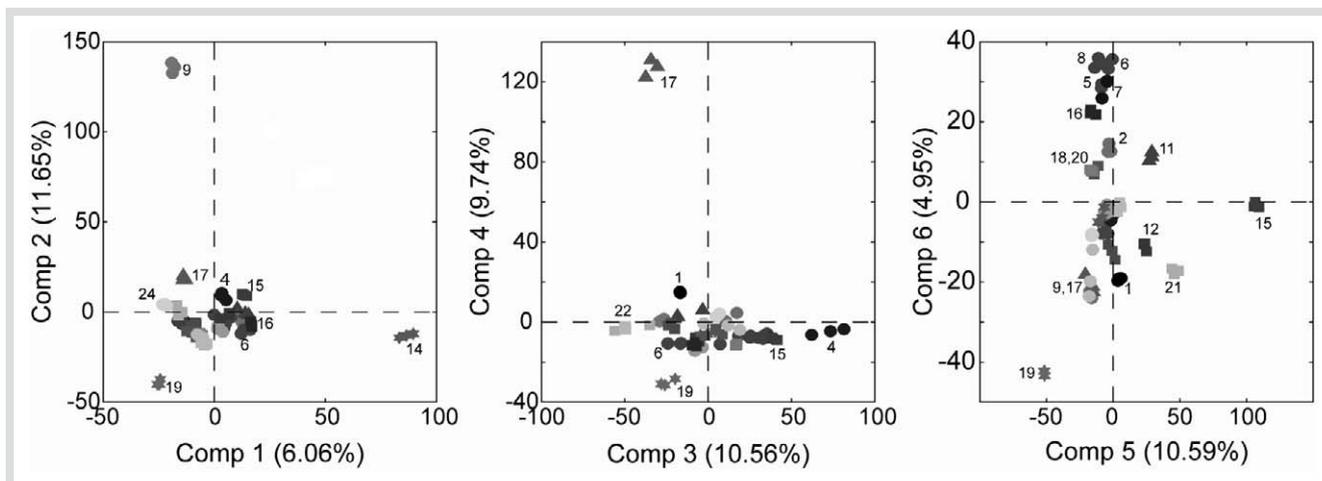


Fig. 2 Score plots of the first six components derived from a rotated PCA model (loadings rotated) based on $^1\text{H-NMR}$ spectra of 24 preparations of St. John's wort.

ces of the pharmaceutical excipients, α - and β -lactose (α - and β -Lac) at $\delta=4.89$ (H-1 Glc, α -Lac), 4.32 (H-1 Glc, β -Lac), 4.20 (H-1 Gal, β -Lac), and 4.18 (H-1 Gal, α -Lac) almost solely influence the second component. Analysis of the same signals in the loadings derived from the non-rotated PCA model reveals that the resonance signal of Suc ($\delta=5.18$) influences the first six components. The resonance signals of α - and β -Lac ($\delta=4.89$, 4.32, 4.20, and 4.18) influence the first, second, fourth, fifth, and the sixth component. Thus, interpretation of these resonances only requires analysis of two components when the rotated PCA model is used for interpretation, whereas interpretation of six components is necessary when the non-rotated PCA model is used. In fact, the loadings corresponding to the second and fourth component of the rotated PCA model provide good approximations of real $^1\text{H-NMR}$ spectra of Suc and α - and β -Lac, respectively. Suc and α - and β -Lac are primarily pharmaceutical excipients and not constituents of St. John's wort. The clustering of extracts of commercial preparation of St. John's wort due to the excipients may seem uninteresting. However, this example illus-

trates the possibility of using rotations of PCA models to obtain unique loadings for outlying samples, which may be a valuable tool for identifying causative sources for outliers. Moreover, in the specific case of medicinal products, this example illustrates that rotations of PCA models can be used to separate clustering due to excipients from that due to genuine constituents of the plant.

An interesting observation in **Fig. 3** is the distribution of the influence of signals in the region around $\delta=2$. The influence of these signals is distributed over the first, third, fifth, and the sixth component in the original as well as the rotated PCA model. The interpretation of the influence of these signals seems more straightforward using the loadings derived from the original PCA model, since mostly the third component is influenced by these signals, whereas the influence of these signals is equally distributed over the above-mentioned four components in the rotated PCA model. As already mentioned, no change in the overall fit of the model occurs upon rotation and the aim is to obtain a more simple structure with a few high loading values and many small (ideally zero) loading values. The cost can be that some loading elements do not change at all or become even more complex upon rotation, even though the overall representation is simpler. Thus, it is not possible to obtain a perfect description of every element in the matrix without changing the overall fit. Therefore, the application of rotated PCA models should be seen as an additional opportunity rather than a replacement of the original PCA model.

To be able to interpret the influence of individual plant metabolites on the observed clustering, a closer look at the loadings is necessary. The influence of quercetin on the observed clustering has been further analyzed by looking at the H-5' signal of quercetin ($\delta=6.89$). Loadings corresponding to this signal are shown in **Fig. 4** for the non-rotated PCA model (**Fig. 4A**) as well as the rotated PCA model (loadings rotated) (**Fig. 4B**).

Comparison of the loadings corresponding to H-5' of quercetin clearly illustrates that interpretation of the influence of quercetin on the observed clustering is facilitated using the rotated loadings (**Fig. 4B**) as compared to the non-rotated loadings (**Fig. 4A**). Interpretation is aided since the influence of quercetin is partitioned over many components in the non-rotated PCA model, whereas in the rotated PCA model the influence of quercetin is described mainly by the sixth and seventh components.

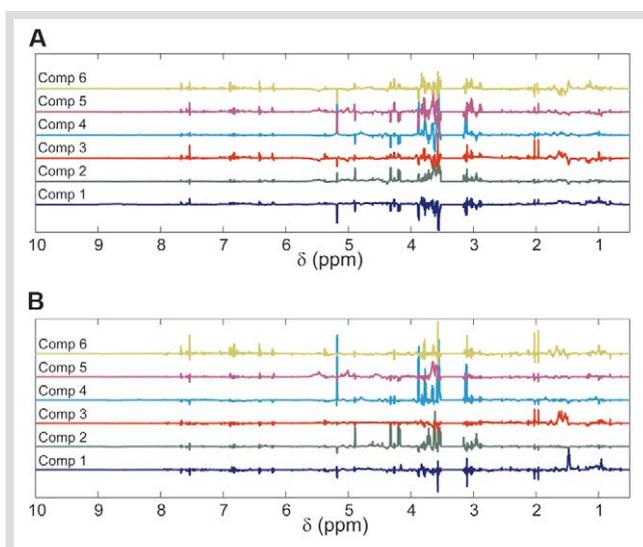


Fig. 3 Back-transformed loadings corresponding to the first six components derived from the PCA model (A) and the rotated PCA model (loadings rotated) (B).

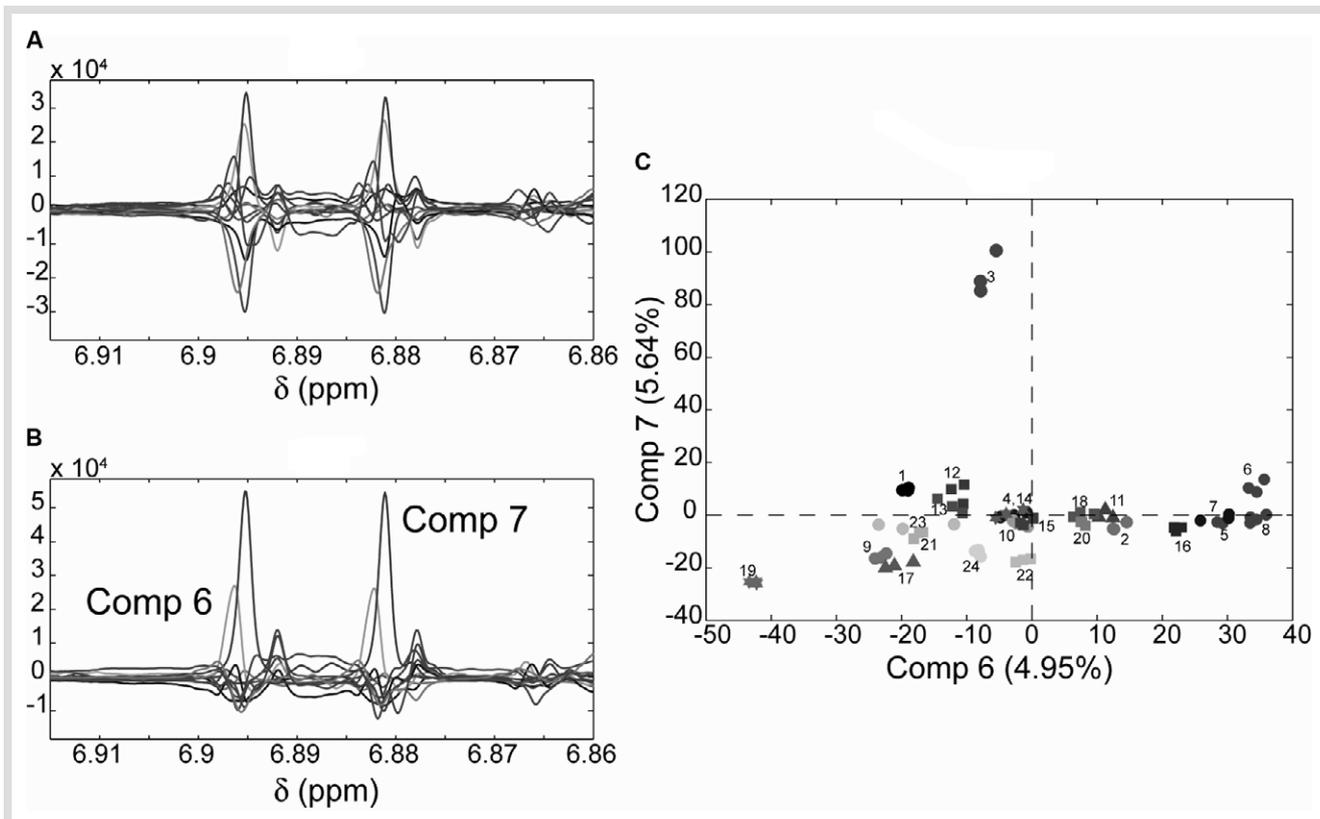


Fig. 4 Back-transformed loadings corresponding to all twelve components derived from the PCA model (A) and the rotated PCA model (loadings rotated) (B). The score plot of the sixth and seventh component of the rotated PCA model shows a tight clustering of preparations.

