Metabolic profiling of volatile organic compounds and enhanced vibrational spectroscopy

A thesis submitted to The University of Manchester
for the degree of Doctor of Philosophy in the
Faculty of Engineering and Physical Sciences

William Cheung

June 2010

Table of Contents

Table of Figures

Abstract

The University of Manchester, William Cheung.

A thesis submitted to The University of Manchester for the degree of **Doctor of Philosophy (PhD)** in the Faculty of Engineering and Physical Sciences.

Metabolic profiling of volatile organic compounds and enhanced vibrational spectroscopy

Nov 2010

Metabolomics is a post genomic field of research concerned with the study of low molecular weight compounds within a biological system permitting the investigation of the metabolite differences between natural and perturbed systems (such as cells, organs and tissues). Rapid identification and discrimination of biological samples based upon metabolic differences and physiological status in microbiology, mammalian systems (particularly for disease diagnosis), plants and food science is highly desirable. Volatile organic compound (VOC) profiling is a novel area of research where the composition of the VOCs emitted by the biological samples can be correlated to its origin and physiological status. The aim of this project was to investigate the applicability of VOC profiling as a potential complementary tool within metabolomics.

In this project the discrimination of bacteria using a novel gas phase separation method was investigated and the development of VOC-based profiling tools for the collections of VOCs emitted from biological samples was also studied. The optimisation and validation of a high throughput method for VOC analysis was achieved and this was used to assess wound healing.

VOC metabolite profiling was further extended to the discrimination of *S. typhimurium* contaminated meat; the study was conducted in parallel with metabolite profiling analysis for the analysis of non-volatile small molecules. Finally, enhanced vibrational spectroscopic techniques were applied to the characterisation and screening of dye molecules in contaminated foodstuffs using Raman spectroscopy.

This thesis clearly demonstrates that VOC metabolic profiling is a complementary tool within the metabolomics toolbox, one of its great attractions is that it permits the characterisation of biological samples in a rapid and non-invasive manner. The technique provides detailed chemical information regarding the VOC composition present above the headspace of the sample and can be used to understand its physiological status and biological origin. VOCs metabolite profiling will become a valuable tool for non-invasive analysis of many biological systems. Raman spectroscopy is a sensitive and non-destructive technique which can generate detailed chemical and structural information regarding the analyte under investigation with little or no sample preparation needed. The effect of the weak Raman signal can be significantly amplified by coupling the analyte molecule to surfaces of nanoparticles and demonstrated that it is ideal for analysing aqueous dye solutions in a quantitative manner.

Copyright statement

Declaration

I declare that no portion of this work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Acknowledgments

I wish to express my sincerest gratitude and appreciation for working with Roy Goodacre and Yun Xu, for their guidance, support and supervision of my research over the past four years. Thank you for giving me the freedom to do my research, you have always been a good supervisor to me chief and I am truly grateful for that.

Yun, it is truly amazing how you can make something so insanely complicated look so simple and effortless.

I am indebted to Samrina Rehman, Dave Ellis and Sam Mabbott for their help and support, writing was never really my strong suit and I do apologises for the mental trauma I put everyone through.

Also I would like to thank Richard Snook, Ian Fleet, Stephen Wong, Paul Begley, Svetlana Riszanskia, Eisa Faria, Ewan Blanch, Alexis Thomas, Iqbal Shadi, Claudia Martins, Daniel Kirton, Sing Ng, William Allwood, Phil Roche, Alfonso Riley-Morales, John Duncan, Mitch Jacobs, Konesh Masilamany, Dong Hyun Kim, John Hildyard, Pedro Juane, Andrew Ballantyne, Shane Ratcliffe, Gavin Potter, Richard O'Connor and Juan Guerrero Leal for their unique insights, discussions and opinions on my "unconventional" ideas and experiments in the MIB and Faraday tower.

Liz Devonport and Eric Warburton for their constant help and laboratory support during my time in the Morton lab, I have lost count how times I have broken and fixed that Varian ion trap MS.

Dedication

媽媽和爸爸，我終於完成我的博士！
另外，我很抱歉，我年紀越大時，我要進一步逃跑。

Mum and dad, I finally finished my PhD! I am sorry that the older I get, the further I have to fly away.

List of Abbreviations

α	=	Selectivity factor
A	=	Multipath constant
ALS	=	Asymmetric least squares
ANN	=	Artificial neural networks
B	=	Longitudinal diffusion
BVOC	=	Biogenic volatile organic compound
C	=	Mass transfer co-efficient
CID	=	Collision induced dissociation
COW	=	Correlated optimised warping
DHS	=	Dynamic Headspace sampling
DMS	=	Differential mobility spectrometry
DNA	=	Deoxyribonucleic acid
ECD	=	Electron capture detectors
EI	=	Electron impact
FID	=	Flame ionisational detectors
FT-ICR	=	Fourier Transform Ion Cyclotron Resonance
FTIR	=	Fourier Transform Infrared Spectroscopy
GC/ITMS	=	Gas chromatography ion trap mass spectrometry
GC/MS	=	Gas chromatography mass spectrometry
GC/TOF-MS	=	Gas chromatography time of flight mass spectrometry
GC	=	Gas chromatography
HETP	=	Height equivalent to theoretical plates
HS	=	Headspace
SHS	=	Static Headspace sampling
QIT	=	Quadrupole ion trap
K'	=	Capacity factor
K	=	Partition co-efficient
LC/MS	=	Liquid chromatography mass spectrometry
LC	=	Liquid chromatography
LLE	=	Liquid-liquid extraction
mRNA	=	Messenger ribonucleic acid
MS	=	Mass spectrometry
N	=	Number of theoretical plates
NMR	=	Nuclear magnetic resonance

NPD = Nitrogen phosphor detectors PARAFAC=Parallel factor analysis

PCA = Principal component analysis
PDMS = Polydimethylsilioxane
PLS = Partial least square
PLS-DA = Partial least square discriminate analysis
Py = Pyrolysis
PY-GC/DMS = Pyrolysis gas chromatograph differential mobility spectrometry

QMF = Quadrupole mass filter

RS = Raman spectroscopy
Rs = Resolution
SD = Stream distillation
SDE = Simultaneous distillation extraction
SERRS = Surface enhance resonance Raman spectroscopy
SERS = Surface enhanced Raman spectroscopy
SFE = Supercritical fluid extraction
SPME = Solid phase microextraction
SVM = Support vector machine

TD = Thermal desorption
TOF = Time of flight
TQ = Triple quadrupole
TVC = Total variable counts

u = Average velocity of the mobile phase

VOC = Volatile organic compound

Chapter 1

1.1 Introduction

The classical view of molecular biology is based upon a hypothetico-reductionist approach, where the emphasis is placed on understanding individual key components of a system. The flow of information within a cell has been assumed to be unidirectional where, the DNA encodes mRNA, which in turn encodes several proteins some of which are enzymes, and will catalyses intermediates involved in metabolic reactions (metabolites), this paradigm is summarised below.

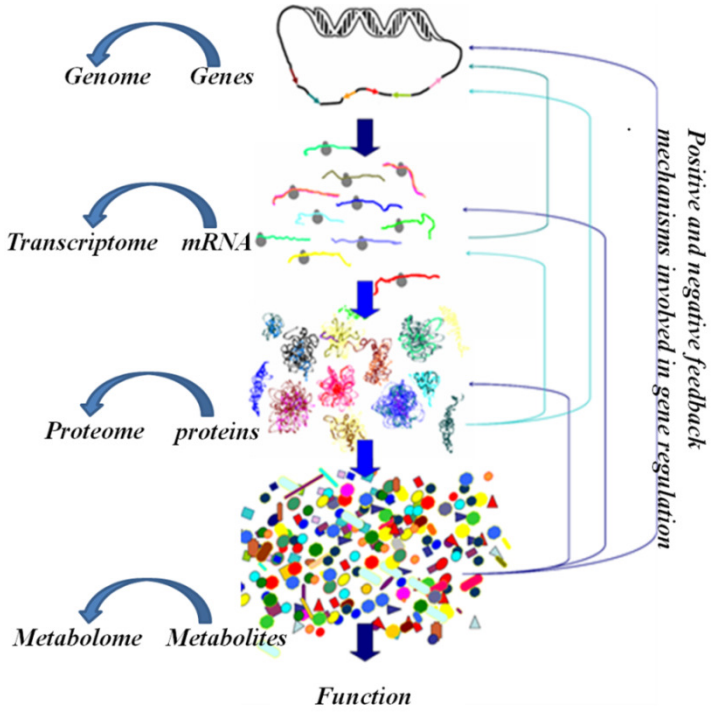


Figure 1 The flow of information from genes to function, the downward arrows showing the direction of influence from one component to the next, the upwards arrows are positive and negative feedback mechanisms in gene regulation.

Since the completion of the human genome project in 2003, there has been a paradigm shift in which we view modern biology, moving away from a traditional unidirectional approach and towards a more holistic view, in which greater importance on understanding the system (the cell, tissue or an organism as a whole) and how these different components e.g. the genome (*the complete sequence of the chemical base pairs of an organism*), transcriptome (*total set of mRNA produced within cell or tissue)*, proteome *(the entire complement of protein expressed by the cell, tissue or organism)* and metabolome (*the sum of the metabolites constituents of*

an organism) interact in relation to each other to give rise to function. The metabolome represents the final downstream product of the genome, both gene expression and protein modification are subject to complex homeostatic control and feedback mechanisms and the net results of these processes are alterations within the metabolome. This is summarized in Figure 1 by the upward arrows. Metabolites are functional entities and they have clear roles in the operation of a biological system, reflecting its surrounding environment (Ryan and Robards 2006). The concept that various disease states can be reflect/result in the changes of the metabolite concentration is central to metabolomics (van der Greef 2004; van der Greef 2007; Nicholson and Lindon 2008).

Metabolomics is the study of the metabolome; the total quantitative collection of low molecular weight compounds (metabolites), present within a cell or organism under set conditions, which are involved in metabolic reactions required for growth, maintenance and normal function (Oliver 1998). Metabolites are generally labile species and have immense chemical diversity existing in a highly dynamic state ranging from picoMolar (10^{-12}) to milliMolar (10^{-3}) concentrations, with a turnover rate for some pathways (e.g. glycolysis) in the order of milliseconds (Weckwerth and Morgenthal 2005; Dunn and Ellis 2005b).

Metabolomics is a hypothesis driven science in which, carefully designed experiments are conducted to generate complex chemical data in conjunction, with inductive reasoning (supervised analysis) to derive a hypothesis which in turn can be tested and validated (Goodacre 2004; Kell and Oliver 2004). The diversity of function and structure of metabolites is greater than that of proteins or genes at present. Many organisms have fewer metabolites than genes or proteins, (e.g. the single cell eukaryote *Saccharomyces cerevisiae* has ~6000 protein-encoding genes with only ~600 low molecular intermediates (Rammsdonk 2001)) and therefore profiling the metabolome might increase the likelihood of identifying and tracking meaningful biochemical changes in relation to external perturbations (van Ravenzwaay 2007). Metabolomics is complementary to transcriptomics and proteomics and reflects more closely to the phenotype as changes are amplified relative to the transcriptome and proteome.

The metabolome is in a state of constant flux and is influenced by a magnitude of intrinsic and extrinsic factors such as health status, age, metabolic rate and nutrient intake. The net results of these perturbations can be reflected within the metabolome (Goodacre 2007).

Metabolomics had been successfully applied to numerous different fields since its inception from the investigation and characterisation of microorganisms (Koek 2006; Khoo and Al-Rubea 2007) to plants, (Kopka 2004; Weckwerth 2005), food and nutrition (Wishart 2008) health and diagnostic purposes (van der Greef 2004; German 2005; Kell 2006; Dieterle 2007; Oresic 2009) and the evaluation of drug effectiveness and its mode of action (Lindon 2006; van der Greef 2007; van Ravenzwaay 2007).

1.1.2 Analytical strategies in Metabolomics

There is numerous analysis methodologies employed in metabolomics. Below are a set of terms and definitions commonly used (Dunn 2008).

(1) **Metabolomics**: non-discriminate quantification and identification of all metabolites present within a biological system.

(2) **Metabolic profiling**: detection of a wide range of metabolites related by classes (such as amino acids, sugars, alcohols, organic acids, vitamins) by employing a single or multiple analytical platforms to obtain maximal coverage of the metabolome. Metabolic profiling can be considered as an untargeted analysis approach, with the emphasis on detecting and quantifying as many metabolites as possible. The relative changes of the metabolite concentrations are used to define any metabolic differences between the samples.

(3) **Metabolic fingerprinting**: rapid and high-throughput analysis of crude sample mixtures with minimal sample preparation. Identification and quantification is limited in this strategy as greater emphasis is placed on discrimination of samples, based on biological origin or physiological status.

(4) **Metabolic footprinting**: analysis of extracellular metabolome, any metabolites secreted from the intracellular complement of the organism into its extracellular growth medium. Sampling and analysis is rapid since

metabolite quenching and extraction is not required. This strategy is commonly applied to the study of microorganisms and cell culture systems.

(5) **Targeted metabolite analysis**: this involves the absolute quantification and identification of a single or set of related metabolites after extensive sample preparation and extraction, prior to analysis by one or more analytical platforms.

In addition to the many analysis methodologies there are also a diverse range of analytical platforms employed in metabolomics, the most predominate analytical platforms include gas chromatography mass spectrometry (GC/MS) (Fiehn 2002), liquid chromatography mass spectrometry (LC/MS), Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy nuclear magnetic resonance spectroscopy (NMR) (Lindon 2003). Each analytical platform has its own advantages and disadvantage towards the detection of different classes of metabolites. Of these analytical platforms Ultra high performance liquid chromatography mass spectrometry (UPLC-MS) has the greatest potential for near-global profiling (Plumb 2005, Theodoridis 2008.) A detailed comprehensive review of the analytical platforms employed in metabolomics is given elsewhere (Dunn 2005a; Dunn and Ellis 2005b).

Due to the complexity, dynamic range and size of the metabolome complete global quantification is extremely challenging and currently unachievable (Dunn 2008). At present there is no single extraction methodology or analytical platform capable of targeting and detecting every single metabolite (Hall 2006). To circumvent this problem multiplex targeted extractions (different extraction methods applied to the same sets of samples) combined with parallel sample analysis on different analytical platforms are employed in an attempt to profile a biological system at the global level with minimal bias. This multiplex approach reduces the effects of individual extraction and instrumental bias towards certain classes of compounds and allows a more comprehensive range of possible metabolites to be detected (maximal metabolite coverage).

1.1.3 Technological innovations in MS-based Platforms

Recent technological innovations in MS-based analytical platforms have led to exponential increases in performance: the introduction of Orbitraps (OT) and Fourier transform ion cyclotron resonance (FT-ICR) (Aharoni 2002, Marshall and Hendrickson 2002) mass analyzers, have enabled a significant step up in sensitivity, selectivity, resolution and hence accuracy. Faster and smaller micro electric devices have allowed MS analyzers to achieve near total 3D ion confinement and ultra fine ion control. This in combination with ultra fast scan rate in modern analyzers in the order of sub-millisecond has enabled a greater number of total metabolites to be simultaneously, accurately and reliably detected. The higher scan rate allows much a better peak shape to be obtained for even ultra low level metabolites, resulting in a more accurate peak deconvolution and subsequent identification.
The use of collision-induced disassociation (Johnson 1988) and chemical fragmentation (Munson1966; Harrison 1992) has also allowed a greater range of structural information to obtained relating to unknown compounds facilitating their identification.

The use of robotic automation in conjunction with shortened sample analysis time (due to improved chromatographic separation techniques) has enable significantly greater throughput, to the point where extremely large amounts of samples can be analysed accurately and reproducibly in a relatively short space of time with minimal supervision from the operator. Currently MS based instruments provide the most reliable and accurate way of identifying and quantifying low concentration metabolites in complex sample matrices (Weckwerth 2005).

1.1.4 Metabolic flux analysis (MFA)

Another important area of metabolomics is based around metabolic flux analysis (MFA), in which the rate of transport of substrates through multiple connected biochemical pathways at steady state is monitored and measured. This is achieved by the use of isotopic labeling experiments and measuring these distributions using either nuclear magnetic resonance (NMR) or MS. The technique involves monitoring changes in the substrate and product concentration or by analysing the flow and redistribution of labeled precursor material into other metabolites (Schwender 2004).

This technique has been successfully applied to plants (Boatright 2004; Heinzle 2007; Matsuda 2007), mammalian cells (Munger 2008) and bacteria (Nöh and Wiechert 2006; Antoniewicz 2007; Nöh 2007). MFA is essential for further understanding of control and regulatory mechanisms within metabolic networks (Fell 1997). Metabolic fluxes are regulated by both gene expression and post transcriptional and post translational modifications, and while such processes may results in only minor changes in enzyme concentrations, these changes can and do have a significant impact upon concentration of a variety of metabolites.

Within the post genomic era of research, despite the advent of transcriptomics, proteomics and metabolomics we have yet to increase our understanding of the metabolic network within a biological system to the point where rational metabolic engineering is a viable option. However MFA has shown to be a valuable tool for the elucidation of metabolic pathways, enhancing our ability to accurately predict the impact of gene knock-outs on mutant flux network and facilitating the development of accurate mechanistic pathway models (Kruger and Ratcliff 2009). As such, MFA may well prove to be a useful complementary tool to metabolomics for the global characterisation of metabolic networks (Cornish-Bowden and Cardenas 2000).

1.2 Methods and materials

1.2.1 Biogenic volatile organic compounds (BVOC)

The traditional definition of volatile organic compounds (VOCs) is extremely broad and misleading due to large overlaps between distinct classes of compounds. Any low molecular weight compound (with a mass less than 400 amu) with a carbon backbone or ring and high vapour pressure, (greater than 0.27 KPa) with a boiling point between 50-260^{o}C and readily exists and as a gas under standard temperature and pressure can be classed as a VOCs (WHO 1989; EPA 1999). This definition makes no distinction between any biological and artificial form of VOCs covering everything from organic solvents such as acetone and dichloromethane, to polyaromatic hydrocarbons (PAH), environmental pollulates such as BTEX (benzene, toluene, ethylbenzene and xylene). Methane from decomposing organic matter through to perfumes, sense and flavour molecules in fruits, food, plants,

bacteria to chemical warfare agents such as phosgene (choking agent) and cyclosarin (nerve agents). Biogenic volatile organic compounds (BVOCs) are a relatively novel term that describes any VOCs which are synthesised by living organisms. The production and emission are dictated by abiotic factors such as light, temperature and nutrient level. In addition to this, the release of such BVOCs is induced as a stress response in plants due to wounding or feeding by insects and herbivores (Arneth and Niinements 2010). This may also act as a defence mechanism for plants reacting to climatic changes (Penuelas and Llusia 2003).

BVOC plays an important role in plant to insect communication (Du 2001; Inui 2007) and also in the mating of insects (Landolt and Philips 1997) and mammals. Specific BVOCs include olfactive and non-olfactive compounds and these can be considered as a class of terminal metabolites (Wang and Zhao 1995; Zini 2001; Zeng 2002); examples include isoprene, terpenes, alkanes, alcohols, esters, carbonyls and acids (Kesselmeier and Staudt 1999). BVOC profiling can be used as an indication of food quality since bacteria will release VOCs. The BVOC profile of fruits and vegetables are often sampled and analysed at different phases of transport for quality control purposes (Krumbein 2004; Georg Schmarr and Bernhardt 2010). The BVOC profile of human skin and breath varies significantly with respect to time and physiological state and holds great potential as a non-invasive diagnostic tool for personalised health care and for disease progression monitoring purposes (Acevedo 2007; Belda-Iniestaa 2007; Riazanskaia 2008; Song 2009).

1.2.2 Volatile organic compound (VOC) sampling

Analysis of VOCs is normally achieved through headspace (HS) sampling with gas chromatography, HS is one of the most important sampling techniques available for the capture of VOCs, requiring minimal sample preparation and good instrumental sensitivity make this technique of the choice for sampling a wide range of analytes and concentrations. The classical theory of HS sampling is explained in extensive detailed in three key texts by Hachenberg, Loffe and Vitenberg, Kolb and Ettre (Hachenberg and Schmidt 1977; Loffe and Vitenberg 1984; Kolb and Ettre 2006). Other sampling methodologies include liquid-liquid extraction (LLE), stream distillation (SD), simultaneous distillation extraction (SDE) and supercritical fluid

extraction (SFE); a detailed review of these techniques is given elsewhere by Zhang (Zhang and Li 2010).

HS sampling can be split into two main categories: static and dynamic, the following sections briefly describes the two different types of sampling methods and the different surfaces used.

1.2.3 Headspace sampling

Solid phase micro extraction (SPME) is widely regarded as the current gold standard for static Head space (HS) sampling. The capturing method is favoured because it is simple to use and demonstrates a high degree of robustness. Extensive literature is available on the theory and applications of SPME (Pawliszyn 2002; Pawliszyn 2003; Musteatan and Pawliszyn 2007). The sample is placed in a sealed vial (the ratio of sample with respect to the HS is kept at a ratio of 1:1) and a small circular air flow is induced by a rotating magnetic bead as this is necessary for representative sampling of a large volume of HS onto a small surface area of the fibre; this also facilitate in introducing more VOCs into the HS. For solid samples, the sample was placed directly into the vials and capped, the VOCs was then allowed to build up, typically for around 45-60min.

The SPME fibre was then inserted into the HS above the sample and exposed for a fixed amount of time; typically 30min is sufficient to collect any VOCs; the sampling procedure is shown in Figure 2. After the sampling is complete the SPME fibre is retracted and removed, it is then inserted into a GC injector port where it is thermally treated to desorb the VOCs which are transferred into the GC to undergo chromatographic separation and subsequent detection. SPME fibres provide an excellent surface for the trapping of VOCs, with several different activated surfaces available for sampling different classes of VOCs. SPME has a good overall sampling versatility and each fibre can last up to 200 injections before degradation is observed. Its throughput can be limited as each fibre needs to be cleaned individually before and after each injection, this may limit its use in large scale experiments. Furthermore the fibre is highly susceptible to contamination and should be cleaned beforehand and used immediately in order to minimise background signature during sampling.

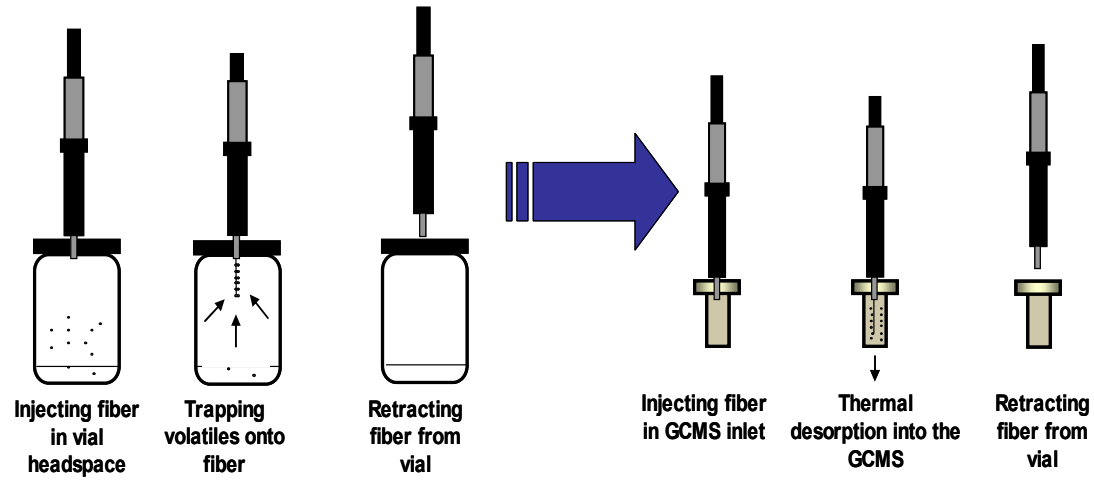


Figure 2 A diagrammatic representation of SPME sampling and introduction of VOC into the GC.

1.2.4 Polydimethylsilioxane elastomer (PDMS)

PDMS patches were originally developed by Riszanskia et al (Riazanskaia 2008) and were design for use in skin BVOC sampling; the patch was modified to act as a passive VOC sampling device. It functions similar to SPME where the patch is suspended over the HS of the sample. After collection the patches are then transferred to a silico-coated thermal desorption tube for GC/MS analysis. PDMS offers good sensitivity and reproducibility for the trapping of polar compounds, large batches can be made and cleaned in advance and they are not particularly prone to degradation. The batches can be stored in a stable manner inside a clean thermal desorption tube for up to several days at time.

1.2.5 Dynamic sampling (DHS)

DHS the volume of the HS within the sample vial is evacuated several times typically using an inert carrier gas such as helium, nitrogen or argon which are passed through an absorbent material such as Tenax-TA or PDMS for the VOCs to be adsorbed on it. DHS allows significant sample enrichment to be achieved as large volumes of air are sampled onto a relatively small surface area; this is ideal for low level concentration VOC analysis (Hagman and Jacobsson 1987). However, this type of sampling is highly problematic as it requires the use of specialised sampling rigs

and other equipment. The complexity of the sampling also increases if, multiple parallel sampling is employed. Analogue needle valve flow controllers are normally used due to their robustness and low cost, but digital flow controllers are commercially available which are more precise however, they are significantly more expensive and durability is often an issue. Suitable flow regulators and controllers are essential in order to obtain reproducible results. Care must be taken during active sampling to ensure consistent flow rate and sampling time, as small variations in either parameters will be amplified into significant differences in the total volume of HS extracted.

1.2.6 Tenax-TA

Tenax-TA is a porous polymer resin of 2,6-diphenylene oxide, it is a highly versatile absorbent material used in DHS. Tenax-TA comes in a granulated form which provides an extremely large surface area for the VOCs to be adsorbed onto, the trapping material is universal for all types of VOCs (polar or non polar) and has good durability for repeated use. Furthermore, Tenax-TA has a very low background signature and if sufficient care is taken during the cleaning process each Tenax-TA packed thermal tube and can be used for up to 250 injections before any degradation is observed. However, it is recommended that the packing material is changed after 250 injections to ensure optimal trapping performance. Tenax-TA does have several drawbacks as it is highly prone to sample degradation and is stable for a few days only after cleaning.

Thermal desorption systems provide a method of sample introduction for transferring VOCs from absorbent material with a large surface area such as Tenax-TA or PDMS, into a GC in a highly reproducible and automated fashion. The Tenax-TA packed thermal desorption tube is first thermally treated to desorbed the VOCs and pre-concentrated it into a smaller volume via a primary cold trap. This cold trap then is thermally heated and the VOCs are then introduced into the GC as a small discrete volume of gas. This method of sample introduction minimise peak broadening and reduces co-elution effects, significantly improving overall resolution and sensitivity of the analysis.

1.3 Instrumentation

1.3.1 Analytical pyrolysis (Py)

Analytical Py (Wampler 1999) is a well established method of sample introduction of nonvolatile materials not amenable to GC analysis. It involves the control of thermal degradation of a sample under an inert atmosphere to generate volatile pyrolysis fragments (pyrolysate) which is then amenable to GC analysis. The sample is applied to the tip of a pyroprobe and inserted into the pyrolysis chamber where it is rapidly heated to the point of thermal decomposition. The pyrolysate is then cryogenically focused onto a primary cold trap; once the initial pyrolysis is completed the cold trap is then heated to desorbs the preconcentrated pyrolsate and introduced into the GC as a small discrete volume. When the experimental parameters for analytical pyrolysis are sufficiently controlled (the temperature, heating rate and duration of heating) the pyrolysis fragment generated will be highly characteristic and indicative of the original parent sample. The technique has been extensively applied in numerous different fields from the analysis of art materials (Shedrinsky 1989; Chiavari 1995), biological samples (Goodacre 1990; Goodacre 1991; Goodacre 1992; Navale 1992; Goodacre 1993; Stankiwicz 1998), food and agriculture (Galletti 1997; Hashimoto 2005) and in particular forensic science for the analysis of paints (Wampler 1997), fibers (Almer 1991) and drugs (Bottcher and Bassmann 1984).

The two most commonly used pyrolysers for GC analysis are Curie-point and resistively heated pyrolysis. In Curie-point pyrolysis the pyrolysis condition is produced by the use of ferromagnetic metal alloys that are rapidly heated using a high frequency induction coil. The pyrolysis temperature is determined by the Curie-point of the metal alloy used and heating ceases when the metal alloys become paramagnetic (the temperature at which the material loses its magnetism), the temperature remains constant until the coil is switched off (Fifield and Kealey 2000). Resistively pyrolysis involves the use of a high voltage current to induce rapid heating of the filament until the desired temperature is reached. The pyrolysis temperature is maintained by reducing the voltage applied, in resistively heating

pyrolysis a greater range of temperature is available to the operator as in Curie-point pyrolysis the pyrolysis temperature is limited by the type of alloy used.

1.3.2 Gas chromatography (GC)

GC is a separation technique suited for the analysis of low molecular weight compounds with a typical mass range of 40 to 600 *m/z*. Chromatographic separation can be defined as the separation of a mixture into its individual components via the interactions with both the stationary and mobile phases. Chromatographic separation in GC (Grob and Barry 2004) takes place within an analytical column, which is a narrow bore open tubular column with a stationary phase immobilised onto its inner wall. The type of stationary phase employed is dictated by the chemical nature of the compounds (i.e. its polarity) and its sample matrix. A small volume of liquid sample (1-5 μL) is injected into a high temperature, pressurised injection port assembly where it is immediately vaporised (Figure 3). The sample is then deposited onto the top of the analytical column, the sample then migrates through the column under the influence of the mobile phase. As this occurs the individual components within the sample will interact chemically with the stationary phase and be retained to varying degrees. As the sample migrates through the column the separation of each individual component within the sample will become more and more pronounced eventually separating into individual discrete bands and is then eluted from the column. Each band or peak represents an individually resolved component within the sample mixture.

The equilibrium constant *K* (also known as the partition co-efficient) is defined as the molar concentration of the analyte within the stationary phases divided by the molar concentration of the analyte within the mobile phase. The time taken from the point of injection to the time of elution is termed the retention time (t_R). Each resolvable component within the sample will have its own unique retention time. The time required for the mobile phases to travel the length of the analytical column is termed the retention of the mobile phase (t_M), (Figure.4).

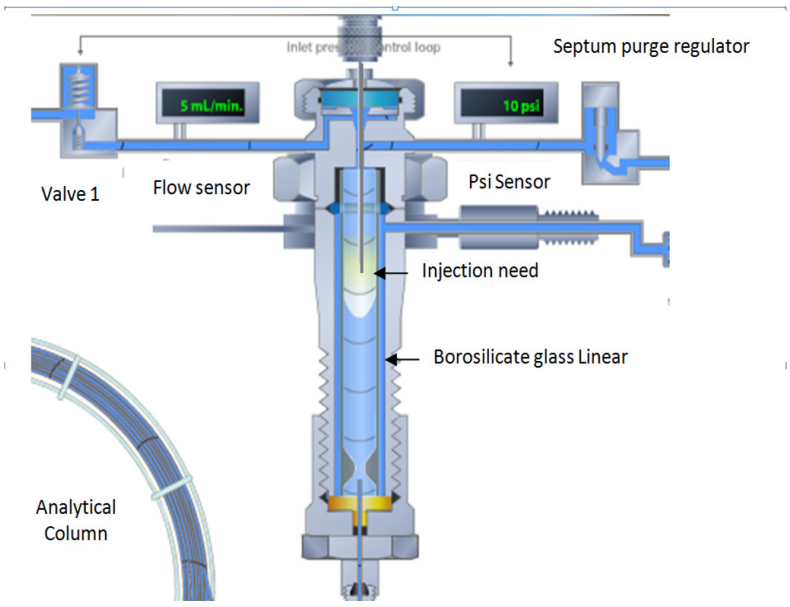


Figure 3 GC Injection port assembly (Restek 2005).

Within GC the distribution of the analyte A is in equilibrium between the two phases

$$A_{Mobile} \rightleftharpoons A_{stationary}$$

The rate of migration of an analyte A through the analytical column can be describe by its capacity factor (***K'***) and is defined as follows:

$$(K'_A) = (t_R - t_M)/\, t_M$$

Where t_M and t_R is obtained from the chromatograph directly. If the analyte has a capacity factor less than 1, the rate of elution will be too fast for the retention time to be determined accurately. The higher the capacity factor (***K'***) the longer it will take for the component to be eluted from the column as it will be more strongly interacting with the stationary phases and therefore be more strongly retained, the capacity factor of most components within a GC normally has a range from 1 to 5 (Skoog 1998).

The separation of the two components (A and B) is given by the selectivity factor ***(α)*** and it is expressed as follows:

$$\alpha = K'_B/K'_A$$

The rate of elution of component A is always faster than component B therefore the selectivity factor will be greater than 1. Chromatographic separation in GC can be described by the plate model in which the analytical column consists of a large number of imaginary layers stacked on top of each other (termed theoretical plates).

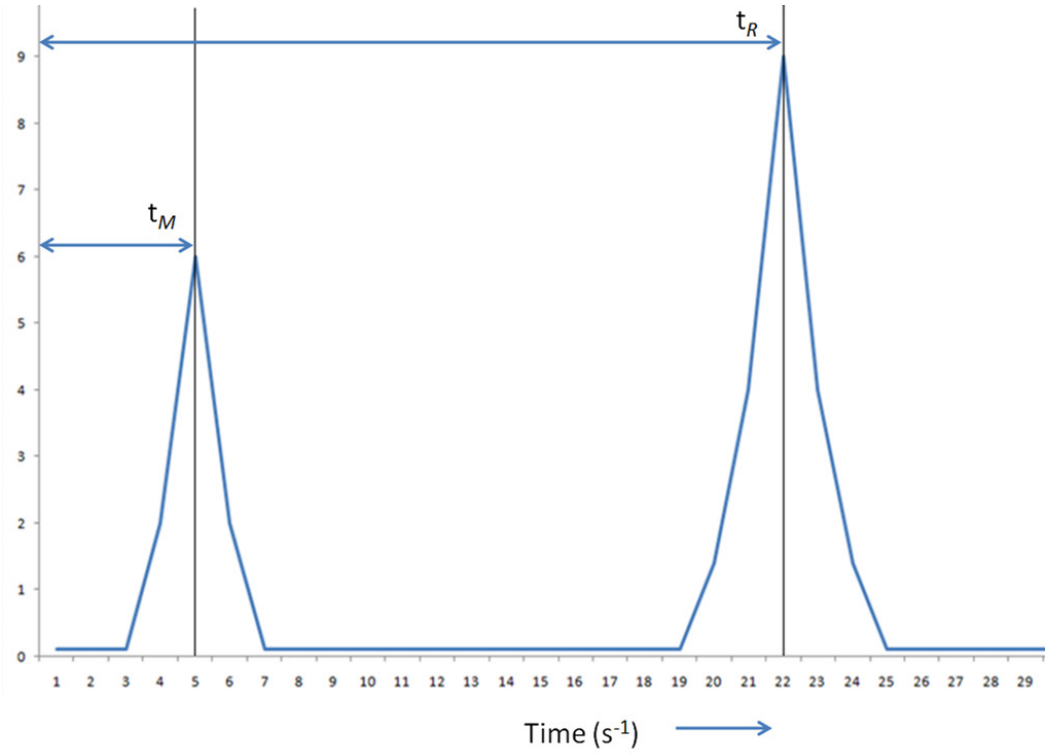


Figure 4 Idealised chromatographs of two peaks.

Each plate is where the separation equilibrium of the samples between the stationary and mobile phase take place. The mixture migrates down the column by the transfer of equilibrated mobile phases from one plate to the next. The plates model is an abstract concept developed to describe the column efficiency by either stating the total number of theoretical plates within the column (***N***), the greater the number the higher the efficiency, or by stating the height equivalent to a theoretical plate (***HETP***) the lower the number the more plates are contained within a given unit of length of column. Therefore for a given column length (***L***), the (***HETP***) can be calculated as:

$$HETP=L/N$$

The number of theoretical plates on the analytical column can be approximated directly from the chromatograph as follows:

$$N=5.55(t_R)^2/(w_{1/2})^2$$

Where t_R is the retention time of a given peak and w is the width of the peak at half height.

The resolution of a column (R_s) provides a quantitative measure of its ability to separate two
analytes. A resolution of 1.5 gives effectively complete separation of two components, the
resolution of the analysis can be improved by increasing the length of column or by decreasing the temperature ramp rate to resolve compounds with similar volatility; however both approaches will result in increased analysis time.
Resolution of separation with respect to peak width and the number of theoretical plates is expressed as follows:

$$R_s = (\sqrt{N})/4 * (\alpha - 1/\alpha) * (K'_B/1 + K'_B)$$

The equation can be rearranged to calculate number of theoretical plates required for a specific resolution as follows:

$$N = 16R_S^2 (\alpha/\alpha - 1)^2 * (1 + K'_B/K'_B)^2$$

1.3.2.1 Van Deemter equation

The plate model however does not take into account the lag time taken for the solute in the gas phase to be equilibrated between the stationary and mobile phase (it is assumed that this occurs at an infinitely fast rate). The band shape of the chromatographic peak is thus dependent upon the rate of elution and different paths available for the solute as they migrate from one plate to next. Various mechanisms that contribute towards peak broadening are described by the Van Deemter equation (Figure 5). The equation show below contains terms which are linearly proportional, inversely proportional and independent of the velocity of the mobile phases (Van Deemter 1956).

$$HETP = A + B/(u + Cu) = HETP = A + B/u + (C_s + C_m)u$$

Where u is the average velocity of the mobile phase given in cm s^{-1}, and (A) is the multiple path term or eddy diffusion which describes the alternative path taken by the solute within the
column, and is linearly proportional to the diameter of the packing material which make up the column. Multipath broadening is partially offset by ordinary diffusion effects at a low flow rate, as the analyte particles can switch between different paths

quickly and this reduces variations in transit time from one plate to the next, the effect of multipath diffusion is independent of the flow rate. It should be noticed that this term has more influence in liquid chromatography LC where the analytical column is packed rather than in GC where the column is an open tubular as this term only has a very small contribution.

B is the longitudinal diffusion term, the concentration of the analyte is lower and at the edges of the band than at its centre, the analyte will therefore diffuse out toward the edges resulting in peak broadening. This effect is inversely proportional to the flow rate. The higher the flow rate, the less time the analyte has to spend on the column and reduces the effect of longitudinal diffusion. This effect is much less pronounced in LC because the longitudinal diffusion term reduces to zero relative to the other terms due the high pressure employed.

C is the mass transfer coefficient, where C_s and C_m is the mass transfer co-efficient of the stationary and mobile phase respectively. The mass transfer co-efficient describe the finite time required for an analyte moving from the mobile phases to the stationary phases and back again if the velocity of the mobile phase is too high and the analyte is strongly retained by the stationary phase, the analyte in the mobile phases will move ahead of the analyte in the stationary phases resulting in band broadening and this effects is proportional to the velocity of the mobile phase. Other factors that will contribute towards column efficiency are the diameter of the packing material and the internal diameter of the analytical column itself. A lower diameter of the packing material will allow lower stable plate height to be achieved with respect to the increasing flow rate; however this effect is much more influential in LC than in GC. Smaller internal column diameter column will have a lower surface area and therefore lower capacity but the effects of band broadening are also reduced.

Temperature is the single most influential factor in GC since chromatographic separation is based upon chemical volatility. Lower boiling point compounds will traverse through the column faster than those with a higher boiling point. Increasing the GC oven temperature results in a increase in the rate of elution of all compounds within the system, as they will have a greater amount of kinetic energy to migrate through the analytical column. Initial temperature or ramp rates (degrees per minute)

that are too high can severely affect the chromatographic separation and subsequent resolution of the analysis.

Variable linear temperature programming is standard practice in GC to solve the general elution problem, where the GC oven temperature is programmed to increase at a set rate over a given period of time (in degrees per minute, commonly known as the ramp rate). This is used to exploit the large differences in the boiling points of different compounds present within a given mixture, the method also minimise peak broadening effect and improves resolution while reducing analysis time compared to isothermal analysis

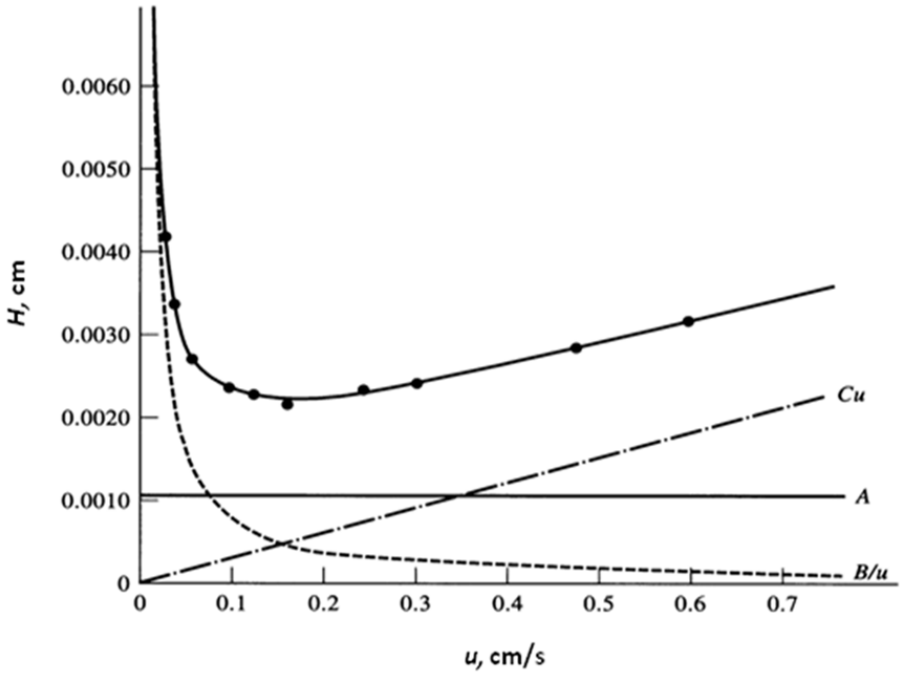


Figure 5 A Van Deemter plot showing the contribution of the various terms, the upper curve are experimental data, Cu is the mass transfer for both phases, (A) is the multipath effect and (B/u) is longitudinal diffusion (Skoog 1998).

After elution from the analytical column, the band or peak is then subjected to detection and quantification by various means. GC can and has been coupled to many different types of detectors, including flame ionization detectors (FID), nitrogen phosphor detectors (NPD), electron capture detectors (ECD) and has been increasingly coupled to mass spectrometers (Adlard 2007) for the analysis of small low molecular weight compounds.

1.3.3 Gas chromatography mass spectrometry (GC/MS)

GC/MS is a hyphenated analytical method widely employed in the field of analytical science and metabolomics. The GC is employed as the front end separation technique in which, complex chemical samples are fractioned into individual components which are then sequentially introduced into a MS detector. The MS typically employs an electron impact source (70eV), to ionise the sample generating a molecular ion which is unstable and will undergo self fragmentation into smaller sub units. This is then introduced into the QMF/QIT cell where, the ion fragments are separated and detected according to their mass to charge ratio (*m/z*). The peak fragmentation profile generated is highly characteristic and indicative of the original parent molecule, by examining the isotopic ratios, distribution and composition of these ion fragments, detailed chemical information regarding the structure and functional group present can be obtained. The peak fragmentation profile is normally matched against a known mass spectral library database such as National Institute of Standards and Technology (NIST) (Stein 1999) to aid in the identification of the unknown compounds.

1.3.3.1 Quadrupole mass filter (QMF)

QMF and the Quadrupole ion trap (QIT) were both developed in parallel during the early 1970s, below is a brief description of both mass analyzers and their principle of operation. A QMF consist of four symmetrically spaced cylindrical electrodes connected electronically in pairs, a positive (U+Vcos (ωt)) potential is applied to one set, while at the same time a second negative - (U+Vcos(ωt)) potential is applied to the other set as shown in Figure 6.

Where (U) is the dc voltage, (Vcos(ωt)) is the ac voltage, (ω) is the angular velocity of alternating voltage and (t) is time. The potential applied to the two set of electrodes is identical in amplitude but are 180^{o} out of phase with respect to each other (this is commonly known as superposition effect), generating a quadruple field which will affect the trajectory of ions in the X-Y plan relative Z axis in the QMF cell.

The motion of the ions within this field is an inherent property associated with its mass to charge ratio (*m/z*), at selected U and Vcos(ωt) potential only ions of a certain *m/z* value will be able to traverse entire length of the QMF cell and be detected (resonant ion), while all other ions of different *m/z* values will be ejected from the QMF or neutralised through collision with the electrodes due to unstable trajectory (non resonance ions).

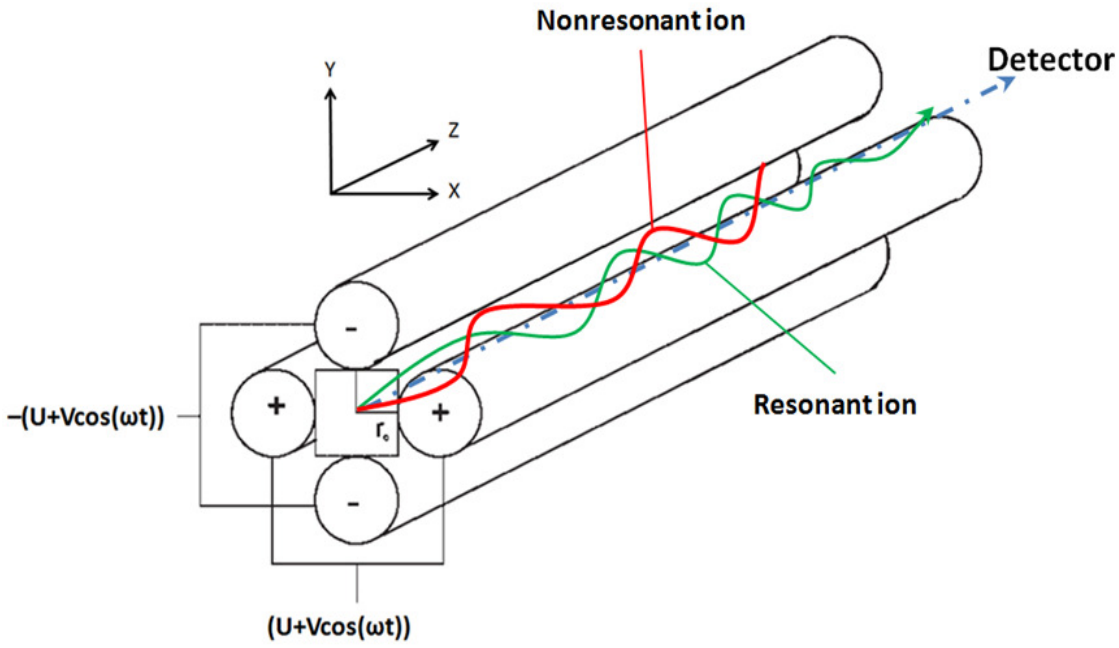


Figure 6 Idealised representation of QMF (see text).

In order for the ion to traverse the entire length of the QMF it must have stable trajectories in both the X-Z (high mass past filter) and Y-Z axes (low mass past filter), this mutual stability condition imposed by the low and high mass pass filter combined to form a band mass filter where, only ions of the correct *m/z* value will be able to pass through to the detector, this is dictated by the ratio of ac and dc potential and its applied amplitude. By increasing magnitude of both the ac and dc potential whilst keeping their ratio constant, ions of different *m/z* values is able pass through the QMF and arrived at the detector generating a mass spectrum (Miller 1986).

1.3.3.2 Quadrupole ion trap (QIT)

QIT is often thought of as a 3 dimensional quadrupole filter; however its design and assembly is somewhat different, the QIT is significantly more compact in terms of

design and construction and it allows greater control and manipulations of ions in comparison to a QMF. A QIT consist of three electrodes, two of these electrodes has a hyperbolic geometry and are often referred to as end cap electrodes due to their shape. At the centre of each end cap electrodes there is an ion aperture which allows either the introduction or exit of ions through the QIT assembly.

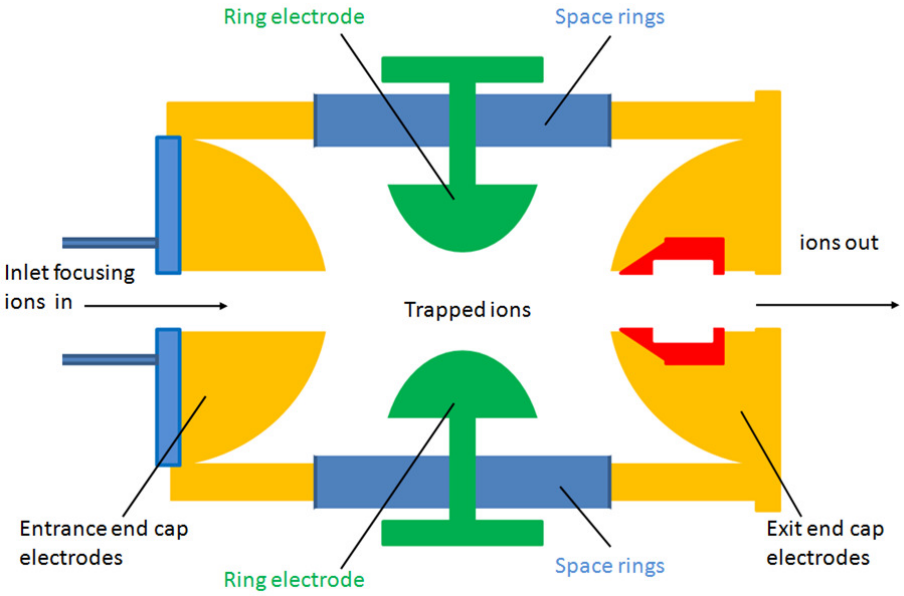


Figure 7 Idealised diagram of the ion trap assembly.

A third ring electrode with an internal hyperbolical geometry (like a hollowed out donut shape) is positioned symmetrically between the two end-cap electrodes, Figure 7 shown an idealised ion trap assembly. As the ions enter the QIT assembly, they are trapped within a quadrupole field generated by the voltages applied to these electrodes. A low amplitude auxiliary oscillating potential is applied to the two end cap electrodes whilst, a RF oscillating drive potential is applied to the central ring electrode; this creates a 3 dimensional rotationally symmetrical quadrupole field trapping the ions in a stable oscillating orbit and in the presence of the carrier gas at the pressure of 10^{-3} Torr the ions are condensed into a cloud in the centre of quadrupole field (March 2000).

The motion of the ions is dependent upon the RF potential applied, by increasing the drive potential the orbital motion of the ions will become progressively more and more energetic, until they are eventually ejected from the field and be focused onto the exit aperture towards the detector. A useful property of this 3D quadrupole field

is that the ions are ejected in order of increasing *m/z* corresponding to an increase in amplitude of the applied RF and dc potential (March 1998). By keeping the ratio of the RF and dc potential constant, but increasing in amplitude, ions of increasing *m/z* are destablised from the QIT and be focused into detector and generate a corresponding mass spectrum.

In tandem MS/MS a survey scan is initially carried out in which all the ions are sequentially destablised and ejected from the QIT in order to indentify the ion of interested. A separate second scan was then conducted which involves the selective trapping of the desired ion. MS/MS operation in QIT is carried out in three separate sequential stages:

(1) Initially the voltages in the QIT was selected so that all other ions apart from the one of interested was ejected from QIT, this was achieved by increasing the RF potential applied to the central electrode while, keeping the supplementary ac voltage to be the same on both end cap electrodes.

(2) The second stage involve the stabilisation of the precursor ions within the QIT, the RF voltage potential to the central electrode was decreased until the isolated precursor ions attend a stable trajectory within the quadrupole field.

(3) The isolated precursor ions are then subjected to CID, where a supplementary time dependent RF potential was applied to both the end cap electrodes, exciting the precursor ions in the field causing them to undergo CID with the surrounding helium carrier gas, resulting in the formation of product ions in the QIT. The ions are then destabilised and ejected from the quadrupole field for mass detection.

The capacity of the ion trap is somewhat limited and its linear dynamic range can be compromised by the over population of ions, where the spatial charge repulsion effect between individual ions arises and reduces the trapping efficiency of the field. In order to minimised this effect, the ions are rapidly counted before they enter the QIT and the time window is dependent upon the ion flux, the greater the ion flux, the shorter the time window and vice visa. Ensuring that, there is a fix amount of ions within the QIT each time and it does not become over populated, the disadvantages of this is that low intensity ions with weak signal are unlikely to be detected in a mix of other ions (March 2000).

1.3.4 Chemical derivatisation

Chemical derivatisation is widely used in the field of metabolomics and GC/MS analysis. It is employed to permit the analysis of compounds not directly amenable in GC analysis due to inadequate volatility or stability. Furthermore it serves to increase the quality of analysis by improving the chromatographic profile and subsequent detectability of the analyte molecule within the sample. The main reason for using chemical derivatisation is to impart volatility to otherwise non volatile compounds, low volatility in large compounds is the result of a large dispersion field encompassing the molecule stabilizing it, in such instance chemical derivatisation cannot be used to increases the molecule's volatility (Knapp 1979). In smaller molecules low volatility may be the result of strong intramolecular attraction between the polar groups present. Masking such functional groups can yield significant increases in the molecule's volatility, polar groups such N-H, O-H and S-H groups undergo hydrogen bonding and have a significant contribution towards the intermolecular attraction. Replacing the active hydrogen in those groups through alkylation, acylation or silylation will dramatically increase its chemical volatility particularly in compounds with multiple polar groups.

The analysis of monosaccharide are a prime examples of this (Wang and Huang 2007), the low molecular weight molecules display little volatility even when subjected to temperatures at the point of decomposition. However by replacing the active hydrogen with methylation groups can generate a derivative which is readily amenable to GC analysis. The opposite is also true in situations where excessive volatility is a problem, derivatisation can be used to yield less reactive products to help minimize sample lost in preparation and improve the chromatographic profile. Also in the case where thermally sensitive compounds are able to be volatilized but undergo partial thermal decomposition during GC analysis, derivatisation can be employed to yield more stable product and improve chromatographic resolution (Knapp 1979).

1.3.5 Differential mobility spectrometry (DMS)

DMS is a gas phase ion separation technique operating at ambient pressure. This analytical technique emerged from ion mobility spectrometry (IMS) (McLean 2005). DMS operates by exploiting the non linear dependency of a compound's ion mobility (ΔK) within an alternating weak and strong electric field. (Krylov 2007;

Krylov and Nazarov 2009). The technique was originally developed for the detection of chemical warfare agents and explosives; it has been applied more recently to the analysis of human breath (Basanta 2007), bacteria (Prasad 2007), forensic analysis (Lu and Harrington 2007) and jet fuel (Rearden 2007).

The DMS is composed of the three main components; an ionisation source, a drift region consisting of two closely spaced parallel metal plates and a detector. Within DMS, the samples are first ionised and then swept into the drift region under the influence of a drift gas (N2) here the ions are subjected to a rapidly alternating asymmetric electric field perpendicular to the ion flow, a property of this applied asymmetric electric field is that it has two specific cycles to it, a high field (strong electric field, $\mathbf{t_1}$) and a low field end (weak electric field $\mathbf{t_2}$) show in Figure 8.

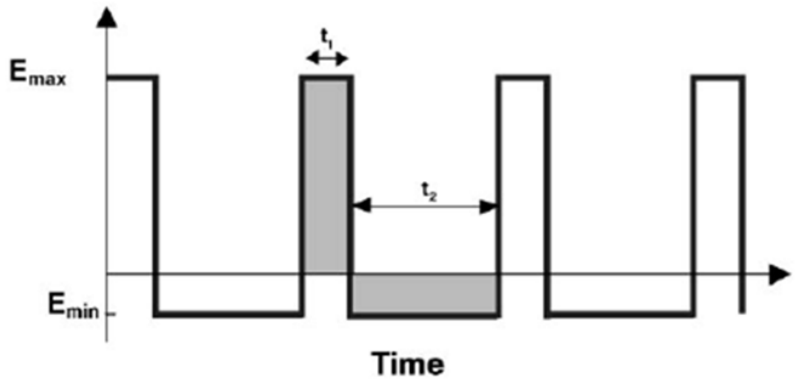


Figure 8 The two different cycle of the asymmetric electric field applied across the drift tube region.

In the presence of a strong electric field (high field end) the ions will move toward the upper electrode plate while the opposite occurs when ions are exposed to a weak electric field (low field end). This results in the ion moving through the drift region in a zigzag trajectory. The asymmetric electric field induces the transverse displacement of ions between the electrode plates. Any non resonance ion will eventually strike either the upper or lower plates and become neutralised in the process (Prasad 2007). The trajectories of any resonance ions as they transverse the drift region, do not necessarily have be in a staggered zigzag manner, it is possible to align the ion's trajectories to be more linear, enabling it to traverse from the drift region to the detector unhindered. This can be achieved by applying a suitable low amplitude DC electric field (termed compensate voltage) in the opposite direction to the applied asymmetric waveform across the drift region. By applying a single fixed

DC voltage only one type of ions will have a resonance trajectory able to traverse the drift region and arrive at the detector unhindered, this is due to each type of ion having its own unique (ΔK) value which is a property associated with the ion's shape and charge (Miller 2001).

The ΔK of the ion corresponds to a particular dc frequency within applied the compensation voltage therefore by applying a fixed single DC voltage, the DMS function becomes similar to a continuous ion filter or alternatively the DMS can scan across the DC range in a sweeping manner producing a spectrum of all ions present within the analyzer (Figure 9), this methodology is termed dispersion field programming and it is commonly employed in DMS to increase the amount of analytical information obtained (Basanta 2007).

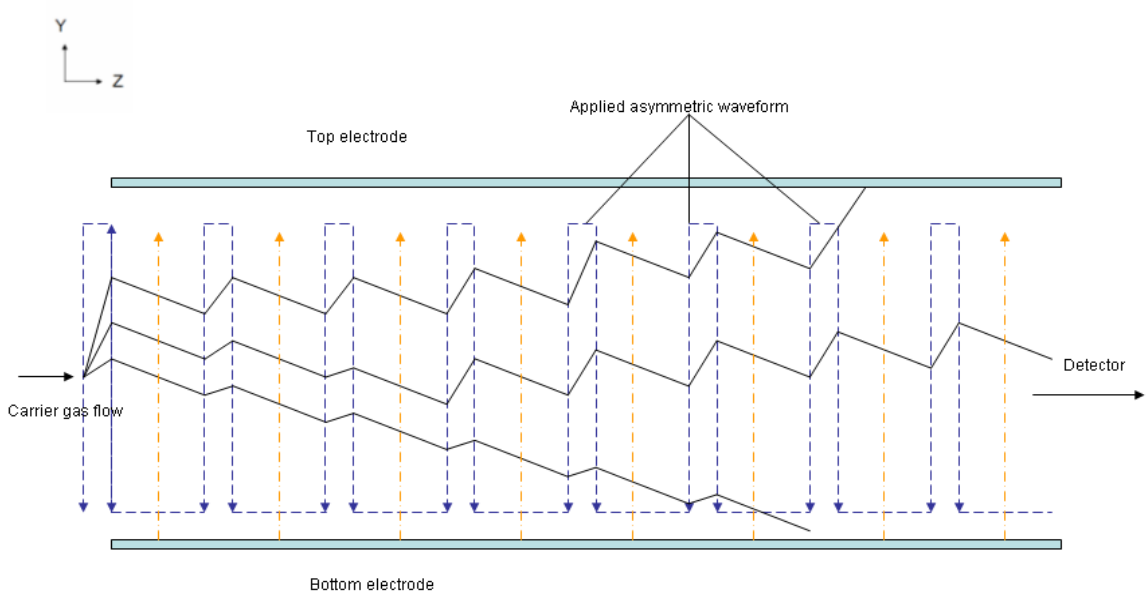


Figure 9 Interactions of ionic species in the presence of a asymmetric applied electrical field (in Blue) and a compensation voltage (CV) (in Orange), when the appropriate CV is applied, ions with corresponding (ΔK) can transverse unhindered towards the detector.

1.3.6 Raman spectroscopy (RS)

RS is a non destructive analytical technique which can be applied to the analysis of different matrices in various physical states (liquid, solid or gas) and in very dilute concentrations, with minimal sample preparation. In RS a monochromatic radiation (laser) is used to irradiate the sample, the photons will neither be scattered elastically or inelastically. The majority of the photons will be scattered elastically where there

is no change in the frequency/wavelength of scatter photons relative to the applied incident radiation (Rayleigh scattering). However, a very small component will be scattered inelastically where there will be a transfer of energy either from the photon to the molecule or vice visa this is known as Raman scattering and can be used to give detailed molecular information about the structure and chemical bonds present within the molecule.

Raman scattering can be further divided into two categories Stokes and anti-Stokes scattering, in Stokes scattering, the molecule gains energy and is promoted to a virtual state and emitted photon which has a lower energy relative to the incident radiation (excitation), while the opposite is true for anti-Stokes scattering. Anti-Stokes scattering arises due to thermal contribution, as some of the molecules will already be present in a virtual state and scattering from this to a lower ground state is termed anti-Stokes scattering (relaxation). The effect of anti-Stokes is extremely weak in comparison to Stokes scattering, as the majority of the molecules will be present in it lowest vibrational ground state and the likelihood of observing a Stokes scattering phenomena will be much greater than anti-Stokes, in addition the effect of anti-Stokes scattering will increase proportionally with increasing in temperature due to greater thermal contributions. Raman scattering is an inherently a weak process where only 1 in every 10^6-10^8 photons will be scattered this way. The weak Raman scattering can be significantly amplified through surface or resonance enhancement to obtained very high level of sensitivity even with very low analyte concentration.

