University of Strathclyde

Department of Pure and Applied Chemistry

Université de Reims Champagne-Ardenne

Ecole Doctorale Sciences Technologie Santé

Rapid Pre-Symptomatic Diagnosis of Sepsis by Vibrational Spectroscopy

By

Lila Lovergne

A thesis presented to the Department of Pure and Applied Chemistry, University of Strathclyde, in fulfilment of the requirements for the degree of Doctor of Philosophy.

January 2018

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Date:

Acknowledgements

Résumé

Le sepsis est une dérégulation de la réponse de l'hôte à une infection et est associé à un dysfonctionnement des organes engageant le pronostic vital du patient. Chaque année, plus de 30 millions de cas de sepsis et 5 millions de décès sont estimés à travers le monde. Le diagnostic du sepsis est basé sur des signes cliniques non spécifiques et la longue procédure d'identification des pathogènes responsables de l'infection. La détection rapide et pré-symptomatique d'un sepsis permettrait l'administration à un stade précoce d'un traitement adapté, maximisant son effet, réduisant la mortalité et les frais de santé associés.

La spectroscopie vibrationnelle permet d'obtenir une ''empreinte moléculaire'' de matériaux biologiques et présente de nombreux aspects avantageux. En effet, cette technique est sans marquage, non invasive, sans contact, rapide, peu coûteuse, simple d'utilisation et sans préparation lourde d'échantillons.

L'objectif de cette étude est de développer et d'évaluer le potentiel de la spectroscopie vibrationnelle appliquée au sérum humain en vue d'améliorer le diagnostic de patients atteints d'un sepsis. Les défis de la spectroscopie appliquée au sérum inhérents à la nature de l'échantillon et à la technique ont été évalués. Différentes modalités de spectroscopies infrarouge et Raman, de même que différentes préparations d'échantillons ont été comparées pour déterminer l'approche méthodologique la plus adaptée à l'analyse du sérum avec un objectif global d'application clinique.

Ensuite, certains aspects de la phase pré-analytique ont été contrôlés dans le but de standardiser les protocoles dans la manipulation et la préparation des échantillons afin d'assurer la qualité et la reproductibilité des données spectrales collectées.

Finalement, à partir de la méthodologie développée, des sérums de patients septicémiques (n=380) collectés avant chirurgie et jusqu'à 3 jours avant le jour du diagnostic ainsi que le jour du diagnostic ont été analysés. Les échantillons du groupe contrôle comprenaient des sérums (n=353) de patients ayant un profil similaire en termes d'âge, de sexe, de procédure chirurgicale subie mais n'ayant pas développé de

sepsis. Les échantillons ont également été obtenus suivant la même cinétique de collecte. De même, des échantillons (n=180) de patients présentant un syndrome de réponse inflammatoire systémique (SRIS), collectés avant chirurgie et le jour du diagnostic ont été analysés. Les données acquises ont été exploitées par méthodes chimiométriques dans le but de discriminer des zones spectrales reflétant des différences de composition moléculaire.

Abstract

Sepsis is a dysregulated host response to an infection that causes life-threatening organ dysfunction. Each year, over 30 million cases and 5 million deaths are estimated worldwide. Diagnosis of sepsis is based on non-specific clinical signs and time consuming positive identification of the causative pathogen(s). Rapid pre-symptomatic detection of sepsis would enable the early administration of therapeutics, maximising their effects, reducing mortality and healthcare costs.

Vibrational spectroscopy can provide a "molecular fingerprint" of biological materials and presents many advantageous aspects. Indeed, this technique is label-free, non-destructive, non-contact, rapid, cost-effective, simple to operate, and requires only simple sample preparation.

The objective of this study was to develop and evaluate the potential of vibrational spectroscopy applied on human serum with the aim to improve diagnosis of patients with sepsis. Challenges of serum spectroscopy inherent to the sample nature and to the technique have been assessed. Different infrared and Raman spectroscopy modalities as well as different serum sample preparations have been compared to determine the most suitable methodology approach with an overall clinical application purpose.

Then, some aspects of the pre-analytical phase have been addressed in order to standardise protocols in sample handling and preparation for spectral acquisitions to ensure quality and reproducibility of spectral data collected.

Finally, based upon the developed methodology, patient serum samples (n=380) collected before surgery, up to 3 days before sepsis diagnosis, and on the day of sepsis diagnosis have been analysed. Control serum samples (n=353) from age/ sex/ procedure-matched patients who did not go on to develop sepsis have been also analysed over similar timeframes post-surgery as well as samples (n=190) from patients with systemic inflammatory response syndrome. Spectral data acquired have been interrogated by chemometric methods to discriminate spectral zones reflecting differences in molecular composition.

Contents

Chapter V. Investigating pre-analytical requirements for infrared spectroscopy of blood-derived products 137

List of abbreviations

AIDS Acquired Immune Deficiency Syndrom

AMI Acute Myocardial Infarction

aPC activated Protein C

ATR Attenuated Total Reflectance

BSA Bovine Serum Albumin

CaF$_2$ Calcium Fluoride

CARS Coherent Anti-Stokes Raman Scattering

CCD Charged-Coupled Device

CLRs C-Type Lectin Receptors

CO$_2$ Carbon Dioxide

CRP C-Reactive Protein

DAMPs Damage-Associated Molecular Patterns

DF Discrete Frequency

DFA Discriminant Function Analysis

DGA Direction Générale de l'Armement

Dstl Defence Science & Technology Laboratory

DTGS Deuterated Triglycine Sulfate

EDTA Ethylenediaminetetraacetic Acid

EFSW Electric Field Standing Wave Artefact

EMSC Extended Multiplicative Signal Correction

FIR Far-Infrared

FT Fourier-Transform

G$^-$ Gram-Negative

G$^+$ Gram-Positive

GBM GlioBlastoma Multiforme

HCA Hierarchical Cluster Analysis

HSA Human Serum Albumin

HT High Throughput

ICU Intensive Care Unit

IL Interleukin

IR Infrared

KBr Potassium Bromide

LDA Linear Discriminant Analysis

LOOCV Leave-One-Out Cross Validation

MCT Mercury Cadmium Telluride

MIR Mid-Infrared

NHS National Health Service

NIR Near-Infrared

NLRs Nucleotide-Binding Oligomerisation Domain-Like Receptors

PaCO$_2$ Partial Pressure of Carbon Dioxide

PAMPs Pathogen-Associated Molecular Patterns

PCA Principal Component Analysis

PCT Procalcitonin

PLS-DA Partial Least Square-Discriminant Analysis

PRRs Pattern-Recognition Receptors

PTSD Posttraumatic Stress Disorder

QCL Quantum Cascade Lasers

QDA Quadratic Discriminant Analysis

qSOFA quick Sequential Organ Failure Assessment

QT Quality Test

RF Random Forest

RH Relative Humidity

RLRs Retinoic Acid Inducible Gene 1-Like Receptors

RNA Ribonucleic Acid

RRS Resonant Raman Scattering

SEIRA Surface-Enhanced Infrared Absorption

SERS Surface-Enhanced Raman Scattering

SIRS Systemic Inflammatory Response Syndrome

SOFA Sequential Organ Failure Assessment

SORS Spatially Offset Raman Spectroscopy

SR Synchrotron Radiation

SRS Stimulated Raman Scattering

STD Standard Deviation

SVM Support Vector Machine

TLRs Toll-Like Receptors

TNF Tumor Necrosis Factor

UK United Kingdom

List of figures

List of tables

Chapter I. Introduction

I.1 Résumé du chapitre I. Introduction

I.1.1 Le sepsis

I.1.1.1 Incidence

Le sepsis est une dérégulation de la réponse de l'hôte à une infection et est associé à un dysfonctionnement des organes engageant le pronostic vital du patient. Le choc septique est défini comme un sous-ensemble du sepsis dans lequel surviennent de profondes anomalies circulatoires, cellulaires et métaboliques qui sont associées à un plus grand risque de mortalité[1,2]. Le sepsis est une cause majeure de décès dans les hôpitaux, notamment dans les services de soins intensifs[3]. Malgré le manque d'études épidémiologiques robustes, en particulier dans les pays en développement, l'incidence du sepsis est estimée à plus de 30 millions de cas à travers le monde par an et serait responsable de plus de 5 millions de décès[4]. L'incidence du sepsis devrait augmenter dans un futur proche en conséquence du vieillissement des populations des pays développés et de l'augmentation des cas de résistance aux antibiotiques[5]. Au Royaume-Uni, environ 147 000 cas sont estimés chaque année dans les hôpitaux, causant le décès de 44 000 personnes et des coûts directs de 830 millions de livres sterling au système de santé britannique, le NHS (National Health Service)[6]. De plus, les patients ayant survécu à un sepsis peuvent souffrir de séquelles physiques et cognitives ou d'un stress post-traumatique nécessitant des soins à long terme et des frais de santé supplémentaires[7–10].

I.1.1.2 Diagnostic

Bien que le sepsis est une incidence de même qu'une morbidité et mortalité élevées, aucune technique ''gold standard'' n'existe à l'heure actuelle pour le diagnostic du sepsis et sa définition est toujours en évolution[1,11,12]. La difficulté du diagnostic d'un sepsis est due sa nature complexe. Le sepsis peut être causé par différents pathogènes, le plus communément par des bactéries mais aussi des virus, champignons ou parasites[13]. De plus, le développement d'un sepsis dépend de l'état de santé général d'un patient, par exemple la coexistence de maladies (chroniques), une déficience du

système immunitaire, la génétique, la prise traitement médicamenteux[5]. Il est aussi connu que le sepsis a une plus forte incidence chez les hommes que chez les femmes, ainsi que chez les enfants et populations âgées[14].

Un autre point crucial dans le diagnostic d'un sepsis est l'identification de l'agent pathogène responsable de l'infection. Malgré de récentes avancées technologiques, la principale méthode d'identification reste la culture d'échantillons cliniques qui est une procédure longue (généralement de 24 à 48h) et est positive dans seulement 30 à 40% des cas. Cela est dû au fait que certains microorganismes ne se développent pas dans ce type de cultures ou le fait que certains patients peuvent recevoir un traitement avant le prélèvement d'échantillons cliniques[20–23]. Ainsi, jusqu'à ce que le pathogène soit identifié, les patients sont traités par l'administration d'antimicrobiens à large spectre pour couvrir les pathogènes les plus communs (bactéries, champignons, virus) avant de cibler le traitement pour réduire les dépenses de santé et prévenir la résistance aux antibiotiques[24]. Prodiguer un traitement dans l'heure suivant l'observation d'une hypotension chez un patient en choc septique est associé à taux de survie de 80%, ce taux décroissant à chaque heure de retard[25–27]. L'administration d'un mauvais traitement initial réduit par 5 les chances de survie d'un patient[28]. Ainsi, l'identification rapide d'un cas de sepsis est d'une importance majeure pour la réussite d'un traitement.

De par la complexité de la pathogénèse et la nature systémique du sepsis, il est certain que de nouvelles stratégies impliquant une combinaison de biomarqueurs devraient être considérées pour le diagnostic, le pronostic et l'élaboration de nouveaux traitements pour le sepsis[34]. Ainsi, le concept de recherche de signatures spectrales dans les échantillons de biofluides de l'hôte pré-symptomatiquement apparaît comme une approche prometteuse pour un diagnostic rapide et afin de permettre une intervention thérapeutique précoce[33].

I.1.2 La spectroscopie vibrationnelle

La spectroscopie vibrationnelle concerne les techniques optiques spécifiques que sont la spectroscopie infrarouge (IR) et Raman. Ces techniques sondent les

rotations et vibrations intramoléculaires d'un échantillon quand celui-ci est irradié avec une source lumineuse[45]. La relation lumière-matière est décrite par la théorie électromagnétique de Maxwell[46]. La spectroscopie vibrationnelle a été utilisée pour analyser un vaste rang d'échantillons pour des applications chimiques, physiques et biologiques. Les liaisons moléculaires chimiques constitutives présentent de nombreuses formes de vibrations qui interviennent à différentes énergies correspondant à différents niveaux de transitions. Les spectroscopies IR et Raman sont des techniques complémentaires qui permettent d'obtenir une ''empreinte'' ou ''signature'' des molécules contenues dans un échantillon selon que leurs liaisons possèdent une activité IR ou Raman. Certaines vibrations permises en Raman, peuvent ne pas se produire en IR et inversement.

I.1.2.1 La spectroscopie infrarouge

La spectroscopie IR est définie comme l'étude des caractéristiques d'absorption découlant du mouvement moléculaire de matériaux dû au déplacement atomique[49] lors d'une interaction intime avec une source infrarouge[50]. En fonction du choix de modalité utilisée, le rayonnement peut être soit transmis, réfléchi en interne, réfléchi, ou transflecté (une combinaison de transmission et de transflection). Lors d'une interaction lumière-matière, la lumière infrarouge provoque l'entrée dans état vibrationnel supérieur due au transfert de ''quanta'' ou ''paquets'' d'énergie à certaines longueurs d'ondes en fonction de la composition de la matière analysée. Ces transitions permettent l'obtention d'un spectre constitué de pics et bandes qui peuvent être interprétés qualitativement (position des pics) et quantitativement (intensité des pics/aires, intensité relative). Dans le cas de la spectroscopie IR, les bandes proviennent d'un changement du moment dipolaire électrique des molécules. L'absorbance des bandes est proportionnelle à la concentration des liaisons moléculaires du matériel étudié comme énoncé par loi de Beer-Lambert.

I.1.2.2 La spectroscopie Raman

L'effet Raman constitue le processus spontané de diffusion de la lumière inélastique de photons à la suite d'une interaction entre une radiation monochromatique (ex: source laser) et un échantillon. Lors de cette interaction les deux processus de diffusion élastique et inélastique prennent place. Une grande proportion de photons est élastiquement diffusée sans changement d'énergie (par conséquent sans information moléculaire). Ce phénomène est connu sous le nom de diffusion Rayleigh[46]. Lorsque les photons transfèrent de l'énergie aux molécules sous forme d'énergie vibratoire, la perte d'énergie des photons diffusés correspond aux niveaux d'énergie vibratoires des molécules. Ce processus est la diffusion Raman-Stokes. Les photons incidents peuvent en retour recevoir de l'énergie des molécules en vibration, ainsi leurs fréquences augmentent, décrivant le principe de la diffusion Raman anti-Stokes. En Raman spontané, la diffusion Stokes est généralement utilisée en raison de sa plus grande sensibilité.

I.1.3 Objectif de l'étude

Le principal objectif de cette étude est de développer et d'évaluer le potentiel de la spectroscopie vibrationnelle appliquée au sérum humain pour le diagnostic rapide et pré-symptomatique du sepsis.

I.2 Sepsis

I.2.1 Incidence

Sepsis is defined as a ''life-threatening organ dysfunction caused by a dysregulated host response to infection'' and septic shock as "a subset of sepsis in which particularly profound circulatory, cellular, and metabolic abnormalities are associated with a greater risk of mortality than with sepsis alone"[1,2]. Sepsis is a leading cause of death in hospitals, especially in the Intensive Care Unit (ICU)[3]. Despite a lack of robust epidemiological studies, particularly in developing countries, the incidence of sepsis is estimated to be over 30 million cases worldwide per year and is responsible for more than 5 million deaths[4]. The incidence of sepsis is believed to increase in the future due to aging populations in developed countries and the increase of multidrug resistance[5]. In the United Kingdom (UK) up to 147,000 sepsis cases are estimated per year in hospitals, causing 44,000 deaths, and direct costs of £830 million to the National Health Service (NHS)[6]. Patient surviving sepsis can also suffer from physical, cognitive impairments or Posttraumatic Stress Disorder (PTSD) requiring long-term treatment and additional healthcare costs[7–10].

I.2.2 Diagnosis

Although sepsis has a high incidence as well as a high morbidity and mortality, there is no gold standard technique to diagnose sepsis and its definition is still evolving[1,11,12]. The difficulty occurring in the diagnosis of sepsis is due to its complex nature. Sepsis can be caused by different pathogens, more commonly by bacteria but also virus, fungi or parasites[13]. Among Gram-positive bacteria, *Staphylococcus aureus* and *Streptococcus pneumoniae* are the most common sources of sepsis as well as *Escherichia coli*, *klebsiella* species, and *Pseudomonas aeruginosa* among Gram-negative bacteria. Pneumonia is the main cause, with intra-abdominal and urinary tract infections[3]. Another difficulty to establish a diagnosis is the fact that sepsis is dependent upon the patients general health conditions such as the presence of co-existing (chronic) diseases, immune system deficiency, genetics, medications[5]. Sepsis

is also known to have a higher incidence in males than females, in children and elderly population[14].

It is established that sepsis triggers both pro- and anti-inflammatory responses also leads to profound modifications of non-immunologic parameters. Therefore, following the recent update of the sepsis definition[1], the use of the Sequential Organ Failure Assessment (SOFA) score is recommended to identify potential cases of sepsis when infection is suspected among patients in the ICU. Criteria of the SOFA score are presented in Table 1.1 and to summarise, respiratory (fraction of inspired oxygen, partial pressure of oxygen), coagulation (platelets), liver (bilirubin), cardiovascular (hypotension), central nervous system (Glasgow Coma Score Scale[15]) and renal (serum creatinine or urine output) factors are assessed during a patient stay in the ICU to evaluate the degree and evolution of organ dysfunction on a score from 0 to 4 for each organ[16]. A high SOFA score is correlated with a high mortality rate[17]. A diagnosis based on at least 2 of the 4 Systemic Inflammatory Response Syndrome (SIRS) criteria based on body temperature, heart rate, respiratory rate, and white blood cell count (Table 1.2) has been declared not enough specific to diagnose a sepsis but may still be of use[18]. Indeed, SIRS criteria are common to patients with different non-infectious conditions such as trauma, pancreatitis, ischemia[19]. Septic shock can be recognised when a patient requires a vasopressor administration to maintain a mean arterial pressure at a minimum level of 70 mmHg. Although the utility of lactate measurement to characterise a septic shock is still debated, a lactate level greater than 2 mmol/L associated with hypotension represents a higher mortality risk population and therefore both criteria can be helpful to identify septic shock patients[2,18].

A simpler version of the SOFA score, the quick SOFA (qSOFA) score, has been adapted to rapidly identify patients with suspected infection, and therefore potential sepsis cases, outside of the ICU. Criteria of the qSOFA score are a minimum respiratory rate of 22/min, an altered mental status and a maximum systolic blood pressure of 100 mmHg. Further investigations should be considered for patients with at least 2 of the criteria described to confirm a sepsis[1]. An overview of the criteria allowing the identification of sepsis and septic shock cases is presented in Figure 1.1.

Table 1.1 *Sequential Organ Failure Assessment (SOFA) score, adapted from ref 16*

Variables	SOFA score				
	0	1	2	3	4
Respiratory					
PaO$_2$/FIO$_2$ (mmHg)	> 400	≤ 400	≤ 300	≤ 200 ⊤	≤ 100 ⊤
Coagulation					
Platelets (x10^3/μL)	> 150	≤ 150	≤ 100	≤ 50	≤ 20
Liver					
Bilirubin (mg/dL)	< 1.2	1.2-1.9	2.0-5.9	6.0-11.9	12
Cardiovascular					
Hypotension	No hypotension	Mean arterial pressure <70 mmHg	Dop ≤ 5 or Dob (any dose)	Dop > 5 Epi ≤ 0.1 or Norepi ≤ 0.1*	Dop > 15 Epi > 0.1 Norepi > 0.1*
Central nervous system					
Glasgow Coma Scale Score	15	13-14	10-12	6-9	< 6
Renal					
Serum creatinine (mg/dL)	< 1.2	1.2-1.9	2.0-3.4	3.5-4.9	> 5
or urine output (mL/day)				or < 500	or < 200

Dob: dobutamine, Dop: dopamine, Epi: epinephrine, Norepi: norepinephrine, FIO$_2$: fraction of inspired oxygen, PaO$_2$: partial pressure of oxygen

⊤ values are with respiratory support

*Adrenergic agents administered for at least 1 hour (doses given are in μg/kg per minute)

Table 1.2 *Systemic Inflammatory Response Syndrome (SIRS) criteria from ref 1*

Variables	SIRS criteria
Temperature	>38°C or < 36°C
Heart rate	>90/min
Respiratory rate	>20/min or $PaCO_2$<32 mmHg
White blood cell count	>12 000/µL or <4000/µL or >10% immature neutrophils
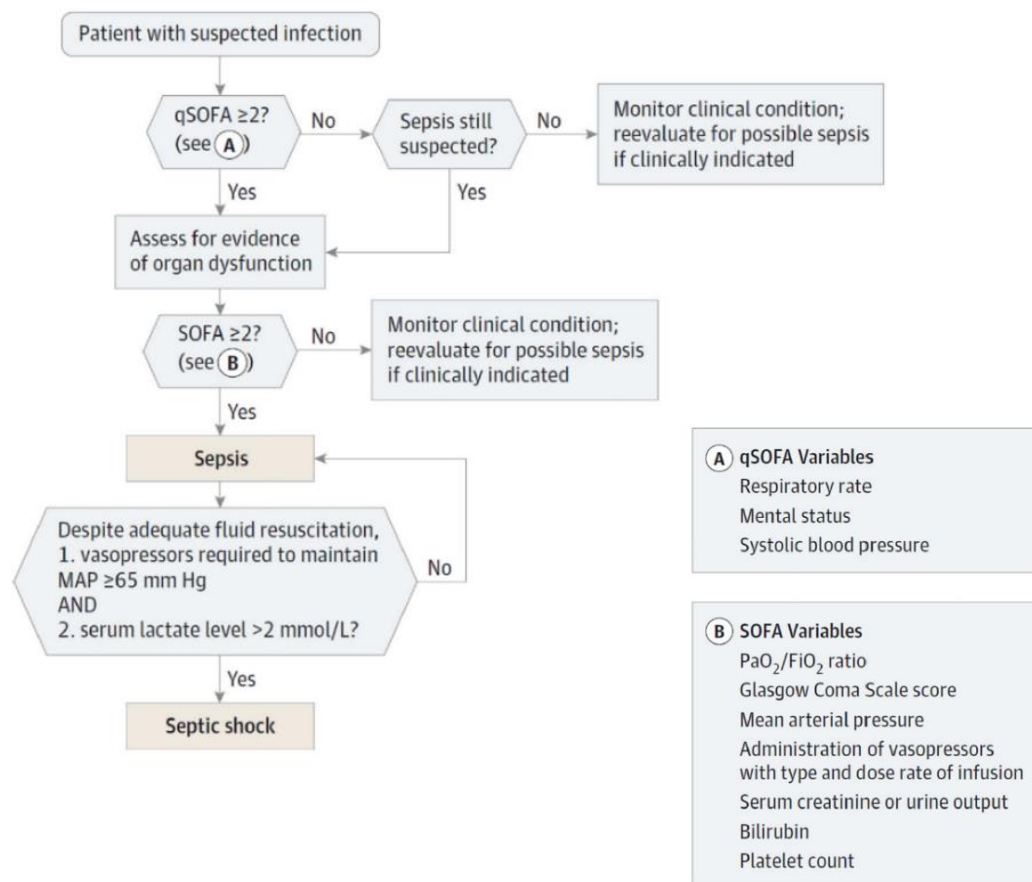
Figure 1.1 *Clinical criteria identifying patients with sepsis and septic shock reproduced from ref 1. SOFA: Sequential Organ Failure Assessment, qSOFA: quick SOFA, MAP: mean arterial pressure.*

A crucial point in the diagnosis of sepsis is the identification of the causative pathogenic agent(s). Despite recent technological advances, the principal method of identification is by culture of clinical samples that are time consuming (usually within 24 to 48h) and are positive in only 30 to 40% of cases. This is due to the fact that some microorganisms do not propagate in cultures or patients have received medication before obtaining clinical samples[20–23]. Therefore until the pathogen is identified, patients are treated via broad-spectrum antimicrobial administration to cover the most common pathogens (bacteria, fungi, virus) before narrowing the treatment. This is to reduce cost and prevent drug resistance[24]. Providing a treatment within an hour after observation of hypotension in septic shock patients is associated with 80% survival, this decreases with every hour of delay[25–27]. Administering the wrong therapy initially decreases by 5-fold the chance of survival[28]. Hence, the successful and rapid identification of sepsis is of paramount importance for successful treatment.

I.2.3 Pathogenesis

During sepsis, pathogen-associated molecular patterns (PAMPs), conserved microbial structures such as endotoxins (lipopolysaccharide), peptidoglycan, lipoteichoic acid, flagellin, mannan or viral ribonucleic acid (RNA), are recognised by the host immune cells through pattern-recognition receptors (PRRs)[29]. Pathogens of different natures are detected by 4 main PRR families, toll-like receptors (TLRs), C-type lectin receptors (CLRs), retinoic acid inducible gene 1-like receptors (RLRs) and nucleotide-binding oligomerisation domain-like receptors (NLRs)[3]. The activation of immune cells triggers 1) the overproduction of pro-inflammatory mediators, in particular cytokines, this phenomenon is termed "cytokine storm" and 2) the systemic activation of the complement system, part of the host innate defences against pathogens, which also promotes the inflammatory response[29–31]. Another consequence of this pro-inflammatory response is the activation of the coagulation pathways resulting in disseminated intravascular coagulation[31]. All these combined effects result in necrotic cell death and host tissue damage that release damage-associated molecular patterns (DAMPs), recognised by the same PRRs of PAMPs,

accentuating the inflammatory process. Moreover, a parallel anti-inflammatory response leading to leukocyte apoptosis (programmed cell death) contribute to immunosuppression and increased patient susceptibility to second infections[3]. A schematic representation of a host response during sepsis is presented in Figure 1.2.
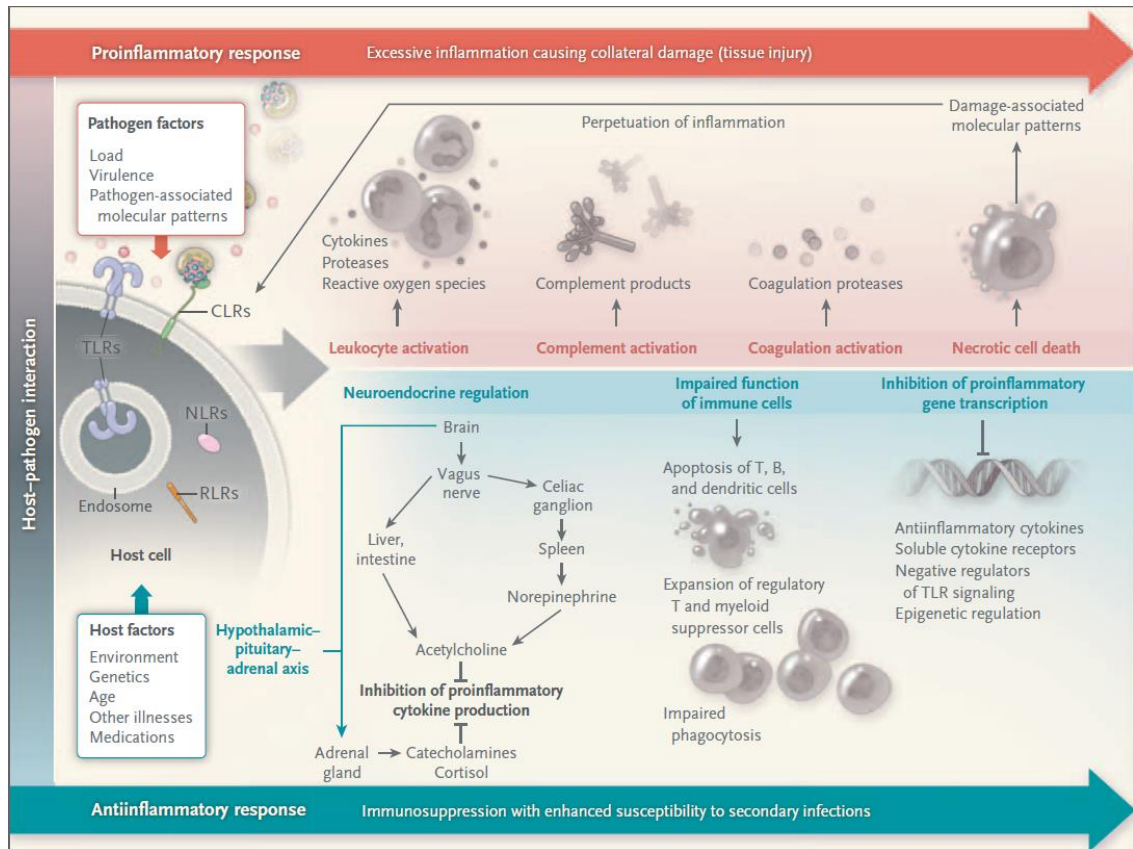
Figure 1.2 *Host response during sepsis, reproduced from ref 3.*

I.2.4 Biomarkers

Biomarkers for the diagnosis of infectious disease may include, genetic, protein and metabolomic materials derived from either the host or the pathogen. From the initial interaction onwards, the majority of biomarkers available to measure are derived from the host since pathogen numbers are very low and the host is able to utilise components of both the innate and adaptive host response to drive an appropriate response. In serious infection, when pathogens are able to overcome the early host response to their presence, their numbers increase at an exponential rate resulting in significant

mortality rates. In such cases, the relative concentration of microbial biomarkers increases over time whilst biomarkers associated with the ongoing, yet ineffective, host response are still readily detectable (Figure 1.3)[33].
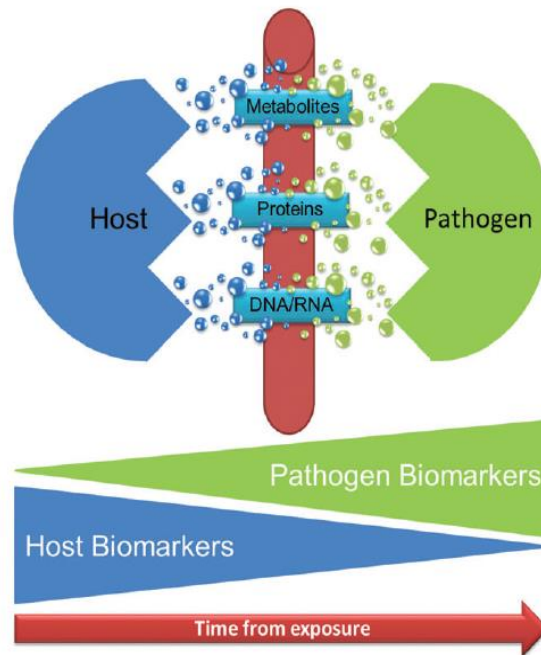
Figure 1.3 The relative contribution of host and microbial derived biomarkers to enable diagnosis of infection, reproduced from ref 33.

A review has reported investigations of over 170 potential host-derived biomarkers for sepsis diagnosis. However, not a single biomarker has been found reliable to diagnose sepsis with enough sensitivity and specificity and no clear biomarker has been identified to differentiate sepsis from sterile SIRS or local infection[34]. Procalcitonin (PCT) and C-reactive protein (CRP) are the most widely used biomarkers of sepsis despite a lack of specificity. Indeed, an increased level of PCT and CRP are observed in non-infectious inflammatory conditions such as trauma or major surgery. Similarly, cytokines such as interleukin-1 (IL-1), interleukin-6 (IL-6), interleukin-8 (IL-8), interleukin-10 (IL-10) or tumor necrosis factor (TNF) have been identified in several studies as potential sepsis biomarkers[34] whereas their levels also increase after trauma and surgery[35,36].

Over the years, many clinical trials have failed to demonstrate the efficacy of new therapeutics in improving sepsis patient survival[5,37]. Only one treatment has been

commercialised in 2001 (Xigris, Eli Lilly), based on the administration to patients with septic shock of recombinant human activated protein C (aPC) which has anticoagulant and anti-inflammatory actions[38]. However, the treatment was withdrawn from the market in 2011, as no significant difference in mortality rate decrease was obtained between septic shock patients supplemented with aPC or treated with a placebo[39].

Due to the complexity of the pathogenesis and the systemic nature of sepsis, it is clear that new strategies implying combination of biomarkers should be considered for the diagnosis, prognostic and elaboration of new treatments for sepsis[34]. For example, sepsis patients were differentiated from SIRS patients thanks to a combination of 4 genes (SeptiCyte[TM] Lab, Immunexpress, Inc.)[40] or with 11 genes (Sepsis MetaScore) in another study[41]. A publication has also demonstrated that when using host-derived biomarkers it is possible to distinguish patients with sterile, viral or bacterial acute respiratory inflammation[42]. A recent article have shown the ability to determine 4 sepsis subtype populations and with the identification of 2 genes for the characterisation of patients with the highest mortality rate risk[43]. Moreover, using 7 biomarkers, preliminary evidence has been produced which indicates that it is possible to identify the presence of an infectious organism through analysis of host biomarker signatures before patients become symptomatic[44]. Thus, the concept of searching for such signatures in host biofluids pre-symptomatically appears as a promising avenue for exploration in order to enable early therapeutic intervention[33].

Note: following sections I.3-I.5 are a modified version of a published article:

Developing and Understanding Biofluid Vibrational Spectroscopy: A Critical Review, M. J. Baker, S. R. Hussain*, L. Lovergne*, V. Untereiner, C. Hughes, R. A. Lukaszewski, G. Thiéfin, G. D. Sockalingum, *Chem. Soc. Rev.*, 2016, 45, 1803-1818. [*equal contribution]

I.3 Vibrational spectroscopy

Vibrational spectroscopy relates to the specific optical techniques of Infrared (IR) and Raman spectroscopy. These techniques probe intramolecular vibrations and rotations of the sample when irradiated with light[45]. The light-matter

relationship is underpinned by the electromagnetic theory postulated by Maxwell[46]. Vibrational spectroscopy has been used for analysing a myriad of samples in chemical, physical and biological applications. Constituent chemical molecular bonds present many forms of vibrations which occur at different energies corresponding to different allowed transitions. IR and Raman spectroscopies are complementary and provide a "fingerprint" or "signature" of the molecules contained within the sample depending on whether their bonds exhibit Raman or IR activities. Certain vibrations that are allowed in Raman may be forbidden in IR and *vice versa*.

Two categories of fundamental vibrational modes can occur, stretching vibrations (symmetrical or asymmetrical) with a variation in molecular bond length (Figure 1.4a) and bending vibrations (in- or out-of-plane) implying bond angle changes (Figure 1.4b). A diatomic molecule possesses only one vibrational mode, stretching vibrations whereas a polyatomic molecules exhibit more complex profiles. Indeed, an N-atomic molecule has 3N-6 fundamental vibration modes with one exception for linear molecules which have 3N-5 modes as no rotation around their axis is possible[47,48].
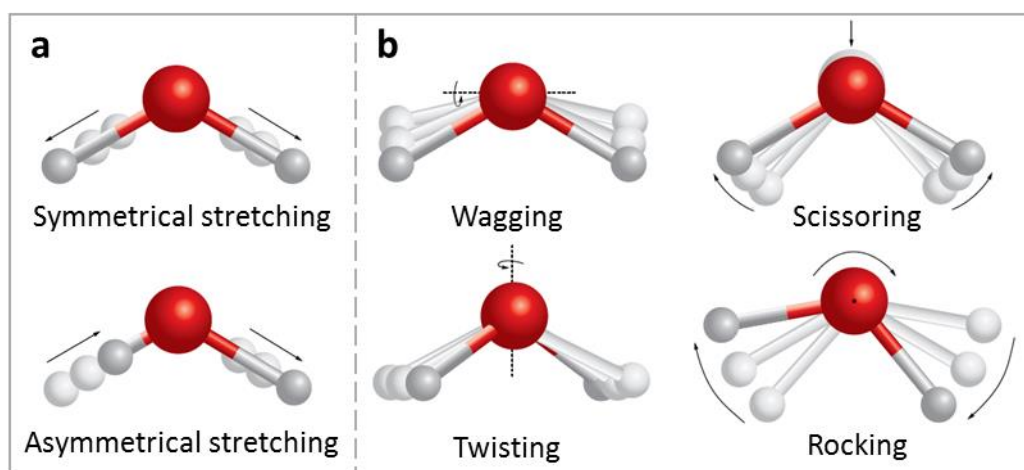
Figure 1.4 *Stretching (**a**) and bending (**b**) molecular vibrational modes. Adapted from ref 48.*

I.3.1 Infrared spectroscopy

Infrared spectroscopy (IR) is broadly defined as the study of absorption characteristics arising from the molecular motion of materials due to atomic displacement[49] upon intimate interaction with an infrared source[50]. Depending on the modality of choice, the radiation can be either transmitted, internally reflected, reflected, or transflected (a combination of transmission and reflectance). During the light-matter interaction, infrared light causes a molecule to enter in a higher vibrational state due to the transfer of "quanta" or "packets" of energy at certain wavelengths dependent upon the composition of the matter under analysis. Figure 1.5 illustrates the energy level transition involved in the IR absorption process compared to Raman scattering showing that vibrational energy levels can be probed with both techniques using different physical processes. These transitions result in a spectrum constituted of peaks/bands that can be interpreted qualitatively (peak position) and quantitatively (peak intensity/area, relative intensity). For IR spectroscopy the bands arise from a change in the electric dipole moment of the molecules. The absorbance of the bands is proportional to the concentration of the molecular bonds of a studied material as expressed by the Beer-Lambert law:

$$A = \log\left(\frac{I_0}{I}\right) = \varepsilon.c.l$$

I_0: incident light intensity

I: transmitted light intensity

ε: molar extinction coefficient (L.mol^{-1}.cm^{-1})

c: concentration of the absorbing material (mol.L^{-1})

l: path length of the light within the absorbing material (cm)

The IR region extends from 12500 cm^{-1} (800 nm) to 10 cm^{-1} (1 mm) and is divided in 3 domains, the near-IR (NIR) region from 12500 cm^{-1} (800 nm) to 4000 cm^{-1} (2.5 μm),

the mid-IR (MIR) region from 4000 cm^{-1} (2.5 µm) to 400 cm^{-1} (25 µm) and the far-IR (FIR) region from 400 cm^{-1} (25 µm) to 10 cm^{-1} (1 mm). The majority of organic compound vibrations are known to occur in the mid-IR region which is thus the preferential domain for biological sample investigations[47].
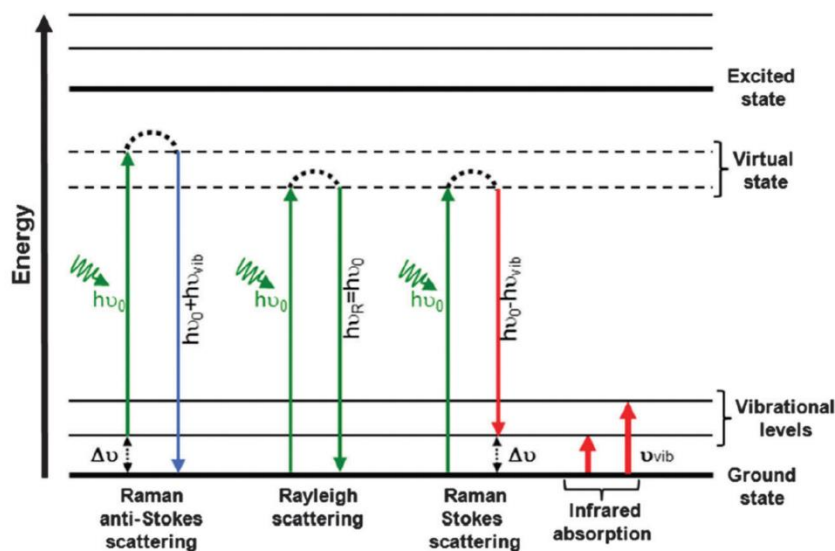
Figure 1.5 *Energy diagram showing transitions involved during infrared absorption, Rayleigh, Raman Stokes and anti-Stokes scattering. This Jablonski diagram shows that the same vibrational states of a given molecule can be probed* via *two different routes; one directly measures the absolute frequency (IR absorption) and the other measures the relative frequency or Raman shift (Stokes and anti-Stokes). hv_o= incident laser energy, hv_{vib} = vibrational energy, Δv = Raman shift, v_{vib} = vibrational frequencies[33].*

I.3.1.1 Instrumentation

A Fourier-Transform (FT) spectrometer is composed by 3 main components, an IR light source, a Michelson interferometer and a detector (Figure 1.6).