Further analysis of the loadings corresponding to the sixth and seventh components reveals that the sixth component is also positively influenced by other flavonoid glycosides (rutin, hyperoside, isoquercetin, quercetrin) and chlorogenic acid. The seventh component only describes the influence of quercetin on the clustering in the positive direction of this component as disclosed by H-5' shown in **Fig. 4** and other resonances of quercetin at $\delta=7.67$ (H-2'), 7.55 (H-6'), 6.42 (H-8), and 6.19 (H-6) (data not shown). A score plot of the sixth and seventh component of the rotated PCA model is shown in **Fig. 4C**. Analysis of this score plot and the corresponding loadings reveals that the clustering of preparations 2, 5, 6, 7, 8, and 16 in the positive direction of the sixth component of the rotated PCA model is due to higher levels of rutin, hyperoside, isoquercetin, quercetrin, quercetin, and chlorogenic acid, whereas the clustering of preparation 3 in the positive direction of the seventh component is due only to higher levels of quercetin as compared with other preparations.

Interpretation of rotated PCA models based on HPLC-PDA profiles of *St. John's wort* extracts

To illustrate the simplified interpretation provided by rotated PCA models, an example using an extremely condensed yet comprehensive data matrix will follow. The data matrix contains relative concentrations of *St. John's wort* plant metabolites derived from PARAFAC analysis of HPLC-PDA profiles. Identification of the plant metabolites represented by the chromatographic peaks was provided by HPLC-PDA-SPE-NMR-MS experiments [10].

Interpretation of the first three components in the derived PCA model has been described in detail in previous work; however, the interpretation of the influence of several plant metabolites involved analysis of several components [10]. The loadings of the derived PCA model were therefore rotated. Score plots of the first four components of the rotated PCA model in association with a loading bar plot are shown in **Fig. 5**. It is apparent that each component explains the influence of individual plant metabolites, and the rotated PCA model facilitates interpretation of the observed clustering.

The first four components of the rotated PCA model describe the influence of guaijaverin, quercetrin, miquelianin, and quercetin 3-O- β -D-(2-O-acetyl)galactoside, respectively (**Fig. 5C**). Thus, the clustering of preparations 14, 17, and 23 in the negative direction of the first component is caused by a higher content of guaijaverin. Preparations 12, 13, 21, and 22, all originating from North America, contain higher levels of quercetrin as compared with other preparations, which cause their separation in the positive direction of the second component (**Fig. 5A**). Higher levels of miquelianin cause the separation of preparations in the positive direction of the third component. In agreement with earlier results [10], higher levels of miquelianin as compared with other preparations influence the clustering of preparations 9 and 23. Preparation 23 displays a more distinct discrimination in the positive direction of the third component (**Fig. 5B**) due to a higher level of miquelianin in this preparation as compared with preparation 9. The influence of quercetin-3-O- β -D-(2-O-acetyl)galactoside on the observed clustering is described in the fourth component. The presence of higher levels of this plant metabolite in preparations 5, 7, 8, 14, and 23

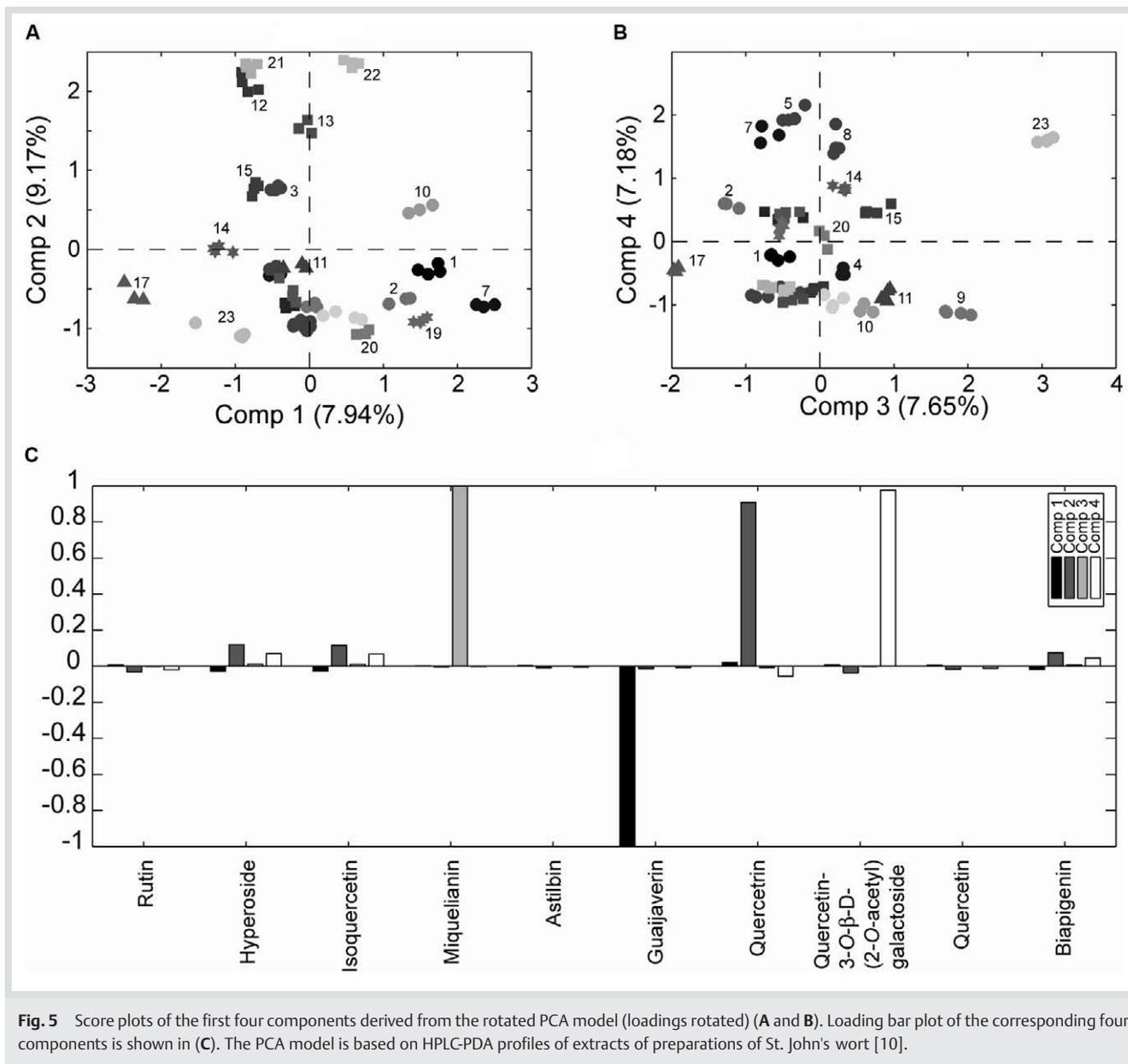


Fig. 5 Score plots of the first four components derived from the rotated PCA model (loadings rotated) (A and B). Loading bar plot of the corresponding four components is shown in (C). The PCA model is based on HPLC-PDA profiles of extracts of preparations of St. John's wort [10].

explains their discrimination from other preparations in the positive direction of the fourth component (● Fig. 5B).

Rotation of loadings thus eases interpretation of the influence of individual plant metabolites on the observed clustering. If, on the other hand, the aim of the study is to gain knowledge about the plant metabolites influencing the clustering of individual samples, rotation of scores can be a valuable tool. To interpret the influence of metabolic profiles on the clustering of individual samples, rotation of scores from the derived PCA model based on the condensed data matrix with relative concentrations has been performed. This aids interpretation of plant metabolites influencing the clustering of individual samples.

Score plots of the first four components of the rotated PCA model (scores rotated) in association with a bar plot of the corresponding loadings are shown in ● Fig. 6. It is apparent that rotation of scores provides discrimination of individual preparations or closely related preparations. Thus, preparation 17 is discriminated in the negative direction of the first component (● Fig. 6A), and therefore interpretation of the loadings corresponding to

this component provides information about plant metabolites influencing the clustering of this preparation. From the loading bar plot it is seen that guaijaverin and biapigenin influence the clustering of preparation 17 (● Fig. 6C), in agreement with earlier results [10]. As opposed to the non-rotated PCA model, which also provided discrimination of preparation 17 in the first component, the loadings derived from rotated PCA model (scores rotated) is not confounded by the influence of other preparations. Interpretation of plant metabolites influencing the clustering of preparation 23 required analysis of several components of the non-rotated PCA model [10]. Rotation of scores provides easier interpretation of plant metabolites influencing the clustering of this preparation by analysis of a single component – the third component of the rotated PCA model (scores rotated) (● Fig. 6B). Thus, plant metabolites influencing the clustering of preparation 23 in the positive direction of the third component are directly seen in the loading bar plot corresponding to this component (● Fig. 6C). This shows that higher levels of miquelianin, guaijaverin, and quercetin 3-O-β-D-(2-O-acetyl)-