1.3.7 Surfaced enhanced resonance Raman spectroscopy (SERRS)

SERRS is a variation of Raman spectroscopy which combines 2 independent mechanisms of enhancement together.(1) The applied excitation frequency is tuned to be sufficiently close to the molecular adsorption maxima of the analyte for the resonance enhancement to take effect, generating significantly increased Raman scattering compared normal Raman scattering. (2) This signal intensification can be further amplified by surface enhancement through coupling to silver/gold nanoparticles (conventional SERS) and the combined enhancement of a factor of 10^{14} can be obtained if, the excitation frequency sufficiently matched to the molecular adsorption maxima (Roger 1996; Nie and Emery 1997). SERRS has been

successfully applied to the analysis of a variety of biological samples, DNA, antibodies, cytochromes and since water scatters light weakly, SERRS can be used to analyse samples in aqueous environments, this holds great potential for bio-sensing applications and *in vivo* measurements (Bizzarri 2002). An additional benefit of SERRS is that any fluorescence generated is efficiently quenched by the metallic nanoparticles and used for subsequent enhancement, this permits the analysis of a wide range of colored dyes natural or synthetic with high fluorescence background as long as it can made to adhere to the metal surface (Graham 2006),

1.3.8 Surfaced enhanced Raman spectroscopy (SERS)

SERS is a relatively new spectroscopic technique used in the analysis of solid and liquid samples. Fleischman and colleagues were the first to observe SERS when investigating pyridine adsorbed onto an electrochemically roughened silver surface. (Fleischmann 1974) Pyridine exhibited an increased Raman scattering effect. At the time these authors were unable to account for this enhancement and attributed the effect solely to the increase in surface area of the silver electrode. Shortly after this two groups independently reported similar observations (Albrecht and Creighton 1977; Jeanmaire and van Duyne 1977) and also highlighted that the increase in surface area of the electrode was not sufficient to account for the signal enhancement.

Two different mechanisms for the observed SERS enhancement were proposed, the electromagnetic enhancement effect by Jeanmaire and Van Duyne (Jeanmaire 1977) and the charge transfer effect (also known as chemical enhancement effect) by Albrecht (Albrecht and Creighton 1977). The development of the electromagnetic theory was built upon and developed further during the early part of the 80s and was investigated by several research groups (Gersten and Nitzan 1980; McCall 1980; Gersten 1980a; Gersten 1980b; Gersten and Nitzan 1981; Kerker 1984) Kneipp *et al*. approximated which metals would be most likely to generate significant enhancement based on their surface plasmonic oscillations properties.

It was also theorized that the enhanced Raman cross section was not solely confined to a roughened silver surface but could be replicated in a similar manner using other metal nanoparticles (Kneipp 2006). The charge transfer theory was further developed

and refined in the early 90s (Mrozek and Otto 1990; Otto 1991; Otto 1992), it was proposed that the analyte is covalently bound to the surface of the SERS substrate achieving Raman spectral enhancement. In understanding the SERS effect it is necessary to describe both proposed theories.

1.3.9 Electromagnetic effect

Surface plasmons are waves of oscillating electrons which exist at the surface of a dielectric material, such as metallic nanoparticles. Their sinusoidal amplitude can be tuned using a monochromatic light source. When radiation of a defined wavelength hits the nanoparticles' surface, localized surface plasmons become excited. When the frequency of the plasmon oscillation (ω_p) is in resonance with the applied monochromatic radiation source, the amplitude of the plasmonic waves will be at their maximum. The plasmon oscillations occurs perpendicular to the plane of the surface. In order to induce Raman scattering there must be a change in polarizability of the analyte, the enhancement effects can be simply explained using the equation below:

$$\boldsymbol{\mu=\alpha E}$$

Where **μ** is the induced dipole moment, **α** is the polarisation, **E** is the electric field.

The oscillating electron waves which surround the nanoparticles work to increase the electric field surrounding the analyte. This causes an exaggerated increase in the polarisability of the analyte and results in the formation of a larger Raman cross section.

Theoretical models such as the generalised Mie theory (Hao and Schatz 2004) allow the plasmonic fields of metallic nanoparticles to be described when the nanoparticle is a single entity or part of an aggregated system, it is evident that the aggregated nanoparticles exhibit a greater number of SERS hotspots, where the electric field is at its greatest (Schatz 2006). Therefore, experimentally it is necessary to bring the nanoparticles into close proximity to ensure maximum enhancement is achieved. Further studies demonstrates that plasmon bands can be altered by varying the shape and size of the nanoparticles and experimental data indicates that for a single nanoparticle, Raman enhancement is less than 10^6 however in dimers, trimers and poly nanoparticle systems, Raman hotspots become more prevalent and enhancement is greater (Schatz 2006). Modeling a nanoparticle as a single spherical entity means that the equation below can be derived as follows:

$$E_r = E_o \cos\theta + g(a^3/r^3)\, E_o \cos\theta$$

Where E_r is the total charge of the electric field, at a given distance r from the spherical surface a is the radius of the sphere, θ is the angle relative to the direction of the electric field and g is the surface plasmon intensity related to the dielectric constant. E_0 is the dielectric of the surrounding medium, E_1 is the dielectric constant of the metallic sphere and v_L is the frequency of the incident radiation:

$$g = (E_1(v_L) - E_0 / E_1(v_L) + 2E_0)$$

The value of g is inversely proportional to the denominator, when g is at maximum; the plasmon resonance frequency is increased resulting in the immediate area surrounding the nanoparticles to experience an increase in localized electron density. This increase in charge density also encompasses the molecule causing the electrons within it to become polarised giving rise to intense Raman scattering (Smith and Dent 2005).

1.3.10 The charge transfer mechanism

The charge transfer model is based around the formation of a metal-analyte covalent bond, allowing communication to and from the metallic surface. Electron mobility along this bond perturbs the analyte's electronic cloud increasing the Raman cross section and its affinity for change in polarisability. The orientation of the molecule and how it is bound to the nanoparticle's surface is dependent upon the structure and functional groups present.

Molecules containing thiols and amines have demonstrated greater adsorption property towards a gold surface whilst molecules containing halogen atoms exhibit similar trends toward a silver surface. This theory is limited to monolayer coverage, as bond formation is essential for the phenomena to occur. Through a combination of mathematics and experimental observations the chemical theory seem to be very well defined however, it only describes a small fraction of all Raman spectral enhancement (Liu 2009; Morton and Jensen 2009).

1.3.11 Nanoparticles

Nanoparticles are dielectric materials composed of a positive charge center with sinusoidal electron waves propagating at its surface. The silver and gold colloids

used in both studies were prepared according the Lee-Meisel method in a one pot synthesis (Lee and Meisel 1982). SERS enhancement is facilitated most commonly by either silver or gold colloids. The size and shape of the nanoparticles can be manipulated by varying the concentration of chemical reactants and also their reactivity. The silver (Ag) and gold (Au) colloid have different optical properties, Ag and Au solution displays absorption maxima of approximately 420nm; Au sol has maxima around 520nm respectively. Aggregated Au colloid show new absorption bands in the region of 500-600 nm and for Ag colloid in the region of 700-900 nm, this is due to plasmonic shift. Nanoparticle suspensions offer a better alternative to a metallic surface due to their increased surface area; furthermore as the nanoparticles are suspended in solution they also benefit from the effect of Brownian motion, therefore the SERS spectra observed represents an averaged effect.

1.4 Data analysis

The high throughput capacity and sensitivity of modern MS-based instruments are not without its drawbacks. The ever increasing complexity and volume of data generated represents a real significant challenge in terms of data storage, curation and analysis. GC/MS and LC/MS data often have a high degree of dimensionality associated with them. This necessitates the need for an accurate and efficient data pre-processing and analysis methodology in order to extract meaningful and interpretable information from the experiments, therefore the use of chemometrics plays an integral part within metabolomics.

1.4.1 Data pre-processing

Data pre-processing is an important step prior to multivariate data analysis, as inappropriate use will severely biased the result obtained. This may include the effects of misalignment, baseline drift, large scale differences between variables and variation in samples or sampling and need be accounted for and minimized. If these technical/non-biological variations are not adequately constrained, the major trend observed would be due to such variations mentioned, rather than any biological differences between the samples.

Hyphenated GC/MS or LC/MS analytical methods used in metabolomics analysis can generates huge quantity of highly complex data, which may require a significant amount of data pre-processing before it is suitable for multivariate data analysis, below is a brief description of the data pre-processing steps applied in this research thesis.

1.4.2. Alignment

In GC/MS, slight variations in the flow rate will result in instrumental drift over time. The traditional univariate approach focusing only on selected parts of the chromatograph for quantitative analysis, are largely not affected by such misalignment. However, in multivariate analysis the entire chromatograph/spectra is normally utilised, therefore it is essential that the chromatographs are properly aligned so that they have the same start and end time; and each corresponding variable within different samples are matched accordingly by retention time. Correlated Optimised Warping (COW) is a alignment algorithm extensively applied for correcting chromatographic shifts (Nielsen 1998, Pravdova 2002), it is a piecewise or segmented data processing method, aim at aligning samples towards a common reference template,. This is done by allowing limited change in the segments lengths (time windows) and predefined shift margins (slack size). The different windows are shifted or warped, so that it optimises the overall correlation between the sample, and the reference template. Mis/unaligned data when subjected to PCA the resulting major trend observed would likely to be due to the differences in alignment or drifting between the samples rather, than to biological differences.

1.4.3 Baseline correction

Differences in background intensity for samples analysed at different time points often has a contributions towards the clustering observed in PCA, baseline correction (Boelens 2004) with asymmetric least square (ALS), is an adaptive baseline estimation algorithm, employed to eliminate/minimise this type of instrumental variations, by adjusting the background intensity of all the samples to be the same prior to multivariate data analysis. The shape and intensity differences of the background relation to the eluent/vibrational peak are taken into account, and are subtracted from the chromatogram/spectra. If data set is not baseline corrected prior to PCA, it will have a significant effect on the resulting scores plot in a similar

manner to unaligned data, where the clustering observed would likely to be based on the differences in baseline intensity between samples rather than biological differences.

1.4.4 Scaling

Scaling in data pre-processing is used to adjust for large scale differences and to account for heteroscedasticity (differences in the variance between variables) within the data set. By adjusting the variances of each variable to be the same initially, this would allow the PCA model to describe/capture as much systematic variations as possible, as every variable will have an equal opportunity to enter the PCA model. There are two types of scaling relevant to two way data; mode one scaling is when every row (samples) is multiplied by a specific number, and mode two scaling is when every column (variables) is multiplied by a specific number. This is commonly applied to the data set prior to PCA. The individual variable are scaled by the inverse of it standard deviations, this in combination with mean centering across the row vector, this is often refer to as autoscaling or scaling to unit variance or column standardization (Bro and Smilde 2003).

1.4.5 Mean centering

Mean centering is applied to remove any common offset present within the dataset and adjust the distribution of the data points to be more symmetrical. Mean centering is achieved by subtracting the mean of each variable from itself, the process also has the following outcomes (1) reduction the Rank of the model, (2) improvements to the fit of the model; mean centering is automatically implemented prior to PCA (Brereton 2003).

1.4.6 Normalisation

Normalisation is used to adjust for sample to sample variations, one such type of normalisation is a min-max normalisation where each variable is divided by the absolute differences between the maximum and minimum intensities, the dataset is scaled so that the influence (weight) of each sample lies within a predefine range with respect to each other (i.e. between 0 and 1), this also avoids large values issues (Xu 2006).

1.5. Unsupervised data analysis

1.5.1 Principal component analysis (PCA)

PCA is one of the most well known and widely used tools in exploratory data analysis (EDA). PCA was first published over 100 years ago (Pearson 1901) and was further developed by Hotelling (Hotelling 1933). PCA is a variance capture method in which it seeks to describe the most significant variations within the data set, by the use of a few latent variables termed principal components (PCs). Through a linear combination of the original features, PCA can be used to highlight underlying trends or clustering relationships within the data set; it is also used in the detection of outliners. In real application it is normal that multiple variables follow a similar trend and can be effectively described using a single PC, therefore PCA can be used to reduced the dimensionality of the data set. Given a data matrix (***X***) the PCA model is as follows:

$$\boldsymbol{X} = \boldsymbol{T} \cdot \boldsymbol{P}' + \boldsymbol{E}$$

Where ***X*** is the original matrix of ***i*** number of columns and ***j*** numbers of rows, where the columns represent the variables and it elements present the value of chemical measurement on the samples (***j***). The number of columns in the matrix ***T*** is equals to the number of columns in the matrix ***P***. ***T*** is termed the score matrix and contains the latent variables, and it contains as many rows as in the original matrix. ***P*** is the loading matrix and contains as many columns as the number of variables in the original data matrix. The ***n***th column of ***T*** and ***n***th column of ***P*** can be represented by vectors ***Tn*** and ***Pn*** and are vector representation of the ***n***th PC.

The product of ***T***x***P'*** can be regarded as a model of the data approximation by the PC to the original dataset and the residual error is represented by ***E***. This general equation can be expanded thus:

$$\boldsymbol{X} = \boldsymbol{t}_1 \cdot \boldsymbol{p}_1' + \boldsymbol{t}_2 \cdot \boldsymbol{p}_2' + \ldots + \boldsymbol{t}_k \cdot \boldsymbol{p}_k' + \boldsymbol{E}$$

The first PC accounts for the largest variation and subsequent PCs have progressively smaller variations. In most cases, the variation caused by noise is smaller than that caused by biological factors. By disregarding the latter PCs, where

little or no information of interest is present, random instrumental noise variation can be effectively removed from the data set.

In order to visualised the overall clustering relationships of the samples or objects within the dataset, the few PCs in ***T*** are plotted against each other (i.e. PC1 vs. PC2, PC1 vs. PC3, and PC2 vs. PC3), this is called the scores plot and permits any major trends, groupings and outliners to be revealed, the Euclidean distance of the objects within the scores plots, can be used to relate them to each other. Objects which lie close together will have a similar multivariate profile whereas; objects which are far apart will be more dissimilar to each other; by assessing the compactness and purity of the clusters within the score plots it is possible to evaluate the reproducibility of replicate analysis as well.

Analogous to the scores plot, the loadings plot which plot the first few PCs in ***P*** against each other displays the relative weight/influence of each individual variable and its contribution towards the separation observed within the score plot. The greater the magnitude of a variable within the loading vector, the more the contribution it has towards the trend observed in the scores plot. An important feature of PCA that the directions of the scores plot correspond to the directions in the loadings plot, this allows the identification and correlation of which variables are responsible for the separation observed between different groups in the scores plot (Lindon 2007).

PCA is used both as an exploratory data analysis (EDA) tool as well as a modelling tool. By examining the cumulative percentage explained variance, with respect to the number of PC extracted, it is possible to assess the proportionality of the data that has been modelled through PCA, the closer it is to 100% the more accurate the data has been modelled.

1.5.2 Parallel factor analysis (PARAFAC)

PARAFAC is another form of unsupervised leaning methodology and it is particularly suited to analysing multiway data with known multiple trends present. PARAFAC has its roots in psychometrics (Carroll and Chang 1970; Bro 1999) and it is a generalisation of PCA. Other methods for analysing multiway data include Tucker3 methods (Kroonenburg 1983; Kroonenberg 2009) or by simply unfolding the 3D array into a 2D matrix prior to PCA analysis (Bro 1997).

In multiway data the structure is composed of a series of matrices stacked in front of one and another arranged in a cube like fashion and a set of variables measured in a nested manner (where the variables are measured repeatedly over different samples and different time points). In such cases where multiple trends maybe present, PCA may not be the most suitable method for analysing the data. This point can be demonstrated by the following example: for an F component PCA solution to a multiway array with a dimension of ***(IxJxK)***, in conventional PCA the 3D matrix is first unfolded into a 2D matrix so that it becomes ***(IxJK)***. The PCA model will consist of ***F(IxJK)*** parameters with the corresponding PARAFAC model consisting of ***F(I+J+K)*** parameters, for a hypothetical array with a dimension of 20x150x20 modeled by a 6 component solution. The PCA model of 6(20 x 3000) array will be unfolded into 360,000 parameters, while the same data set can be modelled by PARAFAC using only 1140 parameters. The resulting PCA loading plots can be misleading and difficult to interpret (Bro 1997).

This is due to the way the multiway data is concatenated into a 2D matrices, where the effect of one variable is no longer associated with one element but with many within the loading vector (termed co-linearity). The primary aim of using multiway method over conventional PCA is not only to obtained a better fit of the data but rather to have a more robust and interpretable model which is less sensitive to noise and give loadings which can be related directly to the different modes of the multiway array. A multiway object with a dimensional of ***(IxJxK)*** is modelled by PARAFAC as the summation over the number of PARAFAC component ***(R)*** outer product of the triads of vectors, where the triad is the tensor product of the three vectors (Burdick 1995).

For each $\boldsymbol{X_k}$ in $\boldsymbol{X}$, each slab within the multiway data (e.g., all the data points within particular time point) PARAFAC decomposes into a product of two loading matrices:

$$\boldsymbol{X}_k = \boldsymbol{F} \cdot D_k \cdot \boldsymbol{A}^{\mathrm{T}} + \varepsilon_k$$

Where $\boldsymbol{F}$ is the loading matrix of the row vector and $\boldsymbol{A}$ is the loading matrix in the column vector, $\boldsymbol{D_k}$ is the scalar weight for the $\boldsymbol{X_k}$ slab (where a slab is 2D array) and $\boldsymbol{\varepsilon_k}$ is the residual error forthe $\boldsymbol{X_k}$ slab. However, PARAFAC requires the size of each

slab to be the same and in practice this cannot always be the case due to sample loss for various reasons even with a well balanced experimental design. Therefore PARAFAC2 was developed in order to cope with such problems and allows one dimension, normally the row vector, to be unequal between different slabs. Instead of having one global loading for all slabs as in PARAFAC1, PARAFAC2 gives $\boldsymbol{K}$ loadings matrices, where each slab has its own individual loading vector as show below. Furthermore each slab may have different row vectors which match the original size of $\boldsymbol{X}$ within the multiway data. Using PARAFAC2 minor sample loss can be tolerated without the need to trim the multiway data so that each slab is of equal size.

$$\boldsymbol{X}_k = \boldsymbol{F}_k \cdot D_k \cdot \boldsymbol{A}^{\mathrm{T}} + \varepsilon_k$$

Within PARAFAC nomenclature the row vector is term the mode 1 and the column vector is mode 2 and the diagonal matrix $\boldsymbol{D_k}$ containing the weights of all the slab is mode 3. PARAFAC models the multiway object ($\boldsymbol{X}$) by minimising the residual error ($\boldsymbol{\varepsilon}_k$) by the use of alternating least square optimization. By examining the loadings of mode 1 the relative distributions of the samples within a particular time point or experiment can be revealed and the loading of mode 2 permits the identification of potential interesting variables and the loading of mode 3 (the weight of each slab) reveals the global change of the samples with respect to the time.

1.6. Supervised data analysis

Supervised data analysis involves the use of a classifier (discriminate function) to generate a rule of classification based on a sample set with a known class membership (training set). The rules are then tested by applying it to unknown samples to asses and evaluate the classifier's predictive ability. In supervised learning the aim is to find a model which will correctly associate the inputs (data) with the target (output) by minimising the error between the known target and the model response. Classifiers such as partial least square (PLS), artificial neural networks (ANNs) and support vector machines (SVMs); these are briefly discuss in the following sections.

1.6.1 Partial least squares (PLS)

PLS is a well known method that has been widely applied in metabolomics. The classifier has certain properties that make it useful for predictive purposes. In PLS regression it is possible to obtain a score matrix that is related to both the instrumental response (***X***) and the concentrations (***c***). In addition, PLS takes into account errors in both the concentration estimates (***c***) and the spectral data (***X***) and does not assume they are error free and that these errors are equally distributed in both the ***X*** and the ***c*** blocks.

PLS was originally first developed by Wold and Martens (Martens and Naes 1989; Wold 1989). The algorithm attempts to relate two types of variables ***X*** block and ***c*** block (where ***X*** represents the instrumental response and ***c*** is the labelling information).
Rather than just modelling exclusively on ***X*** variables, two set of equations are obtained as follow in PLS1

$$\boldsymbol{X} = \boldsymbol{T}\,\boldsymbol{P} + \boldsymbol{E}$$

$$\boldsymbol{c} = \boldsymbol{T}\boldsymbol{q} + \boldsymbol{f}$$

T is the score and ***P*** is the loading and ***E*** is the error, the product of the ***T*** and ***P*** will approximate to spectral data and the product of ***T*** and ***q*** will approximate to the concentration, therefore the common link between the two set of equation is ***T***. ***T*** and ***P*** for PLS are different from those to the ***T*** and ***P*** obtained in PCA, for each component within the dataset, a unique set of ***T*** and ***P*** is generated, therefore if there are 5 compounds within the data set, there will be 5 different sets of ***T***, ***P*** and ***f***, one for each component.

The scores are orthogonal but the loadings are not. In addition, the loadings are not normalized and therefore the sums of squares of each loading vector does not equal to one. The magnitude of the PLS components can be defined by multiplying the sums of squares of both ***t***a and ***p***a together as in PCA. This feature has a property whereby the sum of values of the PLS components for all non-zero components adds up to the sums of squares of the original data set. Therefore it accounts for all the total explained variance of the original data set. The value of each PLS component

does not necessary decrease as successive component are calculated. In PLS it not only try to model the ***X*** data, but also compromise between ***X*** and ***c*** block for regression (Brereton 2003).

1.6.2 Artificial neural network (ANN)

ANN has been extensively applied in numerous different scientific disciplines for non-linear classification and calibration purposes. Extensive and detailed descriptions of ANNs are available in the literature (Smits 1994; Beavis 2000). ANNs are a biologically inspired classifiers consisting of many processing sub units that function similarly to that of the biological neuron.

They receive a signal or stimuli, process it and decide whether or not to produced a response that is then passed to other cells. In ANN the neurons are replaced with a simple processing unit (called node) which can take a numerical input and transform it into an output (this is typically can be achieved by summation followed by a sigmoidal squashing function). The manner in which those processing units are connected and organized themselves is very similar to that of the human brain and they do exhibit similar characteristics (e.g., they can learn from example generalising from past experience and apply it to new ones).

Much of the inspiration for the development of ANN stems from the desire to produce an artificial system which would be capable of intelligent computation similar to the human brain. Multi layer perceptron (MLP) is one of the most widely applied forms of ANN; this is briefly described in following section. One of the reason why MLP is so attractive and widely used, is that it has been shown that a MLP network consisting of only a single hidden layer with a large number of nodes are able to learn and generalise any arbitrary non-linear problem to suitable degree of accuracy (Funahashi 1989; Hornik 1990; White 1990).
The structure of MLP consists of an input layer connected to the output layers via a hidden layer (where the hidden layer may have one or more layers). Each node of the input layer is connected to the node of the hidden layer through abstract connection termed synapse. Each synapse has am associative weight (***wi***) which scales the input (***ii***) passing through them, along with a bias term with a modifiable weight.

The nodes in the hidden layer(s) sum the signal feeding to the synapse and scale it using a squashing function along with the bias term so that it is between 0 and 1, these values are then fed to the output node within the output layer. The output node then sums up signals feeding to it from the hidden layer and also applies a sigmoidal activation function, the results are then feed to the outside world (Beavis 2000). For classification purposes in ANN a sigmodal squashing function is generally used whereas in calibration a summation function is employed instead. Training of MLP typically uses a standard back propagation (BP) algorithm (Werbos 1994), as the input is applied to the network, it is allowed to run until an output is produced at each output node. The difference between the target and the actual output over the entire training set are then fed back through the network in the reverse direction modifying each weight as they go. The process is iteratively repeated until a suitable level of error is obtained, training a MLP network is a very inefficient process as all the connections are adjusted simultaneously and the number of nodes and layers need to be specified before training. This process is very time consuming and computationally intensive to optimised and it is prone to over-training resulting in decreased predictive ability.

1.6.3. Support vector machine (SVM)

SVM is a form of kernel method originally developed in the mid 1990s (Boser 1992; Vapnik 1995) and it has since been widely accepted and applied in different applications for analysing non-linearly separable classification problems. This is due to the fact that many of the problems normally associated with non-linear classifiers such as ANN have been largely resolved by SVM. With fewer parameters to optimised, less prone to over-training and noisy data, and with significant improvement in computational efficiency SVMs are an attractive alternative to ANNs. SVM works by transforming a non-linearly separable problem into a linearly separable one by the use of a suitable mapping function (termed a kernel function), after transformation a linear solution can be found within the feature space defined by the mapping function.

The transformation of using a mapping function allows the determination of a non-linear boundary to be found in a hyperplanar boundary which separates the classes in

a (generally) higher dimension feature space defined by the mapping function. In addition, the optimal hyperplanar boundary should be equidistant from the outer edges of respective each class which maximised the margin between them. SVM therefore focuses on the data points which closest to the optimal separation boundary (so called support vectors) and ignores those that are already well separated. Thus the separation boundary is only based on a subset of samples rather than the whole data set.

SVM can be divided into two categories: hard and soft margin SVMs. Hard margin SVM assumes two classes to be perfectly separable and the goal is to determine the optimal boundary that exactly separates the two classes with a maximum possible margin between them, however this is not always be the case. By contrast, soft margin SVM are able to tolerate a certain degree of misclassifications and are designed to find a compromise between the complexity of the model and generalisation performance of the model; the lower the tolerance for rate of classification error, the more complex the hyperplanar boundary will become and this in turn may lower the generalisation performance of the model (Brereton 2003; Xu 2006).

1.7 Research objectives

The aim of this thesis is to investigate the ability of using BVOC analysis as a profiling tool within metabolomics. It will be necessary to trap the volatile components prior to MS and once this is established and shown to be reproducible this will be used for bacterial discrimination, foodstuff quality assessment, and non-invasive personalised health care applications. The data generated will be complex and so it is also the aim to develop suitable robust data analysis methodologies. The application of SERS and SERRS in rapid food screening analysis was also investigated and this also involved the use of chemometrics analyses. The primary work conducted over the course of this thesis is presented in chapter two to seven, with a final chapter discussing the findings.

Chapter 2

Discrimination of bacteria using pyrolysis gas chromatography-differential mobility spectrometry (Py GC-DMS) and chemometrics

William Cheung, Yun Xu, Christopher. L. P. Thomas and Royston Goodacre.

Author contribution to the work presented within the thesis

William Cheung: Design and implementation of the overall experiment, construction and validation of the hyphenated Py-GC/DMS system. All subsequent maintenance and repairs to the system. Culturing, harvesting and purification of all biological samples and analysis. Data pre-processing and initial unsupervised data analysis.

Dr Yun Xu: Development of the standard operating protocol (SOP) for data analysis: correlated optimised warping for alignment, asymmetric least square for baseline correction and subsequent supervised analysis and validation using PLS-DA.

Discrimination of bacteria using pyrolysis-gas chromatography-differential mobility spectrometry (Py-GC-DMS) and chemometrics

William Cheung,[ab] Yu Xu,[a] C. L. Paul Thomas[bc] and Royston Goodacre*[a]

Received 23rd July 2008, Accepted 25th November 2008
First published as an Advance Article on the web 16th December 2008
DOI: 10.1039/b812666f

Discrimination of bacteria was investigated using pyrolysis-gas chromatography-differential mobility spectrometry (Py-GC-DMS). Three strains belonging to the genus *Bacillus* were investigated and these included two strains of *Bacillus subtilis* and a single *Bacillus megaterium*. These were chosen so as to evaluate the possibility of bacterial strain discrimination using Py-GC-DMS. The instrument was constructed in-house and the long-term reproducibility of the instrument was evaluated over a period of 60 days using a Scotch whisky quality control. To assess the reproducibility further each bacterium was cultured six times and each culture was analysed in replicate to give three analytical replicates. The DMS data were generated in both positive and negative modes, and the data in each mode were analysed independently of each other. The Py-GC-DMS data were pre-processed *via* correlation optimised warping (COW) and asymmetric least square (ALS) to align the DMS chromatograms and to remove any unavoidable baseline shifts, prior to normalisation. Processed chromatograms were analysed using principal component analysis (PCA) followed by supervised learning methodology using partial least squares for discriminant analysis (PLS-DA). It was found that the separations between *B. subtilis* and *B. megaterium* can be readily observed by PCA; however, strain discrimination within the two *B. subtilis* was only possible using supervised learning. As multiple biological replicates were analysed an exhaustive splitting of the training and test sets was undertaken and this allowed correct classification rates (CCRs) to be assessed for the 3375 test sets. It was found that with PLS-DA the negative ion mode DMS data were more discriminatory than the positive mode data.

Introduction

In just about every area of microbiology the rapid identification of bacteria is desirable. For example, being able to identify a pathogen from a patient admitted into hospital would allow targeted antimicrobial therapy and accurate epidemiology studies to be conducted. Physicochemical methods are constantly being investigated and these have focussed mainly on vibrational spectroscopy- and mass spectrometry-based techniques.[1–5] For the latter technique a variety of sample introduction and ionisation methods have been employed and these included fast atom bombardment, pyrolysis, matrix assisted laser desorption ionisation and electrospray ionisation[1,6–9]

In general, MS is vacuum-based which has implications for portability of the instrument. By contrast, differential mobility spectrometry (DMS) is a gaseous phase ionic separation technique operating at ambient pressure, where the separation of ions is achieved by exploiting the difference in the ion mobilities between alternating high and low electric fields within the DMS drift cell.[10–16] During the last decade DMS has been primarily employed for detecting volatile organic compounds (VOCs). The low power consumption, compactness of DMS, coupled to ambient pressure operation with minimal maintenance makes it an attractive alternative to MS for VOC analysis where portability is required. Recently, DMS has been successfully coupled with GC for the analysis of human breath, bacterial odours and for jet fuel analysis.[11–17]

Eiceman *et al.*[11,12,14,15] have extensively studied the suitability of using GC-DMS as an alternative method to MS for bacterial characterisation where the non-volatile bacterial components are introduced into the GC using pyrolysis; a method that has been routinely coupled to MS.[1,7,18] In a series of experiments Eiceman's group has shown that biomarkers can be discovered which are specific to sporulated *Bacillus* and these included crotonic acid (a pyrolysis product of 3-hydroxybutyric acid) from *Bacillus megaterium*[11] and pyridine for *Bacillus subtilis*.[15] Pyridine is a pyrolysis product from dipicolinic acid which is found within bacterial spores and readily identified using Py-MS.[1,3] In addition, these authors have explored the pyrolysis conditions used[14] and found that these are consistent with those adopted for Py-GC-MS.[3] Finally, they have investigated the phenotypic changes that bacteria undergo when cultured at different temperatures, and have shown that the Py-GC-DMS signal is dependent on the organism's phenotype.[16] However, to date,

[a]*School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, UK M1 7DN. E-mail: roy.goodacre@manchester.ac.uk*
[b]*School of Chemical Engineering and Analytical Science, The University of Manchester, P.O. Box 88, Sackville St., Manchester, UK M60 1QD*
[c]*Department of Chemistry, Loughborough University, Loughborough, Leicestershire, UK LE11 3TU*

bacterial discrimination at the sub-species (*i.e.* strain) level has not been successfully reported.

Therefore the purpose of the present study was to investigate whether it is possible to distinguish between different bacteria at the strain level using Py-GC-DMS. Three strains from the genus *Bacillus* were selected, and these included two different strains from *B. subtilis* and *B. megaterium* as a closely related but phylogenetically different species. In order, to assess spectral reproducibility each strain was cultured six times and three analytical replicates recorded from each culture. Chemometric analysis was used to assess reproducibility and the ability to classify these bacteria.

Methods and materials

Culture and harvesting methodology

Three stains belonging to the genus *Bacillus* were studied; these included *B. subtilis* B0014, *B. subtilis* B1382 and *B. megaterium* B0010.[1,19] These strains were cultivated axenically on LabM blood agar base plates at 37 °C for 16 h. Growth was performed independently six times to generate six biological replicates per strain. This was because we are measuring the phenotype of the organism with Py-GC-DMS and since the phenotype = genotype + environment we need to control the latter else the former will be variable and the ability to characterise the different bacteria impaired, a phenomenon noted by all whole organism fingerprinting methods.[1,7,11] After incubation the vegetative biomass was carefully collected using a sterile plastic loop and suspended in 1 mL of physiological saline (0.9% NaCl in H_2O). The bacterial suspensions were centrifuged at 15 871 *g* for 3 min, the supernatants were then discarded, and the pellets were then resuspended in 1 mL of saline solution and centrifuged for 3 min; this process was repeated twice to remove any medium components from the agar. In the final resuspension the biomass concentration was adjusted to lie within an optical density of 2.5–2.7 AU at 600 nm (Biomate 5, Thermo Electron Corporation). As the analytical equipment was housed in a category 1 environment the resulting bacterial pellets were sterilised by autoclaving at 70 °C at 4 psi for 45 min. These were then stored at −80 °C until analysed.

In order to compensate for any systematic drift it is essential that the analysis was randomized by injections rather than by batches (biological specimens). Each bacterial injection was followed by a system blank to check for instrumental and/or environmental artefacts. In addition, after every three bacterial injections, a QC sample (see below) was run to assess the reproducibility and performance of the system. Sample sizes of 1.5 and 2 μL injections were chosen for the bacterial and QC samples respectively, this is equivalent to 3.9 and 2 mg of dry matter.

Quality control (QC) samples

In addition to the bacterial samples as described above, a number of QC samples were also analysed in order to monitor the performance of the system with respect to time. This was to determine whether or not there was any systematic drift within the instrumental response over the course of the experiment. The QC was an in-house whisky mix, 10 mL of a single malt whisky (Glen Moray Classic, Elgin, Spreyside, Scotland) was rotary evaporated down to dryness and left under a vacuum system (5×10^{-2} Torr) overnight to ensured the removal of all volatiles. After this the brown 'slurry' residue was then dissolved in 2 mL of ethanol (analytical grade; Fischer Scientific).

This QC sample had a complex matrix, which upon pyrolysis produced a complicated mixture of pyrolysates that can be used to assess instrumental drift; whisky has been used for this application before and yielded complicated Py-GC and Py-MS spectra.[20,21] The data obtained from these QC samples were treated using the same methodology that is described in the data analysis section below. Principal component analysis (PCA) was also performed on these QC samples to assess whether there was any systematic drift that correlated with injection time.

Instrumentation

As shown in Fig. 1 a CDS 5200 analytical pyrolysis unit (CDS Analytix Ltd., Unit 9 Seaview Workshop, Timber Rd, Horden, Peterlee, Durham, UK) was connected to a HP5890 gas chromatography unit (Mass Spec UK, Regal House, Highfield St., Oldham, UK) *via* a digitally controlled heated transfer line (an insulated silco steel coated inner core) into the front injector port A. The GC was also modified to enable a DMS unit (Sionex® Corp, 8-A Preston Court, Bedford, MA, USA) to be fitted directly to the existing flame ionisation detection housing using an annular heat pipe (30 cm × 0.635 cm inner diameter) maintained at 170 °C.

Pyrolysis unit. A Pt coil pyroprobe and quartz fire tube were used with the walls of the pyrolysis chamber maintained at 150 °C to minimise condensation of the pyrolysate. The pyrolysis chamber was continuously purged with He (5 mL min^{-1}) prior to sealing the chamber. The pyrolysis chamber was sealed and allowed to equilibrate for 60 s when the internal temperature reached 150 °C. Next the chamber temperature was increased from 150 to 300 °C at a rate of 20 °C ms^{-1}. After equilibration the temperature of the Pt coil pyroprobe was increased to 530 °C at a rate of 20 °C ms^{-1}. Pyrolysis took place for 5 s. During this sequence the Py-GC transfer line was maintained at 300 °C with a flow rate of 1 mL min^{-1} of He.

GC. The front injector port A was maintained at 300 °C with a split ratio of 10 : 1. The analytical column was a Restek RTx 5 Sil MS analytical column (30 m × 0.25 mm × 0.25 μm film thickness, with a stationary phase composition of 5% diphenyl/95% dimethylsiloxane). The He flow rate was maintained at approximately 1 mL min^{-1}. The analytical column was connected directly to the DMS unit through an annular heat pipe with the following temperature program: initial temperature: 60 °C (held for 2 min); then increased to 280 °C at a rate of 8 °C min^{-1}; the final temperature of 280 °C was held for 2 min.

DMS. The DMS unit was a Sionex SVAC-1 unit (Sionex® Corp) maintained at 100 °C with a N_2 flow rate of 270 mL min^{-1}, see Table 1 for the instrument parameters. Dispersion field programming methodology was used which increased the maximum dispersion field strength from 10 to 26 kV cm^{-1}. The

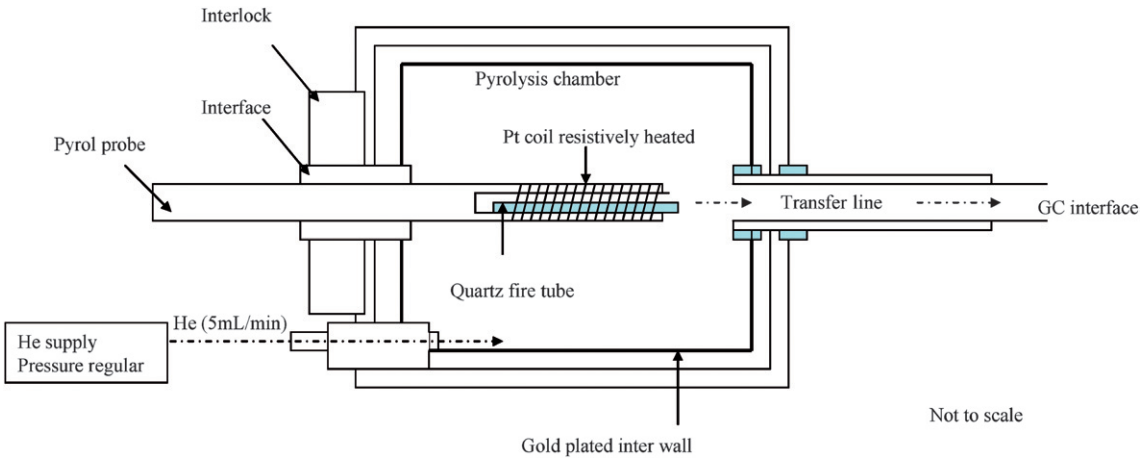


Fig. 1 Schematic diagram of the pyrol probe and pyrolysis chamber interface.

Table 1 DMS settings and parameters

Starting CV/V	−42
End CV/V	15
No. of steps	100
Step duration/mS	10
Step settle time/mS	3
Step to blank	1
Positive gain	High
Negative gain	High
RF step size/V	1
RF steps	0
CV step/V	0.56565

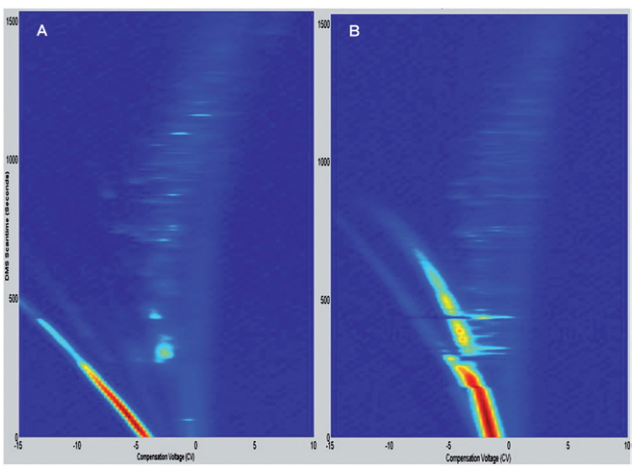


Fig. 2 DMS responses from *B. subtilis* B0014 in the (A) negative and (B) positive modes.

rationale for adopting this methodology has been reported recently.[17] Typical DMS profiles for *B. subtilis* are shown in Fig. 2.

Data analysis

The overall scheme used for data pre-processing and data analysis is detailed in Fig. 3. For each strain there were six biological

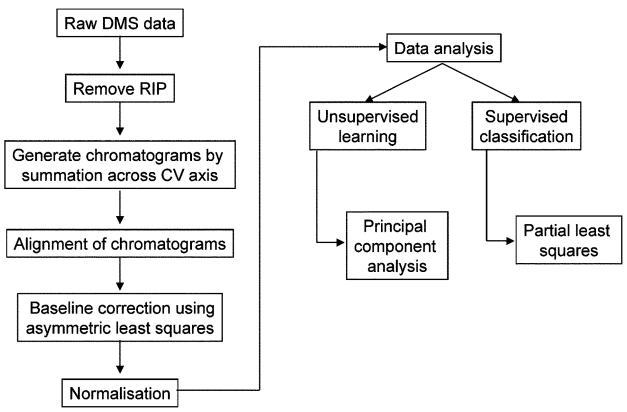


Fig. 3 Flow diagram summarising the data pre-processing (LHS) and data analysis (RHS) methodology.

replicates and each biological replicate was analysed three times (machine/analytical replicates) creating 18 samples per strain. Therefore the data matrix analysed consisted of three strains, containing 54 Py-GC-DMS spectra in total.

Signal processing. The unprocessed DMS data were generated in Microsoft Excel Worksheet format. The resultant files were catalogued into biological, and then analytical replicates, with the positive and negative modes data-processed separately. Preliminary visual inspection of the data enabled dominating and potentially non-reproducible features to be identified and we chose to exclude the reactant ion peak (RIP) from the all Py-GC-DMS data. The original DMS matrix sizes were as follows: the voltage compensation was from −15 to +10 CV and DMS scan time was 0–1534 s^{-1}. After RIP removal the cropped matrix for negative mode included the GC eluent from 9 to 22.5 min retention time and the DMS ranged from −6 to +6 CV with a 500–1350 s^{-1} scan time. Whilst for the positive mode the GC included the retention times from 10 to 23.3 min and the DMS was from −5 to +5 CV and the scan time was 600–1400 s^{-1}. The data were then summed across the compensation voltage (CV) axis producing two DMS chromatograms; for the negative mode this was summed from −5 to +5 CV, and −6 to +6 CV for the

positive mode data. For consistency all Py-GC-DMS data were pre-processed in an identical manner.

The chromatograms were aligned by using correlation optimised warping (COW[22]) to correct the drifting of the peaks between different Py-GC-DMS runs. The segment size was set to 10% of the number of the data points within each chromatogram, and the slack variable was set to 10% of the segment size (*e.g.* as the Py-GC-DMS chromatogram contained 500 data points the segment size was therefore set to 50 and the slack variable was set to 5).

After COW alignment the baseline was corrected by using asymmetric least square (ALS[23]), an adaptive baseline estimation algorithm. Finally, the chromatograms were normalised by using min–max normalisation, whereby each chromatogram was divided by the absolute difference between the maximal and minimal intensities.

Pattern recognition. MATLAB version 2007a (MathWorks, Nantwich, USA) was employed for the data analysis. Principal component analysis (PCA)[24–27] was performed on the processed chromatograms (detailed in Fig. 3) containing 500 data points. PCA is an unsupervised vector space transform method, often used to reduce high dimensional data sets to lower dimensions for analysis, *e.g.* modelling or visualisation. The variance of the data set was captured by a few latent variables called principal components (PCs). The variance captured by each PC is in a nested fashion. The first PC always captures the largest variance of the whole data set and the second PC captures the largest variance of the residues (the unexplained variance of the previous PC) and so on for the later PCs. The data set was column centred before the PCA. The scores of the first few PCs were plotted against each other in order to visualise any natural clustering trends within the data sets. After PCA, the next step was to perform a supervised classification upon the data set to investigate if it was possible to discriminate between the three bacterial strains. Unlike PCA, which is an unsupervised method, supervised classification attempts to build a predictive model based on a subset of samples with known origin (training set). If there are sufficient chemical differences between the bacteria that are detected by Py-GC-DMS the model should be able to predict the class membership of unknown samples. The accuracy of such prediction was assessed by using an independent data set (test set) which was not used during the training stage. In this study, we used partial least squares-discriminant analysis (PLS-DA)[24,26–29] as the supervised classifier.

PLS-DA is a commonly used supervised classification method which is based on a well known regression model called partial least squares or project to latent structure (PLS). Similar to PCA, PLS is also a latent variable-based model but in a supervised manner. Instead of finding a smaller set of latent variables capturing as many variations as possible, PLS finds a linear model describing some predicted variables (*e.g.* concentration levels, class membership, *etc.*) in terms of a set of other observable variables, *e.g.* the Py-GC-DMS data in our case. Similar to PCA, the observable variables were also 'compressed' into a few latent variables, called PLS components, and the fundamental relations between the predicted variables and observable variables were established based on the PLS components. PLS can model one predicted variable, which is usually called a PLS1 model, while it can also model several predicted variables simultaneously, which is usually called a PLS2 model. Although it was originally designed as a regression model, various applications as well as some theoretical studies have proved that it can be a very effective classification model.[24,26–29] For two-class separation problems, both PLS1 and PLS2 models can be employed, while for multiple class classification problems, in general a PLS2 model should be used (although it is also possible to combine multiple PLS1 models). In this study, there are three classes to be separated; hence PLS2 modelling was used. The class memberships of the samples were represented by a Y matrix with three columns and each column corresponds to one different class. Binary encoding was used such that class 1 would be encoded as 1, 0, 0, class 2 as 0, 1, 0, and class 3 as 0, 0, 1.

The PLS model was built on the training set and the number of significant PLS components were determined by using k-fold cross-validation while k is the number of biological replicates (see below). This model was then applied to the test set and the class membership of each sample was determined by using the procedure described by Wu *et al.*[30] For each sample, the predicted vector of $\boldsymbol{y}$ was calculated. The sample was assigned to the class for which the predicted value is the only one higher than 0.5. For instance, if the predicted $\boldsymbol{y}$ is [0.1, 0.9, 0.2], the sample is assigned to class 2. When the prediction is, for example, [0.1, 0.9, 0.8] or [0.1, 0.4, 0.2], the sample was regarded as a misclassified sample, *i.e.* the class membership of that sample cannot be confidently determined.

The training set was created by selecting 66.7% (2/3) of the samples from the original data set for training and the remaining 1/3 were used as the test set. During training set selection each of three machine replicates per biological replicate were considered as a 'whole' rather than as independent samples, the sample selection must account for this else one is merely measuring the reproducibility of the analytical instrument rather than the biological differences. In our case, there were six biological replicates for each strain, so four were used for training and two were used for testing. Since the number of different combinations of selecting four samples out of six is 15, this splitting of training and test set procedure had been repeated exhaustively $15^3 = 3375$ times. The predictions were averaged to give an estimation of the expected accuracy of the classification model. The reason for such exhaustive resampling is to minimise the influence of selecting samples for training or testing on the final outcome of the classification models and avoid the chance that seemingly good results were in fact due to a 'lucky' set of samples being chosen as the training set and another 'lucky' set of samples being chosen as the test set. If the separation is genuine, it should be relatively insensitive to the splitting of training and test set and the results of these models should be similar to each other. There might be some 'fortunate' or 'unfortunate' occasions which yield extremely good or poor results, but such circumstances should be rare.

Results and discussion

In order to show that Py-GC-DMS could be a useful analytical approach for the characterisation and identification of bacteria we designed a robust experiment where multiple biological replicates of three bacteria (*viz.* two strains of *B. subtilis* and one

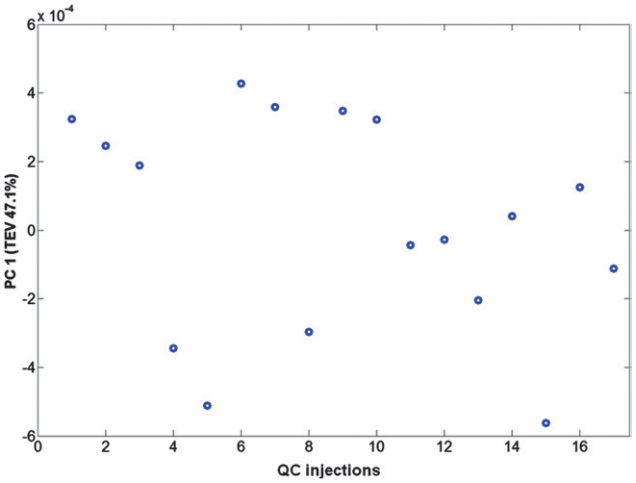


Fig. 4 Plot of injection number for the QC standards *versus* the first principal component score (total explained variance = 47.1%). This shows that there was no systemic time-related trend within the Py-DMS data.

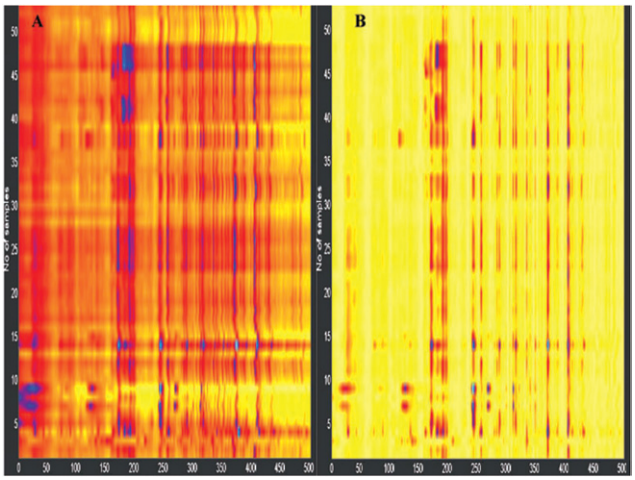


Fig. 5 (A) Chromatograms before alignment, and (B) after alignment using the COW algorithm. In this two-step alignment process the (i) global alignment uses one replicate from each of the three machine replicates for each sample ($n = 18$) aligned to the global mean, followed by (ii) a local second alignment where the remaining two chromatograms were aligned to the above alignment for each of the 18 groups. In the figure the *x*-axis represents the chromatograms (DMS scantime) and the *y*-axis the 54 chromatograms (3 strains × 6 biological replicates × 3 machine replicates).

B. megaterium) were analysed in triplicate using Py-GC-DMS over a period of 60 days. During this analysis analytical blanks were collected and used to assess artefacts in terms of carryover of pyrolysate from one sample to another and inspection of these blanks showed that there were no 'memory effects' observed (data not shown). In addition, after three bacterial analyses a QC sample from dried whisky was analysed so as to assess reproducibility during data acquisition. Following the pre-processing as detailed in Fig. 3 and above the QC samples were analysed using PCA. A plot of the injection number (which is relative to time) *versus* the first PC (which contained 47.1% total explained variance) is shown in Fig. 4 where no correlation with respect to sample injection time is seen, and the same was true for other PCs as well (data not shown). This result suggests that as no systematic trends can be observed in PCA space that the instrument was running in a stable and reproducible manner. This gave us confidence that the analysis of the bacteria by the same system would not be overtly influenced by any analytical artefacts.

Following data collection from the bacterial samples the Py-GC-DMS data were processed as detailed above (and Fig. 3). Initially the Py-GC-DMS data (Fig. 2 for examples) were summed across the CV axis after RIP removal and aligned using a two-step COW alignment. The data before and after the results of the COW alignment are shown in Fig. 5, where it can be clearly seen that the peak drifting is significantly reduced after the alignment, indicating the utility of running this step in the pre-processing sequence of procedures.

Once aligned the data were baseline corrected using ALS and were normalised to min–max per chromatogram. These data were then ready for chemometric analysis. The initial stage of the data analysis strategy was to use unsupervised exploratory data analysis and PCA was employed to discover any natural groups within the data and is also a useful algorithm for discovering any outliers. The results of the PCA are shown in Fig. 6 where it can be seen that both negative and positive modes generated similar trends and there was an obvious separation between *B. megaterium* and the two strains of *B. subtilis* in both PC1 and PC2 for both positive mode and negative mode. However, no obvious separation between the two strains of *B. subtilis* (B0014 *vs.* B1382) can be observed, also for both ionisation modes. Whilst all three bacteria belong to the *Bacillus* genus, in terms of phylogenetics *B. megaterium* is different from *B. subtilis* at the 16S rDNA sequence level,[31] and these genotypic differences are manifest in the organism's phenotype which is what we are measuring using Py-GC-DMS. By contrast the two strains of *B. subtilis* are very closely related and so using this unsupervised learning algorithm cannot be separated; indeed a finding we have previously observed with Raman spectroscopy[31,32] but not electrospray ionisation mass spectrometry,[19] presumably because the latter generates analyte specific information. This therefore presents a true test for Py-GC-DMS. The question arises as to whether there are any significant differences which lie in less obvious variance of the Py-GC-DMS data that can be discovered using supervised learning.

Therefore we chose to use a supervised classifier PLS-DA which was programmed as described above. As it was trained with training data – *i.e.* data from Py-GC-DMS from known bacterial origin – it is important that the classification model is tested independently and we used a third of the data as the independent hold out test set. These training and test sets were selected from the Py-GC-DMS profiles in an exhaustive fashion. As detailed above, this resulted in 3375 splits of the data. The 3375 predictions for the test set were then averaged to give an estimation of the expected accuracy of the classification model.

In most cases, three PLS components appeared to be optimal *via k*-fold cross-validation. The average correct classification rates (CCRs) on the test set along with their standard deviation as well as the minimum and maximum of all 3375 runs are shown in Table 2. In addition to the CCR, the averaged confusion matrices are also shown in Table 3. The results suggested that the

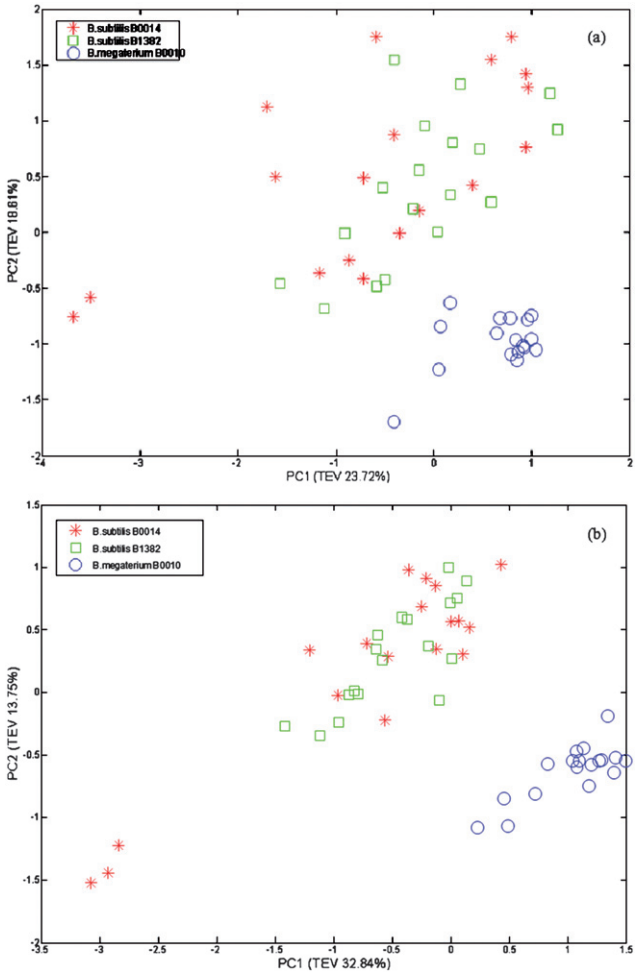


Fig. 6 PCA scores plot of PC1 *vs.* PC2: (a) negative mode; (b) positive mode. See inset legend for which symbols refer to *B. subtilis* strains B0014 and B1382 and *B. megaterium* B0010.

Table 2 Summary of the correct classification rate (CCR) for PLS-DA

	Mean CCR	Standard deviation	Minimum CCR	Maxmum CCR
PLS-DA				
Positive mode	91.90%	0.073%	61.00%	100.00%
Negative mode	97.39%	0.036%	77.78%	100.00%

positive prediction accuracy of *B. megaterium* is 100% and so consistently better than those of the two strains of *B. subtilis*. For the data obtained using the negative mode data from the DMS, the error rate of the prediction of *B. megaterium* is also always 0 for all 3375 runs, and for positive mode data *B. subtilis* B1382 is rarely mis-identified as *B. megaterium*.

Inspection of the confusion matrix indicates that by using PLS-DA it is possible to separate the two strains of *B. subtilis* with very high accuracy. The prediction accuracies are higher in the negative mode where 98.83% on average for *B. subtilis* B0014 and 95.50% average for *B. subtilis* B1382 are predicted. This is particularly encouraging as in PCA (Fig. 6) there were no

Table 3 Confusion matrices for the supervised classification algorithm

	Predicted strain		
Actual strain	*B. subtilis* B0014	*B. subtilis* B1382	*B. megaterium* B0010
PLS-DA on positive mode data			
B. subtilis B0014	85.83%	14.17%	0.00
B. subtilis B1382	9.67%	89.83%	0.50%
B. megaterium B0010	0.00	0.00	100.00%
PLS-DA on negative mode data			
B. subtilis B0014	98.83%	1.17%	0.00
B. subtilis B1382	4.50%	95.50%	0.00
B. megaterium B0010	0.00	0.00	100.00%

obvious separations between these two *B. subtilis* strains; by contrast PLS-DA suggests that there is enough information in the Py-GC-DMS data to allow supervised classification methods to separate these two strains.

It is also evident in Table 2 and Table 3 that the negative mode data yielded better prediction accuracy of the bacterial class than those of the positive mode data.

Conclusions

It has been successfully demonstrated that discrimination at the bacterial strain level is possible using Py-GC-DMS as the analytical technique, but only when coupled with supervised learning methods. The separations between different species can be readily observed by PCA; however, strain discrimination requires more powerful chemometric methods using supervised classifiers such as PLS-DA.

In addition, the data analysis suggests that the positive mode data contained less strain-specific information than the negative mode data in DMS (Table 2 and Table 3) and this is likely to be due to the difference in ionisation chemistry with negative reactant ions, and such a finding is consistent with the work carried out by other researchers.[11] However, rather than disregard the chemical information obtained from the positive mode this should be viewed as an additional orthogonal response to that of the negative mode, and in the future we shall investigate this further as the positive mode in DMS also offers the added option of derivatization to be employed to obtain greater chemical information.

In conclusion, we believe that Py-GC-DMS presents itself as a complementary analytical approach for the rapid characterisation of bacteria, and this is the first study to apply advanced chemometrics for the separation of bacteria at the sub-species level and will be investigated further.

Acknowledgements

Y. X. and R. G. are grateful to the European Commission's BIOTRACER Integrated Project (www.biotracer.org) for financial support. We are grateful to Sionex Corporation for funding W. C. and technical support and maintenance of the DMS SVAC-1 unit, and Dr Ian Fleet for his guidance and support.

References

1 R. Goodacre, B. Shann, R. J. Gilbert, E. M. Timmins, A. C. McGovern, B. K. Alsberg, D. B. Kell and N. A. Logan, *Anal. Chem.*, 2000, **72**, 119–127.
2 D. I. Ellis, D. Broadhurst, D. B. Kell, J. J. Rowland and R. Goodacre, *Appl. Environ. Microbiol.*, 2002, **68**, 2822–2828.
3 A. P. Snyder, J. P. Dworzanski, A. Tripathi, W. M. Maswaden and C. H. Wick, *Anal. Chem.*, 2004, **76**, 6492–6499.
4 J. P. Dworzanski, A. Tripathi, A. P. Snyder, W. M. Maswadeh and C. H. Wick, *J. Anal. Appl. Pyrolysis*, 2005, **73**, 29–38.
5 R. Jarvis, S. Clarke and R. Goodacre, *Topics in Applied Physics*, 2006, **103**, 397–408.
6 B. Michael, R. S. Bordoli, R. D. Sedgwick and A. N. Tyler, *J. Chem. Soc., Chem. Commun.*, 1981, 325–327.
7 H. L. C. Meuzelaar, J. Haverkamp and F. D. Hileman, *Pyrolysis mass spectrometry of recent and fossil biomaterials*, Elsevier, Amsterdam, 1982.
8 S. Vaidyanathan, D. B. Kell and R. Goodacre, *J. Am. Soc. Mass Spectrom.*, 2002, **13**, 118–128.
9 S. Vaidyanathan, D. Jones, D. I. Broadhurst, J. Ellis, T. Jenkins, W. B. Dunn, A. Hayes, N. Burton, S. G. Oliver, D. B. Kell and R. Goodacre, *Metabolomics*, 2005, **1**, 243–250.
10 I. A. Buryakov, E. V. Krylov, A. L. Makas, E. G. Nazarov, V. V. Pervukhin and U. K. H. Rasulev, *J. Anal. Chem.*, 1993, **48**, 114–121.
11 S. Schmidt, F. K. Tadjimukhamedov, K. M. Douglas, S. Prasad, G. B. Smith and G. A. Eiceman, *Anal. Chem.*, 2004, **76**, 5208–5217.
12 P. H. Rearden and P. Harrington, *Hyphenated separations*, LabPlus international, February/March, 2006.
13 P. Rearden, P. B. Harrington, J. J. Karnes and C. E. Bunker, *Anal. Chem.*, 2007, **79**, 1485–1491.
14 S. Schmidt, F. K. Tadjimukhamedov, K. M. Douglas, S. Prasad, G. B. Smith and G. A. Eiceman, *J. Anal. Appl. Pyrolysis*, 2006, **76**, 161–168.
15 S. Prasad, H. Schmidt, P. Lampen, M. Wang, R. Guth, J. Rao, G. B. Smith and G. A. Eiceman, *Analyst*, 2006, **131**, 1216–1225.
16 S. Prasad, K. M. Pierce, H. Schmidt, J. V. Rao, R. Robert Guth, S. Bader, R. E. Synovec, G. B. Smith and G. A. Eiceman, *Analyst*, 2007, **132**, 1031–1039.
17 M. Basanta, D. Singha, S. Fowler, I. Wilson, R. Dennis and C. L. P. Thomas, *J. Chromatogr., A*, 2007, **1173**, 129–138.
18 E. M. Timmins and R. Goodacre, in *Identification of Microorganisms by Mass Spectrometry*, ed. C. L. Wilkins, J. O. Lay and J. D. Winefordner, John Wiley & Sons, New Jersey, 2006, pp. 319–343.
19 S. Vaidyanathan, J. J. Rowland, D. B. Kell and R. Goodacre, *Anal. Chem.*, 2001, **73**, 4134–4144.
20 K. J. G. Reid, J. S. Swan and C. S. Gutteridge, *J. Anal. Appl. Pyrolysis*, 1993, **25**, 49–62.
21 R. I. Aylott, A. H. Clyne, A. P. Fox and D. A. Walker, *Analyst*, 1994, **119**, 1741–1746.
22 N. P. Vest Nielsen, J. M. Carstensen and J. Smedegarrd, *J. Chromatogr., A*, 1998, **805**, 17–35.
23 H. F. M. Boelens, R. J. Djikstra, P. H. C. Eilers, F. Fitzpatrick and J. A. Westerhuis, *J. Chromatogr., A*, 2004, **1057**, 21–30.
24 H. Wold, in *Multivariate Analysis*, ed. P. R. Krishnaiah, Academic Press, New York, 1966, pp. 391–420.
25 R. C. Beavis, S. M. Colby, R. Goodacre, P. Harrington, J. P. Reilly, S. Sokolow and C. W. Wilerson, in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers, John Wiley & Sons, Chichester, 2000, pp. 11558–11597.
26 R. G. Brereton, *Chemometrics Data Analysis for the Laboratory and Chemical Plants*, Wiley, Chichester, 2002.
27 Y. Xu, *Chemometrics pattern recognition with applications to Genetics and Metabolomics data*, PhD Thesis, University of Bristol, 2006.
28 B. K. Alsberg, D. B. Kell and R. Goodacre, *Anal. Chem.*, 1998, **70**, 4126–4413.
29 S. J. Dixon, Y. Xu, R. G. Brereton, H. A. Soini, M. V. Novotny, E. Oberzaucher, K. Grammer and D. J. Penn, *Chemom. Intell. Lab. Syst.*, 2007, **87**, 161–172.
30 W. Wu, Q. Guo, D. Jouan-Rimbau and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1999, **45**, 39–53.
31 E. C. López-Díez and R. Goodacre, *Anal. Chem.*, 2004, **76**, 585–591.
32 R. M. Jarvis, A. Brooker and R. Goodacre, *Faraday Discuss.*, 2006, **132**, 281–292.