I.3.1.1.1 IR light sources

In IR spectroscopy, the most common source of polychromatic light in the MIR region is a Globar source based on thermal radiation resulting from a silicon carbide rod electrically heated. However, the launching of new IR imaging devices incorporating high-intensity tunable quantum cascade lasers (QCL)

could revolutionise the way clinical IR images are acquired[51]. High-throughput IR chemical imaging is now in its early days, and needs to be tested and validated. However, a gain of three orders of magnitude in acquisition time has recently been reported for large samples by Bhargava's group[52]. Combining signal enhancement from surface-enhanced infrared absorption (SEIRA) and fast imaging using a QCL source with small bandwidths, a recent study claimed a ~ 200 fold gain in imaging time[53].
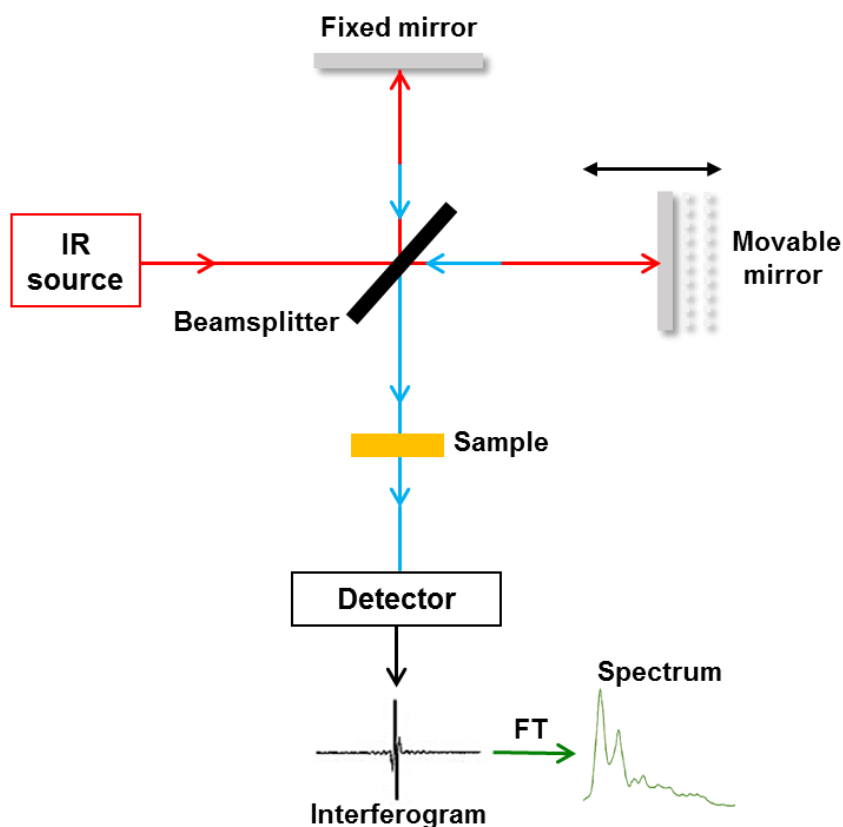
Figure 1.6 *Schematic of an FTIR spectrometer with a Michelson interferometer.*

Synchrotron radiation (SR) sources generate intense beams of infrared, X-rays or ultraviolet (UV) light. Electrons generated are accelerated to reach nearly the speed of light thanks to a linear accelerator (Linac), a booster synchrotron and a storage ring before to be directed towards experimental stations called beamlines (Figure 1.7). Regarding IR spectroscopy, SR source radiation is 100 to 1000 times brighter than conventional IR light sources[54,55].
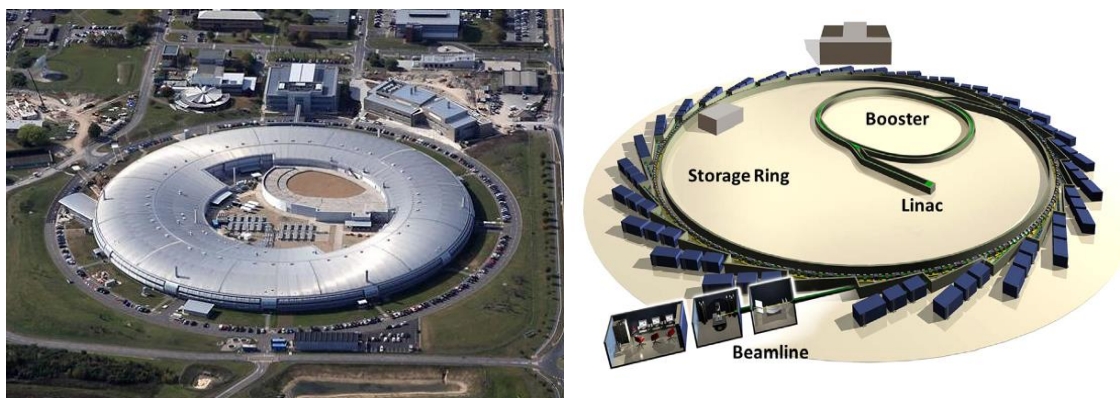
Figure 1.7 *Photograph (left) and Schematic (right) of Diamond Light Source science facility (UK) adapted from ref 55.*

I.3.1.1.2 Michelson interferometer

A Michelson interferometer comprises a beamsplitter, a fixed mirror and a movable mirror (Figure 1.6). The MIR beamsplitter, usually made of calcium fluoride (CaF2) or potassium bromide (KBr), allows to divide the IR light into 2 beams of equal intensity directed towards the 2 mirrors and reflected back to the beamsplitter to be recombined. The difference in the optical pathlength of the 2 beams, due to the displacement of the movable mirror and the fixed mirror respectively, creates destructive and constructive interferences. The resulting signal called an interferogram represents signal intensity as a function of distance of the movable mirror. Then, the recombined IR beam interacts with the sample (or reference) before reaching the detector. Two interferograms recorded with and without sample are mathematically converted by FT to IR spectra, represented by intensity as a function of frequency. Finally, the difference of the 2 spectra generates the representative IR spectrum of the sample[47].

I.3.1.1.3 Detectors

FTIR spectrometers are commonly equipped with Deuterated Triglycine Sulfate (DTGS) or Mercury Cadmium Telluride (MCT) detectors. The DTGS thermodetector has the advantages of operating at ambient temperature, has a good linearity and allows

spectral acquisitions down to 400 cm^{-1}. The MCT photodetector is more sensitive but requires to be cooled with liquid nitrogen and acquisitions in the low wavenumbers are generally limited around 750-600 cm^{-1}. Contrary to these single element detectors allowing a point by point acquisition, array detectors enable to collect simultaneously multiple spectra reducing considerably the acquisition time[47].

I.3.1.2 IR sampling modes

Transmission IR spectroscopy is a common approach for analysing thin samples (<10µm). In this mode, the IR beam is passed through the sample and is measured after its interaction with the matter. Thus, spectra collected are representative of the whole volume sampled. This method requires the use of IR transparent substrates such as CaF_2 or KBr[56].

Attenuated total reflectance (ATR) FTIR spectroscopy is a promising modality for biological sample analysis. The guided IR beam propagates through a high refractive index crystal surface producing an evanescent standing wave that penetrates the sample by a few microns. However for proper use, several issues need to be considered, such as contact between the ATR crystal and the sample, the beam penetration depth and image distortion due to high refractive indices[57,58].

Another method of obtaining an IR spectrum is when the sample is placed on a highly reflecting surface, typically aluminium/teflon coated substrates or a glass slide with tin oxide-based silver reflective coating called low e-slides (e.g., MirrIR). In this case the process is termed transflection because the IR beam passes through the sample, is reflected off the slide and passes again through the sample before detection. These substrates have very low cost but recently they have been shown to cause significant spectral intensity variations, due to an electric field standing wave artefact (EFSW)[59,60] which could be misinterpreted as composition variations while it is the sample thickness variation that could be the cause of spectral differences. The fundamental question when using low e-slides is whether the spectral variations observed are due to the EFSW impact on the discriminant spectral differences. In the case of thin

samples such as air-dried cellular monolayers, recent research by Cao *et al.* has shown that the same classification was obtained when performing transmission and transflection measurements[61]. Moreover, a recent paper have questioned the EFSW effect existence and demonstrated that spectral differences observed result from reflections and interferences at the air-sample interface[62]. Principles of the different IR sampling modes are illustrated in Figure 1.8.
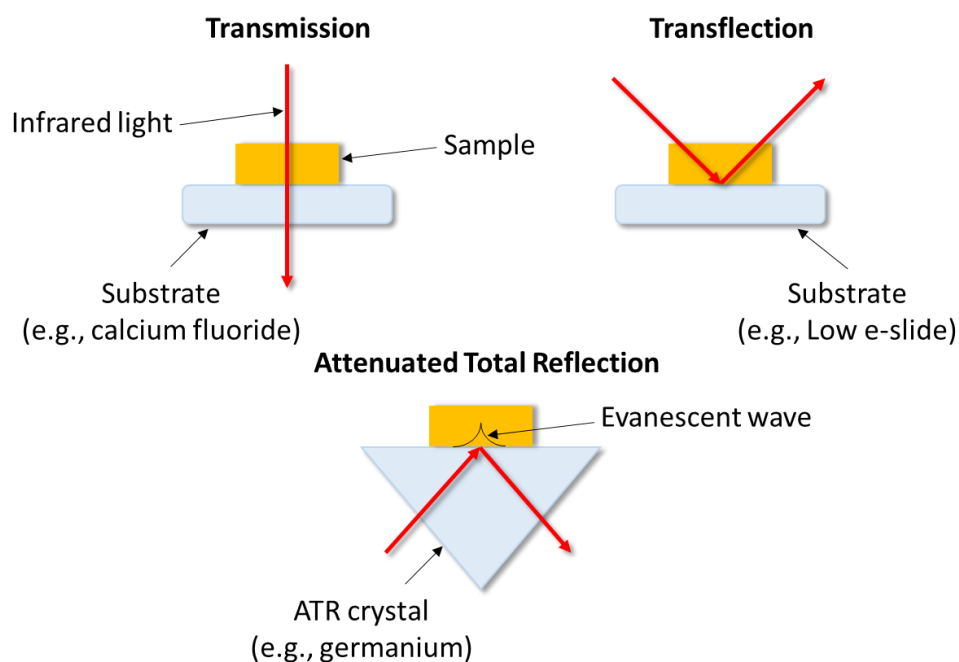
Figure 1.8 *Example of different IR sampling modes, adapted from ref 56.*

I.3.2 Raman spectroscopy

The Raman effect constitutes the spontaneous inelastic light scattering process of photons following the interaction of monochromatic radiation (e.g., laser source) with the sample. During this interaction both elastic and inelastic scattering processes take place. A high proportion of the photons are elastically scattered with no change in energy (so no molecular information), known as Rayleigh scattering[46]. When photons transfer energy to the molecules as vibrational energy, the energy loss of the scattered photons corresponds to the vibrational energy levels of the molecules. This is known as Raman-Stokes scattering. The incident photons can in turn receive energy from vibrating

molecules, and therefore their frequencies increase, described as Raman anti-Stokes scattering. Figure 1.5 shows the transitions involved during these three processes. In spontaneous Raman, Stokes scattering is generally used due to its higher sensitivity.

I.3.2.1 Instrumentation

A Raman spectrometer is constituted of a monochromatic light source (laser) with different possible choices of wavelength frequencies. Lower excitation wavelengths allow to obtain a better Raman diffusion intensity but are also more prone to fluorescence interferences. The light is oriented towards the sample thanks to different mirrors, its intensity can be attenuated with filters to avoid photodamage. Notch or edge filters are necessary to eliminate Rayleigh diffusion and to only transmit the laser wavelength. The filtered light is then focused on the sample through a microscope objective. The scattered light resulting from the sample excitation is collected via an objective and passes through the Notch filter again. The pinhole allows to spatially select a specific volume sample signal. The slit is defining the spectral resolution, a reduction of the aperture gives a better spectral resolution but also involves that the intensity of the light is reduced. A diffraction grating splits the wavenumbers of the light scattered by the sample. This system is characterised by a number of lines per millimetre, a higher number of lines/mm improves the spectral resolution to the detriment of the spectral range. Then, the intensities of the splitted wanumbers are simultaneously collected by a Charged-Coupled Device (CCD) detector, highly sensitive to detect the weak scattering light and allowing to obtain a Raman spectrum[63]. All constituting components of a Raman spectrometer are illustrated in the Figure 1.9.
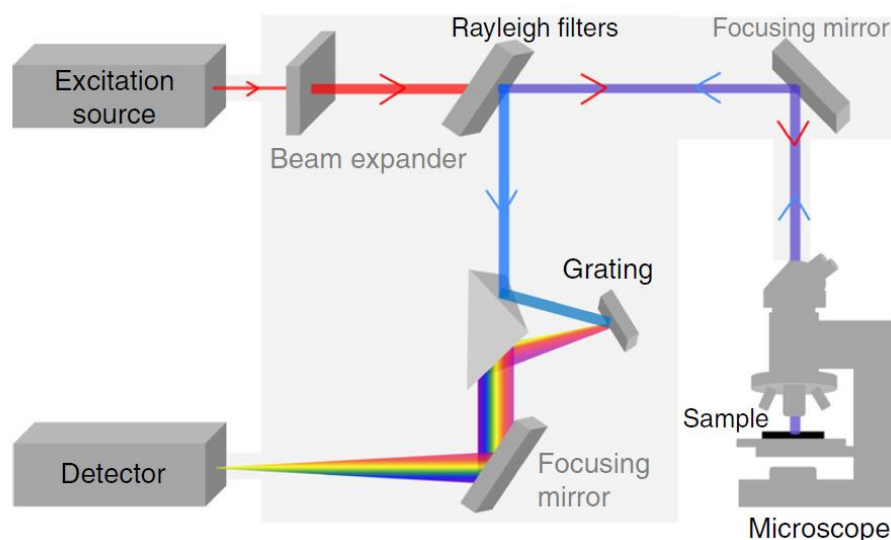
Figure 1.9 *Schematic of a Raman spectrometer reproduced from ref* 63.

I.3.2.2 Raman sampling methods

The Raman shift covers the range between 0 and 4000 cm^{-1}. Raman spectroscopy can be used in a confocal mode and with resonance and surface-enhanced modalities. Applications of Raman microspectroscopy for probing biological systems have been continuously expanding over the years along with IR spectroscopy[64]. Its high spatial resolution (~0.5 μm with green lasers), compatibility with aqueous environment[65,66] and *in vivo* amenability[67–70] makes it a good candidate for biological and biomedical research. Akin to FTIR, it also provides high content biomolecular information. Microspectroscopy with immersion measurements can be used to enhance signal to background ratio enabling higher quality data acquisition as demonstrated by Bonnier *et al.*[71].

Due to its advantages, label-free, high spectral specificity, limited water signal, and the fact that most biological molecules are Raman active, Raman has been deployed to *in vitro* cell and tissue studies, but now significant developments of *in vivo* work due to the compatibility with fibres, has enabled Raman endoscopy in a label-free manner and *in vivo* Raman probes/endoscopes have made direct tissue analysis possible[72].

New fields of measurement and implementation possibilities have multiplied due to recent hardware developments, improved sampling methods, and advances in the design of Raman technology alongside developments and advances in multivariate data analysis. It has been possible to uncover subtle disease-related spectral changes and exploit them in classification models. However, an important drawback of Raman spectroscopy is that the effect is inherently weak as a very small proportion of incident photons are scattered (~1 in 10^8) with a corresponding change in frequency[46]. This together with the fact that to date most of the commercial systems use dispersive configurations adds another limitation compared to fast IR imaging systems, and makes Raman imaging of biological specimens a slower process. These limitations can be partly circumvented with other Raman modalities based on Resonant Raman Scattering (RRS) and Surface-Enhanced Raman Scattering (SERS) to enable gains in detection sensitivity[64]. In SERS technology, the use of functionalised metal nanosurfaces has allowed optimising the enhancement to several orders of magnitude depending on the metal substrate. Metal nanoparticle arrays and single nanoparticles have been utilised for high-throughput detection[73]. SERS has been applied in different areas in the chemical and biological fields[74] and its very high sensitivity has allowed single molecule detection[75]. Until recently, SERS was not widely applied to biomedical research because of issues linked to complexity of the biological medium, biocompatibility, reproducibility, and short shelf life. However, using silver and gold colloids as SERS substrates, Bonifacio *et al.* recently showed that repeatable spectra could be obtained from protein-free blood serum and plasma[76].

Furthermore, non-linear Raman spectroscopy has been developed to be applied to biomedical analysis like Stimulated Raman Scattering (SRS) and Coherent Anti-Stokes Raman Scattering (CARS), for rapid image acquisition (one Raman band at a time) with higher sensitivities than spontaneous Raman[77–79]. For non-linear Raman, it is important to know which marker band(s) are useful, in analogy to the application of Discrete Frequency-IR (DF-IR) as enabled by the use of QCL sources.

Other areas of current interest for Raman spectroscopy are exploring the sampling depth and location of spectral information. For instance, seminal research conducted by Stone, Matousek and collaborators, demonstrated the principle of spatially offset

Raman spectroscopy (SORS) for subsurface analysis towards *in vivo* breast cancer[80,81] and deep Raman measurements using liquid tissue phantoms to mimic non-invasive cancer screening applications *in vivo*[82]. Through-tissue sensitivity was increased *via* SESORS measurements at several millimetre depth, i.e., combining SORS with nano-tagged SERS particles[83–86].

I.4 Multivariate analysis

It is becoming more and more evident that vibrational spectroscopy represents an interesting approach to explore the diagnostic potential of circulating biomarkers/biosignatures in various body fluids[87]. Along with technological development, the front-end sample preparation challenges and approaches, and the data acquisition procedures, the pre-processing and post-processing of spectral data are equally important for the deployment of various biofluids into diagnostic development. Vibrational spectroscopic data are inherently multivariate by nature and their pre- and post-processing require multivariate data analysis approaches.

Different instruments from different manufacturers have different responses and spectral distortions and backgrounds have to be taken into account via pre-processing algorithms in order to compare data from different studies for example. The pre-processing should therefore be able to give accurate, robust and reliable data. These considerations should also include how the sample is prepared and conditioned, the optical substrate used, and the acquisition mode used in order to post-process reliable data. The way the sample is dried or acquired (e.g. transmission or reflection) will also pre-empt the pre-processing procedures. For example, rapid drying of serum can produce a granulating effect which then causes more scattering/dispersion artefacts and a specific correction has to be implemented. It is clear that the pre-processing is not the same for infrared and Raman spectra of biofluids because the physical phenomena involved are respectively absorption and scattering. In FTIR spectroscopy the use of an interferometer ensures an excellent intensity and wavenumber

calibration. In addition, a background signal is regularly recorded and automatically subtracted to obtain the sample spectrum. For Raman, a day-to-day calibration procedure needs to be implemented to correct for instrument response, and to calibrate the wavenumber and intensity axes. Other experimental considerations include the need to subtract substrate contributions and other physical phenomena such as fluorescence and heating. Biofluid vibrational spectra are therefore corrected, derived (or not), then normalized. As a general rule, it is also important to include prior to the pre-processing steps, a quality test to remove spectra with a poor signal/noise ratio (threshold to be defined depending on the sample nature) and a validated outlier removal routine before post-processing.

The post-processing step includes data mining and the construction of classifiers. Very often, the spectral differences between normal and pathological states are very subtle and the next step is to perform data mining, i.e., a process used to extract the salient information from the spectral data. By using specific algorithms, patterns can be found in large batches of data. Thus, such feature selection procedures can help to identify discriminant spectral features to discriminate between patient groups[88]. However, it is important to note that data mining depends on effective data collection, the size of the datasets, and as well as their pre-processing.

To build classification models, several multivariate approaches have been used and as of today there is no general consensus on which method is the best. In other research fields, numerous linear and non-linear supervised algorithms have been evaluated and a combination of methods like Support Vector Machine (SVM) and Partial Least Square-Discriminant Analysis (PLS-DA) has been shown to enhance the sensitivity and specificity of the classifiers[89]. Generally, building the classifier should include a calibration phase (training phase), an internal validation phase, and an external validation phase (blind testing phase). One of the important issues encountered is the size of the datasets used as a small dataset that does not accurately describe the patient population can lead to under- or over-fitting and impact the classifier outcome. For a classifier to be

robust, it is important to have a large number of class-representative patient samples. In addition, the external validation requires a dataset that has not been used in the two previous steps of calibration and internal validation (based upon patient spectra and not replicate spectra from the same patient i.e. a spectrum from the same patient should not be in the calibration/internal validation and external validation phases). The leave-n-out cross validation method is often used for these models. It is important to note that all spectra from a given patient must be removed in this process in order to enable a valid outcome. Considering all individual spectra, mean spectra or median spectra as input datasets of the classifier should also be taken into consideration although it has been found that when spectra are highly reproducible and after applying a quality control test plus an appropriate outlier removal, the results are comparable[90]. The workflow in Figure 1.10 illustrates the different steps, for both IR and Raman spectroscopies, starting from sample preparation to data pre- and post-processing and the building of classifiers for diagnostics.
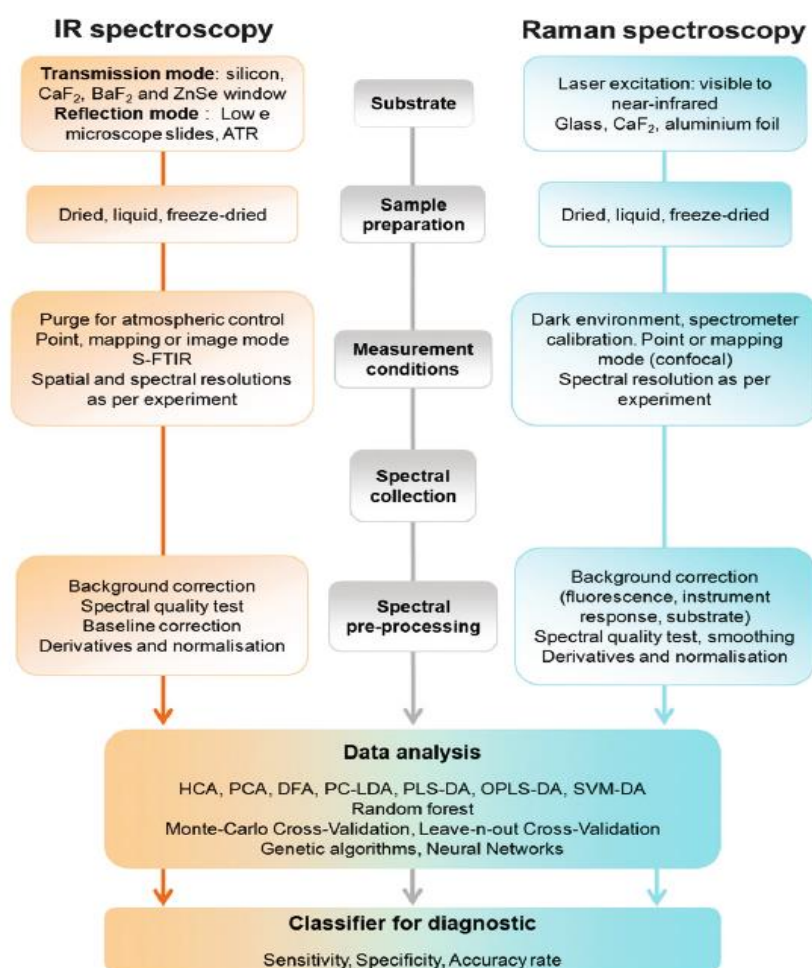
Figure 1.10 *Workflow of biofluid spectroscopy from substrate choice through sample preparation to spectral measurements and data analysis with diagnostic classifiers[33].*

I.5 Biofluid spectroscopy

Regarding biofluids, blood and its constituents appear the most convenient for biomarker/biosignature detection given its ease of availability and the possibility to repeat the test as often as necessary to monitor disease progression or response to treatment.

Blood serum houses more than 10,000 different proteins. It perfuses all body organs meaning it contains a large range of proteomes from surrounding tissues and cells, making it the most complex biofluid[91]. The low molecular weight fraction serum component of blood, known as the "peptidome" is information rich for diagnostic purposes[92]. Other biofluids (bile, urine, sputum, pancreatic juice, and ascitic, pleural, cerebrospinal fluids), in direct contact with the diseased tissue, are of great interest as media to detect biomarkers/biosignatures that are secreted or shedded locally. These are expected to be present in higher concentration in these fluids than in the blood. In addition, their identification may be facilitated by a less complex molecular composition of local biofluids compared with blood. Although some biofluids such as urine share with blood samples ease of availability and repeatability, analysis of other biofluids requires an invasive procedure, which limits their repeated use in the clinical setting. An example is cerebrospinal fluid which requires a lumbar puncture for collection.

Whilst vibrational spectroscopy has been developed initially mainly for cell and tissue analysis, it has been also applied more recently to biofluids for biomarker discovery, generating a number of pilot studies with promising results as presented below. The challenge is now to translate the results of these exploratory studies to the routine clinical practice.

The search for disease markers in biofluids *via* photonic approaches is a fast emerging field and has only been recently explored by vibrational spectroscopic approaches. Biofluids are easily accessible and minimally invasive for patients making large studies feasible. Like cells and tissues, biofluids exhibit vibrational spectra that have characteristic bands reflecting their biomolecular composition. Figure 1.11 compares the FTIR spectra of some dried biofluids (serum, plasma,

and bile) obtained with a high-throughput module in the transmission mode. IR spectra of serum and plasma present very close profiles with subtle differences that are difficult to depict visibly. This is explained by the fact that serum is essentially plasma with the clotting factors of blood removed. The assignment of the main bands is provided in Table 1.3.
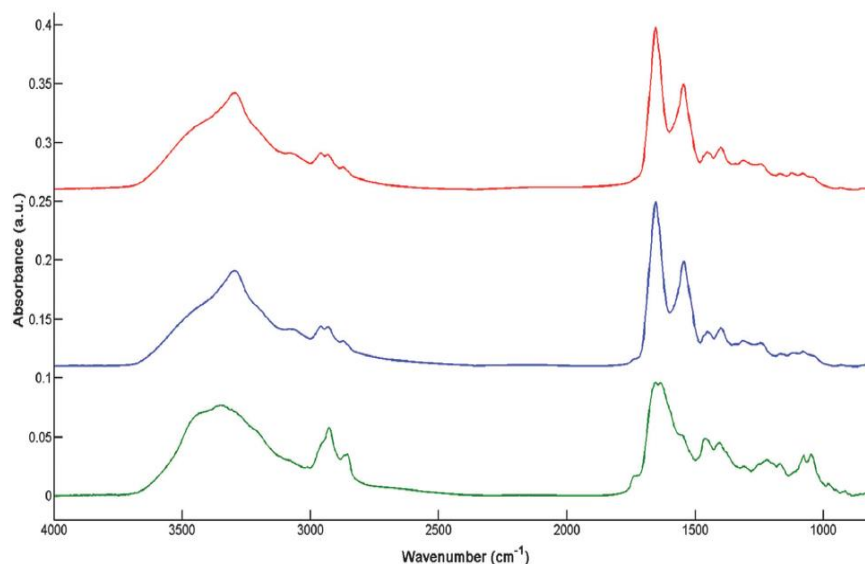
Figure 1.11 *Comparison between HT-FTIR spectra of different biofluids serum (red curve), plasma (blue curve), and bile (green curve). Spectra are background corrected and normalised. Note: Serum and bile were collected in dry tubes while for plasma samples lithium heparin tubes were used[33].*

Table 1.3 *Assignment of the major absorption bands of a plasma FT-IR spectrum*[93]

Bands (cm^{-1})	Major assignments for plasma contents
3300	v(N-H) of proteins (amide A band)
3055-3090	v(=CH) of lipids
2950-2960	v_{as}(CH$_3$) of lipids
2920-2930	v_{as}(CH$_2$) of lipids
2865-2880	v_s(CH$_3$) of lipids
2840-2860	v_s(CH$_2$) of lipids
1730-1760	v(C=O) of fatty acids
1660	v(C=O) of proteins (amide I band)
1550	δ(N-H) of proteins (amide II band)
1400	v(COO$^-$) of animo acids
1240	v_{as}(P=O) of nucleic acids
1170-1120	v(C-O) and v(C-O-C) of carbohydrates

v: stretching vibrations, δ: bending vibrations , s: symmetric, as: asymmetric. Taken from Lacombe *et al.*, Analyst, 2015, **140**, 2280.

The bile spectrum differs through a higher lipid and carbohydrate content and by relative intensity changes of the protein amide I/amide II bands[90].

Raman spectroscopy gives complementary information to IR. Besides the main macromolecules like proteins, lipids, and carbohydrates, other modes originating from amino acids for example are active. The assignment of the main bands is indicated in Figure 1.12 showing an example of a typical Raman serum spectrum taken from a dried drop.
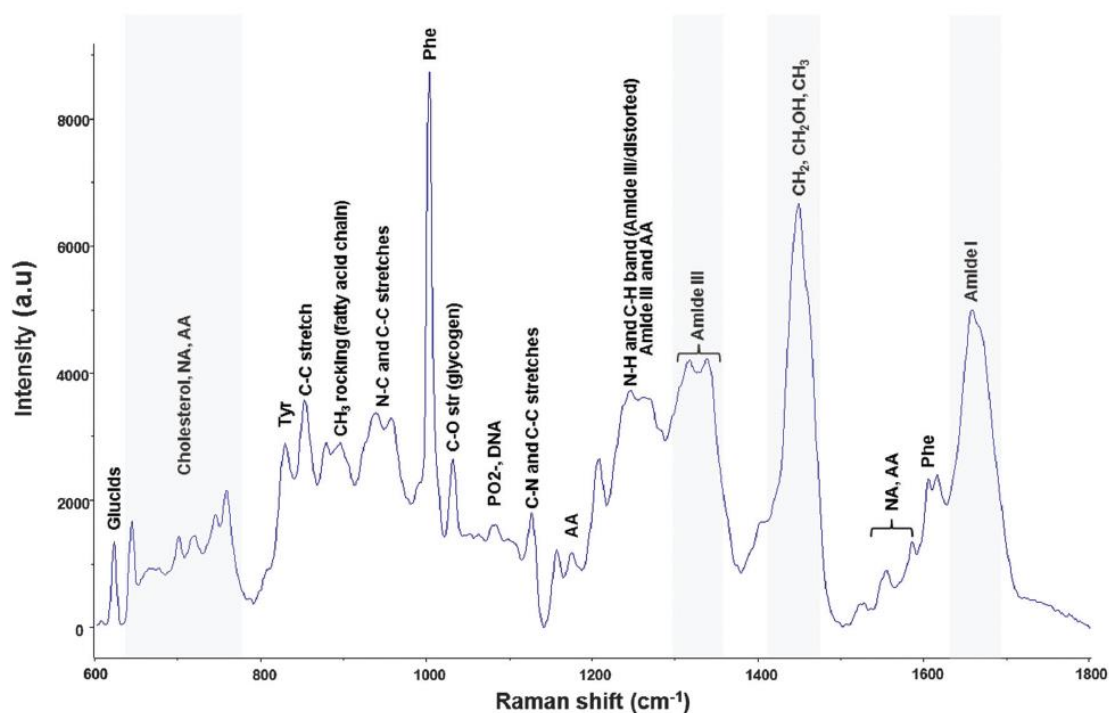
Figure 1.12 *Typical Raman spectrum of dried serum drop with spectral assignments. Spectrum was measured on a calcium fluoride window with a 785 nm laser excitation with an acquisition time of 2x30 seconds[33].*

I.5.1 Serum and plasma

At present, the majority of the biofluid spectroscopy research has focused on serum and plasma. This is most likely due to the prevalence of these types of samples within current biobank stocks or the fact that ethics are already established to collect these samples and all that is required is an addendum stating a separate use of the material.

I.5.1.1 Malignant diseases

Currently, in the field of oncology, most investigations are proof-of-principle studies showing the potentials of FTIR/Raman spectroscopy to identify different types of cancer from serum samples with high degrees of accuracy. HT-FTIR spectroscopy in transmission mode was used to discriminate urinary bladder cancer patients from patients with urinary tract infection with linear discriminant analysis (LDA) or random forest (RF) classifiers[94]. Using blood serum, Backhaus *et al.* distinguished between breast cancer and controls with a very high sensitivity and specificity[95]. Chemometrics combining support vector machine (SVM) and leave-one-out cross validation was employed by Zhang *et al.* to separate cirrhotic patients with or without hepatocellular carcinoma[88]. Equally important is the possibility to identify liver fibrosis stages prior to the development of hepatocellular carcinoma, which are crucial for the clinical management. A study by Scaglia *et al.* revealed that patients with extensive fibrosis (F3/F4 stages) could be distinguished from those with no fibrosis (F0 stage) on the basis of their FTIR serum spectra using a combination of discriminant wavenumbers[96]. Studies using ATR-FTIR spectroscopy, coupled with classification machine discriminated ovarian[97] and endometrial cancers[98]. It also allowed differentiating glioblastoma multiforme (GBM) from healthy controls and low grade gliomas and GBM *versus* healthy controls[99,100].

Applications of Raman spectroscopy to the study of various biofluids from cancer patients are in continuous progress. Sahu *et al.* analysed serum samples and could differentiate oral cancer patients from controls[101]. More recently, they reported that Raman serum spectroscopy was capable to predict the probability of recurrence in this cancer[102]. Other studies have shown the potential of Raman spectroscopy for differentiating normal subjects from patients with breast[103], colorectal[104], or cervical[103] cancers. A proof-of-concept study using micro-Raman spectroscopy applied to the sera of 71 cirrhotic patients showed that it could be an alternative method for discriminating cirrhotic patients with and without hepatocellular carcinoma[105]. On the other hand, SERS of serum or plasma

has also been shown as a promising tool for the diagnosis of various types of cancer such as nasopharyngeal[106–108], digestive[107,109–111], and prostate cancers[112].

I.5.1.2 Non-malignant diseases

Serum and plasma have been also employed to diagnose other diseases using biospectroscopy. For example, Raman serum data allowed to differentiate Alzheimer's disease from other dementia[113] and Carmona *et al.* used plasma Raman spectral data to grade mild, moderate, and severe Alzheimer cases[114]. *Via* FTIR spectroscopy of plasma, Peuchant *et al.* have shown that patients with Alzheimer's disease could be well delineated from normal ageing subjects used as controls[115].

Recent plasma data published by Lacombe *et al.* clearly showed that HT-FTIR spectroscopy could be an interesting alternative technique in neonatal screening of rare diseases such as classic galactosemia. Promising results indicated that healthy/diabetic, healthy/galactosemic, and diabetic/galactosemic patients could be discriminated with good sensitivity and specificity[93].

Few large studies have been reported. An example is the study led by Petrich's group showing the potential of mid-infrared spectroscopy in the triage of patients with acute chest pain[116]. This study included 1,429 serum samples from 389 patients reporting to two US hospitals (Massachusetts General and Latter Day Saints, Utah) consisted of 104 suffering from acute myocardial infarction (AMI), 136 from unstable angina pectoris, and 149 from chest pain of other sources. FTIR measurements were performed in the transflection mode. Using a threshold value generated from a robust linear discriminant analysis, they achieved high sensitivity and specificity enabling triage of patients with AMI, those most at need within the Accident and Emergency setting, compared to the other sources of chest pain. They hypothesise on the involvement of carbohydrates as discriminant features, possibly a glycation reaction. Interestingly, their results were comparable to the performance of routine cardiac laboratory markers within the same study population. They conclude on

the potential of FTIR to aid the diagnostic procedure as early as within the first 6 hours after the onset of chest pain.

Blood plasma from patients has been investigated with Raman spectroscopy as dried drops to identify a reliable biomarker that can differentiate sepsis patients from those with non-infectious systemic inflammatory response syndrome. Neugebauer *et al.* reported on the high sensitivity and specificity that can be achieved[117]. The possibility of separating the two groups of patients is crucial because a stratification of at risk patients can be established for a rapid delivery of appropriate treatment. Finally, following the results obtained in a model of infected cultured cells, SERS appears as a promising approach for malaria parasite detection from whole blood[118].

I.5.2 Other biofluids

Other biofluids non-invasively accessible (urine, saliva, sputum, tears) and invasively accessible (bile, synovial fluid, cerebrospinal fluid, amniotic fluid) have been investigated by vibrational spectroscopy for diagnostic purposes. Somorjai *et al.* were able to distinguish urine samples from normal renal transplants and rejected allografts, applying IR spectroscopy and a three-stage classification strategy[119]. A Raman spectroscopic analysis combined with Principal Component Analysis (PCA) and Quadratic Discriminant Analysis (QDA) performed on urine, has allowed identification of spectral biomarkers predictive of complications and kidney failure in the urine of diabetic and hypertensive patients[120]. Finally, in the field of oncology, Del Mistro *et al.* reported that SERS using Au nanoparticle substrates had the potential to detect in urine spectral biomarkers of prostate cancer[121].

Another approach by FTIR spectroscopy associated with LDA on saliva, has reported the correct classification of diabetic patients from healthy control[122]. SERS of saliva showed the ability to predict lung cancer by monitoring the decrease of proteins and nucleic acids with 80%, 78%, and 83% accuracy, sensitivity, and specificity respectively[123]. A preliminary study using SERS on saliva suggested the possibility of

a quick detection of acquired immune deficiency syndrom (AIDS) but these results obtained on a small number of patients deserve to be confirmed on a larger population[124].

An exploratory study has shown that FTIR spectroscopy applied to sputum could be a useful approach for the diagnostic of the chronic obstructive pulmonary disease[125].

Investigating the potential of human tears for the diagnosis of ocular diseases, Travo *et al.* have shown the discrimination of patients with keratoconus (degenerative disorder affecting the cornea) from healthy control and also between patients at an early or advanced stage of disease by HT-FTIR and PCA[126]. Additionally, Choi *et al.* report that SERS can be used for diagnosis of adenoviral conjunctivitis from tears[127].

Using HT-FTIR spectroscopy in association with support vector machine (SVM) classification and leave-one-out cross validation (LOOCV), Untereiner *et al.* have shown that bile samples of patients with malignant biliary strictures were differentiated from those with benign biliary diseases[90].

Eysel *et al.* using FTIR spectroscopy and LDA with LOOCV on synovial fluid, were able to differentiate samples from joints affected by rheumatoid arthritis, osteoarthritis, spondyloarthropathies, and meniscal injuries[128]. Also from synovial fluid samples, a Raman spectroscopic study associated to a k-means analysis has shown discrimination between patients with osteoarthritis of low or high severity[129].

Liu *et al.* have investigated the amniotic fluid potential for fetal lung development assessments by IR spectroscopy. The lecithin/sphingomyelin (lung surfactants) and lung surfactant/albumin ratio measurements by IR spectroscopy were quantitatively and qualitatively correlated to those obtained by thin-layer chromatography and fluorescence depolarization, two clinical methods used to determine fetal lung surfactant maturity in amniotic fluid[130]. Prenatal disorders from amniotic fluids have also been investigated by ATR-FTIR spectroscopy revealing spectral profile changes

between amniotic fluids from pregnancies with fetal malformations, preterm delivery and healthy term pregnancies[131].

Griebe *et al.* were able by FTIR spectroscopy to distinguish patients with Alzheimer's disease from healthy controls using cerebrospinal fluid[132].

I.5.3 Translation

With a few exceptions, all the mentioned proof-of-concept studies have been carried out on rather small populations and have shown promises for clinical utility and highlight the potential of vibrational spectroscopy for spectral diagnostics. To our knowledge, two major programmes for large scale clinical trials in remote settings are ongoing using hand-held FTIR modalities. The first campaign led by Wood *et al.* concerns the screening of population in Thailand for malarial diagnosis (http://monash.edu/news/show/infrared-light-puts-malaria-to-the-test).

A similar approach is being taken in the UK with the establishment of Glyconics Ltd. Glyconics is using sputum to diagnose Chronic Obstructive Pulmonary Disorder and are moving towards clinical validation of handheld ATR-FTIR on a subset of the UK population (http://www.glyconics.com/technology.asp).

These steps towards actual clinical environment testing is pushing the field to the forefront of the application and will illuminate the utility of these techniques as well as barriers to clinical implementation that need to be overcome.

I.6 Aim

The principal objective of this research is to develop and evaluate the potential of vibrational spectroscopy applied on human serum for the rapid pre-symptomatic diagnosis of sepsis.

The first step to achieve this goal is to determine the most suitable methodology approach for serum analysis by spectroscopy with an overall clinical application purpose. IR spectroscopy is sensitive to water, as a consequence liquid samples are commonly dried for analyses. However, a dried serum drop exhibits an inhomogeneous profile that could affect spectral data collected and introduce potential bias in their interpretation. Challenges of serum spectroscopy inherent to the sample nature and to the technique will be evaluated. Different IR and Raman modalities as well as different serum sample preparations will be compared.

Then, further experimentations will be conducted with the chosen method. Some aspects of the pre-analytical phase will be addressed in order to standardise protocols in sample handling and preparation for spectral acquisitions. Parameters that can impact either the samples (e.g., solvents, collection tubes, freeze-thaw cycles, long-term storage, drying methods) and/or the spectroscopic analysis (e.g., deposit methods, volumes, substrates, humidity, operators dependency) and consequently on the quality and the reproducibility of spectral data will be investigated.