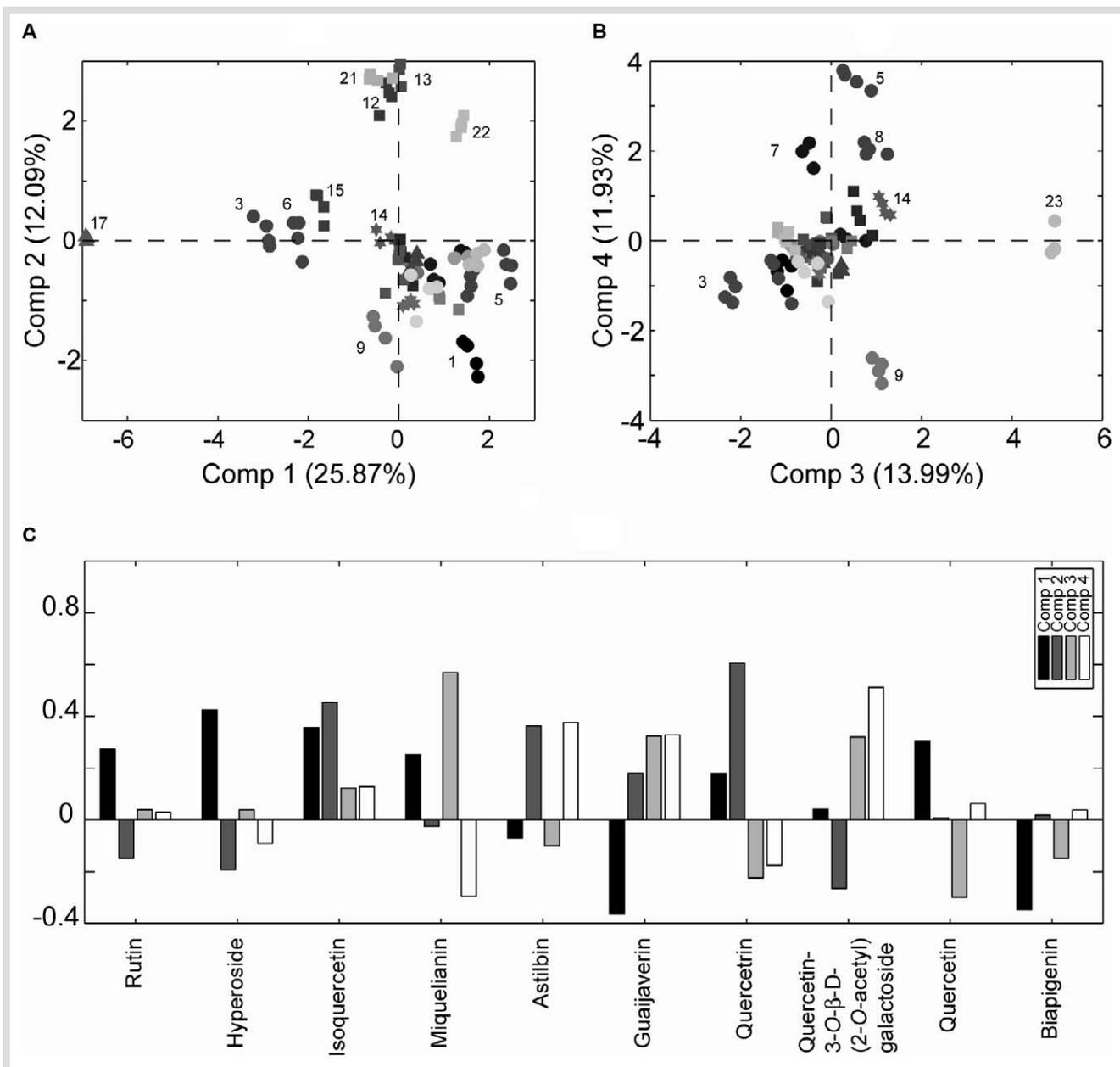


Fig. 6 Score plots of the first four components derived from the rotated PCA model (scores rotated) (A and B). Loading bar plot of the corresponding four components is shown in (C). The PCA model is based on HPLC-PDA profiles of extracts of preparations of St. John's wort [10].

galactoside, and to a minor degree also higher levels of rutin, hyperoside, and isoquercetin, are responsible for the observed clustering of preparation 23, in agreement with earlier results [10].

In conclusion, this study has illustrated the advantages of using rotated PCA models for aiding interpretation of PCA models derived from $^1\text{H-NMR}$ spectra as well as from HPLC-PDA profiles of herbal remedies. Rotation of loadings led to simpler visualizations in terms of interpretation of the influence of individual metabolites on the observed clustering, since the number of components influenced by individual metabolites was reduced as compared to the non-rotated PCA model. For the $^1\text{H-NMR}$ data, only a few components of the rotated PCA model described the influence of quercetin, whereas for the HPLC-PDA data each component of the rotated PCA model described the influence of an individual plant metabolite. Rotation of scores of the PCA

model derived from the HPLC-PDA data set led to conditions, where the whole plant metabolite profiles that are characteristic for individual preparations could be derived from the rotated PCA model. This approach is especially valuable for understanding the clustering of individual preparations or groups of clusters. Rotation of PCA models illustrated in this study is believed to have general applicability in metabonomic, metabolomic, and metabolite profiling studies.

Acknowledgements

▼
Bruker Avance 600 spectrometer used in this work was acquired through a grant from "Apotekerfonden af 1991" (Copenhagen). The technical assistance of Ms. Birgitte Simonsen and Ms. Dorte Brix is gratefully acknowledged.

References

- 1 Martens H, Næs T. Multivariate calibration. New York: Wiley; 1989
- 2 Bailey NJ, Sampson J, Hylands PJ, Nicholson JK, Holmes E. Multi-component metabolic classification of commercial feverfew preparations via high-field ¹H-NMR spectroscopy and chemometrics. *Planta Med* 2002; 68: 734–8
- 3 Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AWM, Verpoorte R. Metabolomic differentiation of *Cannabis sativa* cultivars using ¹H NMR spectroscopy and principal component analysis. *J Nat Prod* 2004; 67: 953–7
- 4 Fr  d  rich M, Choi YH, Angenot L, Harnischfeger G, Lefeber AWM, Verpoorte R. Metabolomic analysis of *Strychnos nux-vomica*, *Strychnos icaja* and *Strychnos ignatii* extracts by ¹H nuclear magnetic resonance spectrometry and multivariate analysis techniques. *Phytochemistry* 2004; 65: 1993–2001
- 5 Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Whitcombe I. Metabolomic strategy for the classification and quality control of phyto-medicine: A case study of chamomile flower (*Matricaria recutita* L.). *Planta Med* 2004; 70: 250–5
- 6 Kim HK, Choi YH, Erkelens C, Lefeber AWM, Verpoorte R. Metabolic fingerprinting of *Ephedra* species using ¹H NMR spectroscopy and principal component analysis. *Chem Pharm Bull* 2005; 53: 105–9
- 7 Holmes E, Tang HR, Wang YL, Seger C. The assessment of plant metabolite profiles by NMR-based methodologies. *Planta Med* 2006; 72: 771–85
- 8 Rasmussen B, Cloarec O, Tang HR, St  rk D, Jaroszewski JW. Multivariate analysis of integrated and full-resolution ¹H-NMR spectral data from complex pharmaceutical preparations: St. John's wort. *Planta Med* 2006; 72: 556–63
- 9 Seger C, Sturm S. Analytical aspects of plant metabolite profiling platforms: Current standings and future aims. *J Proteome Res* 2007; 6: 480–97
- 10 Schmidt B, Jaroszewski JW, Bro R, Witt M, St  rk D. Combining PARAFAC analysis of HPLC-PDA profiles and structural characterization using HPLC-PDA-SPE-NMR-MS experiments: Commercial preparations of St. John's wort. *Anal Chem* 2008; 80: 1978–87
- 11 Kaiser HF. The Varimax criterion for analytic rotation in factor-analysis. *Psychometrika* 1958; 23: 187–200
- 12 Kiers HAL. A comparison of techniques for finding components with simple structure. In: Cuadras CMC, Rao CR, editors. *Multivariate analysis: future directions 2* Amsterdam: Elsevier; 1993: 67–86
- 13 Bundy JG, Sidhu JK, Rana F, Spurgeon DJ, Svendsen C, Wren JF. "Systems toxicology" approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm, *Lumbricus rubellus*. *BMC Biol* 2008; 6: 25
- 14 Rasmussen S, Parsons AJ, Fraser K, Xue H, Newman JA. Metabolic profiles of *Lolium perenne* are differentially affected by nitrogen supply, carbohydrate content, and fungal endophyte infection. *Plant Physiol* 2008; 146: 1440–53
- 15 Soares PK, Bruns RE, Scarminio IS. Statistical mixture design – Varimax factor optimization for selective compound extraction from plant material. *Anal Chim Acta* 2008; 613: 48–55
- 16 Stojanovic K, Jovancicevic B, Vitorovic D, Golovko Y, Pevneva G, Golovko A. Evaluation of saturated and aromatic hydrocarbons oil-oil maturity correlation parameters (SE Pannonian Basin, Serbia). *J Serb Chem Soc* 2007; 72: 1237–54
- 17 Kiers HAL. Simple structure in component analysis techniques for mixtures of qualitative and quantitative variable. *Psychometrika* 1991; 56: 197–212
- 18 Cattell RB. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika* 1944; 9: 267–83
- 19 Malinowski ER. *Factor analysis in chemistry*, 3rd edition. New York: Wiley-Interscience; 2002
- 20 Dien J, Beal DJ, Berg P. Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clin Neurophysiol* 2005; 116: 1808–25
- 21 Kiers HAL. Simplimax – oblique rotation to an optimal target with simple structure. *Psychometrika* 1994; 59: 567–79
- 22 Ferguson GA. The concept of parsimony in factor analysis. *Psychometrika* 1954; 19: 281–90
- 23 ten Berge JMF. Suppressing permutations or rigid planar rotations: A remedy against nonoptimal varimax rotations. *Psychometrika* 1995; 60: 437–46
- 24 Harman HH. *Modern factor analysis*. 3rd edition. Chicago: University of Chicago Press; 1976
- 25 Bilia AR, Bergonzi MC, Mazzi G, Vincieri FF. Analysis of plant complex matrices by use of nuclear magnetic resonance spectroscopy: St. John's wort extract. *J Agric Food Chem* 2001; 49: 2115–24
- 26 European Pharmacopoeia, 5th edition. Strassbourg: Council of Europe; 2005: 2485–6
- 27 United States Pharmacopeia, USP 30: The National Formulary 25. Rockville: The United States Pharmacopoeial Convention; 2007: 978–9

Paper III

Fluorescence Spectroscopy as a Potential Metabonomic Tool for Early Detection of Colorectal Cancer

Reprint from Metabolomics online edition

DOI 10.1007/s11306-011-0310-7

Fluorescence spectroscopy as a potential metabonomic tool for early detection of colorectal cancer

Anders Juul Lawaetz · Rasmus Bro ·
Maja Kamstrup-Nielsen · Ib Jarle Christensen ·
Lars N. Jørgensen · Hans J. Nielsen

Received: 12 January 2011 / Accepted: 6 April 2011
© Springer Science+Business Media, LLC 2011

Abstract Fluorescence spectroscopy Excitation Emission Matrix (EEM) measurements were applied on human blood plasma samples from a case control study on colorectal cancer. Samples were collected before large bowel endoscopy and included patients with colorectal cancer or with adenomas, and from individuals with other non malignant findings or no findings ($N = 308$). The objective of the study was to explore the possibilities for applying fluorescence spectroscopy as a tool for detection of colorectal cancer. Parallel Factor Analysis (PARAFAC) was applied to decompose the fluorescence EEMs into estimates of the underlying fluorophores in the sample. Both the pooled score matrix from PARAFAC, holding the relative concentrations of the derived components, and the raw unfolded spectra were used as basis for discrimination models between cancer and the various controls. Both methods gave test set validated sensitivity and specificity values around 0.75 between cancer and controls, and poor discriminations between the various controls. The PARAFAC solution gave better options for analyzing the

chemical mechanisms behind the discrimination, and revealed a blue shift in tryptophan emission in the cancer patients, a result that supports previous findings. The present findings show how fluorescence spectroscopy and chemometrics can help in cancer diagnostics, and with PARAFAC fluorescence spectroscopy can be a potential metabonomic tool.