Chapter 3

Novel noninvasive identification of biomarkers by analytical profiling of chronic wound using volatile organic compounds

Alexis N. Thomas, Sevetlana Riazanskia, **William Cheung**, Yun Xu, Royston Goodacre, Christopher .L. P. Thomas, Mohamed S. Bagunied and Ardeshir Bayat.

Author contribution to the work presented within the thesis

Alexis Thomas: Clinical sample collection and analysis.

William Cheung: Maintenance and troubleshooting of the TD-GC/ITMS system during the course of the study, supervision and monitoring of all sampling and data collection using PDMS patch. Data pre-processing (alignment and baseline correction) and initial unsupervised data analysis.

Dr Yun Xu: Application of additional multivariate and univariate analysis.

Novel noninvasive identification of biomarkers by analytical profiling of chronic wounds using volatile organic compounds

Alexis N. Thomas, MBChB[1,2]; Svetlana Riazanskaia, PhD[3]; William Cheung, MSc[4]; Yun Xu, PhD[4]; Royston Goodacre, PhD[4]; C. L. Paul Thomas, PhD[4]; Mohamed S. Baguneid, FRCS[2]; Ardeshir Bayat, PhD[1]

1. Plastic & Reconstructive Surgery Research, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester, United Kingdom,
2. Department of Vascular Surgery, University Hospital of South Manchester NHS Foundation Trust, Wythenshawe, Manchester, United Kingdom,
3. The Centre for Instrumentation and Analytical Science, School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, United Kingdom, and
4. Laboratory for Bioanalytical Spectroscopy, School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester, United Kingdom

Reprint requests:
Ardeshir Bayat, Principal Investigator, Plastic & Reconstructive Surgery Research, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom.
Tel: +44 0161 306 5177;
Email: ardeshir.bayat@manchester.ac.uk

Manuscript received: March 22, 2009
Accepted in final form: March 26, 2010

DOI:10.1111/j.1524-475X.2010.00592.x

ABSTRACT

A complex profile of volatile organic compounds (''VOC''s) emanates from human skin, which is altered by changes in the body's metabolic or hormonal state, the external environment, and the bacterial species colonizing the skin surface. The aim of this study was to compare VOC profiles sampled from chronic leg wounds with those from asymptomatic skin. Five participants with chronic arterial leg ulcers were selected. VOC samples were obtained using polydimethylsilicone membranes (''skin-patch method'') and analyzed by gas chromatography-ion trap mass spectrometry. Resultant data were analyzed using multivariate analysis and mass spectral matches were compared against the National Institute of Standards and Technology database. Principal component analysis showed differences in profiles obtained from healthy skin and boundary areas and between profiles from healthy skin and lesion samples ($p < 0.05$). Partial least squares for discriminant analysis gave an average prediction accuracy of 73.3% ($p < 0.05$). Mass spectral matching (verified against microbial swab results) identified unique VOCs associated with each sample area, wound bacterial colonization, and ingested medications. This study showcases a reproducible, robust, noninvasive methodology that is applicable in a clinical setting and may offer a new, hitherto unexplored, class of biochemical markers underpinning the metabolism of chronic wounds.

Chronic wounds affect circa 200,000 people in the United Kingdom at any one time and of these wounds, leg ulcers are highly prevalent affecting up to 2% of the adult population.[1,2] The estimated cost of treating leg ulceration (as defined as a loss of skin that takes more than 6 weeks to heal[3]) to the UK's National Health Service amounts to £400 million–£600 million per year,[4,5] with the overall financial burden of all chronic wounds being estimated at more than £1 billion per annum.[6] Furthermore, patients with leg ulcers have reduced quality of life when compared with age-matched controls due to pain, odor, and decreased mobility.[3]

Identification of the causation of such wounds is problematic, as is ascertaining the most appropriate treatment method. Currently, patients with leg ulcers are assessed via the history, appearance of the lesion, and a vascular assessment (palpation of the pedal pulses, ankle brachial pressure index and duplex ultrasound scanning). The ability to delineate both the underlying cause and the bacterial colonization of such lesions via a noninvasive technique would be helpful in understanding the disease process, aiding appropriate treatment selection (possibly mitigating against the emergence of multiresistant organisms via correct antibiotic choice). The noninvasive technique used in this study is based on the collection and analysis of volatile organic compound (''VOC'') signatures emitted from leg ulcers.

VOC is a generic term used to classify a wide range of molecules with a boiling point of $\leq 300\,^{\circ}C$, for example, alcohols, aldehydes, ketones, isocyanates, sulfides, and hydrocarbons.[7] The human body is known to contain and emit a large number of these substances as essential nutrients and intermediates, waste products of endogenous processes, from the absorption of environmental contaminants, and also via exogenous bacterial metabolism. To date, these compounds have been detected from skin (sebum, sweat, skin emanations, and hair), breath, serum, urine, saliva, cerebrospinal fluid, feces, breast milk, semen, amniotic fluid, and tissue homogenates.[7,8] The VOCs emitted change with the body's metabolic or

NA	Nonadhesive
PC	Principal component
PCA	Principal component analysis
PLS-DA	Partial least square for discriminant analysis
VOC	Volatile organic compound

hormonal state[9] (for example, the metabolic changes associated with diabetes cause the release of acetone on breath[10]), with ingested dietary compounds,[11] by variations in the external environment (VOC emissions from the skin alter diurnally and seasonally[12]), and by alterations in the bacterial species colonizing the skin surface.[12–14]

Aside from the underlying causative factors that result in the formation of ulcers, another important aspect in their chronicity is concurrent bacterial colonization or infection of the ulcerated area.[15] Wounds are usually colonized by the commensal skin flora but pathogenic bacterial species are also commonly implicated.[16,17] Colonization of a wound is not itself a barrier to wound healing[15]; however, repeated infections have been shown to cause increased proinflammatory cytokines and matrix metallopeptidases, decreased tissue inhibition of these peptidases and decreased levels of growth factors—these changes have been hypothesized to be the underlying causes of wound chronicity.[18]

Historically, the degree and/or speed of wound healing were thought to be related to the bacterial count. A study of pressure ulcers showed that significant healing only occurred when the bacterial count was $< 10^6\,mL^{-1}$ bacteria.[19] More recent research indicates that while bacterial density at the wound surface is independently predictive of nonhealing, this is overly simplistic as factors such as bacterial diversity, microbial synergistic interactions, and the underlying host response are contributory.[16,18,20]

Current clinical methods of wound infection analysis are surface swabbing and wound exudate culture, wound tissue biopsy, and the clinician's judgment of the "classic" signs of infection (pain, erythema, edema, heat, and purulence). None of these methods is ideal. Surface swabbing is difficult to undertake reproducibly and reliably[21] and causes trauma to the granulating tissue.[22] Trauma is of greater concern with tissue biopsy, still deemed the "gold standard" method for quantitative wound infection analysis, with some authors finding that the accuracy of culture results obtained is comparable with that of surface swabbing.[20,21,23] Clinicians cannot provide reproducible, consistent, and accurate assessments: purely subjective observations result in a large variation in the sensitivity of results with little interobserver reproducibility.[24,25]

Development of an accurate, noninvasive method for the analysis of chronic wound etiology, infection, and healing would be of clinical use. In this study, we propose to sample the VOC profiles of vascular leg ulcers to demonstrate a novel technique for the future analysis of not only vascular ulcers but also other types of chronic wounds.

METHODS

Participant selection

Participants were selected from a cohort of inpatients under the care of the vascular surgery department at University Hospital of South Manchester NHS Foundation Trust, Wythenshawe, Manchester, UK. The inclusion criteria were that a participant was male, of Caucasian descent, between the ages of 55 and 95 and have a leg ulcer of predominantly arterial etiology (proven by the evidence of significant arterial disease on duplex ultrasound scanning). This was necessary to minimize confounding factors associated with comparing ulcers of variable etiology. A participant was deemed to be of Caucasian descent if his parents and all grandparents were stated as Caucasian—this was relevant because ethnicity and sex are known to alter skin VOC profiles.[26–30] There were no specific exclusion criteria; however, the selected cohort was necessarily limited to five participants due to the prospective nature of the study and the time required to process and analyze the samples—this provided 50 samples, allowing the appropriate use of the chosen statistical tests. For each participant, a comprehensive history was elicited including that of the lesion to be sampled, ingested medication, and the treatment applied to the lesion both historically and before the sampling period. Also noted were any toiletries used and whether specific consumables that affect the emitted VOC profile, e.g., spicy food, coffee had been ingested during 48 hours before sampling.

Equipment preparation

The sampling methodology has been described recently and evaluated.[11] Briefly, polydimethylsilicone skin-sampling patches measuring 20 mm × 15 mm × 0.45 mm (Goodfellow Cambridge Ltd., Huntingdon, UK), were prepared by washing and conditioning at 180 °C in a vacuum oven before storage in inert thermal desorption tubes (Markes International Ltd., Rhondda Cynon Taff, UK). Before utilization, the patches were thermally desorbed to verify that they were free from contamination.

Ethics and participant preparation

This study was conducted in accordance with the ethical principles of Good Clinical Practice and the Declaration of Helsinki. The Local Research Ethics Committee (Manchester, UK) approved the protocol before commencement of the study and all subjects gave written, informed consent. Twenty-four hours before sampling, the dressing covering the participant's lesion was removed, the area was irrigated with distilled water to remove any remnants of prior dressing materials, and a sterile nonadhesive (NA) dressing (Johnson & Johnson Medical Ltd., Ascot, UK) was applied using an aseptic technique. A secondary dressing layer comprising Softform gauze and crepe bandaging was applied over the NA dressing. The same procedure was used to apply a second sterile NA dressing on a more proximal, lesion-free region of the ipsilateral leg, at least 10 cm away from the ulcer edge. This was used to collect the VOC profile of normal skin. The participant was advised to keep the dressings dry and not to use any toiletries until post-sampling 24 hours later.

Sampling procedure

Twenty-four hours after applying the dressings, the prepared thermal desorption tubes were removed from refrigeration and transported to the participant. The participant's local environment was assessed for signs of significant exogenous VOC contamination (cleaning, cooking, or other medical interventions) and other clinical staff and patients were excluded from the sampling locality to reduce exogenous sample contamination.

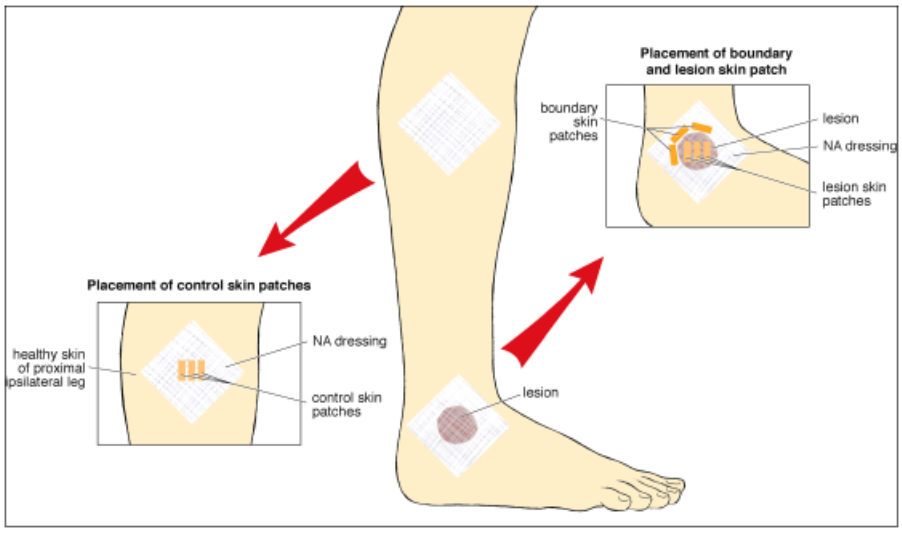


Figure 1. A diagram showing the positioning of the skin patches on a participants' foot. Patches were applied in triplicate for 30 minutes to the lesion, boundary, and control areas above the nonadhesive dressings applied previously.

Forceps used to handle the patches were sterilized with an isopropyl alcohol wipe and air-dried. A cotton-wool pad was located 1–2 m from the participant onto which a sampling patch was placed. This was exposed for 30 minutes, providing a baseline of the environmental VOCs present. Concurrently, the secondary dressings were removed and three skin patches were applied to each sampling region (Figure 1), covered with a cotton-wool pad and left in situ for 30 minutes. At the end of the sampling time, each patch was removed from the sampling site and resealed into its thermal desorption tube. The sampled area was photographed and swabbed for microbiological analysis before being redressed. The samples were placed in storage at 4 °C to await analysis within 24 hours of sample collection.

Sample analysis

Sampled VOCs were recovered from the skin patches via a two-stage thermal desorption procedure (Markes International Ltd.), separated along 60 m of a 0.25 mm diameter capillary column with a 0.25 µm thick 5% phenyl, 95% methyl polysiloxane stationary phase (DB-5MS; Agilent, West Lothian, UK) before passing into the Varian 2200 (Varian Ltd., Oxford, UK) ion trap mass spectrometer, the operating parameters of which are summarized in supporting information Table S1.

Mass spectral searching and data visualization

All chromatographic data were evaluated and checked to ensure reproducibility. The chromatograms were assessed on a peak-by-peak basis and five mass spectra were averaged from consecutive scans for all resolved peaks $> 5\%$ of the maximum peak intensity. The resultant list of mass spectra was searched against the National Institute of Standards and Technology library and provisional assignments were made to the isolated compounds. The assignments were then reviewed and those with incompatible physical chemical characteristics were removed from the list of candidate compounds and labeled as unknown. The compounds were then examined by sample type to assess the variability of the samples obtained from the different sites.

Data preparation

The gas chromatography-mass spectrometric (GC-MS) data files were converted into netCDF format using a conversion program (Palisade MASSTransit, Scientific Instrument Services, Ringoes, NJ, USA) enabling the data to be exported into the multivariate statistical processing software (Matlab, Mathworks, Natick, MA, USA) (Figure 2). Within Matlab, linear interpolation algorithms with resampling were applied to ensure that all data sets were the same size—the accumulation of small variations in the instrument control unit results in GC-MS data surfaces of different dimensions over a range of time off-sets. The resampled data were then aligned using correlated optimized warping and baseline corrected by applying asymmetric least squares. The final process was to remove those parts of the data set that did not contain chemical information—the gaps between chromatographic peaks. This was achieved by studying the chromatogram visually and identifying the level of the base line, along with the standard deviation for the baseline. A standard deviation filter was subsequently applied to the data sets and all data that were within three standard deviations of the base line were removed. The resultant chromatogram was normalized such that the total integrated peak areas of the peaks were equal to unity.[31]

Chemometric analysis

The chromatograms were analyzed for natural trends or outliers using principal component analysis (PCA) and the result of this unsupervised learning analysis was visualized by plotting the first three principal components scores (PC 1, PC 2, and PC 3).[9,32,33] Hotelling's T^2-statistics[33] were applied to the PCA scores to determine the statistical significance of the observed intersite differences.

Partial least squares for discriminant analysis (PLS-DA)[9,31–33] was then used to model the VOC data for its ability to discriminate between different sites (lesion,

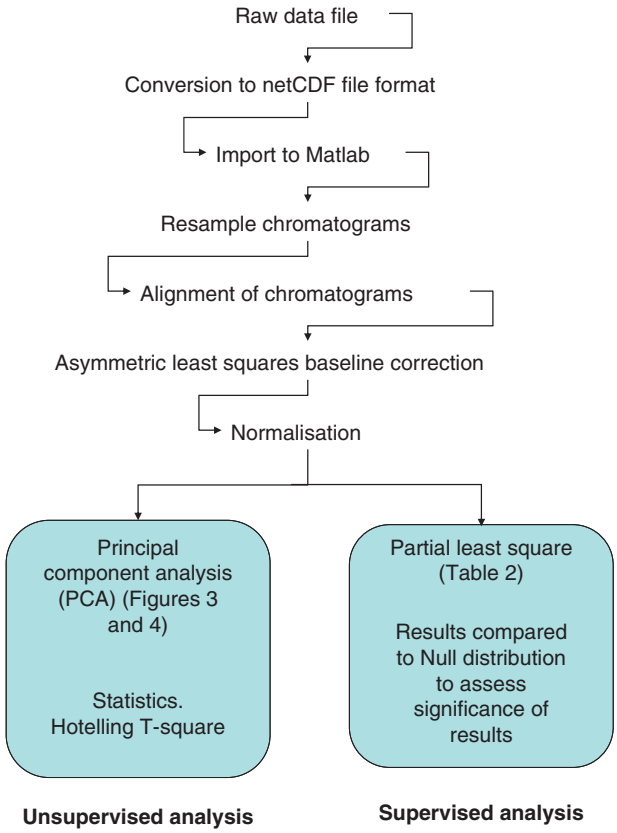


Figure 2. Data handling flow-chart.

control and boundary). The data were randomly divided into two sets: 80% was used to build a training model and the remaining 20% was used to test the accuracy of the model. In order to avoid biasing the analysis (vide infra), the selection of the training set and test set was based on participants rather than samples. That is, if the data from a participant were selected for use as part of the training set, all the other data from this participant were also used in the training set. Consequently, four out of five of the participants' data were used for training and the remaining participant's data were used for testing. This procedure was repeated five times, each time using one different individual. This minimized the influence of the characteristic individual VOC signatures that might have resulted in overoptimistic results if the training and test data contained profiles from the same participant. Because of the limited sample size, the classification results obtained were averaged and further evaluated using a permutation test.[34–36] In this test, 2,500 Monte-Carlo simulations were carried out and in each simulation, the order of labels was shuffled randomly—the same PLS-DA procedure was applied to the data set post-shuffling. The results formed the null distribution and the averaged classification result obtained as described above was compared against it to assess the significance level of the result.

RESULTS

The researchers observed no discomfort to the participants and no participant reported any discomfort arising from the procedure.

All sampled lesions were of mixed etiology: the major component was limb ischemia due to arterial occlusion, but elements of neurogenic injury and superimposed wound colonization were also present. There was heterogeneity in the character of each lesion, varying between dry and necrotic to wet and exudative. The clinical description correlated with bacterial colonization of each lesion (see Table 1 for a summary of the microbiological swab results). The organisms cultured were mainly Gram-negative aerobic bacilli, coagulase-negative staphylococci or Proteus species and coliforms. Methicillin-resistant *Staphylococcus aureus* was cultured in one patient.

Data characterization and visualization

Examples of GC-MS data obtained from lesion samples are represented in Figure 3. The complexity of the chromatogram is striking: the range of intensities of responses spans more than three orders of magnitude, with more than 300 clearly resolved signatures. There were many other nonresolved chromatographic peaks with peak shapes suggesting the presence of significant numbers of

Table 1. Microbiology swab results taken from the center of the lesion, boundary area of the lesion and control (reference) skin site

Participant	Control (reference site)	Boundary area of the lesion	Lesion
1	Mixed coagulase-negative staphylococci (mixed skin-type flora)	Methicillin-resistant *Staphylococcus aureus*	Methicillin-resistant *Staphylococci aureus*
2	No growth	No growth	Mixed coliforms and *Proteus* species
3	Mixed coagulase-negative staphylococci (mixed skin-type flora)	Mixed coagulase-negative staphylococci (mixed skin-type flora)	Mixed coagulase-negative staphylococci (mixed skin-type flora)
4	No growth	Mixed coagulase-negative staphylococci (mixed skin-type flora)	Mixed coagulase-negative staphylococci (mixed skin-type flora)
5	No growth	Mixed coagulase-negative staphylococci and enterococci	Mixed coagulase-negative staphylococci and enterococci

Rotational spot swabbing was utilized due to the limited surface area available on some sites. Unsurprisingly, all lesions were associated with bacterial growth whereas the majority of healthy control skin demonstrated no bacterial growth.

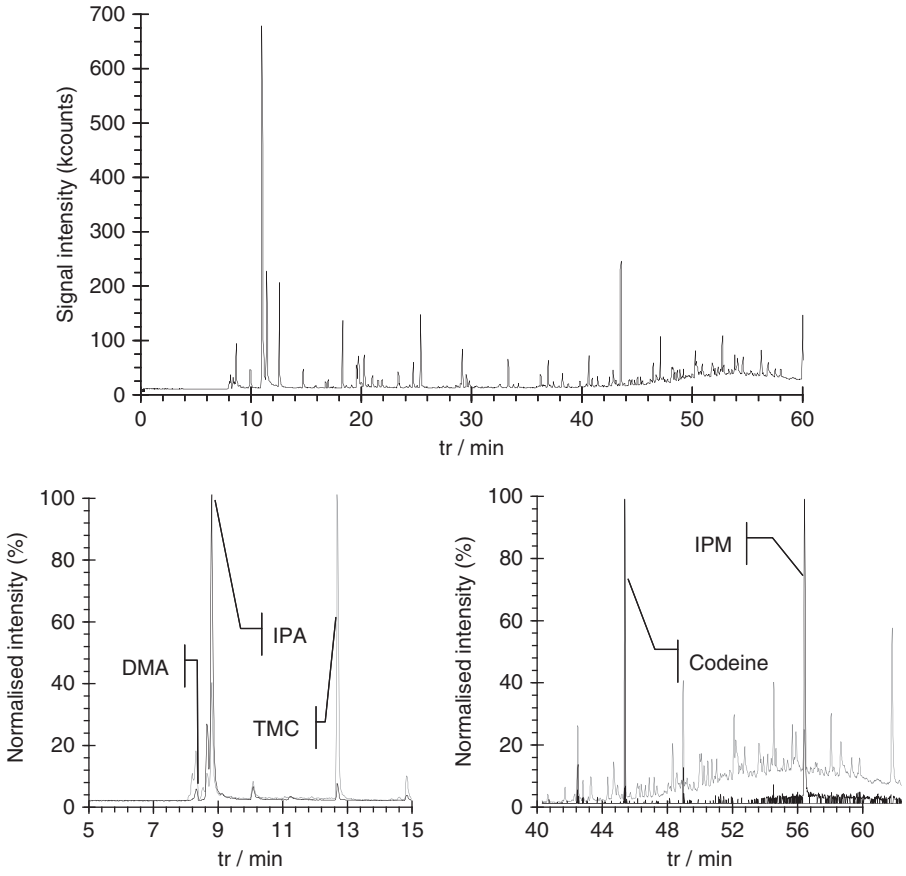


Figure 3. Examples of the gas chromatography-mass spectrometric data obtained from lesion samples. The top trace shows the total ion chromatogram while the two bottom traces show normalized plots of selected ion chromatograms (shown in black), superimposed on the total ion trace (gray lines). The selected ions were *m/z* 45 for the bottom left trace and *m/z* 299, and 228 for the bottom right trace. The complexity of the data observed within the total ion chromatogram in the top trace derives from a variety of sources (examples as follows): Endogenous metabolites—trimethylcarbazole (TMC, at a retention time [T_r] of ca. 12.7 minutes) bottom left; exogenous volatile organic compounds—isopropylalcohol from the sampling protocol (IPA, T_r=ca. 9 minutes) bottom left, isopropylmyristate from personal care products (IPM, T_r=ca. 56.3 minutes) bottom right and Codeine (T_r=ca. 44.8 minutes) bottom right; and from the bacterial activity within the lesion-dimethylamine (DMA, T_r=ca. 8.3 minutes) bottom left.

close and coeluting components. Indeed, a preliminary assessment of the data with deconvolution software (ACD/Labs IntelliXtract, Toronto, Canada) indicates the presence of many hundreds of hitherto undocumented VOCs in the skin samples. Such chemical diversity has been discussed previously[11] and arises from the underlying metabolism of the participant (affected by phenotype, diet and environment); exogenous inputs (medication, environmental contamination); compounds specifically related to the metabolism of bacteria in the lesion; and the associated pathology of tissue damage.

Compounds of varying chemical groups were recovered, including esters, alcohols, thiols, hydrocarbons, carboxylic acids, amines, amides, ketones and siloxanes. Visualization of the data resulted in the list of compounds in Table 2—this summarizes the tentative assignments of compounds that were associated exclusively with each individual sampling site. Although site-unique compounds may be discerned, it is helpful to note that the distribution of compounds across the different sites arises from a variety of mechanisms: exogenous materials may be washed out by wound exudates; VOCs generated within the lesion may diffuse into the blood stream to be released from the skin at the control site; and VOCs released from the affected area into the surrounding air may be present in the environmental control samples. Certainly, a strong odor associated with infected, necrotic tissue was frequently encountered during sampling. Thus, while the compound list in Table 2 is encouraging, the information associated with peak intensities is also important, i.e., compound abundances, and as such, chemometric approaches are required to prospect these data to identify the chemical differentiators.

Chemometric analysis

Chemometric analysis was utilized as this allows the discrimination of statistically significant differences in the VOC profiles acquired from each sample site. Figure 4 shows the results of PCA of the complete data set (comprising 50 samples). There was no separation observed between the lesion and boundary areas ($p > 0.05$). However, significant differences between the control and lesion areas and between the control and boundary areas were observed (both $p < 0.05$). In addition, it was possible to identify which chemical compounds might be responsible for such separation via comparison of the scoring plot and the corresponding loading plot (see Figure 5). Because chromatograms were used for the data analysis, each chromatographic peak (correlating to a specific chemical compound) was represented by a series of adjacent variables. From the loadings plot, the variable clusters on the extremes were mainly responsible for the separation exhibited in the scores plot whereas those close to the origin had little or no contribution to such separation. Examination of the loading plot revealed six peaks with the highest

Table 2. Tentative identification of the extracted compounds that were unique to each of the sampled sites—this list was derived from mass spectral matches against the NIST data base and is subject to further confirmation

Compounds	Molecular weight	Formula
Lesion		
2-Indazol-2-ylphenylamine	209	$C_{13}H_{11}N_3$
1,4-Methanoazulen-3-ol decahydro-1,5,5,8a-tetramethyl-, [1s-(1.α.,3.β.,3a.β.,4.α.,8a.β.)]-	222	$C_{15}H_{26}O$
Z,Z-2,5-Pentadecadien-1-ol	224	$C_{15}H_{28}O$
E-2-Methyl-3-tetradecen-1-ol acetate	268	$C_{17}H_{32}O_2$
Pentanoic acid, 2,2,4-trimethyl-3-carboxyisopropyl, isobutyl ester	286	$C_{16}H_{30}O_4$
Octadecane, 1-chloro-	288	$C_{18}H_{37}Cl$
2,6-Nonadienoic acid, 7-ethyl-9-(3-ethyl-3-methyloxiranyl)-3-methyl-, methyl ester, [2R-[2.α.(2E,6E),3.α.]-	294	$C_{18}H_{30}O_3$
Z-3-Octadecen-1-ol acetate	310	$C_{20}H_{38}O_2$
.α.-Ethylether of 11-epi-dihydroartemisinin	312	$C_{17}H_{28}O_5$
4-Trifluoroacetoxypentadecane	324	$C_{17}H_{31}F_3O_2$
1,2-Benzenedicarboxylic acid, butyl 2-ethylhexyl ester	334	$C_{20}H_{30}O_4$
Boundary area		
2-Propanol, 1-(2-methoxy-1-methylethoxy)-	148	$C_7H_{16}O_3$
2-t-Butyl-5-propyl-[1,3]dioxolan-4-one	186	$C_{10}H_{18}O_3$
3-Decen-1-ol, (E)-	156	$C_{10}H_{20}O$
E-2-Tetradecen-1-ol	212	$C_{14}H_{28}O$
1-Dodecanol, 3,7,11-trimethyl-	228	$C_{15}H_{32}O$
Myristic acid, 9-hexadecenyl ester, (Z)-	450	$C_{30}H_{58}O_2$
Reference skin site (nonlesion)		
(S)-(+)-1,2-Propanediol	76	$C_3H_8O_2$
1,3-Pentanediol, 2,2,4-trimethyl-	146	$C_8H_{18}O_2$
Cyclodecanol	156	$C_{10}H_{20}O$
5,9-Undecadien-2-one, 6,10-dimethyl-, (E)-	194	$C_{13}H_{22}O$
Tetradecanal	212	$C_{14}H_{28}O$
1-Decanol, 2-hexyl-	242	$C_{16}H_{34}O$
1-Nonadecene	266	$C_{19}H_{38}$
1-Eicosanol	298	$C_{20}H_{42}O$
4-Trifluoroacetoxytetradecane	310	$C_{16}H_{29}F_3O_2$
Octadecanoic acid, 4-hydroxy-, methyl ester	314	$C_{19}H_{38}O_3$
Cyclopropaneoctanoic acid, 2-[(2-pentylcyclopropyl)methyl]-, methyl ester, trans,trans-	322	$C_{21}H_{38}O_2$
2-Trifluoroacetoxypentadecane	324	$C_{17}H_{31}F_3O_2$
Dodecane, 1,2-dibromo-	326	$C_{12}H_{24}Br_2$
Background compounds (field blanks)		
Nonanal	142	$C_9H_{18}O$
Cyclodecanol	156	$C_{10}H_{20}O$
7-Tetradecene	196	$C_{14}H_{28}$
1-Propyl-3,6-diazahomoadamantan-9-ol	210	$C_{12}H_{22}N_2O$
2-Methyl-Z-4-tetradecene	210	$C_{15}H_{30}$
Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	216	$C_{12}H_{24}O_3$
Butylated hydroxytoluene	220	$C_{15}H_{24}O$
2-Hexadecanol	242	$C_{16}H_{34}O$
3-tert-butyl-5-chloro-2-hydroxybenzophenone	288	$C_{17}H_{17}ClO_2$
3-Benzoylmethyl-3-hydroxy-5-nitro-2-indolinone	326	$C_{17}H_{14}N_2O_5$
1,3,5-Tris(trimethylsiloxy)benzene	342	$C_{15}H_{30}O_3Si_3$
3-Isopropyl-6a,10b-dimethyl-8-(2-oxo-2-phenyl-ethyl)-dodecahydro-benzo[f]chromen-7-one	396	$C_{26}H_{36}O_3$

NIST, National Institute of Standards and Technology.

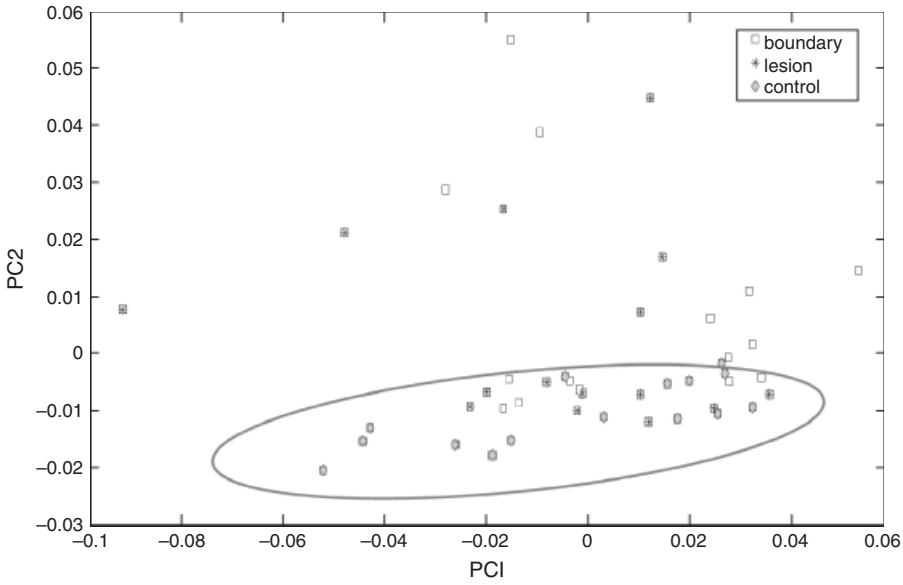


Figure 4. Principal component analysis (PCA) scores biplot of gas chromatography mass spectrometric data from lesion (denoted by asterisk symbol), boundary (denoted by square symbol), and control (denoted by diamond symbol) areas where principal component (PC) 1 and PC 2 exhibit 62.08 and 21.29% of the total explained variance (TEV), respectively. The first two PC scores in PCA were plotted against each other in order to look for underlying trends within the data set, where no obvious patterns were observed for the boundary and lesion areas. However, the reference (control) areas were observed to be localized as a broad cluster at the lower left hand side of the PCA plot (encircled in gray).

variability (see Table 3 for the tentative assignments of these peaks)—the compounds identified in this pilot study are often associated with preservatives typically used in the formulation of creams and gels. However, as the experimental protocol restricted the use of such creams and gels during the experiment and none of the patients reported using such products, the source of these compounds is unclear.

Analysis clearly shows that participant-specific "fingerprints" were identified (see supporting information Table S1). Therefore in the supervised classification, the selected training and test sets had to be based on participants rather than samples in order not to create overoptimistic results. The data from the lesion and boundary samples were combined into a single class, termed the affected class, as the PCA scoring plot showed no clear separation between these samples. A predictive model based on PLS-DA supervised classification was formulated, which gave an average prediction accuracy of 73.3%. When this was compared with the null distribution that showed an average prediction accuracy of 51.1% and when only 122 out of 2,500 simulations ($\sim$4.9%) obtained a better prediction accuracy, it was concluded that our classification results were significant to a confidence level of 95% ($p < 0.05$)

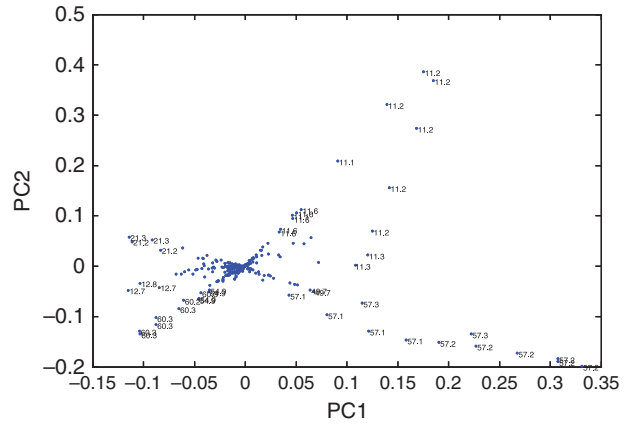


Figure 5. Principal component analysis loadings plot of the reduced chromatographic data. Data points that lay close to origin (zero), have little or no contributions toward separations shown in the scores plot, whereas points that lay further away from the origin (zero) have more significant contributions toward the separations. Because the separation between control and lesion/boundary most appeared in principal component (PC) 2, the variables that show a large diversity in PC 2 are more likely to be the chemicals that differentiate these classes. The numbers represent retention times of the variables, which correspond to one or multiple peaks observed in chromatograms.

Table 3. Tentative identification of unique compounds recovered from the loading plot of the reduced chromatographic data—these compounds were not unique to a single sample area but significantly discriminate between the areas

T_r/minutes	Compound	Molecular weight	Formula
11.20	2-propanol, 1-(1-methyethoxy),	118	$C_6H_{14}O_2$
11.63	Disulfide, dimethyl	94	$C_2H_6S_2$
20.9	3-Carene	136	$C_{10}H_{16}$
21.2	1-Hexanol, 2-ethyl-	130	$C_8H_{18}O$
43.1	Phenol,3,5-bis(1,1-dimethylethyl)-	206	$C_{14}H_{22}O$
43.2	Butylated hydroxytoluene	220	$C_{15}H_{24}O$
43.3	Straight chain hydrocarbon		

They may serve as a focus for future chronic wound VOC analysis.
VOC, volatile organic compound.

Table 4. Summary of partial least square for discriminate analysis (PLS-DA)

	Control vs. lesion/ boundary areas	Lesion vs. boundary areas
Average CCR (%)	73.3	56.67
Minimum CCR (%)	55.56	50
Maximum CCR (%)	88.89	83.33
Standard deviation (%)	14.91	14.91

(Table 4). This shows that the sampling technique generates reproducible, information-rich, complex VOC profiles which, when analyzed with chemometric methods, show statistically significant differences between the affected, i.e., lesion plus boundary areas and the control, i.e., healthy, normal skin areas. The most quantitatively prevalent of these differences were revealed by the loading plot, allowing identification of the key statistically significant differentiating VOC peaks from among the hundreds of compounds delineated through visual characterization alone.

DISCUSSION

The aim of this study was to show for the first time the utility of a novel, noninvasive sampling method for the analysis of chronic wounds and specifically chronic arterial leg ulcers. This was achieved by sampling from three separate areas (the lesion center, a control area of healthy skin and a boundary area between the two) on five individuals with chronic ulcers. VOC profile differentiation was achieved between the control and lesion areas and between the control and boundary areas (to $p < 0.05$) but not between the lesion and boundary areas ($p > 0.05$). VOCs unique to each sampling area were identified, with evidence of VOCs from ingested medication (e.g., Codeine) and secondary to the metabolic processes of colonizing bacteria (e.g., 2-indazole-2ylphenylamine). Although the VOC species cannot be defined as specific biomarkers for wound infection or healing, some may be attributed as the products of reactive oxidative stress while others may be due to preferential absorption of exogenous and ubiquitous environmental contaminants into damaged tissue. Species such as biogenic diamines and thiol compounds may also be related to bacterial degradation of the tissues. Conversely, it may be alterations in the overall VOC profile rather than quantitative analysis of specific VOC species that will give the greatest insight into chronic wound metabolic processes. Overall, it must be noted that it would be premature to definitively attribute the identified compounds to specific wound healing or degenerative processes due to the limited sample size—this would necessitate a larger, precisely described, phenotypically matched cohort of participants.

It is possible to differentiate visually between healthy control skin and diseased lesions but such a technique allows a greater understanding of the chronic wound microenvironment. Each VOC profile reflects the compounds carried to and from the skin within the blood stream, the metabolites of the underlying epidermal and dermal cellular layers, the superimposed metabolites from the normal skin bacterial flora and the environmental VOCs adsorbed at the skin surface. In the case of ulcerated regions, the profiles may be altered by the loss of normal skin cellular layers and their metabolic products and/or by the metabolites associated with the processes of necrosis, healing, and superimposed bacterial infection. Therefore, the technique of VOC profiling can provide a greater insight into the complex processes of skin infection and healing. It may, with further refinement and in conjunction with other existing modalities, become a useful method to assess the lifecycle of chronic wounds, the bioavailability of applied treatments, and in vivo bacterial antibiotic susceptibility. It will also allow for the identification of single VOCs or profiles that could act as biomarkers of specific bacterial wound infections. However, with specific reference to arterial ulceration, while this technique yields extensive chemical information regarding processes occurring at the local skin level that may be used in the assessment of infection and healing, its application in the planning of vascular interventional procedures is likely to remain adjunctive to imaging modalities.

The benefits of such a technique include its noninvasive, painless nature, its accuracy and reproducibility and the fact that it offers the possibility of rapid point-of-care testing at the bedside. At present, however, the laboratory equipment would need further optimization and miniaturization to allow this to occur. In addition, aspects of the sample processing are as yet not fully automated and thus it remains labor intensive. There are currently no VOC libraries available that are specific to chronic wounds with which to rapidly identify the sampled VOCs; certainly this will change with further development in this field of research.

This work has been successful in proving the applicability of the VOC sampling technique in a clinical setting with specific reference to the assessment of chronic wounds. The sample size investigated was adequate to allow chemometric analysis to be undertaken in order to show both the significance of the results in differentiating between VOC profiles and the reproducibility of the method. Despite this, future work would aim to extend the scope of sampling to a larger number of participants with a variety of chronic wounds in individuals of different ethnicities.

Research into the efficacy of VOC analysis in a medical setting has shown promise in a variety of conditions including lung carcinoma,[37] asthma,[38] aerodigestive tract carcinoma,[39] pulmonary tuberclosis,[40] hyperglycemia,[41] heart transplant rejection[42]), gastrointestinal disease (ulcerative colitis, *Clostridium difficile* and *Clostridium jejuni* infections),[43] bacteremia,[44] bacterial vaginosis,[45] and in the detection of an array of microbes. Few studies of wound infection have been undertaken, with those to date utilizing "electronic nose" machines[46,47] rather than GC-MS.

Chronic wounds have a major effect on both morbidity and mortality worldwide but currently are only assessed by dated, invasive methods (i.e., swabbing and biopsy). These methods do allow for an accurate identification of infecting organisms but take no account of the complex interdependent relationship between those cultured bacteria and an individual's genetic susceptibility or response to such an infection. Research has indicated that while bacterial density at the wound surface is independently predictive of wound nonhealing, this view is overly simplistic; bacterial

diversity, microbial synergistic interactions, and the underlying host response all play a role,[16,18,20] the effects of which could be assessed via VOC analysis.

The application of this technique to chronic wounds is still in its infancy but shows great promise. Future work will involve compiling a library of wound bacterial VOCs both in vitro and in vivo with the eventual identification of biomarkers or VOC fingerprints by which future noninvasive diagnosis of wound infection and etiology could be achieved. Another focus of substantial future work will be an investigation into the chronological VOC changes within healing and nonhealing chronic wounds and an analysis of the effect of ingested medications, e.g., antibiotics, on the VOC profiles of such wounds. The ultimate aim will be to minimize the equipment to allow hand-held bedside VOC analysis that would allow rapid, cheap, and painless analysis of chronic wounds and other skin conditions.

In conclusion, this preliminary study has shown how VOC profiles of chronic human skin lesions may be sampled and studied. The data processing methods applied to this study reliably differentiated control profiles from boundary profiles and control profiles from lesion profiles. The methodology has been shown to yield reproducible data from complex, previously intractable sampling environments. The development and extension of this approach may be appropriate for use in the future clinical evaluation of ulcers, wounds, and other skin lesions. An important next step in this area will be the creation of a phenotypically matched library of volatile skin metabolites.

ACKNOWLEDGMENTS

Our gratitude goes to the study participants and nursing staff of South Manchester University Hospitals NHS Trust. S.R. would like to thank Unilever Research and Development, Port Sunlight for their continued assistance throughout this research. Y.X. and R.G. are grateful to the EU project Biotracer for financial support. Final thanks must go to Ian Fleet and Eric Warburton of the Centre for Instrumentation and Analytical Science. The authors state no conflict of interest.

REFERENCES

1. Ruckley CV. Socioeconomic impact of chronic venous insufficiency and leg ulcers. *Angiology* 1997; 48: 67–9.
2. O Nelzén DB, Lindhagen A.The prevalence of chronic lower-limb ulceration has been underestimated: results of a validated population questionnaire. *Br J Surg* 1996; 83: 255–8.
3. Nelson EA, Cullum N, Jones J. Venous leg ulcers. *Clinical Evid* 2005; 13: 2507–26.
4. Simon DA, Dix FP, McCollum CN. Management of venous leg ulcers. *Br Med J* 2004; 328: 1358–62.
5. Coleridge Smith PD. Venous ulcer. *Br J Surg* 1994; 81: 1404–5.
6. Harding KG, Morris HL, Patel GK. Science, medicine, and the future: healing chronic wounds. *BMJ* 2002; 324: 160–3.
7. Zlatkis A, Brazell RS, Poole CF. The role of organic volatile profiles in clinical diagnosis. *Clin Chem* 1981; 27: 789–97.
8. Soini HA, Bruce KE, Wiesler D, David F, Sandra P, Novotny MV. Stir bar sorptive extraction: a new quantitative and comprehensive sampling technique for determination of chemical signal profiles from biological media. *J Chem Ecol* 2005; 31: 377–92.
9. Dixon SJ, Xu Y, Brereton RG, Soini HA, Novotny MV, Oberzaucher E, Grammer K, Penn DJ. Pattern recognition of gas chromatography mass spectrometry of human volatiles in sweat to distinguish the sex of subjects and determine potential discriminatory marker peaks. *Chemometr Intelligent Lab Syst* 2007; 87: 161–72.
10. Galassetti PR, Novak B, Nemet D, Rose-Gottron C, Cooper DM, Meinardi S, Newcomb R, Zaldivar F, Blake DR. Breath ethanol and acetone as indicators of serum glucose levels: an initial report. *Diabetes Technol Ther* 2005; 7: 115–23.
11. Riazanskaia SBG, Harker M, Taylor D, Thomas CLP.The analytical utility of thermally desorbed polydimethylsilicone membranes for in-vivo sampling of volatile organic compounds in and on human skin. *Analyst* 2008; 133: 1020–7.
12. Zhang ZM, Cai JJ, Ruan GH, Li GK. The study of fingerprint characteristics of the emanations from human arm skin using the original sampling system by SPME-GC/MS. *J Chromatogr B Analyt Technol Biomed Life Sci* 2005; 822: 244–52.
13. Di Natale C, Macagnano A, Paolesse R, Tarizzo E, Mantini A, D'Amico A. Human skin odor analysis by means of an electronic nose. *Sensors Actuators, B: Chem* 2000; 65: 216–9.
14. Gardner JW, Shin HW, Hines EL. Electronic nose system to diagnose illness. *Sensors Actuators, B: Chem* 2000; 70: 19–24.
15. Ryan TJ. Infection following soft tissue injury: its role in wound healing. *Curr Opin Infect Dis* 2007; 20: 124–8.
16. Bowler PG, Davies BJ. The microbiology of infected and noninfected leg ulcers. *Int J Dermatol* 1999; 38: 573–8.
17. Chesham JS, Platt DJ. Patterns of wound colonisation in patients with peripheral vascular disease. *J Infect* 1987; 15: 21–6.
18. Tarnuzzer RW, Schultz GS. Biochemical analysis of acute and chronic wound environments. *Wound Repair Regen* 1996; 4: 321–5.
19. Robson MC. Award recipient address: lessons gleaned from the sport of wound watching. *Wound Repair Regen: Official Publication of the Wound Healing Soc Eur Tissue Repair Soc* 1999; 7: 2–6.
20. Davies CE, Hill KE, Newcombe RG, Stephens P, Wilson MJ, Harding KG, Thomas DW. A prospective study of the microbiology of chronic venous leg ulcers to reevaluate the clinical predictive value of tissue biopsies and swabs. *Wound Repair Regen* 2007; 15: 17–22.
21. Gardner SE, Frantz RA, Saltzman CL, Hillis SL, Park H, Scherubel M. Diagnostic validity of three swab techniques for identifying chronic wound infection. *Wound Repair Regen* 2006; 14: 548–57.
22. Towler J. Cleansing traumatic wounds with swabs, water or saline. *J Wound Care* 2001; 10: 231–4.
23. Sullivan PK, Conner-Kerr TA, Hamilton H, Smith EP, Tefertiller C, Webb A. Assessment of wound bioburden development in a rat acute wound model: quantitative swab versus tissue biopsy. *Wounds* 2004; 16: 115–23.
24. Lorentzen HF, Gottrup F. Clinical assessment of infection in nonhealing ulcers analyzed by latent class analysis. *Wound Repair Regen* 2006; 14: 350–3.

25. Gardner SE, Frantz RA, Doebbeling BN. The validity of the clinical signs and symptoms used to identify localized chronic wound infection. *Wound Repair Regen* 2001; 9: 178–86.
26. Doty RL, Applebaum S, Zusho H, Settle RG. Sex differences in odor identification ability: a cross-cultural analysis. *Neuropsychologia* 1985; 23: 667–72.
27. Penn DJ, Oberzaucher E, Grammer K, Fischer G, Soini HA, Wiesler D, Novotny MV, Dixon SJ, Xu Y, Brereton RG. Individual and gender fingerprints in human body odour. *J Royal Soc Interf* 2007; 4: 331–40.
28. Rawlings AV. Ethnic skin types: are there differences in skin structure and function? *Int J Cosmetic Sci* 2006; 28: 79–93.
29. Zeng X-n, Leyden JJ, Lawley HJ, Sawano K, Nohara I, Preti G. Analysis of characteristic odors from human male axillae. *J Chem Ecol* 1991; 17: 1469–92.
30. Zeng X-n, Leyden JJ, Spielman AI, Preti G. Analysis of characteristic human female axillary odors: qualitative comparison to males. *J Chem Ecol* 1996; 22: 237–57.
31. Wold H. Estimation of principal components and related models by iterative least squares. In: Krishnaiaah PR, editor. *Multivariate analysis*. New York: Academic Press, 1966, pp. 391–420.
32. Alsberg BK, Kell DB, Goodacre R. Variable selection in discriminant partial least-squares analysis. *Anal Chem* 1998; 70: 4126–33.
33. Brereton RG. *Chemometrics: data analysis for the laboratory and chemical plant*. John Wiley & Sons, West Sussex, UK, 2002.
34. Golland P, Liang F, Mukherjee S, Panchenko D.Permutation tests for classification. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)2005, pp. 501–15.
35. Mielke PW Jr., Berry KJ. *Permutation methods: a distance function approach*. 1st ed. New York: Springer, 2001.
36. Westerhuis J HH, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F. Assessment of PLSDA cross validation. *Metabolomics* 2008; 4: 81–9.
37. Phillips M, Altorki N, Austin JH, Cameron RB, Cataneo RN, Greenberg J, Kloss R, Maxfield RA, Munawar MI, Pass HI, Rashid A, Rom WN, Schmitt P. Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark* 2007; 3: 95–109.
38. Dragonieri S, Schot R, Mertens BJA, Le Cessie S, Gauw SA, Spanevello A, Resta O, Willard NP, Vink TJ, Rabe KF, Bel EH, Sterk PJ. An electronic nose in the discrimination of patients with asthma and controls. *J Allergy Clin Immunol* 2007; 120: 856–62.
39. Schmutzhard J, Rieder J, Deibl M, Schwentner IM, Schmid S, Lirk P, Abraham I, Gunkel AR. Pilot study: volatile organic compounds as a diagnostic marker for head and neck tumors. *Head Neck* 2008; 30: 743–9.
40. Phillips M, Cataneo RN, Condos R, Ring Erickson GA, Greenberg J, La Bombardi V, Munawar MI, Tietje O. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis* 2007; 87: 44–52.
41. Novak BJ, Blake DR, Meinardi S, Rowland FS, Pontello A, Cooper DM, Galassetti PR. Exhaled methyl nitrate as a noninvasive marker of hyperglycemia in type 1 diabetes. *Proc National Acad Sci USA* 2007; 104: 15613–8.
42. Phillips M, Boehmer JP, Cataneo RN, Cheema T, Eisen HJ, Fallon JT, Fisher PE, Gass A, Greenberg J, Kobashigawa J, Mancini D, Rayburn B, Zucker MJ. Heart allograft rejection: detection with breath alkanes in low levels (the HARD-BALL study). *J Heart Lung Transplant* 2004; 23: 701–8.
43. Garner CE, Smith S, De Lacy Costello B, White P, Spencer R, Probert CS, Ratcliffe NM. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB J* 2007; 21: 1675–88.
44. Scotter JM, Allardyce RA, Langford VS, Hill A, Murdoch DR. The rapid evaluation of bacterial growth in blood cultures by selected ion flow tube-mass spectrometry (SIFT-MS) and comparison with the BacT/ALERT automated blood culture system. *J Microbiol Methods* 2006; 65: 628–31.
45. Chandiok S, Crawley BA, Oppenheim BA, Chadwick PR, Higgins S, Persaud KC. Screening for bacterial vaginosis: a novel application of artificial nose technology. *J Clin Pathol* 1997; 50: 790–1.
46. Parry AD, Chadwick PR, Simon D, Oppenheim B, McCollum CN. Leg ulcer odour detection identifies beta-haemolytic streptococcal infection. *J Wound Care* 1995; 4: 404–6.
47. Rimdeika R, Setkus A, Seniuliene D, Seniulis M. Semiconductor based electronic nose can identify wound pathogens—The WOUNDMONITOR project. *Burns* 2007; 33 (Suppl. 1): S72.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Summary of instrumentation parameters.

Fig. S1. Distance heat map plot. The distance between each pair of samples are represented by a colour as indicated by the colour bar on right (red to blue). The higher the distance, the lower (bluer) the similarity between two samples. The 5 different subjects, each subject has 3 samples from each class, respectively, were labeled by numbers from 1–5 and samples from the sample class (e.g., Boundary, Control or Lesion) were placed together. Since each subject has 3 samples for each class, the 3×3 blocks in the diagonal of the picture represent the similarities between the samples from the same subject and also the same class which thereby show the reproducibility of the sampling methodology.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Chapter 4

VOC-based metabolic profiling for food spoilage detection with the application to detecting *Salmonella typhimurium* contaminated pork

Yun Xu, **William Cheung**, Catherine L. Winder and Royston Goodacre.

Author contribution to the work presented within the thesis

William Cheung: Optimisation of the chromatography and analysis SOP, involved in experimental design and planning, all sample collection and analysis relating to VOC analysis. Data pre-processing (alignment, baseline correction and deconvolution) peak ID using NIST02 and Varian mass library.

Dr Catherine Winder: Help and advice on microbiology.

Dr Yun Xu: Development of deconvolution algorithms for pre-processing VOC data. Multivariate and univariate analysis.

Anal Bioanal Chem (2010) 397:2439–2449
DOI 10.1007/s00216-010-3771-z

ORIGINAL PAPER

VOC-based metabolic profiling for food spoilage detection with the application to detecting *Salmonella typhimurium*-contaminated pork

Yun Xu · William Cheung · Catherine L. Winder · Royston Goodacre

Received: 4 February 2010 / Revised: 19 April 2010 / Accepted: 21 April 2010 / Published online: 16 May 2010

Abstract In this study, we investigated the feasibility of using a novel volatile organic compound (VOC)-based metabolic profiling approach with a newly devised chemometrics methodology which combined rapid multivariate analysis on total ion currents with in-depth peak deconvolution on selected regions to characterise the spoilage progress of pork. We also tested if such approach possessed enough discriminatory information to differentiate natural spoiled pork from pork contaminated with *Salmonella typhimurium*, a food poisoning pathogen commonly recovered from pork products. Spoilage was monitored in this study over a 72-h period at 0-, 24-, 48- and 72-h time points after the artificial contamination with the salmonellae. At each time point, the VOCs from six individual pork chops were collected for spoiled vs. contaminated meat. Analysis of the VOCs was performed by gas chromatography/mass spectrometry (GC/MS). The data generated by GC/MS analysis were initially subjected to multivariate analysis using principal component analysis (PCA) and multi-block PCA. The loading plots were then used to identify regions in the chromatograms which appeared important to the separation shown in the PCA/multi-block PCA scores plot. Peak deconvolution was then performed only on those regions using a modified hierarchical multivariate curve resolution procedure for curve resolution to generate a concentration profiles matrix **C** and the corresponding pure spectra matrix **S**. Following this, the pure mass spectra (**S**) of the peaks in those region were exported to NIST 02 mass library for chemical identification. A clear separation between the two types of samples was observed from the PCA models, and after deconvolution and univariate analysis using *N*-way ANOVA, a total of 16 significant metabolites were identified which showed difference between natural spoiled pork and those contaminated with *S. typhimurium*.

Keywords VOC analysis · Pork · *Salmonella typhimurium* · Peak deconvolution · Metabolic profiling

Electronic supplementary material The online version of this article (doi:10.1007/s00216-010-3771-z) contains supplementary material, which is available to authorized users.

Y. Xu (✉) · W. Cheung · C. L. Winder · R. Goodacre
School of Chemistry, University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK
e-mail: yun.xu-2@manchester.ac.uk

C. L. Winder · R. Goodacre
Manchester Centre for Integrative Systems Biology,
Manchester Interdisciplinary Biocentre, University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK

Introduction

In recent years, non-targeted metabolite profiling-based approaches have been widely applied to many types of sample matrices like plant extractions, biofluids from animals or humans, etc. to address various problems like gene function analysis, disease diagnostic and more [1–3]. However, until now, there were few reports of using metabolite profiling on volatile organic compounds (VOCs). Compared to other types of analysis, VOC analysis has its unique advantage which is that the sampling process is normally rapid, noninvasive and can be done in situ and the collected VOCs can be readily analysed using gas chromatography with little to no sample pretreatment.

The objective of this study was to investigate the possibility of using VOC-based metabolite profiling for food spoilage detection. Conventional methods of microbial

spoilage detection are normally carried out by determining the total viable bacterial counts taken from the surface of the suspected meat; this typically requires an incubation period of 24 to 72 h for bacteria colony formation. Furthermore, this does not include additional time for selective enrichment of the sample and subsequent analysis; therefore, the total time required for the analysis of meat spoilage could be in the orders of weeks rather than days. Since food spoilage is the result of decomposition and the formation of metabolites caused by the growth and enzymatic activity of microorganisms, this is normally accompanied by the emission of unpleasant odour [4, 5]. Therefore, as the odour intensifies with the progression of the spoilage process, the analysis of VOCs potentially provides a rapid and quantifiable estimation of the level of microbial contamination. Previous studies have demonstrated a correlation between the intensity of the odour emitted from spoiled pork and viable counts. It has been reported that [6–8] at levels of 10^7 colony forming units per square centimetre (c.f.u. cm^{-2}), off-odours may become evident in the form of a faint 'dairy'-type aroma, and once the surface population of bacteria has reached 10^8 c.f.u. cm^{-2}, the supply of simple carbohydrates has been exhausted and recognisable off-odours develop, leading to what is known as 'sensory' spoilage. The development of off-odours is dependent upon the extent to which free amino acid utilisation has occurred, and these odours have been variously described as dairy/buttery/fatty/cheesy at 10^7 c.f.u. cm^{-2} through to a sickly sweet/fruity aroma at 10^8 c.f.u. cm^{-2} and finally putrid odour at 10^9 c.f.u. cm^{-2}. VOCs are readily analysed using headspace collection, and this process is noninvasive and so attractive for food sampling. Since VOCs from the spoilage process are mostly caused by the action of microorganisms, it would also be advantageous to establish a method using the analysis of VOCs which allows the discrimination between pathogenic and natural spoilage microorganisms.

In this study, we investigated such a possibility using pork as a test food matrix and *Salmonella typhimurium* as a relevant model pathogen. *S. typhimurium* is commonly associated with microbial spoilage of meat products including pork [9] and is pathogenic to humans; symptoms are gastrointestinal and infection is typified by bloody diarrhoea with mucus, fever, vomiting and abdominal cramps. The sampling of VOCs can be achieved using a variety of methods. The most popular sampling method for VOC analysis is head space sampling, and this includes dynamic head space sampling and static head space sampling [10]. Dynamic head space sampling generally offers better sensitivity, albeit requires more sophisticated equipment which makes it more suitable for the detection of analytes at low levels. Static head space sampling is more often used because of its better adaptivity and lower cost. Head space sampling can be achieved by sampling the air above the samples directly by using an airtight syringe or combined with a pre-concentration device such as solid phase micro-extraction. In this study, we employed a polydimethylsilicone (PDMS) patch-based approach reported by Riazanskaia et al. [11] as the PDMS patch can be easily fit into a glass Petri dish (see below).

Fresh pork chops were purchased from a local supermarket to avoid variation caused by different hygiene status between the pork chops. Each pork chop was butterfly cut into two pieces to provide a sterile surface for the study. The two 'matched' pieces from the same pork chop were either subjected to a natural spoilage process or contaminated with *S. typhimurium*. VOCs from each piece were sampled via head space using a PDMS patch and subsequently analysed using gas chromatography/mass spectrometry (GC/MS). Four time points over a period of 72 h were monitored, and the VOC profiles were analysed using a newly devised chemometrics methodology which combined rapid multivariate analysis on total ion currents (TICs) with in-depth peak deconvolution on selected regions and univariate statistics testing to reveal the natural trend of the whole data set and also identify the potentially interesting metabolites.

Materials and methods

Culture and chemicals

A culture of *S. typhimurium* strain 4/74, whose genome sequence is known, was kindly provided by Professor Tim Brocklehurst (Institute of Food Research, Norwich, UK). This strain was sub-cultured on Lab M LAB028 blood agar plates three times after cold storage to establish phenotypic stability. A single colony was harvested using a sterile plastic loop and transferred to 250 mL of sterile nutrient broth and incubated at 37 °C for 16 h, resulting in a suspension of 5×10^7 c.f.u./mL according to plate counts. An aliquot (5 mL, $\sim2.5\times10^8$ cells) of the suspension solution was transferred to a Falcon tube and centrifuged at $4{,}810\times g$ for 10 min. The supernatant was removed and the pellet resuspended in 50 mL of sterile physiological saline solution (0.9% NaCl) and centrifuged at $4{,}810\times g$ for 10 min. The supernatant was removed and the pellet was resuspended in 50 mL of sterile saline solution. This was repeated three times in total to remove the remaining media. The washed pellet was resuspended in 50 mL of sterile saline solution and used for the artificial contamination of the pork.

Sampling and extraction

A total of 24 boneless pork chops (weight 200–300 g) were purchased from a local supermarket. Each pork chop was

cut to a unified size of 12–14 cm^2 surface area and butterfly cut into two pieces with equivalent surface area. This provided a sterile surface for the study. A digital photo (2,048×1,536 resolution, ~3,000,000 pixels) was taken for each piece of pork alongside a ruler to enable the estimation of pixels per square centimetre and thus the surface areas in the section of pork. The butterfly piece was then placed in a large sterile glass Petri dish (Fisher Scientific, PDS-100-04L) lined with sterilised Whatman grade 40 filter paper (cat no. 90-7501-06) to which 2 mL of sterilised physiological saline solution was added to act as a moisture source to prevent the surface of the meat from drying out. The two 'matched' pieces from the same pork chop were used to produce a control (natural spoilage) and salmonellae-contaminated sample. An aliquot (1 mL) of sterile physiological saline was added to the surface of the meat and spread across the surface using a sterile plastic loop. The sample was incubated as detailed below to allow for the colonisation by the natural spoilage microorganisms. The second section of pork was contaminated with the saline suspension of *S. typhimurium* (1 mL, see above) and also spread across the surface of the meat using a sterile plastic loop. The Petri dishes were wrapped in two autoclave bags individually and sealed to minimise VOC cross-contamination and incubated at 25 °C. The headspace was sampled at a total of four time points: 0, 24, 48 and 72 h after the contamination. At each time point, six pork chops, i.e. 12 pieces, six of each type (naturally spoiled and *Salmonella*-contaminated) were sampled using PDMS patches.