Finally, based upon the developed methodology, patient serum samples (n=380) collected before surgery, up to 3 days before sepsis diagnosis, and on the day of sepsis diagnosis will be analysed. Control serum samples (n=353) from age/ sex/ procedure-matched patients who did not go on to develop sepsis will be also analysed over similar timeframes post-surgery as well as SIRS patient samples (n=190). Spectral data acquired will be interrogated by chemometric methods to discriminate spectral zones reflecting differences in molecular composition and to establish robust classifications.

I.7 References

1   M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. Vincent and D. C. Angus, *JAMA*, 2016, **315**, 801–810.

2   M. Shankar-Hari, G. S. Phillips, M. L. Levy, W. Christopher, V. X. Liu, C. S. Deutschman and D. C. Angus, *JAMA*, 2016, **315**, 775–787.

3   Angus, *N. Engl. J. Med.*, 2013, **369**, 840–851.

4   C. Fleischmann, A. Scherag, N. K. J. Adhikari, C. S. Hartog, T. Tsaganos, P. Schlattmann, D. C. Angus and K. Reinhart, *Am. J. Respir. Crit. Care Med.*, 2016, **193**, 259–272.

5   K. N. Iskander, M. F. Osuchowski, D. J. Stearns-Kurosawa, S. Kurosawa, D. Stepien, C. Valentine and D. G. Remick, *Physiol. Rev.*, 2013, **93**, 1247–1288.

6   N. Hex, J. Retzler, C. Bartlett and M. Arber, *YHEC*, 2017, 1–46.

7   T. J. Iwashyna, W. E. Ely, D. M. Smith and K. M. Langa, *JAMA*, 2010, **304**, 1787–1794.

8   T. J. Iwashyna, C. R. Cooke, H. Wunsch and J. M. Kahn, *J. Am. Geriatr. Soc.*, 2012, **60**, 1070–1077.

9   P. P. Pandharipande, T. D. Girard, J. C. Jackson, A. Morandi, J. L. Thompson, B. T. Pun, N. E. Brummel, C. G. Hughes, E. E. Vasilevskis, A. K. Shintani, K. G. Moons, S. K. Geevarghese, A. Canonico, R. O. Hopkins, G. R. Bernard, R. S. Dittus and E. W. Ely, *N. Engl. J. Med.*, 2013, **369**, 1306–1316.

10   M. Ratzer, O. Brink, L. Knudsen and A. Elklit, *Heal. Psychol. Behav. Med.*, 2014, **2**, 882–898.

11   R. C. Bone, R. A. Balk, F. B. Cerra, P. R. Dellinger, A. M. Fein, W. A. Knaus, R. M. H. Schein and W. J. Sibbald, *Chest*, 1992, **101**, 1644–1655.

12   M. M. Levy, M. P. Fink, J. C. Marshall, E. Abraham, D. Angus, D. Cook, J. Cohen, S. M. Opal, J.-L. Vincent and G. Ramsay, *Intensive Care Med.*, 2003, **29**, 530–538.

13   J.-L. Vincent, S. M. Opal, J. C. Marshall and K. J. Tracey, *Lancet*, 2013, **381**, 774–775.

14   A. E. Barnato, S. L. Alexander, W. T. Linde-Zwirble and D. C. Angus, *Am. J. Respir. Crit. Care Med.*, 2008, **177**, 279–284.

15   G. Teasdale, A. Maas, F. Lecky, G. Manley, N. Stocchetti and G. Murray, *Lancet Neurol.*, 2014, **13**, 844–854.

16   J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter and L. G. Thijs, *Intensive Care Med.*, 1996, 22, 707–710.

17   F. Lopes Ferreira, D. Peres, A. Bross, C. Mélot and J.-L. Vincent, *JAMA*, 2001, **286**, 1754–1758.

18   C. W. Seymour, V. X. Liu, T. J. Iwashyna, F. M. Brunkhorst, T. D. Rea, A. Scherag, G. D. Rubenfeld, J. M. Kahn, M. Shankar-Hari, M. Singer, C. S. Deutschman, G. J. Escobar and D. C. Angus, *JAMA*, 2016, **315**, 762–774.

19   C. M. Robertson and C. M. Coopersmith, *Microbes Infect.*, 2006, **8**, 1382–1389.

20   N. Mancini, S. Carletti, N. Ghidoli, P. Cichero, R. Burioni and M. Clementi, *Clin. Microbiol. Rev.*, 2010, **23**, 235–251.

21   A. Afshari, J. Schrenzel, M. Ieven and S. Harbarth, *Crit. Care*, 2012, **16**, 1–12.

22   A. J. M. Loonen, P. F. G. Wolffs, C. a Bruggeman and A. J. C. van den Brule, *Eur. J. Clin. Microbiol. Infect. Dis.*, 2014, **33**, 1687–1702.

23   O. Opota, A. Croxatto, G. Prod'hom and G. Greub, *Clin. Microbiol. Infect.*, 2015, **21**, 313–322.

24   A. Rhodes, L. E. Evans, W. Alhazzani, M. M. Levy, M. Antonelli, R. Ferrer, A. Kumar, J. E. Sevransky, C. L. Sprung, M. E. Nunnally, B. Rochwerg, G. D. Rubenfeld, D. C. Angus, D. Annane, R. J. Beale, G. J. Bellinghan, G. R. Bernard, J.-D. Chiche, C. Coopersmith, D. P. De Backer, C. J. French, S. Fujishima, H. Gerlach, J. L. Hidalgo, S. M. Hollenberg, A. E. Jones, D. R. Karnad, R. M. Kleinpell, Y. Koh, T. C. Lisboa, F. R. Machado, J. J. Marini, J. C. Marshall, J. E. Mazuski, L. A. McIntyre, A. S. McLean, S. Mehta, R. P. Moreno, J. Myburgh, P. Navalesi, O. Nishida, T. M. Osborn, A. Perner, C. M. Plunkett, M. Ranieri, C. A. Schorr, M. A. Seckel, C. W. Seymour, L. Shieh, K. A. Shukri, S. Q. Simpson, M. Singer, B. T. Thompson, S. R. Townsend, T. Van der Poll, J.-L. Vincent, J. W. Wiersinga, J. L. Zimmerman and P. R. Dellinger, *Crit. Care Med.*, 2017, **45**, 486–552.

25   A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar and M. Cheang, *Crit. Care Med.*, 2006, **34**, 1589–1596.

26   C. W. Seymour, F. Gesten, H. C. Prescott, M. E. Friedrich, T. J. Iwashyna, G. S. Phillips, S. Lemeshow, T. Osborn, K. M. Terry and M. M. Levy, *N. Engl. J. Med.*, 2017, **376**, 2235–2244.

27   R. Ferrer, I. Martin-Loeches, G. Phillips, T. M. Osborn, S. Townsend, P. R. Dellinger, A. Artigas, C. Schorr and M. M. Levy, *Crit. Care Med.*, 2014, **42**, 1749–1755.

28   A. Kumar, P. Ellis, Y. Arabi, D. Roberts, B. Light, J. E. Parrillo, P. Dodek, G. Wood, A. Kumar, D. Simon, C. Peters, M. Ahsan and D. Chateau, *Chest*, 2009, **136**, 1237–1248.

29   O. Takeuchi and S. Akira, *Cell*, 2010, **140**, 805–820.

30   D. Rittirsch, M. A. Flierl and P. A. Ward, *Nat. Rev. Immunol.*, 2008, **8**, 776–787.

31   R. V D'Elia, K. Harrison, P. C. Oyston, R. A. Lukaszewski and G. C. Clark, *Clin. Vaccine Immunol.*, 2013, **20**, 319–327.

32   T. van der Poll, F. L. van de Veerdonk, B. P. Scicluna and M. G. Netea, *Nat. Rev. Immunol.*, 2017, **17**, 407–420.

33   M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszewski, G. Thiéfin and G. D. Sockalingum, *Chem. Soc. Rev.*, 2016, **45**, 1803–1818.

34   C. Pierrakos and J.-L. Vincent, *Crit. Care*, 2010, **14**, 1–18.

35   T. Gutierrez, R. Hornigold and A. Pearce, *Surgery*, 2014, **32**, 149–152.

36   R. Easton and Z. J. Balogh, *Inj. Int. J. Care Inj.*, 2014, **45**, 934–941.

37   J. C. Marshall, *Trends Mol. Med.*, 2014, **20**, 195–203.

38   G. Bernard, J.-L. Vincent, P.-F. Laterre, S. P. Larosa, J.-F. Dhainaut, A. Lopez-Rodriguez, J. S. Steingrub, G. E. Garber, J. D. Helterbrand, W. E. Ely and C. J. Fisher, *N. Eng. J. Med.*, 2001, **344**, 699–709.

39   M. V Ranieri, T. B. Thompson, P. S. Barie, J.-F. Dhainaut, I. S. Douglas, S. Finfer, B. Gårdlund, J. C. Marshall, A. Rhodes, A. Artigas, D. Payen, J. Tenhunen, V. Al-Khalidi, Hussein R Thompson, J. Janes, W. L. Macias, B. Vangerow and M. D. Williams, *N. Engl. J. Med.*, 2012, **366**, 2055–2064.

40   L. McHugh, T. A. Seldon, R. A. Brandon, J. T. Kirk, A. Rapisarda, A. J. Sutherland, J. J. Presneill, D. J. Venter, J. Lipman, M. R. Thomas, P. M. C. Klein Klouwenberg, L. van Vught, B. Scicluna, M. Bonten, O. L. Cremer, M. J. Schultz, T. van der Poll, T. D. Yager and R. B. Brandon, *PLoS Med.*, 2015, **12**, 1–35.

41   T. E. Sweeney, A. Shidham, H. R. Wong and P. Khatri, *Sci. Transl. Med.*, 2015, **7**, 1–33.

42   E. L. Tsalik, R. Henao, M. Nichols, T. Burke, E. R. Ko, M. T. McClain, L. L. Hudson, A. Mazur, D. H. Freeman, T. Veldman, R. J. Langley, E. B. Quackenbush, S. W. Glickman, C. B. Cairns, A. K. Jaehne, E. P. Rivers, R. M. Otero, A. K. Zaas, S. F. Kingsmore, J. Lucas, V. G. Fowler, L. Carin, G. S. Ginsburg and C. W. Woods, *Sci. Transl. Med.*, 2016, **8**, 1–11.

43   B. P. Scicluna, L. A. van Vught, A. H. Zwinderman, M. A. Wiewel, E. E. Davenport, K. L. Burnham, P. Nürnberg, M. J. Schultz, J. Horn, O. L. Cremer, M. J. Bonten, C. J. Hinds, H. R. Wong, J. C. Knight, T. van der Poll, F. M. de Beer, L. D. J. Bos, J. F. Frencken, M. E. Koster-Brouwer, K. van de Groep, D. M. Verboom, G. J. Glas, R. T. M. van Hooijdonk, A. J. Hoogendijk, M. A. Huson, P. M. K. Klouwenberg, D. S. Y. Ong, L. R. A. Schouten, M. Straat, E. Witteveen and L. Wieske, *Lancet Respir. Med.*, 2017, **2600**, 1–11.

44   R. A. Lukaszewski, A. M. Yates, M. C. Jackson, K. Swingler, J. M. Scherer, A. J. Simpson, P. Sadler, P. McQuillan, R. W. Titball, T. J. G. Brooks and M. J. Pearce, *Clin. Vaccine Immunol.*, 2008, **15**, 1089–1094.

45   R. A. Shaw, S. Low-Ying, A. Man, K.-Z. Liu, C. Mansfield, C. B. Rileg and M. Vijarnsorn, *Biomedical Vibrational Spectroscopy*, John Wiley and Sons, Inc, 2008.

46   C. N. Banwell and E. M. McCash, *Fundamentals of molecular spectroscopy*,

McGraw-Hill, London, 1983.

47   G. Gauglitz and T. Vo-Dinh, *Handbook of Spectroscopy*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2003.

48   *Spectroscopy in a Suitcase*, RSC, 2016.

49   M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nat. Protoc.*, 2014, **9**, 1771–1791.

50   G. D. Sockalingum, P. Dumas and J. Sule-Suso, *Trends Biotechnol.*, 2007, **25**, 40–44.

51   G. Clemens, B. Bird, M. Weida, J. Rowlette and M. J. Baker, *Spectrosc. Eur.*, 2014, **26**, 14–19.

52   K. Yeh, S. Kenkel, J.-N. Liu and R. Bhargava, *Anal. Chem.*, 2014, **87**, 485–493.

53   A. Hasenkampf, N. Kröger, A. Schönhals, W. Petrich and A. Pucci, *Opt. Express*, 2015, **23**, 5670–5680.

54   *Diamond Light Source, Ltd*, Http://www.diamond.ac.uk/Home/About/How-Diamond-Works.html.

55   S. Gayadeen and S. R. Duncan, *Control Eng. Pr.*, 2014, **26**, 201–210.

56   K. M. Dorling and M. J. Baker, *Trends Biotech*, 2013, **31**, 327–328.

57   S. G. Kazarian and K. Chan, *Appl. Spectrosc.*, 2010, **64**, 135A–152A.

58   B. Schrader, *Infrared and Raman spectroscopy: methods and applications*, John Wiley and Sons, 2008.

59   P. Bassan, J. Lee, A. Sachdeva, J. Pissardini, K. M. Dorling, J. S. Fletcher, A. Henderson and P. Gardner, *Analyst*, 2013, **138**, 144–157.

60   J. Filik, M. D. Frogley, K. Wehbe, J. K. Pijanka and G. Cinque, *Analyst*, 2012, **137**, 853–861.

61   J. Cao, E. S. D. McNaughton, E. G. Stanley, A. G. Elefanty, M. J. Tobin and P. Heraud, *Analyst*, 2013, **138**, 4147–4160.

62   J. Lee, *Vib. Spec.*, 2017, **90**, 104–111.

63   H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone and F. L. Martin, *Nat. Protoc.*, 2016, **11**, 664–687.

64   F. S. Parker, *in Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry*, Springer, New York, 1983.

65   F. Draux, P. Jeannesson, A. Beljebbar, A. Tfayli, N. Fourre, M. Manfait, G. D. Sockalingum and J. Sulé-Suso, *Analyst*, 2009, **134**, 542–548.

66   I. Notingher and L. L. Hench, *Expert Rev. Med. Devices*, 2006, **3**, 215–234.

67   T. Bakker Schut, M. Witjes, H. Sterenborg, O. Speelman, J. Roodenburg, E.

Marple, B. H and G. Puppels, *Anal. Chem.*, 2000, **72**, 6010–6018.

68   C. Fulljames, N. Stone, D. Bennett and H. Barr, *Ital. J. Gastroenterol. Hepatol.*, 1999, **31**, 695–704.

69   A. Mahadevan-Jansen, M. F. Mitchell, N. Ramanujamf, U. Utzinger and R. Richards-Kortumt, *Photochem. Photobiol.*, 1998, **68**, 427–431.

70   M. G. Shim, L. M. Wong Kee Song, N. E. Marcon and B. C. Wilson, *Photochem. Photobiol.*, 2000, **72**, 146–150.

71   F. Bonnier, S. M. Ali, P. Knief, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. Lee, F. M. Lyng and H. J. Byrne, *Vib. Spectrosc.*, 2012, **61**, 124–132.

72   H. Barr, C. Kendall, J. Hutchings, F. Bazant-Hegemark, N. Shepherd and N. Stone, *Surgeon*, 2011, **9**, 119–123.

73   J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao and R. P. VanDuyne, *Nat. Mater.*, 2008, **7**, 442–453.

74   D. Graham and R. Goodacre, *Chem. Soc. Rev.*, 2008, **37**, 883–884.

75   K. Kneipp, Y. Wang, H. Kneipp, L. T. Perelman, I. Itzkan, R. R. Dasari and M. S. Feld, *Phys. Rev. Lett.*, 1997, **78**, 1667.

76   A. Bonifacio, D. M. S, R. Spizzo, S. Cervo, A. Steffan, A. Colombatti and V. Sergo, *Anal. Bioanal. Chem.*, 2014, **406**, 2355–2365.

77   C. Krafft and J. Popp, *Anal. Bioanal. Chem.*, 2015, **407**, 699–717.

78   I. W. Schie, C. Krafft and J. Popp, *Analyst*, 2015, **140**, 3897–3909.

79   M. Winterhalder and A. Zumbusch, *Adv. Drug Deliv. Rev.*, 2015, **89**, 135–144.

80   A. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe and M. S. Feld, *Cancer Res.*, 2006, **66**, 3317–3322.

81   N. Stone and P. Matousek, *Cancer Res.*, 2008, **68**, 4424–4430.

82   M. Z. Vardaki, B. Gardner, N. Stone and P. Matousek, *Analyst*, 2015, **140**, 5112–5119.

83   M. D. Keller, E. Vargis, N. de Matos Granja, R. H. Wilson, M.-A. Mycek, M. C. Kelley and A. Mahadevan-Jansen, *J. Biomed. Opt.*, 2011, **16**, 77006–77008.

84   P. Matousek and N. Stone, *J. Biophotonics*, 2013, **6**, 7–19.

85   B. Sharma, K. Ma, M. R. Glucksberg and R. P. Van Duyne, *J. Am. Chem. Soc.*, 2013, **135**, 17290–17293.

86   N. Stone, R. Baker, K. Rogers, A. W. Parker and P. Matousek, *Analyst*, 2007, **132**, 899–905.

87   M. J. Baker, *Special Issue: Photonic Biofluid Diagnostics*, Wiley-VCH Verlag GmbH & Co, Weinheim, 2014.

88   X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon and G. D. Sockalingum, *Transl. Res.*, 2013, **162**, 279–286.

89   V. Gaydou, A. Lecellier, D. Toubas, J. Mounier, L. Castrec, G. Barbier, W. Ablain, M. Manfait and G. Sockalingum, *Anal. Methods*, 2015, **7**, 766–778.

90   V. Untereiner, G. D. Sockalingum, R. Garnotel, C. Gobinet, F. Ramaholimihaso, F. Ehrhard, M. D. Diebold and G. Thiéfin, *J. Biophotonics*, 2014, **7**, 241–253.

91   R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads and T. D. Veenstra, *Mol. Cell. Proteomics*, 2003, **2**, 1096–1103.

92   E. F. Petricoin, C. Belluco, R. P. Araujo and L. A. Liotta, *Nat. Rev. Cancer*, 2006, **6**, 961–967.

93   C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, *Analyst*, 2015, **140**, 2280–2286.

94   J. Ollesch, M. Heinze, H. M. Heise, T. Behrens, T. Brüning and K. Gerwert, *J. Biophotonics*, 2014, **7**, 210–221.

95   J. Backhaus, R. Mueller, N. Formanski, N. Szlama, H.-G. Meerpohl, M. Eidt and P. Bugert, *Vib. Spectrosc.*, 2010, **52**, 173–177.

96   E. Scaglia, G. D. Sockalingum, J. Schmitt, C. Gobinet, N. Schneider, M. Manfait and G. Thiéfin, *Anal. Bioanal. Chem.*, 2011, **401**, 2919–2925.

97   G. L. Owens, K. Gajjar, J. Trevisan, S. W. Fogarty, S. E. Taylor, D. Gama-Rose, P. L. Martin-Hirsch and F. L. Martin, *J. Biophotonics*, 2014, **7**, 200–209.

98   K. Gajjar, L. D. Heppenstall, W. Pang, K. M. Ashton, J. Trevisan, I. I. Patel, V. Llabjani, H. F. Stringfellow, P. L. Martin-Hirsch and T. Dawson, *Anal. Methods*, 2013, **5**, 89–102.

99   J. R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R. W. Lea, J. A. McIntosh and M. J. Baker, *Anal. Bioanal. Chem.*, 2013, **405**, 7347–7355.

100   J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D. Jenkinson, R. W. Lea and C. Walker, *J. Biophotonics*, 2014, **7**, 189–199.

101   A. Sahu, S. Sawant, H. Mamgain and C. M. Krishna, *Analyst*, 2013, **138**, 4161–4174.

102   A. Sahu, N. Nandakumar, S. Sawant and C. M. Krishna, *Analyst*, 2015, **140**, 2294–2301.

103   J. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. G.-T. C. Ramírez-Alvarado, G. Gutiérrez-Juárez and C. Medina-Gutiérrez, *Lasers Med. Sci.*, 2007, **22**, 229–236.

104   X. Li, T. Yang and S. Li, *Appl. Opt.*, 2012, **51**, 5038–5043.

105   I. Taleb, G. Thiéfin, C. Gobinet, V. Untereiner, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Hillon, M. Manfait, P. Ducoroy and G. D. Sockalingum, *Analyst*, 2013, **138**, 4006–4014.

106   S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen and H. Zeng, *Biosens. Bioelectron*, 2010, **25**, 2414–2419.

107   S. Li, Y. Zhang, Q. Zeng, L. Li, Z. Guo, Z. Liu, H. Xiong and S. Liu, *Laser Phys. Lett.*, 2014, **11**, 65603.

108   D. Lin, J. Pan, H. Huang, G. Chen, S. Qiu, H. Shi, W. Chen, Y. Yu, S. Feng and R. Chen, *Sci. Rep.*, 2014, **4**, 475.

109   S. Feng, R. Chen, J. Lin, J. Pan, Y. Wu, Y. Li, J. Chen and H. Zeng, *Biosens. Bioelectron*, 2011, **26**, 3167–3174.

110   H. Ito, H. Inoue, K. Hasegawa, Y. Hasegawa, T. Shimizu, S. Kimura, M. Onimaru, H. Ikeda and S.-E. Kudo, *Nanomedicine Nanotechnol. Biol. Med.*, 2014, **10**, 599–608.

111   D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng and R. Chen, *Opt. Express*, 2011, **19**, 13565–13577.

112   S. Li, Y. Zhang, J. Xu, L. Li, Q. Zeng, L. Lin, Z. Guo, Z. Liu, H. Xiong and S. Liu, *Appl. Phys. Lett.*, 2014, **105**, 91104.

113   E. Ryzhikova, O. Kazakov, L. Halamkova, D. Celmins, P. Malone, E. Molho, E. A. Zimmerman and I. K. Lednev, *J. Biophotonics*, 2015, **8**, 584–596.

114   P. Carmona, M. Molina, M. Calero, F. Bermejo-Pareja, P. Martínez-Martín and A. Toledano, *J. Alzheimers Dis.*, 2013, **34**, 911–920.

115   E. Peuchant, S. Richard-Harston, I. Bourdel-Marchasson, J.-F. Dartigues, L. Letenneur, P. Barberger-Gateau, S. Arnaud-Dabernat and J.-Y. Daniel, *Transl. Res.*, 2008, **152**, 103–112.

116   W. Petrich, K. Lewandrowski, J. Muhlestein, M. Hammond, J. Januzzi, E. Lewandrowski, R. Pearson, B. Dolenko, J. Früh and M. Haass, *Analyst*, 2009, **134**, 1092–1098.

117   U. Neugebauer, S. Trenkmann, T. Bocklitz, D. Schmerler, M. Kiehntopf and J. Popp, *J. Biophotonics*, 2014, **7**, 232–240.

118   N. L. Garrett, R. Sekine, M. W. Dixon, L. Tilley, K. R. Bambery and B. R. Wood, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21164–21168.

119   R. Somorjai, B. Dolenko, A. Nikulin, P. Nickerson, D. Rush, A. Shaw, M. Glogowski, J. Rendell and R. Deslauriers, *Vib. Spectrosc.*, 2002, **28**, 97–102.

120   J. A. M. Bispo, E. E. de Sousa Vieira, L. Silveira and A. B. Fernandes, *J. Biomed. Opt.*, 2013, **18**, 87004.

121   G. Del Mistro, S. Cervo, E. Mansutti, R. Spizzo, A. Colombatti, P. Belmonte, R. Zucconelli, A. Steffan, V. Sergo and A. Bonifacio, *Anal. Bioanal. Chem.*, 2015, **407**, 3271–3275.

122   D. A. Scott, D. E. Renaud, S. Krishnasamy, P. Meriç, N. Buduneli, Ş. Çetinkalp and K.-Z. Liu, *Diabetol. Metab. Syndr.*, 2010, **2**, 48.

123   X. Li, T. Yang and J. Lin, *J. Biomed. Opt.*, 2012, **17**, 370031–370035.

124   W. Yan, H. Lin, L. Jinghua, Q. Dian, C. Anyu, J. Yi, G. Xun, L. Chunwei, H. Wen and W. Hong, *IEEE*, 2009, 885–887.

125   S. Whiteman, Y. Yang, J. Jones and M. Spiteri, *Ther. Adv. Respir. Dis.*, 2008,

2, 23–31.

126 A. Travo, C. Paya, G. Déléris, J. Colin, B. Mortemousque and I. Forfar, *Anal. Bioanal. Chem.*, 2014, **406**, 2367–2376.

127 S. Choi, S. W. Moon, J.-H. Shin, H.-K. Park and K.-H. Jin, *Anal. Chem.*, 2014, **86**, 11093–11099.

128 H. Eysel, M. Jackson, A. Nikulin, R. Somorjai, G. Thomson and H. Mantsch, *Biospectroscopy*, 1997, **3**, 161–167.

129 K. A. Esmonde-White, G. S. Mandair, F. Raaii, J. A. Jacobson, B. S. Miller, A. G. Urquhart, B. J. Roessler and M. D. Morris, *J. Biomed. Opt.*, 2009, **14**, 34013–34018.

130 K.-Z. Liu, T. C. Dembinski and H. H. Mantsch, *Am. J. Obs. Gynecol.*, 1998, **178**, 234–241.

131 G. Graça, A. S. Moreira, A. J. V Correia, B. J. Goodfellow, A. S. Barros, I. F. Duarte, I. M. Carreira, E. Galhano, C. Pita and M. do Céu Almeida, *Anal. Chim. Acta*, 2013, **764**, 24–31.

132 M. Griebe, M. Daffertshofer, M. Stroick, M. Syren, P. Ahmad-Nejad, M. Neumaier, J. Backhaus, M. G. Hennerici and M. Fatar, *Neurosci. Lett.*, 2007, **420**, 29–33.

Chapter II. Impact of the "coffee ring" effect on serum vibrational spectroscopy

II.1 Impact de l'effet "rond de café" sur l'étude du sérum par spectroscopie vibrationnelle

II.1.1 Résumé

Les études des biofluides par spectroscopie vibrationnelle ont augmenté au cours de ces dernières années car les biopsies liquides sont de plus en plus analysées dans le but de découvrir des informations moléculaires spécifiques liées à des maladies ciblées. Ce type d'échantillons cliniques est peu invasif pour le patient, facilement collecté et accessible pour des recherches à visée biomédicale. De plus, les biofluides représentent un milieu idéal pour des analyses spectroscopiques car la préparation des échantillons ne nécessite pas de procédures lourdes, contrairement à celle de coupes tissulaires ou pour l'analyse de cellules. L'eau ayant une contribution importante en spectroscopie infrarouge, l'approche la plus commune pour l'analyse de biofluides est de laisser sécher les échantillons avant l'analyse. Cependant, le séchage de gouttes de biofluides est un phénomène complexe qui résulte en un dépôt inhomogène caractéristique, appelé effet ''rond de café''. L'objectif de cette étude est d'évaluer l'influence de cet effet sur les données collectées par spectroscopies infrarouge à transformée de Fourier (IRTF) et Raman pour l'analyse d'échantillons de sérums.

Les analyses ont porté sur un mélange commercialisé de sérums humains et sur un modèle simple constitué par une seule protéine, l'albumine sérique bovine (ASB) étant donné que l'albumine sérique humaine (ASH) est une protéine prédominante du sérum. Les influences de la dilution du sérum et du volume de l'échantillon déposé sur l'effet ''rond de café'' ont été étudiées par microimagerie spectrale, et à l'aide de la source Synchrotron (Diamond Light Source, Didcot, Royaume-Uni) ainsi que par microspectroscopie Raman. Les données acquises ont été analysées par la technique de classification K-means et par Analyse en Composante Principale (ACP).

II.1.2 Conclusions

Cette étude montre que les spectres collectés en mode point dans différentes zones d'une goutte de sérum séché sont affectés par les variations d'épaisseur de l'échantillon, cela questionnant la reproductibilité des acquisitions spectrales. De plus les analyses par microspectroscopie IRTF ont montré que les bandes spectrales associées aux protéines sont saturées dans la zone de l'anneau, rendant l'interprétation de potentiels changements moléculaires en périphérie d'une goutte hasardeuse. Cependant, des spectres collectés à partir de dépôts automatisés de quelques nanolitres semblent moins affectés par l'effet ''rond de café'' et apparaissent donc comme une alternative intéressante.

II.2 Chapter overview

Biofluid vibrational spectroscopy studies have increased over the past few years as liquid biopsies are more and more investigated to search for disease-specific molecular information. They are minimally invasive for the patient, less subject to sampling bias, easily collectable and accessible for biomedical research purposes. Moreover, they represent an ideal medium for spectroscopic analysis as no heavy sample preparation is required unlike tissue sections or cell analysis. However, biofluid dried drop deposition is affected by the naturally occurring "coffee ring" effect, a phenomenon still widely investigated, that leaves a heterogeneous pattern after the liquid evaporation. The aim of this study is to evaluate the impact of the "coffee ring" effect on spectral data collected from serum dried drops. Some factors that can influence this phenomenon, such as the dilution and the volume of sample deposited, have also been considered. Fourier Transform Infrared (FTIR) and Raman microspectroscopy techniques have been compared in order to determine the most suitable methodology approach for serum analysis with an overall clinical application purpose.

II.3 Introduction

Water has an important spectral contribution in FTIR spectra, and as a result a common approach for analysing liquids is to dry the samples. However, the drying process of a liquid is a complex phenomenon depending on several parameters such as temperature, relative humidity, wetting properties of the substrate, and surface roughness[1–5]. The evaporation and surface tension gradient occurring during the drying process of a drop trigger two opposite flows, the capillary flow that induces the migration of the molecules towards the edge of the drop and the Marangoni flow that redistributes the particles towards the centre of the drop[6–8]. Commonly, the combination of both flows causes the formation of a ring at the periphery of a dried drop, also known as the "coffee

ring" effect[9]. This pattern deposition is affected by the composition of the liquid, size and form of solid particles as well as their concentration[10–12].

In biofluids such as serum and plasma, due to their complex composition, it is even more complicated to predict the behaviour of their constituents during the drying process. Plasma hosts more than 10,000 different proteins with the capacity to interact with a vast range of ligands[13,14]. The deposit is also influenced by the different adsorption capacities of proteins to a surface. Usually, low molecular weight proteins initially bound to a surface are replaced by larger proteins with a higher affinity for the substrate surface (Vroman effect)[15–17]. However, the resulting inhomogeneous pattern of a dried serum drop reflects the sample composition. Therefore, a potential biomedical application has raised interest in understanding the formation of crystals in the centre of drops due to the natural presence of salts in serum or cracks occurring during the gel-solid state of the drying process[18–21]. Despite the fact that many studies have been focusing on the physical mechanisms responsible for the uneven drying or the morphology properties of the resulting inhomogeneous pattern, little is known on the pattern composition itself. In the present study, we propose to investigate the "coffee ring" effect by vibrational spectroscopy/imaging, a label-free technique and an ideal approach to determine potential molecular changes in dried serum drops and on a Bovine Serum Albumin (BSA) solution model.

II.4 Material and methods

II.4.1 Study model

Experiments were performed using normal human mixed pooled serum (TCSBiosciences, UK). Dilution tests were realised with normal saline (NaCl 0.9%, Miniversol, Aguettant, France). Drops of 0.5 μL were hand spotted while 0.5 nL drops were obtained by a Piezoelectric dispensing system (Nordson EFD, PICO LV-100) onto a calcium fluoride (CaF_2) window (Crystran Limited, UK)

and analysed by FTIR imaging. Synchrotron based measurements of 0.5 nL drops were acquired on MirrIR low-e reflective substrates (Kevley Technologies, USA).

Liquid Raman spectroscopy tests were conducted by depositing 10, 20 or 30 µL of undiluted pooled serum per well (7 mm diameter) of an aluminium plate or 130 µL in a quartz cuvette. Drops of pooled serum (0.5 µL) dried onto a CaF_2 window were also analysed by Raman microspectroscopy.

BSA used as a simple model (Sigma-Aldrich) was solubilised with distilled water (B. Braun Melsungen AG, Germany), and prepared at concentrations of 5 mg/mL or 100 mg/mL for FTIR and Raman spectroscopy analysis respectively.

II.4.2 FTIR imaging

Spectral images of 0.5 µL and 0.5 nL dried serum drops were acquired in the transmission mode with a PerkinElmer Spectrum Spotlight 400 imaging system (Courtaboeuf, France) at a spatial resolution of 6.25 µm/pixel, a spectral resolution of 4 cm^{-1}, and 4 co-added scans over the spectral range of 4000-750 cm^{-1} (Spectrum-Image software, version 1.6, PerkinElmer).

II.4.3 Synchrotron FTIR microspectroscopy

Spectral maps were collected at the Diamond Light Source MIRIAM B22 beamline (Didcot, UK) with the Hyperion 3000 FTIR spectrometer (Bruker Optics GmbH, Ettlingen, Germany). Spectral acquisitions were performed in the reflectance mode (OPUS software version 7.0, Bruker Optics GmbH) using a 36x objective, 10 µmx10 µm aperture, 8 µm step, over the 4000-600 cm^{-1} range with a spectral resolution of 4 cm^{-1} and 32 co-added scans.

II.4.4 Raman microspectroscopy

Five single point-mode spectra from the periphery and the centre of 4 serum dried drops were collected with the LabRAM ARAMIS Raman microspectrometer and the software LabSpec 5 (Horiba Jobin Yvon, Villeneuve d'Ascq, France). Spectra were

acquired over the spectral range 600-1800 cm^{-1} with a laser excitation at 785 nm, 600 lines/mm grating, a confocal hole of 1000 µm and a 100x microscope objective (Olympus). Spectral acquisition time was 30 seconds and 2 accumulations were co-added.

II.4.5 Data analysis

All spectral images of serum drops were subjected to an atmospheric correction for water vapour and CO_2 using the Spectrum-Image software V1.6 (PerkinElmer, Les Ulis, France). Spectra with a too low signal to noise ratio were discarded using an Extended Multiplicative Signal Correction (EMSC) method. All remaining spectra were rubberband baseline corrected and vector normalised. Corrected spectral images were subjected to a K-means clustering analysis using 5 classes. The white pixels in the K-means cluster images correspond to the spectra eliminated before the pre-processing.

Synchrotron FTIR serum spectra or spectra extracted from BSA FTIR images, were cut in the spectral range 4000-900 cm^{-1}, rubberband baseline corrected, vector normalised and analysed by Principal Component Analysis (PCA).

Raman spectra pre-processing was performed with the Matlab software (The MathWorks, Inc., USA). The procedure was as follows. First, sample spectra were corrected for the instrument response using optics and dark current spectra. Frequency calibration was done using standard spectra from a silicon wafer and from a neon lamp. Intensity calibration was performed using a 785 nm NIST standard (Gaithersburg, MD, USA). Finally, spectra were smoothed with a Savitzky-Golay function (polynomial order 2), baseline corrected using a 5th order polynomial function, and vector normalised before performing a PCA.