Keywords Fluorescence spectroscopy · Colorectal cancer · Chemometrics · PARAFAC · Metabonomics

1 Introduction

The idea of using autofluorescence measurements of blood to discriminate people with cancer from non-cancer was first presented by Leiner, Wolbeis and co-workers in the 1980s. They considered the fluorescence excitation emission matrix (EEM) of a diluted blood serum sample as a base for pattern recognition to monitor the health status of a person. The hypothesis was that, due to the high sensitivity of fluorescence spectroscopy, it would be possible to observe even small deviations in the fluorescence spectrum from “normal” healthy subjects to diseased subjects (Leiner et al. 1983, 1986; Wolfbeis and Leiner 1985). This hypothesis actually fits well into the present theories of metabonomic based diagnostics. Metabonomic based diagnostics explores metabolites in a biological system and its response to a stress situation such as disease. Metabonomics is often based on non-targeted quantitative and qualitative measurements using nuclear magnetic resonance spectroscopy (NMR) or chromatography [liquid (LC) or gas (GC)] combined with mass spectroscopy (MS) (Nordström and Lewensohn 2010; Zhang et al. 2007). In the present study we explore the possibilities for

A. J. Lawaetz (✉) · R. Bro · M. Kamstrup-Nielsen
Department of Food Science, Quality and Technology,
Faculty of Life Sciences, University of Copenhagen,
Rolighedsvej 30, 1958 Frederiksberg C, Denmark
e-mail: ajla@life.ku.dk

I. J. Christensen
Finsen Laboratory, Copenhagen, Denmark

L. N. Jørgensen
Department of Surgery, Bispebjerg Hospital,
University of Copenhagen, Copenhagen, Denmark

H. J. Nielsen
Department of Surgical Gastroenterology, Copenhagen
University Hospital Hvidovre, Hvidovre, Denmark

introducing fluorescence spectroscopy of blood plasma samples as an alternative metabonomic tool for detection of cancer.

Other publications have followed up on the work from Leiner and co-workers or applied other strategies in using autofluorescence on blood to detect cancer (Hubmann et al. 1990; Kalaivani et al. 2008; Leiner et al. 1986; Madhuri et al. 1997, 1999, 2003; Masilamani et al. 2004; Nørgaard et al. 2007; Uppal et al. 2005; Xu et al. 1988). Different approaches have been used; some use extracts or controlled fractions of the plasma, whereas others use the plasma or serum merely diluted or with no sample treatment at all. The studies by Madhuri et al. (1999, 2003) and by Masilamani et al. (2004) use an acetone extract of blood plasma in order to reduce spectral interference in their attempt to measure emission from porphyrins. The results from these studies show elevated levels of porphyrins in cancer patients compared to healthy patients. In the present study we will therefore also have a focus on emission from porphyrins.

Common for almost all of the previous studies was the use of only few or single specific wavelength pairs as opposed to the whole spectral approach combined with chemometrics used in the present study. Only the study from Nørgaard et al. (2007) applied chemometrics in their data analysis, and they got promising results on serum samples from breast cancer patients. The use of chemometrics allows us to use the whole spectrum instead of focusing on single wavelength pairs. Multivariate data analysis/chemometrics is a cornerstone in metabonomics used to extract important information from the complex data output, and hereby hopefully identify specific metabolites with discriminatory or predictive ability (biomarkers) that can be used e.g. for a diagnostic purpose (Ragazzi et al. 2006; Ward et al. 2006). The lack of methods to extract the useful information from the EEMs was exactly a problem for Leiner and co-workers and hence, despite the rather complex EEM measurements, the outcome of their analysis was a simple ratio between two wavelength pairs. In the present study, we apply chemometrics on the fluorescence spectra to discriminate between blood plasma samples from colorectal cancer (CRC) patients and healthy individuals. We apply two different methods of data analysis; one which has been applied previously using the raw spectra as input to the classification model, and one where we extract underlying chemical information from the spectra by Parallel Factor Analysis (PARAFAC) (see materials and methods for a description of PARAFAC). The combination of fluorescence spectroscopy and PARAFAC has not previously been applied in a diagnostic test approach. The combination of PARAFAC and three-way fluorescence data (the EEMs) is especially fruitful, as the parameters of the PARAFAC model can be seen as

estimates of the relative concentrations (scores) and the emission and excitation spectra (loadings) of the fluorophores in the sample (Andersen and Bro 2003; Bro 1997). As for conventional NMR and LC-MS this chemical identification opens for fluorescence spectroscopy as a metabonomic tool.

Fluorescence spectroscopy is widely applied in biomarker research though almost solely in the field of labeled fluorescence, where designed fluorescence probes are used to detect the presence of specific biomarkers (Hamdan 2007). In autofluorescence or intrinsic fluorescence, naturally occurring fluorophores are measured with or without minimal sample preparation (Lakowicz 2006). The number of fluorophores in a blood sample is limited compared to the number of compounds detectable by MS and NMR, though among the fluorophores, biologically important compounds are found. In blood for example, the amino acids tryptophan, tyrosine and phenylalanine and also some cofactors and flavonoids NAP, NAD(P)H, FAD are among the fluorophores (Wolfbeis and Leiner 1985). Compared to MS and NMR, fluorescence spectroscopy is highly sensitive and can thus measure concentrations down to parts per billion (Lakowicz 2006). The fluorescent signal from a fluorophore is dependent on the surrounding environment. For example, tryptophan groups in different proteins or on different positions in the same protein can have different excitation and emission maxima, and can thus be distinguished from each other (Abugo et al. 2000). In fact Leiner et al. (1986) showed a difference in the fluorescence from the amino acid tryptophan in human serum from healthy individuals and patients with gynaecological malignancies.

In the practical data acquisition, fluorescence spectroscopy has some advantages compared to both traditional metabonomic techniques. Sample preparation is limited to a minimum of only diluting the sample, and the time of acquisition can be down to few minutes, depending on the spectral area covered and the integration time. A spectrofluorometer can be small and compact compared to MS and NMR, and the price is often much lower. Compared to standard diagnostic tools such as X-ray, MR and CT scanning, fluorescence spectroscopy is very cheap, but at the present stage not a viable alternative. Compared to targeted methods for single biomarkers based on immunochemical tests the onetime investment in fluorescence spectroscopy is, like in MS and NMR, relatively high, but the running costs are much lower, and fluorescence spectroscopy is faster and easy to use.

Some drawbacks of fluorescence spectroscopy are the instrument dependent results that call for spectral correction before they are globally comparable (DeRose and Resch-Genger 2010). The fluorescence intensity is also highly dependent on the overall absorbance of the sample. At low concentrations of fluorophores (and/or low absorbance), the

linear relation between concentration and intensity known from Lambert-Beers law is also valid in fluorescence spectroscopy. At higher concentrations/high absorbance this relation is broken. This phenomenon is called concentration quenching or the inner filter effect (Lakowicz 2006). Blood plasma is highly absorbent, and thus precautions must be taken to avoid or reduce inner filter effects. In the present study the samples are both diluted and undiluted. For the undiluted samples the pathway of the exciting light is reduced to reduce absorbance.

Colorectal cancer is one of the most frequent malignant diseases for both women and men in the western world. In Denmark in 2008, 4194 cases of CRC were diagnosed, which accounted for more than 12% of all malignant diseases (The Danish Cancer Society 2010; The Danish National Board of Health 2010). The 5-year survival rate of CRC patients is approximately 50%, only ovarian, lung, and pancreas cancers have lower rates (UK, national statistics, 2010). The low rate is primarily due to high recurrence frequencies in some patients undergoing intended curative resection and disseminated disease at the time of diagnosis in other patients. At present fecal occult blood test (FOBT) combined with subsequent colonoscopy in those with positive tests is the method of choice for early detection of colorectal cancer. In recent years national screening programs based on FOBT have been introduced in several countries. The FOBT has been criticized for limited compliance rates, which reduce the advantage of the test, and therefore new, improved screening modalities with high compliance rates are urgently needed (Jenkinson and Steele 2010). The only accepted serum biomarker for CRC is carcinoembryonic antigen (CEA), but with sensitivity and specificity values of 0.34/0.93, this is only accepted for prognosis after detection. Other biomarkers have been suggested with similar or better performance, for example free DNA (Flamini et al. 2006) and plasma lysophosphatidylcholine levels (Zhao et al. 2007). None of these biomarkers have yet been clinically accepted. In search for alternative methods with improved detection rates, and/or better compliance rates in screening for CRC, a metabonomic approach with broad unbiased search for changes in the metabolic profile is a possible solution. Interesting results have been published by Ward et al. (2006) by use of MALDI MS. The present paper will explore whether a solution with fluorescence spectroscopy could be an interesting approach.

2 Materials and methods

2.1 Samples

Human plasma samples (sodium citrate anticoagulant) from 308 individuals were used for the experiment. The

samples are a part of a larger sample set from a multi-centre cross sectional study conducted at six Danish hospitals of patients undergoing large bowel endoscopy due to symptoms associated with CRC (Nielsen et al. 2008). The present sample set is designed as a case control study with one case group (verified CRC) and three different control groups. The three control groups are (1) healthy subjects with no findings at endoscopy, (2) subjects with other, non malignant findings and (3) subjects with pathologically verified adenomas (Lomholt et al. 2009). Each of the groups, case and controls, consisted of samples from 77 individuals. Additional control samples, standardized pooled human citrate plasma, were purchased from 3H-Biomedical AB, Sweden.