PDMS patches were cut from a single silicone elastomer sheet (cat no. 751-624-16; Goodfellow Cambridge Limited) to the following dimensions, 20×15×50 mm. The patches were washed in a 5% Decon 90 solution with ultrapure water (18 MΩ) three times. The patches were then rinsed with a copious amount of ultrapure water and conditioned in a vacuum oven for 15 h at 180 °C (pressure was maintained at −1,000 mbar). Once conditioned, the patches were transferred directly into clean thermal desorption (TD) tubes (cat no. C-TBE10: Markes International Ltd.) and capped. The TD tubes were stored at room temperature within an airtight glass container filled with a layer of molecular sieves (cat no. M2635,8-12 mesh beads, Sigma). All of the tubes were used within 48 h of cleaning and conditioning, and a random selection of 20% of the TD tubes was analysed after cleaning to confirm that the batch was clean and free from contaminants before use. The patches were stuck onto the underside of the front cover of the Petri dish due to its natural adhesiveness. The samples were then resealed and placed back in the incubator for 1 h. The patches were removed and placed back into the TD tubes and analysed by GC/MS. All samples were analysed within <24 h after collection.

GC/MS analysis

A Markes International Unity 1 thermal desorption unit was connected directly through the front injector assembly of a Varian CP 3800 gas chromatograph coupled to a 2200 quadrupole ion trap mass analyzer. The Unity 1 transfer line was connected directly to the analytical column within the GC oven through the use of a Valco zero dead volume micro union (Restek, cat no. 20148). The cold trap packing material was Tenax-TA carbograph 1 TD. The transfer line to the GC was kept at 150 °C isothermally.

The analytical column was an Agilent HP-5 60 m× 0.25 mm (I.D) with a film thickness of 0.25 μm and with a stationary phase composition of 5% phenyl/95% methyl capillary column. To ensure splitless injection efficient desorption within the cold trap within the Mark 1 Unity unit, a minimum flow rate of 1.5 mL/min is required. The on-column pressure was adjusted to 85.5 kPa.

The thermal desorption profile is as follows: sample desorption 180 °C for 3 min, the cold trap kept at −10 °C. After the initial first stage sample desorption, the cold trap was then heated to 300 °C for 3 min.

The GC profile used a 70 °C initial temperature hold for 10 min; this was then ramped to 250 °C at the rate of 3 °C/min and held for 10 min. The transfer line to the MS was maintained at 270 °C isothermally. The MS was maintained at 200 °C with electron impact ionisation source at 70 eV. The mass range used was from 40 to 400 amu with a scan rate of 1.03 scan per second. Cold trap blanks and column blanks were ran after each sample to ensure that the system was free of artefacts before the next sample was analysed.

Data analysis

All the data were recorded as Varian sms files, the files were converted to netCDF files using Palisade mass transit programme (Scientific Instrument Services, Ringoes, USA), and the netCDF files were then loaded into Matlab (Mathworks, MA, USA) using the mexnc toolbox.

The procedure used for data analysis was performed as given below, and a detailed discussion about the reason of using such a procedure is explained fully in "Results and discussion". The multivariate pattern recognition was mainly conducted on the TIC. The TICs were firstly baseline-corrected using alternative least square as described by Paul et al. [12]. Alignment of the data was performed by correlation optimised warping (COW) [13]. The two parameters of COW, the number of segments and the slacking size, were optimised with a simplex optimisation procedure as described by Skov et al. [14]. For each aligned TIC, the results of the warping, i.e. the position of every segment before and after the warping, were recorded and applied to each mass channel of that chromatogram so

that not only the 1D TICs but also the 2D chromatograms were both aligned using the same alignment settings. The baseline-corrected and aligned TICs were further normalised to the surface area of each sample (see above) and $\log_{10}$-scaled. The data were $\log_{10}$-scaled because of the large difference in the peak intensities between different time points, and such scaling is necessary to prevent the multivariate analysis from being dominated by large peaks.

Principal component analysis (PCA) [15, 16] was performed on the pre-processed TICs to reveal the natural trend of the data. As previously demonstrated [17–19], the multi-block PCA models, e.g. consensus PCA (CPCA) [17, 18], are easier to interpret compared to classical PCA when there are two or more influential factors. Hence, two further CPCA models were built on two rearranged multi-block matrices to elucidate the changes caused by the two major factors of interest: (1) natural spoilage vs. salmonella-contaminated spoilage and (2) the changes over time during spoilage. A diagrammatic representation of this rearrangement of the data is shown in Fig. 1. The block loadings plots were used to identify potential interesting time windows which may contain metabolites responsible for the separation shown in the scores plot. These time windows were then selected for peak deconvolution; each time window contains 100–150 data points with 8–15 potential peaks.

On each time window, peak deconvolution was performed on each sampling time point separately due to the large difference in the peak intensities between different time points. Within the same time window, all the chromatograms at the same sampling time point were concatenated to form a matrix **X**. An orthogonal projection approach [20] coupled with Dubin–Watson statistics [21, 22] were performed to provide an initial estimation of the pure spectra matrix $\mathbf{S}_{est}$ and the number of components in the **X**. An initial curve resolution was performed on the **X** using alternating regression (AR) [23, 24] based on $\mathbf{S}_{est}$. Non-negativity and unimodality constraints were applied to concentration profiles, whilst non-negativity constraint was applied to spectra. The resolved concentration profiles matrix **C** and the pure spectra matrix **S** were visually inspected and obvious artefacts, e.g. irregular peak shape, highly variable retention time, etc., were removed. The remaining components in **C** and **S** were multiplied together to produce a reconstructed data matrix, $\mathbf{X}_{est}$, which was superimposed on the original **X** and subjected to visual inspection. The unfitted peaks were identified and the spectra at the apex of those unfitted peaks were selected and appended to **S** to generate the new estimated spectra matrix $\mathbf{S}_{est}$. Another AR curve resolution was then performed using the new *S*. This was repeated until most significant peaks were fitted and no obvious artefacts were noted in the deconvolved **C** and **S** matrices. This procedure is illustrated as a flowchart in Fig. 2. The peak area of each peak was calculated by integrating the concentration profile of each peak.

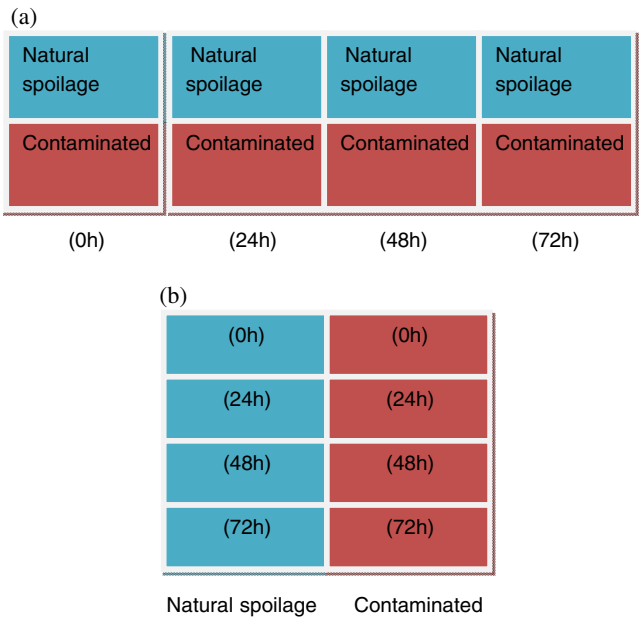


Fig. 1 Graphic illustration of the CPCA models. **a** CPCA model 1, aimed at highlighting the difference between two types of samples. **b** CPCA model 2, aimed at highlighting the changing-over-time effect whilst the spoilage progresses

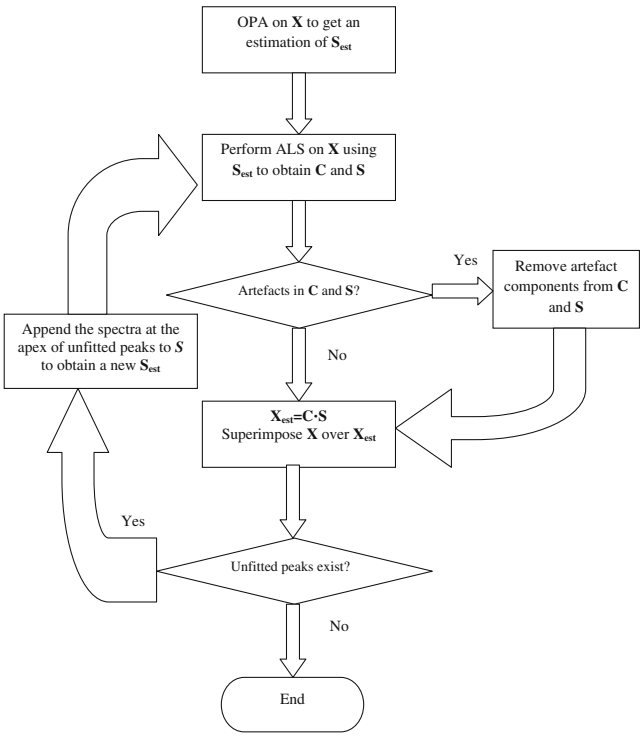


Fig. 2 Flowchart of the GC/MS peak deconvolution process

The deconvolved peaks at different sampling time points were then combined to build an overall peak table. The peaks with high similarity in the deconvolved spectrum (correlation coefficient >0.9) and also similar retention time (±1 data points, ~2 s) were considered the same peak and merged as the same variable. If in one sample no peak can be found to match the above two criteria, this peak is considered as "missing" in this sample. It is noteworthy that during the peak deconvolution, it is not possible to fit all of the peaks in all the chromatograms, and hence, it is safer to treat "not exist" as "missing" rather than an arbitrary "0" because "0" values can have very a significant impact on the statistical testing. Once the peak table was compiled, the values were log-scaled and *N*-way ANOVA (although this is a balanced two-way ANOVA experiment design, the two-way ANOVA cannot be applied directly due to the missing values) was applied to each variable to identify statistically important peaks which showed significant difference between different time points, different types of samples (naturally spoiled or *Salmonella*-contaminated) or both. The deconvolved mass spectra of the significant peaks were exported as NIST text files and imported into NIST 02 mass library for metabolite identification. A tentative chemical identification is made if a match factor (either match or reverse match) is >750 within the NIST library matching software. In our experience, a match factor >750 can generally be considered as a good match, and such a threshold has been used previously [19, 25]. In order to achieve definitive identification, it would be necessary to analyse the pure reference chemical and incorporate at least one additional orthogonal parameter (e.g. retention index) [26]. As most of these chemicals are not readily available, this was not conducted in this study.

Results and discussion

Two categories of methods are routinely employed to analyse the data generated by hyphened chromatography such as GC/MS, LC/MS or CE/MS. One is to use the TIC directly and treat it as a format of spectrum and apply multivariate methods such as PCA and PLS-DA to the TICs directly [27]. The advantage of this method is that less effort on the data pre-processing is needed and the risk of introducing artefacts caused by the data pre-processing into the data is relatively low; therefore, researchers can quickly gain insights into the data. However, the main drawback of using this approach is that the extra dimension brought by using hyphened chromatography, e.g. the mass spectrometric dimension, is completely ignored and the results provide little to no chemical information (i.e. the identification of the chromatographic peaks), which is against the very purpose of using MS compared to FID. By contrast, in another category of methods focused on performing a form of peak deconvolution on the 2D chromatograms using the information provided by the second dimension (i.e. spectrum at each time point), one can deconvolve all (or at least the most significant) the chromatographic peaks and their "pure" spectra to form a peak table with relative concentrations and then apply multivariate analysis on the peak table [24, 28]. This way, in theory, provides the most comprehensive information of the data set. Various methods have been developed during the last few decades; the latest method is called hierarchical multivariate curve resolution (H-MCR), reported by Jonsson et al. [24]. The main advantage of H-MCR is that by stacking multiple chromatograms together and assuming a trilinear data structure, the curve resolution results can be significantly improved compared to previously reported methods which apply curve resolution to each chromatogram separately. Furthermore, it is even possible to resolve co-eluting metabolites which results in fully overlapped chromatographic peaks using H-MCR because in multiple samples, the relative concentrations will be different; this is impossible for methods which deconvolve chromatograms individually. However, this type of method has a much higher requirement on the quality of the data compared to the methods which use TICs directly; else, the deconvolved results will contain artefacts and can be misleading.

In this study, we found that the results obtained by applying H-MCR were far from satisfactory (results not shown). By adopting the end criterion proposed by Jonsson et al., a large number of artefacts were found from the deconvolved peaks, and these manifested themselves as irregular-shaped peaks, whilst a large amount of peaks in the chromatograms appeared left unfitted. There were many cases found wherein when the peak requirement can no longer be met by introducing more components into the deconvolution process, there were still one or more significant peaks in the chromatograms that remained unfitted, and the extra introduced components were modelled as artefacts (e.g. irregular-shaped peaks) in the end results. We believe this is because our data employed a quadrupole mass filter which has a much lower sampling rate and also a mass resolution comparing to the more expensive time-of-flight MS used by Jonsson and colleagues.

This suggested a problem with the data structure rather than the H-MCR algorithms per se, and since many researchers to do not have access to ToF-MS and use quadrupoles as an alternative, devising a new strategy for handling such data is needed. We concluded that H-MCR was still a valid approach to peak deconvolution, but that we could not follow the semi-automatic procedure described by the Jonsson et al. Extensive manual inspection had to be performed on the deconvolved results and some adjustments had to be made on the curve resolution (e.g.

removal of artefacts, insertion of better initial estimations and re-computing the curve resolution) accordingly, and this would be a very time-consuming process given the amount of data we have collected. In addition, it is likely that it is in fact not necessary to perform full deconvolution on all the peaks from all the chromatograms as not every peak in the chromatogram will have biological significance. Therefore, we decided to combine the two categories of the methods described above:

- We firstly performed the multivariate analyses of PCA and multi-block PCA on the TICs directly to reveal the trend of data set. The potentially interesting chromatographic regions were identified through the inspections of the loadings plots.
- Next these selected regions were then subjected to a modified H-MCR procedure as described in "Data analysis" for curve resolution, and the resolved peaks were assessed using statistical test. The resolved mass spectra of those peaks being identified as significant were exported to NIST 02 mass spectra library (National Institute of Standards and Technology, MD, USA) for structure identification.

By adopting this approach, the excessive amount of time performing just curve resolution was avoided, and as reported below, we were still able to gain useful chemical insight into the data, even using a relatively cheap, routinely used GC/MS. More importantly, even though there were some imperfections in the deconvolved data, which were unavoidable, this did not have a negative influence on the results obtained from the multivariate analysis.

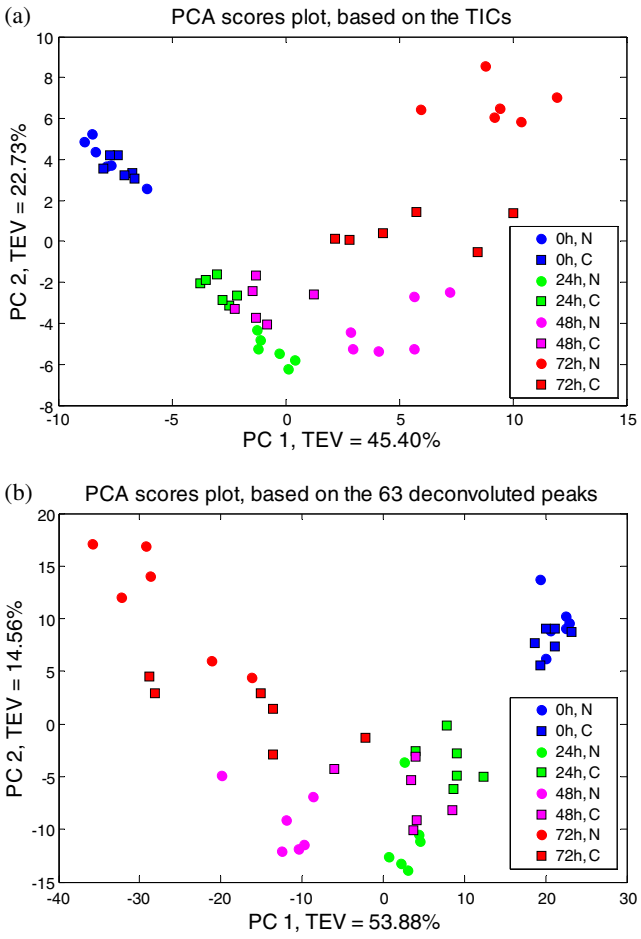


Fig. 3 PCA score plots. **a** Score plots from the PCA on TICs; **b** Score plots from the PCA on 63 deconvoluted unique peaks (VOCs). N natural spoilage, C contaminated with *S. typhimurium*, TEV = total explained variance

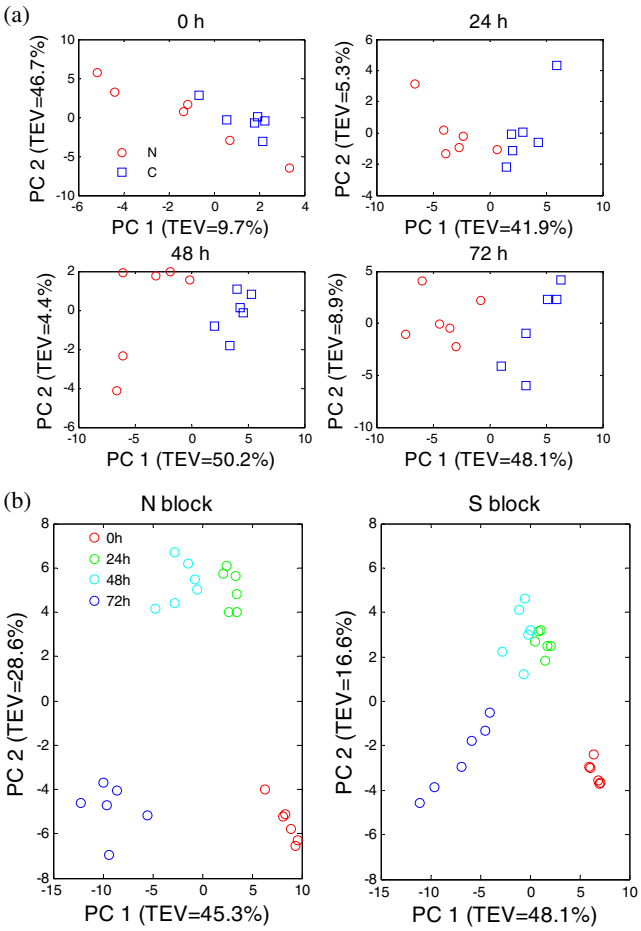
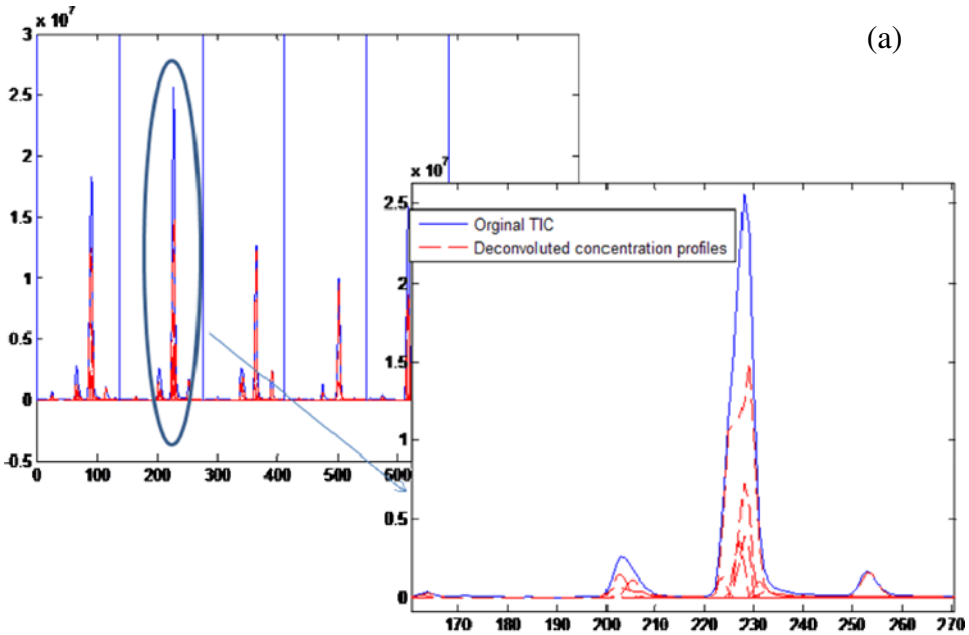
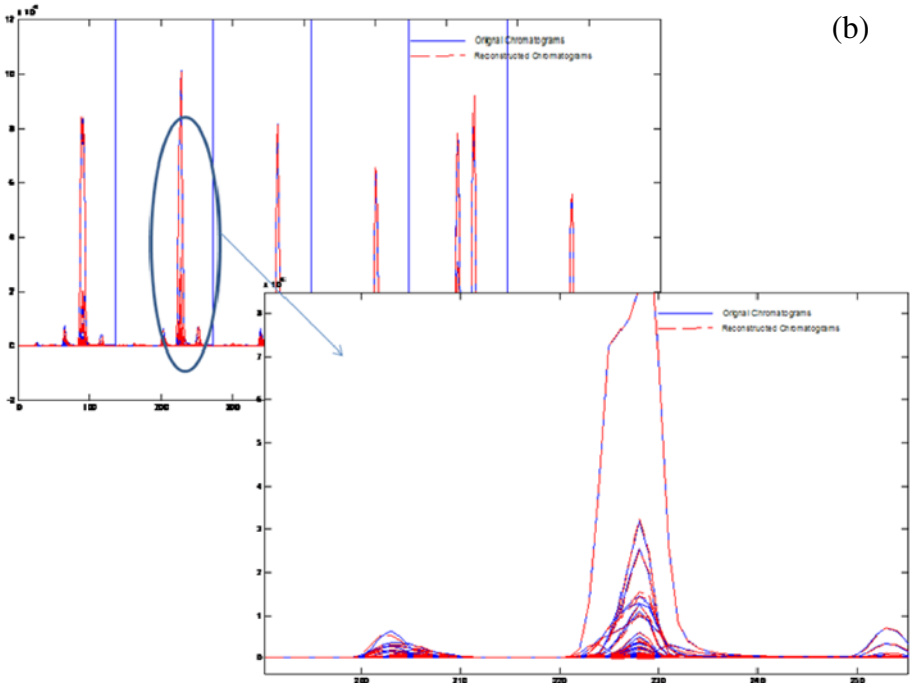
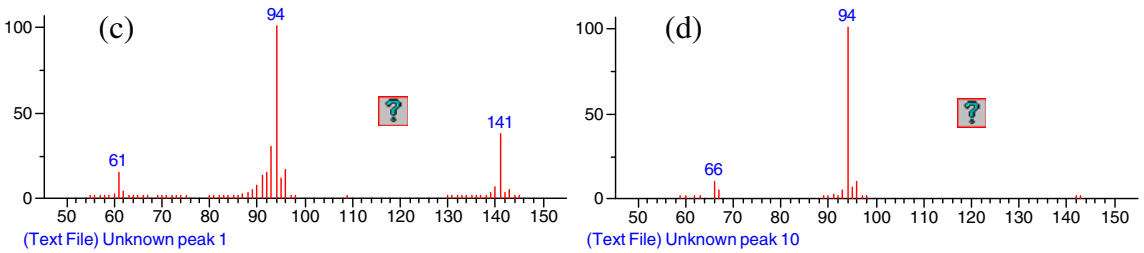


Fig. 4 Block scores plots from the two CPCA models. **a** Block score plots of the first CPCA model; *N block* naturally spoiled samples, *S block Salmonella*-contaminated samples. **b** Block scores of the second CPCA model. *N* natural spoilage, *C* contaminated with *S. typhimurium*

Fig. 5 Example of the results from peak deconvolution. **a** The ▶ deconvoluted concentration profiles superimposed on the TIC. **b** Reconstructed 2D chromatogram obtained by **C** × **S** superimposed on the original 2D chromatogram. **c** Mass spectrum of the peak before the deconvolution. **d** Mass spectrum of the peak after the deconvolution


(a)
x 10^7
Orginal TIC
Deconvoluted concentration profiles


(b)
Orginal Chromatograms
Reconstructed Chromatograms


(c)
94
61
141
(Text File) Unknown peak 1
(d)
94
66
(Text File) Unknown peak 10

As described above, the chromatograms, using TIC alone, were first analysed using classic PCA and the resulting scores plot is shown in Fig. 3a. Inspection of this plot shows that the four sampling time points were clearly separated. In addition, the separation between two types of samples (natural spoilage and salmonella-contaminated) can be observed from 24 h onwards, and seemingly, the discrimination between the sample types improves as the spoilage progresses. Such a trend is even easier to observe from the block scores in the CPCA models once one of the two major influential factors had been suppressed (in the first CPCA model, the influence of the progression of spoilage is suppressed (Fig. 4a), whilst in the second CPCA model, the influence of different type of samples is suppressed (Fig. 4b)) by rearranging the data into a multi-block matrix as illustrated in Fig. 1. In the score plots (Fig. 4a) from the first CPCA model, which highlights the difference between two types of samples, natural spoilage vs. salmonella-contaminated, the sample types were separated across the first two PCs, particularly the first PC, from 24 h which accounted for 41.9% total explained variance (TEV) and increased with the following time points to approximately 50% TEV. The time-dependent trends of the two types of samples are clearly shown in the score plots (Fig. 4b) from the second CPCA model, and it can be seen that the naturally spoiled samples appeared to have a "longer" time trajectory compared to that of salmonellae-contaminated samples; this can also be seen from the classical PCA score plots.

On examination of the loadings plots of the two CPCA models (shown in Electronic supplementary material (ESM) Figs. S-1 and S-2), it appeared that the most important features are within the first 25 min of the chromatogram, with the exception of a predominating peak at ~32 min which was more abundant in the *S. typhimurium*-contaminated samples at 72 h. Hence, we performed peak deconvolution on the first 25 min plus the region between 30–33 min of each chromatogram as described before. A typical example of the results from the peak deconvolution is given in Fig. 5. It can be seen that by superimposing the original and the reconstructed chromatograms, the most significant *m*/*z* were well fitted, and this gives a certain confidence on the quality of the deconvolved mass spectra. The mass spectra before and after the peak deconvolution of one significant peak are also given in this figure. No satisfactory match can be found using the mass spectrum before the deconvolution. By contrast, using the mass spectrum after the deconvolution, it is possible to identify this feature as 5-methylpyrimidine with a high matching factor of 814. A total number of 63 unique peaks (metabolites) were extracted by this deconvolution processing from the first 25 min of the chromatograms, and 48 of them can be identified through mass spectra matching in the NIST 2.0 library search engine with confidence (match factor and reverse match factor >750). The predominating peak at 32 min appeared to be a siloxane-type peak which is very unlikely to be of a biological origin and more likely to be some type of column bleeding material. Its occurrence in one type of sample at only one

Table 1 List of the identified significant metabolites

Identified chemical	Match scores	Reverse match scores	CAS no.	Change[a]
Phenylethyl alcohol	875	914	60-12-8	↑
1-Butanol, 3-methyl acetate	846	847	123-92-2	↑
Dimethyl disulfide	787	887	624-92-0	↑
2-Heptanone	854	862	110-43-0	↑
2,5-Dimethylpyrazine	891	899	123-32-0	↓
Methoxybenzene	919	932	696-62-8	↓
Phenol	925	936	108-95-2	↓
Dimethyltrisulfide	922	929	3658-80-8	↓
2-Octanone	815	861	111-13-7	↓
Butanoic acid, 3-methyl, 2-methylpropylester	887	906	589-59-3	↓
5-Methylpyrimidine	814	885	2036-41-1	↓
4-Methyl-2-oxovaleric acid	778	844	816-66-0	↓
2-Octenal	811	817	2548-87-0	↓
L-5-Propylthiomethylhydantoin	734	753	71100-43-1	↑
Propanoic acid, 2-methyl, 3-methylbutylester	766	855	2050-01-3	↓
Toluene	889	901	108-88-3	↓

[a] An upward arrow means an increase in *Salmonella*-contaminated relative to natural spoilage and vice versa

time point could be purely by chance. Hence, it was excluded from the statistical analysis. In addition, there were 52 missing values (i.e. a peak that could not be found a match in a certain chromatogram) in total, accounting for 1.72% of the total number of the elements in the final peak table, and the correlation coefficients of the matched peaks varied from 0.92 to 0.99, with the majority of the matches having a correlation coefficient higher than 0.96. This suggested that even the peak deconvolution procedure was conducted independently four times on different set of samples (samples from four different time points); the results are highly consistent and can be seen as an evidence showing that the peak deconvolution is robust and reproducible.

N-way ANOVA was applied on the 63 metabolites. It appeared that the majority of these VOCs (52 out of 63) show significant difference with false discovery rate [29] $Q \leq 0.05$ between different time points, and these metabolites generally increased as spoilage progresses. This matches our organoleptic observations that the odour from the spoiled pork intensifies as spoilage progresses. In addition, 19 metabolites were identified as being significant in that they showed differences ($Q \leq 0.05$) between the sample types. Among those 19 significant peaks, 16 of them can be putatively identified through mass spectra matching. These significant peaks with their chemical identifications are listed in Table 1, and the box whisker plots along with the corresponding deconvoluted mass spectra of one metabolite, 5-methylpyrimidine which is also the exemplary peak given in Fig. 5, is shown in Fig. 6. Such plots of other significant metabolites are listed in the supplementary information (ESM Figs. S-3 to S-20).

Once deconvolution was complete, classical PCA was performed on the peak areas of the 63 deconvolved peaks (without the siloxan peak at 32 min), and the scores plot is shown in Fig. 3b. It is clear that the result is almost identical to the one obtained from TICs (Fig. 3a), except that the separation on the last time point (72 h) was not as striking as before; this is likely to be because the artefact siloxane peak was excluded from the analysis. This proved that in using TICs alone, it was still possible to extract valuable information from the data set without the need to go through laborious peak deconvolution process. On the other hand, the predominating siloxane peak found at 32 min in this study also suggested that the results obtained from TICs analysis alone can be misleading because no chemical information was obtained from the TICs, and hence, the findings are prone to false positive errors. We believe that a good approach is to combine the two methodology using TICs to get an initial view of the data set and then perform targeted peak deconvolution and structure identification to confirm the findings from the analysis based on TICs. Employing such a strategy enables researchers to achieve results within a reasonable timescale whilst maximising the knowledge obtained from the data.

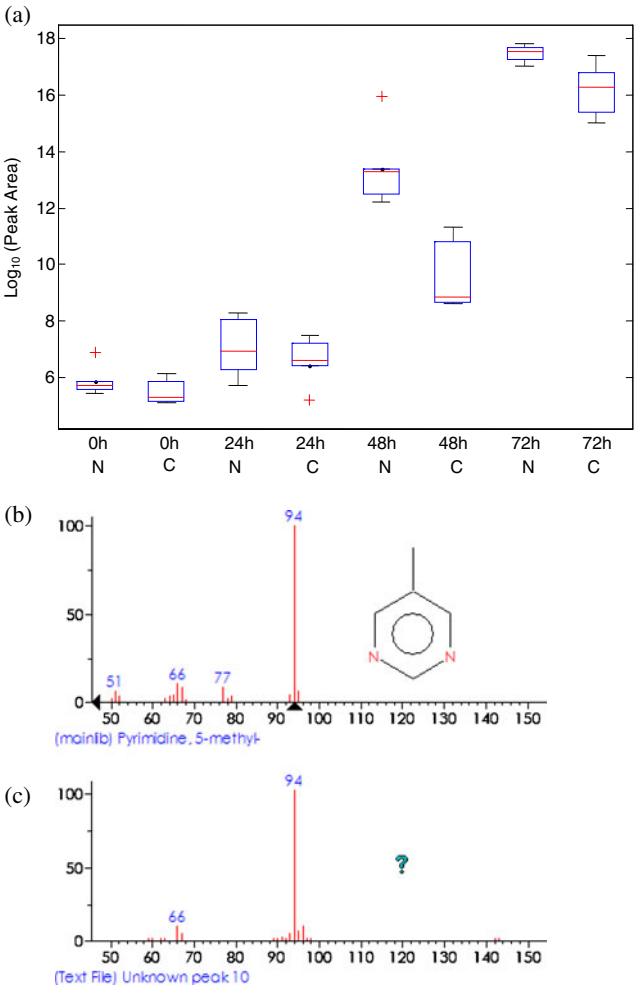


Fig. 6 **a** Box-whisker plot of 5-methylpyrimidine. **b** Standard mass spectrum. **c** Extracted mass spectrum of the target peak

From Table 1, it is clear that five VOCs are significantly increased in *S. typhimurium* relative to pork that has been allowed to spoil naturally. These are phenylethyl alcohol ($C_8H_{10}O$), 1-butanol, 3-methyl acetate ($C_7H_{14}O_2$), dimethyl disulfide ($C_2H_6S_2$), 2-heptanone ($C_7H_{14}O$) and L-5-propylthio methylhydantoin ($C_7H_{12}N_2O_2$).

2-Heptanone is naturally found in certain foodstuffs including beer, bread and various cheeses [30]. As these are all microbially produced food (albeit from fungi), it is perhaps not surprising that we identified this metabolite. Although 2-heptanone has been observed in the headspace of *Enterobacteriaceae* [31], the concentration of the metabolite was reported at a higher concentration in *Escherichia coli* compared to *S. typhimurium*; however, the growth was performed on artificial culture rather than on the surface of meat in which the metabolic source of 2-heptanone has not yet been reported.

It is known that phenylethyl alcohol is a metabolite that is part of the phenylalanine pathway (KEGG: reaction R02611), so this suggests that from the meat substrate, *Salmonella* is utilising the amino acid phenylalanine preferentially for growth compared with other organisms found on the meat surface naturally. In a similar manner, dimethyl disulfide may be present due to degradation of sulphur-containing amino acids such as methionine and cysteine, as has been reported previously for *S. typhimurium* contamination in vegetables rather than meat [32].

With respect to the other volatile species, these have not been extensively reported in the literature, and so it is difficult to be very specific. However, it is known that VOCs are used by microorganisms to give the VOC producer an advantage in terms of colonisation by suppressing the growth of other bacteria [33], so these may be defence chemicals elicited by *S. typhimurium*. It is already known that phenylethyl alcohol has antimicrobial property; hence, this metabolite could also be a part of the defence mechanism of *S. typhimurium* to inhibit the growth of other bacteria.

Conclusions

In this study, we have demonstrated the feasibility of developing a VOC metabolite profiling method based on headspace analysis using PDMS capture followed by thermal desorption into GC/MS. This analytical approach generated information-rich metabolite data that we have shown were able to discriminate successfully *S. typhimurium*-contaminated and naturally spoiled pork during spoilage at 25 °C over 72 h. This suggests that VOC analysis has the potential to become a valuable tool for noninvasive, rapid detection of pathogenic microorganisms.

A set of potentially interesting metabolites have been putatively identified through this study, and this will aid the development of specific and sensitive analytical tools for the detection of pathogens in food. Further work is needed to improve the confidence of the metabolite identification; that is to say to go from putative to definitive identification [26]. Currently, the metabolite databases containing VOCs is still at a very preliminary stage and more international effort is needed to increase the metabolite coverage in these databases to the current level offered by derivative-based GC/M metabolite databases.

In addition, we have also developed a chemometric methodology that combines fast multivariate analysis on TICs with more time-consuming peak deconvolution on selected regions of the chromatograms that contain discriminatory information. Satisfactory results were obtained rapidly from the data generated by a routinely used MS instrument employing a quadrupole mass filter. This makes it particularly suitable for low-cost, rapid pilot studies. If further investigations are desired, more labour-intensive and expensive instrumentation can be used (viz. GC-ToF-MS or GCxGC-ToF-MS) followed by more in-depth deconvolution. In this investigation, we have employed this methodology to investigate the differences in the production of VOCs in the natural and artificial spoilage of pork. We have highlighted several metabolites which are characteristic of spoilage by the pathogenic bacteria and with future work may be used as indicators for noninvasive biomarkers of pathogenic spoilage of meat.

Acknowledgement We are grateful to Professor Tim Brocklehurst for provision of *S. typhimurium* strain 4/74. RG and YX thank the EU Framework 6 programme Biotracer for funding. CLW thanks the BBSRC for financial support. RG also wishes to thank the BBSRC and EPSRC for financial support of The Manchester Centre for Integrative Systems Biology.

References

1. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Nat Biotechnol 18:1157–1161
2. Harrigan GG, Goodacre R (2003) Metabolic profiling: its role in biomarker discovery and gene function analysis. Kluwer, Boston
3. O'Hagan S, Dunn WB, Brown M, Knowles M, Knowles JD, Kell DB (2005) Anal Chem 77:290–303
4. Ellis DI, Goodacre R (2001) Trends Food Sci Tech 12:414–424
5. Ellis DI, Broadhurst D, Kell DB, Rowland JJ, Goodacre R (2002) Appl Environ Microbiol 68:2822–2828
6. Jackson TC, Acuff GR, Dickson JS (1997) In: Doyle MP, Beuchat LR, Montville TJ (eds) Food microbiology: fundamentals and frontiers. ASM, Washington
7. Jay JM (1996) Modern food microbiology. Chapman & Hall, London
8. Stanbridge LH, Davies AR (1998) In: Davies A, Board R (eds) The microbiology of meat and poultry. Blackie Academic & Professional, London
9. Oliveiraa CJB, Carvalhoc LFOS, Fernandesd SA, Tavechiod AT, Domingues FJ Jr (2005) Int J Food Microbiol 2:267–271
10. Mitra S (2003) Sample preparation technique in analytical chemistry. Wiley, New Jersey
11. Riazanskaia S, Blackburn G, Harker M, Taylor D, Thomas CLP (2008) Analyst 133:1020–1027
12. Eilers PHC (2004) Anal Chem 76:404–411
13. vest Nielsen NP, Carstensen JM, Smedegarrad J (1998) J Chromatogr A 805:17–35
14. Skov T, van den Berg F, Tomasi G, Bro R (2006) J Chemometr 20:484–497
15. Wold S (1987) Chemom Intell Lab Syst 2:37–52
16. Brereton RG (2003) Data analysis for the laboratory and chemical plant. Wiley, Chichester
17. Westerhuis JA, Kourti T, MacGregor JF (1998) J Chemometr 12:301–321
18. Smilde AK, Westerhuis JA, de Jong S (2003) J Chemometr 17:323–337
19. Biais B, Allwood JW, Deborde C, Xu Y, Maucourt M, Beauvoit B, Dunn WB, Jacob D, Goodacre R, Rolin D, Moing A (2009) Anal Chem 81:2884–2894
20. Cuesta Sanchez F, Toft J, van den Bogaert B, Massart DL (1996) Anal Chem 68:79–85

21. Durbin J, Watson GS (1950) Biometrika 37:409–428
22. Durbin J, Watson GS (1951) Biometrika 38:159–179
23. Tauler R (1995) Chemom Intell Lab Syst 30:133–146
24. Jonsson P, Johansson AI, Gullberg J, Trygg J, Jiye A, Grung B, Marklund S, Sjöström M, Antti H, Moritz T (2005) Anal Chem 77:5635–5642
25. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis RM, Stephens GM, Goodacre R (2008) Anal Chem 80:2939–2948
26. Sumner LW, Amberg A, Barrett D, Beger R, Beale MH, Daykin C, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, Hardy N, Higashi R, Kopka J, Lindon JC, Lane AN, Marriott P, Nicholls AW, Reily MD, Viant M (2007) Metabolomics 3:211–221
27. vest Nielsen NP, Smedsgaard J, Frisvad JC (1999) Anal Chem 71:727–735
28. Dixon SJ, Xu Y, Brereton RG, Soini HA, Novotny MV, Oberzaucher E, Grammer K, Penn DJ (2007) Chemom Intell Lab Syst 87:161–172
29. Benjamini Y, Hochberg Y (1995) J Roy Statist Soc Ser B 57:289–300
30. Maeda T, Kim JH, Ubukata Y, Morita N (2009) Eur Food Res Tech 228:1438–2377
31. Siripatrawan U (2008) Sensor Actuator B Chem 133:414–419
32. Siripatrawan U, Harte BR (2007) Anal Chim Acta 581:63–70
33. Kai M, Haustein M, Molina F, Petri A, Scholz B, Piechulla B (2009) Appl Microbiol Biotechnol 81:1001–1012

Chapter 5

Metabolic profiling of meat: assessment of pork hygiene and contamination with *Salmonella typhimurium*

Yun Xu, **William Cheung**, Catherine L. Winder, Warwick B. Dunn and Royston Goodacre.

Author contribution to the work presented within the thesis

William Cheung: Experimental design, logistics and implementation of the spoilage study; processing of all samples for metabolite profiling (extraction and purification).

Dr Catherine Winder: Help and advice on microbiology.

Dr Warwick Dunn: Derivitisation and analysis of samples using GC/TOF-MS.

Dr Yun Xu: Experimental design, growing of *S. typhimurium*, sample collections, data analysis using PARAFAC and all subsequent univariate analysis for the identification of potential discriminate metabolites.

Metabolic profiling of meat: assessment of pork hygiene and contamination with *Salmonella typhimurium*†

Yun Xu,[a] William Cheung,[a] Catherine L. Winder,[a] Warwick B. Dunn[ab] and Royston Goodacre*[ab]

Received 10th June 2010, Accepted 2nd November 2010
DOI: 10.1039/c0an00394h

Spoilage in meat is the result of the action of microorganisms and results in changes of meat and microbial metabolism. This process may include pathogenic food poisoning bacteria such as *Salmonella typhimurium*, and it is important that these are differentiated from the natural spoilage process caused by non-pathogenic microorganisms. In this study we investigated the application of metabolic profiling using gas chromatography-mass spectrometry, to assess the microbial contamination of pork. Metabolite profiles were generated from microorganisms, originating from the natural spoilage process and from the artificial contamination with *S. typhimurium*. In an initial experiment, we investigated changes in the metabolic profiles over a 72 hour time course at 25 °C and established time points indicative of the spoilage process. A further experiment was performed to provide in-depth analysis of the metabolites characteristic of contamination by *S. typhimurium*. We applied a three-way PARAllel FACtor analysis 2 (PARAFAC2) multivariate algorithm to model the metabolic profiles. In addition, two univariate statistical tests, two-sample Wilcoxon signed rank test and Friedman test, were employed to identify metabolites which showed significant difference between natural spoiled and *S. typhimurium* contaminated samples. Consistent results from the two independent experiments were obtained showing the discrimination of the metabolic profiles of the natural spoiled pork chops and those contaminated with *S. typhimurium*. The analysis identified 17 metabolites of significant interest (including various types of amino acid and fatty acid) in the discrimination of pork contaminated with the pathogenic microorganism.

Introduction

Food spoilage is the result of the activities of microorganisms on the food matrix resulting in the decomposition of carbohydrates and proteins,[1,2] and the synthesis and subsequent release of malodorous substances and metabolites from the breakdown of meat products or microorganisms. The quality of food is usually defined with respect to organoleptic changes that make the meat unacceptable to the consumer and these include the development of off-odours, colour changes or slime formation;[3] this deterioration eventually makes the food unsuitable to eat and this occurs at around 10^7 colony forming units (cfu) per square centimetre.[4] It is particularly hazardous when the food is contaminated by pathogenic microorganisms, for example *Salmonella* bacteria, as these can attack the stomach and intestine and in severe cases result in blood poisoning. Various tools have been developed to monitor the hygiene status of foods, from general flavours and aroma analysis to some very specific tool including PCR assays and immunosensors among others for detecting specific hazardous pathogen(s); for a review see Ellis and Goodacre.[2] In recent years, the approach of spectral "fingerprinting" of the metabolic profile of a biological system has attracted significant attention, especially in the field of metabolomics. This is because of the rapidity of the sampling which leads to immediate action being taken if a contamination is detected. In addition, methods that measure the phenotype of the cell are advantageous as this shows how the organism is responding to an environment rather than a somewhat static readout such as PCR. The metabolic profile is a 'holistic' method to determine metabolites from a diverse range of metabolic pathways, so to define system-wide differences in metabolism. By utilizing powerful analytical tools such as gas chromatography-mass spectrometry, a wide range of metabolites can be detected and identified and, at least, semi-quantified. This provides abundant information to further the understanding of the biological process under investigation and paves the way for developing more specific and sensitive tools for pathogen detection.[5–8]

Given the advantages of metabolic profiling approaches, we believe that such approach can be adopted as a valuable tool to improve the traceability of the contaminations of microorganism. There are a few reports in the literatures investigating rapid natural meat spoilage detection by proton transfer reaction mass spectrometry[9] and FT-IR spectroscopy[10,11] yet these studies were based on a single isolated experiment on naturally spoiled meat. In this study we explore the use of a metabolic profiling approach to characterise the microbial contamination of pork and, in addition, discriminate samples colonised with natural spoilage microorganisms from those contaminated

[a]*School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK. E-mail: roy.goodacre@manchester.ac.uk; Fax: +44 (0)161 3064519; Tel: +44 (0)161 3064480*
[b]*Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK*

† Electronic supplementary information (ESI) available: The box-whisker plots of the identified significant metabolites. See DOI: 10.1039/c0an00394h

by *Salmonella typhimurium*. Two objectives were investigated: (1) the temporal progression of the spoilage process and (2) to determine metabolic differences associated with contamination by *Salmonella* or natural spoilage microorganisms. Two independent experiments were conducted; the first experiment monitored the progression of the spoilage process during the incubation period following contamination by *S. typhimurium*. Metabolic profiles were generated at multiple time points ($n = 11$) across a 72 hour period for each experimental condition (natural spoiled and *Salmonella* contaminated). The second experiment provided robust statistical testing and identified specific metabolites to differentiate between the experimental conditions. It is worth noting that even *S. typhimurium* is not a commonly observed microorganism found in food 'spoilage'; however, when this food 'poisoning' organism is found, it usually leads to very serious consequences.[12,13]

Experimental

Culture and chemicals

S. typhimurium strain 4/74 was kindly provided by Professor Tim Brocklehurst (The Institute of Food Research, Norwich, UK). The strain was sub-cultured on Lab M LAB028 blood agar plates (Lab M Ltd., Lancashire, UK). A single colony was inoculated into nutrient broth (250 mL) and incubated at 37 °C for 16 h, which resulted in a culture of $\sim 5 \times 10^7$ cfu mL^{-1}. An aliquot (5 mL) of the culture was harvested by centrifugation at 4810*g* for 10 min. The supernatant was removed and the pellet was re-suspended in 50 mL of sterile saline solution and centrifuged at 4810*g* for 10 min, the supernatant was removed and this process was repeated two further times. The pellet was re-suspended in 50 mL of sterile saline solution and used for the artificial contamination of the pork.

Sample collection

A total of 33 boneless pork chops (weight 200–300 g) were used in the initial experiment. They were purchased from a local supermarket. Each pork chop was sampled to provide a unified surface area of 12–14 cm^2 followed by a process of butterflying into two pieces with equal surface area. In butterflying the meat is cut laterally and as spoilage organisms are not intramuscular this provides a virtually sterilized surface for the study. For each piece of pork, a digital photo was taken alongside with a ruler. The number of pixels on the pork surface was counted and divided by the number of pixels within one cm^2 area (estimated from the ruler in the same photo), thus providing an estimation of the surface area of the pork. Each piece of pork was placed in a large glass Petri dish lined with sterilized filter paper (Whatman grade 40 cat. no. 90-7501-06) to which 2 mL of sterilized saline solution (0.9% NaCl, w/v) was added to act as a moisture source and prevent the surface of the meat from drying out. The matched pieces of meat from the same pork chop were used as either a control or an artificial contamination surface. To the control sample, an aliquot (1 mL) of sterilized saline solution was added and the sample was incubated to allow the natural spoilage to progress. To the other piece, 1 mL of saline suspension of *S. typhimurium* (*vide supra*) was added and spread by using a sterile plastic loop. The Petri dishes were covered and incubated at 25 °C for various time points. A total of 11 time points were monitored: 0, 12, 24, 28, 32, 36, 40, 44, 48, 60 and 72 h during the incubation. At each time point, the biomass from the surface of 3 pork chops, *i.e.* 6 pieces were harvested using sterile swabs. Two swabs were used for each piece of pork and both were transferred directly into 1 mL of ice-cold methanol stored on dry ice (−48 °C). The suspension was extracted with three freeze–thaw cycles (frozen in liquid nitrogen and allowed to thaw on dry ice). The suspensions were then centrifuged at 16 060*g*, at −9 °C for 5 min. The supernatants were immediately lyophilised and subjected to GC-MS analysis (see below). A second experiment was performed approximately 6 months after the first experiment, the sample collection procedures were the same with the exception that only 4 time points were monitored: 0, 24, 48 and 72 h after the contamination. Samples from 6 pork chops were collected for each time point with 50% of the samples being applied as the controls and the other 50% applied as contaminated with *Salmonella*. Total viable counts (TVCs) were also performed in this experiment. This was performed by harvesting the microbial load from a 1 cm^2 area (randomly selected) of each pork piece using a sterile swab and transferred it into 1 mL of sterilized saline solution. The TVC data were hereby obtained using the plate count method.[3]

GC/MS analysis

An aliquot of 1000 μL of each metabolite extract was spiked with 100 μL of internal standard solution (0.19 mg mL^{-1} succinic d_4 acid, 0.27 mg mL^{-1} malonic d_2 acid and 0.22 mg mL^{-1} glycine d_5 in HPLC grade water) and then lyophilised in a HETO VR MAXI vacuum centrifuge attached to a HETO CT/DW cooling trap (Thermo Life Sciences, Basingstoke, UK). Samples were subsequently derivatized in two stages. An aliquot (50 μL) of 20 mg mL^{-1} *O*-methylhydroxylamine solution in pyridine was added and heated at 60 °C for 45 min followed by adding an aliquot (50 μL) of MSTFA (*N*-acetyl-*N*-(trimethylsilyl)-trifluoroacetamide) and then heating at 60 °C for 45 min. A retention index solution was added for chromatographic alignment (20 μL, 0.6 mg mL^{-1} $C_{10}/C_{12}/C_{15}/C_{19}/C_{22}$ *n*-alkanes).

The samples were analyzed in a random order by employing a GC/TOF-MS (Agilent 6890 GC coupled to a LECO Pegasus III TOF mass spectrometer) using a previously described method for yeast footprint samples.[14] Raw data were processed using LECO ChromaTof v2.12 and its associated chromatographic deconvolution algorithm, with the baseline set at 1.0, data point averaging of 3 and average peak width of 2.5. A reference database was prepared, incorporating the mass spectrum and retention index of all metabolite peaks detected in a random selection of samples (one sample per contamination or time-point class) so to allow detection of all metabolites present, whether or not expected from the study of bibliographic data. Each metabolite peak in the reference database was searched for in each sample and if matched (retention index deviation < ±10; mass spectral match > 750) the peak area was reported and the response ratio relative to the internal standard (peak area-metabolite/peak area-succinic d_4 acid internal standard) calculated. These data (matrix of N samples × P metabolite peaks) representing normalised peak lists were exported in ASCII format for further analysis. Metabolites were definitively

identified[15] by matching the mass spectrum and retention index of detected peaks to those present in the mass spectral library constructed at the University of Manchester.[16] A match is defined as a match factor greater than 750 and a retention index ±10.

Data analysis

We employed a three-way model named PARAllel FACtor analysis 2 (PARAFAC2)[17,18] to model the data generated by the two experiments. PARAFAC2 is an extension of the well known PARAFAC model,[19–21] a generalization of principal component analysis[22] to the situation where a series of matrices to be analysed (*e.g.* a three-way matrix). In PARAFAC, it models a 3 dimensional matrix $\underline{\boldsymbol{X}}$ with a dimension of ($I \times J \times K$) as a summation over R outer products of triads of vectors where R is the number of PARAFAC components. For the k^{th} ($k = 1, 2,\ldots, K$) "slab" (*e.g.*, all the sample at one certain time point) $\boldsymbol{X}_k$ in $\underline{\boldsymbol{X}}$, PARAFAC decompose it into a product of two loadings matrices as shown in eqn (1):

$$\boldsymbol{X}_k = \boldsymbol{F}\boldsymbol{D}_k\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{\varepsilon}_k \quad (1)$$

where $\boldsymbol{F}$ is the loadings matrix of the row units (*e.g.*, samples), $\boldsymbol{A}$ is the loadings matrix of the column units (*e.g.*, the peak areas of the metabolites), D_k is the weight scalar for the k^{th} slab and $\boldsymbol{\varepsilon}_k$ is the residue matrix of the k^{th} slab. In PARAFAC related literatures, row units is normally named mode 1, column units is named mode 2 and a diagonal matrix $\boldsymbol{D}$, which contains the weights of all the slabs on its diagonal, is named mode 3, while $\boldsymbol{F}$, $\boldsymbol{D}$ and $\boldsymbol{A}$ are called the loadings of mode 1, 3 and 2 respectively. PARAFAC fits the three-way matrix X by minimizing each of the ε_k and this is typically achieved by using an alternative least square optimization.

However, PARAFAC requires that the size of each slab has to be exactly the same which sometimes cannot be met, even with a balanced experiment design (*e.g.*, a few samples may have failed to be derivatized and cannot generate good GC/MS data). PARAFAC2 was developed to cope with such problem and allows one dimension (normally rows, containing the samples) to be unequal between different slabs. The PARAFAC2 model is given by

$$\boldsymbol{X}_k = \boldsymbol{F}_k\boldsymbol{D}_k\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{\varepsilon}_k \quad (2)$$

Instead of giving a global loadings matrix for mode 1, PARAFAC2 gives K loadings matrices, one for each slab, and each matrix may have different rows which matches the original size of $\underline{\boldsymbol{X}}$. By using PARAFAC2, a minor sample lose can be afforded without the need to "trim" the whole 3-D data matrix to make each slab equal.

In this study, two samples were lost during the first experiment and therefore a flexible PARAFAC2 was needed. We re-arranged the data matrix into a three-way matrix as illustrated in Fig. 1, then employed a PARAFAC2 model to fit the data. Using this model, the relative distribution of the samples at each time point can be revealed at loadings plots of the mode 1, the loadings of mode 2 can help identifying potential interesting variables (*i.e.*, metabolites), while the loadings of mode 3 (*i.e.*, the weight of each slab) reveal the global change of metabolic profiles over time. The number of PARAFAC2 components was 2 as once 3 or more components were used, the results appeared to be very unstable (*i.e.*, different runs generated different results), indicating too many components were introduced causing distortions. Once the model was built, the separation between two types of samples, *viz. S. typhimurium* contaminated and natural spoiled samples, at each time point was measured using a Hotelling T^2 statistic[23] using the model 1 loadings. Two PARAFAC2 models were built, one for each experiment, with the same parameter setting and the results were compared.

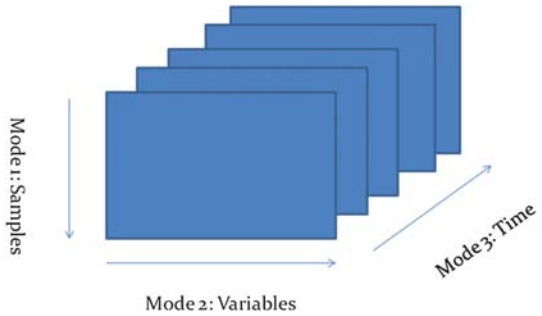


Fig. 1 Illustration of PARAFAC2 model. Mode 1 details information about the samples that have been analysed (*e.g.* naturally spoiled and contaminated with a food poisoning pathogen), Mode 2 details the variables (in this case identified metabolites), and Mode 3 the time course of the experiment (*e.g.* 11 time points from 0–72 h in the first experiment).

Although mode 2 of the PARAFAC2 model can help identifying potentially interesting metabolites as stated above, it provides no statistical assessment of the metabolites. Hence we also employed univariate statistics test to identify statistically significant metabolites from the two experiments. Since the two experiments have different sample compositions, different statistical tests have to be used. For the first experiment, we monitored a fairly large number of time points (11 in total) while the number of samples at each time point is rather limited (6 in total, 3 for each type). In this dataset, the peak areas of each metabolite of the same type of samples at the same time point were averaged and resulted in 11 pairs of averaged peak areas indicating the difference of the abundance of a certain metabolite at each time point. A paired two-sample Wilcoxon signed rank test[24] was then employed to identify the metabolites which showed significant difference between two types of samples. Since the number of levels (time points) is larger than the number of samples at each level in this experiment, it is difficult to identify the metabolites which showed significant difference between different levels based on the data obtained from this experiment. For the second experiment, a reduced number of time points (4 in total) was monitored while the number of samples at each time point was doubled which makes it more suitable for statistics test. We employed a non-parametric two-way ANOVA test (Friedman's test)[25] to identify metabolites which showed significant differences between different types of samples or different time points or both. Since the statistics tests were used for multiple hypothesis testing (*i.e.* simultaneously testing a family of variables/metabolites), the significant threshold of p-values was adjusted by using false discovery rate (FDR)[26] to control the increased chance of type I error (false positive).

Results and discussion

In both experiments, a distinct odour can be smelt after 24 h of incubation and such odour intensified rapidly afterwards. After 48 h of incubation the odour developed into a very unpleasant putrid smell, indicating the onset of proteolysis. Transparent to light green slimes could be seen on the surface of the pork from 36–48 h onwards. Few visual differences were observed between naturally spoiled pork and those contaminated by *Salmonella* after 48 hours. The total viable counts at 0, 24, 48 and 72 h in the second experiment are provided in Table 1. The TVCs increase as spoilage progresses to a level of approximately 10^8 to 10^9 cm^{-2} after 72 h of incubation. The increase in TVC was different for the two types of spoilage with the naturally spoiled meat showing higher TVCs at each time point.

A total number of 126 peaks were detected from the first experiment while 116 peaks were detected from the second experiment from the GC-MS analysis. In multivariate analysis, principal component analysis (PCA) is the most commonly used tool to gain an intuitive view of the multivariate data. However, when there are 2 or more underlying influential factors PCA is not always the best method to reveal the influence of these factors. In this study there are two factors of interest, these are the progression of spoilage and the type of samples (natural *vs.* *Salmonella* spoilage). As suggested by Cattell,[27] when classical factor analysis models like PCA are applied to such dataset, the latent factors obtained will be "neither clear species differentiator nor optimal individual differentiators". A three-way model is most suited to the data generated in this study such that the influential parameters *i.e.*, time, experimental conditions (natural spoilage and *Salmonella* contamination) and progression of spoilage can be investigated individually. The influence of time, *i.e.* the progression of spoilage, can be explicitly modelled by adding a third dimension into the model (mode 3) whilst separating it from the influence of the experimental condition *i.e.* natural spoilage and *salmonella* contamination which are modelled in mode 1 (Fig. 1). The plots of the loadings of mode 1 which models the relative position of the samples at each time point from the PARAFAC2 model based on the data from the first experiment are shown in Fig. 2. For brevity, only 4 time points: 0, 28, 32 and 72 h of the mode 1 loadings were shown in Fig. 2(a) and this illustrates that the separation of the sample types (*S. typhimurium* contaminated and natural spoiled) becomes greater as the spoilage progresses. This trend is easier to see by plotting the Hotelling's T^2 statistics which measures the separation between two types of samples against time as shown in Fig. 2(b). The Hotelling's T^2 statistics allows a *p*-value to be calculated. If one considers $p < 0.05$ as "significant", it appears that in this experiment between 32 and 36 h (and later time points) after the contamination, the sample types (contaminated *vs.* naturally spoiled) can be "significantly" separated. Note that at each time point the TVCs (Table 1) are equivalent so this is not

Table 1 Average total viable counts per cm^2 from the 4 time points monitored[a]

	0 h	24 h	48 h	72 h
N[b]	0–500	1.9×10^7 (1.8×10^6)	3.9×10^8 (5.1×10^7)	1.5×10^9 (9.5×10^7)
C[c]	4.8×10^5 (5.7×10^4)	1.4×10^7 (5.1×10^6)	2.3×10^8 (9.5×10^6)	5.3×10^8 (1.3×10^7)

[a] The standard deviations of the TVCs are given in the parentheses. [b] N: naturally spoiled samples. [c] C: *S. typhimurium* contaminated samples.

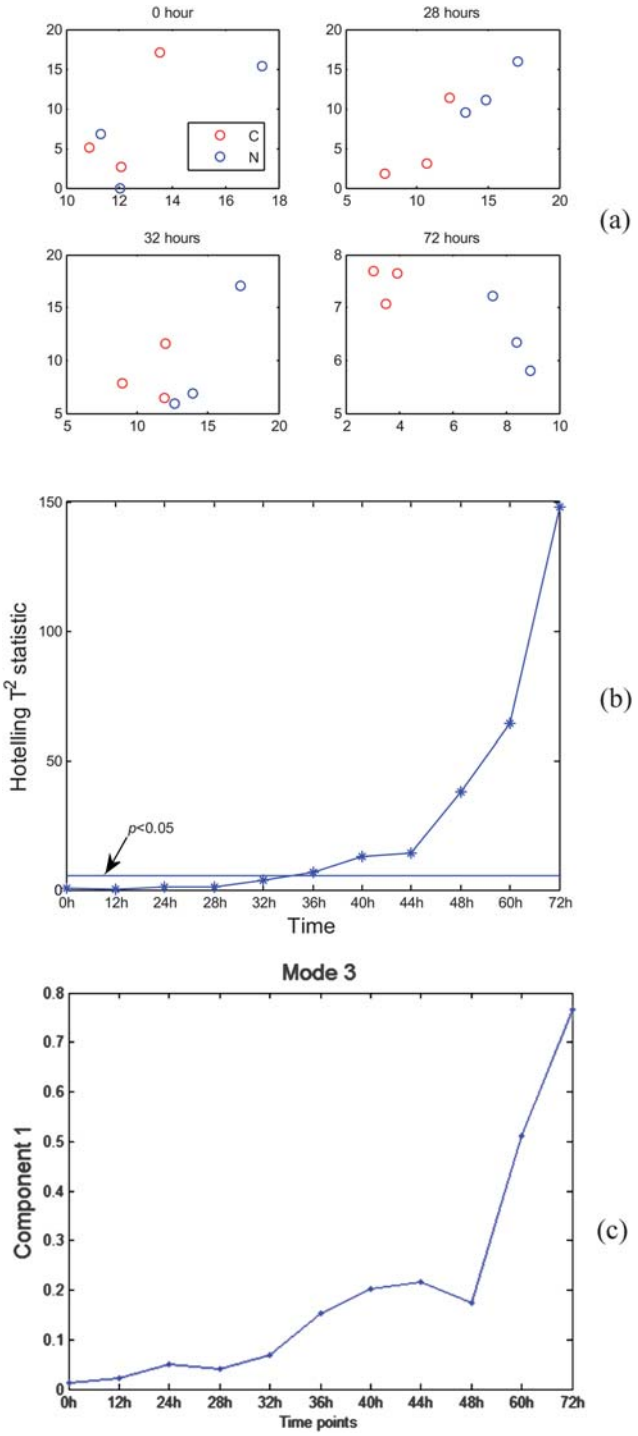


Fig. 2 (a) PARAFAC2 mode 1 loadings plot from the first experiment: C = *Salmonella* contaminated samples; N = natural spoiled samples; (b) Hotelling's T^2 statistics *versus* time, the value corresponds to $p = 0.05$ was marked by a horizontal line and (c) the first component in the PARAFAC2 mode 3 loadings plotted against time.

a bacterial load influence but a metabolic difference, *i.e.* different types of samples appeared to have different metabolic profiles. Due to the restricted sample size, the statistical implication of the *p*-value derived from the Hotelling's T^2 statistics is limited and can only be used for reference purpose, therefore in the second experiment we increased the number of samples at each sampling point. The mode 3 loadings plot (Fig. 2(c)) showed an approximately monotonically increasing trend which suggests that the abundance of the majority of the metabolites was increasing as the spoilage progressed, and thus can be used as indicator for the progression of the spoilage. The increase in the general abundance of metabolites is most likely caused by the increase in the abundance of the microorganisms on the surface of the meat as shown in the TVC data (Table 1).