II.5 Results

II.5.1 Assessing the "coffee ring" effect by FTIR microspectroscopy /imaging

II.5.1.1 FTIR microimaging

Serum drops after drying show an inhomogeneous surface with the presence of a ring at the periphery of the drop and cracks as well as fern-pattern crystals in the centre (Figure 2.1, white light images). IR imaging demonstrates that a higher absorbance is measured in the ring compared to the centre (Figure 2.1, spectral images). K-means clustering was used to assess the possible spectral molecular changes in the different physical structures exhibited after the drying of the serum drop with or without dilution. It can be seen on the pure 0.5 μL serum drop, that most of the spectra from the ring were discarded because of a too high absorbance (Figure 2.1a, K-means clustering image). The edge of the drops are represented by the class 4 (orange). It can be observed that the amide I band is shifted ($1642\ cm^{-1}$) in the corresponding centroid spectrum (Figure 2.2) compared to the amide I band ($1653\ cm^{-1}$) of class 2 and 5, probably resulting from a spectral artefact due to the shape of the drop edge. The ring structure is characterised by the classes 1 and 3 in blue and green respectively (Figure 2.1a-d, K-means clustering images) however the centroid spectra (Figure 2.2) corresponding to these 2 classes are exhibiting abnormal flat amide I and II bands, indicating that the spectra are saturating in the ring region even after a 4-fold dilution. The centre of the drops is represented by the classes 2 (light blue) and 5 (pink). The class 2 is associated with the crystal features and is more important in the 4-fold diluted (41.3%) than in the 3-fold diluted (11.2%) serum drop (Figure 2.1d, K-means clustering image) as expected from the dilution of serum with normal saline.
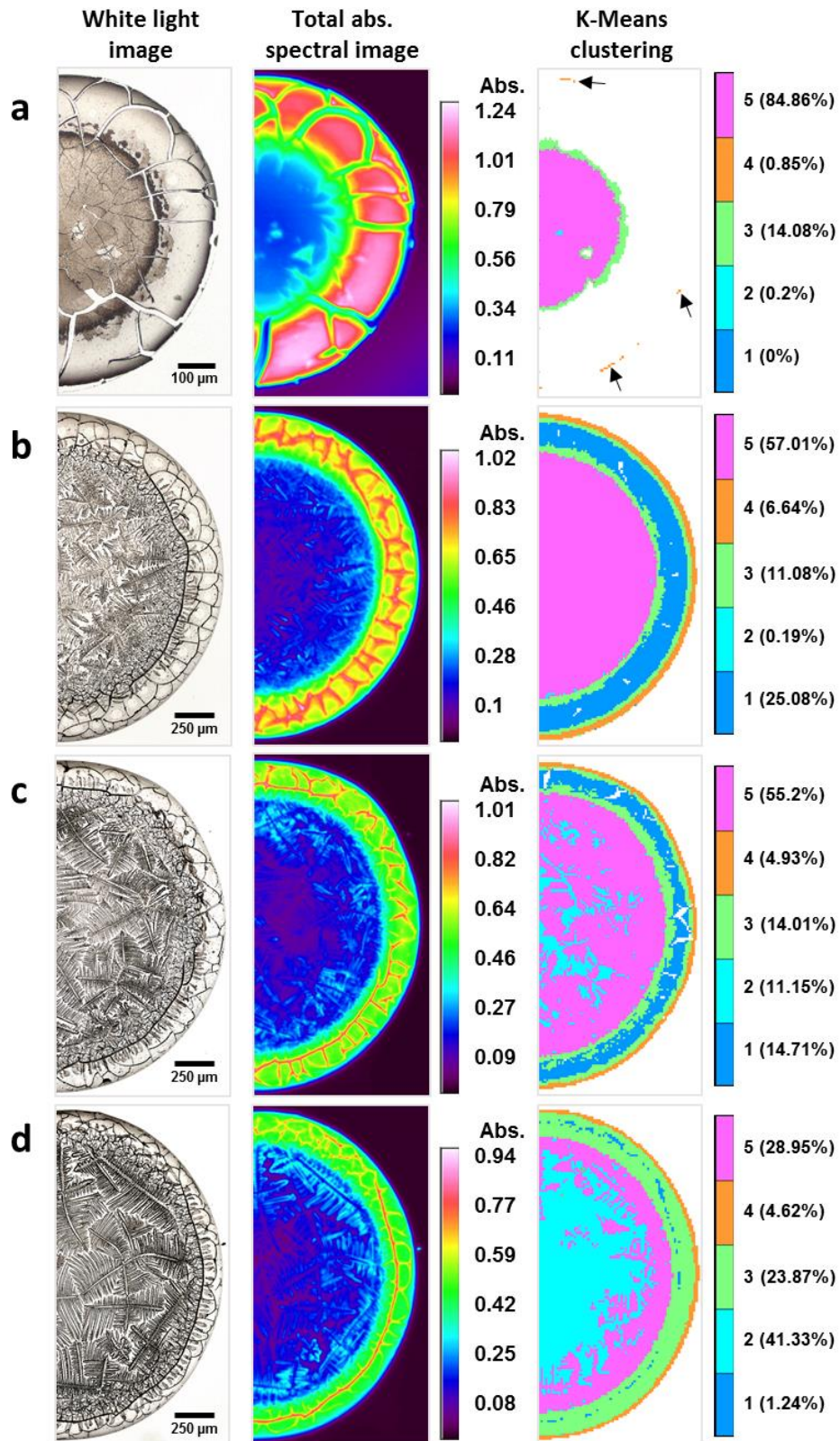
Figure 2.1 *Assessing the ''coffee ring'' effect by IR spectral imaging and K-means clustering. White light images (1ˢᵗ column), total absorbance spectral images (2ⁿᵈ column) and K-means cluster images (3ʳᵈ column) of pure, 1/2, 1/3, 1/4 diluted 0.5 µL serum drops respectively (**a-d**). Color bars represent the 5 classes of the K-mean clustering and their associated percentages. Arrows indicate edges of the drop.*

Spectral differences are visible all over the "fingerprint region" between the centroid spectra of the 2 centre classes, especially in the carbohydrates region (~1000 cm^{-1}) and lipids at 1730 cm^{-1}, suggesting molecular changes due to a different distribution of the serum components towards the drops, some constituents being segregated during the crystal formation or the spectral differences can be artefacts from a difference in thicknesses of the deposit. These results suggest that the single point measurement mode is not appropriate in order to ensure spectral quality and reproducibility due to a high variability of the sample physical properties.
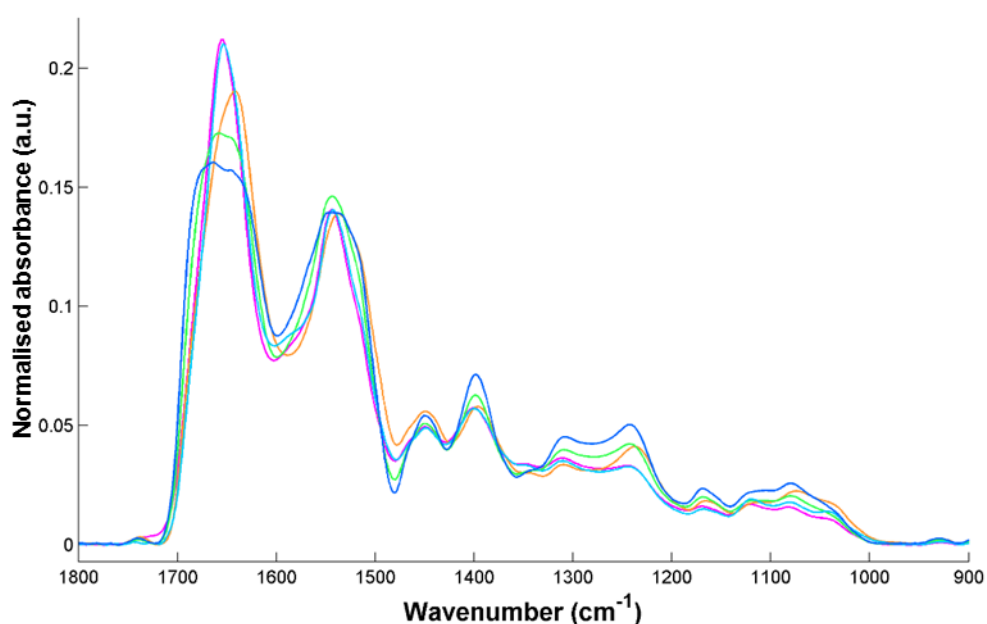
Figure 2.2 *Centroids of the K-means clustering results in 5 classes of Figures 2.1 and 2.3.*

In comparison, it can be seen in Figure 2.3 from the spectral image of a 0.5 nL serum drop, a thousand times smaller volume that the "coffee ring" effect still occurs. However, none of the spectra is affected by saturation even though the serum sample was not diluted. The surface of the drop appears more homogeneous (Figure 2.3, white light image) despite some striations resulting from the automated sampling. Spectral reproducibility is increased; 84% of the spectra belonging to class 5 despite the fact that 16% of the spectra remain affected by an edge effect.
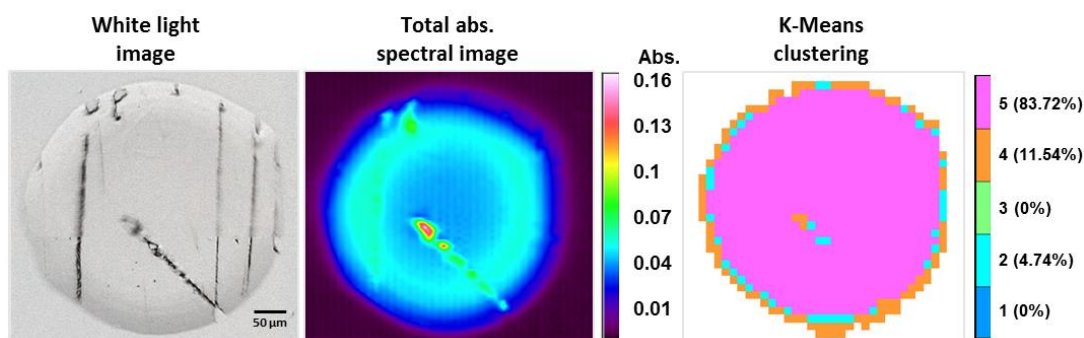
Figure 2.3 *Assessing the ''coffee ring'' effect by IR spectral imaging and K-means clustering. White light image (left), total absorbance spectral image (middle) and K-means cluster image (right) of a pure 0.5 nL serum drop. Color bars represent the 5 classes of the K-mean clustering and their associated percentage.*

A less complex model, a solution of BSA, has been studied to understand the origins of these spectral variations. Human Serum Albumin (HSA) is the most abundant protein in the serum, accounting for about 50% of the total proteins[22] and BSA is commonly used as a surrogate for HSA. A spectral image of a BSA drop at 5 mg/mL was acquired (Figure 2.4a). Spectra extracted from 7 zones of the drop exhibiting different levels of absorbance can be well separated by PCA (Figure 2.4b). This suggests that the difference in sample thicknesses is responsible for the spectral variations observed. It should be noted that at a BSA concentration higher than 5 mg/mL, the amide bands were saturated in spectra from the ring (data not shown). The level of proteins in serum is known to be between 60 to 80 mg/mL[22], as such serum samples should be diluted 16 fold to avoid spectral saturation from proteins in the ring formation. As a result, it would make impossible the analysis of other serum constituents of interest originally present at low concentrations.
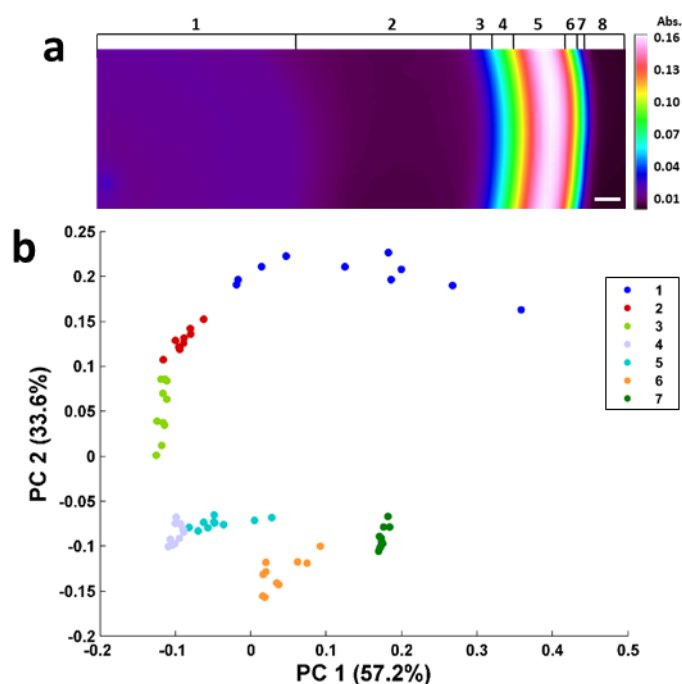
Figure 2.4 (a) *Spectral image of total absorbance of a BSA drop section at a concentration of 5 mg/mL. Numbers represent different zones of interest. Zones 1 and 2 correspond to the centre, and 3 to 7 to the ring of the drop. Zone 8 is the CaF$_2$ substrate. The white scale bar represents 100 μm.* ***(b)*** *PCA plot of IR spectra from the different regions (1 to 7) of the BSA drop (a).*

II.5.1.2 Synchrotron FTIR microspectroscopy

A map of a 0.5 nL dry drop acquired in the reflectance mode confirms that the ''coffee ring'' effect is clearly visible (Figure 2.5a). Moreover, spectra collected at the periphery and at the centre can be distinguished by PCA (Figure 2.5b). The PC1 loading responsible for the separation of the spectral data (Figure 2.6a) shows differences in the amide I/II bands and in the high wavenumber region. It can also be observed on the mean spectra of the 2 groups a shift in the amide I band; 1649 cm^{-1} and 1645 cm^{-1} for spectra from the periphery and the centre of the drop respectively (Figure 2.6b). A baseline distortion was also observed in the raw spectra acquired in the centre of the drop (data not shown), suggesting that the spectral differences are due to the sample thickness variations or scattering effects.
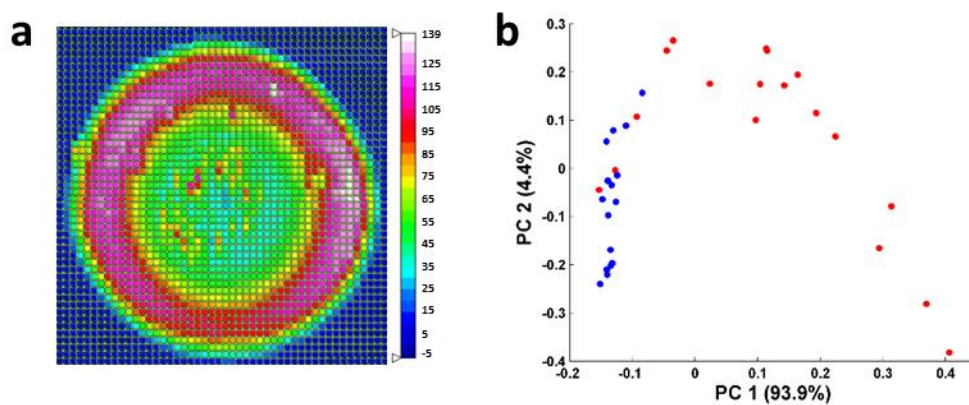
Figure 2.5 (a) *Spectral map total absorbance of a 0.5 nL drop acquired by Synchrotron FTIR microspectroscopy.* **(b)** *PCA plot of IR spectra from the periphery (blue) and centre (red) of a nL serum drop.*
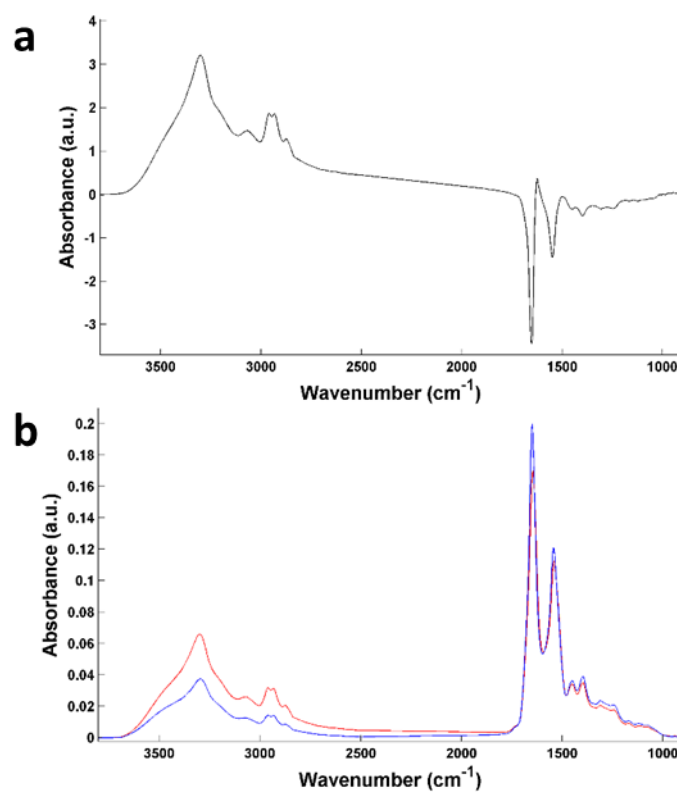
Figure 2.6 (a) *PC 1 loading from the PCA Figure 2.5b.* **(b)** *Mean pre-processed IR spectra from periphery (blue) and centre (red) of a nL pooled serum drop.*

81

II.5.2 Assessing the "coffee ring" effect by Raman spectroscopy

To consider the heterogeneity of dried serum drops by Raman microspectroscopy, 5 spectra were respectively acquired from the centre and from the periphery of the 4 drops. Then, a PCA was performed (Figure 2.7) and it can be seen that spectra of the 2 zones are distinguishable. The PC3 loading responsible for the separation of the spectral data (Figure 2.8a) shows differences all over the spectral range covered for the analysis as also seen on the mean spectra of the 2 groups (Figure 2.8b) and especially the amide I band (1670 cm^{-1}), CH$_2$ (1415 cm^{-1}), phenylalanine (1004 cm^{-1}), C-C stretch (854 cm^{-1}) peaks. Again, the spectral differences observed could be associated to molecular changes due to an inhomogeneous distribution of the serum components or to artefact factors coming from differences of thickness of the sample.
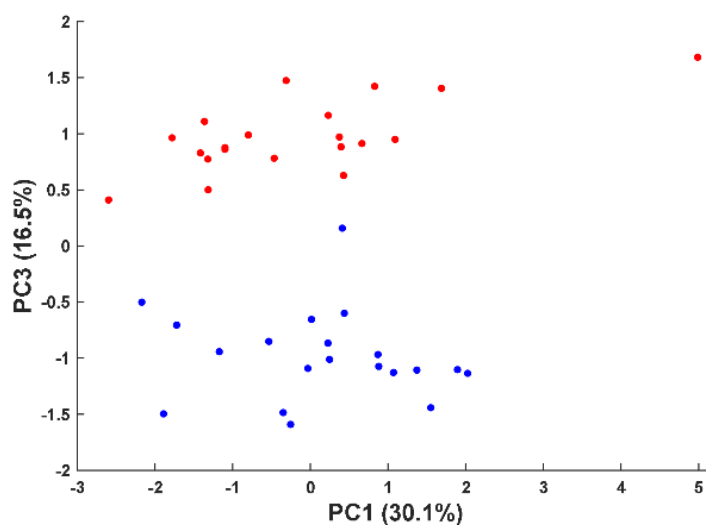
Figure 2.7 *PCA plot of Raman spectra collected in the periphery (blue) and centre (red) of pooled serum drops (0.5 µL).*

As previously, spectra were acquired in the centre or in the ring formation of BSA drops (Figure 2.9a). Again, spectra from the 2 zones were distinguishable by PCA (Figure 2.9b). This observation confirms that the spectral variations observed are artefacts due to the difference of thickness of the samples.
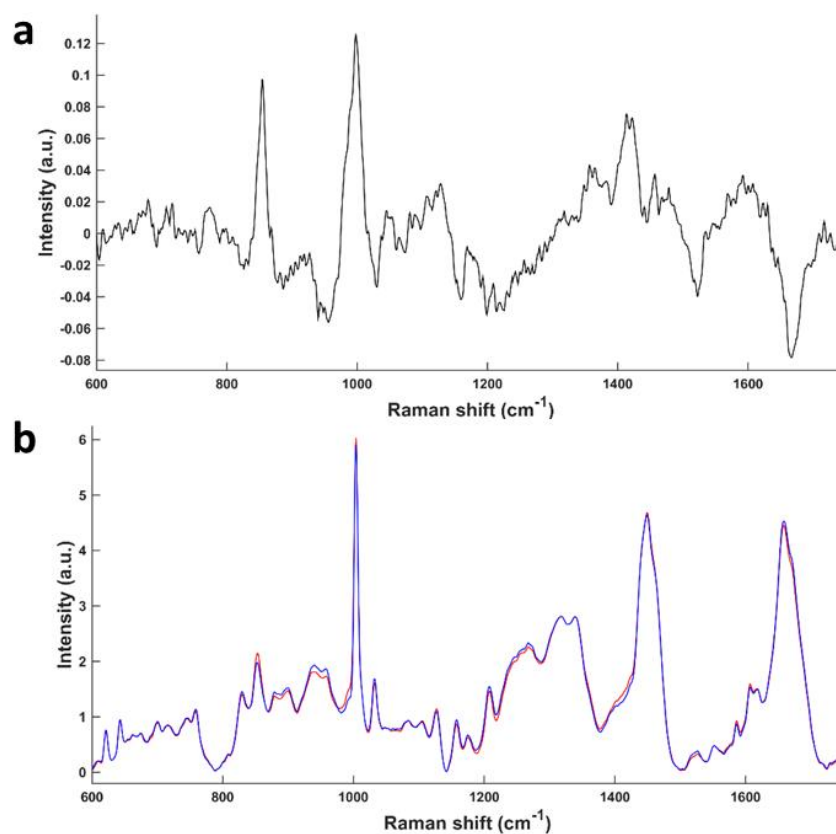
Figure 2.8 *(**a**) PC 3 loading from the PCA Figure 2.7. (**b**) Mean pre-processed Raman spectra collected in the periphery (blue) and centre (red) of pooled serum drops (0.5 µL).*
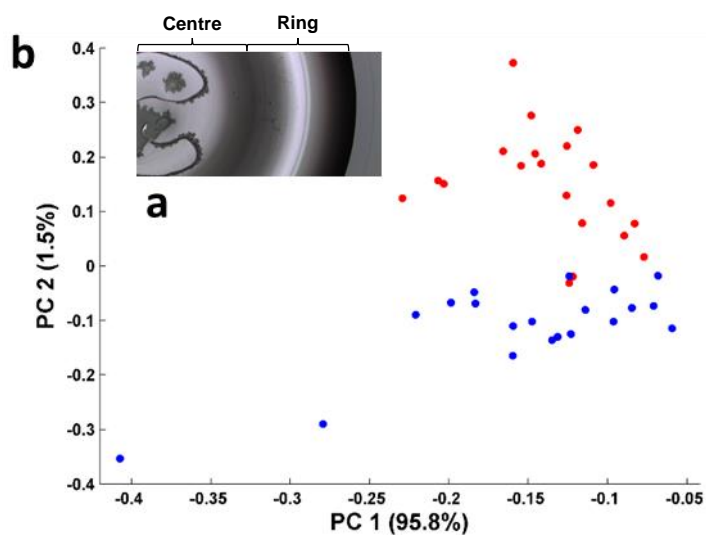
Figure 2.9 *(**a**) White light image of a BSA drop section at a concentration of 100 mg/mL. (**b**) PCA plot of Raman spectra from the ring (blue dots) and the centre (red dots) of BSA drops.*

II.5.3 Liquid serum Raman microspectroscopy

The drying of serum drops resulting in an inhomogeneous surface to analyse, leads to the question whether acquisitions can be performed directly on liquid samples. Raman spectroscopy appears as the suitable technique in this situation as water is a weak scatterer. This approach has been considered and different volumes (10, 20, 30 µL) pipetted into wells of an aluminium plate (96 wells) were tested. The 2 lower volumes were discarded as the samples were drying very quickly under the heating of the laser during the acquisition. Spectra obtained at the surface of 30 µL drops (maximal capacity of a well) were of poor quality. The amide I band and phenylalanine peak were hardly discernable (Figure 2.10). This result can be explained by the fact that the liquid was evaporating during the acquisition, resulting in the loss of laser focus. Then, acquisitions were performed at 2 µm under the surface of the drop, increasing the signal obtained but the signal to noise ratio was considered too low to exploit the data (Figure 2.10).
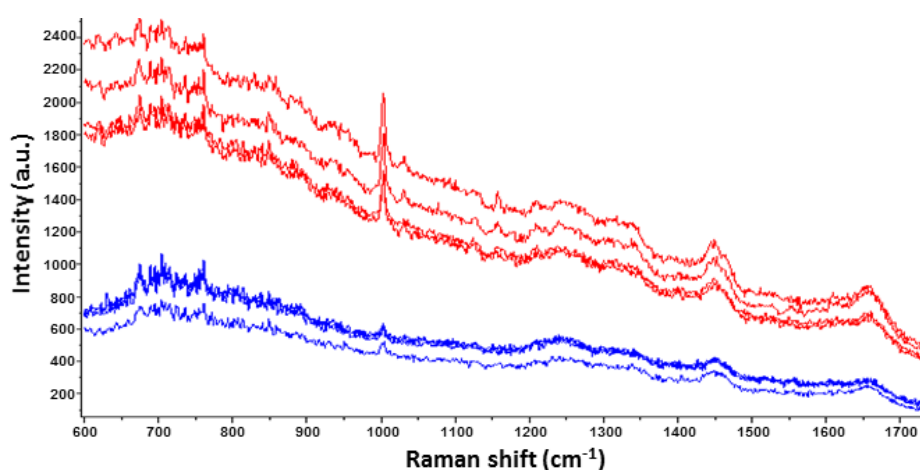
Figure 2.10 *Liquid serum Raman spectroscopy. Raw spectra collected from 30 µL pipetted in a well of an aluminium plate, on the surface of the liquid serum (blue) and 2 µm under the surface (red).*

Another option was assessed, 130 µL in a quartz cuvette were analysed at different depths without obtaining a satisfying spectral signal (Figure 2.11), probably because of a too thick cuvette walls and therefore a too important loss of laser energy going through the sample. Other parameters of acquisition could have been considered in order to improve the spectral signal at the expense of the spectral time of acquisition,

which was initially about 2 min per spectrum. However, it would not be efficient in terms of a clinical application.
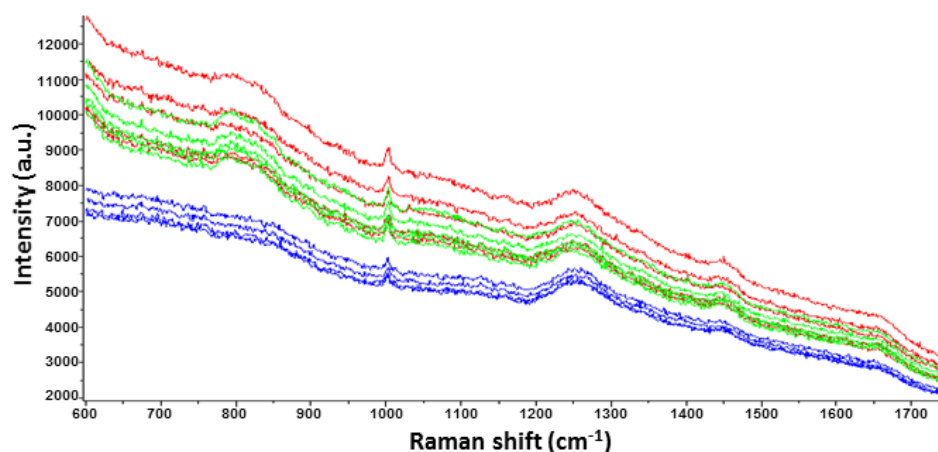
Figure 2.11 *Liquid serum Raman spectroscopy. Raw spectra collected from 130 μL in a quartz cuvette at different depths 0.8 (green), 1 (red) or 2 mm (blue).*

II.6 Conclusions

This study shows that spectra collected in different parts of a dried serum drop were affected by the thickness variations of the samples both in IR and Raman microspectroscopies, thus questioning spectral reproducibility. Moreover, the IR spectroscopic analysis has shown that protein bands are saturated in the ring formation, making any conclusion hazardous regarding potential molecular changes in periphery of serum drops. The predominant presence of high molecular weight proteins could mask other entities of diagnostic interest. Therefore, ultra-filtration to remove high molecular weight protein like HSA is a promising approach[23–26]. In this particular case, the concentration of the remaining low molecular weight constituents by the naturally occurring "coffee ring" effect could be an advantage as seen in a Raman spectroscopic study of tears, a biofluid 7 times less concentrated in proteins than serum[27]. Automated deposits of small serum amounts (nL) could be also a promising alternative, as it appears to diminish the impact of the "coffee ring" effect on spectra collection. The benefits of analysing thin films composed of nanoliter droplets in term of spectral reproducibility[28] has been reported. In order to meet clinical

expectations, High-throughput FTIR macrospectroscopy designed to obtain spectra from a multi-well silicon plate appears more appropriate for biofluid analysis, even if serum and plasma samples have to be diluted to avoid saturation[29,30]. Although Attenuated Total Reflection (ATR)-FTIR spectroscopy is not an automated technique, it allows to collect spectra in the same manner through a single reflection crystal, assuring spectral reproducibility and without sample dilution[26]. ATR-FTIR spectroscopy could also be used on liquid samples as well as Raman spectroscopy but as the drying process is a really dynamic phenomenon, spectra collection of a representative composition sampling on a wet drop is still questionable.

II.7 References

1   Y. Li, Q. Yang, M. Li and Y. Song, *Sci. Rep.*, 2016, **6**, 1–8.

2   W. B. Zeid and D. Brutin, *Colloids Surf. A Physicochem. Eng. Asp.*, 2013, **430**, 1–7.

3   W. Bou-Zeid and D. Brutin, *Colloids Surf. A Physicochem. Eng. Asp.*, 2014, **456**, 273–285.

4   W. D. Ristenpart, P. G. Kim, C. Domingues, J. Wan and H. A. Stone, *Phys. Rev. Lett.*, 2007, **99**, 1–4.

5   W. Xu, R. Leeladhar, Y. Tae Kang and C. Choi, *Langmuir*, 2013, **29**, 6032–6041.

6   H. Hu and R. G. Larson, *J. Phys. Chem. B*, 2006, **110**, 7090–7094.

7   Y. Tsoumpas, S. Dehaeck, A. Rednikov and P. Colinet, *Langmuir*, 2015, **31**, 13334–13340.

8   C. Diddens, J. G. M. Kuerten, C. W. M. van der Geld and H. M. A. Wijshoff, *J. Colloid Interface Sci.*, 2017, **487**, 426–436.

9   R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel and T. A. Witten, *Nature*, 1997, **389**, 827–829.

10   T. A. Yakhno, V. V Kazakov, A. G. Sanin, O. B. Shaposhnikova and A. S. Chernov, *Tech. Phys*, 2007, **52**, 515–520.

11   P. J. Yunker, T. Still, M. A. Lohr and A. G. Yodh, *Nature*, 2011, **476**, 308–311.

12   P. J. Yunker, M. A. Lohr, T. Still, A. Borodin and D. J. Durian, *Phys. Rev. Lett.*, 2013, **110**, 1–5.

13   V. Nanjappa, J. K. Thomas, A. Marimuthu, B. Muthusamy, A. Radhakrishnan,

R. Sharma, A. Ahmad Khan, L. Balakrishnan, N. A. Sahasrabuddhe, S. Kumar, B. N. Jhaveri, V. Sheth Kaushal, K. R. Kumar, P. G. Shaw, S. M. Srikanth, P. P. Mathur, S. Shankar, D. Nagaraja, R. Christopher, S. Mathivanan, R. Raju, R. Sirdeshmukh, A. Chatterjee, R. J. Simpson, H. C. Harsha, A. Pandey and T. S. K. Prasad, *Nucleic Acids Res.*, 2014, **42**, D959–D965.

14 A. di Masi, V. Trezza, L. Leboffe and P. Ascenzi, *J. Control. Release*, 2016, **228**, 191–205.

15 L. Vroman, A. L. Adams, G. C. Fischer and P. C. Munoz, *Blood*, 1980, **55**, 156–159.

16 M. Rabe, D. Verdes and S. Seeger, *Adv. Colloid Interface Sci.*, 2011, **162**, 87–106.

17 F. Poncin-Epaillard, T. Vrlinic, D. Debarnot, M. Mozetic, A. Coudreuse, G. Legeay, B. El Moualij and W. Zorzi, *J. Funct. Biomater.*, 2012, **3**, 528–543.

18 C. C. Annarelli, J. Fornazero, J. Bert and J. Colombani, *Eur. Phys. J. E*, 2001, **5**, 599–603.

19 W. Bou Zeid, J. Vicente and D. Brutin, *Colloids Surf. A Physicochem. Eng. Asp.*, 2013, **432**, 139–146.

20 B. Sobac and D. Brutin, *Colloids Surf. A Physicochem. Eng. Asp.*, 2014, **448**, 34–44.

21 K. Sefiane, *Adv. Colloid Interface Sci.*, 2014, **206**, 372–381.

22 J. N. Adkins, S. M. Varnum, K. J. Auberry, R. J. Moore, N. H. Angell, R. D. Smith, D. L. Springer and J. G. Pounds, *Mol. Cell. Proteomics*, 2002, **1**, 947–955.

23 R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads and T. D. Veenstra, *Mol. Cell. Proteomics*, 2003, **2**, 1096–1103.

24 C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N. W. Clarke and P. Gardner, *J. Biophotonics*, 2014, **7**, 180–188.

25 F. Bonnier, M. J. Baker and H. J. Byrne, *Anal. Methods*, 2014, **6**, 5155–1160.

26 J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D. Jenkinson, R. W. Lea, C. Walker and M. J. Baker, *J. Biophotonics*, 2014, **7**, 189–199.

27 J. Filik and N. Stone, *Anal. Chim. Acta*, 2008, **616**, 177–184.

28 J. Ollesch, S. L. Drees, M. H. Heise, T. Behrens, T. Brüning and K. Gerwert, *Analyst*, 2013, **138**, 4092–4102.

29 L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum and M. J. Baker, *Anal. Methods*, 2015, **7**, 7140–7149.

30 L. Lovergne, P. Bouzy, V. Untereiner, R. Garnotel, M. J. Baker, G. Thiéfin and G. D. Sockalingum, *Faraday Discuss.*, 2016, **187**, 521–537.

Chapter III. Investigating optimum sample preparation for infrared spectroscopic serum diagnostics

III.1 Recherche de la préparation d'échantillons optimale pour un diagnostic sérologique par spectroscopie infrarouge

III.1.1 Résumé

Les biofluides, tels que le sérum et le plasma, représentent un milieu idéal pour le diagnostic de maladies de par leur implication dans les fonctions physiologiques fondamentales chez l'Homme et leur facilité de collecte. La possibilité d'obtenir un diagnostic rapide avec une sensibilité et spécificité élevées est essentielle pour exploiter les avancées dans le développement de nouveaux traitements. De plus, la possibilité d'identifier rapidement une maladie sans équipement médicale imposant (IRM), permettrait le suivi rapproché d'un patient, réduisant ainsi la mortalité. De par ces raisons, la spectroscopie vibrationnelle appliquée aux biofluides a fait l'objet de nombreuses études pour évaluer son potentiel en tant qu'outil diagnostic et a montré des résultats prometteurs. Cependant, la préparation optimale des échantillons et la méthodologie la plus appropriée sont encore mal connues. Cette étude s'intéresse par conséquent à la répétabilité et reproductibilité des données spectrales acquises avec deux modes différents de spectroscopie infrarouge (IR), la réflectance totale atténuée (RTA) et la transmission, appliqués à l'analyse du sérum.

Dans un premier temps, les analyses ont été réalisées à partir d'un mélange commercialisé de sérums humains. Différents facteurs de dilutions (1/2, 1/3, 1/4, 1/5) ont été testés pour évaluer leur influence sur le dépôt final d'une goutte de sérum séché et sur les acquisitions spectrales associées par microimagerie et macrospectroscopie IR à haut débit. Dans un second temps, des échantillons de sérum de patients (n=70) collectés avant et après un acte chirurgical ont été analysés par spectroscopie IR en mode RTA et transmission.

III.1.2 Conclusions

L'analyse des spectres acquis sur des gouttes séchées de sérum par spectroscopie RTA et macrospectroscopie à haut débit a montré une bonne reproductibilité des données obtenues avec les deux approches. Dans le cas de la macrospectroscopie à haut débit réalisée en transmission, un facteur de dilution du sérum au tiers a été déterminé comme étant le plus approprié puisque des facteurs de dilutions inférieurs ne sont pas suffisants pour éviter la saturation des bandes spectrales associées aux protéines. A l'inverse des facteurs de dilution supérieurs pourraient diminuer de manière trop importante la contribution des molécules au faible poids moléculaire et présentes naturellement en faibles quantités dans le sérum. Ensuite, les deux approches comparées et associées à une analyse multivariée ont montré des capacités équivalentes dans la discrimination d'échantillons de patients collectés avant et après chirurgie. Cette réponse IR basale ne permet pas de distinguer complètement les échantillons mais est importante à prendre en compte car celle-ci reflète potentiellement l'inflammation due à l'acte de chirurgie et non à la maladie d'intérêt ciblée. Comparée à la macrospectroscopie à haut débit, la spectroscopie RTA ne nécessite pas de diluer les échantillons de sérum et des modèles de machines portables sont désormais commercialisés permettant une approche de diagnostic personnalisé. Cependant, la macrospectroscopie permet d'analyser un échantillon sur la surface totale d'un puits (3 mm de diamètre) limitant ainsi l'influence de l'effet ''rond de café'' et également de réaliser une analyse automatisée qui est un critère important pour la transposition de la technique en milieu clinique.

III.2 Chapter overview

It has been shown in the previous chapter that spectra collected from serum samples by microspectroscopy techniques were not reproducible due to thickness variations of a dried serum drop caused by the "coffee ring" effect. Thus, macroanalysis such as High-Throughput Fourier Transform Infrared (HT-FTIR) and Attenuated Total Reflection (ATR)-FTIR spectroscopies appear more appropriate for biofluid analysis. This study examines the repeatability and reproducibility of these two sampling modes and their associated serum sample preparation with spectral standard deviation of 0.0015 (post pre-processing) achievable for both sampling modes proving the collection of robust spectra. In addition, this study investigates the optimum serum sample dilution factor for use in high throughput transmission mode analysis with a 3-fold dilution proving optimum. This study exposes as well, the use of ATR and transmission mode spectroscopy to illuminate similar discriminatory differences in a patient study. This fundamental research provides proof of robust spectral collection that will be required to enable clinical translation of serum spectroscopic diagnostics.

III.3 Introduction

Rapid detection of diseases enables the early administration of a therapeutic strategy when the treatment is most effective, thus saving health expenditure and lives. For this purpose Fourier Transform Infrared (FTIR) spectroscopy is a suitable technique as it is non-destructive, label-free, rapid, cost-effective, easy to operate and requires simple sample preparation. Moreover, the use of serum spectroscopy for diagnostics has the advantage for patients to be relatively non-invasive compared to current diagnostic methodologies such as biopsies.

Over the past few years many studies have shown the successful use of vibrational spectroscopy for the analysis of biological samples[1,2]. Recent studies have investigated the potential of vibrational spectroscopy as a diagnostic tool for the analysis of biofluids. A pilot study on saliva using FTIR spectroscopy

associated with linear discriminant analysis (LDA), has reported the correct classification of diabetic patients from healthy control (overall accuracy of 100% on the training set and 88.2% on the validation test)[3]. The potential of human tears for the diagnosis of ocular diseases has been evaluated in preliminary studies, Travo *et al.* have shown significant discrimination (p<0.0001) of healthy people from patients with keratoconus (degenerative disorder affecting the cornea) and also between patients at an early or advanced stage of disease by high throughput (HT)-FTIR and principal component analysis (PCA)[4-6]. A study on synovial fluid by Raman spectroscopy and k-means analysis, has shown the discrimination of patients with osteoarthritis of low or high severity (74% sensitivity and 71% selectivity)[7]. Bile patient samples with malignant biliary strictures were differentiated from samples of patients with benign biliary diseases (sensitivity between 82% and 95%, specificity between 85% and 100%) by HT-FTIR spectroscopy in association with support vector machine (SVM) classification and leave-one-out cross validation (LOOVC)[8]. A Raman analysis combined with PCA and quadratic discriminant analysis (QDA) performed on urine has allowed the discrimination of patients suffering from diabetes mellitus and hypertension with low, high risk of kidney failure or with kidney failure (under dialysis) with a 70% overall classification rate (89% for the control group and 81% for both high risk and renal failure groups)[9]. However, the most popular biofluids analysed by vibrational spectroscopy are blood-derived products such as plasma[10-14] and serum[15-18]. Proof-of-principle studies have shown the ability of FTIR spectroscopy to identify different types of cancer from serum samples. HT-FTIR spectroscopy in transmission mode was used to discriminate urinary bladder cancer patients from patients with urinary tract infection with an accuracy between 85 and 92% achieved with LDA or random forest (RF) classifier[19], breast cancer from healthy donors with 96% sensitivity and 93% specificity with cluster analysis[20], cirrhotic patients with or without hepatocellular carcinoma with 87% sensitivity and 85% specificity by SVM-LOOCV[21]. Studies using attenuated total reflection (ATR)-FTIR spectroscopy, coupled with classification machine analysis obtained an accuracy of 95% and 82% for the discrimination of ovarian

and endometrial cancer[22] and allowed to achieve 87.5% sensitivity and 100% specificity for the differentiation of glioblastoma multiforme (GBM) *versus* healthy control as well as on average of 94% sensitivity and 96.5% specificity for the diagnosis of low grade gliomas and GBM *versus* healthy control[23,24].

Vibrational spectroscopy provides information on the composition and structure of matter. The principle of IR spectroscopy is based on the interaction of IR light with matter. Molecular bonds absorb the IR radiation at the resonant frequency of the bond or group, exciting vibrational modes. The resultant spectrum is a biochemical fingerprint of the analysed sample, each absorption peak/band corresponds to a specific vibration or combination of vibrations of a molecular bond. This absorption phenomenon obeys the Beer-Lambert law, thus allowing to obtain both quantitative and qualitative information[25]. In order to realise the full potential of IR spectroscopy as a leading healthcare tool, some issues need to be understood prior to clinical translation[26]. One issue in particular related to biofluid spectroscopy is due to the strong IR activity of water. As such, the most common protocol for analysing liquids such as biofluids is the drying of drop deposits. However, it has been shown by optical and spectroscopic assessment that this deposition is not homogenous[27-31]. Thus, reproducibility and reliability of drop-dried spectroscopic results are still questioned. This is even more pronounced when using microspectroscopy. In addition, when analysing biofluids with IR transmission mode it is common to dilute the sample, due to saturation of spectral peaks, with water prior to drying. The effect upon the spectrum and the optimum dilution is unknown and has not been previously investigated. In order to take full advantage of the potential of biofluid spectroscopy to provide a rapid and responsive diagnostic environment, this study reports the use of HT-FTIR, FTIR imaging and ATR-FTIR to investigate and understand the effect of sample preparation upon the IR spectrum and provide an optimum sample preparation for biofluid spectroscopy.

III.4 Materials and methods

III.4.1 Serum samples

Normal human mixed pooled serum purchased from TCS Biosciences (UK) was used for dilution tests. Serum samples (n=105) from 70 patients needing high risk surgery (e.g. liver and renal surgery, lung resection, large bowel resection, gastrectomy, Whipple's procedure) were provided by the Defence Science & Technology Laboratory (Dstl) Porton Down. The study has been approved by Southampton & South West Hampshire Research Ethics Committee (A), 1st Floor, Regents Park Surgery, Shirley, Southampton, Hants, SO16 4RJ (reference number 06/Q1702/152). A consent form was signed by all patients recruited into the study at the medical care centres involved (Liverpool Royal and University Hospital, University College Hospital, London, Bristol Royal Infirmary, Guy's & St Thomas' Hospital, London, Birmingham Queen Elizabeth Hospital and University Hospital, Frankfurt). Clotted blood samples (4mL) were centrifuged and the extracted serum samples were stored at -80°C until needed for analysis. The sera collected include 35 pre-operative samples, 35 post-operative samples from the same patients the day following surgery and 35 additional pre-operative samples from different patients. The study included 35 males and 32 females (the gender of 3 patients from records were not provided) between 21 and 79 years old with an average age of 60 years.

III.4.2 Sample preparation for HT-FTIR and FTIR imaging

Four dilutions (1/2, 1/3, 1/4, 1/5) were performed on normal pooled serum with physiological water (NaCl 0.9%, B. Braun). Five microliters per replicate and 10 replicates per sample including diluted, non-diluted normal serum and physiological water as control, were pipetted onto each well of a silicon plate (384 wells). Drops were dried at room temperature before the spectral data acquisition. Serum samples from patients were 3-fold diluted and 8 replicates were prepared using the same protocol.

III.4.3 Infrared spectroscopy and imaging

Spectra of diluted, non-diluted pooled serum, physiological water (30 spectra per condition, 180 spectra in total) and patient serum samples (840 spectra in total) were collected with a high-throughput screening HTS-XT extension coupled to a Tensor 27 spectrometer (both from Bruker Optics GmbH, Ettlingen, Germany). Spectral acquisition was performed in the transmission mode, controlled via OPUS v6.5 software (Bruker Optics GmbH, Ettlingen, Germany) and collected at a spectral resolution of 4 cm^{-1}, co-adding 32 scans over the spectral range of 4000-400 cm^{-1}. A background spectrum was acquired before each sample measurement and automatically removed from sample spectra. One spectrum is obtained over the whole surface of a well (~4 mm diameter) of the 384 well silicon plate used in the HTS-XT extension.