2.2 Sample handling and data acquisition

Before measurements, the samples were defrosted on wet ice (0°C) for app. one hour, or until thawed, and each sample was divided in four aliquots of 200 µL to 1 mL for different analytical methods. The divided samples were immediately refrozen at -80°C. The standardized plasma samples were received in 50 mL aliquots, and stored at -80°C. Before use they were thawed at 0°C and divided into aliquots of 300 µL, and refrozen at -80°C. For fluorescence measurements, the samples were defrosted on wet ice (0°C) for app. 40 min.

The samples were measured both undiluted and in a hundred fold dilution in Phosphate Buffered Saline (PBS) (pH 7.4). The diluted samples were prepared immediately after the samples were thawed, and then stored on wet ice (0°C) until measured (app. 20 min). The non diluted fractions of the samples were measured as fast as possible after thawing. Fluorescence spectra were acquired on an FS920 spectrometer (Edinburgh Instruments) with double monochromators and a red sensitive photomultiplier (R928P, Hamamatsu) in a cooled detector house. The EEMs were acquired for the samples using the following settings. Diluted and undiluted samples were measured with excitation from 250 to 450 nm with a 5 nm increment, and emission from 300 to 600 nm with a 1 nm increment. Integration time was 0.05 s. This spectral area consists of light in both the ultra violet and visual area. The ultra violet area is dominated by excitation and emission from the aromatic aminoacids tyrosine and tryptophan hence the fluorescence from proteins. The visual area covers among other things excitation and emission from vitamins and cofactors (for example riboflavin and NAD(P)H) (Wolfbeis and Leiner 1985). In an attempt to capture emission from porphyrins, additional EEMs were acquired from the undiluted samples with excitation wavelengths from 385 to 425 nm with a 5 nm increment and emission wavelengths from 585 to 680 nm with a 1 nm increment, and an

integration time of 0.2 s. Every day a spectrum of the PBS used for dilution was measured with the same settings as the diluted samples. Excitation and emission slit widths were set at 4 nm for all measurements. The fluorescence data were corrected for the wavelength dependent excitation intensity by an internal reference detector in the spectrometer. Likewise the spectra were corrected for instrument dependent emission spectral biases by a correction factor supplied with the instrument. Total time spent for measuring all three EEMs was app. 40 min.

Diluted samples were measured in a 10 × 10 mm quartz cuvette. To reduce inner filter effect in the undiluted samples, these were measured in a 2 × 10 mm quartz cuvette with the 2 mm in the emission direction.

An external cooling system was mounted on the spectrometer keeping the measurement temperature constant at 15°C. To monitor the performance of the fluorescence instrument, a standard plasma sample was measured every day. All spectra were saved as ASCII and exported to Matlab® by an in-house routine. The raw spectra are available for download at <http://www.models.life.ku.dk/>.

2.3 Data analysis

Some samples were discarded due to either obviously erroneous measurements, or too little sample material. From the three different EEMs acquired, the numbers of samples ready for data analysis were then 301, 295 and 300 from low wavelength undiluted, high wavelength undiluted and diluted, respectively. Before the actual data analysis, the data were subjected to certain signal processing steps meant to appropriately handle and minimize the influence from non-relevant artifacts. When measuring fluorescence EEMs, non-chemical phenomena such as Rayleigh scatter and second order fluorescence may be present (Lakowicz 2006). These were removed and replaced with missing data and zeros using in-house software (Andersen and Bro 2003). For the diluted samples, a background spectrum of the solute PBS, measured the same day as the sample, was subtracted from each sample in order to remove possible Raman scatter (McKnight et al. 2001). All samples were intensity calibrated by normalizing to the integrated area of the water Raman peak of a sealed water sample measured each day prior to the measurements. This converts the scale into Raman units and allows comparison of intensity of samples measured on other fluorescence spectrometers (Lawaetz and Stedmon 2009).

A data reduction/decomposition of the fluorescence EEMs to less complex features was performed using the multi-way decomposition method called PARAFAC. A set of fluorescence EEMs can be seen as a three-way data array ($I \times J \times K$), where I is the number of samples measured (objects), J the number of emission wavelengths, and K the

number of excitation wavelengths. Just as PCA is decomposing a two-way data matrix, a three-way data structure can be decomposed by PARAFAC into a number of latent PARAFAC components, by minimizing the sum of squared residuals e in the PARAFAC model (equation below).

$$X_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk}$$

a_{if} is the i th element of the score vector, b_{jf} the j th element of the loading vector of the emission mode and c_{kf} the k th element of the loading vector for the excitation mode, for the f th PARAFAC component. If the correct number of PARAFAC components is used to decompose data with an approximately true trilinear structure and an appropriate signal to noise value, the solution from the PARAFAC model will give estimates of the true underlying profiles of the variables. This makes PARAFAC perfect for fluorescence spectroscopy when applied on EEMs. The loadings and scores can be treated as estimates of the excitation and emission spectra, and relative concentrations of the fluorophores in the samples respectively (Andersen and Bro 2003; Bro 1997).

PARAFAC models were fitted applying nonnegativity constraints on all parameters in the model (Andersen and Bro 2003); hence the estimated parameters were found in such a way that they would not be negative. Models were validated by split-half analysis (Harshman and DeSarbo 1984) combined with trained judgment of the loadings. PARAFAC models were fitted separately to each of the three sets of EEMs. The score matrices from the PARAFAC analyses were pooled to one matrix with 19 variables, which were subjected to further data analysis. PCA was fitted to get a preliminary overview of the data. Classification models were built using PLS-DA, a PLS regression with the pooled PARAFAC scores as independent X variables and a dummy matrix as the dependent Y variable with ones for samples belonging to the class, and zeros for samples not belonging to the class (Wold et al. 2001). Forward selection was applied for variable selection. For all classifications, the data sets were divided into training and test sets (10–30% in test set). The training sets were used for model building, and the test samples were used for validating the models. During model building of the training sets, the models were cross validated with 10% of the samples randomly removed in each segment and averaging over ten repetitions for each cross-validation run. The test sets for subsequent model validation were randomly selected from the data with the same relative number of samples removed from each class.

As an alternative to building classification models on the three combined PARAFAC score matrices, classification was tried directly with the raw spectra as the independent

variables. Variable selection was applied using Interval PLS (iPLS) (Nørgaard et al. 2000). Before the direct classification the three-way array of EEMs were unfolded to a two-way matrix.

All data analyses were performed in Matlab R2010[®] (The Mathworks Inc.) and chemometric analyses were performed in PLS_Toolbox v.5.8.2 (Eigenvector Research, Inc).

3 Results and discussion

Spectra from the three setups are seen in Fig. 1. Comparing the spectra from one undiluted sample and a sample diluted 100 times (leftmost and rightmost spectra respectively in the figure) the effect of dilution is clear. In both the raw undiluted sample and in the diluted, the major peak is in the region with fluorescence from the aromatic amino acids tryptophan and tyrosine (phenylalanine is also among the fluorescing amino acids, but it has excitation/emission maximum outside the measured area). For the undiluted sample there are two distinct peaks in that area, whereas in the diluted sample there is only one distinct peak. Furthermore in the undiluted sample a distinct peak is seen with emission maximum at a higher wavelength. The complex peak structure indicates that it is a mixture of several peaks, which could reflect analytes such as NAD(P)H, FAD, Riboflavin etc. (Wolfbeis and Leiner 1985). This peak structure is not apparently visible in the diluted sample.

It is also worth noticing that the intensity of the diluted sample is higher than the raw. This shows that even though the raw sample is measured in a micro cuvette, it still suffers from inner filter effect. Though it was also observed that the dilution in PBS buffer had an effect besides the reduced inner filter effect, a slight blue shift was observed in emission following excitation at 295 nm in the diluted samples. This might be explained by a slight change in the configuration of the proteins, which can change the emission profile.

The high wavelength area of the undiluted samples was measured separately in order to capture possible fluorescence from porphyrins. In the diluted samples this area gave no signal and was therefore not measured. In Fig. 1, middle plot, the high wavelength area primarily shows the descending tail of a peak with maximum outside the measured area, but a closer inspection of the EEM reveals a little bump at app. 405/610 nm which is in accordance with literature values of porphyrin fluorescence (Madhuri et al. 2003).

In order to monitor the performance of the fluorescence spectrometer, a standard plasma sample was measured every day. The standard deviation among these standard

samples was up to five times lower than the standard deviation for the real samples, indicating good performance of the instrument and consistent sample handling, and at the same time revealing a large biological variation among the real samples.

On each of the three measured areas, a PARAFAC model was fitted. Due to the high complexity of the plasma matrix and the large biological variation in the samples, a large number of PARAFAC components was expected, which makes modelling more challenging. For the undiluted samples in the main spectral area (excitation from 250 to 450 nm, emission from 300 to 600 nm), ten PARAFAC components were chosen. To the spectra from the diluted samples, a model of six PARAFAC components was fitted. Only a reduced area of the spectra from the diluted samples was used, as the highest emission and excitation wavelengths did not contribute positively to the model. To the last selected area, the high wavelength area of the undiluted samples, a three component PARAFAC model was fitted. The number of PARAFAC components reflects the chemical rank of the system. For each component we get a set of loadings and scores, which are estimates of the excitation and emission profiles for the underlying chemical compounds. The excitation and emission loadings for the three models are seen in Fig. 2. Many of the components can be identified chemically but some are more difficult and even impossible to assign to specific chemical analytes. Despite the large number of PARAFAC components it is possible that some of these peaks reflect more than one chemical compound and the non-Gaussian peak shape of some of the loadings supports this.