The mode 1 loadings plot from the PARAFAC2 model based on the second experiment is shown in Fig. 3(a) which showed a similar trend as observed in the first experiment (Fig. 2). The inclusion of multiple replicates increases the statistical power of the Hotelling's T^2 statistics, and allows a clear discrimination of the early sample points (24 h) after the contamination ($p < 0.05$) as shown in Fig. 3(b).

By applying a paired *t*-test on the data from the first experiment, 16 peaks were identified as significant when the false discovery rate (FDR) was set to 0.05; 12 of these peaks were chemically identified by the matching of mass spectra to those in metabolomics-specific mass spectral libraries. The two-way ANOVA test with the FDR set to 0.05 on the data from the second experiment suggested that nearly half of the metabolite features (55 out of 116 detected peaks) showed significant differences between different time points. The majority of peaks showed a monotonically increasing trend which matched the trend shown in the mode 3 loadings plot from the PARAFAC2 model based on the data from the first experiment (Fig. 2 (c)). In addition, 16 peaks were identified as significant and 11 of them were definitively identified through mass spectra matching.[15,16] Among these 11 metabolites with chemical identifications, 6 of them were also detected and identified as significant in the first experiment. In addition, there is one compound (inosine) which was identified as significant in the first experiment but not significant in the second experiment. It seems that this compound was not detected in 33 of 48 samples analysed in the second experiment. This is most likely a result of the metabolite concentration being lower than the analytical limit of detection. The 17 significant (identified as significant in either of the 2 experiments) and identifiable metabolites are detailed in Table 2 with an indication as to if they are increased in natural spoilage relative to *S. typhimurium* contaminated meat, or *vice versa*; the time trajectories for these metabolites are shown in the ESI†. In addition to these a further 9 metabolite features (peaks) were significant but remain unidentified. Ignoring the differences between natural *vs.* contaminated meat, many of the metabolites identified are due to proteolysis occurring after 24 h. This is associated with the post-glucose utilization of amino acids by pseudomonads and is coincidental with organoleptic changes indicating the production of malodorous protein breakdown products.[2] Several amino acids (valine, tryptophan, aspartic acid, lysine and glycine), amino acid derivatives (hydrocinnamic acid from microbial action in the phenylalanine pathway) and creatinine (a breakdown product from general enzymatic proteolysis) all increase (see ESI†) after 24 h. Methylmalonic acid is a product from the metabolism of fat and protein and also increases after 24 h. In addition, several fatty acids increased during the spoilage process and these included dodecanoic acid, tetradecanoic acid, hexadecanoic acid, octadecenoic acid and ethanolamine (which produces an ammoniacal odour). The production of fatty acids is likely to result from the breakdown of meat and incorporation into the lipid membrane of the bacteria that are growing on the meat surface.

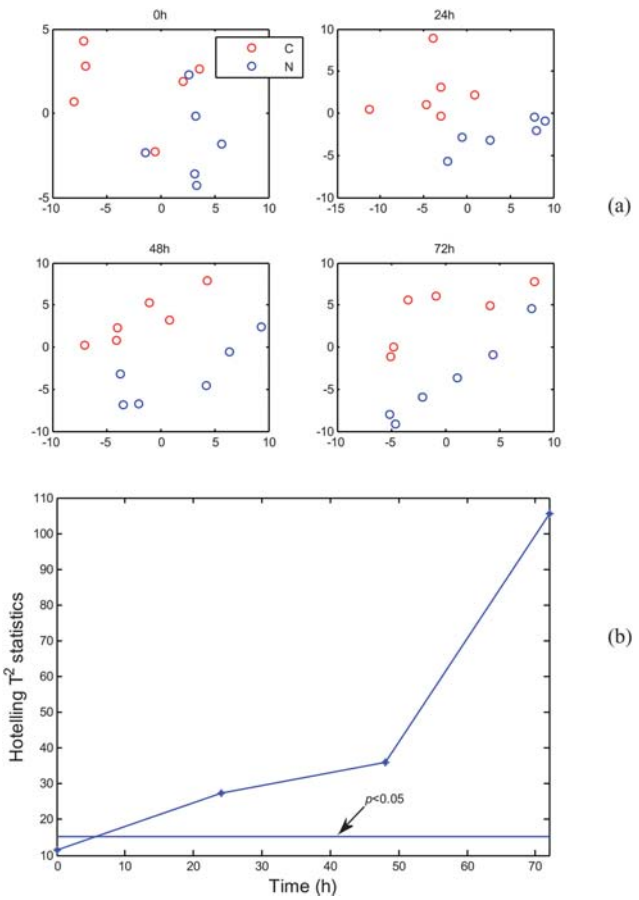


Fig. 3 (a) PARAFAC2 mode 1 loadings plot from the second experiment: C = *Salmonella* contaminated samples; N = natural spoiled samples; (b) Hotelling's T^2 statistics *versus* time.

With respect to differences between *S. typhimurium* contamination and natural spoilage it is clear that several amino acids are increased in *Salmonella* contamination relative to natural spoilage and these include aspartic acid, lysine and glycine, whilst lysine, valine and tryptophan are increased during natural spoilage. It is known that bacteria utilise amino acids preferentially and in a species specific manner[28] and this information could be used to differentiate this food poisoning bacterium from background harmless microflora. In addition, the same is also the case for fatty acids content in bacteria and hexadecanoic acid could be a marker for *S. typhimurium* contamination. However, from this study there is no direct evidence of this and future studies will be targeted at the amino acid utilisation pattern from *S. typhimurium* strain 4/74, and also undertake a fatty acid profile.

Table 2 The significant metabolites identified in the discrimination of the natural spoiled samples and those contaminated with *S. typhimurium*

Name	*p*-value in exp. 1	*p*-value in exp. 2	More abundant in	HMDB i.d.
Valine	0.0119	0.0219	N[c]	HMDB00883
Methylmalonic acid	0.0163	N/A[b]	N	HMDB00202
Phosphate	0.0204	0.0091	N	—
Ethyl pipecolinate	0.0165	N/A	N	—
Hydrocinnamic acid	0.0153	N/A	C[d]	HMDB00764
Creatinine	0.0073	0.00035	C	HMDB00562
Dodecanoic acid[a]	0.0086	N/A	N	HMDB00638
Tetradecanoic acid[a]	0.0197	0.0034	N	HMDB00806
Hexadecanoic acid[a]	0.0169	0.0099	C	HMDB00220
Octadecenoic acid[a]	0.0199	0.012	N	HMDB00207
Tryptophan	0.0239	N/A	N	HMDB00929
Inosine	0.0188	0.162	N	HMDB00195
Ethanolamine	N/A	0.0209	C	HMDB00149
Aspartic acid	N/A	0.0007	C	HMDB00191
Lysine	N/A	0.0077	N	HMDB00182
Myo-inositol	N/A	0.0199	N	HMDB00211
Glycine	N/A	0.0226	C	HMDB00123

[a] Multiple possible isomers exist, only one typical HMDB i.d. was given. [b] N/A: not detected. [c] N: naturally spoiled samples. [d] C: *S. typhimurium* contaminated samples.

Conclusion

In this study we have conducted two independent experiments which were separated by a period of 6 months to explore the possibility of applying metabolic profiling to characterise the hygiene status of pork chops which undergo a spoilage process. The results from the PARAFAC2 analysis on the two experiments appeared to be highly consistent. The mode 3 loadings from the PARAFAC2 analysis as well as the non-parametric two-way ANOVA test (Friedman test) suggested that the intensities of the majorities of the metabolites generally increased over time and reflects the fact that as the spoilage progresses, the microbial load increases and hence the metabolites from the cells become more abundant in the extraction solvent. More interestingly, the mode 1 loadings from the PARAFAC2 suggested that the *salmonella* contaminated samples have metabolic profiles which are *different* from those acquired from naturally spoiled meat. Such differences can be observed as early as 24 h after contamination at room temperature and the findings were consistently observed in the two independent experiments. By employing statistical tests, 17 metabolites were discovered which showed significant differences between two types of samples and 6 of these were identified as significant in both experiments. These included valine, creatinine, tetradecanoic acid, hexadecanoic acid, and octadecenoic acid. In further studies we shall employ a targeted approach to quantify the changes of the 6 metabolites highlighted in this investigation and also test their validity as biomarkers against the contamination of other closely related strains (*e.g. Escherichia coli*). Our investigation demonstrates the use of metabolic profiling as a tool to discriminate between natural spoilage and pathogenic microorganisms without the need for time-consuming and laborious methods traditionally employed in microbiology to detect pathogenic microorganisms. The outlined method provides valuable information for tracing the source of accidental or intentional contamination of food products by pathogenic and natural spoilage microorganisms. Although this study focused on *S. typhimurium* detection, our metabolic profiling technique could readily be applied to detect other commonly observed foodborne pathogenic bacteria that are involved in food poisoning such as *Campylobacter jejuni*, *E. coli*, *Bacillus cereus*, *Staphylococcus aureus*, and *Listeria monocytogenes*.

Acknowledgements

We are grateful to Professor Tim Brocklehurst for provision of *S. typhimurium* strain 4/74. RG and YX thanks EU Framework 6 programme Biotracer for funding (FP6-2006-FOOD-036272; http://www.biotracer.org/). CLW thanks the BBSRC for financial support and WBD and RG also wish to thank the BBSRC and EPSRC for financial support of The Manchester Centre for Integrative Systems Biology.

References

1 J. M. Jay, *Modern Food Microbiology*, Chapman & Hall, London, 5th edn, 1996.
2 D. I. Ellis and R. Goodacre, *Trends Food Sci. Technol.*, 2001, **12**, 414–424.
3 D. I. Ellis, D. Broadhurst, D. B. Kell, J. J. Rowland and R. Goodacre, *Appl. Environ. Microbiol.*, 2002, **68**, 2822–2828.
4 T. C. Jackson, G. R. Acuff and J. S. Dickson, in *Food Microbiology: Fundamentals and Frontiers*, ed. M. P. Doyle, L. R. Beuchat and T. J. Montville, ASM Press, Washington, DC, 1997.
5 R. Goodacre, D. Broadhurst, A. Smilde, B. S. Kristal, J. D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D. B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg and F. Wulfert, *Metabolomics*, 2007, **3**, 231–241.
6 K. Hollywood, D. R. Brison and R. Goodacre, *Proteomics*, 2006, **6**, 4716–4723.
7 D. I. Ellis, W. B. Dunn, J. L. Griffin, J. W. Allwood and R. Goodacre, *Pharmacogenomics*, 2007, **8**, 1243–1266.
8 W. B. Dunn, *Phys. Biol.*, 2008, **5**, 11001.
9 D. Mayr, R. Margesin, E. Klingsbichel, E. Hartungen, D. Jenewein, F. Schinner and T. D. Märk, *Appl. Environ. Microbiol.*, 2003, **69**, 4697–4705.
10 E. Z. Panagou, F. R. Mohareb, A. A. Argyri, C. M. Bessant and G.-J. E. Nychas, *Food Microbiol.*, DOI: 10.1016/j.fm.2010.05.014.
11 A. A. Argyri, E. Z. Panagou, P. A. Tarantilis, M. Polysiou and G.-J. E. Nychas, *Sens. Actuators, B*, 2010, **145**, 146–154.

12 V. Delpech, J. McAnulty and K. Morgan, *Aust. N. Z. J. Public Health*, 1998, **22**, 243–246.
13 K. Mølbak, D. L. Baggesen, F. M. Aarestrup, J. M. Ebbesen, J. Engberg, K. Frydendahl, P. Gerner-Smidt, A. M. Petersen and H. C. Wegener, *N. Engl. J. Med.*, 1999, **341**, 1420–1425.
14 S. O'Hagan, W. B. Dunn, M. Brown, J. D. Knowles and D. B. Kell, *Anal. Chem.*, 2005, **77**, 290–303.
15 L. W. Sumner, A. Amberg, D. Barrett, R. Beger, M. H. Beale, C. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, N. Hardy, R. Higashi, J. Kopka, J. C. Lindon, A. N. Lane, P. Marriott, A. W. Nicholls, M. D. Reily and M. Viant, *Metabolomics*, 2007, **3**, 211–221.
16 M. Brown, W. B. Dunn, P. Dobson, Y. Patel, C. L. Winder, S. Francis-McIntyre, P. Begley, K. Carroll, B. Broadhurst, A. Tseng, N. Swainston, I. Spasic, R. Goodacre and D. B. Kell, *Analyst*, 2009, **134**, 1322–1332.
17 H. A. L. Kiers, J. M. F. ten Berge and R. Bro, *J. Chemom.*, 1999, **13**, 275–294.
18 R. Bro, C. A. Andersson and H. A. L. Kiers, *J. Chemom.*, 1999, **13**, 295–309.
19 R. B. Cattell, *Psychometrika*, 1944, **9**, 267–283.
20 J. D. Carroll and J. J. Chang, *Psychometrika*, 1970, **35**, 283–319.
21 R. A. Harshman, *UCLA Working Papers in Phonetics*, 1970, vol. 16, p. 84.
22 K. Pearson, *Philos. Mag.*, 1901, **2**, 559–572.
23 H. Hotelling, *Ann. Math. Stat.*, 1931, **2**, 360–378.
24 F. Wilcoxon, *Biometrics Bull.*, 1945, **1**, 80–83.
25 M. Friedman, *Ann. Math. Stat.*, 1940, **11**, 86–92.
26 Y. Benjamini and Y. Hochberg, *J. Roy. Stat. Soc. B*, 1995, **57**, 289–300.
27 R. B. Cattell, *Biometrics*, 1965, **21**, 405–435.
28 G. J. E. Nychas, E. H. Drosinos and R. G. Board, in *The Microbiology of Meat and Poultry*, ed. A. Davies and R. Board, Blackie Academic & Professional, London, 1998.

Chapter 6

Quantitative analysis of methyl green using surface-enhanced resonance Raman scattering (SERRS)

Iqbal T. Shadi, **William Cheung** and Royston Goodacre.

Author contribution to the work presented within the thesis

William Cheung: Sample preparation, initial pH profiling and concentration study, validation of the SOP.

Dr Iqbal Shadi: Experimental deign, sample preparation, data collection and analysis. Operation and maintenance of the Raman system.

Anal Bioanal Chem (2009) 394:1833–1838
DOI 10.1007/s00216-009-2896-4

ORIGINAL PAPER

Quantitative analysis of methyl green using surface-enhanced resonance Raman scattering

Iqbal T. Shadi · William Cheung · Royston Goodacre

Received: 30 December 2008 / Revised: 28 May 2009 / Accepted: 4 June 2009 / Published online: 21 June 2009

Abstract Surface-enhanced resonance Raman scattering (SERRS) spectra of aqueous solutions of the triphenylmethane dye methyl green have been obtained for the first time by use of citrate-reduced silver colloids and a laser excitation wavelength of 632.8 nm. Given the highly fluorescent nature of the analyte, which precluded collection of normal Raman spectra of the dye in solution and powdered state, it was highly encouraging that SERRS spectra showed no fluorescence due to quenching by the silver sol. The pH conditions for SERRS were optimised over the pH range 0.5–10 and the biggest enhancement for SERRS of this charged dye was found to be at pH 2.02, thus this condition was used for quantitative analysis. SERRS was found to be highly sensitive and enabled quantitative determination of methyl green over the range 10^{-9} to 10^{-7} mol dm^{-3}. Good fits to correlation coefficients were obtained over this range using the areas under the vibrational bands at 1615 and 737 cm^{-1}. Finally, a limit of detection of 83 ppb was calculated, demonstrating the sensitivity of the technique.

Keywords pH · Methyl green · Silver sol · Surface-enhanced resonance Raman spectroscopy (SERRS)

Introduction

Methyl green is a triphenylmethane dye used for identification of DNA, RNA, and other cell components; in histology it is a particularly useful counterstain that stains nuclei light green [1–7]. This dye has also been used for routine staining of biopsies from lymphoid glands to enable clinical pathologists to identify pyroninophilic cells [8] and has also been used to reveal sites of developing cartilage in human embryos [9].

Surface-enhanced resonance Raman scattering (SERRS) spectra of aqueous solutions of methyl green have been obtained in order to investigate the sensitivity of the technique as an alternative probe for the dye, and the concentration dependence of the signal intensities for quantitative analysis.

The theory of the SERRS effect has been outlined by Mullen et al. [10]. Briefly, SERRS signal enhancement arises from a combination of signal intensification via resonance Raman scattering (RRS) [11] and surface-enhanced Raman scattering (SERS) mechanisms [12–14] which in combination can increase the efficiency of the Raman scattering process by 10^{10}-fold or greater [15, 16]. SERS and SERRS have found wide utility as sensitive analytical and bioanalytical tools [17–20]. For maximum sensitivity, SERRS requires controlled aggregation of the colloidal sol used [21]. Surface enhancement of the Raman signals depend on the size of the colloidal particles [22], pH [23, 24], and the excitation wavelength employed. The surface plasmon absorption bands of metals such as silver and gold show wavelength-dependent shifts with metal particle size and a surface enhancement effect is achieved by choosing the Raman excitation wavelength to lie within the contour of the plasmon band [25].

The inability to interrogate intrinsically highly fluorescent molecules (for example methyl green) by Raman spectroscopy in the visible to near infrared (here defined as 488–830 nm because of the ready availability of lasers) can be overcome by using several alternative

I. T. Shadi · W. Cheung · R. Goodacre (✉)
School of Chemistry and Manchester Interdisciplinary Biocentre,
University of Manchester,
131 Princess St,
Manchester M1 7DN, UK
e-mail: roy.goodacre@manchester.ac.uk

Raman analytical techniques. These include, but are not restricted to, the use of Fourier-transform (FT) Raman at 1064 nm excitation or deep UV resonance Raman below 260 nm [26, 27], anti-Stokes vibrational bands, shifted subtractive Raman spectroscopy [28], and the use of the Kerr gate effect in time-resolved Raman spectroscopy [29]. In addition, surface-enhanced methods (*viz*. SERS and SERRS) can also be used to overcome both fluorescence and the analytically (relative) low sensitivity generally associated with Raman scattering of molecules in solution. Different kinds of media/substrates can be used to obtain SERS and/or SERRS signals of analytes. These include, for example, roughened electrodes [30], silver and gold colloids [31], nanoshell colloids [32], and so-called "gel-colls" which employ a hydrophilic swelling polymer (such as polyacrylic acids) in combination with a metal colloid [33].

Although the average SERS signal enhancement of molecules adsorbed on SERS-active silver surfaces has normally been measured to lie in the range 10^5–10^6, it has been known for some time that the SERS enhancement can be 10^{10} (or greater) in specific "hot spots" [16]. The observed SERS signal enhancement is currently hypothesized to be caused by a combination of effects: "classical electromagnetic field (EM) enhancement" and "the chemical effect". The former refers to surface-plasmon polariton resonances in colloidal systems such as silver colloidal particles (and nanoparticles of a few other metals) used as SERS substrates. The chemical effect, actually due to a combination of effects, is primarily a "first-layer effect" caused by dynamic (i.e. optically excited) electron (or hole) transfer from the metal to the LUMO (or HOMO) of the adsorbed analyte(s) in the first monolayer of the metal (e.g. silver) sol [34, 35].

In previous investigations we carried out SERS and SERRS studies on a wide range of dyes belonging to different structural classes and illustrated the power of Raman spectroscopy, when coupled with a surface enhancement effect, for characterisation of these dyes [23, 36–40]. In the current investigation reported herein we investigated SERRS spectroscopy, with laser excitation at 632.8 nm, to generate information-rich spectra from the triphenylmethane dye methyl green (Fig. 1). We investigate the pH dependence of the dye solutions for generating good SERRS. Following this optimisation we used these conditions for quantitative analysis of the dye.

Experimental

Reagents

Methyl green (Aldrich), poly(l-lysine) hydrobromide, M_r 4,000–15,000 (Sigma), silver nitrate (BDH), trisodium citrate, sodium hydroxide, and hydrochloric acid (Fisher) were of analytical grade. The dye was used without further purification. Double-deionized water was used for all experiments.

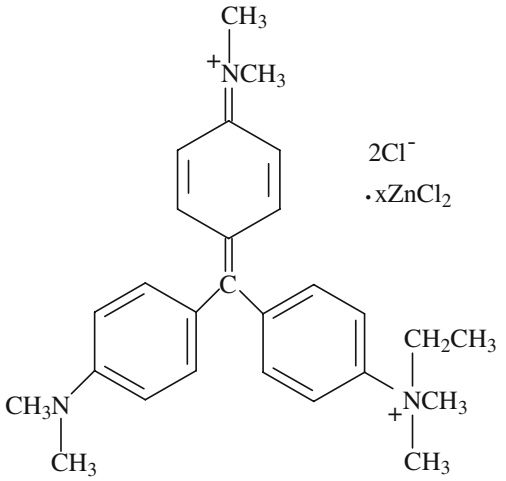


Fig. 1 Schematic diagram of the structure of methyl green

Methyl green solutions

Aqueous solutions of the dye in the concentration range 10^{-9}–10^{-3} mol dm^{-3} were prepared in double-deionised water. Samples were always made up fresh, immediately before analysis was carried out.

Colloid preparation

A silver colloid was prepared according to a modified Lee-Meisel procedure [25, 41]. All glassware was acid-washed with aqua regia (HNO_3-HCl, 1:3*v*/*v*) followed by gentle scrubbing with a soap solution. Silver nitrate (90 mg) was suspended in 500 mL deionised water at 45 °C and rapidly heated to boiling before addition of a 1% solution of trisodium citrate (10 mL) under vigorous stirring. The solution was held at boiling for 90 min with continuous stirring. The quality of the resulting colloid was checked by determining the wavelength of the absorption maximum in the visible region on a Perkin-Elmer Lambda-2 UV-visible spectrometer. Good quality silver colloids for SERS apparently have an absorption maximum at approximately 404 nm and this peak has a full width half height (FWHH) of ~60 nm [41] which demonstrates a narrow distribution of the particles. In addition, a Jeol JEM-200CX transmission electron microscope was used to inspect the colloids and revealed a mean of diameter of 50 nm (±1 nm standard deviation) for citrate-reduced sols (see Fig. 2 for an example TEM image). The nature of the Lee-Meisel colloid [15, 42], has been examined using visible absorption, photon

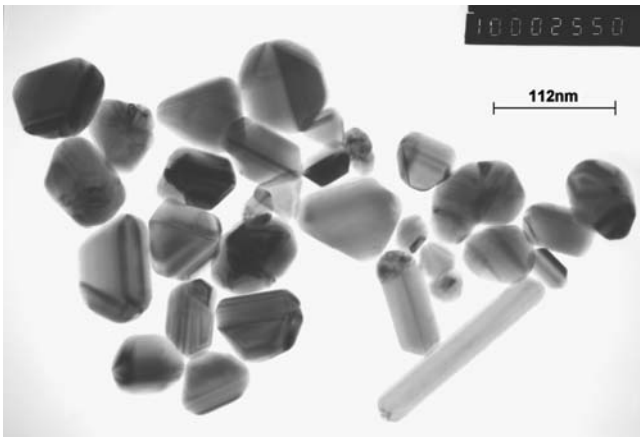


Fig. 2 Representative TEM image of citrate-reduced sols (magnification × 50,000)

correlation, and NMR spectroscopic techniques (data not shown) which confirm that the surface of the silver particles are covered with a layer of citrate with pendent negatively charged groups. As the methyl green dye itself is negatively charged (Fig. 1) and would repel the colloid, the subsequent addition of poly(l-lysine) coats the surface with pendent positively charged groups on the colloidal surface [43] thus allowing colloid and dye to aggregate. We thus adopted this approach to obtain good SERRS signals from the methyl green dye.

Sample preparation

Aggregation of the silver colloid particles was induced by poly(l-lysine). An aqueous solution of poly(l-lysine) (0.01%, 150 μL) was added to 1 mL silver colloid which had been diluted with 1 mL deionised water, followed by 150 μL methyl green solution and 35 μL of a 1 mol dm^{-3} aqueous solution of either HCl or NaOH. For quantification (*vide infra*) all the SERRS spectra were collected three times using the same analyte, colloid, and aggregating agent prepared three times; all measurements were made on the same day.

Instrumentation

Raman spectra were obtained using a Renishaw (Old Town, Wotton-under-Edge, Gloucestershire, UK) System 2000 Raman microscope, with a resolution of ~6.5 cm^{-1}. A 20-mW 632.8 nm HeNe laser source was used for all measurements and there was ~2 mW power at the sample. All SERRS spectra were collected by using 180° back-scattering geometry. An Olympus microscope objective of magnification 50× was used both to focus the incident laser light and to collect the back-scattered Raman light.

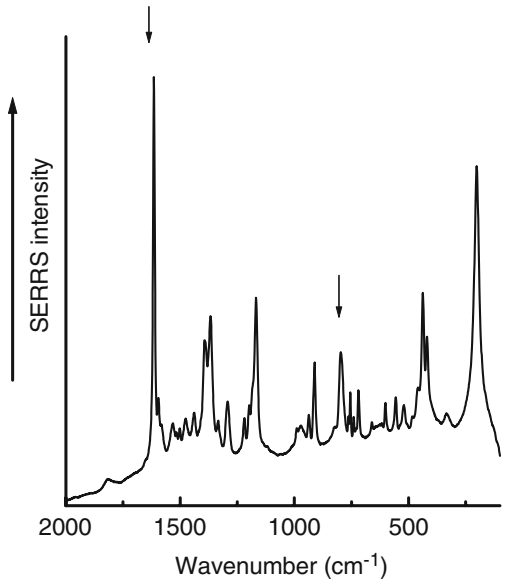


Fig. 3 A typical unprocessed (raw data) SERRS spectrum of 8.1×10^{-6} mol dm^{-3} methyl green using an excitation wavelength of 632.8 nm and a laser power at the sample of 2 mW. Vibrational bands at 797 cm^{-1} and 1615 cm^{-1} used for quantitative analysis are indicated by *arrows*

SERRS pH dependence

A series of SERRS spectra from 10^{-7} mol dm^{-3} methyl green dye were collected over the pH range 0.5 to 10.0. The optimum pH conditions were determined by plotting the logarithm (log_{10}) of peak area of the vibrational band at 1615 cm^{-1} from methyl green *vs*. pH.

Reproducibility/time dependence

Following mixing of the dye solution with the silver sol and aggregation of the colloidal particles, SERRS signal intensities of methyl green grew with time until after ~5 min when

Table 1 Tentative SERRS[a] vibrational band assignments for methyl green

Wavenumber (cm^{-1})	Tentative band assignments
797	CH_3 stretching vibration
1369	CH_3 sym deformation
1400	CH_3 antisym deformation
1445	CH_3 antisym deformation
1479	CH_2 scissoring
1603	Ring stretch
1615	Ring stretch
1620	C=C stretch

[a] λ_{exc} 632.8 nm

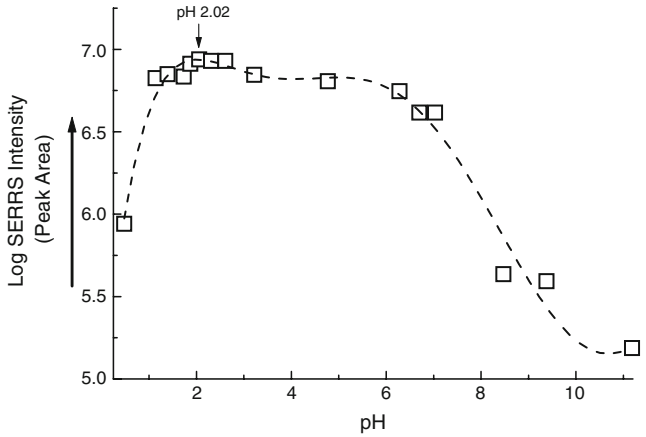


Fig. 4 SERRS pH profile for methyl green using a dye concentration of 5.5×10^{-7} mol dm^{-3}. Logarithm (Log_{10}) intensity (peak area) of the 632.8 nm-excited SERRS vibrational band at 1615 cm^{-1} of methyl green is plotted against the pH. The optimum SERRS response is indicated by an *arrow*. *Points* represent averages from three measurements

they remained constant. Therefore all spectra were collected 5 min post-aggregation.

Quantification of methyl green

The concentration-dependence of the SERRS signal from methyl green was determined by plotting the logarithm intensity (peak area) *vs*. the logarithm of methyl green concentration. For quantitative analysis the area under the vibrational bands at 1615 and 797 cm^{-1} were used; this was because all the other bands showed very similar trends and similar limits of detection (data not shown).

Results and discussion

All Raman and SERRS spectra were collected using exciting radiation of 632.8 nm on a Renishaw 2000 Raman microspectrometer. Because of the highly fluorescent nature of the methyl green dye normal Raman spectroscopy of the dye in the solution and solid state revealed strong fluorescence backgrounds (data not shown) and an absence of any Raman vibrational bands. By contrast, strong vibrational bands were observed at 201, 483, 755, 797,

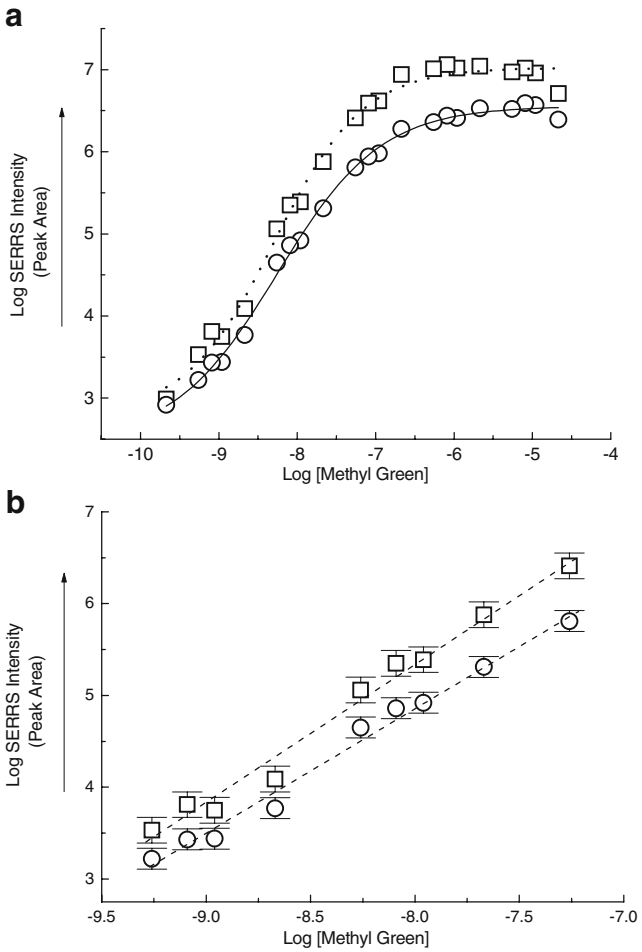


Fig. 5 (**a**) Methyl green in the concentration range 2.14×10^{-10} to 2.20×10^{-4} mol dm^{-3}. SERRS vibrational bands used for analysis are 1615 cm^{-1} (*squares*) and 797 cm^{-1} (*circles*). (**b**) Linear region used for semi-quantitative analysis in the concentration range 10^{-10} to 10^{-7} mol dm^{-3} (see Table 1). The excitation wavelength was 632.8 nm, and the laser power at the sample was 2 mW. *Points* represent averages from three measurements and *error bars* represent the standard deviation

1218, 1369, and 1615 cm^{-1} (Fig. 3) from the dye in solution. Tentative vibrational band assignments are given in Table 1. The excitation wavelength of 632.8 nm lies within the strong electronic absorption band centred on 635 nm, so these spectra are resonance-enhanced and surface enhanced from the silver colloid. In addition, it

Table 2 Analytical data obtained from linear regression for analysis of methyl green using SERRS at 632.8 nm

	Linear regions for calculations below							
Band	Slope	Intercept	Correlation coefficient (R)	Concentration range (mol dm^{-3})	RSD (±)	Orders of magnitude	X^2	LOD[a] (ppb)
1615 cm^{-1}	1.51	17.46	0.992	10^{-9}–10^{-7}	0.139	2	0.016	117
797 cm^{-1}	1.36	15.76	0.993	10^{-9}–10^{-7}	0.114	2	0.06	83

[a] Limits of detection (LOD) were determined by taking three times the standard deviation of the intercept (of non-log plots) and dividing by the slope, as adopted by Womack et al. [45]

was also encouraging to observe in these and all the SERRS spectra collected that the fluorescence was completely quenched.

As one can observe from the chemical structure of methyl green (Fig. 1) this dye is a charged molecule and its ionisation state depends on the pH of the surrounding medium. It has been recently demonstrated that the SER(R)S response is dependent on the pH of the solution of the analyte [23, 24], and therefore we decided to conduct a "pH profile" of the methyl green in order to optimise the signal enhancement. In this optimisation all conditions were kept constant except the pH which was adjusted using NaOH and HCl to lie between pH 0.5 and 10.0. The dye concentration was 5.5×10^{-7} mol dm^{-3} and the largest surface and resonance-enhanced peak at 1615 cm^{-1} was used to assess signal enhancement. Figure 4 shows the pH profile from this set of SERRS spectra and it was interesting to note that good SERRS spectra can be obtained for methyl green over the whole pH profile. Moreover, the optimum pH was found to be 2.02 and this was used for quantification of methyl green.

SERRS spectra of methyl green were collected over the concentration range 5×10^{-9} to 5×10^{-4} mol dm^{-3}. Plots of the area under two SERRS peaks at 1615 cm^{-1} and 797 cm^{-1} that gave good reproducibility are shown in Fig. 5a; these bands are chosen for illustrative purposes because very similar results were obtained for all other bands (data not shown). In this figure it can be seen that the signal enhancement reaches a plateaux after 10^{-7} mol dm^{-3} and this was because of the coverage of the colloidal silver particles being in excess of full monolayer of the dye, a phenomenon that we have observed previously [44]. The region from 10^{-9} to 10^{-7} mol dm^{-3} shows very good linearity (in this log-log plot) over two orders of magnitude (Fig. 5b) and this is confirmed using correlation coefficients and error measurements on the repeated acquisition (three in total) over this range (see Table 2 for the statistics). In Fig. 5b the mean peak areas of each vibrational band at 1615 cm^{-1} and 797 cm^{-1} are shown, with standard deviation error bars. The percentage difference for the three replicates was ~3% which is excellent for SERRS. We note that the slope of these two measurements was slightly different (Table 2); for the band at 1615 cm^{-1} this was 1.51 whereas for the 797 cm^{-1} band it was 1.36. This could indicate different mechanisms for the two lines (i.e., a mixture of electromagnetic *versus* chemical enhancement) but there was no direct evidence of this.

The limits of detection (LOD) were determined by taking three times the standard deviation of the intercept (of non-logged plots) divided by its slope, as adopted by Womack et al. [45] and used by us previously [23, 36–40]. These calculations are shown in Table 2 for the two bands and it is clear that the LOD for SERRS was 83 ppb (10^{-9} mol dm^{-3}) for the band at 797 cm^{-1}.

In conclusion, this is the first study to show SERRS from the triphenylmethane dye methyl green. SERRS has been shown to be a very sensitive technique for quantitative determination of methyl green with the advantage of fluorescence quenching by the silver colloids. The benefits of the SERRS technique are clearly demonstrated in that it was not possible to obtain normal Raman spectra of the dye in the solution and solid states. For quantitative analysis good linear fits over the 10^{-9} to 10^{-7} mol dm^{-3} range were observed. For this molecule SERRS spectra could also be obtained across a broad pH range, and the optimum pH was calculated to be 2.02. Thus we have clearly demonstrated excellent SERRS spectra of aqueous solutions of methyl green in order to investigate the sensitivity of the technique as an alternative probe for the dye. Future work will concentrate on investigating SERRS of this dye in histological sections.

Acknowledgments We are very grateful to the UK BBSRC for financial support.

References

1. Augulis V, Sepinwall J (1971) Stain Technol 46:137–143
2. Kurnick NB (1950) J Gen Physiol 23:243–264
3. Luo L, Liu J, Wang Z, Yang X, Dong S, Wang E (2001) Biophys Chem 94:11–22
4. Mori Y, Lennert K (1969) Electron microscopic atlas of lymph node cytology. Springer, Berlin Heidelberg New York
5. Pearse AGE (1968) Histochemistry theoretical and applied. Churchill, London
6. Scott JE (1967) Histochemie 9:30–47
7. Venkatesh S, Smith TJ (1998) Biotechnol Appl Biochem 27:265–267
8. Ahlquist J, Anderson L (1971) Stain Technol 47:17–22
9. McCann JA (1972) Stain Technol 46:263–265
10. Mullen KI, Wang D, Gayle Crane L, Carron KT (1992) Spectroscopy 7:24–32
11. Soga T (2000) Spectrochim Acta A 56:79–89
12. Kneipp K, Kneipp H, Itzkan I, Dasari RR, Feld MS (1999) Curr Sci 77:915–924
13. Champion A, Kambhampti P (1998) Chem Soc Rev 27:241–249
14. Brolo AG, Irish DE, Smith BD (1997) J Mol Struct 405:29–44
15. Rodger C, Smith WE, Dent G, Edmondson M (1996) J Chem Soc Dalton Trans 5:791–799
16. Nie SM, Emery SR (1997) Science 275:1102–1106
17. Graham D, Faulds K (2008) Chem Soc Rev 37:1042–1051
18. Jarvis RM, Goodacre R (2008) Chem Soc Rev 37:931–936
19. Bailo E, Deckert V (2008) Chem Soc Rev 37:921–930
20. Qian X-M, Nie SM (2008) Chem Soc Rev 37:912–920
21. Creighton JA, Blatchford CG, Albrecht MG (1979) J Chem Soc Dalton Trans 75:790–798
22. Gupta R, Dyer MJ, Weimer WA (2002) J Appl Phys 92:5264–5271
23. Shadi IT, Chowdhry BZ, Snowden MJ, Withnall R (2004) J Raman Spectrosc 35:501–510
24. Whitney AV, Van Duyne RP, Casadio F (2006) J Raman Spectrosc 37:993–1002

25. Lee PC, Meisel D (1992) J Phys Chem 86:3391–3395
26. López-Díez EC, Goodacre R (2004) Anal Chem 76:585–591
27. McCreery RL (2000) Raman spectroscopy for chemical analysis. Wiley-Interscience, New York
28. Bell SEJ, Bourguignon ESO, O'Grady A, Villaumie J, Dennis AC (2002) Spectrosc Eur 14:17
29. Matousek P, Towrie M, Stanley A, Parker AW (1999) Appl Spectrosc 53:1485–1489
30. Fleischmann M, Hendra PJ, McQuillan AJ (1974) Chem Phys Lett 26:163–166
31. Smith WE (1993) Methods Enzymol 226:482–495
32. Jackson JB, Westcott SL, Hirsch LR, West JL (2003) Appl Phys Lett 82:257
33. Bell SEJ, Sirimuthu NMS (2004) Analyst 129:1032–1036
34. Otto A, Mrozeck I, Grabhorn H, Akeman W (1992) J Phys Condensed Matter 4:1143–1212
35. Komiha N, Kabbaj OK, Chraibi M (2002) J Mol Struct (Theochem) 594:135–145
36. Shadi IT, Chowdhry BZ, Snowden MJ, Withnall R (2000) Appl Spectrosc 54:384–389
37. Shadi IT, Chowdhry BZ, Snowden MJ, Withnall R (2001) Anal Chim Acta 450:115–122
38. Shadi IT, Chowdhry BZ, Snowden MJ, Withnall R (2003) Spectrochim Acta A 59A:2201–2206
39. Shadi IT, Chowdhry BZ, Snowden MJ, Withnall R (2003) Spectrochim Acta A 59A:2213–2220
40. Shadi IT, Chowdhry BZ, Snowden MJ, Withnall R (2004) J Chem Commun 1436–1437
41. White PC, Munro CH, Smith WE (1996) Analyst 121:835–838
42. Munro CH, Smith WE, Garner M, Clarkson J, White PC (1995) Langmuir 11:3712–3720
43. Munro CH, Smith WE, Armstrong DR, White PC (1995) J Phys Chem 99:879–885
44. Clarke SJ, Littleford RE, Smith WE, Goodacre R (2005) Analyst 130:1019–1026
45. Womack JD, Vickers TJ, Mann CK (1987) Appl Spectrosc 41:117–119

Chapter 7

Quantitative Analysis of the Banned Food Dye Sudan-1 Using Surface Enhanced Raman Scattering (SERS) with Multivariate Chemometrics

William Cheung, Iqbal T. Shadi, Yun Xu and Royston Goodacre.

Author contribution to the work presented within the thesis

William Cheung: Development of overall experimental design, optimisation and troubleshooting of the instrumental response for collections of the SERS data, pH profiling, concentration study, spiking study of sudan-1 in chilli methanol extracts. All data collection and subsequent data processing, initial unsupervised analysis.

Dr Iqbal Shadi: Advice on Raman and SERS spectroscopy

Dr Yun Xu: Multivariate calibration of SERS data using ANN, SVM and K-PLS.

Quantitative Analysis of the Banned Food Dye Sudan-1 Using Surface Enhanced Raman Scattering with Multivariate Chemometrics†

William Cheung, Iqbal T. Shadi, Yun Xu, and Royston Goodacre*

School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7DN United Kingdom

Received: September 14, 2009; Revised Manuscript Received: January 8, 2010

Sudan-1 has been used for coloring food. However, recent alarms worldwide about the carcinogenic and mutagenic properties of azo-compounds have led to concerns over their human consumption. In the U.K. in 2005, over 570 products were found to be contaminated with the azo dye Sudan-1 and this and the health risks associated with this dye resulted in the subsequent international ban of this additive in all foodstuff, at all levels, relating to human consumption. These incidents have also necessitated the need for high throughput low cost reliable approaches for the detection and quantification of food contaminated by such azo compounds. While there are a small number of analytical techniques that can be considered portable, many lack sensitivity. By contrast, we show that employing a portable Raman spectrometer, using surface enhanced Raman scattering (SERS), can provide good sensitivity, such that Sudan-1 can be quantified in a complex food matrix reliably over the range of 10^{-3} to 10^{-4} mol L^{-1}. We also demonstrate that a variety of multivariate approaches including principal components analysis (PCA), partial least-squares (PLS) regression, artificial neural networks (ANNs), and support vector regression (SVR) can be employed for the chemical analysis of this dye in a quantitative manner. Compared to the commonly used univariate approaches, where the area under a single band in assessed, the advantage of using multivariate approaches is that these algorithms can analyze the full spectra directly and the laborious task of selecting and integrating marker appropriate quantitative spectral bands can be avoided thus greatly simplifying and speeding up data analysis.

Introduction

Sudan-1 (1-phenylazo-2-naphthol; Figure 1) is a synthetic azo compound with orange red appearance, widely used as a coloring agent with many commercial applications. It is commonly found in printing ink, color waxes, oils, petrol, solvents, plastics and polishes, as well as foodstuffs.[1] However, azo dyes have demonstrated significant carcinogenic and mutagenic properties, inducing tumors in the liver and bladder of mice, rats, and rabbits.[1−7] It has therefore been classified as a third category carcinogen by the international agency for research on cancer (IARC).

However despite this, several high profile international incidents involving the use of Sudan-1 were documented in the past decade. In India in 2003 Sudan-1 was detected in hot chilli products resulting in the U.K.'s food standard agency (FSA) to implement measures for the detection of Sudan-1 for all chilli products originating from India. In China in 2005, Sudan-1 was found in numerous foodstuffs from chilli oil and powder, chicken fast food shops, and turnip pickles causing widespread panic.[8] In the same year in the U.K., a Worcestershire sauce product was found to be contaminated by Sudan-1, which was latter traced back to adulterated chilli powder used in its production. This product was widely used in many supermarket products such as pizzas and in numerous ready meal products, leading to over 570 products having to be taken off the shelves. More recently Sudan-1 has also been detected in several dried spices products in South Africa forcing them to implement measures similar to India, China, and U.K. for the detection of Sudan-1 within their foodstuffs. Despite the international ban, these incidents have highlighted the need for a rapid, low cost, reliable method for the detection of Sudan-1 in contaminated foods.

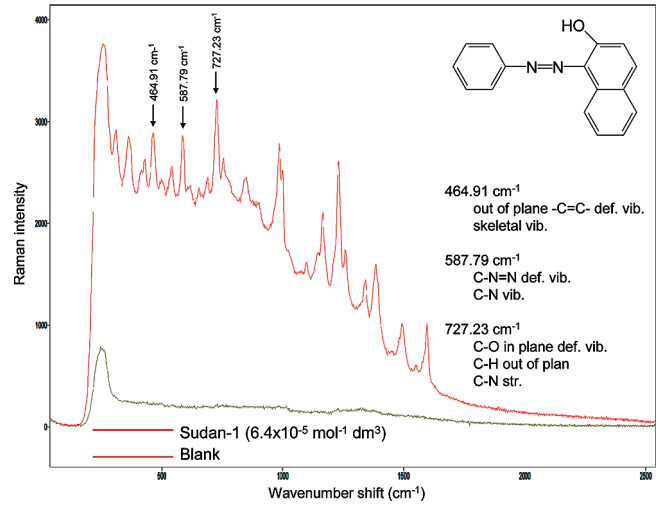


Figure 1. Raw SERS spectrum of Sudan-1 (6.4×10^{-5} mol^{-1} dm^3) with a blank of colloid. Inset is the chemical structure of this azo dye, with tentative assignments for the three highlighted bands, where vib. = vibration, def. = deformation, and str. = stretch.

Extensive literature is available for the detection of Sudan-1 in contaminated foodstuffs, most of which involve the use of liquid chromatography-based instrumentation for chromatographic separation of Sudan-1 from its food matrix and subsequent detection by various detectors. These include HPLC-

† Part of the "Martin Moskovits Festschrift".

* To whom correspondence should be addressed. E-mail: roy.goodacre@manchester.ac.uk. Tel: +44 (0) 161 3064480.

10.1021/jp908892n
Published on Web 02/03/2010

UV/vis,[9,10] HPLC fluorimetry,[11] flow injection chemiluminescence,[12] HPLC/APCI-MS/MS,[13,14] and HPLC/ESI-MS/MS.[15,16] In addition, a novel method for sample extraction has also been reported using molecularly imprinted solid phase extraction,[17] and the use of a glass carbon electrode has also been applied for the detection of Sudan-1 in foodstuffs.[8]

These LC-based methods mentioned for the detection of Sudan-1 are highly robust, reproducible, have relatively high sample throughput through automation, and have very good detection limits. However, they also have some serious draw backs, including sample pretreatment and extractions which can be labor intensive; they are also expensive and not currently portable. The direct analysis of Sudan-1 within the food matrix without sample preparation would hence be very desirable.

Raman spectroscopy is a physicochemical technique, which like the other vibrational spectroscopic method infrared absorption spectroscopy gives chemical information on the molecular make up of a molecule in a nondestructive manner.[18] Raman spectroscopy offers several advantages over mid-IR and near-IR spectroscopy; little or no sample preparation is required, as water is a weaker Raman scatterer no special accessories are needed for measuring aqueous solutions, and recently there have been many manufacturers producing small portable instrumentation. In combination, this makes this technique ideal for both qualitative and quantitative applications that involve the analysis of organic and inorganic chemicals. A main drawback of Raman spectroscopy is its low sensitivity. Raman scattering only comprises a very small fraction of those photons that are scattered, about 1 in 10^7, of the incident photons; the rest are Rayleigh scattered.[19] However, such weak Raman effects can be significantly amplified by coupling the analyte molecules to the surfaces of nanometer-sized nanometal structures, a technique termed surface enhanced Raman scattering (SERS).[20] Under normal conditions enhancements are on the order of 10^4-10^6,[21] as demonstrated for bacterial identification[22] and for DNA analysis,[23] but for certain molecules factors up to 10^{14} or even greater have been observed leading to the idea of single molecule detection.[24,25]

The objectives of this study were to combine SERS with chemometric analysis[26,27] for the detection and quantification of Sudan-1; the advantages of multivariate analysis over peak peaking (univariate) shall also be explored. In addition, we shall demonstrate the use of portable Raman spectroscopy for the quantification of Sudan-1 combined with SERS and chemometrics and further show that good signals from Sudan-1 can be generated when this azo dye is spiked into the complex food matrix of chilli powder.

2. Experimental Section

Reagents. The following reagents were provided as indicated and used without further purification: Sudan-1, poly (L)-lysine hydrobromide, (M_r: 4000−15,000), gold tetrachloride, trisodium citrate were all of analytical grade, purchased from Sigma Aldrich Ltd. (Gillingham, U.K.). Hydrochloric acid was purchased from Fisher Scientific (Loughborough, U.K.). Chilli powder was bought from a local supermarket.

Raman Instrumentation. Spectra were collected using an Ahura defender system (Ahura scientifics, Wilmington, MA) which is a hand-held, portable Raman spectrometer (using back scattered geometry). This instrument has a built in 4 mL sample chamber which provided up to 300 mW laser power at 785 nm ± 0.5 nm laser excitation and 2 cm^{-1} line width. All spectra were collected with an integration time of 5 s with the power setting set to low (30 mW). The spectrometer has a monochromator spectral range of 781−1014 nm, a Raman spectrum range of 250−2875 cm^{-1}, and a spectral resolution of 7−10.5 cm^{-1} (full width at half-maximum (fwhm)) across the range. Built into the Ahura spectrometer are Rayleigh rejection filters of OD 7, a silicon CCD 2048 pixels detector in direct dispersive detection mode, and a dispersion mode which is a single pass spectrometer (1200 grooves/mm blazed at 900 nm).

Surface Enhanced Raman Scattering. A gold colloid solution was prepared according to ref 18. All glassware was soaked in aqua regia (HNO_3:HCl 1:3 *v*/*v*) overnight. The colloid was prepared by reduction of $HAuCl_4$ by sodium citrate: 250 mL of 1 mM $HAuCl_4$ was first brought to the boil, after the addition of 25 mL of 38.8 mM sodium citrate, the mixture was allowed to boil for a further 30 min. The solution was then left to cool before use.

Aggregation of the citrate-reduced gold colloid particles was induced by poly(L)-lysine. The overall composition of the SERS solution was as follows: 1 mL of the gold colloid was diluted with 1 mL of ultra pure water, followed by the addition of 150 μL (0.01%) of poly(L)-lysine solution, 150 μL of sample solution, and finally 35 μL of the pH modifier (*vide infra*). The solution was vortex for 3 s between each successive addition. To allow for aggregation to occur, the solution was allowed to rest within the sample chamber for 1 min before spectral acquisition.

pH Profiling of SERS of Sudan-1. In order to determine the optimal pH which gave the largest SERS response, a stock solution of Sudan-1 (1×10^{-3} mol L^{-1}) was used for pH profiling. In pH profiling the pH of the SERS sample was adjusted between pH 1.4 to pH 5.3 by adding 35 μL of a pH acid modifier with the appropriate pH into the solution. Three replicate SERS spectra were taken at each pH. Following this the peak areas of the vibrational band at 725 cm^{-1} were plotted as a function of pH in order to find the maximum SERS response. This was found to be 2.0 (*vide infra*) and was used for all future SERS analyses.

Extraction of Sudan-1 from Spiked Chilli Powder. To evaluate the viability for the detection of Sudan-1 in a common complex foodstuff, we used chilli powder as the test matrix; this was chosen as chilli powder has been found in the past to be contaminated with Sudan-1.[8] In order to have accurate levels of Sudan-1 in chilli, methanol extractions from unadulterated chilli were prepared. This involved measuring 0.1 g of chilli powder and transferring this to a Falcon test tube to which 10 mL of analytical grade methanol was added and vortexed for 10 min. After, this the solution was then allowed to stand for another 5 min. Following this the brownish red supernatant was carefully removed using a pipet and used as the matrix into which Sudan-1 was to be spiked.

Known solutions of the dye dissolved in water were used to spike chilli powder to give a series of Sudan-1 standards with the following concentration range: 2×10^{-3}, 1×10^{-3}, 8×10^{-4}, 4×10^{-4}, 2×10^{-4}, and 1×10^{-4} mol L^{-1}. Six samples were prepared for each concentration, and each sample was measured three times by SERS. Note that for actual concentrations of the dye being analyzed by SERS one should be aware that these are 15.56 times more dilute than the stock. This is because, for example, in our sample preparation 1 mL of sol is diluted with 1 mL of H_2O followed by aggregation by 150 μL of poly(L)-lysine and 150 μL of the analyte and 35 μL of acid/base.

3. Data Analysis

For univariate analysis we performed a classical linear regression analysis between the areas under the curve (AUC) of the three selected vibrational bands within the SERS spectrum at 464.91, 587.79, and 727.23 cm^{-1} (peak apexes). These three bands were chosen as they appeared to have relatively high intensity with little or no background interference. Prior to regression, for each sample the AUCs from the three replicate SERS spectra from the six identical solutions were averaged prior to the analysis. Three regression models were built, one for each vibration band.

For multivariate analysis, the data set were baseline corrected by using asymmetric least-squares (ALS[28]), an adaptive baseline estimation algorithm. The unsupervised learning algorithm principal components analysis (PCA)[26,29,30] was used to visualized any natural variations within the data set.

Prior to multivariate regression the three replicated spectra SERS spectra from the six identical solutions were averaged prior to the analysis to reduce unavoidable variation caused during collection. For multivariate regression we employed partial least-squares (PLS) regression[26,29,31,32] to establish a multivariate linear regression model between the whole baseline corrected Raman spectra and the concentrations of the Sudan-1 in the solution. Considering the fact that the response of SERS intensity with the concentration of the analyte may not necessarily be linear, we also employed two nonlinear methods: ε-support vector regression (SVR) programmed in Matlab[33] and artificial neural networks (ANNs) using an in house program.[34] For ANN analysis a multilayer perceptron with a topology of 650−8−1 (650 input Raman scatters, eight nodes in the hidden layer and a single output node for the concentration of Sudan-1) was used employing a learning rate of 0.2 and a momentum of 0.8.[35]

For all three multivariate methods the performances of these three models on our data set was compared. The generalization performance of the multivariate regression models were assessed by using a double cross-validation scheme as detailed in refs 36−38. Each time we left out all of the six samples of one concentration which was used as the blind test set and the remaining samples were used as the training set. The model parameters, e.g., the number of PLS components, the kernel parameters of ε-SVR, etc., were tuned by using another inner k-fold cross-validation procedure on the training set only, where k is the number of concentration levels in the training set. Following this the model was built on the full training set using the optimal model parameter found by the inner cross-validation and used to predict the concentration of the samples in the test sets.

This was repeated until all the samples of each concentration level have been left out and predicted once and the root-mean-square error of cross-validation (RMSECV) as well as the predictive squared correlation coefficient (Q^2)[26] was calculated using the samples in the test sets according to the following two equations:

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}}$$

$$Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where n is total number of samples used for testing; y_i is the concentration of sample i ($i = 1, 2, 3, ..., n$); $\hat{y}_i$ is the predicted concentration of sample i, and $\bar{y}$ is the mean of the concentrations of all the samples used for testing.

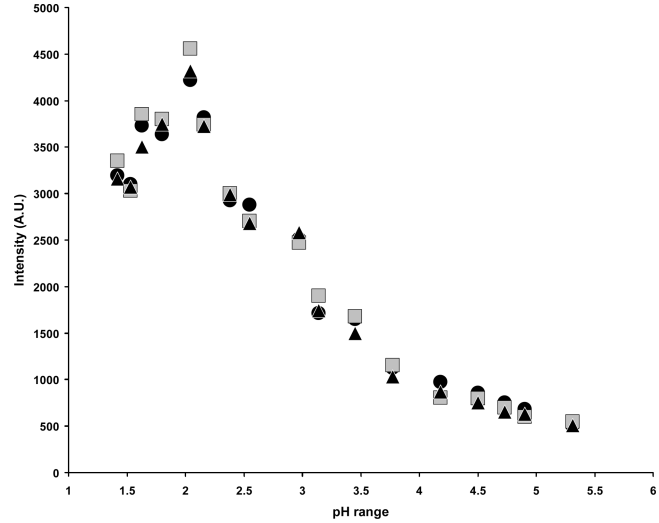


Figure 2. Peak area of the vibrational band at 725 cm^{-1} plotted against the pH profile from pH 1.4 to 5.3. The concentration of Sudan-1 in this preliminary experiment was 6.4×10^{-5} mol^{-1} dm^3, and the data were collected in triplicate.

RMSECV and Q^2 are the 2 most used and unbiased metrics to assess the predictive ability of a predictive model in regression analysis. Generally speaking, RMSECV gives an unbiased estimation of the error of the prediction when the model is used to predict the concentration of an unknown sample. Smaller the RMSECV is, the better the model is. Also, Q^2 is a scale independent, and also unbiased, metric to quantify the predictive ability of the model with values similar to the commonly used squared correlation coefficient R^2 in univariate regression analysis. The closer the value of Q^2 is to 1, the better the model is.

4. Results and Discussion

SERS and pH Profiling. A typical SERS spectrum of the pure Sudan-1 dye is shown in Figure 1 along with this azo dye's chemical structure. As can be seen the SERS spectrum is information-rich and the main bands that were used for quantification are 464.91, 587.79, and 727.23 cm^{-1} and these bands have been tentatively assigned in this figure.[39,40]

It has been shown that the SERS response of many analytes is dependent on the pH of the solution in which the analyte resides,[41,42] therefore we conducted a "pH profile" of the Sudan-1 in order to optimize the signal enhancement. The pH was therefore adjusted from pH 1.4 to pH 5.3 and SERS measurements taken. Using the strong band at 727.23 cm^{-1}, we constructed a pH profile (Figure 2), and it can be clearly seen that there is a general trend of decreasing intensity toward neutral pH and that it appears that at a pH of 2.0 the highest SERS response was seen. Therefore these conditions were used for the quantitative analysis of Sudan-1 in chilli powder.

Assessment of Spiked Chilli Solutions Using Univariate Analysis. As detailed above 2×10^{-3}, 1×10^{-3}, 8×10^{-4}, 4×10^{-4}, 2×10^{-4}, and 1×10^{-4} mol L^{-1} Sudan-1 was spiked into a methanol extract from chilli powder. Chilli powder was chosen as it has been found to be contaminated by Sudan-1; note we took preliminary spectra of the chilli powder and extract and no bands belong to Sudan-1 were seen (data not shown).

In order to see if the SERS signal from Sudan-1 was reproducible enough to allow its quantification the averaged peak areas of the three vibrational bands at 464.91, 587.79, and 727.23

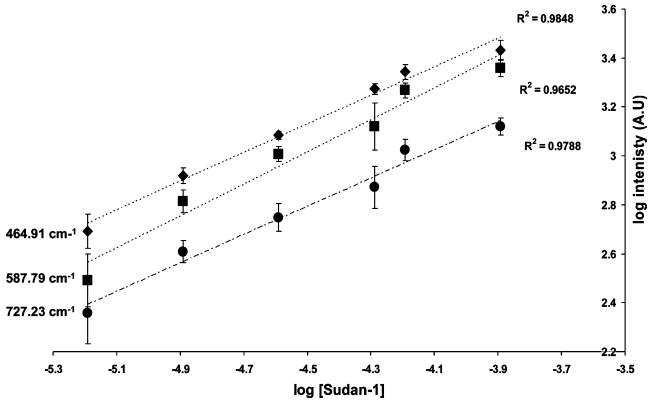


Figure 3. Peak area of three different vibrational bands plotted against the $\log_{10}$ of concentration of Sudan-1; this is equivalent to a concentration range of 6.4×10^{-6} to 1.3×10^{-4} mol^{-1} dm^3. The R^2 values are also shown for the three different bands. The average peak areas for three measurements are shown and the error bars show standard deviations.

TABLE 1: Comparison between Univariate Calibration and Multivariate Chemometric Methods for the Quantification of Sudan-1 from SERS

	univariate analysis from peak areas			multivariate chemometric methods		
	464.91 cm^{-1}	587.79 cm^{-1}	727.23 cm^{-1}	PLSR	SVR	ANN[a]
RMSECV	0.127	0.134	0.137	0.123	0.095	0.084
Q^2	0.911	0.906	0.903	0.916	0.961	0.977

[a] The values are the averages of 10 independent neural networks.

cm^{-1} were plotted against the log of the concentrations of Sudan-1 spiked into the chilli powder (Figure 3). Linear regression analysis was carried out on the areas of each of the selected vibration band in these log−log plots and showed very good correlation with the concentration and resulted in the R^2 values of 0.985, 0.965, and 0.979 respectively for 464.91, 587.79, and 727.23 cm^{-1} (Figure 3). As detailed in the data analysis section, these AUC for the bands were also assessed by using a double cross-validation scheme and the RMSECV and Q^2 are detailed in Table 1.