To study the drop heterogeneity, spectral images of the diluted serum drops were acquired with a PerkinElmer Spectrum Spotlight 400 imaging system (Courtaboeuf, France). Three spectral images were performed on non-diluted, diluted pooled serum and physiological water (18 spectral images in total) at a spatial resolution of 25 µm/pixel, a spectral resolution of 4 cm^{-1}, 8 co-added scans, and covering the spectral range of 4000-750 cm^{-1} (Spectrum-Image software, version 1.6, PerkinElmer).

Patient serum sample spectra were also collected using a Cary 600 series FTIR spectrometer from Agilent technologies, USA, equipped with a Pike Miracle ATR Accessory. After thawing, 1 µL of each sample was pipetted onto a germanium (Ge) crystal and dried for 8 minutes at room temperature. Three spectra were collected per biological replicate and 3 replicates performed for each sample (9 spectra per patient sample and 945 spectra for all samples in total). Spectra were acquired at a resolution of 4 cm^{-1}, over the spectral range of 4000-400 cm^{-1} and averaged over 32 co-added scans (Agilent software, Agilent technologies, USA). A background absorption spectrum of the crystal was measured for atmospheric correction before each replicate sample analysis and removed automatically from every sample spectrum. One average spectrum is obtained over the whole surface of the ATR crystal in contact with a sample.

The crystal was washed between each replicate analysis. Experimental processes are detailed in the experimental schematic shown in Figure 3.1.
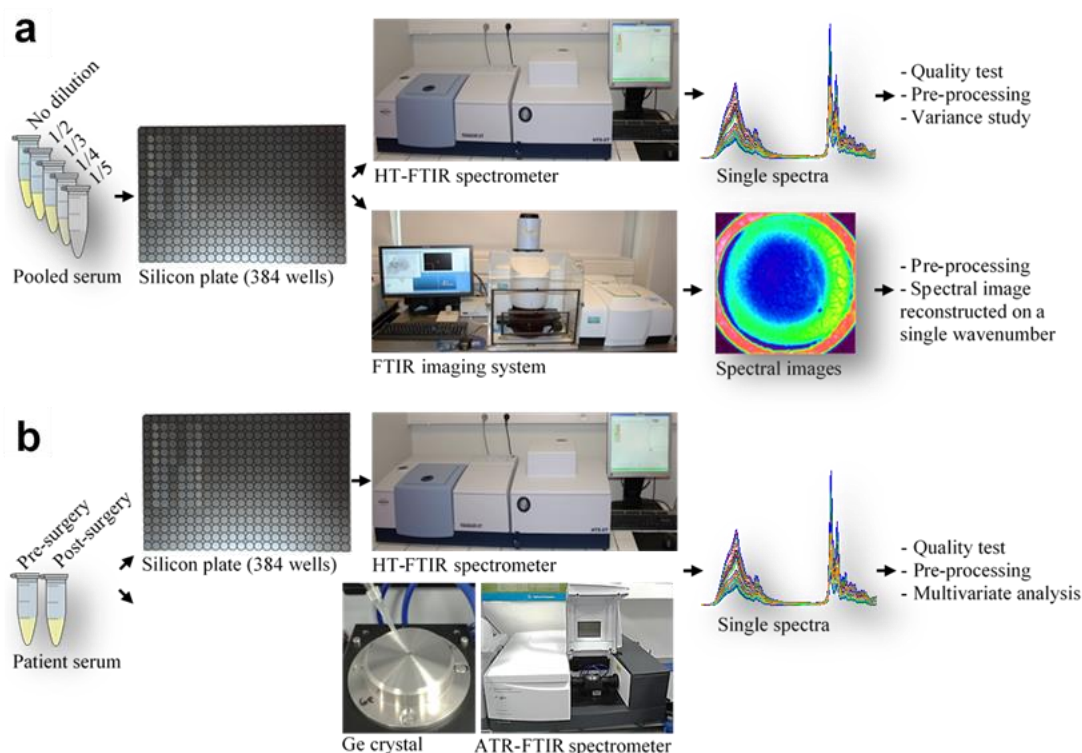
Figure 3.1 *Schematic methodology representation. **(a)** Dilution test study performed on human pooled serum by high throughput-Fourier transform infrared (HT-FTIR) spectroscopy and FTIR imaging. **(b)** HT-FTIR and total attenuated reflection (ATR)-FTIR spectroscopy comparison study performed on pre- (n=70) and post- (n=35) surgery patient serum samples. Ge: germanium.*

III.4.4 Pre-processing and data analysis

A quality test as previously described[21] was performed on raw spectra to control absorbance intensity, signal-to-noise ratio and water vapour content (OPUS v6.5 software, Bruker Optics GmbH, Ettlingen, Germany). Spectra with a maximum absorbance higher than 1.8 or lower than 0.35 in arbitrary units (a.u.) were rejected. Signal-to-noise ratio was calculated taking the maximum absorbance of the amide I band between 1700 and 1600 cm^{-1} (S1) and between 1260 and 1170 cm^{-1} (S2) respectively divided by the noise intensity evaluated between 2100-2000 cm^{-1} (N). Spectra were rejected when S1/N was less than 50 and S2/N was less than 10. Water vapour content (W) was evaluated between 1847-1837

cm^{-1}. Spectra were rejected when S1/W was less than 20 and S2/W was less than 4.

Following the quality test, data reproducibility was assessed on both raw and pre-processed spectra of diluted normal pooled serum *via* a variance study which consists in calculating the mean spectrum and the standard deviation (STD) for each sample dilution condition (Table 3.1). Pre-processing of HT-FTIR serum dilution test spectra was performed with OPUS v6.5 software (Bruker Optics GmbH, Ettlingen, Germany). Spectra were cut to the spectral range of 4000-900 cm^{-1}, baseline corrected and vector normalised.

Spectral images of serum drops were first subjected to an atmospheric correction for water vapour and CO_2 by the built-in Spectrum-Image software V1.6 (PerkinElmer, France) of the Spotlight 400 imaging system. Spectral images were then reconstructed, based on a specific chosen single wavenumber in order to study the wavenumber-related component distribution within drops (Spectrum-Image V1.6).

HT-FTIR and ATR-FTIR patient data pre-processing and multivariate analysis were realised with Matlab software version 7.11.0 (The Math Works, Inc., USA). All spectra were cut to the spectral range of 1800-900 cm^{-1} (fingerprint region). A principal component based noise reduction using the first 30 principal components (PCs) of the data was performed on the spectra to improve the signal-to-noise ratio. All spectra were then base-line corrected and vector normalised. Principal Component-Discriminant Function Analysis (PC-DFA) was performed on pre-processed data over the spectral region of 1200-950 cm^{-1}.

III.5 Results and discussion

III.5.1 Serum dilution effect

Four dilutions (2-fold, 3-fold, 4-fold and 5-fold) of serum samples were prepared with physiological water in order to determine the most appropriate

dilution for spectral acquisition with HT-FTIR spectroscopy as resulting spectra of non-diluted serum samples were saturated (Figure 3.2) at the protein amide I and amide II bands (1700-1500 cm^{-1}) and H-O-H stretching band at 3280 cm^{-1}. In the 2-fold dilution sample group, 60% of the raw spectra failed the quality test, mainly due to a high maximum absorbance (> 1.8 a.u.) observed in the amide I band. On the contrary with a 5-fold dilution, 10% of the spectra were rejected because of a too low maximum absorbance (<0.35 a.u.) of the amide I band. With the 3- and 4- fold dilution, all spectra passed the quality test with a maximum absorbance of 1.6 and 0.9 respectively.
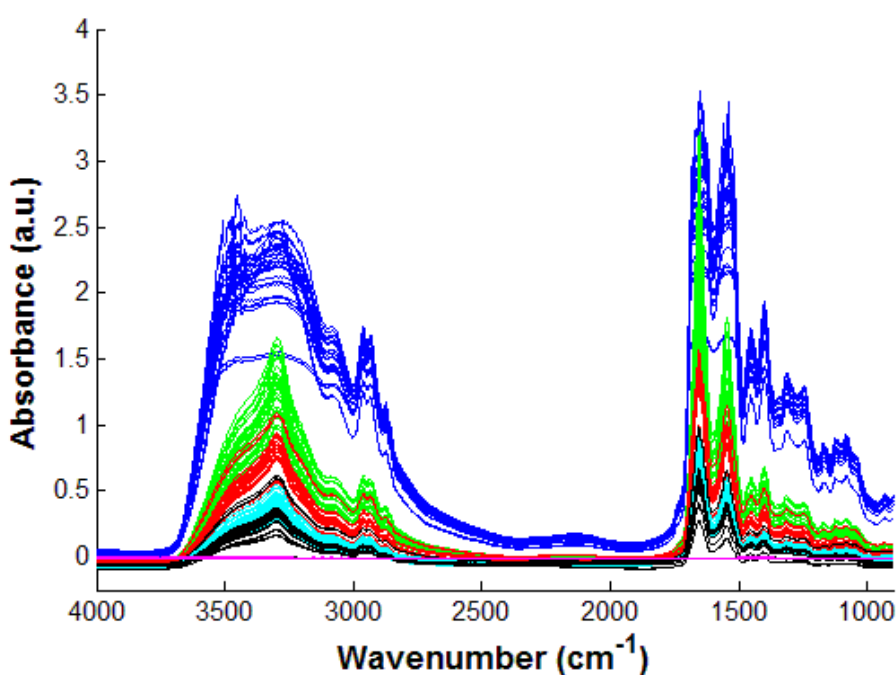
Figure 3.2 *HT-FTIR raw spectra of non-diluted human pooled serum (blue), diluted samples (2-fold: green; 3-fold: red; 4-fold: cyan; 5-fold: black) and physiological water as negative control (pink).*

A variance study on both raw spectra and pre-processed spectra was carried out on the spectra that passed the quality test. Standard deviations (STD) of spectra from each dilution (raw and pre-processed) are reported in Table 3.1. Pre-processed spectra presented a lowest average STD compared to their corresponding raw data (Figure 3.3). The pre-processing has improved the reproducibility of the spectra. The 2-fold diluted pre-processed spectra showed the lowest average STD due to the low number of remaining spectra after the quality test (40%). The 5-fold diluted pre-processed spectra presented the highest average STD (0.0022). The STD results of the 3 and 4 fold dilution are similar with both providing low average STD after pre-processing (0.0012 and 0.0013 respectively). Globally, the results show a good reproducibility of the spectra, similar to the reproducibility of human pooled serum spectra reported by Hands *et al.* via ATR-FTIR spectroscopy (average STD of 0.0137 and 0.0015 for raw and pre-processed spectra respectively)[24].

Table 3.1 *Quality test and variance study results performed on HT-FTIR spectra of normal pooled serum with different dilutions.*

Dilution	Right spectra after quality test (%)	Mean STD (raw spectra)	Mean STD (pre-processed spectra)
1/2	40	0. 015	0. 0004
1/3	100	0. 0474	0. 0012
1/4	100	0. 02	0. 0013
1/5	90	0. 0342	0. 0022

Combining the quality test and the variance study results, the 3-fold dilution was determined as the most suitable dilution for a serum HT-FTIR analysis with a good reproducibility and absorbance intensity. The 3-fold dilution was chosen as a higher dilution (e.g. 4-fold) will induce the risk of a low spectral response, especially for the analysis of molecules present at a low concentration within the serum. No molecular changes were visible on collected spectra, assuming that there was no effect of the dilution with water on the composition of the serum. Similar observations have been reported in a study by Raman spectroscopy on plasma dilution tests. They concluded that the plasma dilution did not affect the chemical composition and molecular structure of proteins within dried drops[30].
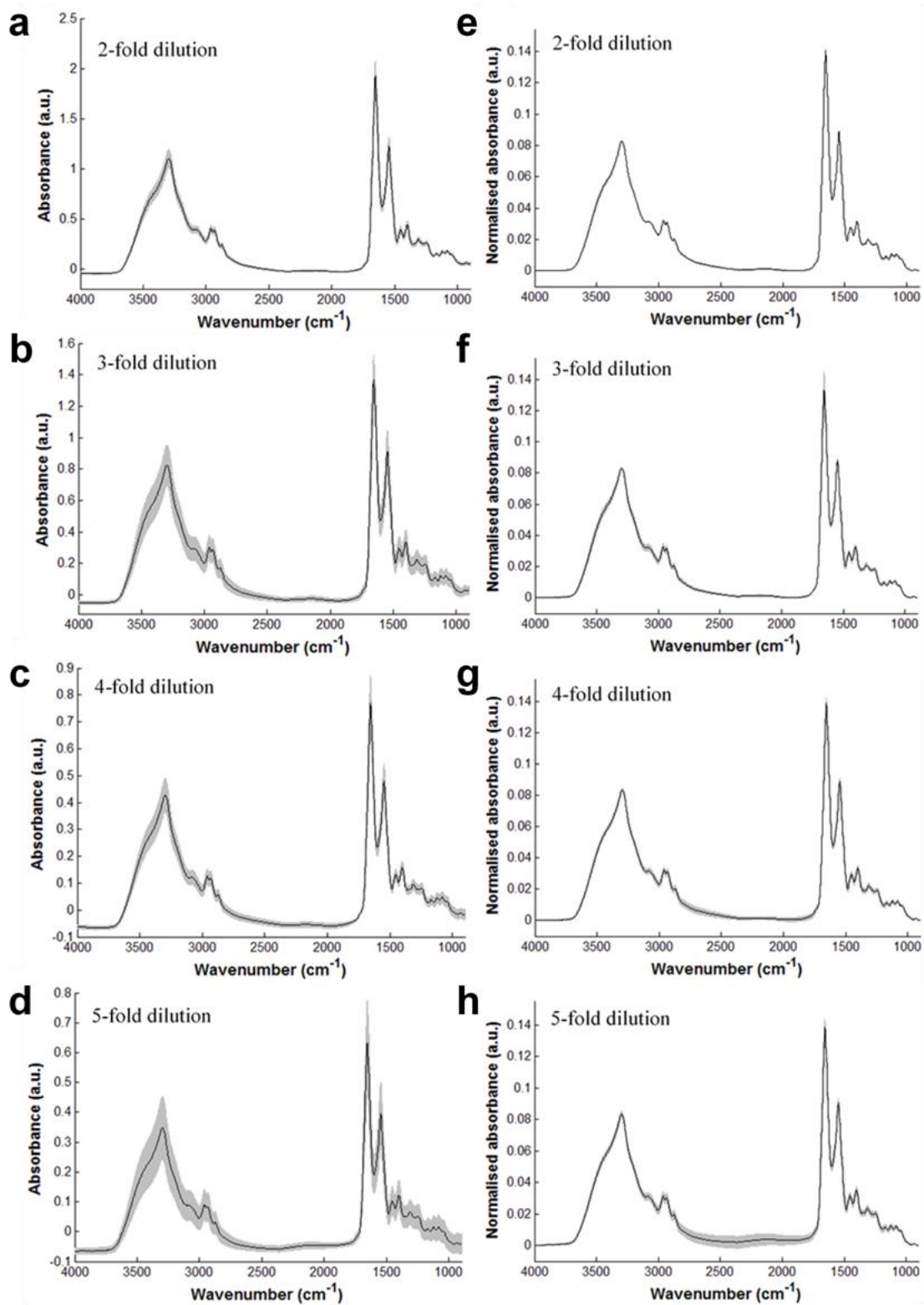
Figure 3.3 *Variance study performed on HT-FTIR raw (**a; b; c; d**) and pre-processed (**e; f; g; h**) mean spectra (black) of diluted human pooled serum. Standard deviation is represented in grey.*

Spectral images were acquired to better understand the distribution of serum molecular components within dried drops as physical differences were observed between non-diluted (presence of cracks) and diluted serum samples (hollow shape) suggesting a drop composition heterogeneity (Figure 3.4, white light images). It can be seen on the total absorbance spectral image of the droplets that the absorbance intensity varies within the same drop, with a highest absorbance intensity in the drop edge compared to the centre. This translates as a concentration gradient of serum components from the centre to the periphery of the drops, showing an uneven drying process across the serum droplets. As expected, a global absorbance intensity decrease was observed for diluted samples compared to non-diluted samples (Figure 3.4, total absorbance spectral images). Spectral images based on a single wavenumber reconstruction enabled the determination, within a drop, of the distribution of molecules associated with the targeted peak. It was observed (Figure 3.4, reconstructed spectral images) that serum components such as proteins (amide I and amide II bands around 1647 cm^{-1} and 1542 cm^{-1}), lipids (1442 cm^{-1}) and nucleic acids (1078 cm^{-1}) have a tendency to migrate towards the periphery of the drops. This phenomenon is referenced in the literature as the well-known coffee-ring effect[32-34]. This effect was more visible when the dilution factor increased, the ring appearing thinner at the drops' edges and the centres appearing more hollowed compared to non-diluted serum samples (Figure 3.4, white light images). The deposit and the resulting drop morphology has also been shown to be influenced by the substrate hydrophobicity but without impact on the chemical composition of samples[30]. To avoid the coffee-ring effect, a study has recommended to perform an analysis on a film composed of reduced-size drops (217 drops of 200 pL each per well of a 384 wells silicon plate) obtained by an automatic sampling and by controlling the drying process (vacuum drying). They obtained a higher reproducibility of spectral data compared to a non-automatic sampling[35].
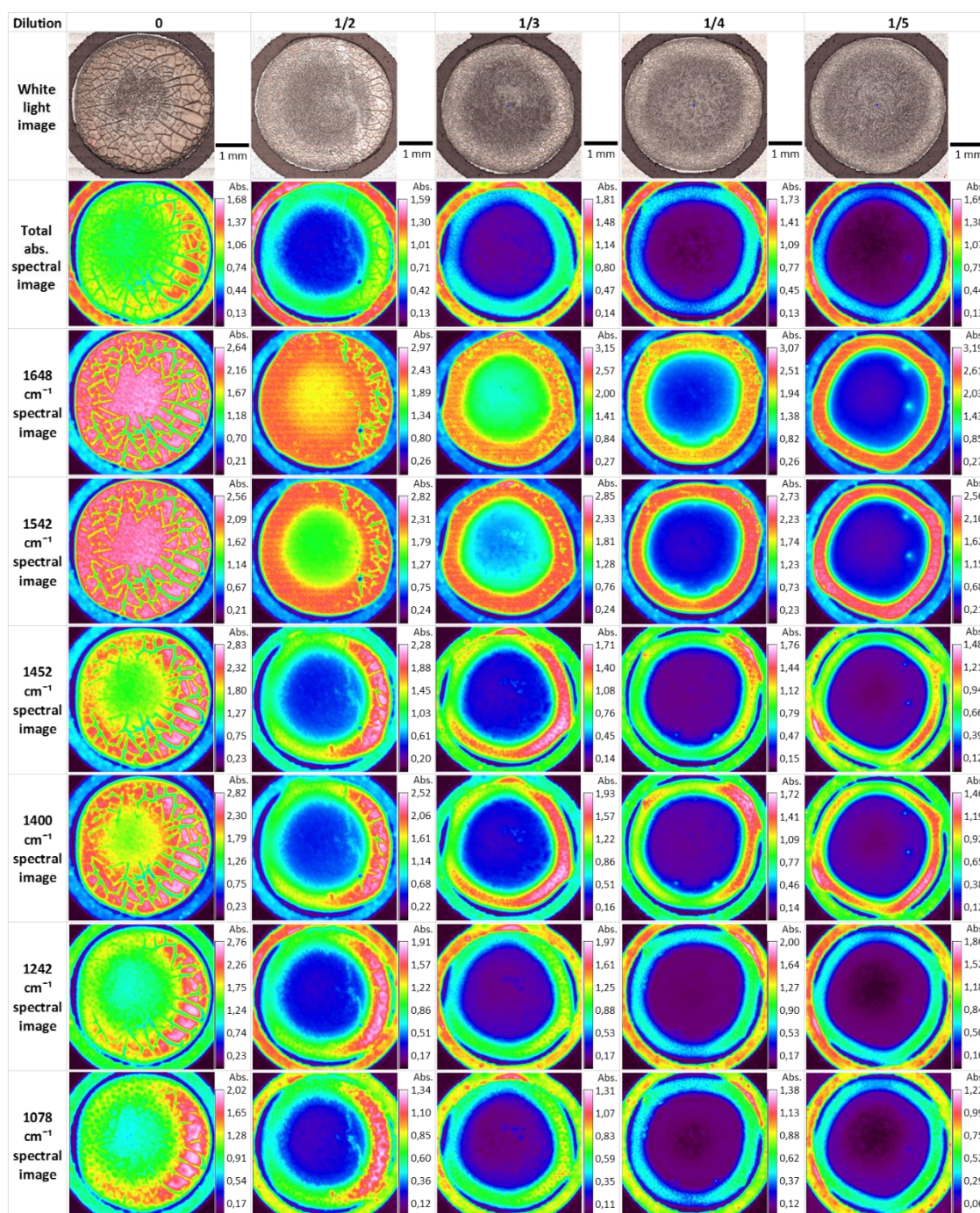
Figure 3.4 *Serum dilution study by FTIR imaging of dried drops. Original spectral images were atmospheric corrected. Other spectral images were reconstructed on a single wavenumber. Abs: absorbance.*

The presence of cracks has been observed on non-diluted sample drops and was less visible/absent on diluted samples (Figure 3.4, white light images). It has been reported that spectra acquired from sample areas containing cracks suffer from spectral distortions resulting in an abnormal amide I and II bands and baseline[29]. In this study,

the dilution of the serum samples has provided visually a dried drop with greater homogeneity over the surface. Moreover, one averaged spectrum is acquired on the whole surface of a well with the HT-FTIR technique, resulting in a representative spectrum of the entire serum sample contrary to spectra collected in a single point mode as serum dried drops spectra have shown spatial but also chemical heterogeneity[28].

III.5.2 ATR-FTIR and HT-FTIR spectroscopy comparison

An analysis was carried out on pre- and post-operative serum samples from patients having undergone surgery for different clinical reasons. This set of samples (n=105) was used to compare the two techniques ATR-FTIR and HT-FTIR spectroscopy and also to enable the profiling of a baseline IR response when considering only surgery. ATR-FTIR is an excellent technique for the analysis of serum since a sample can be directly deposited and dried on the ATR crystal before the spectral acquisition[36]. However, there are currently no multi-well sample accessories to enable rapid automated collection of spectra. HT-FTIR enables the rapid automated collection of spectra but there is an increased requirement for further sample preparation via sample dilution. As such it takes 6 hours to collect 360 spectra from 360 samples on a multi-well plate via HT-FTIR spectroscopy and in the same amount of time 40 spectra would have been able to be collected via the current ATR-FTIR approach (the addition sample preparation time for the HT-FTIR takes approximately 1 hour for the deposit of 360 samples and 30 minutes for the samples drying). The sample preparation regarding the analysis of serum by ATR-FTIR has been reproduced from the methodology applied by Hands *et al.*[24]. They have determined that a serum drop of 1 µL is dry after 8 minutes at room temperature and that the resultant spectra acquired exhibited a high reproducibility. In the present study, spectra from human serum samples were collected from 1 µL and 5 µL (3-fold diluted serum) for the ATR-FTIR and HT-FTIR spectroscopy respectively. These spectra were reproducible and repeatable thus conserving the serum resource by using minute

amounts. Spectral differences are visible on the average spectra of the two pre- and post-operative sample groups especially between 950 and 1200 cm^{-1} for both ATR-FTIR and HT-FTIR techniques, further differences are discernable between 1250 and 1700 cm^{-1} but only with the ATR-FTIR method (Figures 3.5a and 3.5d). A different ratio and a peak shift is observable for the amide I and amide II peaks between ATR and transmission sampling modes (Figures 3.5a, 3.5d and 3.5g). This is similar to the ratio and peak shift differences observed by Ollesch *et al.*[35] on dried serum sample comparison between ATR and transmission due to dispersion effects. Associated with PC-DFA, both ATR-FTIR and HT-FTIR spectroscopies show similar scores plots and ability to discriminate serum spectra over the range of 1200-950 cm^{-1} of patients prior to surgery and serum spectra after surgery in spite of similarities resulting in an overlap. Discriminant Function 1 (DF1) in both cases is responsible for the separation of the two spectral groups (Figures 3.5b and 3.5e) and DFs along the y-axis are only shown for visual purposes. Loadings on DF1 contain only the spectral information that is responsible for the discrimination between spectral groups (Figure 3.5c and 3.5f). With both techniques, the same peaks have been identified but have opposite directions on the loadings plots. Positive peaks in DF1 loadings for the ATR-FTIR spectroscopy analysis are negative in the DF1 loadings for the HT-FTIR spectroscopy study and vice versa. Spectra from the two groups are also separated in an inverse way against the *x*-axis of the scatter plots between the two techniques. Proposed peaks assignments are presented in Table 3.2. Major peaks of the loadings on DF1 were in both cases 1040-1023 cm^{-1} corresponding to glycogen and 1078 cm^{-1} assigned as C–O stretch (DNA/RNA). It can be noted that a complete separation of the two groups by PC-DFA was not expected since it is known that a surgical procedure triggered physiological and immunological reactions in order to maintain the body homeostasis but the stress induced by surgery and the associated body responses are dependent on many factors such as the kind of procedure and its relative injury severity, pre-existing pathologies, genetic predisposition of patients and the difference observed will not be as significant as diseased *versus* non-diseased patient samples[37-39].
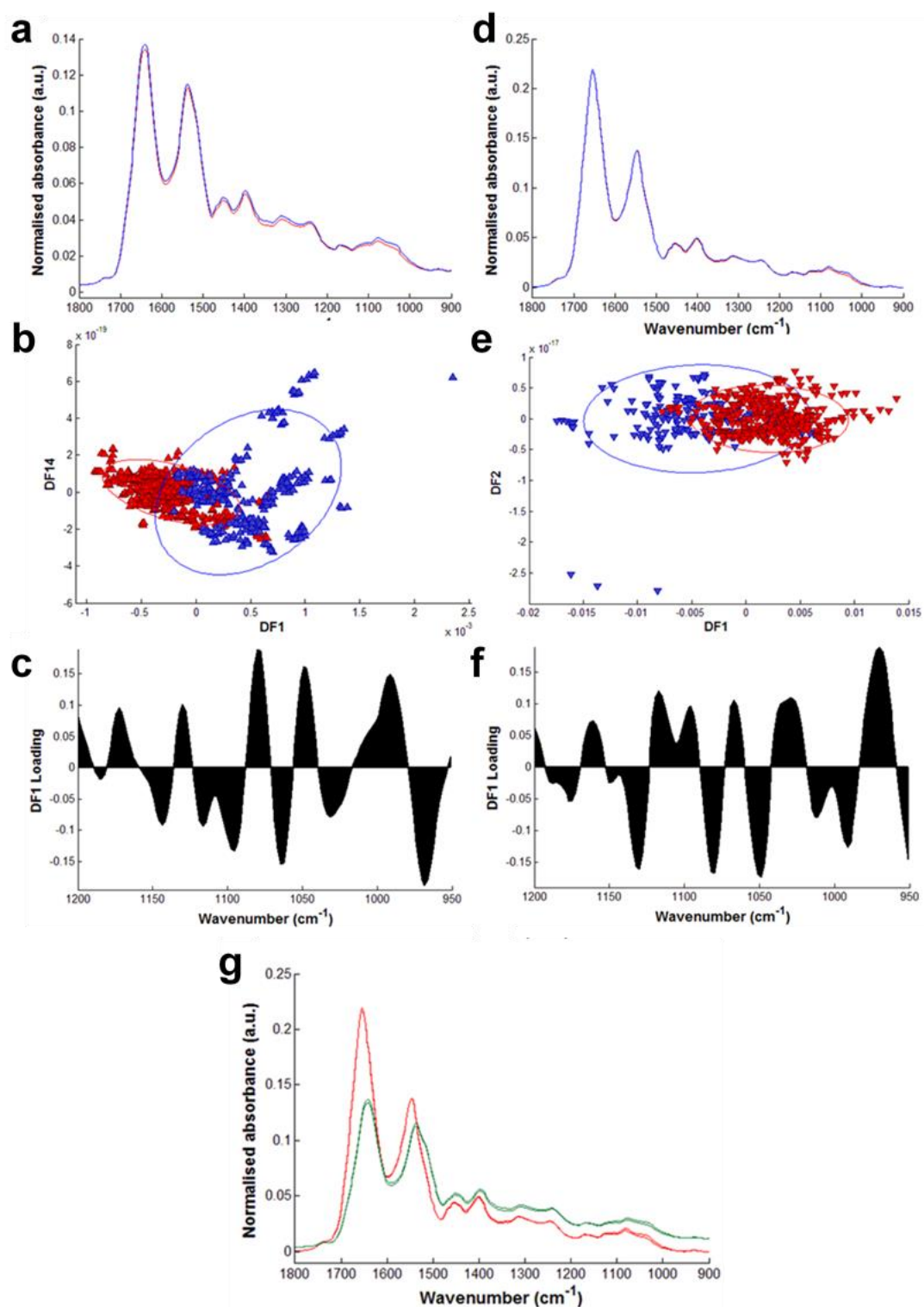
Figure 3.5 *ATR-FTIR and HT-FTIR comparison study on pre and post-surgery serum samples. ATR-FTIR (**a**) and HT-FTIR (**d**) pre-processed mean spectra of patient serum prior to surgery in red and after surgery in blue. ATR-FTIR (**b**) and HT-FTIR (**e**) PC-DFA scatter plot, pre-processed serum spectra of patients prior to surgery are represented by red triangles and after surgery by blue triangles over the spectral range of 1200-950 cm⁻¹ with their respective 90% confidence ellipses. ATR-FTIR (**c**) and HT-FTIR (**f**) corresponding DF1 loading from whole patient serum. (**g**) HT-FTIR and ATR-FTIR spectra (from a and d) in red and green respectively.*

Table 3.2 *HT-FTIR and ATR-FTIR PC 1 loading peaks assignments*[1, 23, 24, 40, 41]

Frequency (cm⁻¹)	Assignments
1170	Ester C-O asymmetric stretch
1150	C-O stretch (carbohydrates)
1125	C-C asymmetric stretch
1110	CN stretch, CH bend
1100	PO_2^- symmetric stretch (nucleic acid)
1078	C-O stretch (deoxyribose/DNA-RNA)
1065	Chain C-C
1050	C-O stretch
1023-1040	Glycogen
1006	Phenylalanine (ring breathing)
990	CO-O-C stretch
970	C-C, CN stretch, PO_3^{2-} stretch (DNA)

III.6 Conclusions

HT-FTIR and ATR-FTIR techniques have both shown promising results for serum dried-drops analysis with a high reproducibility of the spectra collected on normal human pooled mixed serum samples. Moreover, both approaches associated with multivariate analysis have shown an equivalent capability to distinguish spectra of patient serum collected before and after surgical procedures. This basal IR response to surgery should be taken into consideration when performing a diagnostic study as molecular changes observed in spectra might be due to the surgery/inflammation itself and not to the pathology targeted, however this study has shown that these responses are not significant enough to enable full differentiation of serum samples. Compared to the HT-FTIR technique where serum samples have to be diluted because of a saturation effect, with the ATR-FTIR method, samples can be directly applied onto the ATR crystal and can be portable. However, this implies no possible sample storage for further analysis as the crystal has to be cleaned after each sample measurement and no automation of the analysis is yet available. Regarding these aspects, with the HT-FTIR technique the storage of samples is in theory feasible but no studies have investigated the effect of the conservation over the time of

serum properties integrity. A 3-fold dilution was shown to be optimum based upon a variance study, quality test and the ability to detect low concentration biochemicals present compared to higher dilutions and is the recommendation for HT-FTIR analysis. HT-FTIR spectroscopy has the advantage to allow an automated analysis that is an important criterion for a transposition of IR spectroscopy into a clinical application. This automation advantage can be seen in the time comparison with HT-FTIR being able to collect 360 spectra, in the same time it would allow to collect 40 spectra in the current ATR-FTIR approach. An automated approach for ATR-FTIR analysis would be beneficial as there is a comparable spectral reproducibility between ATR-FTIR and HT-FTIR spectroscopy and a comparable ability to discriminate and illuminate differences in patient spectra but with an automated ATR-FTIR approach, there would not be a need for dilution.

III.7 References

1   M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nat. Protoc.,* 2014, **9**, 1771–1791.

2   D. I. Ellis and R. Goodacre, *Analyst*, 2006, **131**, 875–885.

3   D. A. Scott, D. E. Renaud, S. Krishnasamy, P. Meriç, N. Buduneli, S. Çetinkalp and K. Liu, *Diabetol. Metab. Syndr.,* 2010, **2**, 48–56.

4   J. Filik and N. Stone, *Anal. Chim. Acta*, 2008, **616**, 177–184.

5   C. Lin, K. Ming-Tse and C. Hsien-Chang, *J. Med. Biol.*, 2010, **30**, 343–354.

6   A. Travo, C. Paya, G. Déléris, J. Colin, B. Mortemousque and I. Forfar, *Anal. Bioanal. Chem.*, 2014, **406**, 2367–2376.

7   K. A. Esmonde-White, G. S. Mandair, F. Raaii, J. A. Jacobson, B. S. Miller, A.G. Urquhart, B. J. Roessler, and M. D. Morris, *J. Biomed. Opt*., 2009, **14**, 034013.

8   V. Untereiner, G. D. Sockalingum, R. Garnotel, C. Gobinet, F. Ramaholimihaso, F. Ehrhard, M. Diebold and Gérard Thiéfin, *J. Biophotonics*, 2014, **7**, 241–253.

9   J. A. Bispo, E. E. De Sousa Vieira, L. Silveira and A. B. Fernandes, *J. Biomed. Opt.*, 2013, **18**, 087004.

10   V. Erukhimovitch, M. Talyshinsky, Y. Souprun and M. Huleihel, *Vib. Spectrosc.*, 2006, **40**, 40–46.

11   K. Singh, K. Lee, D. Lee, Y. K. Kim and K. C. Kim, *J. Mech. Sci. Technol*, 2010, **24**, 1661–1668.

12   K. W. C. Poon, F. M. Lyng, P. Knief, O. Howe, A. D. Meade, J. F. Curtin, H. J. Byrne and J. Vaughan, *Analyst*, 2012, **137**, 1807–1814.

13   U. Neugebauer, S. Trenkmann, T. Bocklitz, D. Schmerler, M. Kiehntopf and Jürgen Popp, *J. Biophotonics*, 2014, **7**, 232–240.

14   C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, *Analyst*, 2015, **140**, 2280–2286.

15   J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez and C. Medina-Gutiérrez, *Laser. Med. Sci.*, 2007, **22**, 229–236.

16   W. Petrich, K. B. Lewandrowski, J. B. Muhlestein, M. E. H. Hammond, J. L. Januzzi, E. L. Lewandrowski, R. R. Pearson, B. Dolenko, J. Früh, M. Haass, M. M. Hirschl, W. Köhler, R. Mischler, J. Möcks, J. Ordóñnez–Llanos, O. Quarder, R. Somorjai, A. Staib, C. Sylvén, G. Werner and R. Zerback, *Analyst*, 2009, **134**, 1092–1098.

17   E. Scaglia, G. D. Sockalingum, J. Schmitt, C. Gobinet, N. Schneider, M. Manfait and G. Thiéfin, *Anal. Bioanal. Chem.*, 2011, **401**, 2919–2925.

18   M. Khanmohammadi, A. B. Garmarudi, M. Ramin and K. Ghasemi, *Microchem. J.,* 2013, **106**, 67–72.

19   J. Ollesch, M. Heinze, H. M. Heise, T. Behrens, T. Brüning and K. Gerwert, *J. Biophotonics*, 2014, **7**, 210–221.

20   J. Backhaus, R. Mueller, N. Formanski, N. Szlama, H. G. Meerpohl, M. Eidt and P. Bugert, *Vib. Spec.*, 2010, **52**, 173–177.

21   X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon and G. D. Sosckalingum, *Transl. Res.*, 2013, **162**, 279–286.

22   K. Gajjar, J. Trevisan, G. Owens, P. J. Keating, N. J. Wood, H. F. Stringfellow, P. L. Martin-Hirsch and F. L. Martin, *Analyst*, **138**, 3917–3926.

23   J. R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R. W. Lea, A. J. S. McIntosh and M. J. Baker, *Anal. Bioanal. Chem*, 2013, **405**, 7347–7355.

24   J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D. Jenkinson, R. W. Lea, C. Walker and M. J. Baker, *J. Biophotonics*, 2014, **7**, 189–199.

25   *Handbook of intrumental techniques for analytical chemistry*, ed. F. A. Settle, Prentice Hall PTR, 1997, ISBN 0-13-177338-0.

26   H. J. Byrne, M. Baranska, G. J. Puppels, N. Stone, B. Wood, K. M. Gough, P. Lasch, P. Heraud, J. Sulé-Suso and Ganesh D. Sockalingum, *Analyst*, 2015, **140**, 2066–2073.

27   J. Filik and N. Stone, *Analyst*, 2007, **616**, 177–184.

28   F. Bonnier, F. Petitjean, M. J. Baker and H. J. Byrne, *J. Biophotonics*, 2014, **7**, 167–179.

29   C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N. W. Clarke and P. Gardner, *J. Biophotonics*, 2014, **7**, 180–188.

30   K. A. Esmonde-White, F. W. L. Esmonde-White, M. D. Morris and B. J. Roessler, *Analyst*, 2014, **139**, 2734–2741.

31   I. Taleb, G. Thiéfin, C. Gobinet, V. Untereiner, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Hillon, M. Manfait, P. Ducoroy and G. D. Sockalingum, *Analyst*, 2013, **138**, 4006–4014.

32   R. D. Deegan, T. F. Dupont, G. Huber, S. R. Nagel and T. A. Witten, *Nature*, 1997, **389**, 827–829.

33   W. Ristenpart, P. Kim, C. Domingues, J. Wan and H. Stone, *Phys. Rev. Lett.*, 2007, **99**, 234502.

34   P. J. Yunker, T. Still, M. A. Lohr and A. G. Yodh, *Nature*, 2011, **476**, 308–311.

35   J. Ollesch, S. L. Drees, H. M. Heise, T. Behrens, T. Brüning and K. Gerwert, *Analyst*, 2013, **138**, 4092–4102.

36   K. M. Dorling and M. J. Baker, *Trends Biotechnol.*, 2013, **31**, 437–438.

37   E. Lin, S. E. Calvano and S. F. Lowry, *Surgery*, 2000, **127**, 117–126.

38   P. V. Giannoudis, H. Dinopoulos, B. Chalidis and G. M. Hall, *Injury*, 2006, **37S**, S3–S9.

39   T. Gutierrez, R. Hornigold and A. Pearce, *Surgery*, 2011, **29**, 93–96.

40   G. L. Owens, K. Gajjar, J. Trevisan, S. W. Fogarty, S. E. Taylor, B. Da Gama-Rose, P. L. Martin-Hirsch and F. L. Martin, *J. Biophotonics*, 2014, **7**, 200–209.

41   F. Bonnier, M. J. Baker and H. J. Byrne, *Anal. Methods*, 2014, **6**, 5155–5160.

Chapter IV. Biofluid infrared spectro-diagnostics: pre-analytical considerations for clinical applications

IV.1 Diagnostics à partir de biofluides et par spectroscopie infrarouge: considérations pré-analytiques en vue d'une application clinique

IV.1.1 Résumé

De nombreuses études au stade preuve de concept sur l'application de la spectroscopie vibrationnelle à des biofluides ont démontré que la technique a un potentiel prometteur en tant que futur outil diagnostic clinique. Cependant, ces études montrent aussi le manque de standardisation des protocoles dans la manipulation et la préparation des échantillons avant l'analyse spectroscopique. La source des erreurs analytiques la plus courante est la phase pré-analytique. Dans le but de transférer la technique en milieu clinique, il est évident qu'un protocole strict et standard doit être établi pour de tels échantillons biologiques. Cette étude porte sur certains aspects de la phase pré-analytique dans le développement de l'application de la spectroscopie infrarouge à transformée de Fourier aux biofluides les plus communs tels que le sérum, le plasma ou la bile. Des considérations pré-analytiques qui peuvent avoir un impact sur à la fois les échantillons (solvants, anticoagulants, cycles de congélations/décongélations) et/ou l'analyse spectroscopique (préparation d'échantillons tels que le séchage, la méthode de déposition, les volumes, les substrats, la dépendance opérateurs) et en conséquence sur la qualité et la reproductibilité des données spectrales sont discutés dans cette étude.