In case of “just” making a model to discriminate between cancer and non cancer the issue would be to; objectively and in an unsupervised manner reflect the underlying variation, and then chemical assignment is of secondary concern. On the other hand if we at the same time want to gain knowledge about the reason for the discrimination and hereby move fluorescence spectroscopy into the world of metabonomics, chemical identification is an important parameter. A perfect PARAFAC model will give loadings which are estimates of the underlying excitation and emission spectra, and therefore we expected more unambiguous loadings with better options for chemical assignment. The reason for such non-ideal behaviour can be a low signal of some analytes, correlation between different compounds or non-linear behaviour due to quenching and similar phenomena. Given the relatively low number of samples and that some of the samples are not diluted, it is actually impressive that the PARAFAC models come out as chemically interpretable as they do. Still, we anticipate that the interpretability would be possible to improve if many more samples were included in

Fig. 1 Different EEMs recorded on one sample. *Left:* undiluted sample in main spectral area. *Middle:* undiluted sample in high wavelength area (notice the axes are different from the two other). *Right:* sample diluted 100 times in PBS

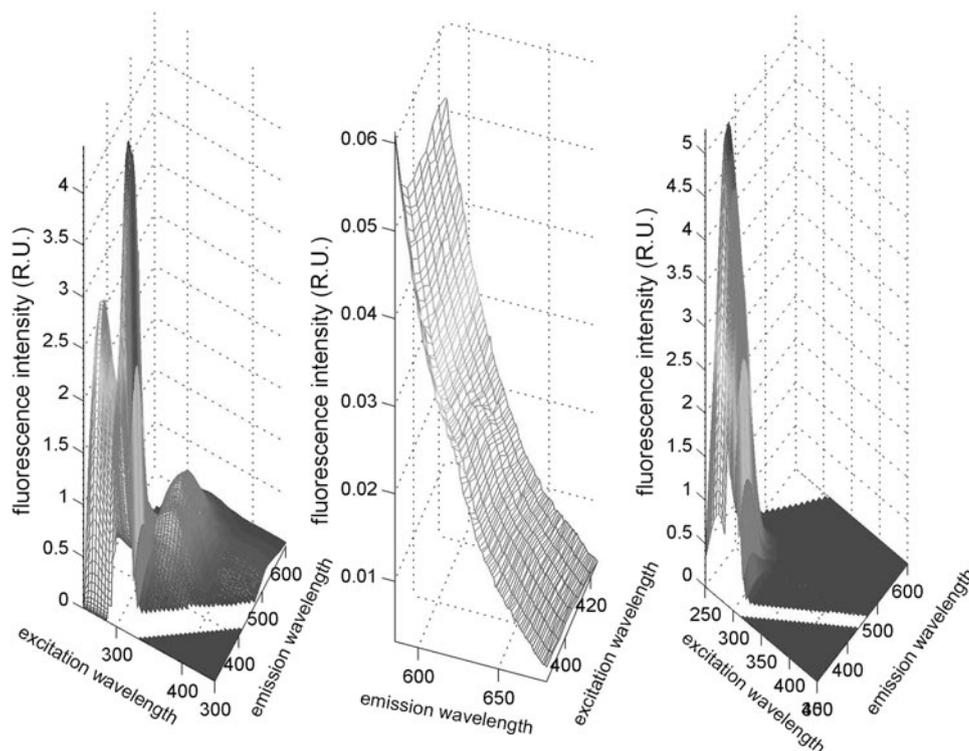
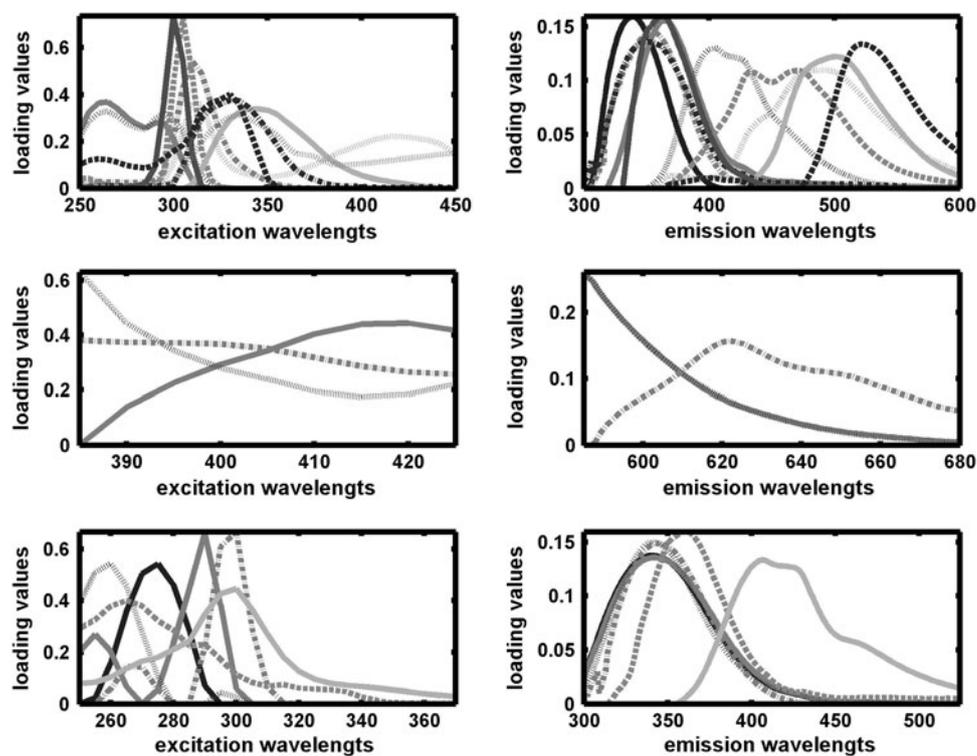


Fig. 2 PARAFAC excitation and emission loadings from the three datasets. *Upper:* undiluted main area. *Middle:* undiluted high wavelength area. *Lower:* diluted main area



the model and possibly also by using targeted standard addition of hypothesized analytes in the modelling phase.

Qualified presumptions on the chemical origin of some of the loadings are made. In both the undiluted and the

diluted samples, several loadings are seen with excitation maximum from 250 to 305 nm, and emission maximum from app. 330 to 350 nm. In this region, fluorescence from protein-bound tryptophan is strong. The emission from

tryptophan can shift when the polarity of the microenvironment changes, hence tryptophan which is bound to different proteins, or at the internal or external parts of a protein, can give rise to different emission maxima. In fact, literature values are reported for tryptophan emissions from 307 to 355 nm (Vivian and Callis 2001). This can explain the numerous peaks for tryptophan emission. Some of the excitation loadings fit well with excitation of tyrosine (app. 265 nm) whereas there is no emission loading supporting the presence of tyrosine emission (app. 300 nm). Energy transfer from excited state tyrosine to tryptophan is a known phenomenon and a reasonable explanation of the absent emission from tyrosine (Lakowicz 2006).

The peaks with maximum at higher wavelengths in both the undiluted and diluted samples can possibly be assigned to compounds such as NAD(P)H, FAD and FMN. In the model from the high wavelength region, it is worth noticing that the little, hardly visible “bump” in the pure spectra gives a clear component with excitation/emission maximum at 400/620 which is in agreement with literature values for porphyrins. There are two other components in this model. One has excitation maximum at 420 nm, but emission maximum outside the measured area, and the other has both excitation and emission maxima outside the measured area. The loadings are in agreement with some of the peaks in the undiluted “main” area (two rightmost peaks in Fig. 2 upper right), and could be tentatively assigned to compounds such as NADH or flavins.

The score matrices from the three PARAFAC models are “pooled” into one common score matrix. This matrix now contains all the quantitative information extracted from the fluorescence measurements. Thus we have reduced the complex spectra with several thousand variables to a matrix with 19 variables consisting of estimated relative concentrations of the underlying chemical compounds of the plasma samples. This matrix is now the input to a classification analysis. Note that absolutely no information about the health status of the patients has been used for building the PARAFAC models. This is important from a validation point of view, as it ensures that the matrix is simply an unbiased representation of the raw data.

3.1 Classification

The combined score matrix is used for building classification models. An initial exploratory PCA analysis of the score matrix explains 52% of the variation in the first three components and needs more than 12 components to explain 95% of the variation. The somewhat low explained variation is most likely due to the biological variation in the data and shows that the 19 PARAFAC scores are not overly redundant. No clear separation of cancer and control samples is found by the PCA analysis. There is thus no

unsupervised direction in the variable space directly separating cancer from controls and hence the major part of the variation in the data is not related to the cancer/non cancer issue at all. Supplementary information such as age, gender, smoking habits, and co-morbidity could not explain further of this variation either. It is most likely just individual differences.

The score matrix with 19 variables was used as input to a PLS-DA classification model. During model building, some samples were removed as outliers based on evaluation of residuals and Hotellings T^2 (Jackson 1991). Classification models were built for all combinations of cancer and control and also control/control. Models are cross validated and the models are tested on a set of samples left out during model building. The huge biological variation from the raw data is still reflected in the extracted 19 variables in the score matrix. Therefore it makes sense to apply variable selection to select those variables of the 19 that reflect the variation relevant for discriminating cancer and non-cancer. We applied forward selection on the calibration data to find the optimal variables for classification. In the different models the number of variables was reduced from 19 variables to between five and 15 variables.

Results from the different models with sensitivity and specificity values for the cross validated and the tested models as well as area under the receiver operating characteristic (ROC) curve are seen in Table 1. A PLS-DA model with all the three control groups pooled to a common control versus the cancer patients gives an area under the ROC curve of 0.69 with optimal sensitivity and specificity values of 0.70 in the cross validated model, and similar values of 0.73 and 0.77 validated on new samples. Similar values are obtained on models with cancer vs. controls from the group of healthy individuals with no findings, and cancer vs. other non malignant findings. These models give areas under the ROC curves of 0.75 and 0.77, and sensitivity and specificity values between 0.73 and 0.80. In the models of cancer vs. adenomas, the area under the curve, sensitivity and specificity values are at the same level as the model with all controls. The results are to some extent surprising as one would expect it to be easier to discriminate between individuals with no findings and cancer, than between individuals with adenomas and cancer. Models of the different controls against each other give poor models with area under the curve values of 0.5–0.6. Even though they have different imbalances (adenomas or other non malignant findings), the controls are thus not much different from a fluorescence point of view. This result is important for future work of building better diagnosis models, as it underlines that the essential differences found in this study are related to cancer, non-cancer. In a different study on the same samples searching for differences in plasma levels of soluble urokinase

Table 1 PLS-DA models for classification of different classes based on the PARAFAC scores

Groups	Sensitivity CV	Specificity CV	AUC CV	Sensitivity predict	Specificity predict
Crc vs. no	0.68	0.84	0.75	0.73	0.77
Crc vs. onf	0.79	0.73	0.76	0.79	0.73
Crc vs. ade	0.73	0.74	0.77	0.92	0.63
Ade vs. no	0.57	0.55	0.50	0.45	0.43
Ade vs. onf	0.47	0.75	0.57	0.47	0.47
Onf vs. no	0.63	0.58	0.59	0.53	0.40
Crc vs. all controls	0.70	0.70	0.69	0.74	0.71

Crc cancer, No no findings, Onf other non malignant findings, Ade adenomas, All all three control groups, CV cross validated

plasminogen activator receptor (suPAR), the level of discrimination between cancer and other non malignant findings was better than between cancer and no findings. The discrimination between cancer and adenomas was less significant in this study (Lomholt et al. 2009).