Principal Components Analysis (PCA). While the RMSECV and Q^2 values for the univaraite analysis are encouraging this is some room for improvement. We therefore decided to assess the spectra employing multivariate analyses. Initially PCA was used on the SERS spectra from all six Sudan-1 concentrations along with a blank (methanol extract from the chilli powder). The first two principal components (PCs) captured over 88% of total variance of the data, and the scores plot of PC1 and PC2 are shown in Figure 4A. It can be seen from this plot that the first PC (which accounts for 84.75% of the total explained variance (TEV)) captures the variation caused by different concentrations of spiked Sudan-1 and the second PC (which only accounts for 4.16% TEV) captures the minor experimental variations. The PCA loadings plot from the first principal component (Figure 4B) clearly shows that many spectral features relating to Sudan-1 (Figure 1) are important for this quantitative separation in PC1. It is noteworthy that the clusters of high concentrations show some overlap, however the separation improves as the concentration of the Sudan-1 decrease; in addition, the separation with respect to Sudan-1 concentration is not wholly explained in the first PC and the second one also has some minor influence as a parabolic trajectory from low to high concentration in both PC1 and PC2

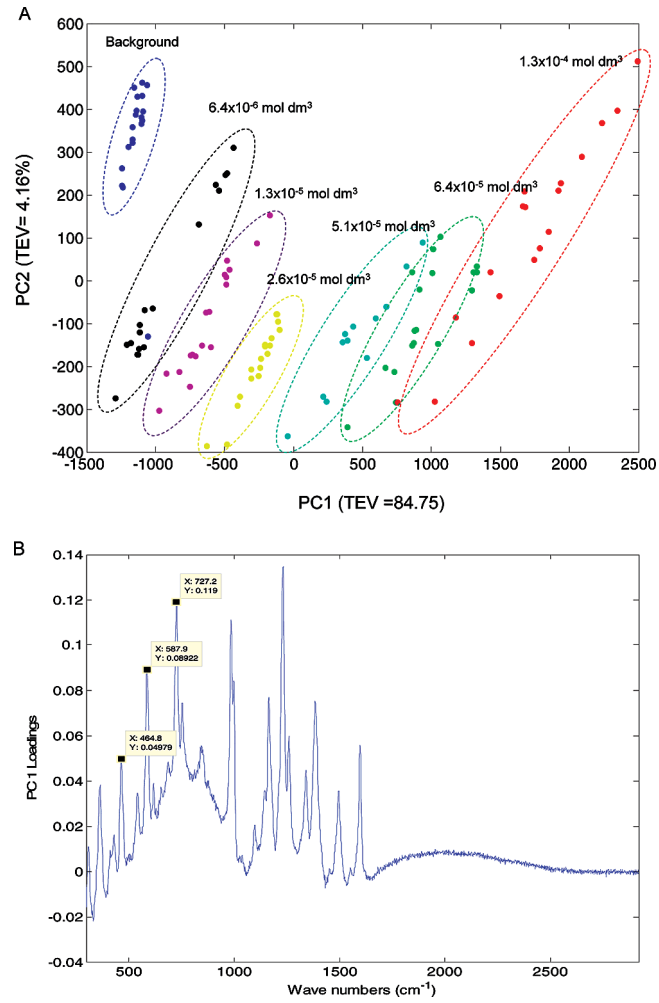


Figure 4. (A) PCA score plot from SERS spectra acquired from spiking Sudan-1 into chilli powder. The ellipses are drawn as a guide and have no statistical meaning. (B) The corresponding PCA loadings plot from the first principal component illustrating which spectral features are important for separation in PC1; highlighted are the three peaks used for univariate quantification.

(Figure 4). In combination this suggests that the SERS response of Sudan-1 may be nonlinear, particularly at high concentrations. Nevertheless, the results from the unsupervised PCA suggests that it is possible to use multivariate regression models to model the relationship between the SERS response of Sudan-1 and it concentration level in the chilli extract using the whole Raman spectrum.

Multivariate Regression. We therefore constructed multivariate models using PLS regression, ε-SVR and ANNs, and Figure 5 shows the predictions from the test data only and this clearly shows that the predictive ability of all three models (Figure 5a-c) was excellent as these predictions lie on the expected $y = x$ line.

We also made comparisons between the univariate regression and multivariate regressions by performing the same type of cross-validation on both the univariate and multivariate regression. The RMSECV and Q^2 obtained from the double cross-validation as described in data analysis section of the univariate regressions on 3 selected vibration bands are also shown in Table 1 along with the same metrics for PLS regression, ε-SVR and ANNs.

From the prediction metrics we can see that two nonlinear multivariate methods (*viz.*, ε-SVR and ANNs) obtained better results compared with those obtained by using the linear

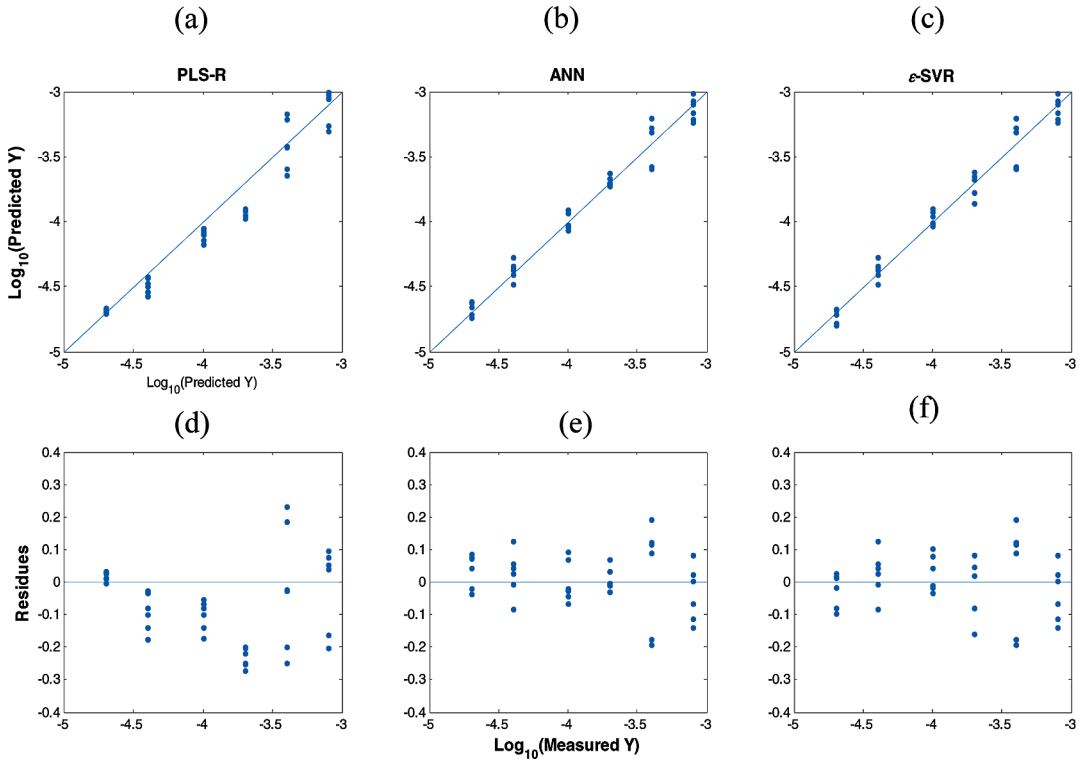


Figure 5. Plots of PLS-R results of quantification of Sudan-1 and residual plots. (a−c) Predicted concentrations against actual concentrations using PLS-R, ANN and ε-SVR respectively; (d−f) Plots of the residues (i.e., the difference between the actual concentrations and the predicted concentrations) against the actual concentrations from the same three models. Note the predictions do not include any dilution factors and refer to the $\log_{10}$ of concentration of Sudan-1 in the stock solution; after dilution the real concentrations are 15.56 times more dilute.

regression methods of PLS and AUC regressions. The results of ANN were the average from 10 independent multilayer perceptrons since even under exactly the same parameter setting ANN still gives (slightly) different results from different runs due to its approximation methodology.[43,44] The results from PLS is marginally better than that from the univariate regression on the vibration band at 464.91 cm^{-1} while the results of univariate regression on the vibration bands at 587.79 cm^{-1} and 727.23 cm^{-1} are the poorest, probably due to the fact that these bands were partially overlapped with neighboring bands and hence resulted in larger integration error when calculating the AUCs. This further justifies the needs of using multivariate regression for calibration. From the residue plots (Figure 5d−f) it can be seen that the residues of the linear regression method PLS are not all distributed around 0 while for other 2 nonlinear regression methods, ε-SVR and ANN, the residues are mostly distributed around 0. This further suggests that nonlinearity exists in the relation between the SERS response and the concentration level of the analyte.

Conclusions

A gold citrate reduced sol was employed in this study as it is recognized to be more homogeneous compared to silver sols.[45] The results obtained show that this is viable as a suitable substrate for semiquantitative analysis of Sudan-1, as good correlations coefficients of $R^2 = 0.985$, 0.965, and 0.979 were observed for the vibrational bands at 464.91, 587.79, and 727.23 cm^{-1} respectively and with standard deviations of only 13.3%.

The recent introduction of portable Raman spectrometers for chemical analysis, *in situ*, and the widespread interest of Raman spectroscopy require that investigations be made to determine the sensitivity of such instrumentation. Although portability is a distinct advantage this has only been shown to allow for qualitative analysis of analytes. However the enormous interest in applying the SERS technique to both qualitative and quantitative chemical analysis requires that tangible working models be available to demonstrate the viability of compact/portable Raman spectrometers. The data reported herein demonstrated that not only is it possible to get good sensitivity by employing the SERS technique using a portable Raman spectrometer (an Ahura defender system), but that multivariate chemometrics is essential for providing excellent sensitivity and quantitative data compared to classical univariate area under specific vibrational bands approaches. The limit of detection of Sudan-1 spiked into a complex matrix is calculated as 48 μg/kg chilli powder. This is currently an order higher than HPLC-based approaches which use considerable sample clean up prior to analysis,[9,10] and an area of work will be to optimize analyte extraction from the sample matrix prior to SERS.

Acknowledgment. I.T.S. and R.G. are very thankful to UK BBSRC for funding and Y.X. and R.G. are grateful to the European Commission's BIOTRACER Integrated Project (www.biotracer.org) for financial support.

References and Notes

(1) Moller, P.; Wallin, H. *Mutat. Res.* **2000**, *462*, 13.
(2) Martinek, V.; Stiborov, M. *Collect. Czech. Chem. Commun.* **2002**, *67*, 1883.
(3) Stiborova, M.; Martinek, V.; Rydlova, H.; Hodek, P.; Frei, E. *Cancer Res.* **2002**, *62*, 5678.
(4) Garner, R. C.; Martin, C. N.; Clayson, D. B. Carcinogenic aromatic amines and related compounds. Searle, C. E., Ed.; *Chemical Carcinogens*, 3rd ed.; ACN Monograph 182; American Chemical Society: Washington, DC, 1984; Vol. 1.
(5) Westmoreland, C.; Gatehouse, D. G. *Carcinogenesis (Lond.)* **1991**, *12*, 1403.
(6) Childs, J. J.; Clayson, D. S. *Biochem. Pharmacol.* **1966**, *15*, 1247.

(7) Marmion, D. M. *Handbook of US colorants for foods*; John Wiley: New York, 1979.

(8) Meiju, D. U.; Xiaogang, H.; Zihao, Z.; Shouguo, W. *Food Chem.* **2007**, *105*, 883–888.

(9) Daood, H. G.; Biacs, M. A. *J. Chromatogr. Sci.* **2005**, *43*, 461–465.

(10) Sproll, C.; Ruge, W.; Strichow, N.; Attig, D.; Marx., G. *Deutsche Lebensmittel-Rundschau* **2005**, *101*, 481–484.

(11) Pielesz, A.; Baranowska, I.; Rybak, A.; Wlochowicz, A. *Ecotoxicol. Environ. Safety* **2002**, *53*, 42.

(12) Gao, X.; Liu, H.; Song, Z.; He, X.; Dong, F. *Spectrosc.-Int. J* **2007**, *21*, 135.

(13) DiDonna, L.; Maiuolo, L.; Mazzotti, F.; Luca, D. D.; Sindona, G. *Anal. Chem.* **2004**, *76*, 5104–5108.

(14) Tateo, F.; Bononi, M. *J. Agric. Food Chem.* **2004**, *52*, 655.

(15) Zhang, Y. P.; Zhang, Y. J.; J.; Gong, W.; Gopalan, A. I.; Lee, K. P. *J. Chromatogr. A* **2005**, *1098*, 183.

(16) Calbiani, F.; Elviri, M. C. L.; Mangia, A.; Pistara, L.; Zagnoni, I. *J. Chromatogr. A* **2004**, *1042*, 123.

(17) Puoci, F.; Iemma, C. G. F.; Muzzalupo, R.; Spizzirri, U. G.; Picci, N. *Food Chem.* **2005**, *93*, 349.

(18) Chalmers, J. M.; Griffiths., P. R. *Handbook of Vibrational Spectroscopy*; John Wiley & Sons, Ltd.: Chichester, U.K., 2002; Vol. 1.

(19) Ferraro, J. R.; Nakamoto. K. *Introductory Raman Spectroscopy*; Academic Press: London, 1994.

(20) Fleischmann, M.; Hendra, P. J.; McQuillan, A. J. *Chem. Phys. Lett.* **1974**, *26*, 163.

(21) Moskovits, M. *Rev. Mod. Phys.* **1985**, *57*, 783.

(22) Jarvis, R. M.; Goodacre, R. *Chem. Soc. Rev.* **2008**, *37*, 931.

(23) Graham, D.; Faulds, K. *Chem. Soc. Rev.* **2008**, *37*, 1042.

(24) Kneipp, K.; Kneipp, H.; Tzkan, I.; Dasari, R. R.; Feld, M. S.; Dresselhaus, M. S. Nonlinear Raman probe of single molecules attached to colloidal silver and gold clusters. *Optical Properties of Nanostructured Random Media*; Springer-Verlag: Berlin, 2002; Vol. 82; p 227.

(25) Moskovits, M.; Tay, L. L.; Yang, J.; Haslett, T. SERS and the single molecule. *Opt. Prop. Nanostruct. Random Media* **2002**, *82*, 215.

(26) Brereton, R. G. *Cheometrics Data analysis for the laboratory and chemical plants.*; Wiley: Chichester, U.K., 2002.

(27) Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman., L. *Chemometrics: A Textbook*; Elsevier Science Publishers B V: Netherlands, 1988.

(28) Boelens, H. F. M.; Djikstra, R. J.; Eilers, P. H. C.; Fitzpatrick, F.; Westerhuis, J. A. *J. Chromatogr. A* **2004**, *1057*, 21.

(29) Wold, H. *In Multivariate Analysis*; Krishnaiah, K. R., Ed.; Academic Press: New York, 1966; p 391.

(30) Beavis, R. C.; Colby, S. M.; Goodacre; R. Harrington; P. Reilly; J. P. Solokolow, S.; Wilerson., C. W. *Encyclopedia of Analytical Chemistry*; Meyers, R. A., Ed.; Wiley: Chichester, U.K., 2000; p 11558.

(31) Alsberg, B. K.; Kell, D. B.; Goodacre, R. *Anal. Chem.* **1998**, *70*, 4126.

(32) Dixon, S. J.; Xu, Y.; Brereton, R. G.; Soini, H. A.; Novotny, M. V.; Oberzaucher, E.; Grammer, K.; Penn, D. J. *Chemom. Intell. Lab. Syst.* **2007**, *87*, 161.

(33) Cortes, C.; Vapnik, V. *Machine Learn.* **1995**, *20*, 273.

(34) McGovern, A. C.; Broadhurst, D.; Taylor, J.; Kaderbhai, N.; Winson, M. K.; Small, D. A.; Rowland, J. J.; Kell, D. B.; Goodacre., R. *Biotechnol. Bioeng.* **2002**, *78*, 527.

(35) Wasserman, P. D. *Neural Computing: Theory and Practice*; Van Nostrand Reinhold: New York, 1989.

(36) Cheung, W.; Xu, Y.; Thomas, C. L. P.; Goodacre, R. *Analyst* **2009**, *134*, 557.

(37) Filzmoser, P.; Liebmann, B.; Varmuza, K. *J. Chemom.* **2009**, *23*, 160.

(38) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smild, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. *Metabolomics* **2008**, *4*, 81.

(39) Degen, I. A. *Tables of chracteristic group frequencies for the interpretation of infrared and Raman spectra*; Acolyte Publications: Harrow, U.K., 1997.

(40) Socrates, G. *Infrared and Raman characteristic group frequencies: tables and charts*; John Wiley and Sons: Chichester, U.K., 2001.

(41) Shadi, I. T.; Chowdhry, B. Z.; Snowden, M. J.; Withnall, R. *J. Raman Spectrosc.* **2004**, *35*, 800.

(42) Shadi, I. T.; Cheung, W.; Goodacre, R. *Anal. Bioanal. Chem.* **2009**, *394*, 1833.

(43) Goodacre, R.; Kell, D. B.; Bianchi, G. *Nature* **1992**, *359*, 594.

(44) Goodacre, R.; Timmins, É.M.; Burton, R.; Kaderbhai, N.; Woodward, A. M.; Kell, D. B.; Rooney, P. J. *Microbiology* **1998**, *144*, 1157.

(45) Jarvis, R. M.; Johnson, H. E.; Olembe, E.; Panneerselvam, A.; Malik, M. A.; Azad, M.; O'Brien, P.; Goodacre, R. *Analyst* **2008**, *133*, 1449.

JP908892N

Chapter 8

8.1 Discussion

The studies presented in this research thesis (Chapter 2 to Chapter 7) involve investigating a combination of different analytical techniques and multivariate data analysis methodologies. Due to the diverse nature of the work presented it is more appropriate to discuss the following chapters as three separate areas of work rather than as a single body of work. Accordingly, this discussion section is divided into three sections: section 8.2 will discuss the results generated in Chapter 2 and Chapter 3, which focused on the discrimination of bacteria ~~to~~ based on VOC profiling of wound infection; section 8.3 will discuss Chapter 4 and Chapter 5, VOC profiling of contaminated meat spoilage using *S. typhimurium* as a model pathogen in conjugation with parallel metabolite profiling; finally section 8.4 will discuss Chapter 6 and Chapter 7 on the development of SERRS and SERS base analysis for the characterization of triphenylmethane and azo dyes.

8.2 Bacterial discrimination using DMS and VOC profiling in medical applications

In Chapter 2 the applicability of DMS as an alternative analytical detection technique for the discrimination of bacteria was investigated, as present bacterial identification is carried out by traditional microbiological techniques, such as the API identification system (Nucer 2006), in which the first step involves obtaining a pure culture. This is followed by a variety of morphological, serological, nutritional and biochemical tests, many of these tests require further incubation steps and are followed by interpretations and ultimately identification of the microorganism. Therefore in microbiology bacterial identification is an inherently slow process dictated by the need to generate sufficient biomass for analysis, the time scale involved is dependent upon the microorganism's growth rate (in certain extreme case members of the slow growing mycobacterial group may take to three to six weeks to grow), therefore in nearly all areas of microbiology a rapid and accurate method of bacterial identification is highly desirable (Goodacre 1994). Various research groups have investigated alternative methods of bacterial identification using spectroscopic (FT-IR/Raman) (Helm 1991,Yu and Irudavaraj 2005, Jarvis and Goodacre 2004,

Lopez-Diez and Goodacre 2004) and MS-based analytical methods (Goodacre 1994, Goodacre and Kell 1996, Sauer 2008, Sauer and Kliem 2010).

Initially rather than analysing the head space (HS) of the bacterial cultures, it was decided to assess the analytical potential of using DMS directly to see if bacterial discrimination could be achieved based on chemical information obtained from the bacterial biomass.

Pyrolysis-MS has been investigated in the past and found to be a reproducible method of sample introduction (Goodacre 1994, Goodacre and Kell 1996). An in-house Py-GC/DMS hyphenated system was therefore constructed. Validation and monitoring of the system were conducted using a whisky quality control (QC), the composition of the QC, and its subsequent instrumental responses, provide a useful means of tracking any instrumental drift that may be occurring during batch analysis and can be used to detect any data points that appear as outliers (i.e., atypical due to technical or biological artefacts).

The volatile species were generated using pyrolysis and the resultant pyrolysate were then separated using GC prior to DMS detection. Whilst this approach for bacterial discrimination has been studied before (Schmidt 2004; Shnayderman 2005, Prasad 2006), sub-species discrimination using Py-GC/DMS had not been successfully reported. Due to the complexity and high dimensionality of the Py-GC/DMS data, insufficient data pre-processing and multivariate analysis methodologies were implemented, and these authors focussed on using only selective sections of the DMS spectra (i.e., specific compensation voltages (CV) and scantimes) rather than utilising the entire DMS spectrum for discrimination. These issues of data pre-processing and data analysis have subsequently been addressed in later studies by the same groups (Prasad 2007). The use of dispersion field programming was not possible in this thesis due to less a sophisticated DMS analyser being employed, instead fixed CV scanning was used throughout these studies.

The DMS data were recorded in both positive and negative modes, reflecting the chemical reactions of the positive and negative adduct ions generated between the pyrolysate GC eluent and the positive $H^+(H_2O)_n$ and negative $O_2^-(H_2O)_n$ reactant ion present within DMS drift region (Schmidt 2004). The result from the Py-GC/DMS study represented therefore a three dimensional cube of scantime (matched to GC

elution) *versus* CV versus ion counts (see Fig.11 for examples). These data cubes were unfortunately complex and in order to use these for bacterial characterisation and identification purposes, data pre-processing and the application of chemometrics was necessary. This was achieved for spectral alignment using correlation optimised warping (COW) followed by baseline correction with asymmetric least squares (ALS). Exploratory data analysis involved PCA and subsequent bacterial identification from the data cubes with PLS-DA. This process is highlighted in the flow diagram (Fig.10) and will be further expanded on below.

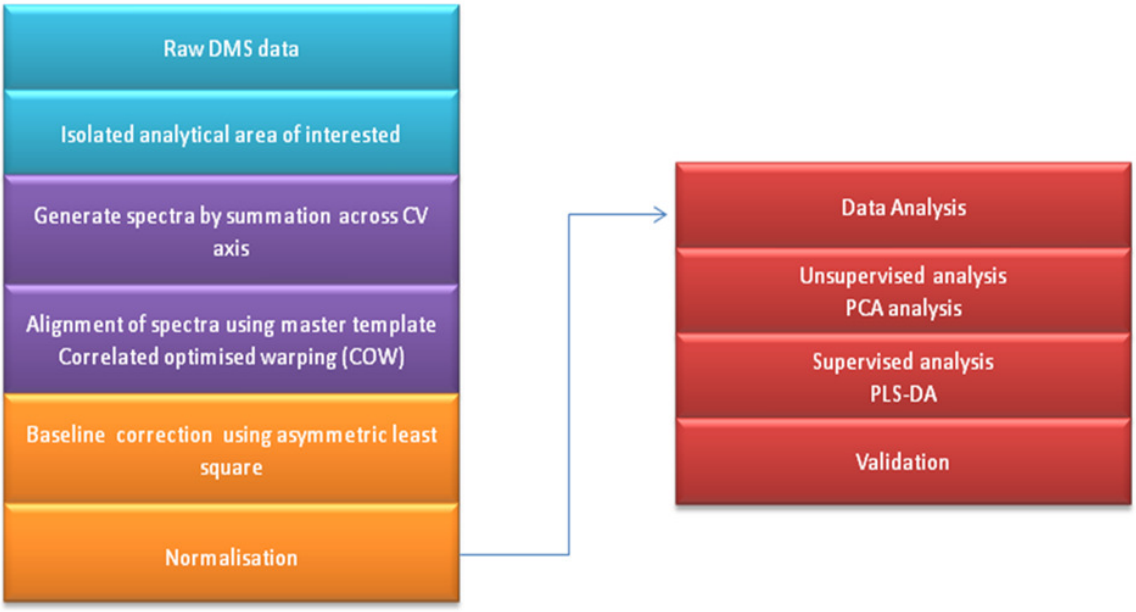


Figure 10 Flow diagram of the data processing methodology for Chapter 2.

The data are normally visualised as contour maps (Fig.11), where the colour intensity represents the peak intensity normalised to the biggest peak (Z-axis). In these contour maps which are for different samples the reactant ion peak (RIP) is variable due to the amount of pyrolysate generated and this is impossible to control. Therefore the RIP region was removed from the analysis and a subset ('slabs') containing the most information rich area of the DMS were used for analysis. Both modes of the DMS data were processed independently from each other. Once these slabs had been isolated they were then summed cross the CV axis, turning the three dimensional contour map into a two dimensional spectra (scantime *versus* intensity). These spectra were then grouped into the appropriate classes (strain identity), the

initial result of this is shown below (Fig 12A) along with the aligned version of the entire array (Fig 12B).

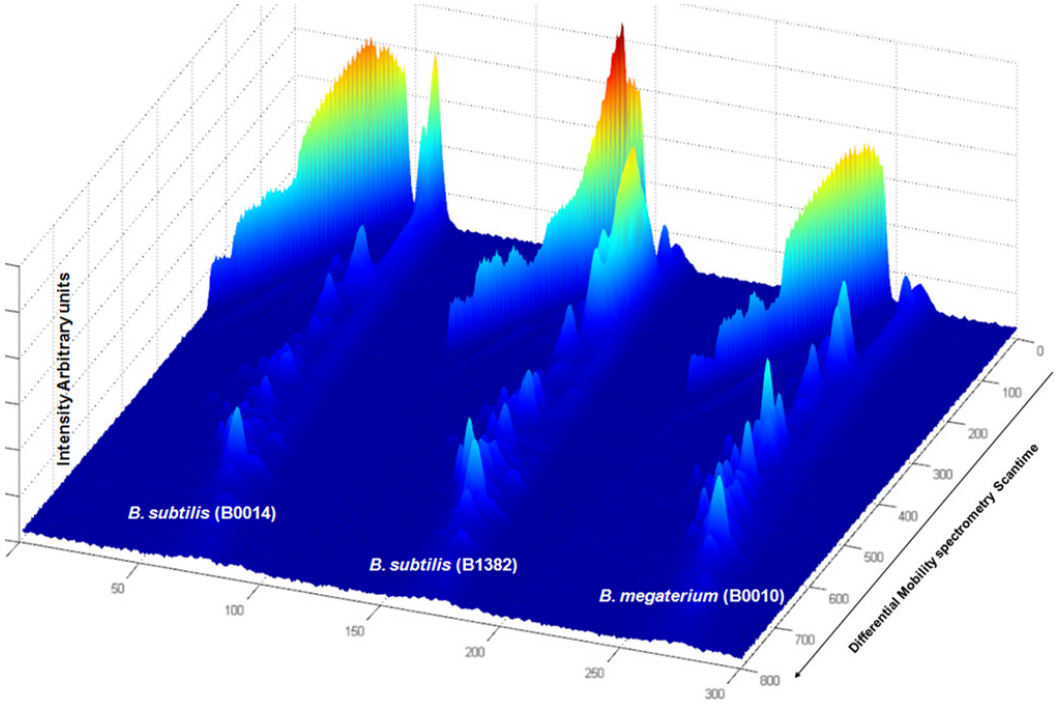


Figure 11 DMS negative mode response of the three bacteria used in the Py-GC/DMS study.

.

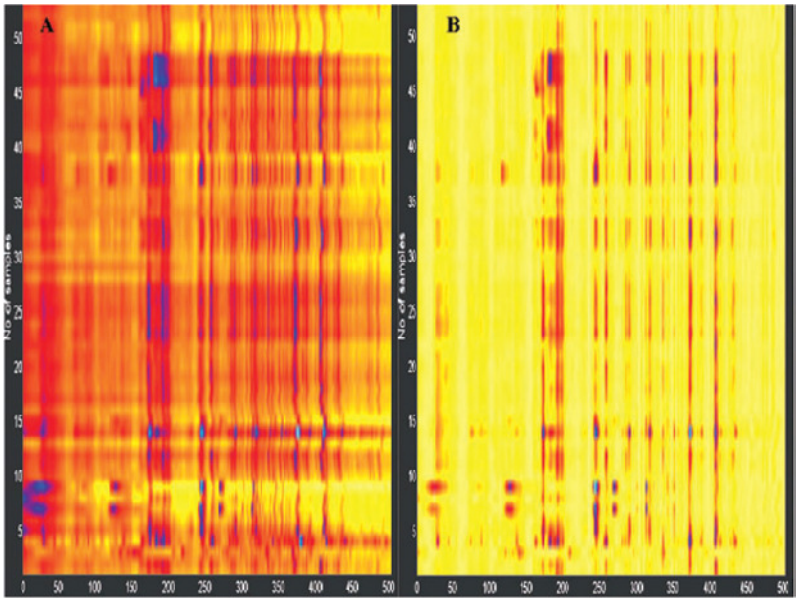


Figure 12 The effect of (A) before and (B) after alignment using correlate optimised warping (negative mode only).

The array was first aligned using COW. In this process a reference template was generated by combining the mean responses from all three strains; this acted as a

master template with features common to all three groups for the array to be aligned against. The whole array was aligned simultaneously to the master template. The array was split initially into different time windows to assess the optimal window for alignment. It was found that the optimum window size for these data was 10% of the original data points (i.e., if the data set has 500 data points then the windows was set to 50 data points). In order to allow peaks that span between different windows a 'slack sides' was also defined (again this was optimised to10% of the window size; i.e., 5 data points). The result of the alignment process is shown in the figure above (Fig.12B). The entire array was then baseline corrected using ALS followed by min-max normalisation prior to PCA and PLS-DA.

Due to the way the DMS data were summed across the CV axis prior to multivariate data analysis, this could potentially reduce the amount of chemical information obtained from the data. This was found to be necessary due to difficulties with aligning the contour maps in 3 dimensions. Alternative methods of data analysis should be investigated in the future to take advantage of the extra dimension of information provided by the CV axis.

Unfortunately peak identification in DMS is very difficult currently as there are limited libraries available. Consequently, the identification of significant variables from the loading plots, from either PCA or PLS-DA, that would appear to contribute towards discriminations, was highly difficult, subjective and misleading. In the future standard reference materials should be analysed to construct DMS and Py-GC/DMS libraries.

Despite the above problems with data processing and spectral interpretation, this study demonstrates that successful sub-species discrimination was possible using Py-GC/DMS combined with supervised learning methodology. Future studies should include more diverse sets of bacteria with greater phenotypic differences to assess the full potential of DMS. DMS is an interesting portable technology and has the ability to operate at ambient pressure. It is particularly well suited to the rapid screening of VOCs with sensitivity down to the ppt level (Krebs 2005), the system is robust, low cost, compact and its modular design allows it to be potentially mobile since it requires minimal laboratory and technical support. This makes DMS an

attractive alternative to MS for the rapid screening of VOCs and this was investigated further in Chapter 3.

One of the main disadvantages of analytical pyrolysis is that the molecules in the sample are thermally fragmented into pyrolysate which reduces the amount of structural information available. Therefore it would be desirable to have a sample introduction method that will leave the molecular ions intact allowing more chemical information to be obtained. In addition, significant amounts of time are spent on sample processing to generate sufficient biomass for analysis by Py-GC/DMS and it would also be highly preferable to have a rapid direct sampling approach. Static Head Space (SHS) sampling would permit a more direct sampling method, circumventing the need for additional culturing and harvesting procedures. Thus the suitability of thermal desorption into GC-MS with SHS sampling was investigated for *in situ* VOC measurements. Fortunately the opportunity arose to profile pathogenic bacteria from wound patients directly. Thus the second part of this discussion section focuses on a pilot study for the development of a VOC profiling protocol for wound analysis. Wound infection is the result of colonisation by pathogenic bacteria and the presence or absence of these pathogens may be evaluated directly by sampling the HS above the infected area. This method of SHS sampling was previously developed by Riazanskaia *et al.* (2008) and was adopted to profile the HS composition from lesion, boundary and control areas from five different subjects with leg ulcers. In these subjects significant arterial disease had been confirmed by ultrasound scanning.

As was shown in Chapter 3 GC-ion trap MS permits detailed VOC profiles from each of the sampling areas to be obtained, as well as capturing the subject's current physiological state. PCA indicated that partial discrimination between the control area, and boundary and lesion areas was possible. However, discrimination between the lesion and boundary area was not observed, and this may be due to the high similarity of the two sites and their close spatial proximity. The loadings biplot indicated that 2-propaniol 1-(1-methyethoxy) and methyl disulfide seemed to have contributed towards discrimination between the control and lesion regions (Chapter 3 Fig.5 and table 3). The VOC profile of each patient was highly characteristic to that individual; these profiles possessed a very strong fingerprint signature generated from the subject's own underlying metabolism, as well as other exogenous factors.

The limitations of the study in terms of disease diagnosis arose by a combination of low sample number and this high biological variation, which made discriminations based on TIC alone not only challenging but extremely difficult.

Despite these limitations, this study did demonstrate that PDMS is a robust SHS sampling device, and combined with thermal desorption, offers a reproducible method of sampling and introduction. This can be observed by examining the distance heat map of all samples plotted against each other (Fig.13).

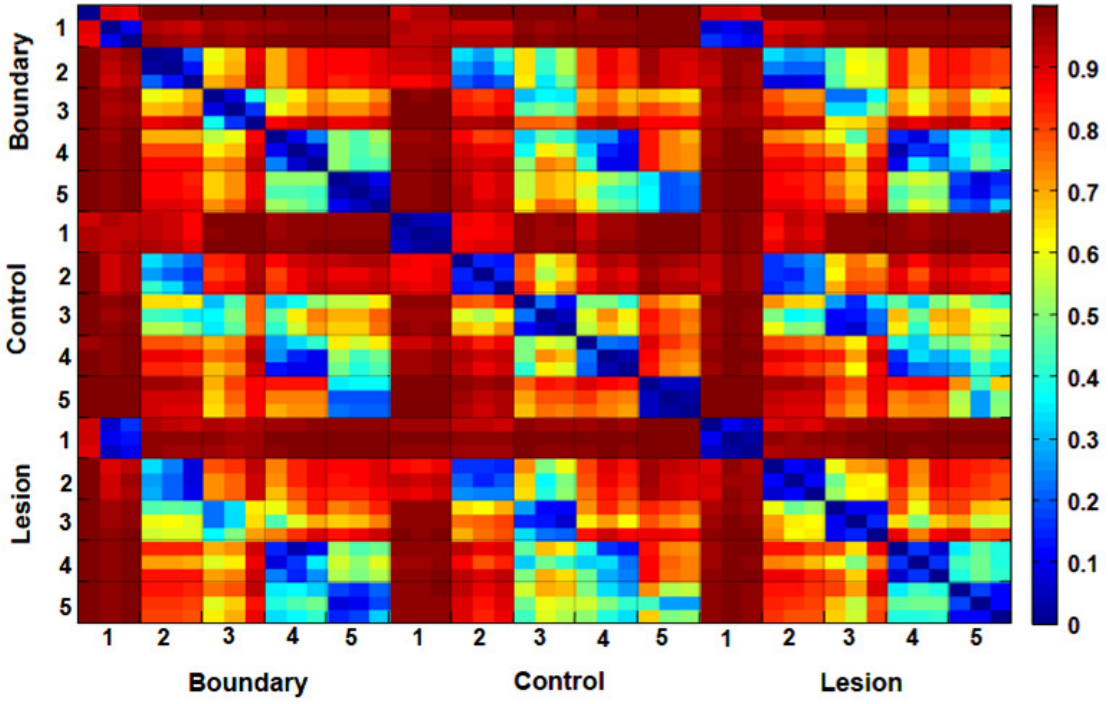


Figure 13 Heat map of the samples plotted against each other in the wound profiling study.

The blue diagonal relationship areas indicate regions of high similarity, whereas the red areas indicate region of low similarity and high variation. It can be seen that replicates from the same subject sampled in the same sites display very high similarity demonstrating excellent analytical reproducibility.

It should be recognised that increasing sample replications is not always possible, particularly for clinical samples where patients are under treatment and any new diagnostic test is a secondary goal. Therefore, there needs to be greater emphasis on more robust experimental design. For example, adequate matching of patients with disease to those who are healthy, in terms of gender, age, BMI, and ethnicity, as well

as other extrinsic factors such as metabolic rates, pharmaceutical use and diet, is important to avoid sample bias (Goodacre, 2007). In addition, diurnal effects should be considered, as should the oestrous cycle or fertility. All of these factors will significantly influence the subject's physiological profile, and these need to be carefully considered when generating a robust experimental design.

Data deconvolution steps for GC-MS will need to be incorporated in any future data analysis protocols, where the pure component (i.e. fragmentation profile) of each resolvable peak in the TIC spectra is extracted and its relative concentration level calculated. This process also allows the removal of any artefacts such as non-reproducible or erroneous peaks (e.g. siloxanes) from the data set, resulting in a peak table of all the common metabolites detected and their relative concentrations. This would increase the quality of the data obtained significantly, allowing a more detailed analysis and interpretation to be conducted. Currently these data pre-processing protocols are still under development.

Compared with other alternative screening methods such as electronic noses (E-noses), the data obtained from TD-GC/ITMS offers higher resolution data due to the incorporation of chromatographic separation prior to MS detection. However, the trade off is a decrease in sample throughput due to the extended sampling and analysis times. E-noses (Berna 2009) are small portable sensor platforms, using either metal oxide sensors (MOS) or conducting polymers (CP), which are selective towards particular classes of VOCs. The detector is usually an array of MOS/CP sensors operating in parallel. When the sensors are exposed to the sample, changes in the current across the individual sensors are recorded as function of concentration, and the combined response of each sensor is used to generate a very distinct signature profile of the sample (Dutta 2002). Like GC-MS measurements on VOCs, E-nose data require the use of supervised learning methodology to associate the response obtained to the original sample; however, unlike GC-MS, E-noses do not readily allow individual components of the VOC to be identified (Laothawornkitkul, 2008).

DMS can also be employed as an alternative method of VOC screening. It would offer an attractive alternative to GC-MS due to less intensive laboratory support, low cost of operation, and higher throughput capacities. In addition, its portability would

be particularly suited to on-site detection and point-of-care analysis. All of these methods are complementary and these methods could be combined where E-noses or DMS would be used for initial screening before more detailed VOC analysis using MS.

The initial result from this pilot study is encouraging as it demonstrates PDMS to be a reproducible and non-invasive method of SHS sampling, with the potential in the future for *in situ* real time sample profiling. Furthermore, the PDMS patches were found to be stable after sampling with no short-term sample degradation observed. Within a clinical context, the eventual aim would be to develop a method of rapid at patient screening, which would aid in the accurate detection of early wound infection and thus allow suitable targeted treatment to be administered.

8.3 VOC and metabolite profiling for the early detection and discrimination of S. *typhimurium* contaminated meat.

This second discussion section will compare and contrast the results of two orthogonal studies (VOC and metabolite profiling), conducted to investigate the effect of *S. typhimurium* on contaminated meat *vs.* natural spoilage.

S. typhimurium is associated with contamination of meat products and these bacteria are pathogenic to humans. Conventional method of microbial screening may be ineffective, due to the time involved in culturing and harvesting sufficient amount of cells required for identification. It is assumed that *S. typhimurium* proliferate utilising different metabolic pathways, in comparison to normal spoilage, and the VOCs emitted from the meat surface or metabolites found on the meat surface could be utilised for both rapid screening and discrimination.

Some of the deficiencies highlighted in the previous section, relating to VOC profiling using PDMS were subsequently addressed, such as: (1) robust experimental design for a time course study and sample analysis; (2) incorporation of a deconvolution step to improve the quality of the MS data obtained followed by; (3) MS library matching with NIST 02 database for peak identification, resulting in the generation of VOC metabolite peak table for multivariate data analysis. The

spoilage/contamination study contains multiple significant trends incorporated into an experimental design (such as the types of spoilage organisms and the progression rate of microbial growth with respect to time). For such types of data, the same variables are measured repeatedly at different time points in a nested fashion. Conventional PCA alone may not be a suitable method of analysis as the effect of one variable will no longer be associated to one. Rather multiple elements in the loading vector (co-linearity effect) will make subsequent interpretation from the loadings vector difficult and misleading. Therefore consensus PCA (CPCA) was utilised in order for these trends to be investigated individually, a simplified pipeline of the data processing steps is outline (Fig. 14).

In deconvolution, a time window was first predefined within the TIC chromatograph, within this time window all the chromatographs from the same sampling time point were stacked together. Orthogonal projection approaches, combined with Dubin-Waston statistics, were then used to provide an initial estimate of the pure spectra and the number of components observed within the time window.

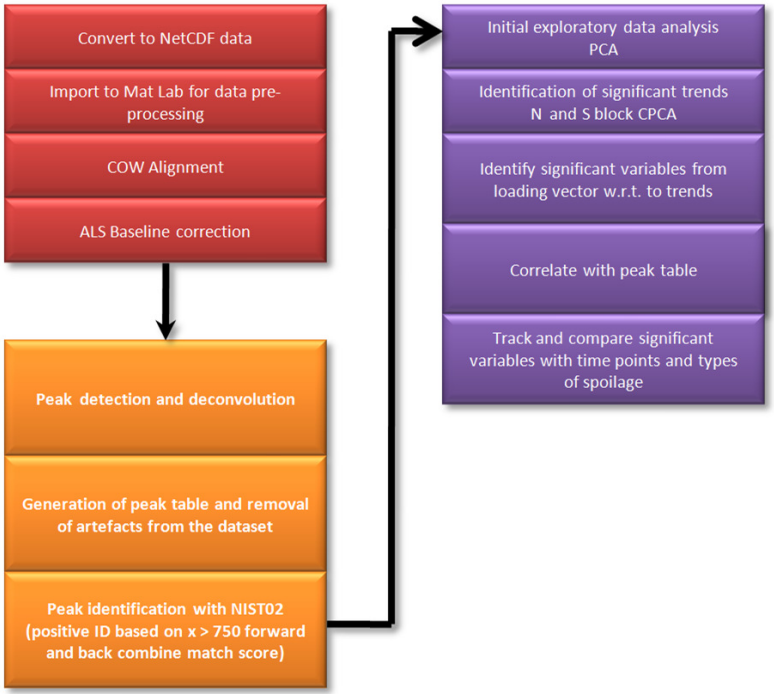


Figure 14 Pipeline for the preprocessing of TD-GC/ITMS data.

Alternating least squares (ALS) was then used to deconvolute the concentration profiles. The pure mass spectra with non-negativity constraints were applied to both the concentration profile and spectra, and unimodality constraints were then applied

to the concentration profile only (Chapter 3 Fig. 2). The deconvoluted concentration profile and the pure component (i.e. MS fragmentation profile) were then visually inspected and manually matched against a NIST 02 MS library database. This process was iteratively repeated for the all remaining time windows until the entire TIC chromatograph was deconvoluted. This is a highly time consuming and labour intensive process as certain parameters need to be repeatedly optimised to obtained satisfactory results (i.e. the length of the specify time window, and number of pure component.)

In order to save time, during the initial exploratory data analysis, only areas of analytical interest deemed to be of statistical significance were deconvoluted and analysed, but for the final data analysis the entire TIC spectra were fully deconvoluted.

Any siloxane peaks or peaks with irregular shape, low MS match score (less than 600 forward and backward combined match MS score) and non-reproducible retention times were removed from the data set prior to multivariate analyses.

The low scan rate of the QIT mass analyzer (1 scan per s), necessitated the need for the deconvolution step to be incorporated, in order to have better confidence on the peaks detected and also facilitate their identification. The final result is a list of all commonly detected metabolites across all time points and types of spoilage/contamination with their calculated concentration values; this was then subjected to multivariate data analysis.

The initial results from PCA suggest that spoilage progressions with respect to time are captured by PC1 and the differences between the two types of spoilage/contamination are largely captured by PC2. It was evident from this that the separation between the two types of spoilage increases with time; indicating *two distinctively different processes* are taking place (Chapter 4 Fig. 3). Results from N block and S block CPCA also confirmed this, all time points and types of spoilage can be clearly discriminated after 24 h. There was a larger degree of overlap between data points for 24 and 48 h for *S. typhimurium* contamination compared to natural spoilage; however, all time points except for 0 h irrespective of spoilage types can be discriminated by both PC1 (24-48 h) and PC2 (48-72 h) in the N and S block CPCA (Chapter 4 Fig. 4).

This study demonstrates the applicability of a VOC screening approach and its potential as a rapid method for the detection of *S. typhimurium* contamination. Discrimination can be readily observed within 24 h from the onset of contamination, compared to traditional microbial identification techniques such as API tests. These biochemical tests require a minimum of 48-72 h incubation time plus additional time for interpretation of the results, by then it may be too late as the contaminated meat product may have already left the shop floor (Ellis and Goodacre 2001). It is clear from the data above that the VOC screening method developed could allow more immediate action to be taken. This type of SHS sampling approach could lead to the eventual development of low cost disposable chemical sensor based on enzyme assays incorporated into the food packaging itself. This works by detecting specific VOC markers associated with *S. typhimurium* contamination. Positive identification would result in a visual change in the colour of the packaging warning that the meat product is no longer suitable for human consumption.

Although discrimination was possible after 24 h using SHS sampling with PDMS, discrimination was not observed at earlier time points, this may be due to insufficient bacterial biomass on the meat surface whereby the concentrations of the emitted off odours were too dilute to be captured effectively using SHS sampling. This could be further improved on by implementing a dynamic Head Space (DHS) sampling methodology, In DHS the entire volume of the HS inside the sample chamber is evacuated and passed through an absorbent material, this allows for significant enrichment in the levels of trace VOCs permitting a greater sampling sensitivity, and may allow discrimination to be achieved at earlier time points.

Based on the type of VOCs identified from this study, the trapping performance of the PDMS seems to suggest it is predominately suited for the trapping of polar compounds and has little or no affinity towards non-polar chemical species. Alternative adsorbent material such as Tenax-TA would allow trapping of both polar and non-polar VOCs permitting a much greater range of VOCs to be captured, thereby generating a more detailed snap shot of the VOCs present in the HS during spoilage/contamination.

Finally with respect to VOC analyses, the PDMS patches give a very characteristic background signature. These features, whilst not desirable (siloxane peaks), were

found to be highly reproducible in terms of intensity and retention time All the siloxane peaks were identified and matched by retention time and fragmentation profiles, these were then used as land marks for chromatographic alignments, after which they were all removed from the data set prior to analysis. In the future, the use of internal standards spiked directly in the HS prior to sampling should be introduced in order to account for small variations in adsorption capacity between PDMS patches and would also aid in identifying erroneous samples.

The second part of this discussion section summarises the result of the metabolite profiling study conducted in parallel with the VOC work. As mentioned previously it is assumed that *S. typhimurium* proliferate using different metabolic processes in comparison to natural spoilage. Therefore metabolite profiling was employed to investigate the intracellular metabolic differences between *S. typhimurium* and natural spoilage for discrimination and elucidation of the possible metabolic pathways that are involved.

A variation of parallel factor analysis (PARAFAC2) and a paired two sample Wilcoxon signed rank test were employed to analyse the data, in order to identify any significant changes in the metabolite profile between the two types of spoilage. Initial findings suggest that discrimination was possible after 36 h. However, due to the large number of time points (classes) and low number of observations (biological replications), the statistical weight of the result was somewhat limited. Therefore for greater statistical significance to be obtained, a second more detailed experiment with only 4 time points (0, 24, 48, 72 h), but with greater analytical and biological replicates, was conducted

As mentioned earlier in this discussion section the spoilage study has multiple significant trends incorporated in its experimental design. Therefore, a three way model would be more appropriate to analyse this type of multi-way data, permitting the effect of influential trends to be investigated individually. PARAFAC2 was utilised rather than CPCA which was employed in the VOC profiling study. This was because the initial metabolite profiling study was analysed using PARAFAC2 therefore, for the sake of consistency and comparison with the first study, it was also employed for the final spoilage experiment.

Half of the metabolites were shown to be significantly different across time points. This is likely to be the effect of increasing bacterial population on the surface during the course of the experiment. This trend is easier to visualise using the Hotelling T^2 statistics where a significant difference between the two types of spoilage/contamination is observed after 24 h with a $p < 0.05$ (Chapter 5 Fig. 3). This trend was consistence with the result obtained from the parallel VOC measurements and strongly inferred that two distinct metabolic processes were taking place between *S. typhimurium* compared to natural spoilage microflora. In each time point the TVC between the two types of spoilage were equivalent to each other, therefore the observed separation was not due different bacterial load influence but rather as a result of different metabolite profiles being detected.

This work clearly demonstrates the utility of metabolite profiling as a potential tool for the early detection of *S. typhimurium* contamination without the need for conventional microbiology approaches. The data were consistent with the VOC profiling and demonstrated that discrimination can be achieved using either approach within 24 h from the onset of contamination. In order to further improve our understanding of *S. typhimurium* proliferation, the next step would be how best to integrate and correlate the two different levels of metabolite data (VOC and metabolites profiling) together into a single coherent framework. Moreover the information acquired from the metabolite profiling study could be used to help identify potential metabolic pathways used specifically by *S. typhimurium*, as both studies confirmed that *S. typhimurium* utilised distinctly different metabolic process compared to natural spoilage.

8.4 Comparison of SERRS and SERS for the analysis of aqueous dyes.

Following on from the theme of food analysis in the previous section, banned or illicit chemicals (which may cause potential health problems) are sometimes intentionally spiked into foodstuffs to enhance their visual appearance and hence also increase their commercial value. This final discussion section will focus on the analysis of aqueous dyes using enhanced vibrational spectroscopy. The first part details a method development protocol using surface enhanced resonance Raman

scattering (SERRS) for the characteristion of the triphenylmethane dye methyl green in water. This analysis protocol was then further adopted and chemometrics incorporated for the quantification of the banned azo dye Sudan-1 using surface enhanced Raman scattering (SERS).

SERRS and SERS are variations of Raman spectroscopy (RS) in which the normally weak Raman scattering signal is significantly amplified by several orders of magnitude. In SERRS two separate mechanism of enhancement are involved. In the first the applied excitation frequency is tuned to be sufficiently close, to the molecular absorption maxima of the analyte. The subsequent Raman signal is significantly amplified via the resonance effect to give greater signal intensity compared to conventional Raman scattering (Mikhonin and Asher 2006, Tuschel 2010). In the second mechanism this signal intensification is further amplified by coupling the analyte to metallic nanoparticles to give surface enhancements, permitting a combined enhancement factor of approximately 10^{14} to be achieved for certain analytes (Roger 1996, Nie and Emery 1997).

Methyl green is a triphenylmethane dye widely used in the field of histology for the straining of DNA, RNA, nuclei and other cellular components (Umemura 2003); it has been used in biopsies of pyroninophilic cells (Zhu 2007) and for the investigation of cartilage development in human embryos (Lee 2000). Literature regarding the characterisation of methyl green is somewhat limited; this may be due to the difficulty of acquiring an Raman spectra because of the inherently high fluorescence background from this dye.

Preliminary result demonstrates that the protocol was sufficiently robust and can be successfully applied to the analysis and characterisation of methyl green with a LOD in the ppb range. The fluorescence effect within the methyl green SERRS spectra was significantly quenched by the silver nanoparticles and utilised for subsequent enhancement. Other spectroscopic methods for the detection of methyl green have recently been reported in the literature using resonance Raman spectroscopy (RRS) and tungstate ion resonance Rayleigh scattering with a LOD reported to be 800ppb and 18 ppb respectively (Yu 2010).

Triphenylmethane dyes are extensively used in the textile industry (Culp and Beland 1996), in particular malachite green is used for the colouring of wool, silk and nylon.

It is also widely employed as a disinfectant in both the aquaculture (Schnick 1988) and poultry industries (Girigoswami and Mandal 2002), this is due to its effectiveness in controlling fungal growth, its low price and commercial availability. However, because of its ubiquitous used concerns regarding its environmental impact and possible introduction into the human food chain has been questioned. The toxicity and mutagenic properties of triphenylmethane dyes have been well documented since the early 1990s (Nilsson 1993), various triphenylmethane dyes namely gentian violet and malachite green have been shown to be highly carcinogenic (Culp 2005, Docampo and Moreno 1990) inducing lung, renal and hepatic tumours in mice. In 2005 malachite green was subsequently banned in the USA and Europe (Srivsatava 2004); however, it is still illegally used in some regions around the world, therefore the development of detection methods are necessary in order to monitored its relative concentrations in aquacultural and poultry live stock deem for human consumption (Jiang 2009).

The work presented concerning the analysis and characterization of methyl green could in theory be applied to the detection of other triphenylmethane dyes with little modification. Each dye may have a slightly different pH requirement for generating optimal SERRS response, but this can be identified in the initial pH profiling step and modifications can be made to the protocol accordingly. Due to the similar chemical structure of all triphenylmethane dyes is it likely that most if not all will exhibit similar high fluorescence backgrounds, the use of SERRS has shown to be effective at quenching these effects and capable to generating detailed SERRS spectra.

The next section details the application of the same analysis protocol for the characterization of azo dye Sudan-1 (1-phenylazo-2-naphthol) using an alternative Raman system. Sudan-1 is a bright red dye, is also a known human carcinogen and has been subsequently banned from human consumption worldwide. However despite this, several high profile health scares in India, China and England involving this dye have been well documented. Many screening methods for the detection of Sudan-1 have been reported but few are spectroscopic based and require samples to be sent back for laboratory analysis using predominately HPLC-based methods. As such they require some degree of sample pre-processing and extraction protocols;

here we present an alternative method for the rapid screening of Sudan-1 using a portable Ahura defender Raman spectrometer.

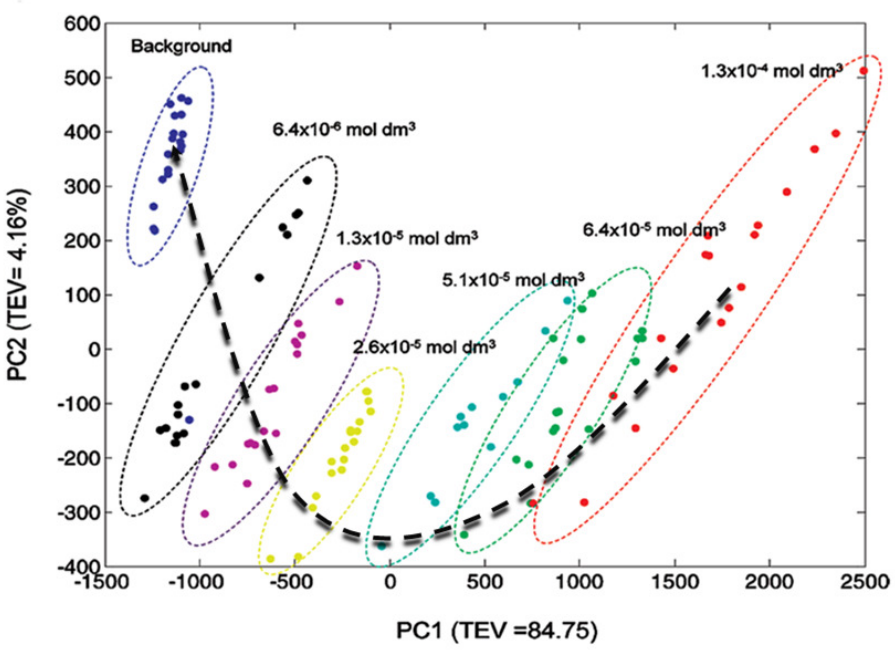


Figure 15 PCA scores plot of the Sudan-1 SERS responses. The arrow indicates a parabolic non linear relationship was observed from high to low concentration.

The PCA scores plot (Figure.15) displays a parabolic relationship from high to low concentration (x-axis from left to right). This separation with respect to concentration could not be fully accounted for within PC1, and PC2 demonstrates some minor influence; in addition, the technical variations of the sample analysis were also captured by PC2. The initial loading from PC1 (Chapter 7, Figure 4B) indicates that many vibrational bands within the Sudan-1 SERS spectra may be important for quantitative analysis. Thus, the SERS responses observed in this study shows a non-linear relationship and this may be due to saturation of the nanoparticles at higher concentrations of the dye.

This study went on to demonstrate that the detection and quantification of Sudan-1 using a portable Raman spectrometer with SERS was success. In addition, this study also highlighted the advantages of using multivariate approach over traditional linear regression method for calibration. Future work will include the optimisation of the extraction procedures from complex food matrices prior to detection to reduce background interference and hence to improve sensitivity.

The gold nanoparticles used in this study were recognised to have more homogeneous morphologies when compared with silver, and this should be better for reproducible SER(R)S. The LOD of Sudan-1 using the Ahura defender with SERS was calculated to be 480 ppb, this was an order of magnitude higher than HPLC and lab-based Raman systems using SERRS/SERS (Shadi 2001; Shadi 2003). However, this is more of a technical issue regarding the portable system design rather than low sensitivity, and is likely to be due to insufficient laser power. Increased sensitivity could be achieved by utilising longer integrated times and increasing laser intensity; however, due to the limited dynamic range of the detector, saturation at high concentrations was often a problem, therefore there was a compromise between the system's upper and lower dynamic range, the overall laser excitation intensity and sample concentrations suitable for quantification. Improvements in the system's software interfaces such as greater flexibility with the acquisition parameters and laser attenuation controls would have further enhanced sensitivity. Incorporation of SERS databases and online data pre-processing step (baseline correction) to aid in the identification and comparison of unknown chemicals would also extend the range and application of the system.

Despite the above, it was demonstrated that a portable Raman Ahura system combined with SERS is capable of onsite detection of Sudan-1 in the ppm to ppb range with little sample preparation required; moreover, the portable system is robust, rapid and straight forward to use.

The optimal pH for providing the most intense Raman signal in both SERRS/SERS studies was determined to be approximately pH 2.0; the reason for this is unclear as the calculated pKa and pKb are very different (Figure 16) but several explanations can be offered to account for this:

(1) Poly-L-lysine was used in both studies. This polymer induces aggregation of the nanoparticles whilst bringing the analyte into close proximately to the metal through electrostatic attraction. The enhancement factor observed with poly-L-lysine are likely to be the consequence of increase aggregation, providing a larger surface area for which the analyte can bind to and promote increased localised hot spot formations for signal enhancement.

(2) Within the acidic environment of the SERRS/SERS solution, the amine and hydroxyl groups of the Sudan-1 and the secondary and tertiary amines

group of the methyl green will become protonated. The positively charged amine groups present in both dye molecules will bind to soft metals such as silver or gold due to the charge attraction effects when sufficiently close to each other in solution.

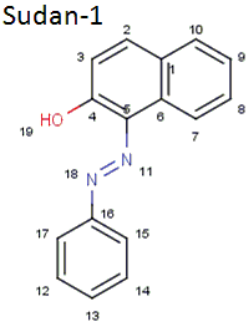
(3) The negatively charged citrates surrounding the nanoparticles will become more readily displaced within a proton rich environment. When analytes are in close proximately to the nanoparticles (brought about by the aggregation process), the citrates will be easily displaced therefore facilitating formation of analyte-metal complex. Similar observations have been reported in previous studies of dyes employing SERS/SERRS where a pH dependence of those analyte was also demonstrated (Shadi 2001; Shadi 2003; Shadi 2004).

SERS and resonance SERS are rapid and sensitive techniques for the analysis for aqueous dyes in very dilute concentrations with minimal sample pre-processing. The Raman spectra provide a unique fingerprint of the analyte which gives specific molecular information. The use of multivariate data analysis for exploratory analyses and quantification offer more robust results to be obtained compared to conventional univariate approaches. This is because the entire spectrum is utilised rather than only focusing on a selective vibrational band.

SERRS is the combined effect of both resonance and surface enhancements. Whilst it may offer higher signal amplification compared to conventional SERS, it is not as widely applicable as the compatibility with the applied excitation frequency and the analyte's molecular adsorption maxima cannot always to be matched. However, SERRS does not require complete overlap of the applied excitation wavelength and adsorption maxima of the analyte, and some SERRS enhancements can be observed even if there is only a partial overlap. In addition, SERRS has significant advantages for the analysis of chromophores which are often difficult to characterise by RS due to their high fluorescence background. The use of silver or gold nanoparticles in combination with SERRS shows that fluorescence can be efficiently quench by the nanoparticles (Graham 2006).

Atom	pK_a	pK_b
C1		
C2	43.0	
C3	43.0	
C4		
C5		
C6		
C7	43.0	
C8	43.0	
C9	43.0	
C10	43.0	
N11		-1.5
C12	43.0	
C13	43.0	
C14	43.0	
C15	43.0	
C16		
C17	43.0	
N18		-0.1
O19	11.8	-6.4

Sudan-1



Atom	pK_a	pK_b
C1		
C2	43.0	
C3	43.0	
C4		
C5	43.0	
C6	43.0	
N7		
C8	51.0	
C9	51.0	
C10		
C11	43.0	
C12	43.0	
C13		
C14	43.0	
C15	43.0	
C16		
N17		
C18	51.0	
C19	51.0	
C20	51.0	
C21	51.0	
C22	43.0	
C23		
C24	43.0	
C25	43.0	
C26		
C27	43.0	
N28		4.5
C29	51.0	
C30	51.0	
H31		
H32		

Methyl green

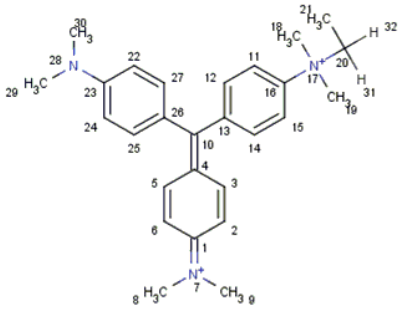


Figure 16 Chemical structure of Sudan-1 and Methyl green with it calculated pKa and pKb values.

The manner in which the metal particles are synthesised and their resultant composition has a substantial effect on the subsequent enhancement observed (Jain 2006), as variations in particle sizes and structural morphology from batch to batch will have a significant impact upon the reproducibility of the data. Therefore it is crucial to have a reproducible and robust method of synthesising nanoparticles. Experimentation in Raman spectroscopy is often required to be conducted using the same batch of nanoparticles and if practical within the same day to minimises any technical and temporal variations. However in large scale experiments, involving high number of samples, this may not always be possible. The experimental parameters in which aggregations was induced prior to signal acquisition, aggregation rate, pH environment and solution composition will all have a significant contributions towards the reproducibility of the Raman spectra (Lee 1982).