Les analyses ont été menées à partir d'échantillons de sérum, plasma et bile de patients. Une dilution au tiers des échantillons de sérum et de plasma a été déterminée la plus appropriée avec une solution saline pour éviter la précipitation des protéines contrairement à une dilution avec de l'eau distillée. La bile étant naturellement moins concentrée en protéines, celle-ci ne nécessite pas d'être diluée avant analyse. Les dépôts de ces 3 types d'échantillons sans étalement sont préférables pour la reproductibilité des résultats. Un séchage d'au moins 90 min est requis avant analyse dans le cas d'une utilisation d'une plaque de silicium 384 puits. La collecte du sang pour l'obtention du plasma est recommandée avec des tubes contenant de l'héparine de lithium car l'acide éthylènediaminetétraacétique, autre anticoagulant

communément employé pour l'obtention du plasma, interfère avec le signal du plasma par la présence de pics supplémentaires. Il a été démontré également que des spectres acquis à partir d'échantillons frais de sérum ou plasma pouvaient être distingués de spectres collectés à partir d'échantillons congelés puis décongelés. Par ailleurs, les spectres d'échantillons ayant subi jusqu'à 5 cycles de congélations/décongélations ne se distinguent pas, suggérant un faible impact de décongélations répétées sur la composition de l'échantillon détectable par spectroscopie. Les expériences menées ont démontré que la collection spectrale n'était pas dépendante des expérimentateurs, et est reproductible de jour en jour.

IV.1.2 Conclusions

Le développement de toute technique analytique de laboratoire, incluant la spectroscopie et ayant le potentiel d'être transférée en milieu clinique requiert une évaluation systématique des potentiels facteurs pré-, intra et post-phase analytique qui peuvent avoir un impact sur la qualité des données spectrales et sur l'interprétation générale des résultats. De plus, de futures recherches seront nécessaires pour évaluer les variabilités intra et inter centres d'étude, de même que d'autres facteurs de la phase pré-analytique tels que la reproductibilité entre les mêmes appareils d'un fabricant. Cette étude est centrée sur l'application de la spectroscopie infrarouge à haut débit en mode transmission, il serait également intéressant d'effectuer le même type d'études avec d'autres modalités de spectroscopie vibrationnelle.

IV.2 Chapter overview

Several proof-of-concept studies on vibrational spectroscopy of biofluids have demonstrated that the technique has promising potentials as a clinical diagnostic tool. However, these studies also show that there is lack of a standardised protocol in sample handling and preparation prior to spectroscopic analysis. One of the most important sources of analytical errors is the pre-analytical phase. For the technique to be translated into clinics, it is clear that a very strict protocol needs to be established for such biological samples. This study focuses on the development of High-Throughput Fourier Transform Infrared (HT-FTIR) spectroscopy which was found the most appropriate approach for serum analysis among the different modalities tested in the two previous chapters. Different aspects of the pre-analytical phase of serum analysis and some of the most common biofluids such as plasma and bile are investigated. Pre-analytical considerations that can impact either the samples (solvents, anti-coagulants, freeze-thaw cycles) and/or spectroscopic analysis (sample preparation such as drying, deposit methods, volumes, substrates, operator dependence) and consequently on the quality and reproducibility of spectral data will be discussed in this chapter.

IV.3 Introduction

During the last decade, a number of studies have highlighted the potential of vibrational spectroscopy applied to biofluids for screening, diagnostic and prognostic applications in the biomedical field[1–9]. These studies exploited the fact that biofluids exhibit spectral characteristics reflecting their biomolecular composition, which allows, through chemometric analysis, the identification of patterns reflecting sensitive and specific spectral biomarkers in various pathological conditions. Some biofluids such as serum and plasma appear as ideal media for routine clinical use as they are easily accessible, collectable by minimally invasive method and repeatedly available for monitoring disease progression or therapeutic response. Accordingly, most clinical studies have been performed using blood components and the quest for spectral blood biomarkers has been widely explored in patients with malignant and non-malignant diseases, as reviewed recently by Baker *et al.*[10]. Other biofluids such as bile or

cerebrospinal fluid are less accessible but of particular interest as their proximity with diseased tissue may facilitate the identification of the spectral signatures of the disease[11,12]. Based on these data, biofluid vibrational spectroscopy appears as an innovative and promising facet of the concept of "liquid biopsy".

However, in spite of promising results, biofluid vibrational spectroscopy is still far from the routine clinical application. The great majority of studies are proof-of-concept studies based on a small number of subjects with and without the disease of interest. The results of these pilot studies should be analysed by cross validation methods and interpreted cautiously until validated in clinical trials performed in independent cohorts with a large number of patients and using golden standard diagnostic methods or complementary techniques. Most importantly, at the time where multicenter studies are being considered, there is a crucial need to standardise and validate the modalities of sample collection, storage and spectral acquisition. Standardisation of sample-related factors is the first step of pre-analytical validity of any technology intended to large scale and multicenter clinical applications. It is well-established that most errors in laboratory testing occur in the pre-analytical phase[13] and the need for standardisation of the procedures has been demonstrated in other high-throughput technologies such as proteomics[14,15], metabolomics[16,17] or genomics[18].

The work presented here will address the issue of pre-analytical variability in the field of high-throughput biofluid infrared transmission spectroscopy with a particular focus on blood components preparation. Factors that may affect the IR spectral quality and profile of biofluids will be discussed including biofluid dilution, volume and deposition modalities, repetition of freeze-thaw cycles, drying conditions, types of blood collection tubes used, intra- and inter-operator reproducibility.

IV.4 Experimental

The studies were performed from a bank of serum and plasma samples stored at -80°C, originally taken for routine biochemical check-up at the Biochemistry Laboratory of Reims University Hospital. Serum was obtained by centrifugation of freshly clotted

blood and plasma by centrifugation of blood collected in tubes containing ethylenediaminetetraacetic acid (EDTA) or lithium heparin as anticoagulants. Bile was obtained during endoscopic retrograde cholangiopancreatography after selective biliary cannulation and before any injection of contrast agent. After centrifugation, the cellular pellet was removed for routine cytological examination and the remaining supernatant was retrieved and stored at -80°C at the Pathology Department of Reims University Hospital. An informed consent was obtained from all patients for performing a diagnostic and/or therapeutic endoscopic retrograde cholangiography.

IV.4.1 Impact of biofluid volume and dilution on spectral quality test

The experimental procedures are detailed in Figure 4.1. In brief, serum and plasma experiments were performed on pure and after 2-fold, 3-fold and 4-fold dilutions using normal saline (0.9% NaCl). As we have previously shown that bile can be analysed without dilution[12], only pure bile samples were used. Samples were deposited onto silicon plates, either 96-well plate with wells of 7 mm diameter or 384-well plate with wells of 3 mm diameter. Sample volumes of 5 and 10 µL were studied using the 96-well plate and 3 and 5 µL using the 384-well plate. A deposit of 1 µL without dilution was also tested with the 384-well plate. The deposits were spread drops onto the entire surface of the well or non-spread drops in the middle of the well. Ten replicates per sample were studied. After 1h drying at room temperature, the different plates were inserted into a high-throughput module (HTS-XT) coupled to an FTIR-spectrometer (Tensor 27, Bruker Optics GmbH, Ettlingen, Germany). Spectra were acquired in the transmission mode (OPUS v.6.5 software, Bruker Optics GmbH) in the 4000-400 cm^{-1} wavenumber range with a 4 cm^{-1} spectral resolution and 32 scans. A background reference spectrum was recorded before each sample analysis. The reference spectrum was automatically subtracted to obtain the final absorbance spectrum. A zero-filling factor of 2 and a Blackman-Harris-3-term function were applied for Fourier transformation.
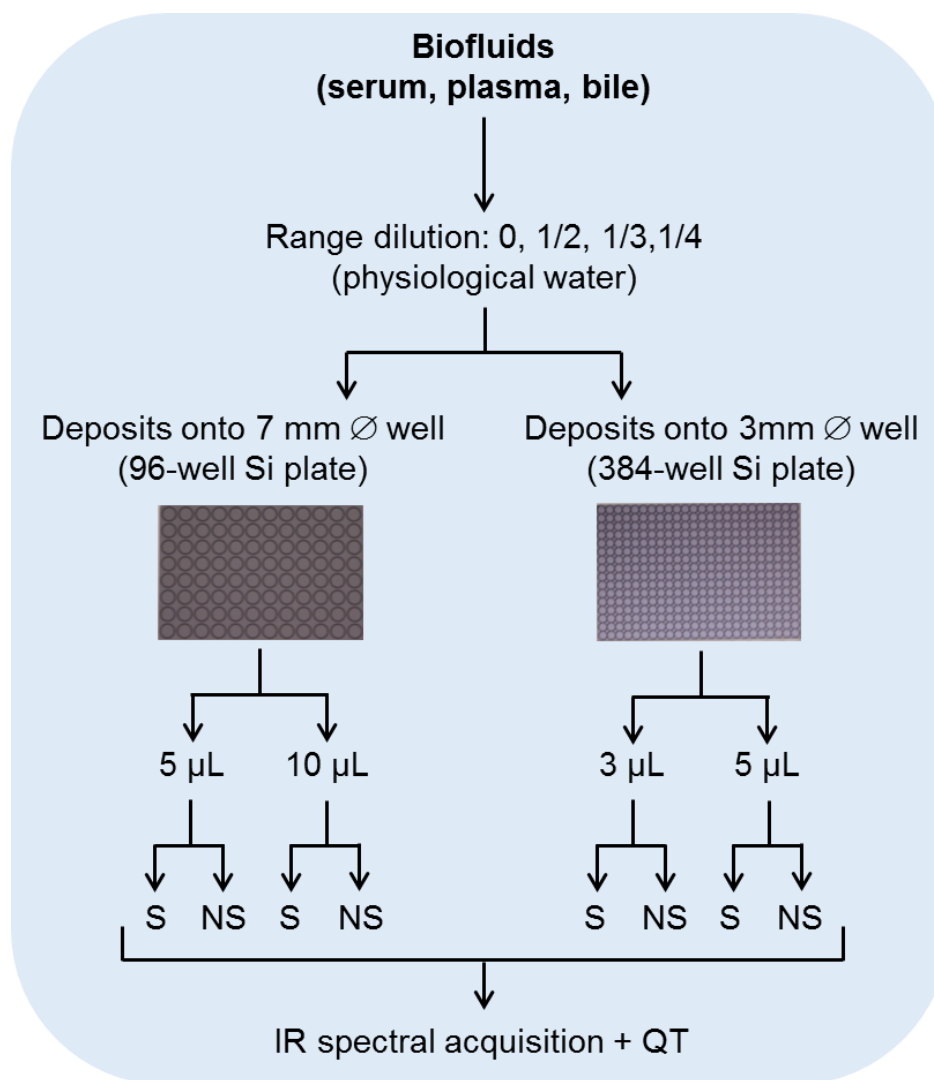
Figure 4.1 *Experimental protocol for studying the impact of dilution and volume on spectral quality test. S: spread dried drop, NS: non-spread dried drop, QT: quality test.*

The following quality tests (QT) were applied to each spectrum: 1) absorbance intensity: spectra were validated when the maximum absorbance (amide I band *c.a.* 1655 cm^{-1}) was below 1.8 arbitrary units; 2) signal-to-noise ratios were calculated according to the manufacturer's instructions for biological materials (Bruker Optics GmbH): the signals were calculated as the difference between the maximum and the minimum 1st derivative values in the range 1700-1600 cm^{-1} (signal 1 or S1) or as the difference between the maximum and the minimum 1st derivative values determined in the range 1260-1178 cm^{-1} (signal 2 or S2). Noise intensity (N) was calculated as the difference between the maximum and the minimum 1st derivative values in the range

2100-2000 cm^{-1}. Spectra met this quality test when ratios S1/N and S2/N were above 50 and 20 respectively; 3) signal-to-water ratio: water signal (W) was evaluated using the difference between the maximum and the minimum 1st derivative values in the range 1842-1837 cm^{-1}. Ratio thresholds for spectral validation were S1/W and S2/W above 10 and 4 respectively. Individual samples were considered as QT-validated for analysis when at least 8 replicate spectra out of 10 met the quality tests.

In order to analyse the impact of drop spreading, we performed experiments by successively using three-fold diluted serum and four-fold diluted plasma with spread and non-spread drops of 10 µL on 7 mm diameter wells. Spectra that passed the quality test were pre-processed on the wavenumber range from 1800 to 800 cm^{-1}. After baseline correction using a rubberband function, spectra were converted to second derivatives using the Savitsky-Golay algorithm, with 9-point smoothing to increase the spectral feature contrast. Then, their second derivatives were vector normalised. This spectral pre-processing was performed using the OPUS software (v6.5, Bruker Optics GmbH). Last, the processed spectra were classified using hierarchical cluster analysis (HCA) and principal component analysis (PCA) unsupervised methods, with spectra from spread and non-spread drops recorded in the same conditions. HCA consisted of grouping the spectra according to their degree of similarity. The method was based on Euclidean distance calculation between all the data sets by using the Ward's algorithm[19]. The merging process used the spectral information contained in the region between 1800-800 cm^{-1}. The result is displayed as a dendrogram showing the grouping of spectra in clusters according to a heterogeneity scale. In PCA, data are reformulated as a linear combination of uncorrelated principal components (PCs)[20]. The first PC calculated accounts for the highest variance in the dataset. The second PC is uncorrelated with the first one *i.e.*, perpendicular to it and explains the next highest variance of the dataset, and so on. The PCs are also called loading vectors and their weights PC scores. PCA was performed in the same spectral range to analyse the spectra from spread and non-spread drops and to determine if the two sets of spectra are separated.

IV.4.2 Impact of the dilution solvent: distilled water *versus* physiological water

In this set of experiments, the impact of the solvent used for dilution, distilled water or physiological water, was analysed. Three-fold diluted serum and plasma were prepared with each solvent. Then, 5 µL of samples were deposited onto a 384-well plate. After deposition, drops were air-dried at room temperature for 1h. Infrared spectra were acquired and pre-processed as described above. Finally, the pre-processed spectra were classified by HCA.

IV.4.3 Impact of the type of anti-coagulant for plasma collection

Plasma samples were obtained by centrifugation of blood collected in tubes containing EDTA or lithium heparin as anticoagulants. After three-fold dilution, EDTA and lithium heparin plasma samples of 5 µL were deposited onto a 384-well plate and air-dried at room temperature for 1h. Then, spectra were acquired and subjected to quality test as described above. Spectra that passed the quality test were pre-processed and classified by HCA. A dendrogram was constructed to illustrate the level of similarity between spectra of plasma samples from EDTA and lithium heparin collection tubes. In order to evaluate the impact of EDTA (MW: 292 Da) on the spectral profile, FTIR spectra of plasma samples were also analysed after dialysis with a 10 kDa membrane to filter out EDTA molecules.

IV.4.4 Spectral reproducibility

IV.4.4.1 Impact of freeze-thaw cycles

The potential impact of repeated freeze-thaw cycles on quality test and reproducibility was studied. Fresh serum and serum samples after up to five consecutive freeze-thaw cycles (1h between each cycle) were studied. After thawing and three-fold dilution, serum samples of 5 µL were deposited onto a 384-well plate and dried at room temperature for 1h. Then, spectra were acquired and subjected to quality test as

described above. Spectra that passed the quality test were pre-processed and classified by HCA to analyse the level of similarity between spectra according to the number of freeze-thaw cycles. In addition, a PCA was performed and a PCA score plot was constructed to determine if separation of the different sets of spectra could be obtained.

IV.4.4.2 Inter-operator spectral reproducibility

Inter-operator reproducibility was studied by comparing the spectra obtained by three different operators analysing the same sample under the same experimental conditions. Three-fold diluted serum was used for this purpose with 5 µL deposits onto a 384-well plate. After 1h drying at room temperature, spectra were acquired and those passing the quality test were pre-processed and analysed by HCA and PCA as described in the previous sections.

IV.4.4.3 Intra- and inter-plate spectral reproducibility, day to day reproducibility

Reproducibility tests were performed using 5 µL of three-fold diluted serum spread by the same operator on 384-well plate. After 1h drying at room temperature, spectra were acquired and subjected to the QT as described above.

Intra-plate reproducibility was studied by depositing 360 replicates of the same sample under the same experimental conditions.

Inter-plate and day to day reproducibility was studied by comparing serum spectra obtained from three different patients using 3 different silicon plates on three different days.

Spectra that passed the quality test were pre-processed and classified by HCA to analyse the level of similarity between spectra.

IV.4.5 Impact of drying modalities on spectral reproducibility

Since spectra from spread or non-spread drops are acquired after a drying period, the impact of drying duration on the QT and on the reproducibility was studied. Three-

fold diluted serum and plasma as well as pure bile were used. Samples of 5 μL were deposited onto a 384-well plate. Then, they were dried for 45, 60, 120 min and 24h before FTIR measurements. The experiments were performed at 25°C in a close environment such as an incubator and at room temperature. Spectra were acquired and subjected to QT as described above. Spectra that passed the QT were pre-processed and classified by HCA and PCA as described above.

IV.5 Results and discussion

Due to the development of high-throughput molecular technologies and associated bioinformatics, the number of studies dedicated to the identification of new diagnostic biomarkers has increased exponentially over the last two decades. However, only a limited number of biomarkers have been validated for use in medical practice[21]. This is mainly explained by methodological flaws. Plebani and Carraro reported that most errors (68%) occurred in the pre-analytical phase and only 13% and 19% errors occured in the intra-analytical and post-analytical phases respectively[22]. As for other high-throughput technologies, vibrational spectroscopic investigations are subject to the same problems although no statistics exist to date. The huge amount of data generated by spectroscopic analysis exposes this analysis to a significant risk of false positive findings. The risks should be minimised by rigorously controlling sample and patient related factors in the exploratory phase and by standardising the conditions of spectral acquisition in the pre-analytical step. Thus, care should be taken to validate sample-related factors and patient-related factors. Standardisation of specimen collection and storage is crucial to reach experimental reproducibility not only in an individual laboratory but also between different laboratories. In addition, investigators should be aware of the risks of contamination during sample handling.

In the context of developing a high-throughput FTIR (HT-FTIR) transmission method for clinical diagnostics, we have investigated some important aspects involved in biofluid sample handling and preparation, which can impact on the spectral data and consequently on the outcome, and that we believe should be considered in the pre-analytical step prior to clinical applications. The study is focused on serum and plasma,

and to a lesser extent bile, as they are the most encountered biofluids in clinical diagnostics.

IV.5.1 Impact of biofluid volume and dilution on spectral quality test

Absorption FTIR spectroscopy is directly related to the Beer-Lambert law and the amount of analysed molecules has an impact on the signal quality. Dilution and volume of samples as well as size of the wells have a significant impact on the spectral QT. Spectra from pure serum and pure plasma using a volume ≥ 3 µL did not meet the QT because the amide I band is saturated with maximum absorbance intensity above 2 arbitrary units (Table 4.1). However, when the volume of pure serum was reduced to 1 µL, a majority of spectra was validated for non-spread deposits (Table 4.1). This is a point of great interest for clinical application as the deposit of a biofluid without dilution and without spreading would be the ideal approach for a high-throughput technology since it minimizes the number of operator's interventions and thus reduces risks of variability in the results.

However, pipetting such low volumes can introduce additional errors and impact on the reproducibility of spectral data (AI.1). The results deserve to be confirmed on a larger number of samples. Furthermore, the approach has not been generalised to plasma samples (Table 4.1). The reason why the great majority of spectra from a pure plasma sample of 1 µL did not meet the QT is an open question but is probably related to the presence of coagulation proteins in plasma and not in serum. When plasma and serum samples were diluted, the amide I band absorbance decreased to the acceptable range due to the lower concentration of proteins (AI.2) in the sample and the proportion of QT-validated spectra increased. The two types of plates that are currently used in HT-FTIR spectroscopy are the 96-well and 384-well silicon substrates. Both types of plates have been tested in this study. Using the 96-well Si plate with 7 mm diameter wells, the optimal dilution for serum and plasma appears to be two-fold for spread drops and four-fold for non-spread drops (Table 4.1). Using the 384-well plate with 3 mm diameter wells, the optimal dilution was three-fold for serum and three- or four-fold for plasma depending on the volume (Table 4.1 and Figure 4.2). Overall,

higher dilution and lower volume appear necessary to meet the QT when using the 3 mm diameter wells compared with the 7 mm diameter wells. This underlines that analyte amount in the deposit area is a limiting factor for obtaining QT validated spectra. As an example, in our experiments a serum volume of 10 µL 2-fold diluted and deposited on a 7 mm diameter well resulted in a QT-validated spectrum as did a volume of 3 µL 3-fold diluted and deposited on a 3 mm diameter well. However, given the specificity of each instrument, it should be underlined that each set up necessitates a proper pre-analytical validation.

Table 4.1 *Results of quality test on FTIR serum and plasma spectra from spread and non-spread dried drops at various dilutions and volumes.*

Si plate	7 mm diameter well (96-well plate)								3 mm diameter well (384-well plate)								
Drop volume (µL)	5				10				1	3				5			
Dilution	0	1/2	1/3	1/4	0	1/2	1/3	1/4	0	0	1/2	1/3	1/4	0	1/2	1/3	1/4
Serum spread drops																	
*Sample QT validation**	R	G	G	R	R	G	G	G	R	R	R	G	G	R	R	G	G
*Validated spectra/sample***	0	9	10	0	0	10	10	9	7	0	5	10	10	0	1	10	10
Serum non-spread drops																	
*Sample QT validation**	R	R	G	G	R	R	R	G	G	R	R	G	G	R	R	G	G
*Validated spectra/sample***	1	2	9	10	0	0	2	10	8	0	4	10	10	0	1	10	10
Plasma spread drops																	
*Sample QT validation**	R	G	G	G	R	G	G	G	R	R	G	G	G	R	R	G	G
*Validated spectra/sample***	0	10	10	8	0	10	10	10	0	0	9	10	10	0	0	9	10
Plasma non-spread drops																	
*Sample QT validation**	R	R	R	G	R	R	R	G	R	R	R	G	G	R	R	R	G
*Validated spectra/sample***	0	1	6	10	0	0	0	10	2	0	0	10	10	0	0	5	10

* The sample was validated (green dot) when at least 8 replicates out of 10 met the quality test (QT) and discarded when less than 8 replicates met the QT (red dot).

** 10 replicates per sample were deposited.
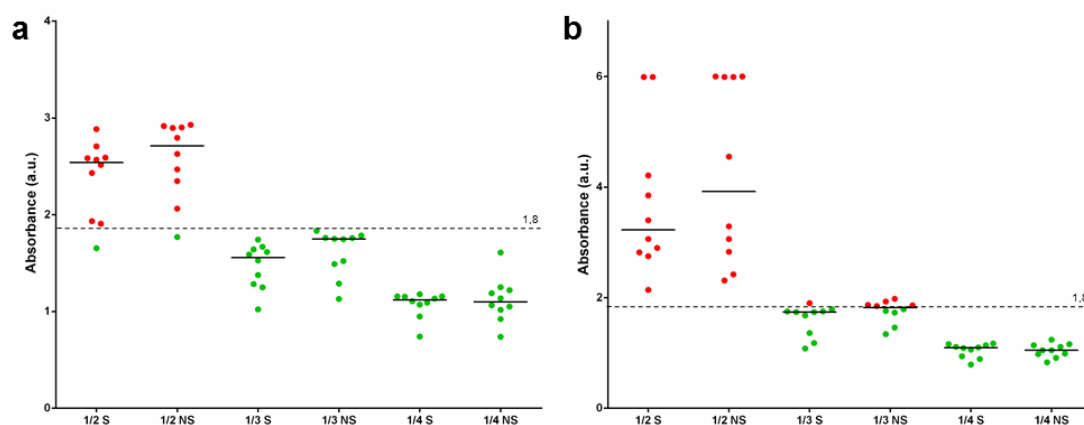
Figure 4.2 *Scatter plots based on the amide I absorbance value of serum (**a**) and plasma (**b**) replicate FTIR spectra from spread and non-spread dried drops (5 µL deposits onto 3 mm diameter well) at various dilutions. Validated and discarded spectra after the quality test are represented by green and red dots respectively. The median absorbance value is represented by the black bar. S: spread dried drop, NS: non-spread dried drop.*

122

From HCA and PCA analyses, it can be noted that pre-processed spectra from 2-, 3- and 4-fold dilution spread drops on 7 mm diameter wells are separated in three different groups (AI.3) whereas pre-processed spectra from 3- and 4-fold dilution spread drops on 3 mm diameter wells are not differentiated (AI.4). The above results indicate a higher heterogeneity with the 7 mm diameter wells compared to the 3 mm diameter wells although spectra were from the same sample and were normalised. It should be mentioned that a 5 µL deposit onto a 3 mm diameter wells covers the whole area of the well, which avoids drop spreading and represents an advantage in terms of pre-analytical validity. In this context and depending on the dilution factor, a bigger volume is needed for the 7 mm diameter wells, which will induce a longer drying time. This is illustrated in Figure 4.3 where photographs of different volumes of serum dried on the two types of plates are shown.
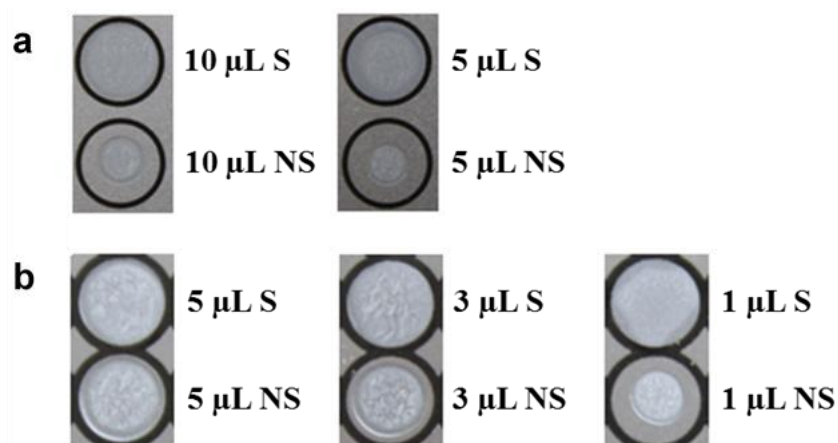
Figure 4.3 *Photographs of different volumes of dried serum drops on 7 mm diameter wells (a) and 3 mm diameter wells (b). S: spread drop, NS: non-spread drop.*

The impact of serum drop spreading on the spectral profile was studied on both 7 and 3 mm diameter wells and data were analysed by HCA and PCA. For the 7 mm diameter wells, the HCA plot showed two well-separated clusters of spectra corresponding to spread and non-spread drops indicating that the modalities of drop deposition on the surface of the plate had an influence on the spectral profile (Figure 4.4a). These results were confirmed by PCA analysis (Figure 4.4b). In addition, PC1 and PC2 loadings tend to indicate modifications in the protein spectral profile (Figure 4.4c).
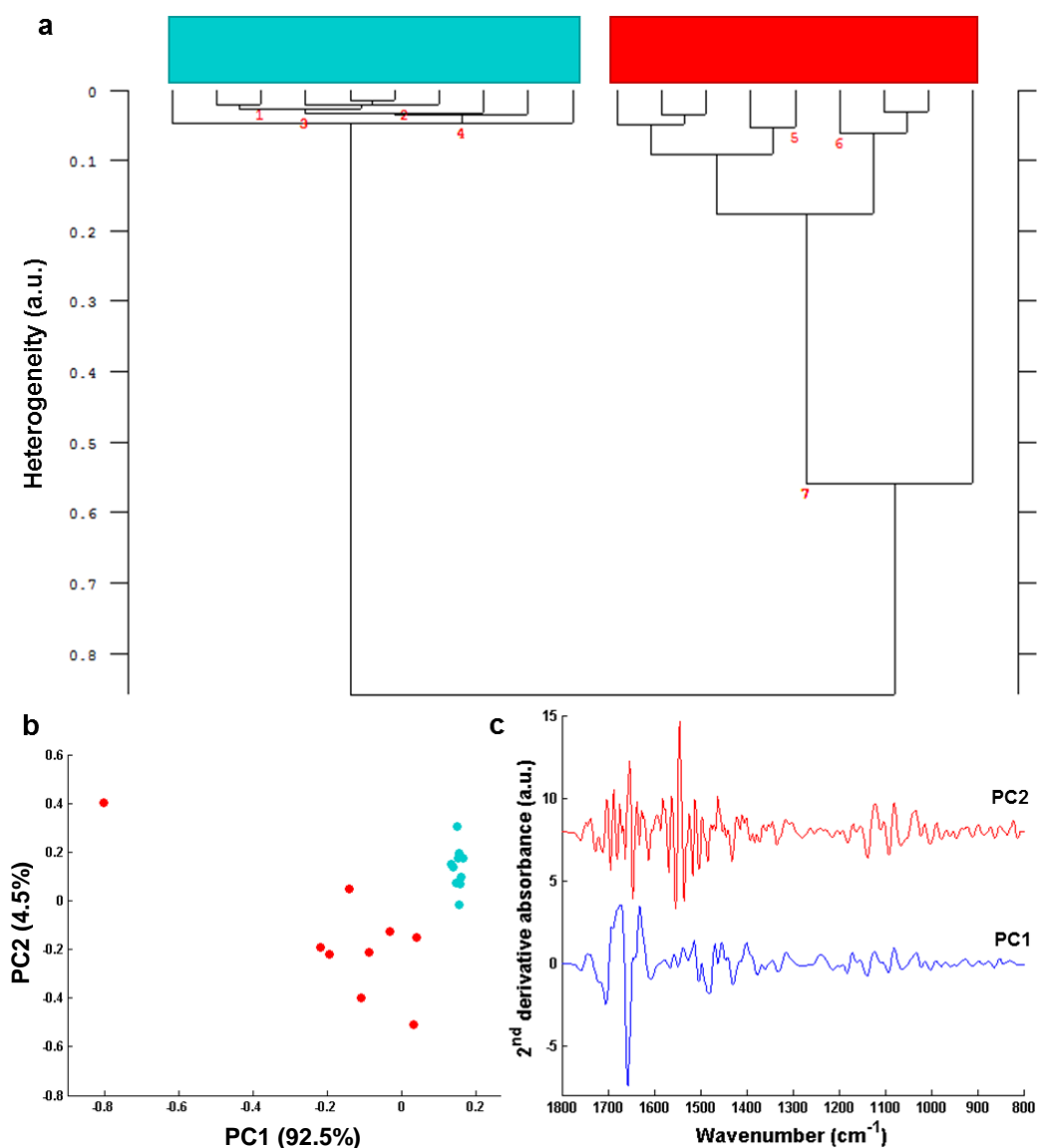
Figure 4.4 (a) *Hierarchical cluster analysis of FTIR pre-processed spectra from 3-fold diluted serum (5 µL) spread (light blue) or non-spread (red) over the surface of 7 mm diameter wells.* ***(b)*** *PCA scatter plot of the spectral dataset from (a),* ***(c)*** *PC1 and PC2 loadings from (b).*

In contrast, for the 3 mm diameter wells, both HCA and PCA analyses showed no clear distinction between the spread and non-spread dried drops spectra (Figure 4.5a, b, c). This supports the above mentioned argument that the 5 µL spread or non-spread deposit onto a 3 mm diameter wells covers the whole area of the well. Similar results were obtained for plasma samples (data not shown). Thus, 5 µL of three-fold diluted serum or plasma appear as the ideal volume for high-throughput analysis. Figures 4.2a and 4.2b confirm these latter results respectively for serum and plasma *via* the scatter

plots based on amide I absorbance by comparing the dilution factor (2-, 3- and 4-fold) as well as the spread and non-spread procedures.
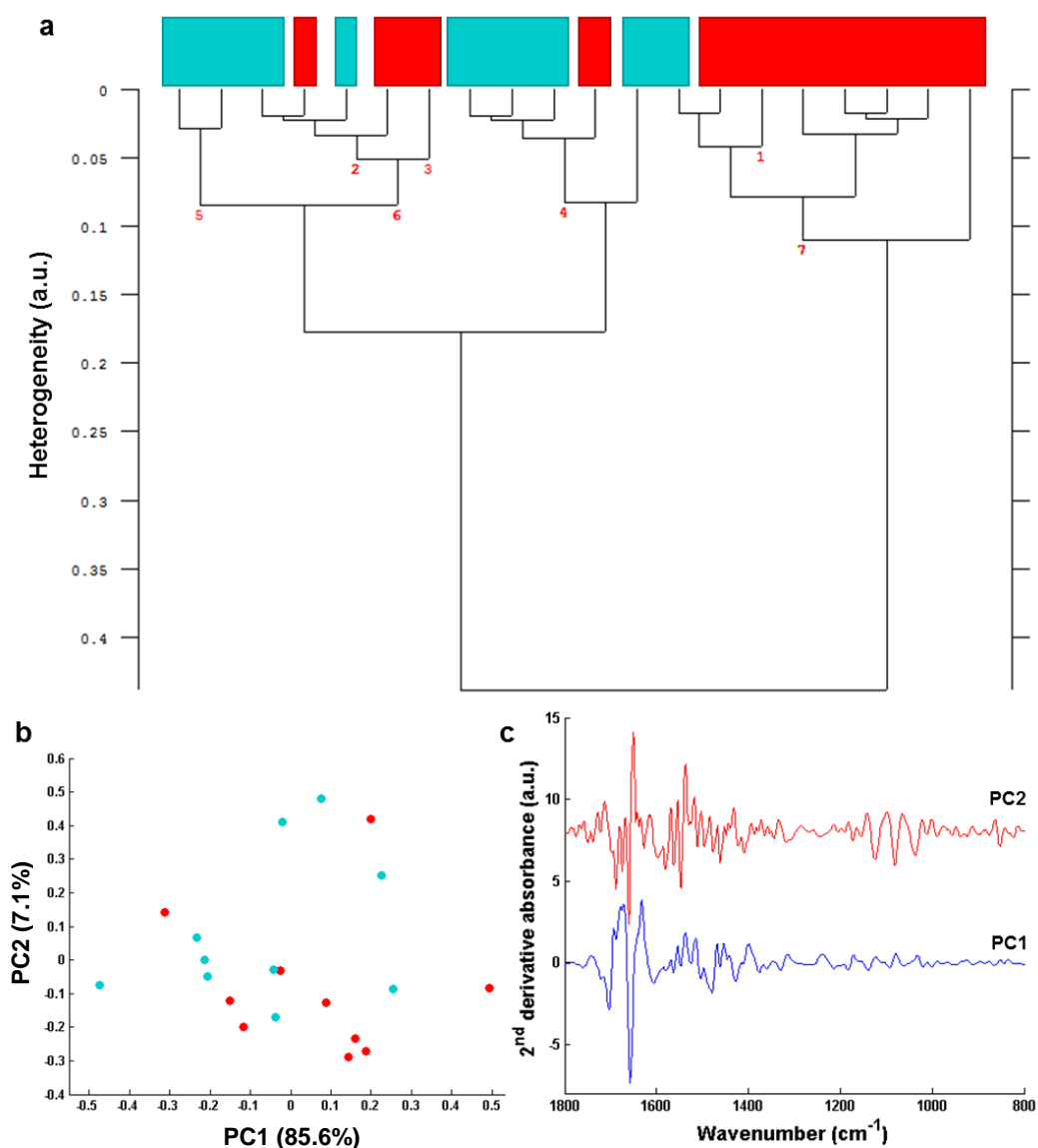
Figure 4.5 *(**a**) Hierarchical cluster analysis of FTIR pre-processed spectra from 3-fold diluted serum (5 µL) spread (light blue) or non-spread (red) over the surface of 3 mm diameter wells. (**b**) PCA scatter plot of the spectral dataset from (a), (**c**) PC1 and PC2 loadings from (b).*

Bile has been treated separately from serum and plasma because its composition is different. This is mainly due to lower concentrations of metabolites in bile compared to serum or plasma. Protein concentration in bile is about 5 g/L[23] as opposed to 60 to 80 g/L in serum. These specificities allow bile samples to be measured in their pure form (AI.5). The low absorption, in particular the amide I band, required the implementation of an adapted QT. Our results show that all pure bile spectra (except for 1 µL) were QT-validated for spread dried drop. All spectra from non-spread dried drops were rejected (Table 4.2).

Table 4.2 *Results of quality test on FTIR bile spectra from spread and non-spread dried drops at various volumes.*

Silicon plate	7 mm diameter well (96-well plate)		3 mm diameter well (384-well plate)		
Drop volume (µL)	5	10	1	3	5
Dilution	0	0	0	0	0
Bile spread drops					
*Sample QT validation**	🟢	🟢	🔴	🟢	🟢
*Validated spectra/sample***	10	10	0	10	10
Bile non-spread drops					
*Sample QT validation**	🔴	🔴	🔴	🔴	🔴
*Validated spectra/sample***	6	6	0	6	7

* The sample was validated (green dot) when at least 8 replicates out of 10 met the quality test (QT) and discarded when less than 8 replicates met the QT (red dot).
** 10 replicates per sample were deposited.

IV.5.2 Impact of the dilution solvent: distilled water *versus* physiological water

Distilled water and physiological water are often used as solvents in many analytical assays. Since biofluids contain high protein concentrations, we wanted to investigate whether these solvents could have an effect on spectral analysis. Results are displayed in Figure 4.6a where the dendrogram shows that spectra are clearly divided into two main clusters corresponding to each solvent used for dilution. This indicates that the solvent has a major impact on the spectral characteristics. It is known that distilled water induces protein precipitation (turbid solution), an effect which is not observed with physiological water (photograph insert, Figure 4.6b). If sample dilution is necessary, it is recommended using physiological water rather than distilled water.
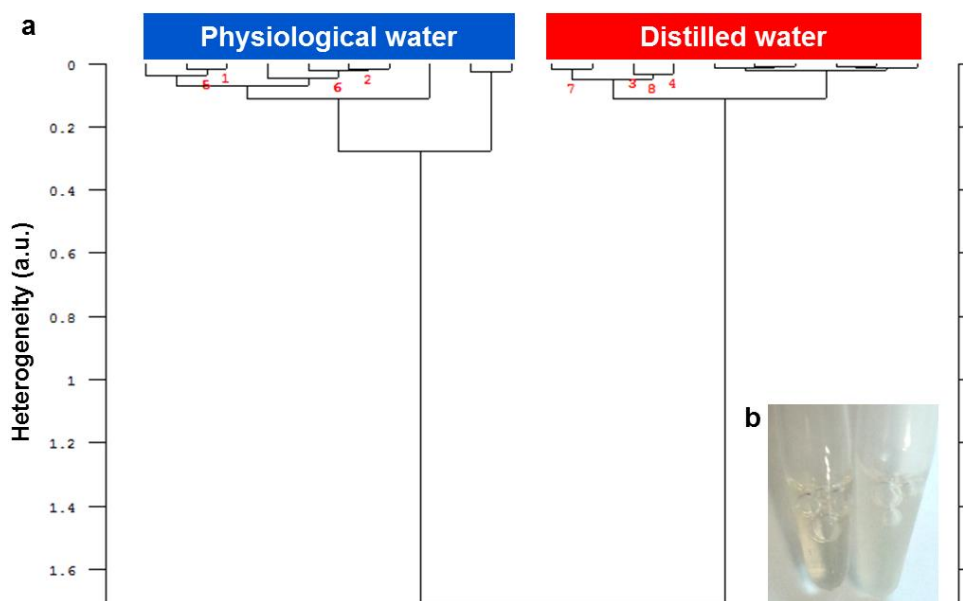
Figure 4.6 (a) Hierarchical cluster analysis of FTIR pre-processed spectra from dried spread serum drops 3-fold diluted using either physiological or distilled water (deposits of 5 μL onto 3 mm diameter wells). (b) Photograph of serum diluted with physiological water (left) and distilled water (right).