The sensitivity and specificity values in Table 1 are found as the optimal value (maximizing the sum of the two). In diagnostic models, a high specificity value is often preferred as this reduces the number of false positives. For the models cancer vs. other non malignant findings and cancer vs. no findings we get sensitivity values of 0.48 and 0.43 at specificity values of 0.9. The result achieved by use of fluorescence spectroscopy and PARAFAC is thus comparable to the performance of the known biomarkers for CRC; CEA that has sensitivity and specificity values of 0.34 and 0.93.

The table above shows the results of the different classification models. The different models are based on different data, and thus use different variables for classification. A score and a loading plot for the classification model of cancer vs. other non malignant findings based on the PARAFAC scores are seen in Fig. 3. As expected from the sensitivity and specificity values, there is not a perfect separation between the two classes. However, there is a tendency towards separation along the diagonal from the second to fourth quadrant in the score plot of the first vs. third PLS-DA component. From the loading plot we can see which variables are important for this separation. The loadings are likewise separated along a diagonal, with samples that are positively correlated to the “cancer direction” and samples negatively correlated to the “cancer direction” or positively correlated to the control samples; in this case the samples with other non malignant findings. A similar exercise can be done for all models.

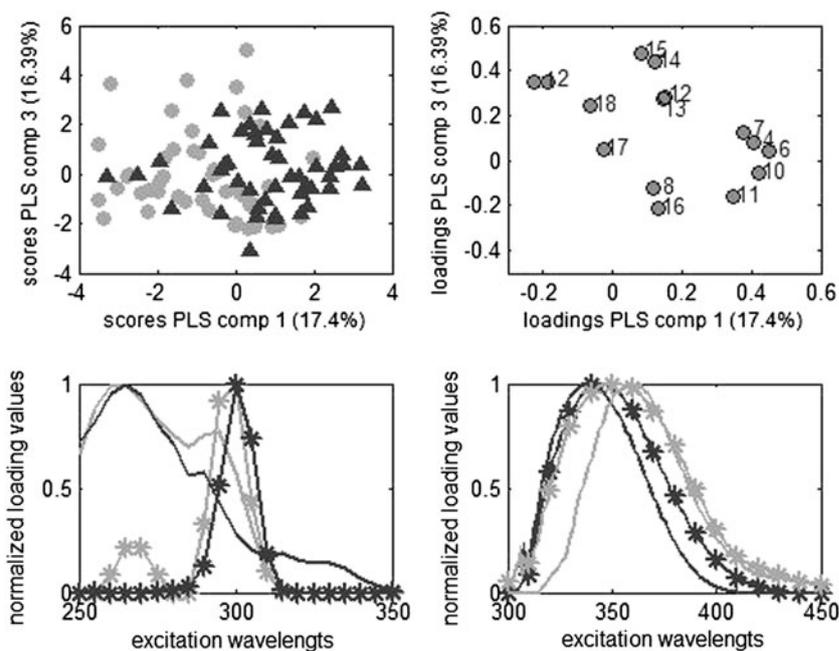
Common for the models with cancer vs. one or all groups of controls is that the variables 1, 2, 8, 16 and 19 for several of the models are negatively correlated to the cancer direction, and likewise variables 6, 7 and 10 are positively correlated to the cancer direction. These variables are thus important in the discrimination between

cancer and controls, though a model based on only those variables does not perform as well as models with more variables. The excitation and emission loadings from components seven and 10 which are positively correlated to cancer and likewise from components eight and 17 which are positively correlated to the controls are shown in Fig. 3 (lower plot). From the excitation and emission loadings these variables can most likely be assigned to tryptophan (variables 7 and 17) or tyrosine, with energy transfer to tryptophan (variables 1 and 4). They have pair wise similar excitation loadings, but the tryptophan emissions in the “cancer variables” are all shifted to shorter wavelengths (blue shift) compared to the “control variables”. This confirms the findings from Leiner et al. (1986) who also experienced a blue shift in tryptophan emission in blood from cancer patients.

As opposed to what was expected, variable 3 (excitation/emission at 400/620), which corresponds to porphyrin, was not correlated to cancer. Several studies have shown elevated porphyrin levels in the blood from cancer patients (Madhuri et al. 2003; Masilamani et al. 2004; Xu et al. 1988). In this study all the subjects were included due to symptoms associated with CRC, and thus, even though three of four do not have cancer, some cellular biochemical imbalance might be expected, and therefore elevated levels could be expected in some of these controls. Additionally, the studies showing porphyrin to be important used acetone extracts of either blood plasma or cells, and not pure blood plasma as in the present study.

In the above models, PARAFAC scores were included from measurements on both diluted and undiluted samples, and as explained earlier there are some important effects of dilution. Fluorescence measurements on the undiluted samples may suffer from inner filter effect due to the high absorbance from the plasma samples. Diluting the samples induce physical/chemical changes in the plasma causing blue shift in the spectra. We found that variables from both the diluted and undiluted measurements were important for detecting cancer. Modelling only on scores from the diluted or undiluted samples gave similar but slightly worse

Fig. 3 Upper left: PLS-DA score plot of the first vs. third PLS-DA component from the model cancer vs. other non malignant findings on PARAFAC loadings. Triangles are cancers and circles are controls. Upper right: corresponding loading plot. Lower: selected PARAFAC excitation (left) and emission (right) loadings. Dark gray line (loading #7) and dark grey with asterisk (loading #10) are correlated with cancer, light gray (#8) and light gray with asterisk (#17) are correlated with control samples



models compared to the combination of scores from the diluted and undiluted samples, thus predictive power is gained by including both. From an analytical point of view, measuring only on the undiluted samples would be preferred as it makes the measurements faster and simpler to perform. Additionally there is a risk that the changes in sample matrix due to dilution could break some of the cancer specific correlations/interactions and thus make discrimination more difficult. A more thorough study addressing this could be interesting. In fact in analysis of the raw spectra (see below) better models were obtained using only the undiluted samples.

3.2 Classification on the raw data

A study similar to this on breast cancer by Nørgaard et al. (2007) applied discrimination only on the raw spectra. The authors did recommend applying more advanced techniques such as PARAFAC on the spectra but did not pursue this. Recall that we have used PARAFAC here, in order to provide more direct chemical information on how a possible classification can come about. Nevertheless, it is interesting to see whether we have gained anything from a quantitative point of view by applying PARAFAC on the data. Hence, classification models were built directly on the raw spectra as well. We have analyzed both diluted and undiluted samples individually and combined, and achieved similar results. However, the results from the undiluted measurements were slightly better than the alternative results, and are thus the only ones presented below. In Table 2 the results from the classifications based on the raw spectra are shown. Compared to the results

based on the PARAFAC scores, these classification models perform equally well and these results are thus also comparable to the performance of CEA. Again the models on control vs. control perform worse than the cancer vs. control models. As for the models based on the PARAFAC scores we have applied variable selection on the models. Different variables are used for the models, but some of the same areas are represented in all four models.

Although it is possible to trace the original wavelengths behind the variables, these do not give the same intuitive information compared to the PARAFAC loadings. The scores and loadings for the model classifying cancer and other non malignant findings (Fig. 4) show a fairly good separation between the two groups in the first and fifth components. The loadings can be traced back to wavelengths around maxima for tryptophan, and the loading for the fifth component has a second derivative-like shape, which can be connected to the shift in the spectra from control to cancer that was shown above in the models based on PARAFAC scores. The results are thus similar, which was expected as it is originally the same data. Still, the extracted features by PARAFAC make the interpretation more straight forward and more comprehensive.

4 Conclusion

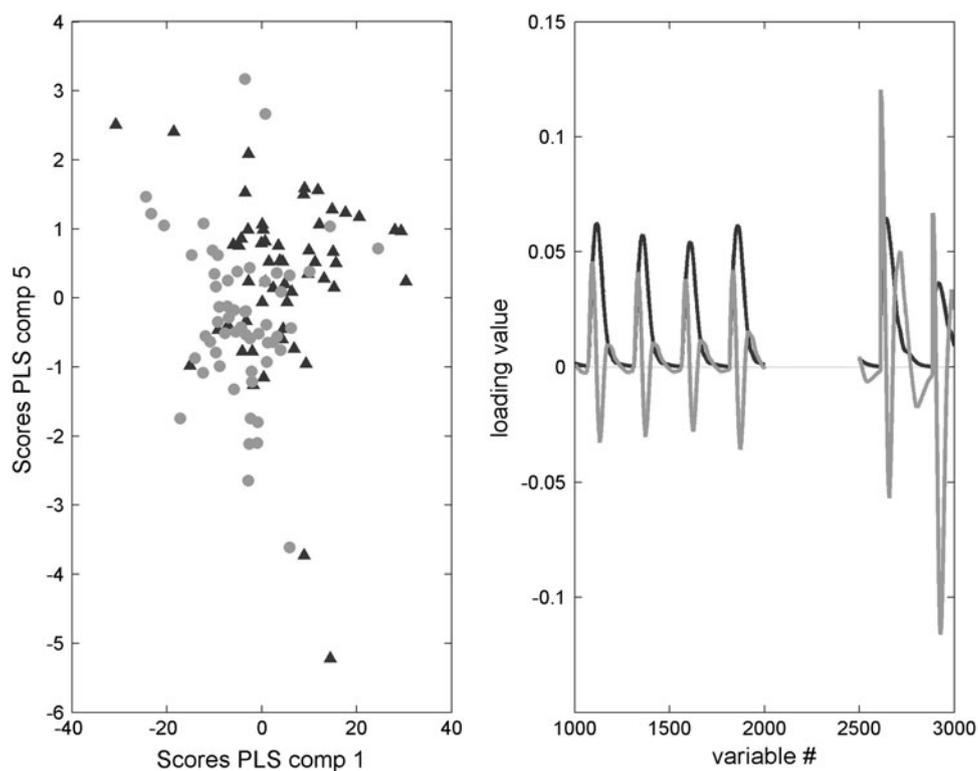
We have introduced excitation emission matrix fluorescence measurements on human blood plasma combined with multivariate data analysis as a potential alternative method to discriminate CRC patients from healthy controls, and controls with other cellular imbalances than cancer. With