Much effort has been focused on the development of a suitable substrate surface for signal enhancements in SERRS/SERS studies, however the substrate surface which

offers higher enhancement often suffer from reproducibility issues as the formation of localised surface hot spots is not always homogeneous (Ruan 2007,Gopinath 2009). Whereas substrates which offer moderate enhancements generally have higher reproducibility as the conditions can be more easily replicated. The most important metal substrates used in SERRS/SERS studies at present are generally gold and silver based. Enhancements using silver nanoparticles trend to be greater, generating 10 to 10^3 fold higher signal compared to gold nanostructures (Garica de Abajo 2007), therefore it is typically used for most small molecule characterisation studies while gold is more often utilised in bio-analysis and structural studies (Kneipp 2006, Sha 2008) due to its greater biocompatibility. The excitation frequency is different for the two types of nanoparticles; silver can be excited by a broader frequency compared to gold (Zhao, 2008). Gold nanoparticles exhibit a surface plasmon frequency at ~ 500-550 nm (Jain 2006) while silver exhibits a lower surface plasmon frequency at ~ 400-415nm (Guzmán 2009), upon aggregation the plasmonic shift in the gold nanoparticles display new absorption bands of higher frequency in the region of 530 nm while for silver display this shifts to lower frequencies in the region of 390 nm. The nanoparticle suspensions offer a better alternative to a metallic surface due to their increased surface area; since the nanoparticles are suspended in solution they also benefit from the effect of Brownian motion, where the SERRS/SERS spectra observed represents an averaged effect.

8.5 Conclusion

Metabolic VOC profiling is a novel field of analysis within metabolomics. The VOCs emitted from biological samples not only provide a variety of detail chemical information but also permits correlation to biological origin as well as physiological status. Metabolic VOC profiling is a non-invasive method and this analysis can be readily applied to real time profiling of a large range of samples and concentration types. This approach has as a variety of promising applications in microbiology, plant science, food analysis, agriculture, disease diagnostics and personalized healthcare.

The PDMS membranes employed in Chapters 3 and 4 are inexpensive, robust and highly versatile; hundreds of these can be cleaned and stored in a stable manner for

several days. Parallel sampling can also be readily achieved using PDMS. The surface composition of the PDMS and its observed trapping performance indicates that it is predominately suited for the collection of polar compounds (Sanchez-Palomo 2005; Riazanskaia 2008). Alternative surfaces such as Tenax-TA are available and these permit the capture of both polar and non polar VOCs; however, they are more labour intensive as they require the use of DHS methodology with specialised equipment.

VOC profiling is a challenging method of analysis, where greater emphasis must be placed on robust and reproducible HS sampling (Jia 2007). Once the VOCs are collected the trapped analytes are prone to degradation and contamination. Sufficient care must be taken to minimise this (Riazanskaia 2008) as the samples are only stable for a few days. Therefore all samples must be analysed as soon as possible (less than 48 h). This necessitates a robust sampling and introduction procedure to ensure good biological and analytical reproducibility. In HS sampling the use of thermal desorption has allowed a reasonable degree of high throughput capacity to be achieved through semi-automation, which helps towards ensuring reproducibility.

As discussed in Chapter 2 DMS is a promising technique for the rapid analyses of volatile materials. As Py-GC/DMS was studied, technically speaking the pyrolysate generated using Py-GC/DMS are not VOCs but thermally decomposed fragments generated under inert atmosphere within a highly controlled condition. However, DMS could potentially be the method of choice for VOC screening surpassing MS-based platforms; this is due to its ability to operate at ambient pressure with good sensitivity (Krebs 2005). Nevertheless, at present DMS is more suited for qualitative rather than quantitative analysis due to its comparative low resolution compare to MS and issues with peak identification.

Robust and coherent theoretical descriptions of the ion chemistry and the mode of operation for DMS are extensively available (Nazarov 2006; Krylov 2007; Krylov 2009), and experimental data are highly consistence with predictive models (Krylov 2010). The technique is still in development but is fast maturing, the small size of the DMS detector (the current generation are less than $1cm^2$) and its low power consumption permits portability and holds great potential for deployment as a field

sensor for real time and on site monitoring (Griffin 2006). Recently DMS has been successfully incorporated into MS-based platforms functioning as an orthogonal method of chromatographic separation in a similar manner to GC or LC with promising results (Levin 2007; Coy 2010; Manard 2010). DMS originally evolved from ion mobility spectrometry (IMS) and these two techniques are closely linked, IMS has also been successfully integrated with modern MS systems (Wu 1999) and offers a real alternative to GC (Dwivedi 2010).

Currently mass spectrometry is one of the most developed fields of analysis in metabolomics and is highly powerful tool generating detailed structural and chemical information. The detector employed in Chapters 3-5 was relatively obsolete by comparison to modern MS detectors, it had a significantly lower scan rate, capacity and linear dynamic range, as such low intensity ions were not likely to be detected. Whilst more sophisticated mass analyzers would allow considerably higher resolution mass spectra to be obtained, MS based analytical methods are not without drawbacks, MS has a lower throughput capacity and requires significant pre-processing prior to data analysis. This point is clearly highlighted in Chapters 3-6 which lead to the development of appropriate data preprocessing methodologies.

Raman spectroscopy is an alternative analytical approach which is rapid and non-destructive, and able to generate detailed spectra of the sample in various matrices. Raman spectroscopy is useful when target analytes are in very low concentrations and when quantification is necessary. The weak Raman signal can be significantly amplified by resonance and surface enhancement effects, as demonstrated by the work presented in Chapters 6 and 7. SERRS and SERS are ideal for analysing aqueous solutions in a rapid and quantitative manner. The recent development of tip enhanced Raman scattering (TERS) (Pettinger 2007; Chaigneau 2010) in which an ultra fine tip composed of a SERS active metal is brought within close proximity to the sample (within the nanometer range), results in signal enhancement factors of up to 5 $x10^9$ compared to conventional Raman scattering (Neacsu 2006). This novel variation of Raman spectroscopy holds great potential for generating highly detailed spatial SERS measurements and could also be used for the characterization of peptides, DNA and RNA (Domke 2007; Bailo and Beckert 2008).

The current analytical platforms employed in metabolomics generate an immense quantity of complex chemical data and conventional methods like simple visual inspection are no longer feasible or even practical. The use of multivariate analyses in combination with univariate statistics is becoming a standard practice in metabolomics to derive valid analyses and hence robust results. The intensive use of chemometrics at all levels presented in Chapters 2-7 has allowed rapid and efficient processing of immense quantities of complex data in a relatively short period of time. The use of unsupervised cluster analysis allows the visualisation and identification of potentially significant trends within data sets which are consistent with the experimental design and biology. These may be subsequently validated through supervised analysis methodologies.

The application of suitable chemometrics to different types of data structures, is essential in order to obtain valid, reliable and objective results. By taking into consideration the advantages and limitations of the data analysis techniques, appropriate improvement/adjustment in the experimental design can be implemented.

Chemometrics is however not an end in itself, as no amount of detailed understanding, experience and smart application of it can address an ill-conceived experimental design with poorly collected data (e.g., inadequate sampling procedure, atypical responses, lack of metadata and incorrect class information), with inadequate biological and analytical replications. In most cases the number of samples per class is lower than one would want and this often results in a large biological variation which cannot be accounted for or constrained; such variation can and does have significant effects upon the quality of data and the statistical relevance of the results obtained. Therefore robust and suitable experimental design, with diligent sampling and data collection in conjugation with suitable data analysis methodology, is crucial to yielding valid and interesting results.

The analysis of VOCs is a complementary profiling tool within the metabolomics toolbox, and one of its great attractions is that it permits the characterisation of biological samples in a rapid and non-invasive manner. The technique provides detailed chemical information regarding the VOC composition present above the headspace of the sample and these molecules can be used to interrogate the

biological system under analysis. In conclusion, the work presented in this thesis has demonstrated the utility of VOC profiling and like all metabolomics analyses its value is only realised when chemometrics processing is applied. I believe that VOC analysis is a valuable additional approach for non-invasive analysis of many biological systems.

Appendix

Supplementary information for Chapter 3

Supplementary Material

Table A Summary of instrumentation parameters

Thermal desorber	
Primary desorption flow	$50cm^3 min^{-1}$
Primary split	Splitless
Primary desorption temperature	180˚C
Primary desorption time	5 min
Cold trap volume	19 µl
Cold trap temperature	-10˚ C
Cold trap packing	Tenax TA/Carbograph 1TD
Secondary desorption flow	$1cm^3 min^{-1}$
Secondary split	50:1
Trap heating rate	$100˚C\ min^{-1}$
Secondary desorption temperature	300˚C
Secondary desorption time	3min
Gas Chromatograph	
Column flow	$1cm^3\ min^{-1}$
Initial column temperature	70 °C
Time that initial column temperature was held	1min
Column heating rate	$3°C\ min^{-1}$
Final Column temperature	230°C
Time that final column temperature was held	10 min
Ion Trap Mass Spectrometer	
Ionisation EI mode	10µA
Scan range	30 to 400m/z
Scan time	1s
Trap temperature	200°C
Manifold Temperature	90°C
Transfer line temperature	260°C
Filament and Multiplier delay	30s

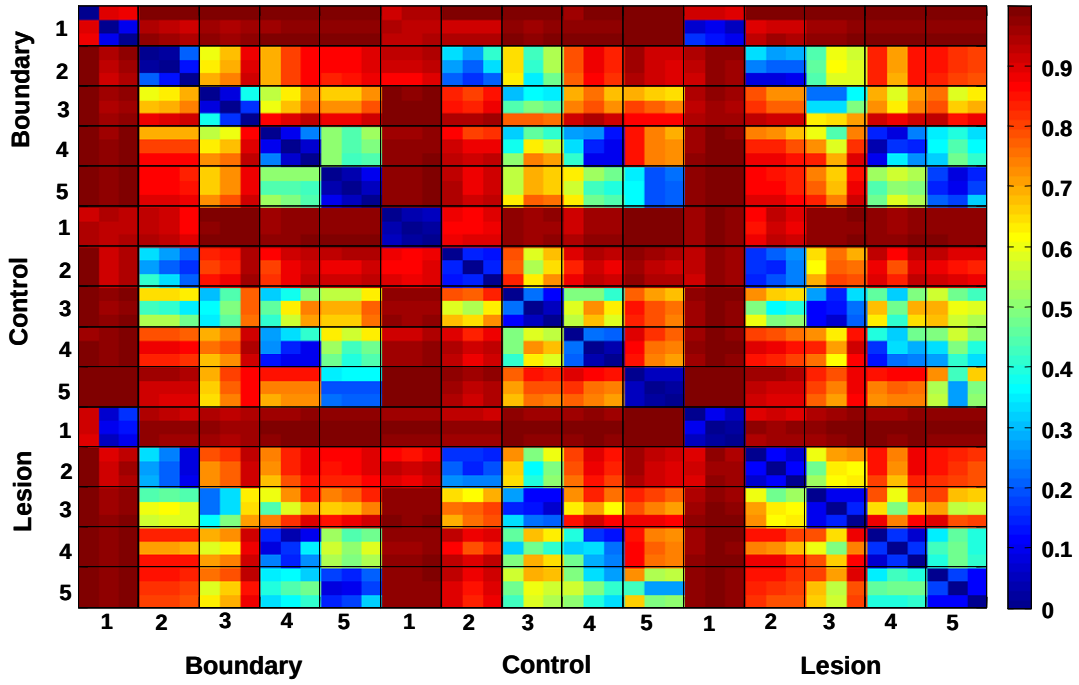


Figure 6. **Distance heat map plot. The distance between each pair of samples are represented by a colour as indicated by the colour bar on right (red to blue). The higher the distance, the lower (bluer) the similarity between two samples. The 5 different subjects, each subject has 3 samples from each class respectively, were labeled by numbers from 1-5 and samples from the sample class (e.g. Boundary, Control or Lesion) were placed together. Since each subject has 3 samples for each class, the 3x3 blocks in the diagonal of the picture represent the similarities between the samples from the same subject and also the same class which thereby show the reproducibility of the sampling methodology.**

Supplementary information for Chapter 4

Analytical and bioanalytical Chemistry

Electronic Supplementary Material

VOC-based metabolic profiling for food spoilage detection with the application to detecting Salmonella typhimurium contaminated pork (Supplementary information)

Yun Xu, William Cheung, Catherine L Winder, and Royston Goodacre

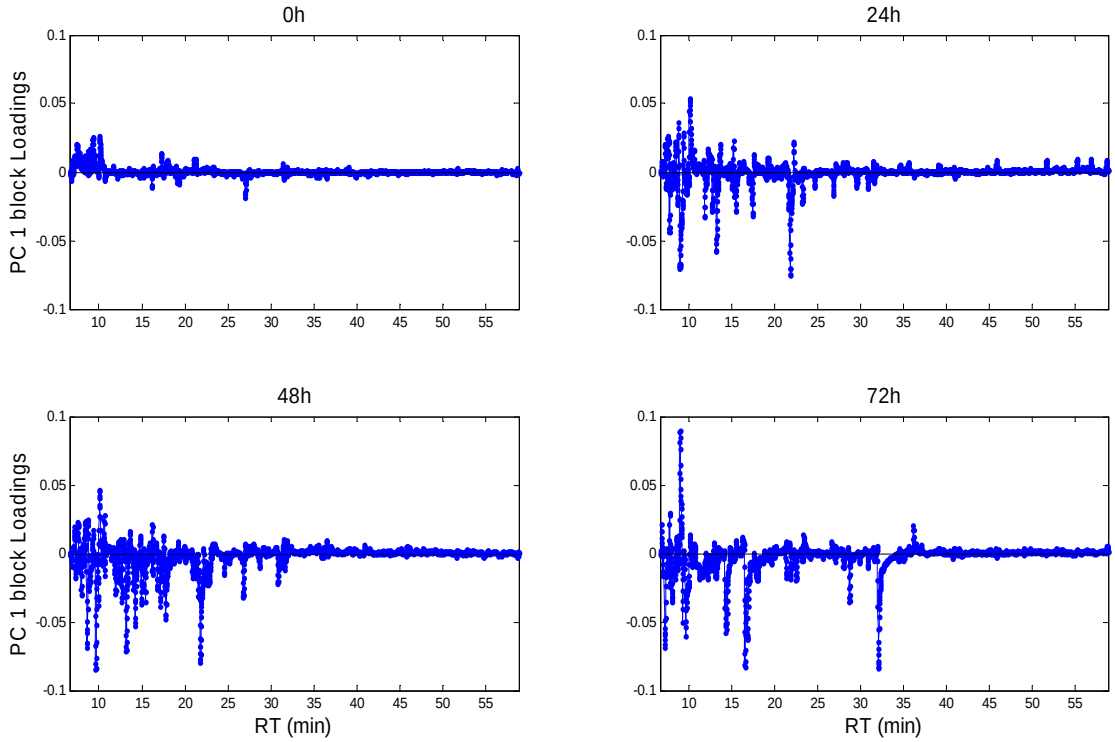


Fig. S-1: Block loadings plot of the first CPCA model.

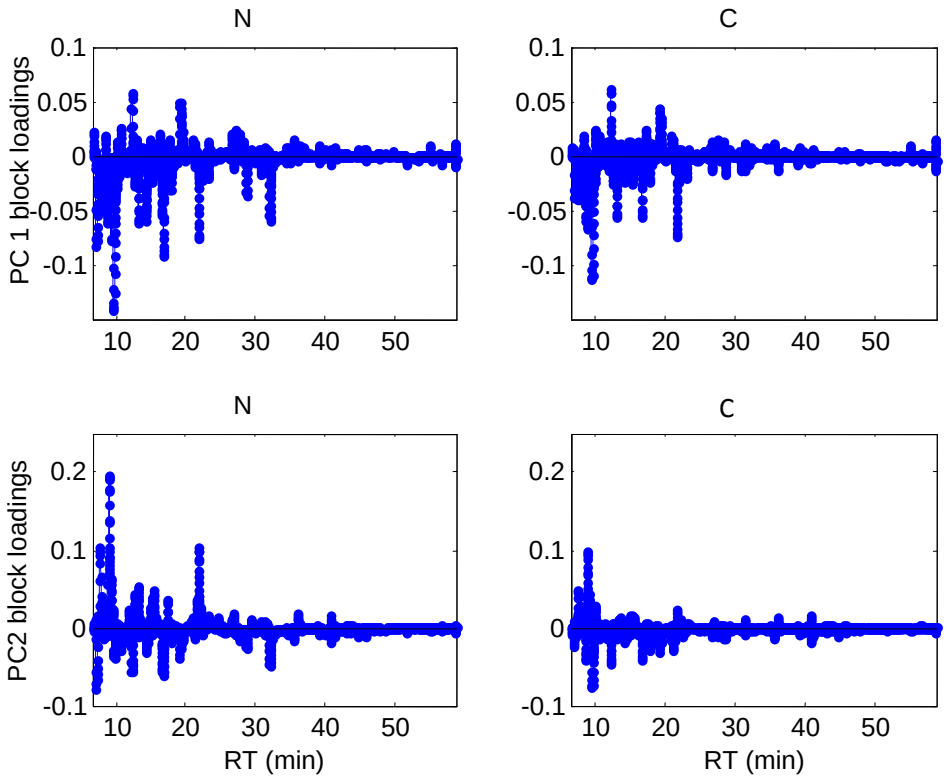
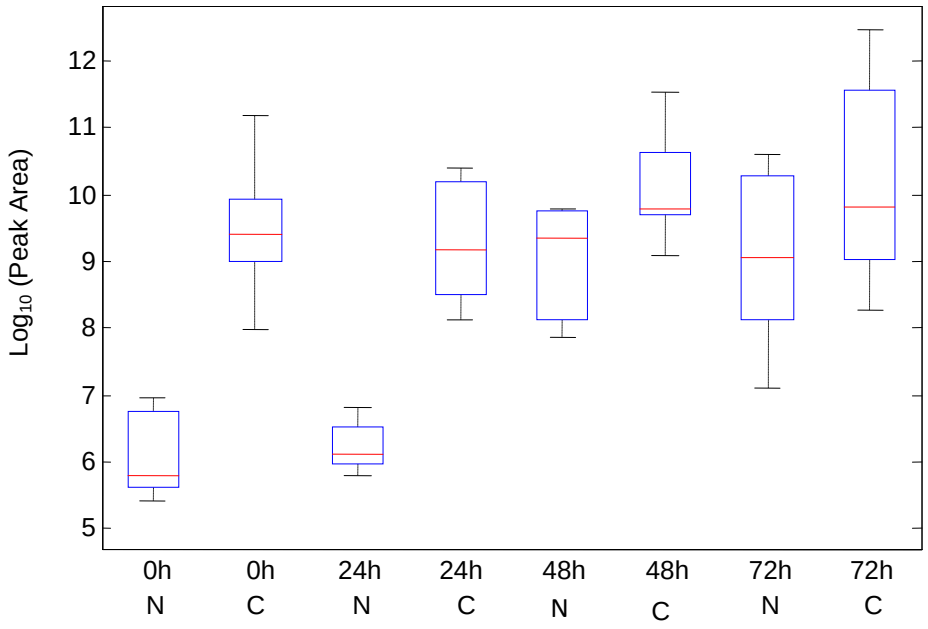
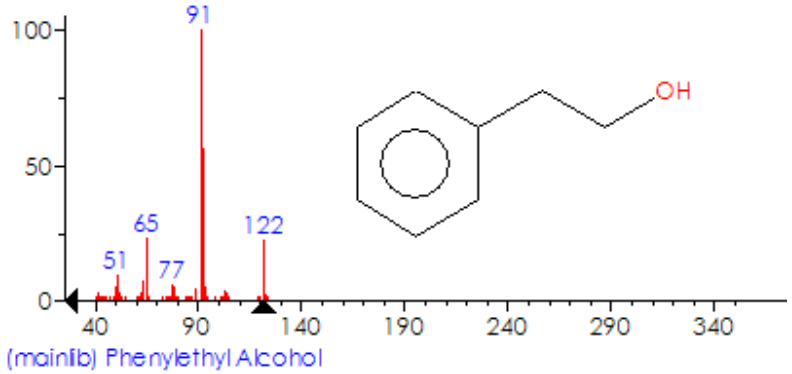


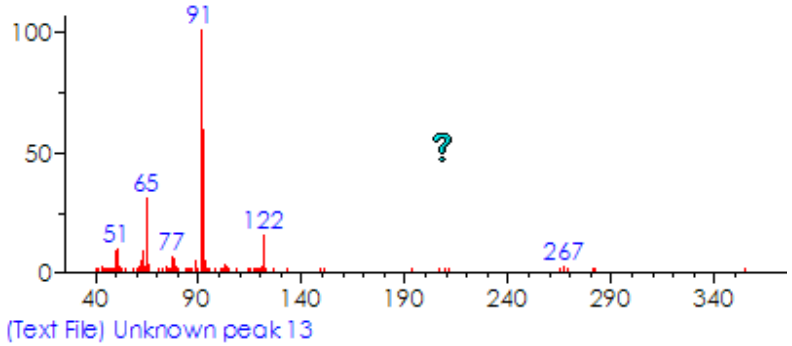
Fig. S-2: Block loadings plot of the second CPCA model.



(a)

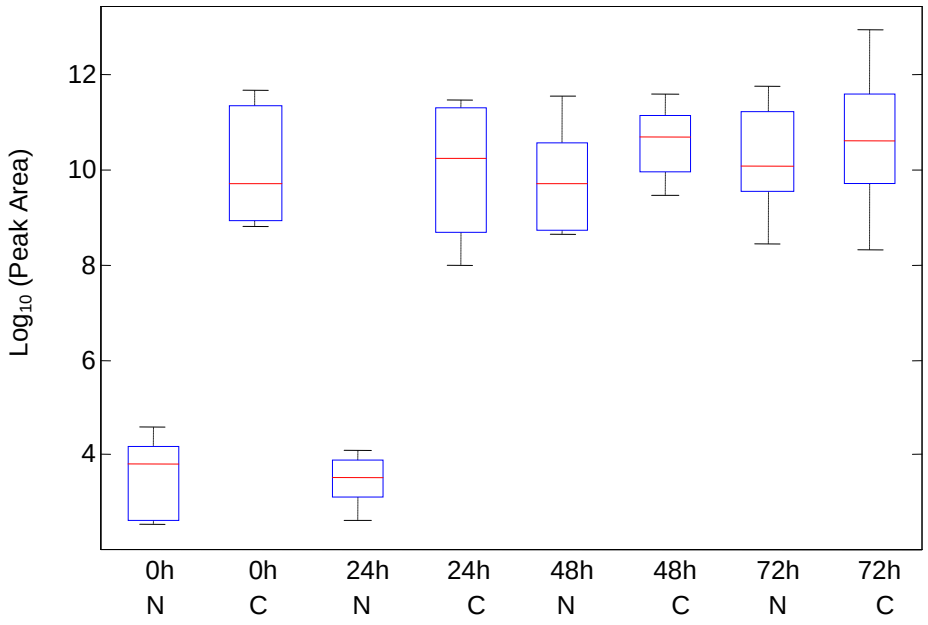


(b)

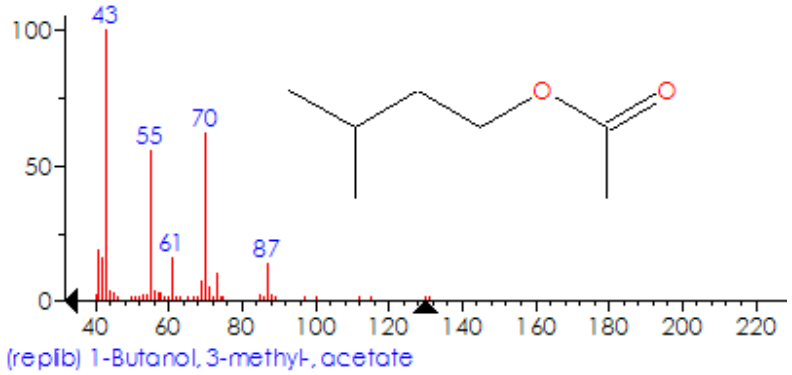


(c)

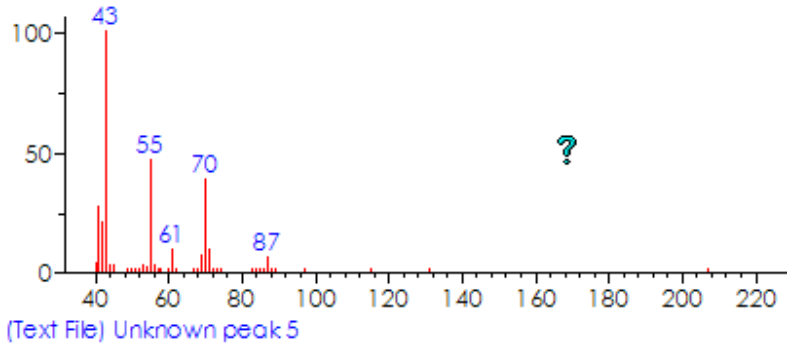
Fig. S-3: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



(a)

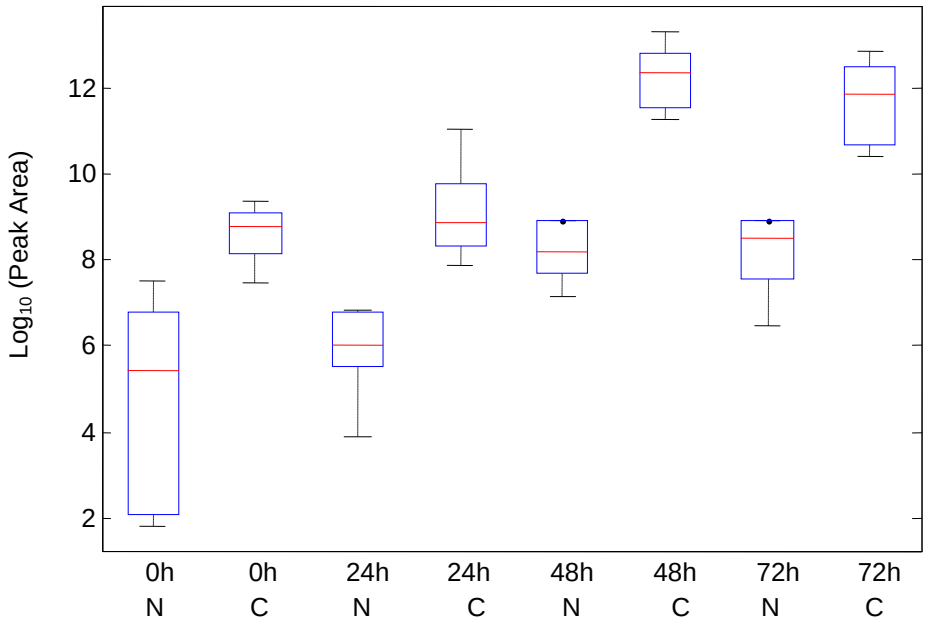


(b)

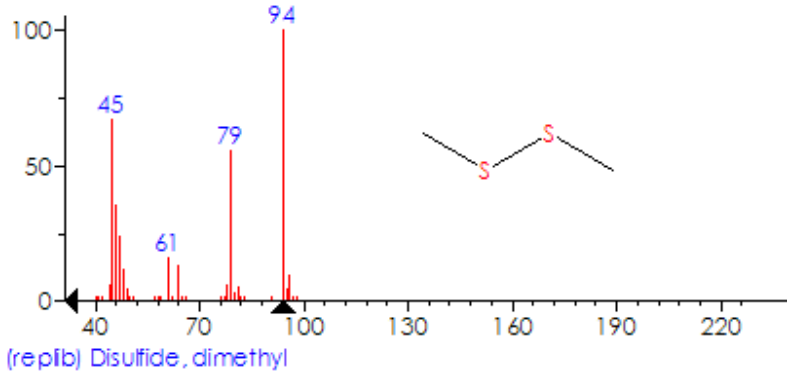


(c)

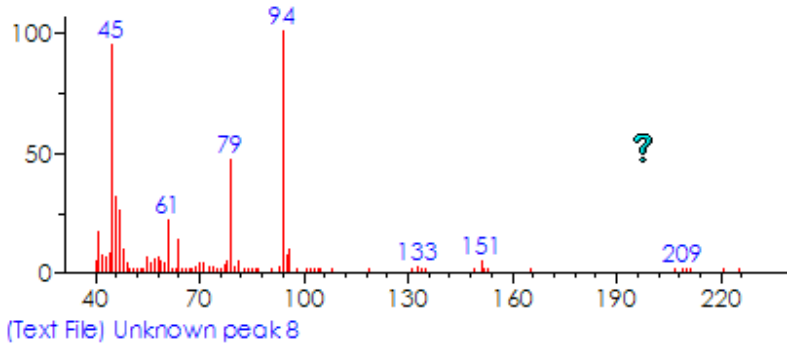
Fig. S-4: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



(a)

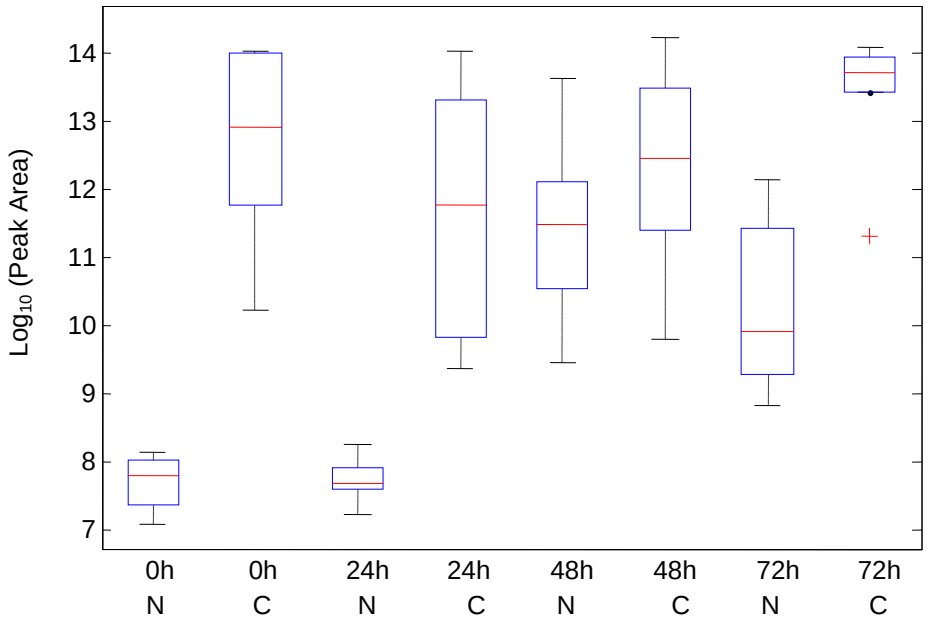


(b)

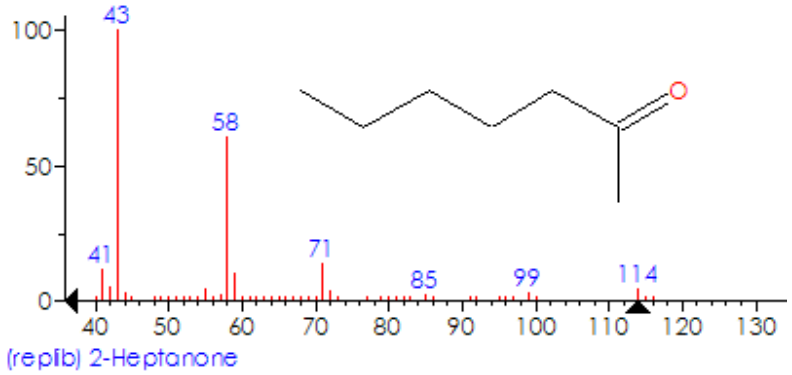


(c)

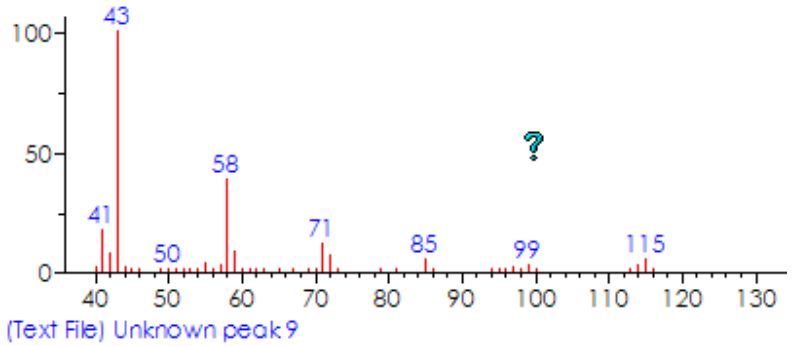
Fig. S-5: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



(a)



(b)



(c)

Fig. S-6: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

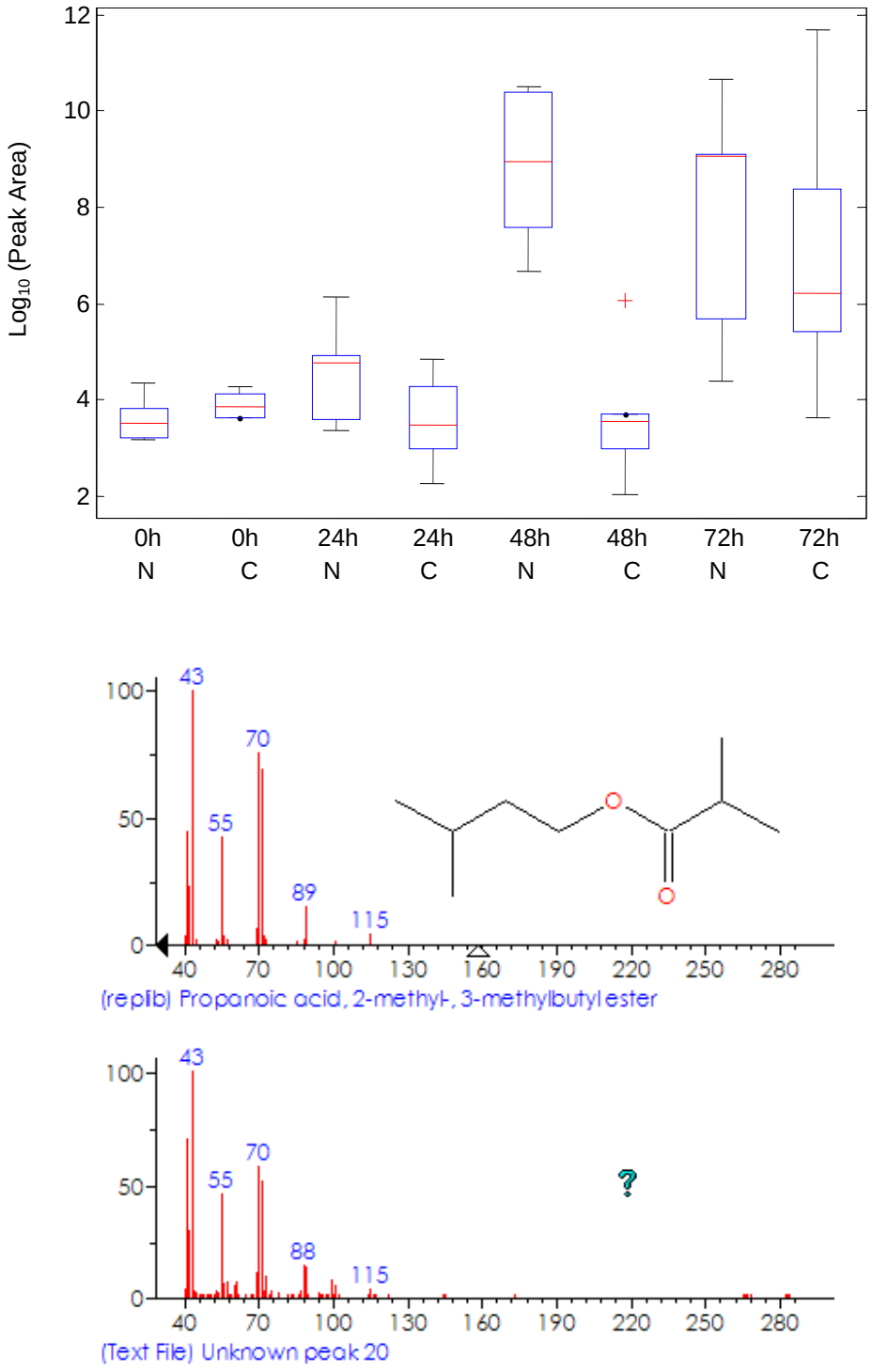
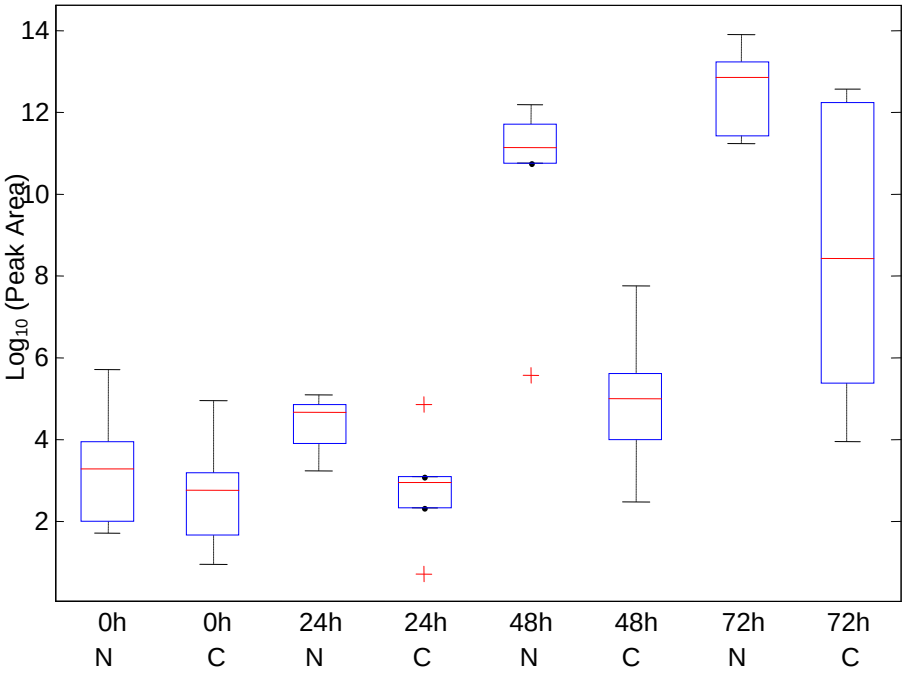
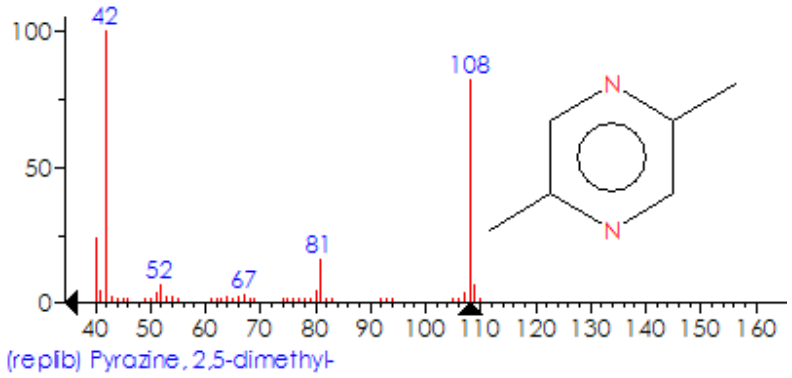


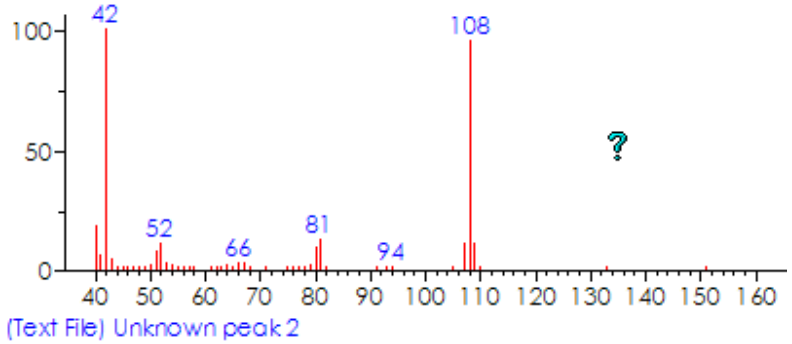
Fig. S-7: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



(a)



(b)



(c)

Fig. S-8: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

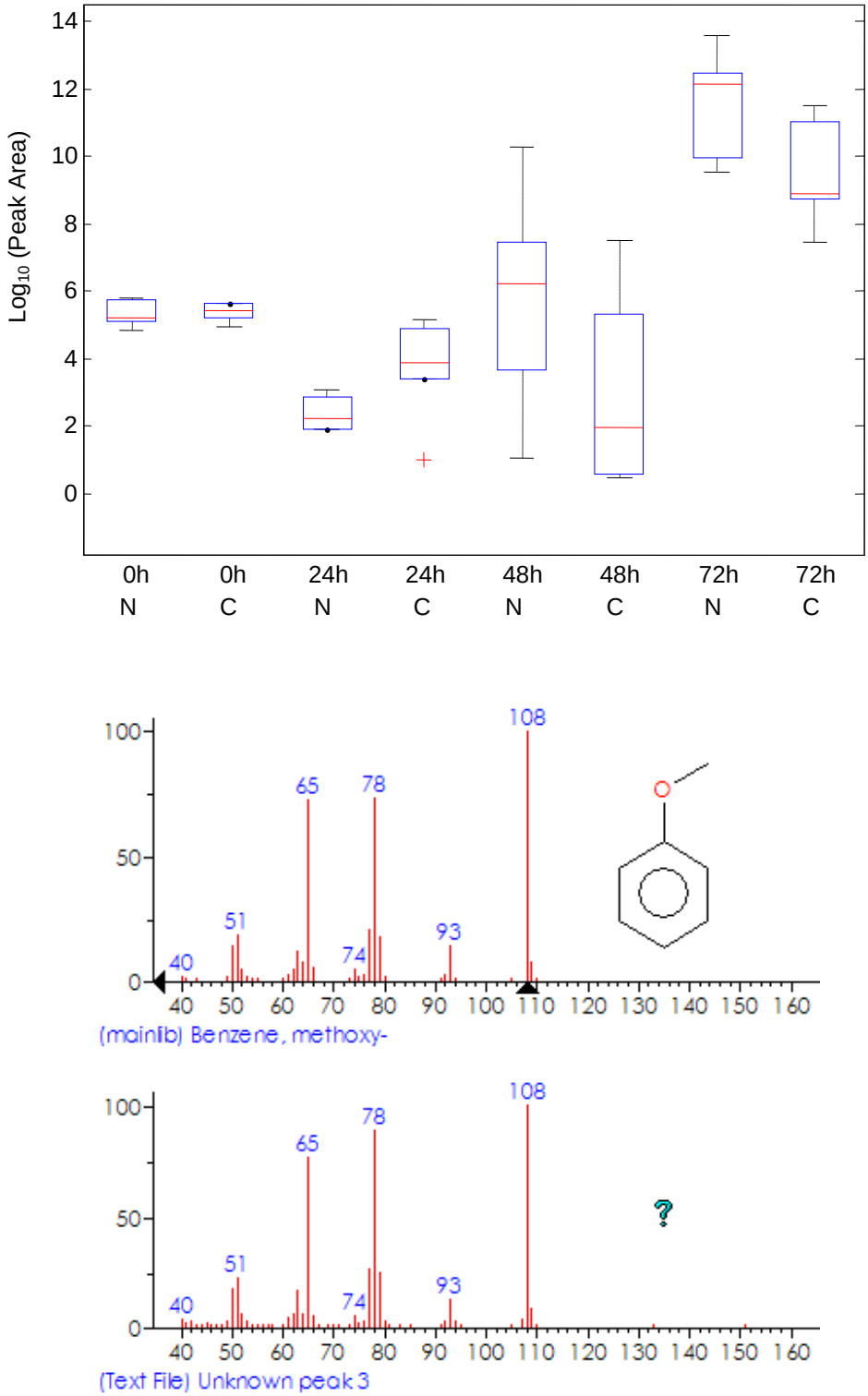


Fig. S-9: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

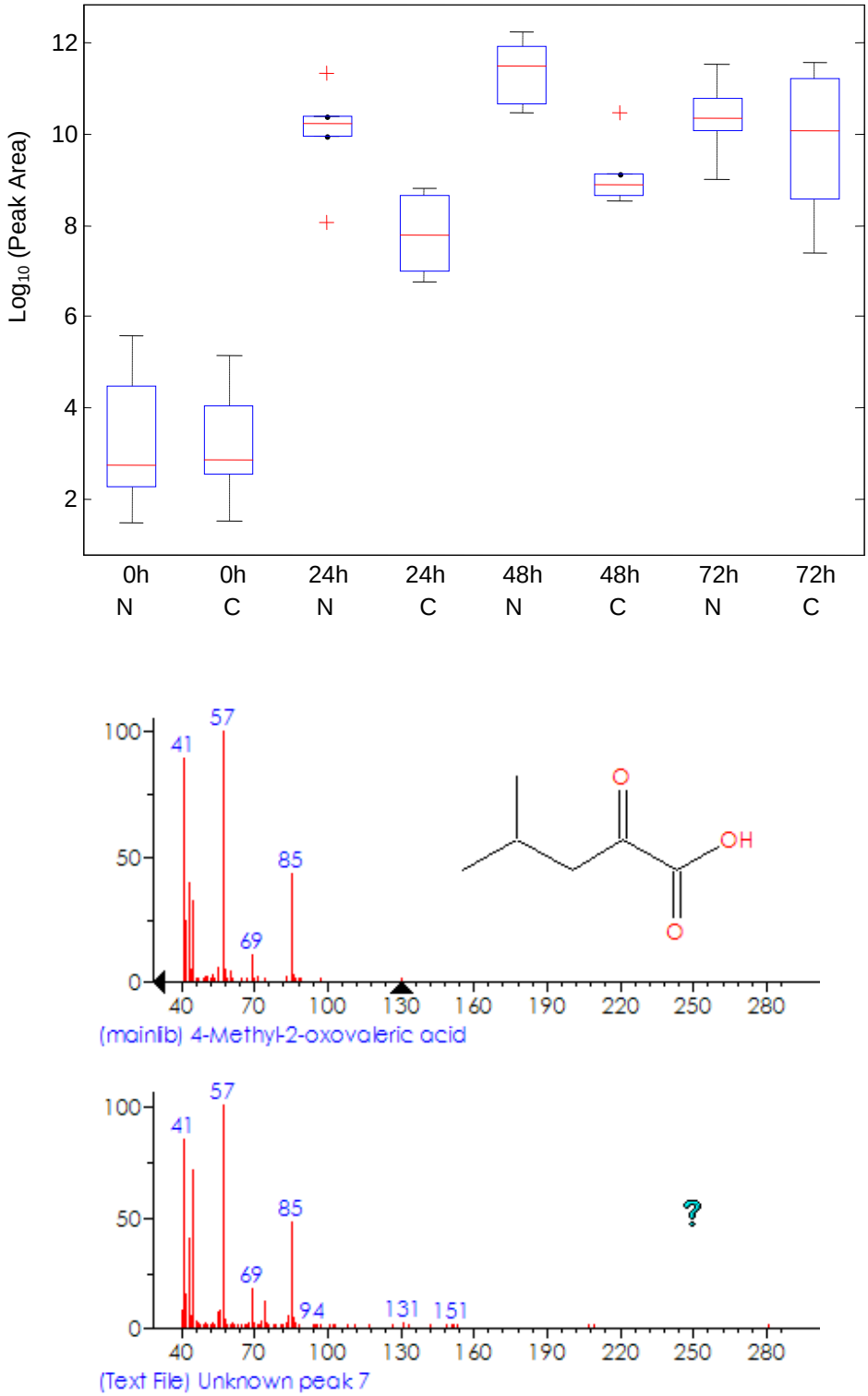


Fig. S-10: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

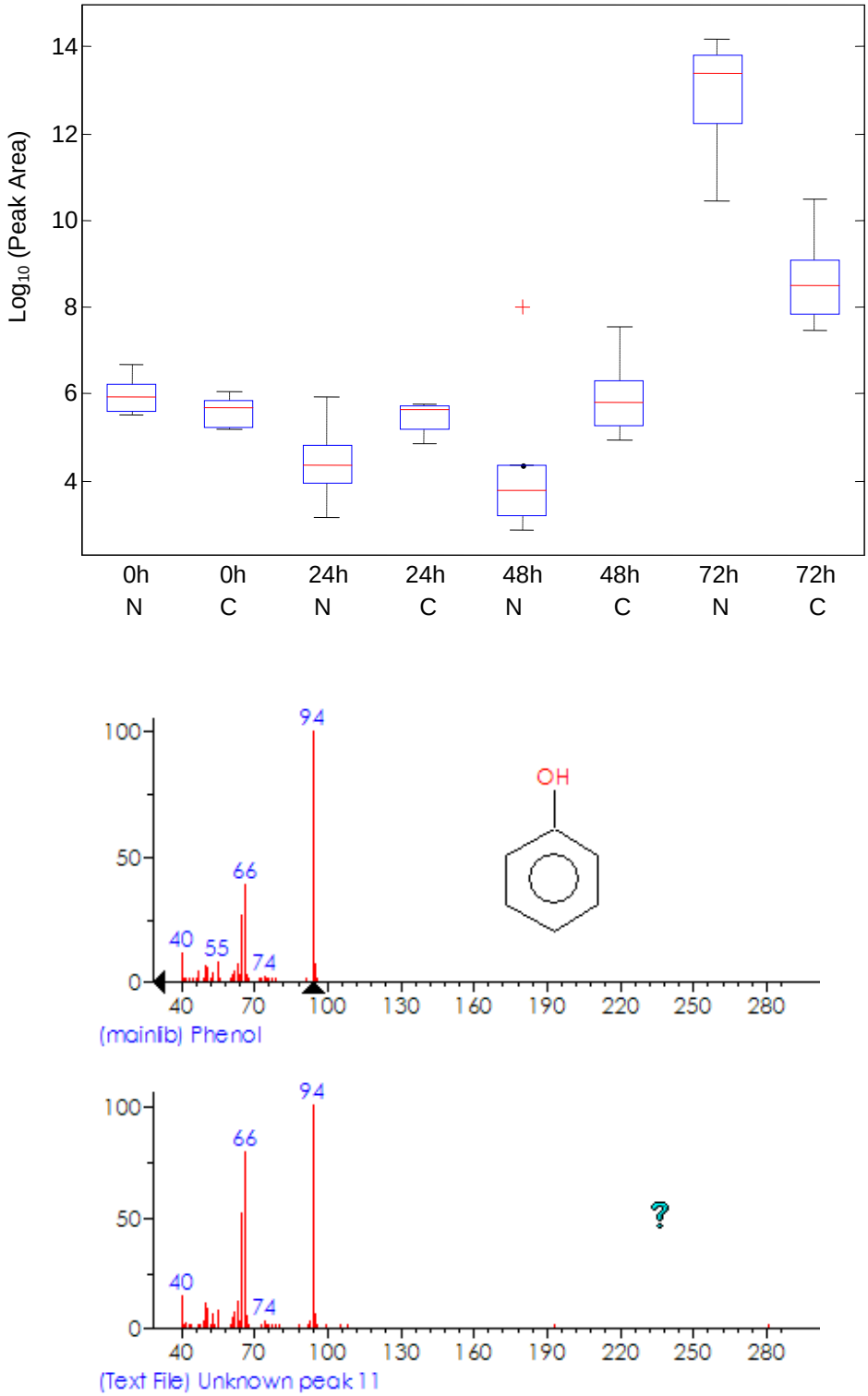


Fig. S-11: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

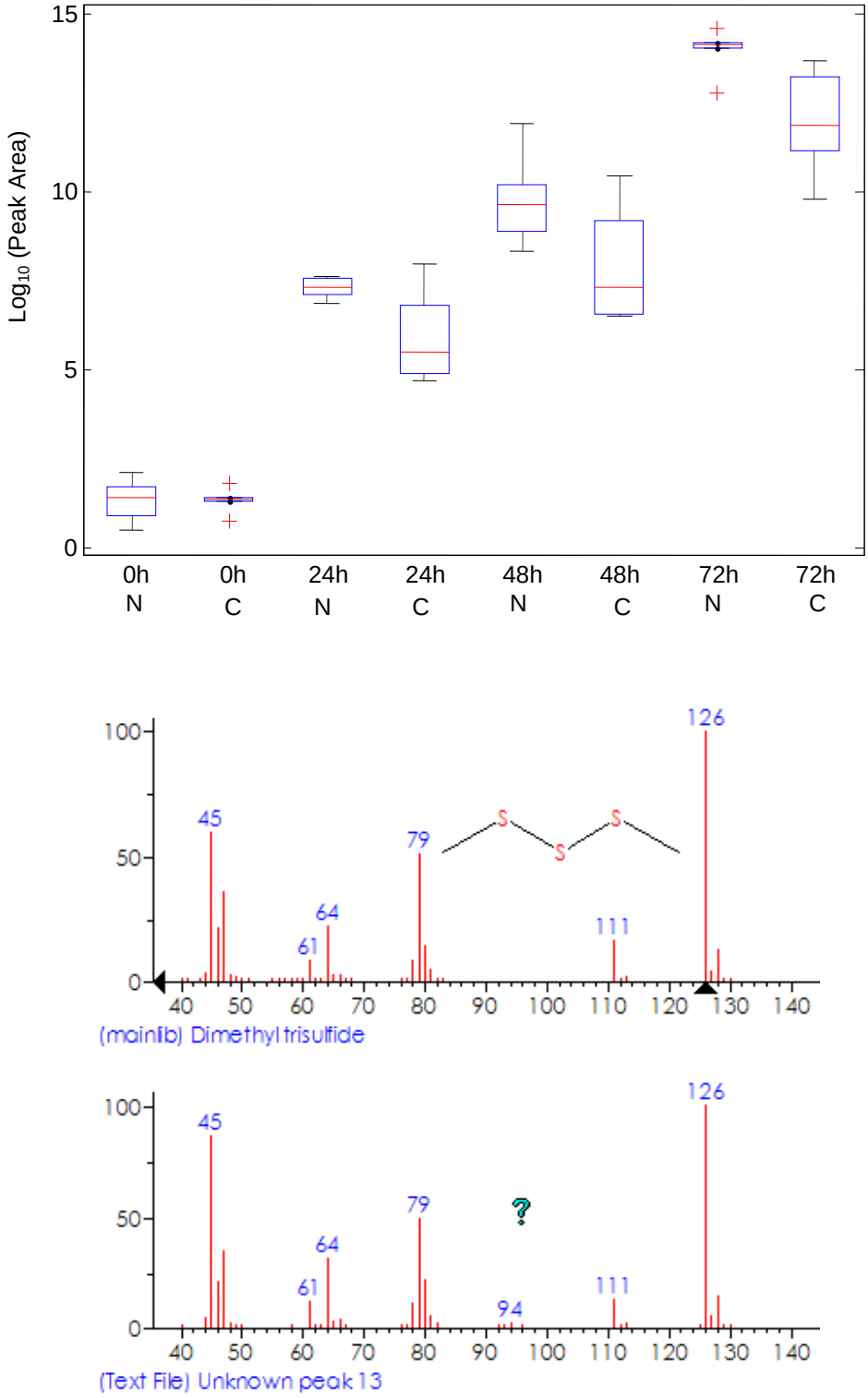


Fig. S-12: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

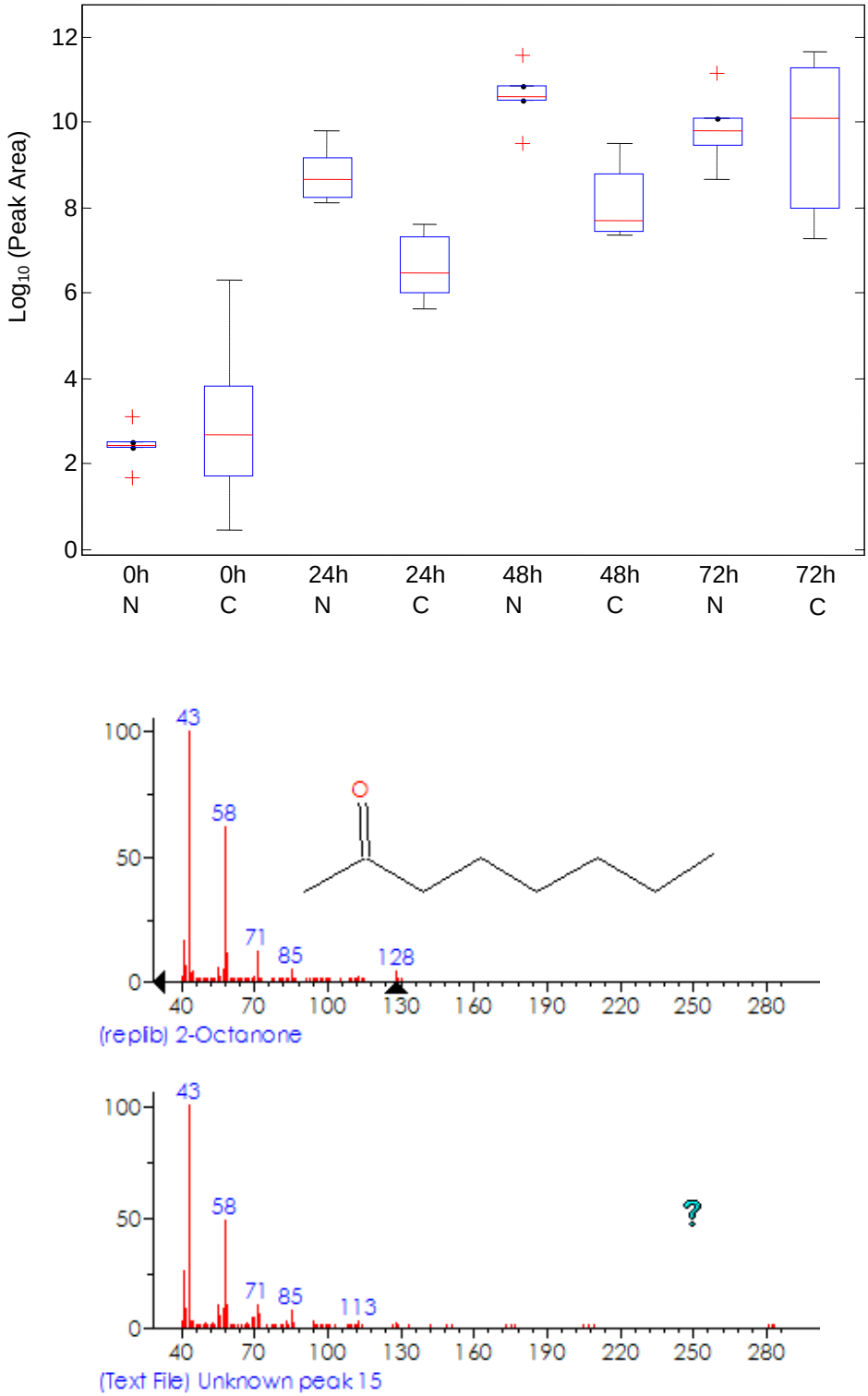


Fig. S-13: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

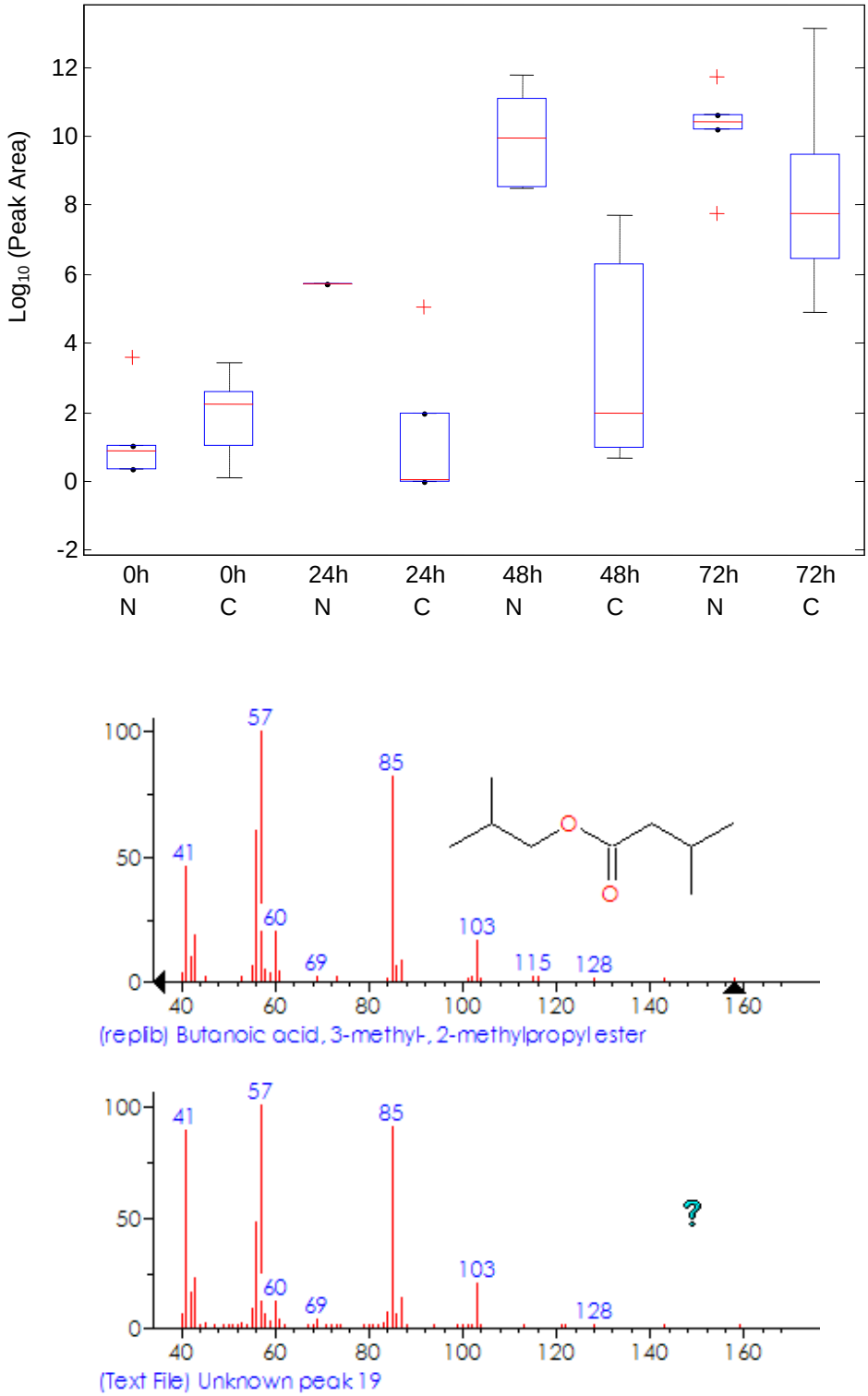


Fig. S-14: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.