IV.5.3 Plasma collection: impact of the type of anti-coagulant

For biochemical assays used routinely in clinics, blood samples are collected using either lithium heparin or EDTA tubes depending on the targeted analyses. However, can these anti-coagulants have an effect on spectral data? To answer this question, spectra of plasmas from lithium heparin and EDTA tubes were acquired and compared in Figure 4.7a. It can be observed that plasma from EDTA tubes exhibit additional spectral features compared to plasma from lithium heparin tubes, specifically in the fingerprint region (1800-800 cm^{-1}). After a dialysis using a 10 kDa membrane, these artifactual peaks originating from small EDTA molecules can be removed (Figure 4.7b). HCA classification displayed a complete separation between spectra from EDTA and lithium heparin plasma samples before dialysis. This clustering was no longer observed after dialysis (data not shown). Given these data, it appears more appropriate to use lithium heparin tubes for spectroscopic analysis of plasma. Unless there is a specific requirement for using plasma samples, serum appears as the sample of choice for biofluid spectroscopy. Use of plasma is important in specific diseases

127

(e.g., hemophilia) where coagulating proteins are necessary for biochemical and hemostasis assays. In contrast with plasma, there is no need for using anticoagulants to collect serum making sample handling easier.
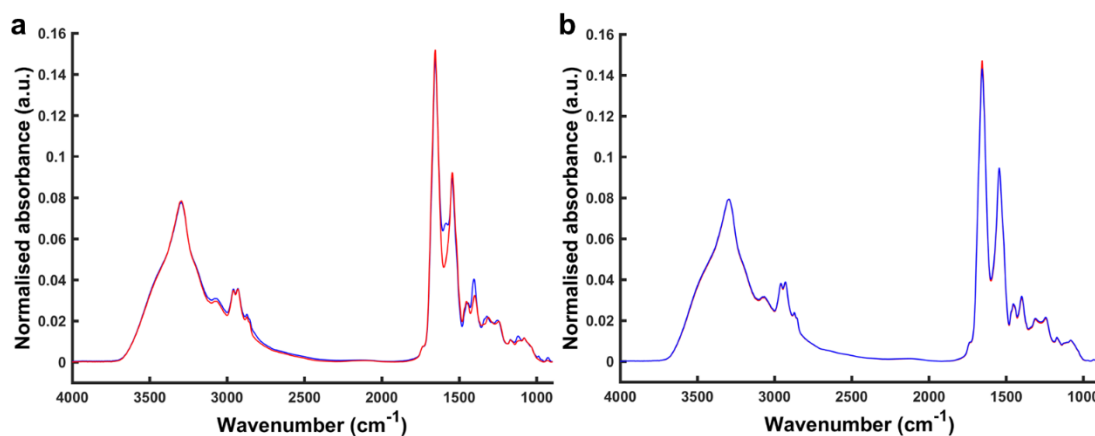
Figure 4.7 *Comparison of FTIR pre-processed mean spectra of plasma collected in tubes containing lithium heparin (red) or EDTA (blue) as anticoagulants without dialysis (**a**) or after dialysis with a 10 kDa membrane (**b**).*

IV.5.4 Spectral reproducibility

IV.5.4.1 Impact of freeze-thaw cycles on spectral reproducibility

The impact of repeated freeze-thaw cycles before biofluid sample analysis is a matter of concern in biomedical technologies. Although they can be obtained non-invasively, serum and plasma are very valuable sample materials and they are stored at -80°C where they are supposed to be relatively stable. However, for clinical studies, repeated freeze–thaw cycles are unavoidable when the sample volume is limited. Thus, inadequate conditions for sample storage can have an impact on measured data as it has been reported in a recent metabolomics study[17]. Taking into account this finding, we have studied the effect of freeze-thaw cycles on spectral data. Fresh serum and serum samples after up to five consecutive freeze-thaw cycles were analysed and their spectral profiles compared. Figure 4.8a displays the HCA results obtained in the 1800-800 cm^{-1} spectral range and clearly demonstrates a clear-cut delineation between fresh and freeze-thawed serum samples even after 1 cycle.
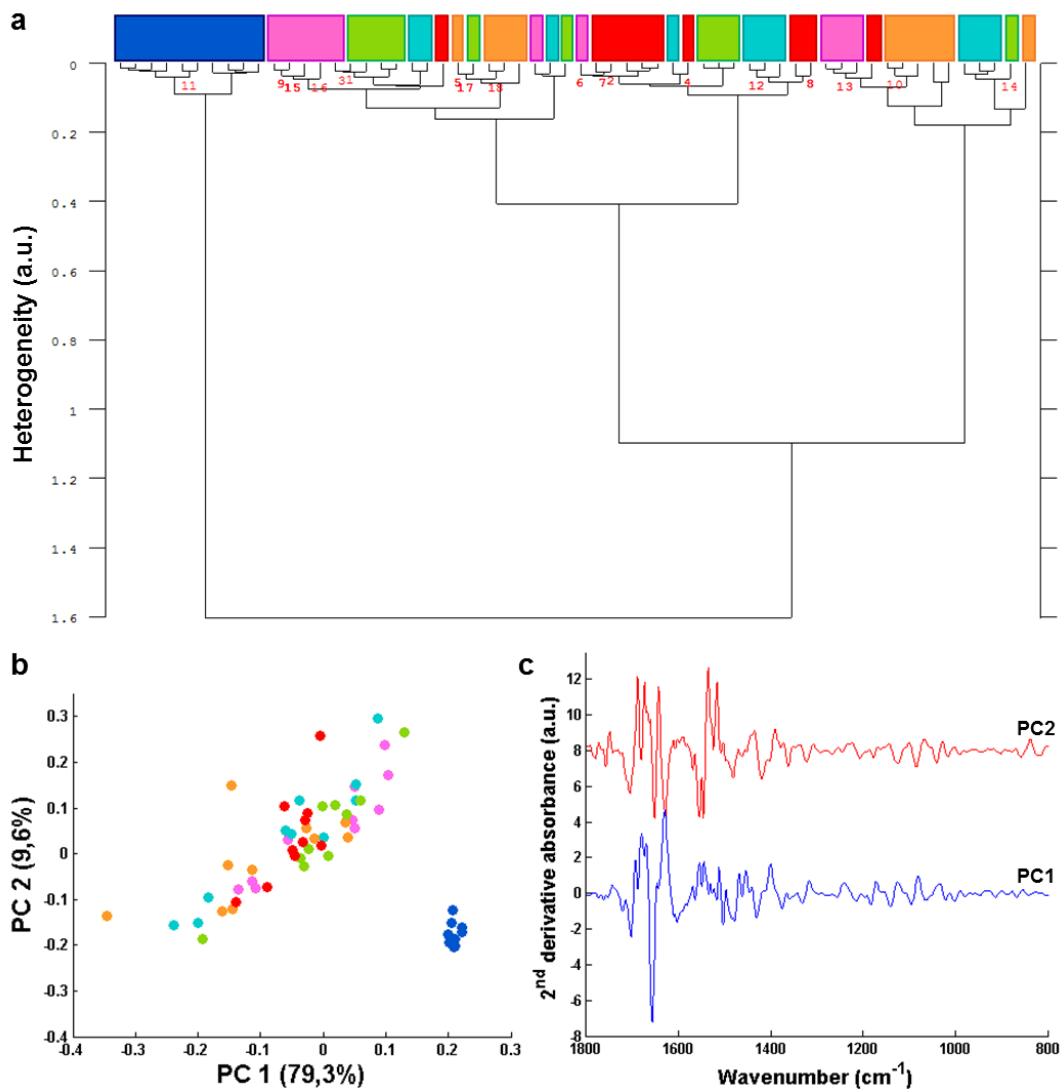
Figure 4.8 (a) *Hierarchical cluster analysis of FTIR pre-processed spectra from fresh serum and after up to 5 freeze-thaw cycles (deposits of 5 µL, 3-fold diluted onto 3 mm diameter wells).* ***(b)*** *PCA scatter plot of the spectral dataset from (a),* ***(c)*** *PC1 and PC2 loadings from (b). Fresh serum (dark blue), freeze-thaw cycle 1 (pink), 2 (light blue), 3 (orange), 4 (green), 5 (red).*

The results were confirmed by PCA analysis showing the fresh samples as a completely separate group (Figure 4.8b). Both analyses show that all freeze-thawed data were mixed. Furthermore, the loadings of PC1 and PC2 tend to indicate that the most prominent modifications occur in the protein region (Figure 4.8c). When the fresh samples were removed, data analysis did not reveal any structures in the spectral datasets of the five freeze-thaw cycles (data not shown). Similar results have been obtained when fresh plasma was compared with plasma samples after up to five

consecutive freeze-thaw cycles (data not shown). These findings suggest that FTIR spectroscopy is sensitive enough to distinguish fresh from freeze-thawed serum and plasma samples but cannot differentiate between the five freeze-thaw cycles. On the other hand, metabolomics techniques being more sensitive have shown changes in metabolite ions[24] and macromolecules[25,26] after a few freeze-thaw cycles. Since sample freezing is unavoidable (large collection, multicentric studies, transportation…), it is therefore recommendable to avoid multiple freeze-thaw cycles of the same sample and favour multiple pure sample aliquots. However, our study has not taken into account the effect of freezing time on the spectral data, in particular for long term storage, and it would be interesting to investigate if FTIR spectroscopy could monitor any content deviations in metabolite ions and macromolecules like proteins, DNA and RNA.

IV.5.4.2 Intra- and inter-plate spectral reproducibility, day to day reproducibility

In this study, intra- and inter-plate as well as day to day reproducibility has been investigated. For intra-plate reproducibility, 360 replicates of the same sample were deposited and measured by the same operator. The HCA analysis did not show any specific clustering and all spectra were completely mixed and the replicates were distributed over the whole cluster (data not shown).

Day to day reproducibility was studied by analysing serum from three patients on three consecutive days. The dendrogram obtained after HCA analysis (Figure 4.9) shows that the clustering is only based on patient information. This is an important finding because the patient to patient variation is greater than day to day variation. This experiment has been reproduced on two other plates and no significant inter-plate spectral variability has been observed (data not shown).
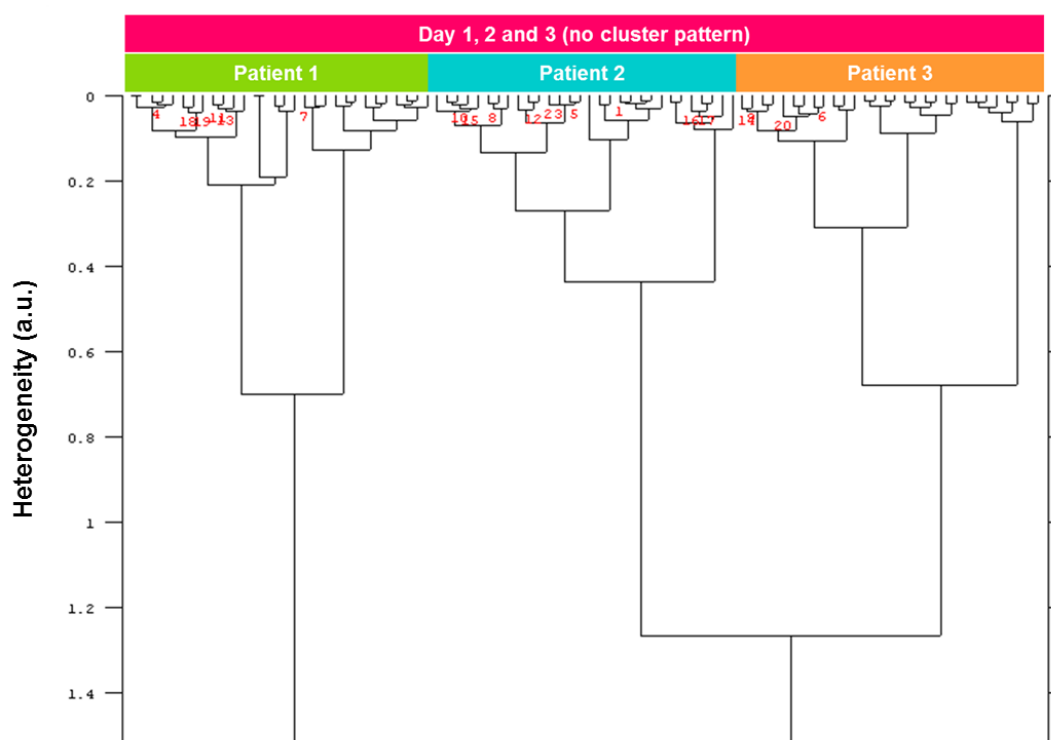
Figure 4.9 *Hierarchical cluster analysis of FTIR pre-processed spectra from serum of three patient samples deposited on a same silicon plate on three different days (deposits of 5 µL, 3-fold diluted onto 3 mm diameter wells).*

IV.5.4.3 Inter-operator spectral reproducibility

The inter-operator reproducibility was investigated by three different operators using the same instrumentation to analyse the same sample on the same plate and on the same day. This is to mimic a normal working day procedure in a clinical setup where the instrument could be used by several users. HCA and PCA have been used to visualise any structure in the dataset. Both HCA and PCA results show that the spectral reproducibility is operator-independent (AI.6). This is an important result for clinical implementation provided that the protocol is strictly followed.

The question is open whether this level of inter-operator reproducibility can be reached on different instruments of the same manufacturer or different manufacturers.

IV.5.5 Impact of drying modalities on spectral reproducibility

High-throughput FTIR transmission spectroscopy of biofluids necessitates a drying phase for both spread and non-spread drops. This phase can impact on the QT and on the reproducibility of the spectral data. This effect has been studied on three-fold diluted serum and plasma as well as on pure bile. Five microlitre drops were dried at room temperature and at 25°C during 45, 60, 120 min and 24h before FTIR measurements. Table 4.3 summarises the results obtained for serum, plasma and bile samples and shows that a drying time of 45 min is not sufficient to meet the QT conditions (red dots). For all three biofluids, a minimum of one hour drying time is necessary at room temperature to pass the QT as opposed to two hours at 25°C. However, it is important to note that it would be more appropriate to dry for a longer duration at room temperature to make sure that the drying process is complete.

Table 4.3 *Quality test results on FTIR raw spectra of 3-fold diluted spread drops (5 µL) of serum (S), plasma (P) and pure spread bile (B) drops (5 µL) left dried at room temperature (RT) or 25°C for 45, 60, 120 minutes and 24 h*

	45 min			60 min			120 min			24 h		
	S	P	B	S	P	B	S	P	B	S	P	B
Drying at RT												
*Sample QT validation**	🔴	🔴	🔴	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢
*Validated spectra/sample***	3	0	1	10	10	10	10	9	9	9	10	9
Drying at 25°C												
*Sample QT validation**	🔴	🔴	🔴	🔴	🔴	🔴	🟢	🟢	🟢	🟢	🟢	🟢
*Validated spectra/sample***	0	0	0	0	0	0	10	10	10	10	10	10

* The sample was validated (green dot) when at least 8 replicates out of 10 met the quality test (QT) and discarded when less than 8 replicates met the QT (red dot).

** 10 replicates per sample were deposited.

The spectra obtained after 2h drying are differentiated by HCA from those recorded after 1h and 24h (Figure 10a). Additionally, PCA analysis shows a separation between 1h and 2h drying and between 2h and 24h drying at room temperature (Figure 10b and 10d). The PC loadings clearly show that water absorption contributes to the spectral variance (Figure 10c and 10e). The proximity of the clusters corresponding to data obtained after 1h and 24h drying suggests that samples rehydration may have occurred overnight. In a clinical application after the drying phase, it is therefore important to

define the drying time and not to wait too long before recording spectra. When comparing spectra of drops dried at room temperature and at 25°C, we found that HCA shows no clustering according to these two conditions (data not shown).This is an interesting aspect for the clinical use because sample drying can be performed at room temperature without the need of any additional laboratory equipment for controlling drying conditions.
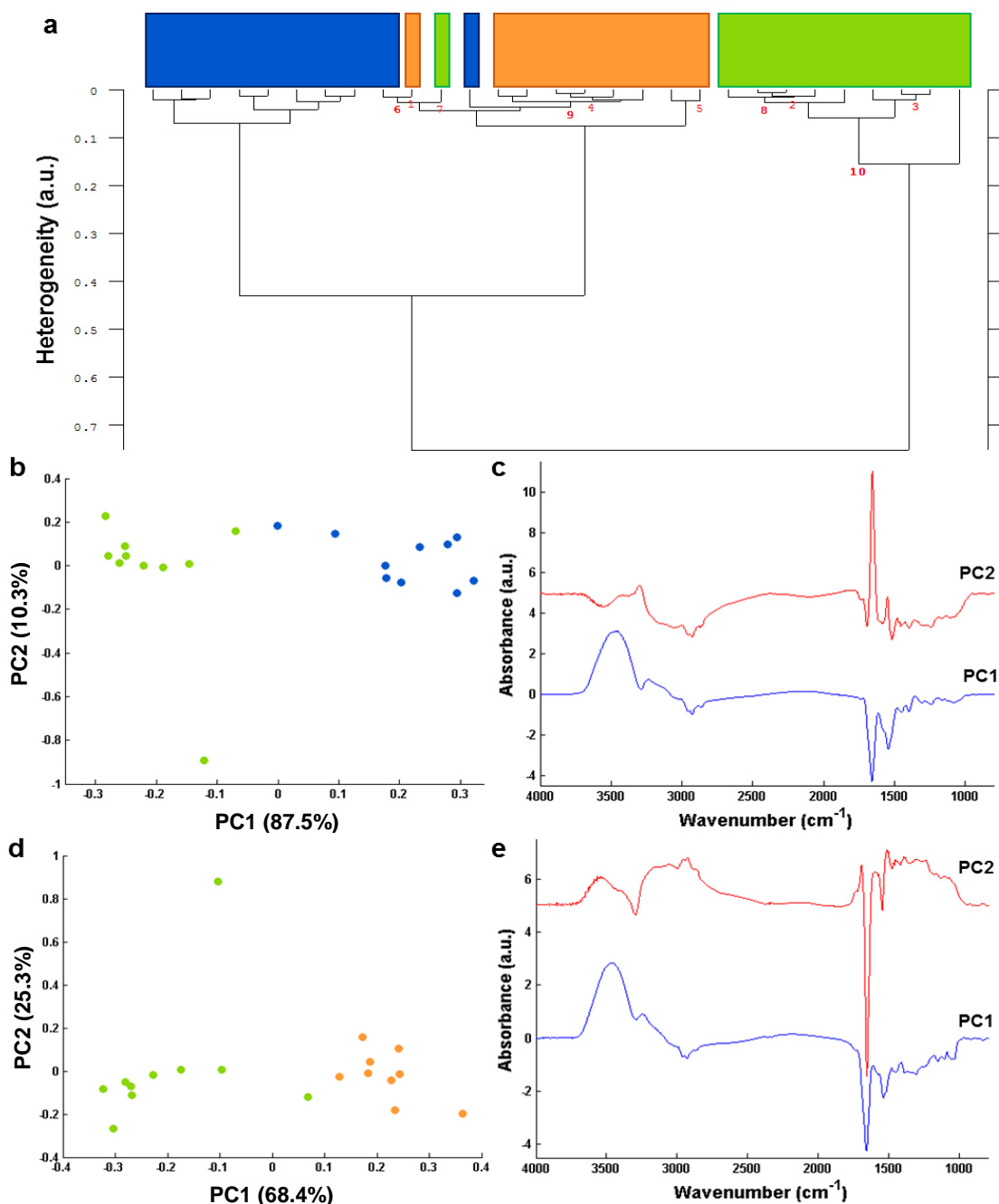
Figure 4.10 (*a*) *HCA of FTIR spectra from 3-fold diluted serum (5 µL) deposited onto 3 mm diameter wells and left dried at room temperature 1h (blue), 2h (green) or 24h (orange).* (*b*) *PCA scatter plot of the spectra of serum dried at room temperature 1h (blue) and 2h (green).* (*c*) *PC1 and PC2 loadings from (b).* (*d*) *PCA scatter plot of the spectra of serum dried at room temperature 2h (green) and 24h (orange).* (*e*) *PC1 and PC2 loadings from (d).*

IV.6 Conclusions

The development of any laboratory analytical technique, including vibrational spectroscopy, that has the potential of being translated to clinical applications requires an evaluation of the potential factors of the pre-, intra- and post-analytical phases that can impact on data quality and consequently on the result outcome. It has been reported that in the procedures described for the laboratory analysis of a patient's sample, two-thirds of the issues originate from the pre-analytical procedures. In this context, we have investigated using HT-FTIR spectroscopy some aspects of the pre-analytical procedures and their impact on the data quality and reproducibility. The results show that the type of biofluid to be analysed will condition the analytical procedures to be implemented. In a non-exhaustive manner, we have shown that the sample collection modality, the type of substrate, the dilution factor, the volume of sample deposited, the way the sample is deposited, the drying conditions, the inter-operator, the day-to-day variabilities are some of the aspects that need to be investigated for error tracking in the pre-analytical phase. Indeed, more work is needed to evaluate intra- and inter-centre variabilities as well as the effect of other factors involved in the intra- and post-analytical phases. The study has focused on the HT-FTIR methodology but it will be interesting to have similar studies on other vibrational spectroscopic modalities.

IV.7 References

1   F. Bonnier, F. Petitjean, M. J. Baker and H. J. Byrne, *J. Biophotonics*, 2014, **7**, 167–179.

2   J. L. González-Solís, J. C. Martínez-Espinosa, L. A. Torres-González, A. Aguilar-Lemarroy and P. Jave-Suárez, L F Palomares-Anda, *Lasers Med. Sci.*, 2104, **29**, 979–985.

3   C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N. W. Clarke and P. Gardner, *J. Biophotonics*, 2014, **7**, 180–188.

4   C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, *Analyst*, 2015, **140**, 2280–2286.

5   A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin and P. L. Martin-hirsch, *J. Biophotonics*, 2014, **165**, 153–165.

6   J. Ollesch, S. L. Drees, M. H. Heise, T. Behrens, T. Brüning and K. Gerwert, *Analyst*, 2013, **138**, 4092–4102.

7   W. Petrich, K. Lewandrowski, J. Muhlestein, M. Hammond, J. Januzzi, E. Lewandrowski, R. Pearson, B. Dolenko, J. Früh and M. Haass, *Analyst*, 2009, **134**, 1092–1098.

8   I. Taleb, G. Thiéfin, C. Gobinet, V. Untereiner, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Hillon, M. Manfait, P. Ducoroy and G. D. Sockalingum, *Analyst*, 2013, **138**, 4006–4014.

9   X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon and G. D. Sockalingum, *Transl. Res.*, 2013, **162**, 279–286.

10   M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszewski, G. Thiéfin and G. D. Sockalingum, *Chem. Soc. Rev.*, 2016, **45**, 1803–1818.

11   M. Griebe, M. Daffertshofer, M. Stroick, M. Syren, P. Ahmad-Nejad, M. Neumaier, J. Backhaus, M. G. Hennerici and M. Fatar, *Neurosci. Lett.*, 2007, **420**, 29–33.

12   V. Untereiner, G. D. Sockalingum, R. Garnotel, C. Gobinet, F. Ramaholimihaso, F. Ehrhard, M. D. Diebold and G. Thiéfin, *J. Biophotonics*, 2014, **7**, 241–253.

13   E. Schleicher, *Anal. Bioanal. Chem.*, 2006, **384**, 124–131.

14   A. Aguilar-Mahecha, M. Kuzyk, D. Domanski, C. H. Borchers and B. M, *PLoS One*, 2012, **7**, e38290–e38290.

15   J. A. Bons, D. de Boer and W. K. van Dieijen-Visser, M P Wodzig, *Clin. Chim. Acta.*, 2006, **366**, 249–256.

16   W. B. Dunn, D. Broadhurst, P. Begley, E. Zelena, N. Francis-McIntyre, S Anderson, M. Brown, J. D. Knowles, A. Halsall and J. N. Haselden, *Nat. Protoc.*, 2011, **6**, 1060–1083.

17   P. Yin, R. Lehmann and G. Xu, *Anal. Bioanal. Chem.*, 2015, **17**, 1–14.

18   Y. Zheng, T. Qing, Y. Song, J. Zhu, Y. Yu, W. Shi, L. Pusztai and S. L, *Biomark. Med.*, 2015, **9**, 1265–1272.

19   J. H. Ward, *J. Am. Stat. Assoc.*, 1963, **58**, 236–244.

20   H. Abdi and L. J. Williams, *Wiley Interdiscip. Rev. Comput. Stat.*, 2010, **2**, 433–459.

21   E. P. Diamandis, *J. Natl. Cancer Inst.*, 2010, **102**, 1462–1467.

22   M. Plebani and P. Carraro, *Clin. Chem.*, 1997, **43**, 1348–1351.

23   A. Reuben, *Hepatology*, 1984, **4**, 46S–50S.

24   P. Yin, A. Peter, H. Franken, X. Zhao, S. S. Neukamm, L. Rosenbaum, M. Lucio, A. Zell, H.-U. Häring and G. Xu, *Clin. Chem.*, 2013, **59**, 833–845.

25   B. L. Mitchell, Y. Yasui, C. I. Li, A. L. Fitzpatrick and P. D. Lampe, *Cancer*

Inform., 2005, **1**, 98–104.

26   A. J. Rai, P. Stemmer, Z. Zhang, B. l Adam, W. T. Morgan, R. E. Caffrey, V. N. Podust, M. Patel, L. Lim and N. V Shipulina, *Proteomics*, 2005, **5**, 3467–3474.

Chapter V. Investigating pre-analytical requirements for infrared spectroscopy of blood-derived products

V.1 Etude des exigences pré-analytiques pour l'analyse de produits dérivés du sang par spectroscopie infrarouge

V.1.1 Résumé

La spectroscopie infrarouge est une technique rapide, facile d'utilisation, sans marquage et peu coûteuse. De nombreuses études sur les biofluides ont montré son application prometteuse en tant qu'outil diagnostic clinique. Cependant, avant de considérer son transfert en milieu clinique, une attention particulière dans le but de standardiser les protocoles pour l'application de la spectroscopie infrarouge aux biofluides est nécessaire. Pour atteindre cet objectif, des études pour identifier et traquer les erreurs qui peuvent être établies dans la phase pré-analytique sont une étape cruciale. Dans ce rapport est détaillé les résultats de recherches sur des facteurs pré-analytiques qui peuvent affecter la qualité des données spectrales acquises, tels que le stockage des échantillons de sérum et plasma à long terme (9 mois) à -80°C, ou la reproductibilité de l'analyse spectroscopique au cours de mois. L'impact de l'humidité relative sur les échantillons et les acquisitions spectrales, de même que des corrections des expériences et des données pour une variation de contenu en eau sont également rapportés et discutés.

Cette étude a été réalisée sur 3 sérums et 3 plasmas de sujets sains, répartis selon 3 modalités différentes, 50 µL dans des tubes plastiques d'une capacité de 0.2 mL ou 1.5 mL et 100 µL dans des tubes plastiques d'une capacité de 1.5 mL pour évaluer l'impact du volume et le type de tube pendant la conservation des échantillons à long terme à -80°C. Chaque mois sur une période de 9 mois, 2 aliquots de sérum et de plasma ont été décongelés et dilués au tiers avec une solution saline. Les dépôts ont séché à température ambiante ou en conditions sous vide pendant 90 min. La température et l'humidité relative ont été relevées à chaque expérience. Les données ont été corrigées par l'approche EMSC (Extended Multiplicative Signal Correction) associée à un spectre de solution saline et analysées par classification hiérarchique ascendante et analyse en composante principale.

V.1.2 Conclusions

Les expériences réalisées sur les 3 sérums et plasmas de mêmes sujets sains ont permis d'estimer et d'identifier des erreurs de la phase pré-analytique dans l'application de la spectroscopie aux biofluides. Il a été observé que la conservation des échantillons de sérums/plasmas sur une période de 9 mois à -80°C n'a pas d'influence sur la signature spectrale des échantillons analysés. De même, l'utilisation de différents tubes plastiques pour la conservation des biofluides n'a pas d'impact sur les spectres collectés. Cependant, lors de l'étude de reproductibilité de l'analyse spectrale au cours des mois, il a été observé que le séchage des échantillons à température ambiante conduisait à une différence de contenu en eau des échantillons. Cela démontrant que le séchage du sérum et du plasma dépend de l'humidité relative de l'environnement d'analyse. Des équipements supplémentaires tels que des systèmes de purge pourrait être implémentés pour contrôler l'humidité de l'air ambiante pendant le séchage et l'analyse des échantillons mais l'ajout d'un tel dispositif augmente le coût total de l'analyse spectrale. Une autre approche est de directement corriger les données spectrales avec différents contenus en eau. Entre les différentes méthodes évaluées, il peut être conclu que la meilleure correction est une combinaison d'une approche EMSC basée sur un spectre de solution saline et le retrait de la composante principale responsable d'une discrimination basée sur le contenu en eau des échantillons. Cependant, il est à noter que cette méthodologie a été développée sur un modèle simple de 3 échantillons pour comprendre les variations spectrales induites par des facteurs environnementaux. Cela implique que la variabilité naturelle entre échantillons de patients d'un large set a été minimisée et pourrait influencer l'efficacité de la correction appliquée. En conclusion, pour réaliser des analyses spectroscopiques sur des biofluides dans le but d'une application clinique, une conception minutieuse de l'expérience mais surtout une évaluation des erreurs de la phase pré-analytique sont nécessaires afin de s'assurer de cibler des informations moléculaires liées à une pathologie donnée et non à des interférences environnementales.

V.2 Chapter overview

Similar to chapter IV, this study exposes the results of further investigations into pre-analytical factors that can affect the quality of spectral data acquired in order to standardise protocols for serum analysis by High Throughput Fourier Transform (HT-FTIR) spectroscopy. The aim of this work is to evaluate the impact of long-term freezing time (9 months) of serum and plasma sample storage and the month to month reproducibility of the spectroscopic analysis. The impact of relative humidity on samples and spectral acquisitions as well as experimental and data corrections for variations in water content are also reported and discussed.

V.3 Introduction

Infrared spectroscopy applied to biofluids presents several advantages primarily as these are common diagnostic media easily collectable and routinely analysed in clinical settings. Spectroscopy can provide diagnostic results within minutes, giving access to a full molecular profile of a patient sample rather than current techniques targeting a specific biomarker. Therefore, spectroscopy could be an excellent competitive alternative to expensive and time-consuming methods[1,2]. Serum based spectroscopy has proved its potential as a clinical diagnostic tool in several proof-of-concept studies[3]. However, these articles have also highlighted the lack of consensus towards a standardised protocol regarding sample handling and preparation methodology in order to perform a spectroscopic analysis[4]. It is known that the pre-analytical phase has a greater source of sample variance than the analytical and post-analytical phases[5–8]. Related erroneous results are a waste of time and money but could also be detrimental to the patient health[9]. It is therefore crucial to identify and track errors that can occur in the pre-analytical phase.

Using High Throughput-Fourier Transform Infrared Spectroscopy (HT-FTIR), we have previously investigated aspects of the pre-analytical phase that can be a source of variance. Sample preparation parameters that can impact the results of the spectroscopic analysis have been assessed such as the use of solvents to dilute the

samples, normal saline (NaCl 0.9%) is preferable to distilled water to avoid protein precipitation. Lithium-heparin anti-coagulant is recommended for plasma obtention, as ethylenediaminetetraacetic acid (EDTA) plasma spectra presented interference peaks. During freeze-thaw cycles it has been shown that serum and plasma samples were affected mainly in the protein region upon the first freeze-thaw cycle but no further changes have been observed after additional freeze-thaw cycles. Operator dependence and the day to day reproducibility assessment of the IR spectroscopy technique, have not highlighted any issue[10]. In the present study, modalities of the sample storage such as the use of different aliquot plastic tubes, long-term freezing time storage and the spectroscopic analysis month to month reproducibility have been evaluated to pursue the aim of standardising protocols for a clinical translation.

V.4 Material and Method

V.4.1 Blood-derived product samples

This study was performed on 3 sera and plasma from 3 healthy subjects, obtained from routine biochemical check-up at the Biochemistry Laboratory of Reims University Hospital. Freshly collected blood was left to clot at room temperature (30 min) and centrifuged at 3000 rpm for 15 min to obtain serum. Blood collected in tubes containing lithium heparin as anticoagulant, was centrifuged at 3000 rpm for 15 min to obtain plasma. Samples were aliquoted among 3 different modalities, 50 µL in 0.2 mL or 1.5 mL capacity plastic tubes and 100 µL in 1.5 mL capacity plastic tubes to assess the impact of the sample volume, the type of tube during a long-term storage. All aliquots were stored at -80°C until experiments were performed.

V.4.2 Experimental design

Every month during a 9 month period, 2 sera and plasma sample aliquots per condition described above were thawed and 3-fold diluted with normal saline (NaCl 0.9%, Miniversol, Aguettant, France) as previously demonstrated[11,12]. Ten instrumental replicates of 5 µL were deposited onto a silicon plate (384 wells) and left dry at room

temperature for 90 minutes (Figure 5.1) or under vacuum drying conditions. In both cases, the temperature and relative humidity (RH) were recorded for each experiments (E1-9) performed (Table 5.1). Spectra of normal saline were also acquired as a reference in the same spectral acquisition parameters (Figure 5.2).
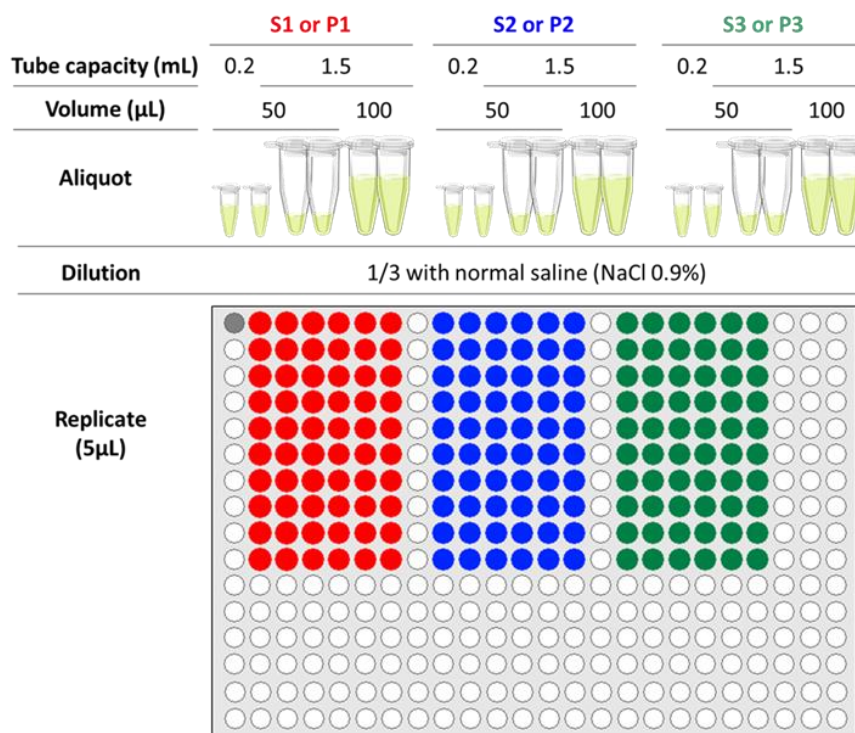
Figure 5.1 *Schematic of the experimental study design, instrumental serum replicates are deposited onto a silicon plate (384 wells). S: serum, P: plasma.*

Table 5.1 *Temperatures and Relative Humidity (RH) recorded during the different experiments performed.*

Experiment	RT drying		Vacuum drying	
	Temperature (°C)	RH (%)	Temperature (°C)	RH (%)
1	21,4	39	NA	NA
2	21,9	30	NA	NA
3	22,6	36	NA	NA
4	21,8	30	NA	NA
5	23,9	60	NA	NA
6	24,4	45	NA	NA
7	25,9	43	23,5	14
8	25	59	23,6	8
9	22,5	50	22,7	8

RT: room temperature

142
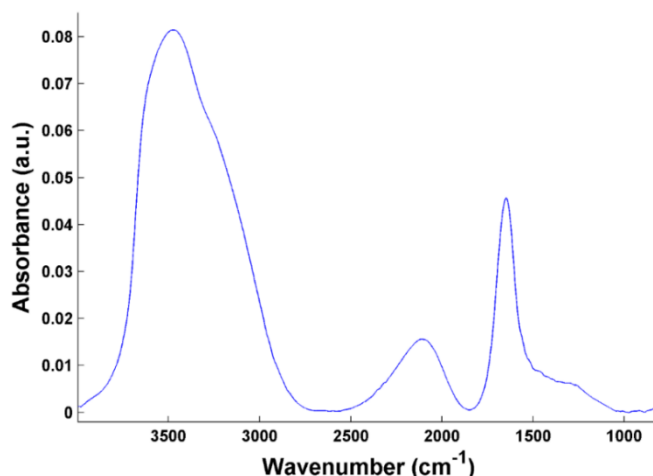
Figure 5.2 *Normal saline normalised absorbance spectrum.*

V.4.3 Spectral acquisition parameters

Spectra were collected with a high-throughput module (HTS-XT) coupled to a FTIR-spectrometer (Tensor 27, Bruker Optics GmbH, Ettlingen, Germany). The spectral acquisition was performed in the transmission mode (OPUS v.6.5 software, Bruker Optics GmbH) in the 4000-400 cm^{-1} range with a spectral resolution of 4 cm^{-1} and 32 scans. For each sample, a background was recorded and automatically subtracted to obtain the final absorbance spectrum. A spectral quality test (QT) as previously described was applied to each spectrum[11,10]. All raw spectra which passed the quality test are presented in AII.1. In addition, serum (n=52) and plasma (n=6) spectra with an absorbance presenting an abnormal decrease in the spectral region 3700-3320 cm^{-1} translating a too high content of water in the sample replicates were also discarded (AII.2). On a total of 3240 spectra collected, 199 spectra were discarded.

V.4.4 Spectral data analysis

Following the QT, spectral data were pre-processed. Spectra were cut over the spectral range 4000-800 cm^{-1}, a rubberband baseline correction or a second derivative was applied before a vector normalisation. Then, serum and plasma spectral reproducibility inter aliquots and long-term storage was assessed by Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA)[13,14].

To overcome the presence of water in the serum and plasma sample spectra, an Extended Multiplicative Signal Correction (EMSC)[15] was tested on the spectral data based on the methodology described by Afseth *et al.*[16]. The mean spectrum of all normalised absorbance spectra was defined as the reference spectrum, and a spectrum of normal saline was used as the interference spectrum to be removed from the dataset.