Table 2 Results from the PLS-DA on the raw unfolded spectra

Groups	Sensitivity CV	Specificity CV	AUC CV	Sensitivity predict	Specificity predict
Crc vs. no	0.64	0.79	0.73	0.73	0.67
Crc vs. onf	0.73	0.79	0.75	0.73	0.73
Crc vs. ade	0.78	0.71	0.74	0.64	0.87
Ade vs. no	0.68	0.61	0.63	0.33	0.63
Ade vs. onf	0.84	0.34	0.55	0.70	0.33
Onf vs. no	0.45	0.82	0.62	0.20	0.82
Crc vs. all controls	0.69	0.7	0.73	0.67	0.83

Crc cancer, No no findings, Onf other non malignant findings, Ade adenomas, All all three control groups

Fig. 4 Left: score plot of the first component vs. the fifth component for the PLS-DA model on cancer (triangles) vs. other non malignant findings (circles) on the raw spectra. Right: loadings from the first component (dark gray) and the fifth component (light gray)



sensitivity and specificity values of app 0.75 on a test set, the results are comparable to the known biomarker CEA. Previous studies with fluorescence spectroscopy have obtained similar results on other types of cancer but with a smaller number of samples. We obtained similar results in regards to discrimination whether we applied classification directly on the raw unfolded spectra or extracted estimates of the underlying fluorophores by use of PARAFAC. By the latter method, however, we obtained better conditions for a chemical interpretation/understanding of the results. We could see a blue shift in the tryptophan emission from cancer patients as one of the reasons for discrimination, a phenomenon described earlier in the literature. The use of PARAFAC on the fluorescence data to extract qualitative and quantitative chemical information from the human

blood plasma samples, and base classification on this information is an example on how fluorescence spectroscopy can be used as a tool for metabonomic research. Compared to biomarker tests, fluorescence spectroscopy is an inexpensive alternative, and with minor sample preparation it is easy to perform the analysis. Further research is needed but we believe that there is room for fluorescence spectroscopy as metabonomic tool in cancer research.

Acknowledgments The VILLUM FOUNDATION is thanked for funding Anders Juul Lawaetz. Abdelrhani Mourhib is thanked for his laboratory assistance. Knud Nielsen, Randers Hospital, Søren Laurberg, Aarhus Hospital, Jesper Olsen, Glostrup Hospital and Hans B Rahr, Odense Hospital, are acknowledged for their contribution to the original protocol.

References

- Abugo, O. O., Nair, R., & Lakowicz, J. R. (2000). Fluorescence properties of rhodamine 800 in whole blood and plasma. *Analytical Biochemistry*, *279*, 142–150.
- Andersen, C. M., & Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data 1. *Journal of Chemometrics*, *17*, 200–215.
- Bro, R. (1997). PARAFAC. Tutorial and applications 1. *Chemometrics and Intelligent Laboratory Systems*, *38*, 149–171.
- DeRose, P. C., & Resch-Genger, U. (2010). Recommendations for fluorescence instrument qualification: The new ASTM standard guide. *Analytical Chemistry*, *82*, 2129–2133.
- Flamini, E., Mercatali, L., Nanni, O., Calistri, D., Nunziatini, R., Zoli, W., et al. (2006). Free DNA and carcinoembryonic antigen serum levels: An important combination for diagnosis of colorectal cancer. *Clinical Cancer Research*, *12*, 6985–6988.
- Hamdan, M. H. (2007). *Cancer biomarkers*. Hoboken: John Wiley and sons.
- Harshman, R. A., & DeSarbo, W. S. (1984). An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H. G. Law, et al. (Eds.), *Research methods for multimode data analysis* (pp. 602–642). New York: Praeger.
- Hubmann, M. R., Leiner, M. J. P., & Schaur, R. J. (1990). Ultraviolet fluorescence of human sera. I. Sources of characteristic differences in the ultraviolet fluorescence-spectra of sera from normal and cancer-bearing humans 1. *Clinical Chemistry*, *36*, 1880–1883.
- Jackson, J. E. (1991). *Operations with group data*. Hoboken: John Wiley & Sons, Inc.
- Jenkinson, F., & Steele, R. J. C. (2010). Colorectal cancer screening—methodology. *Surgeon-Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*, *8*, 164–171.
- Kalaivani, R., Masilamani, V., Sivaji, K., Elangovan, M., Selvaraj, V., Balamurugan, S. G., et al. (2008). Fluorescence spectra of blood components for breast cancer diagnosis. *Photomedicine and Laser Surgery*, *26*, 251–256.
- Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*. New York: Springer.
- Lawaetz, A. J., & Stedmon, C. A. (2009). Fluorescence intensity calibration using the Raman scatter peak of water. *Applied Spectroscopy*, *63*, 936–940.
- Leiner, M. J., Schaur, R. J., Desoye, G., & Wolfbeis, O. S. (1986). Fluorescence topography in biology. III: Characteristic deviations of tryptophan fluorescence in sera of patients with gynecological tumors. *Clinical Chemistry*, *32*, 1974–1978.
- Leiner, M., Schaur, R. J., Wolfbeis, O. S., & Tillian, H. M. (1983). Fluorescence topography in biology. 2. Visible fluorescence topograms of rat sera and cluster-analysis of fluorescence parameters of sera of Yoshida ascites hepatoma-bearing rats. *IRCS Medical Science-Biochemistry*, *11*, 841–842.
- Lomholt, A. F., Hoyer-Hansen, G., Nielsen, H. J., & Christensen, I. J. (2009). Intact and cleaved forms of the urokinase receptor enhance discrimination of cancer from non-malignant conditions in patients presenting with symptoms related to colorectal cancer. *British Journal of Cancer*, *101*, 992–997.
- Madhuri, S., Aruna, P., Summiya Bibi, M. I., Gowri, V. S., Koteeswaran, D., Schaur, R. J., et al. (1997). Ultraviolet fluorescence spectroscopy of blood plasma in the discrimination of cancer from normal. *Proceedings of SPIE*, *2982*, 41–45.
- Madhuri, S., Suchitra, S., Aruna, P., Srinivasan, T. G., & Ganesan, S. (1999). Native fluorescence characteristics of blood plasma of normal and liver diseased subjects. *Medical Science Research*, *27*, 635–639.
- Madhuri, S., Vengadesan, N., Aruna, P., Koteeswaran, D., Venkatesan, P., & Ganesan, S. (2003). Native fluorescence spectroscopy of blood plasma in the characterization of oral malignancy. *Photochemistry and Photobiology*, *78*, 197–204.
- Masilamani, V., Al-Zhrani, K., Al-Salhi, M., Al-Diab, A., & Al-Ageily, M. (2004). Cancer diagnosis by autofluorescence of blood components. *Journal of Luminescence*, *109*, 143–154.
- McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T., & Andersen, D. T. (2001). Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography*, *46*, 38–48.
- Nielsen, H. J., Brunner, N., Frederiksen, C., Lomholt, A. F., King, D., Jorgensen, L. N., et al. (2008). Plasma tissue inhibitor of metalloproteinases-1 (TIMP-1): a novel biological marker in the detection of primary colorectal cancer. Protocol outlines of the Danish-Australian endoscopy study group on colorectal cancer detection. *Scandinavian Journal of Gastroenterology*, *43*, 242–248.
- Nordström, A., & Lewensohn, R. (2010). Metabolomics: Moving to the Clinic. *Journal of Neuroimmune Pharmacology*, *5*, 4–17.
- Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy*, *54*, 413–419.
- Nørgaard, L., Soletormos, G., Harrit, N., Albrechtsen, M., Olsen, O., Nielsen, D., et al. (2007). Fluorescence spectroscopy and chemometrics for classification of breast cancer samples—a feasibility study using extended canonical variates analysis. *Journal of Chemometrics*, *21*, 451–458.
- Ragazzi, E., Pucciarelli, S., Seraglia, R., Molin, L., Agostini, M., Lise, M., et al. (2006). Multivariate analysis approach to the plasma protein profile of patients with advanced colorectal cancer. *Journal of Mass Spectrometry*, *41*, 1546–1553.
- The Danish Cancer Society. (2010). <http://www.cancer.dk>.
- The Danish National Board of Health. (2010). *National screening for tyk- og endetarmskræft*. The Danish National Board of Health, 1.
- Uppal, A., Ghosh, N., Datta, A., & Gupta, P. K. (2005). Fluorimetric estimation of the concentration of NADH from human blood samples 1. *Biotechnology and Applied Biochemistry*, *41*, 43–47.
- Vivian, J. T., & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal*, *80*, 2093–2109.
- Ward, D. G., Suggett, N., Cheng, Y., Wei, W., Johnson, H., Billingham, L. J., et al. (2006). Identification of serum biomarkers for colon cancer by proteomic analysis. *British Journal of Cancer*, *94*, 1898–1905.
- Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130.
- Wolfbeis, O. S., & Leiner, M. (1985). Mapping of the total fluorescence of human-blood serum as a new method for its characterization. *Analytica Chimica Acta*, *167*, 203–215.
- Xu, X. R., Meng, J. W., Hou, S. G., Ma, H. P., & Wang, D. S. (1988). The characteristic fluorescence of the serum of cancer-patients. *Journal of Luminescence*, *40–1*, 219–220.
- Zhang, Xuewu, Li, Lin, Wei, Dong, Yap, Yeeleng, & Chen, Feng. (2007). Moving cancer diagnostics from bench to bedside. *Trends in Biotechnology*, *25*, 166–173.
- Zhao, Z., Xiao, Y., Elson, P., Tan, H., Plummer, S. J., Berk, M., et al. (2007). Plasma Lysophosphatidylcholine Levels: Potential Biomarkers for Colorectal Cancer. *Journal of Clinical Oncology*, *25*, 2696–2701.