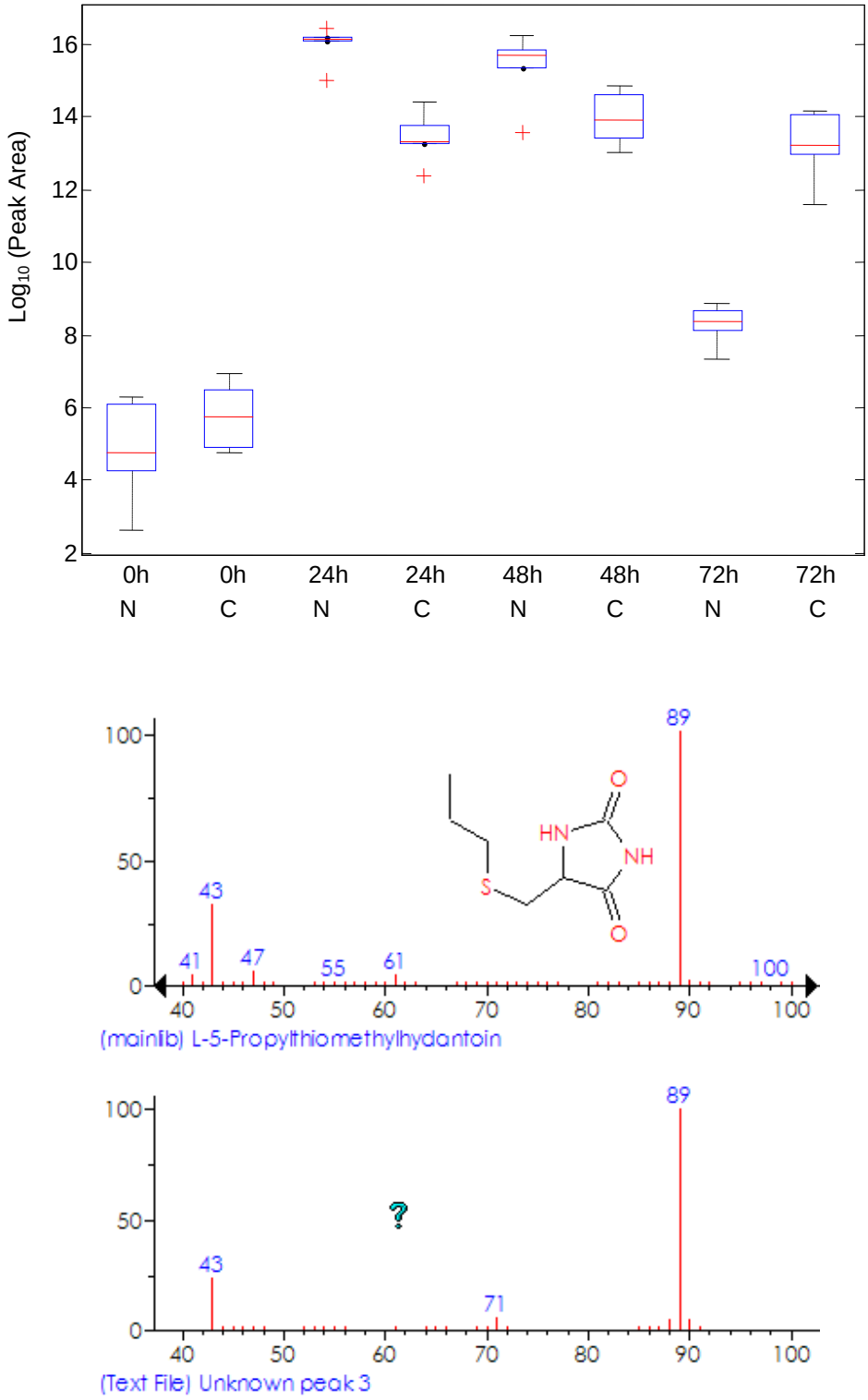
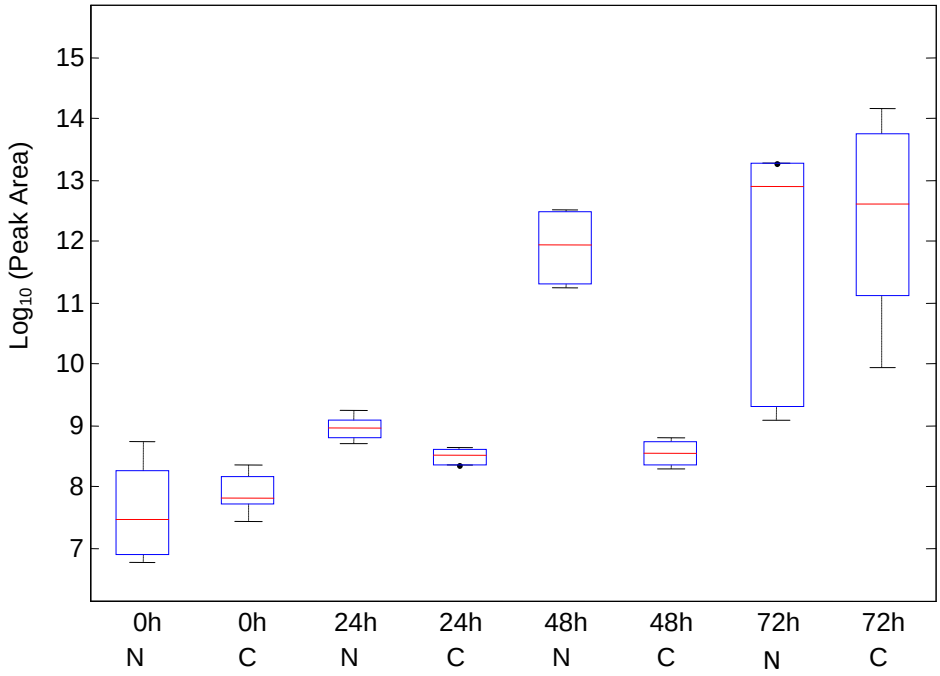
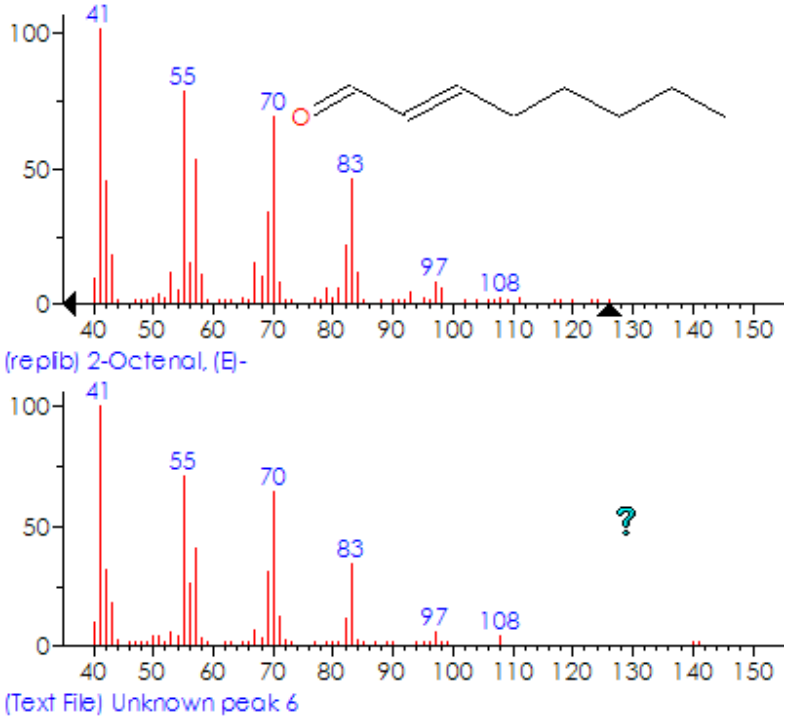


Fig. S-15: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



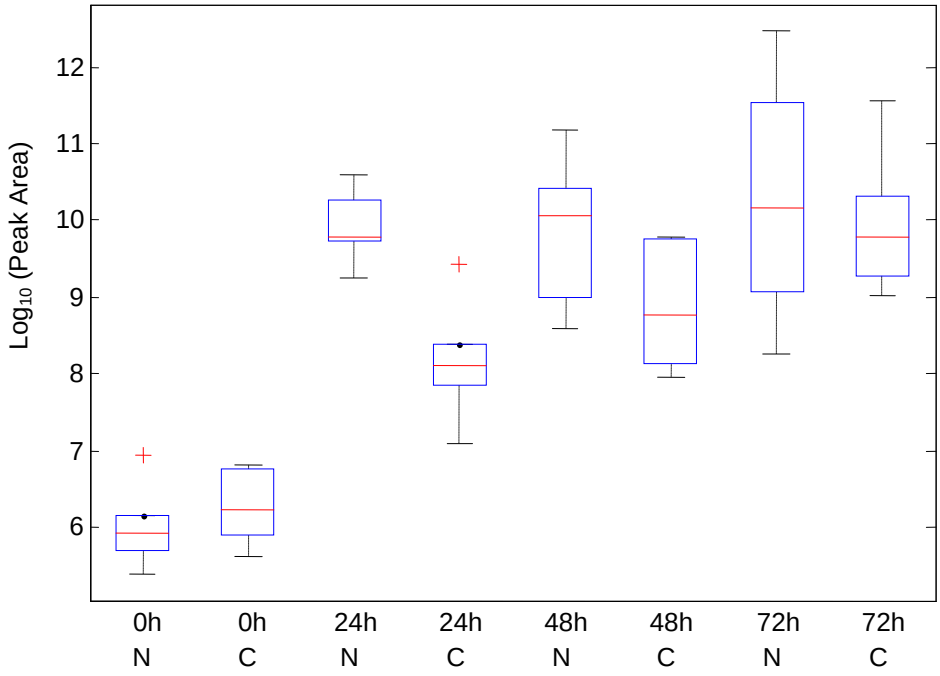
(a)



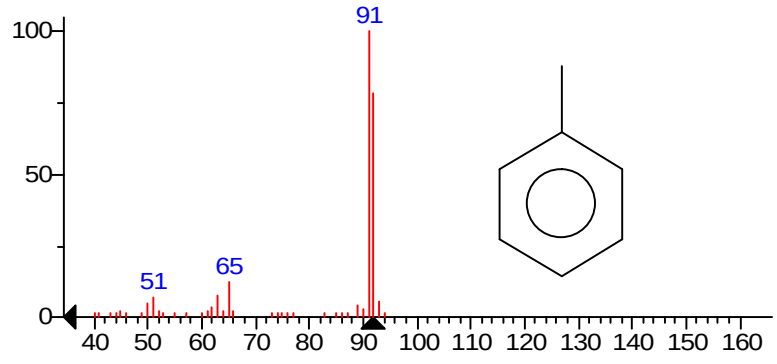
(b)

(c)

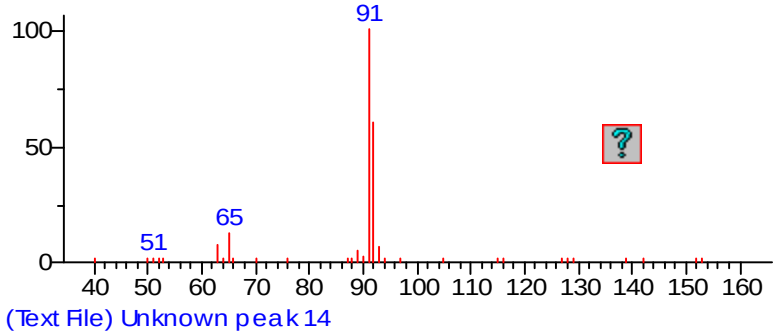
Fig. S-16: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



(a)

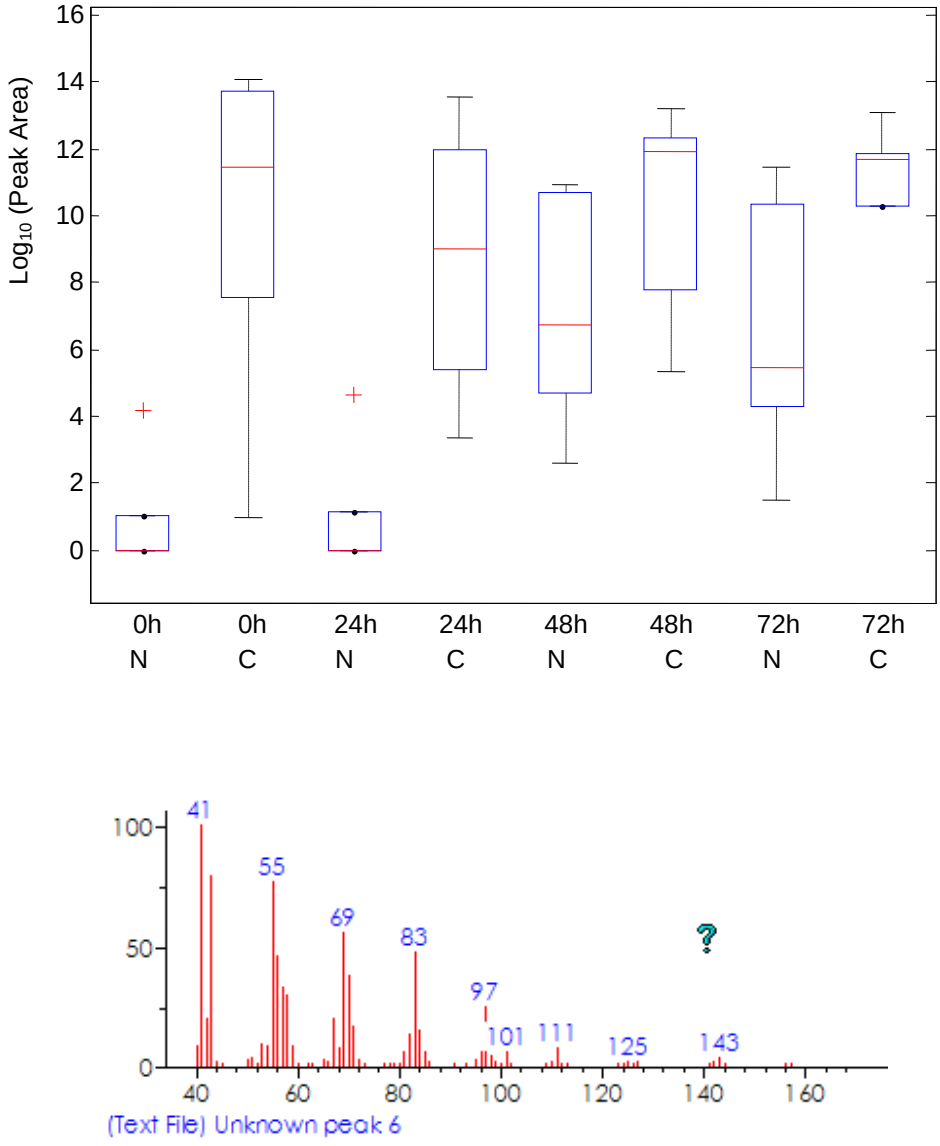


(b)



(c)

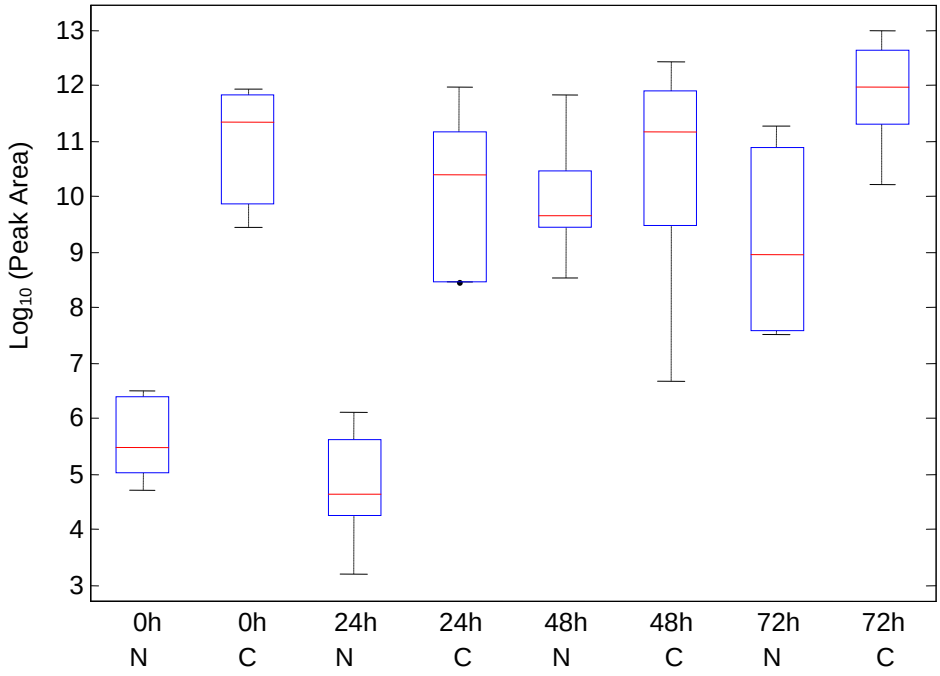
Fig. S-17: (a) Box-whisker plot; (b) Standard mass spectrum; (c) Extracted mass spectrum of the target peak.



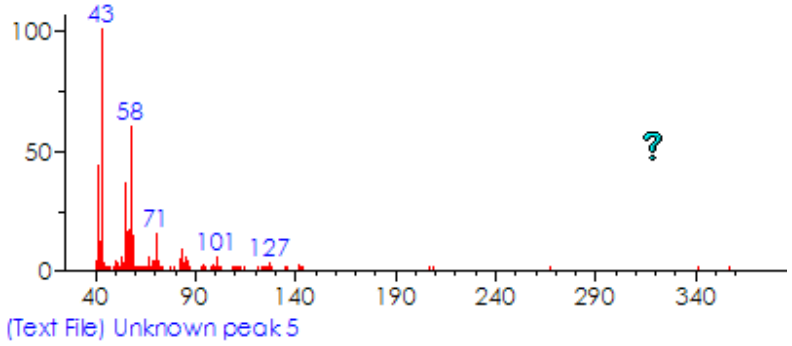
(a)

(b)

Fig. S-18: (a) Box-whisker plot; (b) Extracted mass spectrum of the target peak. No satisfactory matching can be found.

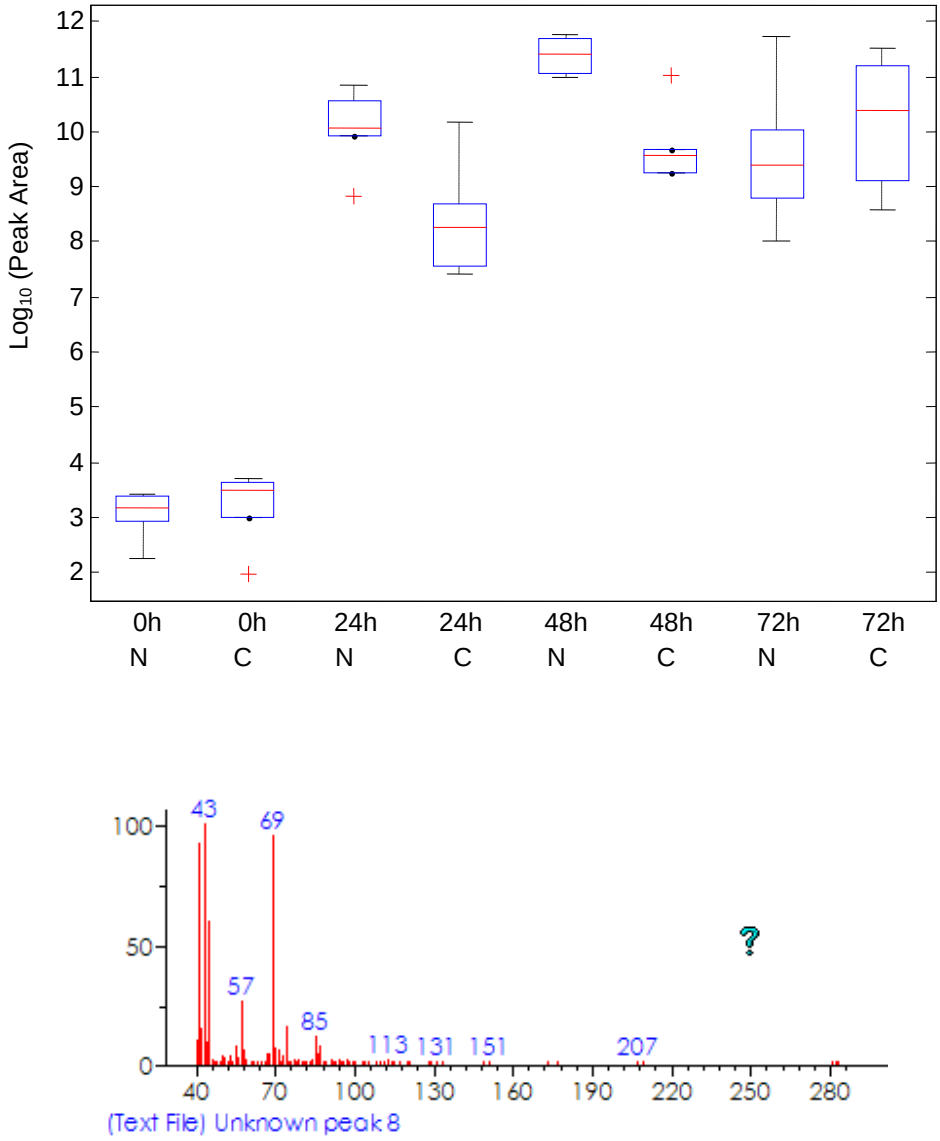


(a)



(b)

Fig. S-19: (a) Box-whisker plot; (b) Extracted mass spectrum of the target peak. No satisfactory matching can be found.



(a)

(b)

Fig. S-20: (a) Box-whisker plot; (b) Extracted mass spectrum of the target peak. No satisfactory matching can be found.

Reference

Acevedo, C.A., Sanchez, E.Y., Reyes, J.G., Young,M.E.(2007). "Volatile Organic Compounds Produced by Human Skin Cells." *Biological Research* (40): 347-355.

Adlard, E.R. (2007). "CHROMATOGRAPHY: Gas Detectors., Encyclopedia of separation." *Science* 456-464.

Aharoni, A., Ric de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., Goodenowe, D.B. (2002). "Non-targeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry." *OMICS* (6): 217-234.

Albrecht, M.G. and Creighton, J.A. (1977). "Anomously Intense Raman-Spectra of Pyridine at a Silver Electrode." *Journal of the American Chemical Society* 99(15): 5215-5217.

Almer, J.(1991)."Subclassification of polyacrylonitrile fibres by pyrolysis capillary gas chromatography." *Journal of the Canadian society of Forensic science* 24(1): 51-64.

Anderson, M.S. (2000). "Locally enhanced Raman spectroscopy with an atomic force microscope." *Applied Physics Letters* (76): 3130.

Antoniewicz, M.R., Kraynie, D.F., Laffend, L.A., González-Lergier, J., Kelleher, J.K., Stephanopoulos, G. (2007). "Metabolic previous term flux analysis next term in a non-stationary system: fed-batch fermentation of a high yielding strain of *E.coli* producing 1,3-propanediol. " *Metabolic Engineering* (9): 277-292.

Arneth, A. and Niinemets, U. (2010). "Induced BVOCs: how to bug our models?." *Trends in Plant Science* 15(3): 118-125.

Bailo, E. and Beckert, V. (2008). "Tip-Enhanced Raman Spectroscopy of Single RNA Strands: Towards a Novel Direct-Sequencing Method." *Angew. Chem. int. Ed* (47): 1658-1661.

Basanta, M., Singh, D., Fowler, S., Wilson, I., Dennis, R., Thomas, C.L.P. (2007). "Increasing analytical space in gas chromatography-differential mobility spectrometry with dispersion field amplitude programming." *Journal of Chromatography A* (1173): 129-138.

Beavis, R.C., Colby, S.M., Goodacre, R., Harrington, P., Reilly, J.P., Sokolow, S. and Wilkerson, C.W. (2000). "Artificial Intelligence and Expert Systems in Mass Spectrometry." Chichester, John Wiley& Sons Ltd.

Belda-Iniestaa, C., de Castro Carpeñoa, J., Carrascob, J.A., Morenoa, V., Casado Sáenza, E., Feliua, J.,Serenoa, M., García Ríoc, F., Barriusoa, J., Baróna, M.G. (2007). "New screening method for lung cancer by detecting volatile organic compounds in breath." *Clinical and Translational Oncology* (9): 364–368.

Berna, A. Z., Anderson, A. R., Trowell, S. C. (2007). "Bio-benchmarking of electronic nose sensors. " *PLoS ONE* 4(7): e6406.

Boatright, J., Negre, F., Chen, X., Kish, C.M., Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D.,Dudareva, N. (2004). "Understanding in vivo benzenoid metabolism in petunia petal tissue." *Plant Physiology* (135):1993-2011.

Boser, B.E., Guyon, I.M, Vapnik, V. N. (1992). "A training algorithm for optimal margin classifiers. " Haussler, D., editior, 5th Annual ACM Workshop on COLT. pp 144-152., Pittsburgh, PA., ACM Press.

Bottcher ,J. and Bassmann, H. (1984). "Identification of conjugated matabolites of antipyrine in urine by pyrolysis electron impact mass spectrometry." *Journal of Analytical and Applied Pyrolysis* (6): 1-18.

Brereton, R. G. (2003). "Data Analysis for the Laboratory and Chemical Plant." Chichester, Wiley.

Bro, R. (1997). "PARAFAC. Tutorial and applications." *Chemometrics and Intelligent Laboratory Systems* (38): 149-171.

Bro, R., Anderson, C.A., Kiers, H.A.L. (1999). "PARAFAC2: Part II. Modeling chromatographic data with retention time shift." *Journal of Chemometrics* (13): 295-309.

Brown, M., Dunn, W.B., Dobson, P., Patel, Y., Winder, C.L., Francis-McIntyre, S., Begley, P., Carroll,K., Broadhurst, B., Tseng, A., Swainston, N., Spasic, I., Goodacre, R and Kell, D.B. (2009). "Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics." *Analyst* (134): 1322-1332.

Burdick, D. (1995). "An introduction to tensor products with applications to multiway data analysis." *Chemometrics and Intelligent Laboratory Systems* (28): 229–237.

Carroll, J.D. and Chang, J. (1970). "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young" decomposition." *Psychometrika* (35): 283-309.

Chaigneau, M., Picardi, G., Ossikovski, R. (2010). "Tip enhanced Raman spectroscopy next term evidence for amorphous carbon contamination on gold surfaces." *Surface Science* 604(7-8): 701-705.

Chernushevich, I.V., Loboda, A.V., Thomson, B.A. (2001) "An introduction to quadrupole–time-of-flight mass spectrometry" *Journal of mass spectrometry* (36): 849-865

Chiavari, G., Fabbri,D., Mazzeo, R., Bocchini, P., Galletti, G.C. (1995). "Pyrolysis gas chromatographymass spectrometry of natural resins used for artistic objects." *Chomatographia* (41): 273-281.

Cornish-Bowden, A. and Cardenas, M.L. (2000). "From genome to cellular phenotype, a role for metabolic flux analysis?." *Nat. Biotechnology* (18): 267-268.

Coy, S. L., Krylov, E.V., Schneider, B. B., Covey, T. R., Brenner, D. J., Tryburski, J. B., Patterson, A. D., Krausz., K.W., Fornace, A.J., Nazarov, E.G. (2010). "Detection of radiation-exposure biomarkers by differential prefiltered mass spectrometry (DMS–MS)." *International Journal of Mass Spectrometry* (291): 108-117.

Dieterle, F., Schlotterbeck, G., Binder, M., Ross, A., Suter, L., Senn, H. (2007). "Application of Metabonomics in a Comparative Profiling Study Reveals N-Acetylfelinine Excretion as a Biomarker for Inhibition of the Farnesyl Pathway by Bisphosphonates." *Chemical Research in Toxicology* (20): 1291-1299.

Domke, K. F., Zhang, D., Pettinger, B. (2007). "Tip-Enhanced Raman Spectra of Picomole Quantities of DNA Nucleobases at Au(111)." *Journal of the American Chemical society* (21): 6708-6709.

Du, J.W. (2001). "Plant–insect chemical communication and its behavior control." *Acta Phytophysiologica Sinica*., (27): 193–200.

Dunn, W.B. (2008). "Current trends and future requirements for the mass spectrometric investigation of mammalian and plant metabolomes." *Physical biology* 5(1): 1-24.

Dunn, W.B., Bailey, N.J.C., Johnson, H.E. (2005a). "Measuring the metabolome: current analytical technologies." *Analyst* (130): 606-625

Dunn, W.B. and Ellis, D.I. (2005b). "Metabolomics: Current analytical platforms and methodologies." *Trends in Analytical Chemistry* 24(4): 285-294.

Dutta, R., Hines, E. L., Gardner, J. W., Boilot, P. (2002). "Bacteria classification using Cyranose 320 electronic nose." *Biomedical engineering online* 1(4):1-7.

Ellis, D.I. and Goodacre, R. (2001). "Rapid and quantitative detection of microbial spoilage of muscle foods: current status and future trends."*Trends in Food science & Technology* (12): 414-424.

EPA. (1999). "Compendium method TO-15: Determination of volatile organic compounds (VOCs) in ambient air using specially prepared canisters with subsequent analysis by gas chromatography." In: Compendium of methods for the determination of toxic organic compounds in ambient air, 2nd edn. US Environmental Protection Agency http://www.epa.gov/ttn/amtic/files/ambient/airtox/to-15r.pdf.

Fell, D.A. (1997). "Understanding the Control of Metabolism." London, Portland Press.

Fiehn, O. (2002). "Metabolomics: the link between genotypes and phenotypes."*Plant molecular biology* (48): 155-171.

Fifield, F.W. and Kealey, D. (2000). "Principles and Practice of Analytical Chemistry 5th edition." Oxford, UK,Blackwell.

Fleischmann, M., Hendra, P. J., McQuillan, A.J. (1974). "Raman spectra of pyridine adsorbed at a silver electrode." *Chemical Physics Letters* 26(2): 163-166.

Funahashi, K. (1989). "On the Approximate Realization of Continuous-mappings by Neural Networks." *Neural Networks* (2): 183-192.

Galletti, G.C., Reeves, J.B., Bocchini, P., Muscarella, C.I. (1997). "Compositional Differentiation of Maize Hybrid Stovers Using Analytical Pyrolysis and High-Performance Liquid Chromatography." *Journal of Agriculture and Food Chemistry* 45(5): 1715-1719

Garcia de Abajo, F.J. (2007). "Colloquium: Light scattering by particle and hole arrays." *Rev. Mod. Phys.* (79): 1267–1290. (doi:10.1103/RevModPhys.79.1267).

Geethakrishnan, T. and Palanisamy, P. K. (2006). "Degenerate four-wave mixing experiments in Methyl green dye-doped gelatin film." *Optik* (117): 282-286.

Georg Schmarr, H. and Bernhardt, J. (2010). "Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques." *Journal of Chromatography A* (1217): 565-574.

German, J.B., Watkins, S.M., Fay, L.B. (2005). "Metabolomics in Practice: Emerging Knowledge to Guide Future Dietetic Advice toward Individualized Health." *Journal of the American Dietetic Association* 105(9): 1425-1432.

Gersten, J. I. (1980a). "The effect of surface-roughness and surface enhanced Raman-scattering." *Journal of Chemical Physics* 72(10): 5779-5780.

Gersten, J. I. (1980b). "Rayleigh, Mie, and Raman-scattering by molecules adsorbed on rough surfaces." *Journal of Chemical Physics* 72(10): 5780-5781.

Gersten, J. I. and Nitzan, A. (1980). "Electromagnetic theory of enhanced Raman-scattering by molecues adsorbed on rough surfaces." *Journal of Chemical Physics* 73(7): 3023-3037.

Gersten, J. I. and Nitzan, A. (1981). "Spectroscopic properties of molecules interacting with small dielectric particles." *Journal of Chemical Physics* 75(3): 1139-1152.

Goodacre, R. (2007). "Metabolomics of superorganism." *Journal of Nutrition* (137), 259s-266s.

Goodacre, R. (1994). "Characterization and quantification of microbal systems using pyrolysis mass spectrometry: introducing Neural Network to analytical pyrolysis." *Microbiology Europe* 2(2): 16-22.

Goodacre, R. and Berkeley, R.C.W. (1990). "Detection of small genotypic changes in Escherichia coli by pyrolysis mass spectrometry." *FEMS Microbiology Letters* (71): 133-138.

Goodacre, R., Edmonds, A.N., Kell, D.B. (1993). "Quantitative analysis of the pyrolysis-mass spectra of complex mixtures using artificial neural networks: application to casamino acids in glycogen." *Journal of Analytical and Applied Pyrolysis* (26): 93-114.

Goodacre, R., Hartmann, A., Beringer, J.E., Berkeley, R.C.W. (1991). "The use of pyrolysis mass spectrometry in the characterization of Rhizobium meliloti." *Letters in Applied Microbiology* (13): 157-160.

Goodacre, R. and Kell, D. (1996). "Pyrolysis mass spectrometry and its applications in biotechnology." *Current Opinion in Biotechnology* (7):20-28.

Goodacre, R., Kell, D.B., Bianchi, G. (1992). "Neural networks and olive oil." *Nature* 359: 594.

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., Kell, D.B. (2004). "Metabolomics by numbers: acquiring and understanding global metabolite data." *Trends in Biotechnology* 22(5): 245-252.

Gopinath, A., Boriskina, S.V., Reinhard, B.M., Negro, L.D., (2009). "Deterministic aperiodic arrays of metal nanoparticles forsurface-enanced Raman sactering (SERS)." *Optics express* 17(5): 3741-3753.

Graham, D., Faulds, K., Smith, W.E. (2006). "Biosensing uing silver nanoparticles and surface enhanced resonance Raman scattering." *Chem. Commun* (43): 64-4371

Griffin, M.T. (2006). "Differential mobility spectroscopy for chemical agent detection." *Chemical and Biological Sensing VII., Proc. of SPIE* (6218): 6218061-6218069.

Grob, R.L. and Barry, E.F. (2004). "Modern practice of gas chromatography 4th Edition." New Jersey, Wiley and Sons.

Guzmán, M.G., Dille, J., Godet, S. (2009). "Synthesis of silver nanoparticles by chemical reduction method and their antibacterial activity." *International journal of chemical and biological engineering* 2(3): 104-111.

Hachenberg, H. and Schmidt, A.P. (1977). "Gas Chromatographic Headspace Analysis (D.Verdin, Trans.)." London, Hayden and Son.

Hagman, A. and Jacobsson, S. (1987). "Analysis of volatile organic compounds in polymers by dynamic headspace multidimensional gas chromatography mass spectrometry." *Journal of Chromatography A (*395): 271-279.

Hall, R.D. (2006)."Plant metabolomics: from holistic hope, to hype, to hot topic." *New Phytologist* (169): 453–468.

Hao, E and Schatz, G.C., (2004) "Electromagnetic fields around silver nanoparticles and dimers." *Journal of chemical physics*, (120): 357 -366.

Harrison, A. G. (1992). "Chemical Ionization Mass Spectrometry (2nd ed.)." Toronto, Ontario, Canada, CRC Press.

Hashimoto, K., Inoue, T., Nagaoka, K., Sumida, Y., Terada, S., Watanabe, C. (2005). "Rapid characterization of spices and herbs by direct heating sample introduction using a curie-point pyrolyzer." *Journal of high resolution Chromatography* 11(4): 347-350.

Heinzle, E., Matsuda, F., Miyagawa, H., Wakasa, K., Nishioka, T. (2007). "Estimation of metabolic previous fluxes expression levels and metabolite dynamics of a secondary metabolic pathway in potato using label pulse-feeding experiments combined with kinetic network modelling and simulation." *Plant Journal* (50): 176-187.

Helm, D., Labischinski, H., Schallehn, G., Naumann, D. (1991). "Classification and identification of bacteria by fouier-transform infrared spectroscopy." *Journal of general Microbiology* (137): 67-79.

Hornik, K., Stinchcombe, M., White, H. (1990). "Universal Approximation of an UnknownMapping and Its Derivatives Using Multilayer Feedforward Networks." *Neural Networks* (3):551-560.

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." *Journal of Educ. Psychology* (24): 417-441

Inui, Y. and Itioka, T. (2007). "Species-specific leaf volatile compounds of obligate Macaranga Myrmecophytes and host-specific aggressiveness of symbiotic crematogaster ants." *Journal of Chemical Ecology* (33): 2054–2063.

Ioffe, B.V. and Vitenberg, A.G. (1984). "Head-Space Analysis and Related Methods in Gas Chromatography (I.A. Mamantov, Trans.)." NewYork, John Wiley and Sons.

Jackson, T.C., Acuff, G.R., Dickson, J.S. (1997) In: Doyle M.P, Beuchat L.R, Montville T.J (eds) Food microbiology: fundamentals and frontiers. ASM, Washington.

Jarvis, R.M. and Goodacre, R. (2004). "Discrimination of bacteria using surface-enhanced Raman spectroscopy." *Analytical Chemistry* 76(1): 40-47.

Jarvis, R.M and Goodacre, R., (2004). "Ultra-violet resonance Raman spectroscopy for rapid discrimination of urinary tract infection bacteria." *FEMS Microbiology Letters* (232):127-132.

Jeanmaire, D.L. and van Duyne, R.P., (1977). "Surface Raman Spectroelectrochemistry .1. Heterocyclic, Aromatic, and Aliphatic-amines adsorbed on anodised silver electrodes." *Journal of Electroanalytical Chemistry* 84(1): 1-20.

Jia, C., Batterman, S., Godwin, C. (2007). "Continuous, intermittent and passive sampling of airborne VOCs." *Journal of Environmental Monitoring* (9): 1220-1230

Jiang, Y., Xie, P., Liang, G. (2009). "Distribution and depuration of the potentially carcinogenic malachite green in tissues of three freshwater farmed Chinese fish with different food habits." *Aquaculture* 288 (1-2):1-6

Johnson, R.S., Martin, S.A., Biemann,K. (1988). "Collision-induced fragmentation of $(M + H)^+$ ions of peptides. Side chain specific sequence ions." *International Journal of Mass Spectrometry and Ion Processes* (86):137-154.

Kell, D.B. (2006). "Systems biology, metabolic modelling and metabolomics in drug discovery and development." *Drug discovery today* (11): 1085-1092.

Kell, D.B. and Oliver, S.G. (2004). "Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era." *Bioessays* (26): 99-105.

Kerker, M. (1984). "Electromagnetic model for surface-enhanced Raman-scattering (SERS) on metal colloids." *Accounts of Chemical Research* 17(8): 271-277.

Kesselmeier, J. and Staudt, M. (1999). "Biogenic Volatile Organic Compounds (VOC): An Overview on Emission, Physiology and Ecology." *Journal of atmospheric chemistry* (33): 23-88.

Khoo, S.H.G. and Al-Rubeai, M. (2007). "Metabolomics as a complementary tool in cell culture." *Biotechnology and applied biochemistry* (47):71-84.

Knapp, D.R. (1979)."Handbook of analytical derivatization reactions. " New york, Wiley.

Koek, M.M., Muilwijk, B., van der Werf, M.J., Hankemeier, T. (2006). "Microbial Metabolomics with Gas Chromatography-Mass Spectrometry." *Analytical Chemistry* (78): 1272-1281.

Kolb, B. and Ettre, L.S. (2006). "Static Headspace Gas Chromatography Theory and Practice, 2nd edition." New York, John Wiley and Sons.

Kopka, J., Fernie, A.,Weckwerth, W., Gibon, Y., Stitt, M. (2004). "Metabolite profiling in plant biology: platforms and destinations." *Genome Biology* 5(6): 1091-1099.

Krebs, M.D., Zapata, A.M. Nazarov, E.G., Miller, R.A., Costa, I.S., Sonenshein, A.L., Davis, C.E. (2005) "Detection of biological and chemical agents using differential mobility spectrometry (DMS) technology." *IEEE Sensor Journal* 5(4): 696-703.

Kroonenberg, P.M., Harshman, R.A., Murakami, T. (2009). "Analysing Three-way Profile Data Using the Parafac and Tucker3 Models Illustrated with Views on Parenting." *Applied multivariate research* 13(1): 5-41.

Kroonenburg, P.M. (1983). "Three-mode principal component analysis. Theory and applications." Leiden, DSWO Press.

Kruger, N.J. and Ratcliffe, R.G. (2009). "Insights into plant metabolic networks from steady-state metabolic flux analysis." *Biochimie* 91(6): 697-702.

Krumbein, A., Peters, P., Brückner, B. (2004). "Flavour compounds and a quantitative descriptive analysis of tomatoes (Lycopersicon esculentum Mill.) of different cultivars in short-term storage." *Postharvest Biology and Technology* (32): 15-18.

Krylov, E.V., Nazarov, E.G., Miller, R.A. (2007). "Differential mobility spectrometer: Model of operation." *International Journal of Mass Spectrometry* (266): 76-85.

Krylov, E.V., Coy, S.L., Vandermey, J., Schneider, B.B., Covery, T.R., Erkinjon, G.N. (2010). "Selection and generation of waveforms for differential mobility spectrometry." *Review of scientific instruments* 81, 024101 1-10

Krylov, E.V. and Nazarov, E.G. (2009). "Electric field dependence of the ion mobility." *International Journal of Mass Spectrometry* (285): 149-156.

Landolt, P.J., Phillips, T.W. (1997). "Host plant influences on sex pheromone behaviour of phytophagous insects." *Annual Review of Entomology* (429): 371–391.

Laothawornkitkul, J., Moore, J.P., Taylor, J.E., Possell, M., Gibson, T.D., Hewitt, C.N., Paul, N.D., (2008). "Discrimination of plant volatile signatures by an electronic nose: a potential technology for plant pest and disease monitoring." *Environmental Science and Technology* (42): 8433–8439.

Lee, P.C. and Meisel, D. (1982). "Adsorption and surface-enhanced Raman of dyes on silver and gold sols."*Journal of Physical Chemistry* 86(17): 3391-3395.

Lee, S.K., Chi, J.G.., Park, S.C., Chung, S.I. (2000). "Transient Expression of Transglutaminase C During Prenatal Development of Human Muscles." *Journal of Histochemistry and Cytochemistry* (48): 1565-1574.

Levin, S.D., Miller, R.A., Nazarov, E.G., Vouros, P. (2007). "Using a Nanoelectrospray-Differential Mobility Spectrometer Mass Spectrometer System for the Analysis of Oligosaccharides with Solvent Selected Control Over ESI Aggregate Ion Formation." *Journal of the American Chemical society* 18(3): 502-511.

Lindon, J.C., Holmes, E., Nicholson, J.K. (2003). "So what's the deal with metabonomics? Metabonomics measures the fingerprint of biochemical perturbations caused by disease, drugs, and toxins." *Analytical Chemistry* (75): 384-391.

Lindon, J.C., Holmes, E., Nicholson, J.K. (2006). "Metabonomics Techniques and Applications to Pharmaceutical Research & Development." *Pharmaceutical Research* (23): 1075-1088.

Liu, Y.C., Yang, K.H., Hsu, T.C. (2009). "Direct evidence of chemical effect of surface-enhanced Raman scattering observed on electrochemically prepared rough gold substrates." *Analytica Chimica Acta* 636(1):13-18.

Lopez-diez, E.C. and Goodacre, R. (2004). "Characterization of microorganism using UV Resonnance Raman spectroscopy and chemometrics." *Analytical Chemistry* 76(3):585-591.

Lu, Y. and Harrington, P.B. (2007). "Forensic Application of Gas Chromatography–Differential Mobility Spectrometry with Two-Way Classification of Ignitable Liquids from Fire Debris." *Analytical Chemistry* 79(17): 6752-6759.

Manard, M.J. Trainham, R., Weeks, S., Coy, S.L., Krylo, E.V., Nazarov, E.G. (2010). "Differential mobility spectrometry/mass spectrometry: The design of a new mass spectrometer for real-time chemical analysis in the field." *International Journal of Mass Spectrometry* doi: 10.1016/j.ijms.2010.03.011.

March, R. E. (2000). "Quadrupole Ion Trap Mass Spectrometer." Chichester, Wiley and Sons.

March, R.E. (1998). "Quadrupole Ion Trap Mass Spectrometry: Theory, Simulation, Recent Developments and Applications." *Rapid communication in Mass spectrometry* (12): 1543-1554.

Marshall, A.G. and Hendrickson, C.L. (2002). "Fourier transform ion cyclotron resonance detection: principles and experimental configurations." *International Journal of Mass Spectrometry* (215): 59-75.

Martens, H. and Naes, T. (1989). "Multivariate calibration." Chichester, Wiley & Sons, Ltd.

Matsuda, F., Wakasa, K., Miyagawa, H. (2007). "Metabolic flux analysis in plants using dynamic labelling technique: application to tryptophan biosynthesis in cultured rice cells." *Phytochemistry* (68): 2290-2301.

McCall, S.L., Platzman, P.M., Wolff, P.A. (1980). "Surface enhanced Raman-scattering." *Physics Letters A* 77(5): 381-383.

McLean, J.A., Ruotolo, B.T., Gillig, K.J., Russell, D.H. (2005). "Ion mobility–mass spectrometry: a new paradigm for proteomics." *International Journal of Mass Spectrometry* (240): 301-315.

Mikhonin, A.V. and Asher, S.A. (2006). "Direct UV Raman monitoring of 310-Helix and π-Bulge premetlting during α-Helix unfolding." *Journal of the American chemical society* 128, (42): 13789-13795.

Miller, R.A., Nazarov, E.G., Eiceman, G.A., King, T.A. (2001). "A MEMS Radio-Frequency Ion Mobility Spectrometer for Chemical Agent Detection." *Sensor and Actuators A. Physical* (91): 301-312.

Morton, S.M. and Jensen, L. (2009). "Understanding the Molecule-Surface Chemical Coupling in SERS."*Journal of the American Chemical Society* 131(11): 4090-4098.

Moskovits, M. (1985). "Surface-enhanced spectroscopy." *Review of modern physics* (57): 783-826.

Mrozek, I. and Otto, A. (1990). "Quantitative Separation of the Classical Electromagnetic and the Chemical Contribution to Surface Enhanced Raman-Scattering." *Journal of Electron Spectroscopy and Related Phenomena* (54): 895-911.

Munger, J., Bennett, B.D., Parikh, A., Feng , X.J., McArdle, J., Rabitz, H.A., Shenk, T., Rabinowitz, J.D. (2008). "Systems-level metabolic previous term flux next term profiling

identifies fatty acid synthesis as a target for antiviral therapy." *Nature Biotechnology* (26): 1179-1186.

Munson, M.S.B. and Field, F.H. (1966). "Chemical Ionization Mass Spectrometry: I General Introduction."*Journal of the American Chemical Society* 88(12): 2621–2630.

Musteata, F.M. and Pawliszyn, J. (2007). "Bioanalytical applications of solid-phase microextraction." *Trends in Analytical Chemistry* 26(1): 36-45.

Navale, V. (1992). "A study of chemical transformation of glycerol ether lipids to hydrocarbons by flash and hydrous pyrolysis." *Journal of Analytical and Applied Pyrolysis* (23): 121-133.

Nazarov, E.G., Coy, S.L., Krylov, E.V., Miller, R.A., Eiceman, G.A. (2006). "Pressure Effects in Differential Mobility Spectrometry." *Analytical Chemistry* (78): 7697-7706.

Neacsu, C.C., Dreyer, J., Behr, N., Raschke, M.B. (2006). "Scanning-probe Raman spectroscopy with single-molecule sensitivity." *Physical Review B* (73): 1934061-193406.

Nicholson, J.K. and Lindon, J.C. (2008). "System biology: metabonomics". *Nature* (23): 1054-1156.

Nie, S.M. and Emery S.R. (1997). "Probing Single molecules and single nanoparticles by surface enhanced Raman scattering." *Science* (275):1102–1106.

Nöh, K., Grönke, K., Luoa, B.,Takors, R., Oldiges, M., Wiechert, W. (2007). "Metabolic previous term flux analysis next term at ultra short time scale: isotopically non-stationary 13C labeling experiments." *Journal of Biotechnology* (129):249-267

Nöh, K. and Wiechert, W. (2006). "Experimental design principles for isotopically instationary 13C labeling experiments." *Biotechnology and Bioengineering* (94): 234–251.

Nucera, D.M., Maddox, C.W., Hoien-Dalen, P. and Weigel, R.M. (2006). "Comparison of API 20E and invA PCR for Identification of Salmonella enterica Isolates from Swine Production Units." *Journal of Clinical Microbiology*. 44 (9): 3388–3390

Oliver, S.G., Winson, M.K., Kell, D.B., Baganz, F. (1998). "Systematic functional analysis of the yeast genome." *Trends in Biotechnology* (16): 373-378.

Oresic, M. (2009). "Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction." *Nutrition, Metabolism & Cardiovascular Diseases* (19): 816-824.

Otto, A. (1991). "Surface-Enhanced Raman-Scattering of Adsorbates." *Journal of Raman Spectroscopy* 22(12): 743-752.

Otto, A., Mrozek, I., Grabhorn, H., Akemann, W.(1992). "Surface-Enhanced Raman-Scattering." *Journal of Physics-Condensed Matter* 4(5): 1143-1212.

Pawliszyn, J. (2002). "Ed. Sampling and Sample Preparation for Field and Laboratory. " Amsterdam Elsevier.

Pawliszyn, J. (2003)."Sample Preparation: Quo Vadis?." *Analytical Chemistry* (75): 2543-2558.

Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space." *Philos. Mag* (2): 559-572.

Penuelas, J. and Llusia, J. (2003). "BVOCs: plant defense against climate warming?." *Trends in Plant Science* (8):105-109.

Pettinger, B., Domke, K.F., Zhang, D., Picardi, G., Schuster, R. (2007). "Tip-enhanced Raman scattering: Influence of the surface geometry on optical resonance and enhancement. " *Surface Science* 603(10-12): 1335-1341.

Plumb, R.S., Granger, J.H., Stumpf, C.L., Johnson, K.A., Smith, B.W., Gaulitz, S., Wilson, I.D. Castro-Perez, J. (2005). "A rapid screening approach to metabolomics using UPLC and oa-TOF mass spectrometry: application to age, gender and dinurnal variation in normal/Zucker obese rats and black, white and nude mice." *Analyst* (130): 844-849

Prasad, S., Pierce, K.M., Schmidt, H., Rao, J.V., Guth, R., Bader, S., Synovec, R.E., Smith, G.B., Eiceman,G.A. (2007). "Analysis of bacteria by pyrolysis gas chromatographya dependence of growth temperature." *Analyst* (132): 1031-1039.

Prasad, S., Schmidt, H., Lampen, P., Wang, M., Guth, R., Rao, J., Smith, G.B., Eiceman, G.A. (2006). "Analysis of bacterial strains with pyrolysis-gas chromatography/differential mobility spectrometry." *Analyst* (131): 1216-1225.

Rammsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K., Oliver, S.G. (2001). "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations." *Nature biotechnology* (19): 45-50.

Rearden, P., Harrington P.B., Karnes, J.J., Bunker, C.E. (2007). "Fuzzy rule-building expert system classification of fuel using solid-phase microextraction two-way gas chromatography differential mobility data." *Analytical Chemsitry* 79(4): 1485-1491

Riazanskaia, S., Blackburn, G., Harker, M., Taylor, D., Thomas, C.L.P (2008). "The analytical utility of thermally desorbed polydimethylsilicone membranes for in-vivo sampling of volatile organic compounds in and on human skin." *Analyst* (133): 1020–1027.

Rodger, C., Rutherford, V., Broughton, D., White, P.C., Smiths, W.E. (1998). "Towards Quantitative Surface Enhanced Resonance Raman Scattering (SERRS): a Study of Aggregation and Concentration for Two Rhodamine Dyes." *Journal of Raman spectroscopy* (29): 601-606.

Rodger, C., Smith, W.E., Dent, G., Edmondson, M. (1996) "Surface-enhanced resonance Raman scattering: an informative probe of surface." *J Chem Soc Dalton Trans* 791–799.

Ruan, C., Eres, G., Wang, W., Zhang, Z., Gu, B. (2007) "Controlled fabrication of nanopillar arrays as active substrates for surface enhance Raman spectroscopy." *Langmuir (*23): 5757-5760.

Ryan, D. and Robards, R. (2006). "Metabolomics: The Greatest Omics of Them All?" *Analytical Chemsitry* (78):7954-7958.

.

Sanchez-Palomo, E., Diaz-Maroto, M.C., Perez-Coello, M.S.(2005). "Rapid determination of volatile compounds in grapes by HS-SPME coupled with GC–MS." *Talanta* 66(5): 1152-1157.

Sauer, S., Freiwald, A., Maier, T., Kube. M., Reinhardt, R., Kostrzewa, M., Geider, K., (2008). "Classification and Identification of Bacteria by Mass Spectrometry and Computational Analysis." *PloS ONE* 3 (7): e2843.

Sauer, S. and Kliem, M. (2010) "Mass spectrometry tools for the classification and identification of bacteria." (2010). *Nature review Microbiology* (8): 74-82.

Schatz, G.C. (2006). "Electromagnetic enhancement of SERS. Surface-Enhanced Raman Scattering: Physics and Applications." Kneipp, M.M.K., Kneipp, H., Berlin Heidelberg New York, Springer (103): 19-45.

Schmidt, H., Tadjimukhamedov, F., Mohrenz, I.V., Smith, G.B., Eiceman, G.A. (2004). "Microfabricated Differential Mobility Spectrometry with Pyrolysis Gas Chromatography for Chemical Characterization of Bacteria." *Analytical Chemistry* (76): 5208-5217.

Schwender, J., Ohlrogge, J., Shachar-Hill, Y. (2004). "Understanding flux in plant metabolic." *Current opinion in Plant Biology* 7(3): 309-317.

Shadi, I.T., Chowdhry, B.Z., Snowden, M.J., Withnall, R. (2001). "Semi-quantitative trace analysis of nuclear fast red by surface enhanced resonance Raman scattering." *Analytical Chimica Acta* (450): 115-122.

Shadi, I.T., Chowdhry, B.Z., Snowden, M.J., Withnall, R. (2004). "Semi-quantitative analysis of alizarin and purpurin by surface-enhanced resonance Raman spectroscopy (SERRS) using silver colloids." *Journal of Raman spectroscopy* 35(8-9): 800-807.

Shadi, I.T., Chowdhury, B.Z., Snowden, M.J., Withnail, R. (2003). "Semi-quantitative analysis of indigo carmine, using silver colloids, by surface enhanced resonance Raman spectroscopy (SERRS)." *Spectrochimca Part A* (59): 2201-2206.

Shedrinsky, A.M., Wampler, T.P., Indictor, N., Baer, N.S. (1989). "Application of analytical pyrolysis to problems in art and archaeology: A review." *Journal of Analytical and Applied Pyrolysis* (15): 393- 412.

Shnayderman, M., Manfield, B., Yip, P., Clark, H.A., Krebs, M.D., Cohen, S.J., Zeskind, J.E., Ryan, E.T., Dorkin, H.L., Callahan, M.V., Stair, T.O., Gelfand, J.A., Gill, C.J., Davis, C.E. (2005) "Species-specifics bacteria identification using differential mobility spectrometry and bioinformatics pattern recognition." *Analytical Chemistry* (77): 5930-5937.

Sirpatrawan, U. and Harte, B.R. (2007). "Solid phase microextraction/gas chromatography/mass spectrometry integrated with chemometrics for detection of *S. typhimurium* contamination in a packaged fresh vegetable." *Analytica Chimica Acta* 58(1): 63-70.

Smith, E. and Dent, G. (2005). "Modern Raman spectroscopy: A practical approach." Chichester, J.Wiley and Sons.

Smits, J.R.M., Melssen, W.J., Buydens, L.M.C., Kateman, G. (1994). "Using artificial neural networks for solving chemical problems Part I. Multi-layer feed-forward networks." *Chemometrics and Intelligent Laboratory Systems* (22): 165-189.

Song, G., Qin, T., Liu, H., Xu, G.B., Pan ,Y.Y., Xiong ,F.X., Gu, K.S., Sun ,G.P., Chen ,Z.D. (2009)."Quantitative breath analysis of volatile organic compounds of lung cancer patients." *Lung Cancer* (67): 227–231.

Srivsatava,S., Sinha, R., Roy, D. (2004). "Toxicological effects of malachite green."*Aquatic Toxicology* (66); 319-329.

Stanbridge, L.H. and Davies, A.R. (1998). "In: Davies A, Board R (eds) The microbiology of meat and poultry." Blackie Academic & Professional, London.

Stankiwicz, B.A., Hutchins, J.C., Thomson, R., Briggs, D.E.G., Evershed, R.P. (1998). "Assessment of bog-body tissue preservation by pyrolysis-gas chromatography/mass spectrometry." *Rapid Communcations in Mass spectrometry* 11(17): 1884-1890.

Stein, S.E. (1999). "An integrated method for spectrum extraction and compound identification from gas chromatography/ mass spectrometry. " *Rapid Communication in Mass Spectrometry* (13): 279-284.

Stockle, R.M., Suh, Y.D., Deckert, V., Zenobo, R. (1999). "Nanoscale chemical analysis by tip-enhanced Raman spectroscopy." *Chemical Physics Letters* (318): 131-136.

Tuschel, D.D., Mikhonin, A.V., Lemoff, B.E., Asher, S.A. (2010). "Deep Ultraviolet Resonance Raman Excitation enables explosives detection." *Applied Spectroscopy* 64(4): 425-432.

Theodoridis, G., Gika, H.G., Wilson, I.D. (2008) "LC-MS-based methodology for global metabolite profiling in metabolonomics /metabolomics." *Trends in Analytical Chemistry* 27 (3): 251-260.

Umemura, S., Itohm J., Takekoshi, S., Hasegawa, H., Yasuda, M., Osamura, R.Y., Watanabe, K. (2003). "Methyl green staining and DNA Strands in Vitro: High Affinity of methyl green dye to Cytosine and Guanine." *Acta Histochem. Cytochem*. 36 (4): 361-366.

Van Deemter, J.J., Zuiderweg, F.J. and Klinkenberg, A. (1956). "Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography." *Chemical engineering Science* (5): 271-289.

van der Greef, J., Martin, S., Juhasz, P., Adourian, A., Plasterer, T., Verheij, E.R., McBurney, R.N. (2007). "The Art and Practice of Systems Biology in Medicine: Mapping Patterns of Relationships." *Journal of proteome research* (6): 1540-1559.

van der Greef, J., Stroobant. P., van der Heijden, R. (2004). "The role of analytical sciences in medical systems biology." *Current Opinion in Chemical Biology* 8(5): 559-565.

van Ravenzwaay, B., Coelho-Palermo Cunha, G., Leibold, R., Looser, R., Mellert, W., Prokoudine, A., Walk, T., Wiemer, J. (2007). "The use of metabolomics for the discovery of new biomarkers of effect." *Toxicology letters* (172): 21-28.

Vapnik, V.N. (1995). "The nature of statistical learning." Berlin, Springer.

Wampler, T.P. (1999). "Introduction to pyrolysis–capillary gas chromatography." *Journal of Chromatography A* (842): 207-220.

Wampler, T.P., Bishea, G.A., Simonsick, W.J. (1997). "Recent changes in automotive paint formulation using pyrolysis-gas chromatography/mass spectrometry for identification." *Journal of Analytical and Applied Pyrolysis* (40-41): 79-89.

Wang, X.C. and Zhao, Y. L. (1995). "Structural theory of odor molecule." *Chinese Journal of Chemical Education* (8): 1–3.

Wang, Z., He, Y., Huang, L. (2007). "An alternative method for the rapid synthesis of partially Omethylated alditol acetate standards for GC–MS analysis of carbohydrates." *Carbohydrate Research* (342): 2149-2151.

Weckwerth, W. and Morgenthal, K. (2005). "Metabolomics:from pattern recognition to biological interpretation." *Drug discovery today* 10(22): 1551-1558.

Werbos, P.J. (1994). "The Roots of Back-propagation: From Ordered Derivatives to Neural Networks and Political Forecasting." Chichester, John Wiley.

White, H. (1990). "Connectionist Nonparametric Regression –Multilayer Feedforward Networks Can Learn Arbitrary Mappings." *Neural Networks* (3): 535-549.

Wishart, D.S. (2008). "Metabolomics: applications to food science and nutrition research." *Trends in Food Science & Technology (*19): 482-493.

Wold, S. (1989). "In Computer Applications in Chemical Research and Education." Heidelberg.

Xu, Y., Zomer, S., Brereton, R. (2006). "Support Vector Machines: A Recent Method for Classification in Chemometrics." *Critical Reviews in Analytical Chemistry* (36): 177-188.

Yu, C. and Irudayaraj, J., (2005). "Spectroscopic characterization of microorganisms by fourier transform infrared microspectroscopy." *Biopolymers* (77): 368-377.

Yuan, J., Liao, L., Lin, Y., Deng, C., He, B. (2008). "Determination of Sudan I in chilli powder from solvent components gradual change–visible spectra data using second order calibration algorithms."*Analytica Chimca Acta*: 161-167.

Zeng, X.B. (2002). "The essence of the theory and characteristics of smell, Chinese." *Journal of Basic Medicine in Traditional Chinese Medicine* (8): 101–102.

Zhang, Z. and Li, G. (2010). "A review of advances and new developments in the analysis of biological volatile organic compounds." *Microchemical Journal* 95 (2): 127-139

Zhu, J., Qiu, Y., Valobra, M., Qiu, S., Majumdar, S., Matin, D., De Rose, V., Jeffery, P.K. (2007). "Plasma Cells and IL-4 in Chronic Bronchitis and Chronic Obstructive Pulmonary Disease." *American Journal of Respiratory critical care medicine*. (175): 1125–1133

Zini, C.A., Augusto, F., Christensen, E., Smith, B.P., Caramão, E.B., Pawliszyn, J. (2001). "Monitoring biogenic volatile compounds emitted by Eucalyptus citriodora using SPME." *Analytical Chemistry* (73): 4729–4735.