V.5 Results and discussion

V.5.1 Serum and plasma inter-aliquot reproducibility

Once a month over 9 months, 2 aliquots of sera and plasma samples (different volumes/tubes) per experiment were thawed and 3-fold diluted with normal saline (NaCl 0.9%). The spectral reproducibility of the 20 serum or plasma instrumental replicates deposited was assessed by HCA for each single sample on the spectral ranges 4000-800 cm^{-1} (Figure 5.3). Spectra of the sample aliquots were not distinguishable in any cases, showing no issue in the spectral reproducibility and therefore in sample handling before the spectral analysis.
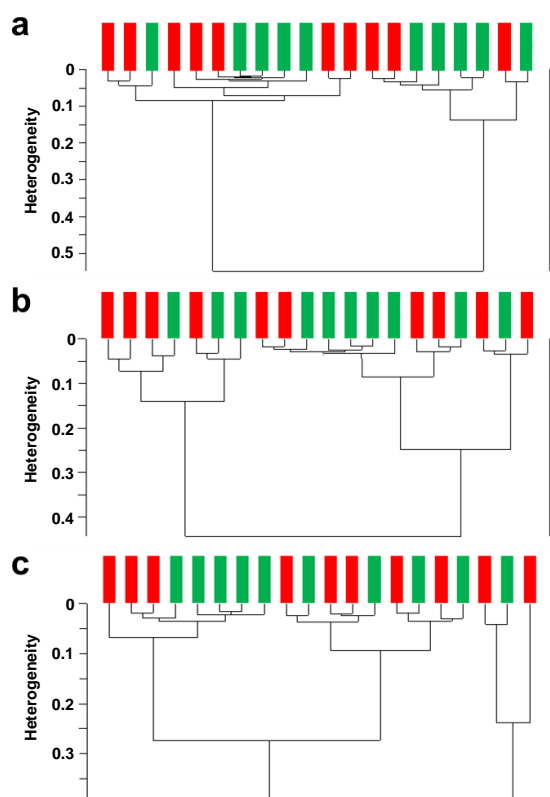
Figure 5.3 *Example of the inter-aliquot reproducibility assessment by HCA of 50 µL of serum 1 (**a**), 2 (**b**) and 3 (**c**) into 0.2 mL capacity plastic tube. Second derivative instrumental replicate spectra of aliquots 1 and 2 are represented in red and green respectively over the spectral range 4000-800 cm^{-1}.*

V.5.2 Serum and plasma long-term storage impact

Careful considerations must be taken when using different types of plastic tubes for the biofluid storage due to the different adsorption characteristics of the tube wall that can impact the entities concentrations measured[17] or the possibility of a contamination by polymer[18]. Potential molecular changes due to the use of different type of tubes and volumes over the month freezing time were therefore tested. These parameters were evaluated for each experiment by PCA each month on the spectral range 4000-800 cm[-1]. Again, the comparison of the spectra of the different volumes/tubes per healthy subject has shown no evidence of a differentiation of the spectra over the months (Figure 5.4 and AII.3). Furthermore, the serum or plasma spectra of the 3 subjects combined exhibit the same behaviour on each single experiment performed. The spectra of the serum or plasma 1 are well separated from the spectra of the serum/plasma 2 and 3 (Figure 5.5 and AII.4). From these observations, it can be concluded that neither the volume of sample stored or the type of tube have an impact on spectral collection, at least over a period of 9 months at -80°C. These results are consistent with conclusions of proteomics studies observing no significant changes in serum and plasma long-term storage at -80°C[19–21].
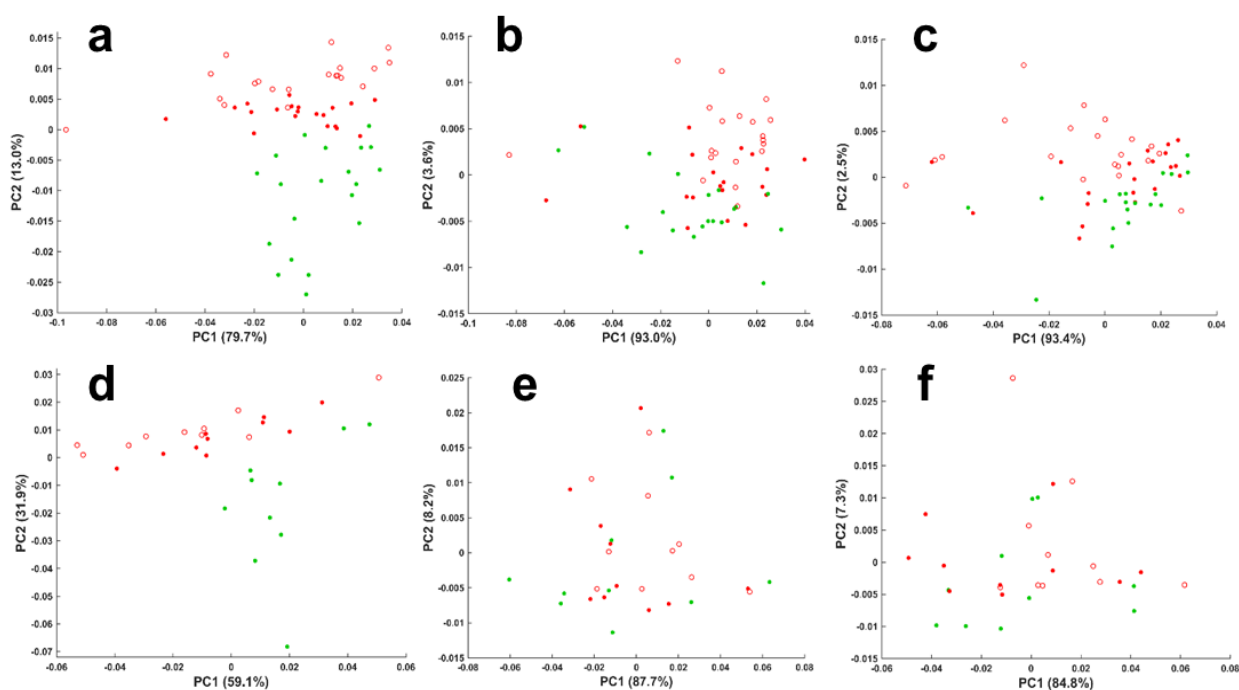
Figure 5.4 *Example of the volume and tube impact assessment by PCA of serum 1 (**a, d**), 2 (**b, e**), 3 (**c, f**) on the first (a-c) and last experiment (d-f) performed. Spectra of 50 µL of serum stored in 0.2 mL capacity tubes are represented by green dots. Spectra of 50 µL or 100 µL of serum stored in 1.5 mL capacity tubes are represented by red dots and circles respectively.*
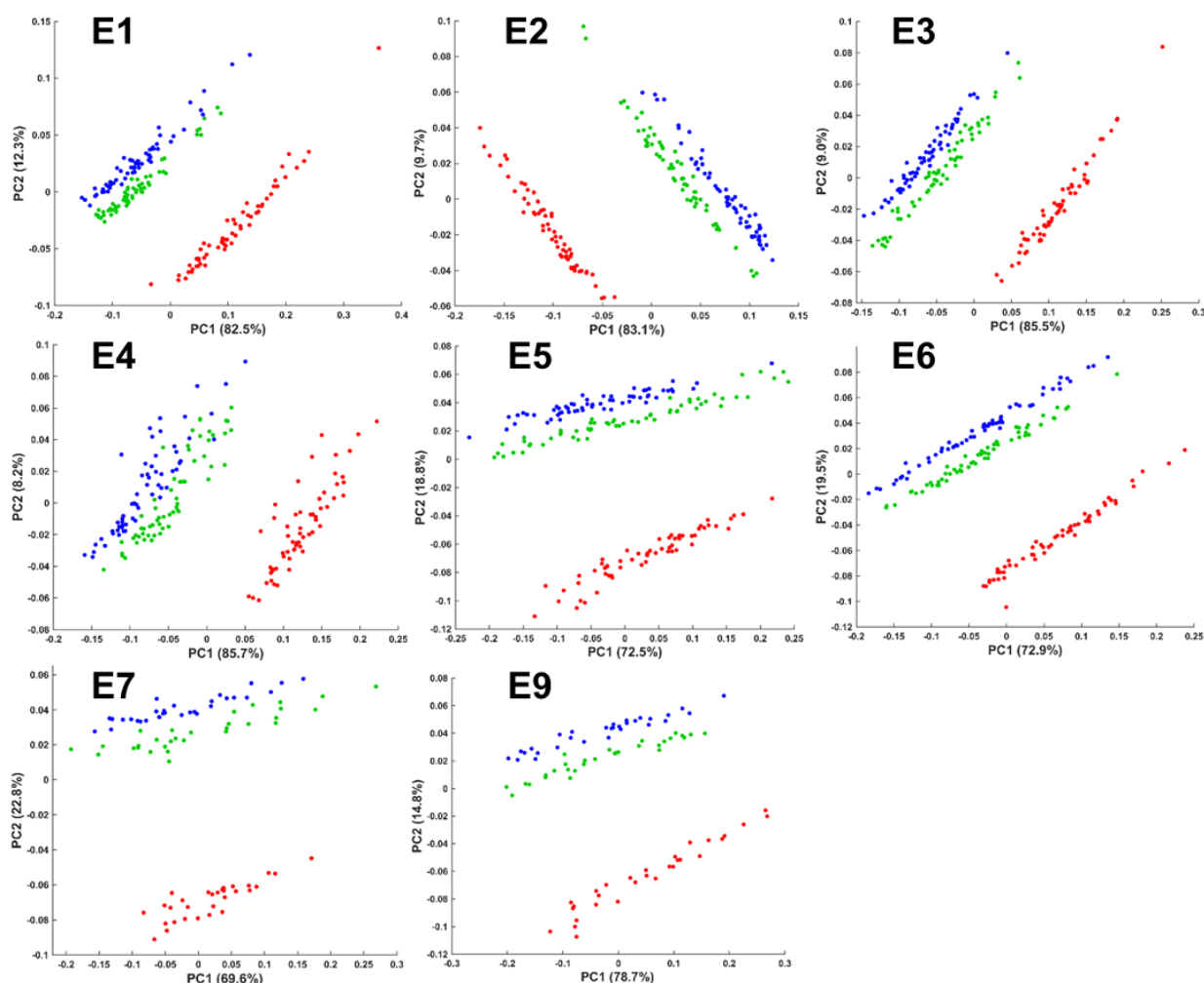
Figure 5.5 *PCA performed on 3 subject serum spectra (S1: red, S2: blue, S3: green) in the spectral range 4000-800 cm^{-1} over a 9 month experiments (**E1-E9**) to assess the impact of a long-term storage. PCA on E8 data was not performed as the majority of the spectra of serum 1 and 2 were discarded.*

From the 2 previous results attesting the spectral reproducibility inter aliquots and conditions of storage, all serum normalised absorbance spectra (Figure 5.6a) were combined and analysed by Principal Component Analysis (PCA) over the spectral range 4000-800 cm^{-1}. The differentiation expected between the sera samples (Figure 5.6b) is not observed. The data of the 3 sera are mixed, cluster pattern shifts are also visible for each serum (Figure 5.6b). To understand this phenomenon, the same PCA was performed highlighting the spectra of the 9 experiments (E1-9) performed rather than the subject spectra (Figure 5.6c). It can be seen from the data that they are separated based upon the dates of experiment and the shift is correlated to a gradient of RH recorded during the experiments (Table 5.1). This result suggests that there is a

difference in water content in the instrumental replicates of the serum samples responsible for this data shift amongst the date of experiments. To confirm this hypothesis, PC loadings of the data were studied. From the PC1 loading (Figure 5.6d) responsible for the data separation among the RH, it can be seen that the major negative band at 3470 cm^{-1} is attributed to water bond vibrations[22]. It can be noticed that the same results were obtained from the plasma samples (AII.5).
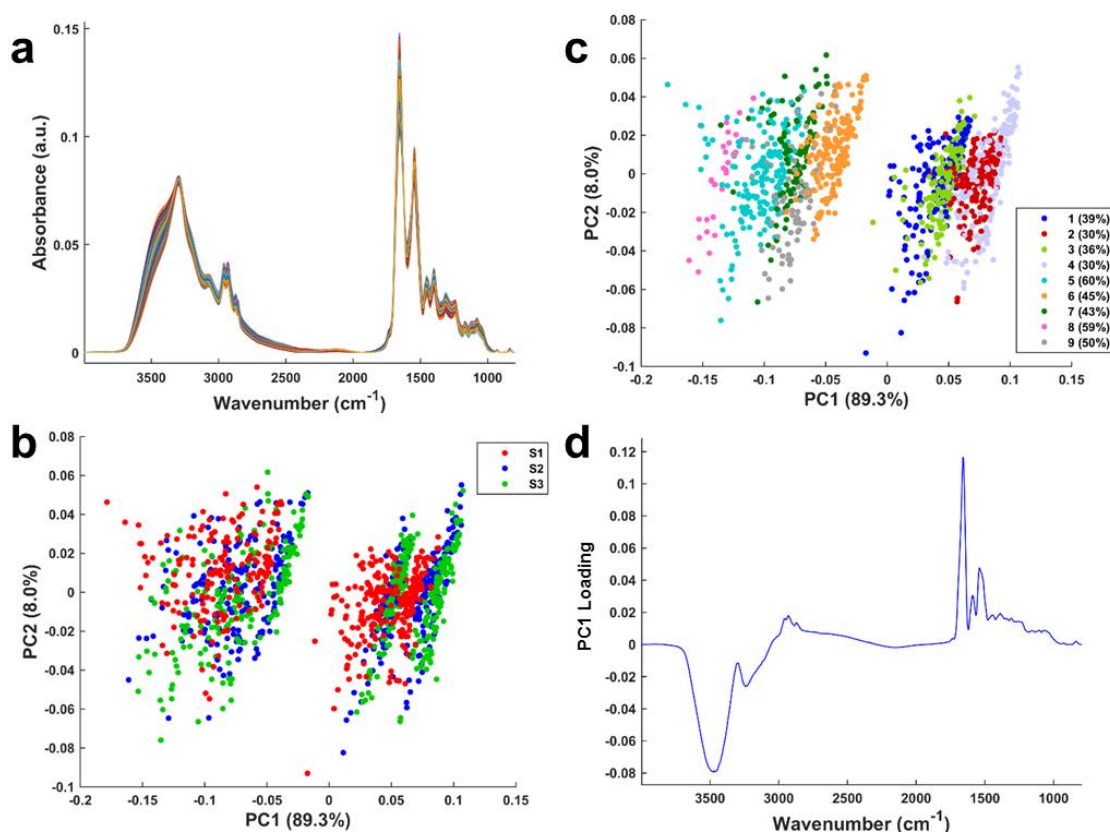
Figure 5.6 *Month to month spectral data collection reproducibility. **(a)** All normalised absorbance serum spectra of 3 subjects (S1, S2, S3). **(b)** PCA scatter plot of the 3 subject serum spectra. **(c)** PCA scatter plot of the same serum spectra highlighting the 9 experiments performed with associated relative humidity (%). **(d)** PC1 loading from (b) and (c).*

V.5.3 Experimental correction

In an attempt to experimentally remove the spectral variance caused by water bond vibrations, serum and plasma samples were left to dry under vacuumed conditions (90 min) in order to achieve a controlled and constant low RH prior to the spectral acquisition and compared with the same samples dried at room temperature (90 min)

with different RH recorded. The spectra collected were analysed using PCA, the Figure 5.7a shows that the spectra of the subject samples obtained under vacuum drying are still separated by the date of experiments (Figure 5.7b) rather than the subjects following the same pattern as the data collected from samples dried at room temperature even though both datasets can be distinguished showing that the vacuum drying has an effect on the samples especially at 1655 cm^{-1} (amide I band). This result suggests that the samples were rehydrated before and/or during the spectral acquisition at the same RH recorded at room temperature despite the presence of a desiccant in the spectrometer. Similar results were obtained from plasma samples (AII.6).
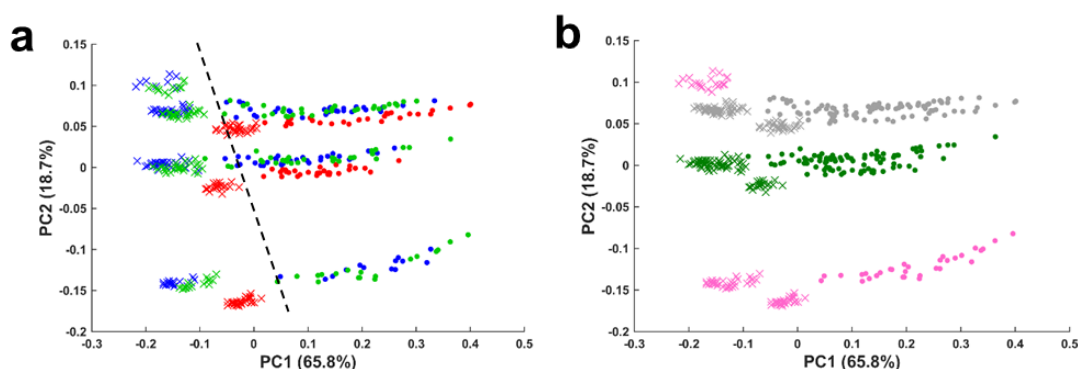
Figure 5.7 *Serum spectral collection experimental correction by sample vacuum drying. **(a)** PCA scatter plot of the 3 serum spectra (red, blue, green) obtained after samples dried at room temperature (dots) or under vacuumed conditions (crosses). **(b)** PCA scatter plot of the same serum spectra highlighting the 3 experiments at a RH of 43% (green), 50% (grey) and 59% (pink).*

V.5.4 Spectral data correction

From the PCA scatter plot (Figure 5.6c), it can be seen that the PC1 is responsible for the data separation based upon the RH recorded during the serum analysis rather than from sample molecular changes. The loading of the PC1 (Figure 5.6d) shows a strong contribution of water at 3470 cm^{-1}. In order to remove the residual water content from the serum and plasma spectral data, the influence of the PC1 was removed from the normalised absorbance data and spectra were reconstructed with all remaining PCs. The spectra obtained after correction are presented in Figure 5.8a. A PCA was then performed, the 3 different sera can be distinguished among the PC2 (Figure 5.8b) and the data are no longer separated among the RH recorded during experiments (Figure

5.8c). The loading of the PC2 (Figure 5.8d) responsible for the serum sample separation does not exhibit a contribution of water anymore. This correction approach allows to reduce the mean standard deviation (Table 5.2) of a factor 16 over the spectral range 3600-3320 cm^{-1}, as well as of a factor 2.5 over the spectral ranges 1720-1600 cm^{-1} (amide I band) and 1600-1500 cm^{-1} (amide II band).
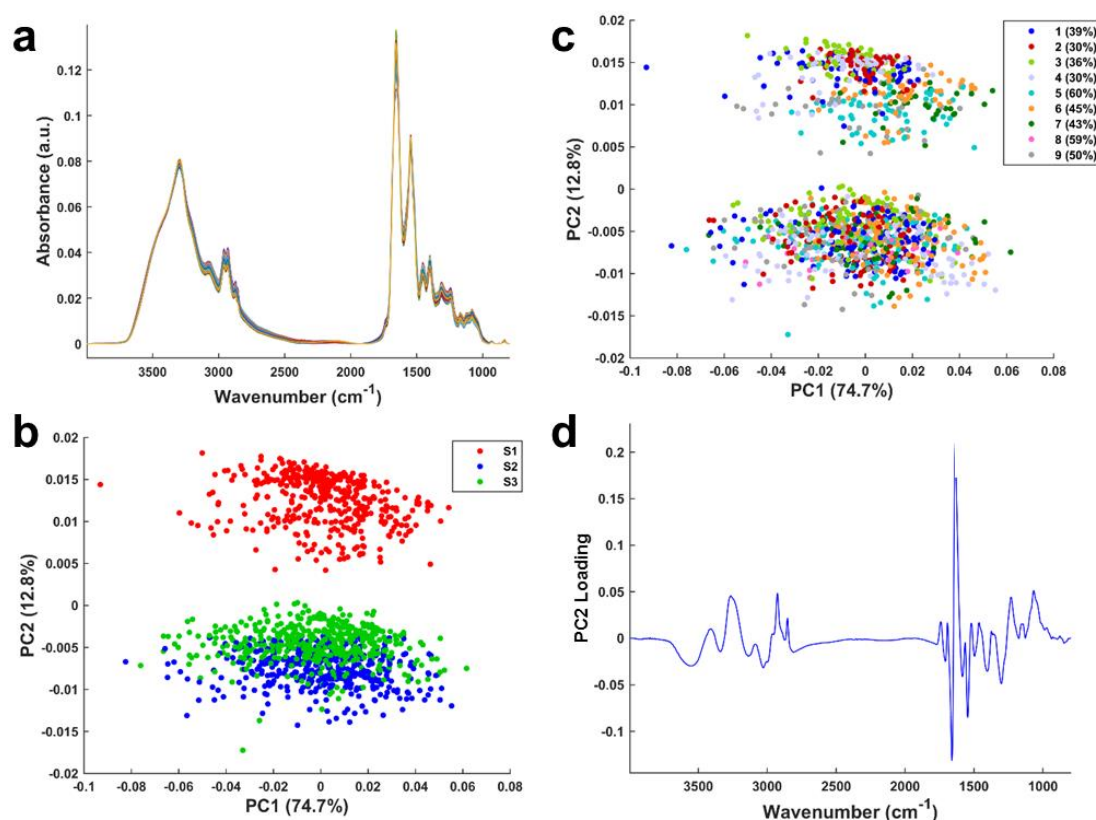
Figure 5.8 *Spectral data water correction by PC removal. (**a**) All normalised absorbance serum spectra of the 3 subjects reconstructed without the PC1 from Figure 6d. (**b**) PCA scatter plot of the 3 subject serum spectra. (**c**) PCA scatter plot of the same serum spectra highlighting the 9 experiments performed with associated relative humidity (%). (**d**) PC2 loading from (b) and (c).*

Table 5.2 *Mean standard deviation (STD) calculated on 3 spectral regions according to different serum spectral water correction methods.*

Correction	Mean STD		
	3600-1320 cm^{-1}	1720-1600 cm^{-1}	1600-1500 cm^{-1}
No correction	4.47e^{-03}	3.84e^{-03}	2.71e^{-03}
PC removal	2.80e^{-04}	1.53e^{-03}	1.09e^{-03}
EMSC$_{NS}$	8.03e^{-04}	2.70e^{-03}	1.30e^{-03}
EMSC$_{NS}$+ PC removal	3.70e^{-04}	1.40e^{-03}	9.55e^{-04}

NS: Normal Saline

An EMSC approach based on a constituent spectrum of normal saline (Figure 5.2) was also considered to correct the data[16]. However, it can be seen in the 3600-3320 cm^{-1} region that this approach does not compensate enough the spectra water content (Figure 5.9a) even if the standard deviation was reduced by a factor of 6 (Table 5.2). PCA results show that patient spectra are not distinguishable (Figure 5.9b) and that the data are separated based upon the date of analysis (Figure 5.9c). The PC1 loading exhibits remaining water spectral features in the 3600-3320 cm^{-1} region (Figure 5.9d).
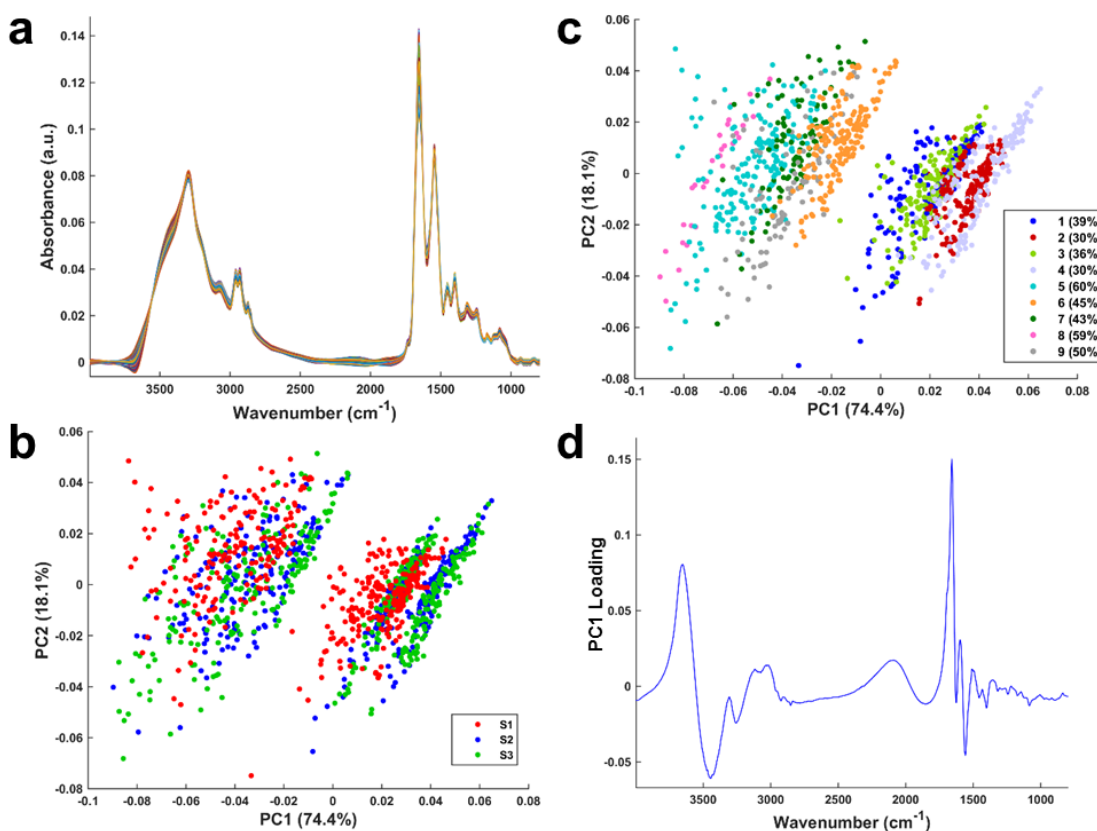
Figure 5.9 *Spectral data correction by EMSC and a constituent spectrum. **(a)** All normalised absorbance serum spectra after an EMSC with a normal saline constituent spectrum. **(b)** PCA scatter plot of the 3 subject serum spectra. **(c)** PCA scatter plot of the same serum spectra highlighting the 9 experiments performed with associated relative humidity (%). **(d)** PC1 loading from (b) and (c).*

As the EMSC approach is improving the standard deviation of the spectra in the water region, a combination with the PC1 (Figure 5.9d) removal was evaluated. The standard deviation of the spectra (Figure 5.10a) for the 3 zones assessed (Table 5.2), is comparable to the PC1 removal strategy (Figure 5.8). It can be observed that the data

are separated based upon patient spectral signatures (Figure 5.10b) and not on the RH recorded during experiments (Figure 5.10c). A better separation between the patient data is visible, the variance inter spectra replicates being decreased, suggesting that other spectral zones than those controlled in Table 5.2 were affected by this correction compared to the PC1 removal approach (Figure 5.8). However, the PC2 loading features (Figure 5.10d) responsible for the patient discrimination are similar to those obtained by removing only the PC1 (Figure 5.8d). The same results were obtained with the plasma samples (AII.7-10).
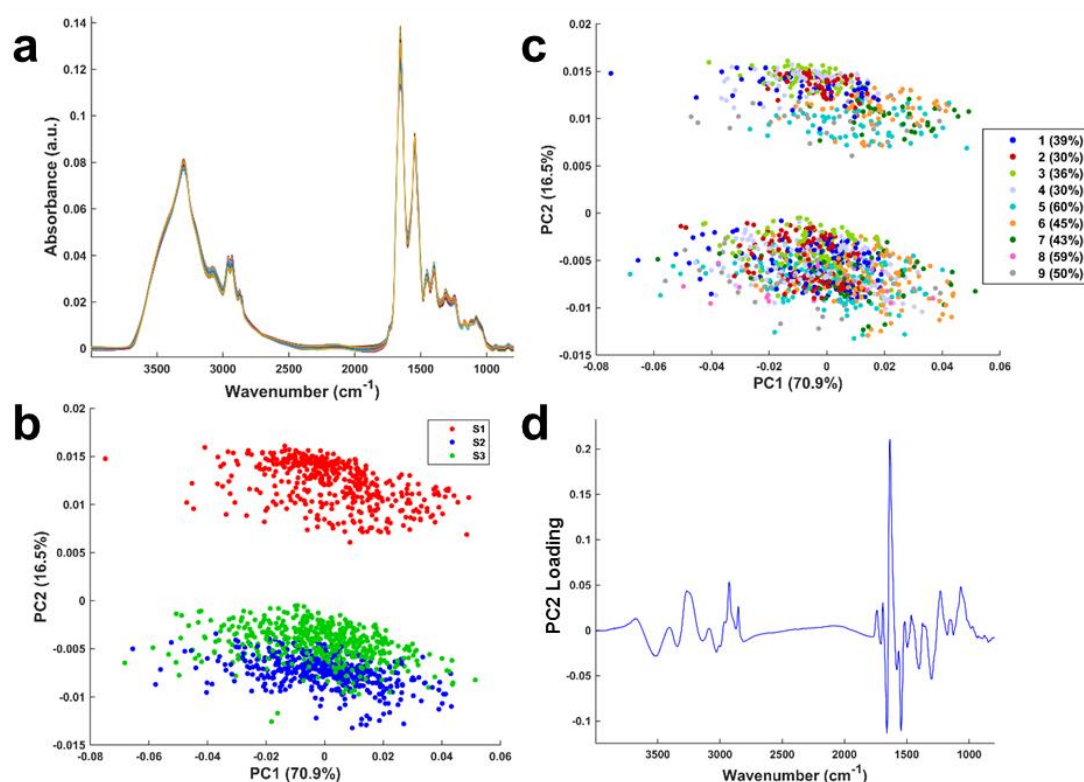
Figure 5.10 *Spectral data correction by a constituent spectrum based EMSC and PC removal. (a) All normalised absorbance serum spectra after an EMSC with a water constituent spectrum and without the PC1 from Figure 9d. (b) PCA scatter plot of the 3 subject serum spectra. (c) PCA scatter plot of the same serum spectra highlighting the 9 experiments performed with associated relative humidity (%). (d) PC2 loading from (b) and (c).*

V.6 Conclusions

The experiments performed on 3 sera and plasma of the same subjects have allowed to estimate and identify pre-analytical phase errors for biofluid spectroscopy applications. It has shown that a serum/plasma long-term freezing period (-80°C) of storage has no impact on the spectral signal of the samples analysed. Moreover in our study, the use of different plastic tubes for the biofluid storage has no influence on the spectra collected. However, while evaluating the month to month reproducibility, it has been observed that the sample drying at room temperature resulted in a difference in the sample water content. It demonstrates that the drying of the serum and plasma samples depend on the relative humidity of the analysis environment. Additional devices such as purging systems could be implemented to control the ambient air humidity during sample drying and analysis but such implementation increases the total cost of the spectral analysis. Another approach is to directly correct the spectral data with a different water content. From the different methods evaluated, it can be concluded that the best data correction performance was the combination of an EMSC solvent based constituent spectrum and the removal of the PC responsible for the discrimination based on the sample water content. However, it has to be noted that this methodology has been developed on a simple model of 3 samples to understand the spectral variations induced by environmental factors, implying that the natural variability inter patients from a large dataset has been minimised and could influence the efficiency of the correction applied.

In conclusion, to perform biofluid spectroscopy with a clinical application aim, careful experiment design and most importantly pre-analytical phase error evaluation and tracking is mandatory to ensure to target the correct molecular information rather than environmental interferences.

V.7 References

1   A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin and P. L. Martin-hirsch, *J. Biophotonics*, 2014, **165**, 153–165.

2   K. Spalding, R. Board, T. Dawson, M. D. Jenkinson and M. J. Baker, *Brain*

Behav., 2016, **502**, 1–8.

3   M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszewski, G. Thiéfin and G. D. Sockalingum, *Chem. Soc. Rev.*, 2016, **45**, 1803–1818.

4   H. J. Byrne, M. Baranska, G. J. Puppels, N. Stone, B. Wood, K. M. Gough, P. Lasch, P. Heraud, J. Sulé-suso and G. D. Sockalingum, *Analyst*, 2015, **140**, 2066–2073.

5   P. Bonini, M. Plebani, F. Ceriotti and F. Rubboli, *Clin. Chem.*, 2002, **48**, 691–698.

6   E. Schleicher, *Anal. Bioanal. Chem.*, 2006, **384**, 124–131.

7   M. Plebani, *Ann. Clin. Biochem.*, 2009, **47**, 101–110.

8   M. Plebani, *Clin. Chem. Lab. Med.*, 2016, **54**, 1881–1891.

9   S. F. Green, *Clin. Biochem.*, 2013, **46**, 1175–1179.

10   L. Lovergne, P. Bouzy, V. Untereiner, R. Garnotel, M. J. Baker, G. Thiéfin and G. D. Sockalingum, *Faraday Discuss.*, 2016, **187**, 521–537.

11   X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon and G. D. Sockalingum, *Transl. Res.*, 2013, **162**, 279–286.

12   L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum and M. J. Baker, *Anal. Methods*, 2015, **7**, 7140–7149.

13   J. H. Ward, *J. Am. Stat. Assoc.*, 1963, **58**, 236–244.

14   H. Abdi and L. J. Williams, *Wiley Interdiscip. Rev. Comput. Stat.*, 2010, **2**, 433–459.

15   H. Martens and E. Stark, *J. Pharm. Biomed. Anal.*, 1991, **9**, 625–635.

16   N. K. Afseth and A. Kohler, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 92–99.

17   A. Perret-Liaudet, M. Pelpel, Y. Tholance, B. Dumont, H. Vanderstichele, W. Zorzi, B. ElMoualij, S. Schraen, O. Moreaud, A. Gabelle, E. Thouvenot, C. Thomas-Anterion, J. Touchon, P. Krolak-Salmon, G. G. Kovacs, A. Coudreuse, I. Quadrio and S. Lehmann, *Clin. Chem.*, 2012, **58**, 787–789.

18   S. Lista, F. Faltraco and H. Hampel, *Prog. Neurobiol.*, 2013, **101**–**102**, 18–34.

19   B. L. Mitchell, Y. Yasui, C. I. Li, A. L. Fitzpatrick and P. D. Lampe, *Cancer Inform.*, 2005, **1**, 98–104.

20   M. E. Hassis, R. K. Niles, M. N. Braten, M. E. Albertolle, E. H. Witkowska, C. A. Hubel, S. J. Fisher and K. E. Williams, *Anal. Biochem.*, 2015, **478**, 14–22.

21   C. Ellervik and J. Vaught, *Clin. Chem.*, 2015, **934**, 914–934.

22   S. Y. Venyaminov and F. G. Prendergast, *Anal. Biochem.*, 1997, **248**, 234–245.

Chapter VI. Impact of water interferences in the diagnosis of sepsis patients by serum infrared spectroscopy

VI.1 Impact de l'interférence de l'eau sur le diagnostic sérique par spectroscopie infrarouge de patients atteints d'un sepsis

VI.1.1 Résumé

Le sepsis est une dérégulation de la réponse de l'hôte à une infection et est associé à un dysfonctionnement des organes engageant le pronostic vital du patient[1]. Chaque année, plus de 30 millions de cas de sepsis et 5 millions de décès sont estimés à travers le monde[2]. Le diagnostic du sepsis est basé sur des signes cliniques non spécifiques et la longue procédure d'identification des pathogènes responsables de l'infection[1,3]. Il a été récemment reconnu que la recherche d'une combinaison de biomarqueurs plutôt qu'une cible spécifique serait plus efficace dans le cas du diagnostic du sepsis[4,5]. La spectroscopie vibrationnelle peut fournir une ''empreinte moléculaire'' de matériaux biologiques et présente de nombreux aspects qui en font un candidat idéal pour le diagnostic de maladies. En effet, cette technique est sans marquage, non destructive, sans contact, simple d'utilisation, et ne requiert pas une préparation d'échantillons lourde[6]. Dans cette étude la technique de spectroscopie infrarouge à haut débit a été utilisée pour acquérir le profil moléculaire complet d'échantillons de sérums humains dans le but d'améliorer le diagnostic de patients atteints d'un sepsis.

Les échantillons de sérums (n=913) ont été obtenus de 294 patients ayant subi une chirurgie à haut à risque (résections au niveau du foie, des reins, des poumons, intestins, estomac ou pancréas). Le groupe de patients septicémiques comprend des échantillons de sérums (n=380) collectés avant chirurgie et jusqu'à 3 jours avant le jour du diagnostic, ainsi que le jour du diagnostic. Les échantillons de sérums (n=353) du groupe contrôle ont été obtenus de patients ayant un profil similaire en termes d'âge, de sexe, de procédure chirurgicale subie mais n'ayant pas développé de sepsis. Les échantillons ont également été obtenus suivant la même cinétique de collecte. De même, des échantillons (n=180) de patients atteints d'un syndrome de réponse inflammatoire systémique (SRIS) collectés avant chirurgie et le jour du diagnostic ont aussi été analysés.

VI.1.2 Conclusions

Il a été observé que la différence de contenu en eau des échantillons de sérums de patients après séchage, due à une variation de l'humidité relative dans l'air lors des différentes expériences, a une forte influence sur l'acquisition et l'analyse des données spectrales. Cela montre l'importance d'une conception expérimentale robuste et d'une standardisation des protocoles pour éviter tout biais pouvant amener à la mauvaise interprétation des données spectrales et à des résultats erronés.

Concernant les différents types de corrections de l'eau testés, l'approche par EMSC (Extended Multiplicative Signal Correction) basée sur un spectre constituant de solution saline est inefficace car celle-ci ne permet de corriger que les molécules d'eau non-liées de l'échantillon. Or, il est connu que l'eau interagie avec les protéines[29,30]. A l'inverse, les corrections basées sur une composante principale (CP) sont plus efficaces pour résoudre le problème de différence de contenu en eau des échantillons car elles reflètent la présence de molécules d'eau liées aux protéines. Cependant, la diminution presque totale de chaque marqueur identifié par analyse Random Forest, liée à la correction basée sur la suppression de la composante principale 1 groupe par groupe, est probablement due à une perte d'informations spectrales liées aux patients qui sont incluses dans les CPs. Les sensibilités et spécificités élevées sont ainsi questionnables et montrent les limitations des corrections basées sur des CPs. De plus, le contenu en eau des échantillons de sérums peut être directement lié à leur composition et être utile à leur discrimination. La meilleure façon de confirmer les hypothèses précédentes serait de réaliser à nouveau les expériences dans un environnement contrôlé pour permettre le séchage et la collecte de spectres dans les mêmes conditions.

Dans cette étude, des profils spectraux tels que les bandes amide I et II, les glucides et acides nucléiques, ont été mis en évidence à plusieurs reprises dans la discrimination des patients septicémiques et contrôles avec des sensibilités et spécificités supérieures à 70%. De plus, les résultats d'analyses de spectres d'échantillons collectés jusqu'à 3 jours avant et le jour du diagnostic d'un sepsis, ont montré des sensibilités et spécificités similaires dans la distinction de patients septicémiques et contrôles, lorsque ceux-ci sont comparés à un jeu de données ne comprenant que des échantillons

collectés un seul jour avant et le jour du diagnostic. Cela suggère que les profils spectraux identifiés peuvent être de potentiels biomarqueurs pour le diagnostic précoce du sepsis. De plus amples recherches seront nécessaires pour confirmer ces résultats.

VI.2 Chapter overview

Sepsis is a dysregulated host response to an infection and is associated with life-threatening organ dysfunction[1]. Each year, over 30 million cases and 5 million deaths are estimated worldwide[2]. Diagnosis of sepsis is based on non-specific clinical signs and time consuming positive identification of the causative pathogen(s)[1,3]. It has been recently recognised that the search for a combination of biomarkers rather than a single target would be more efficient for sepsis diagnosis[4,5]. Vibrational spectroscopy can provide a "molecular fingerprint" of biological materials and presents many aspects that make it an ideal candidate for disease diagnosis purposes. Indeed, this technique is label-free, non-destructive, non-contact, rapid, cost-effective, simple to operate, and requires only simple sample preparation[6]. In this study we used High-Throughput Fourier Transform Infrared (HT-FTIR) spectroscopy, following the methodology developed throughout the previous chapters, to acquire the whole human serum sample molecular profile with the aim to improve diagnosis of patients with sepsis.

VI.3 Introduction

Despite its high incidence and mortality rate, no gold standard technique exists for sepsis diagnosis due to complex underlying pathogenesis mechanisms[7,8]. Over a hundred potential biomarkers have been assessed for the diagnosis of sepsis and yet none have shown enough sensitivity and/or specificity for a clinical application[4,5]. Sepsis patients are not easily differentiated from patients with Systemic Inflammatory Response Syndrome (SIRS), caused by non-infectious conditions (e.g., trauma, burn, pancreatitis) and which requires different medical care[9]. However, time to treatment is a crucial factor for sepsis patient survival[10–12]. A rapid and early diagnosis would enable a reduction in mortality and morbidity rates among sepsis patients. Administration of targeted therapeutics at an early stage would also enable to reduce unnecessary antibiotic use, preventing multidrug resistance[13] and would improve patient outcome by limiting the severity of associated sequelae[14,15]. Overall, it would allow the reduction of healthcare costs[16].

In the recent field of biofluid spectroscopy, several studies have proved the potential of the technique for the diagnosis of different cancer types as well as other various diseases[17]. Regarding sepsis, a proof of concept study using Raman spectroscopy applied on plasma samples, has shown the differentiation of 31 sepsis patients from 39 sterile SIRS patients with 100% sensitivity and 82% specificity[18].

In the present study, High-Throughput Fourier Transform Infrared (HT-FTIR) spectroscopy has been evaluated for the rapid pre-symptomatic diagnosis of sepsis from serum samples. Previous investigations have demonstrated that HT-FTIR is a technique of choice as it allows to perform an automated analysis of a large number of samples quickly compared to other IR systems available[19]. This method presents the advantage to be a macro-analysis which can ensure the reproducibility of spectral acquisition from serum samples which have to be dried to overcome the high contribution of water in IR spectroscopy but which have also shown to exhibit inhomogeneous patterns after drying (chapter II and III). Then, in an attempt to standardise protocols in sample handling and preparation for spectral acquisitions several parameters that can impact either the samples (e.g., solvents, collection tubes, freeze-thaw cycles, long-term storage, drying methods) and/or the spectroscopic analysis (e.g., deposit methods, volumes, substrates, humidity, operators dependency) have been assessed (chapter IV and V). A major drawback found was that serum sample drying was dependant on the relative humidity (RH) in the air, ending up in variations in sample water content and consequently in spectra collected. Furthermore, in a simple model of 3 serum samples it has been seen by Principal Component Analysis (PCA), an unsupervised multivariate analysis, that discrimination of samples was based upon water content rather than on patient spectral information. Different data corrections based on Principal Component (PC) removal and/or Extended Multiplicative Signal Correction (EMSC) were successful to overcome the variation of water content of the 3 same sera recorded at different RH. However, efficacy of such corrections and the impact of water variations on the discrimination of patients from a large sample set are unknown. This study investigates the influence of water interferences on the discrimination of sepsis patients by PCA and Random Forest (RF), a supervised multivariate analysis.

VI.4 Material and methods

VI.4.1 Serum samples

Serum samples (n= 913) were obtained from 294 patients needing elective high risk surgery (e.g. liver and renal surgery, lung resection, large bowel resection, gastrectomy, Whipple's procedure) and were provided by the Defence Science & Technology Laboratory (Dstl), Porton Down, United Kingdom. Patients aged under 18 and above 80, pregnant and any whom were immunosuppressed were excluded from the study. A consent form was signed by all patients recruited at the medical care centres involved (Liverpool Royal and University Hospital; University College Hospital, London; Bristol Royal Infirmary; Guy's & St Thomas' Hospital, London; Birmingham Queen Elizabeth Hospital and University Hospital, Frankfurt). The project has been approved by Southampton & South West Hampshire Research Ethics Committee (A), 1st Floor, Regents Park Surgery, Shirley, Southampton, Hants, SO16 4RJ (reference number 06/Q1702/152).

Blood samples were collected before surgery and daily for up to 7 days post-surgery or until patient discharge (control group) or diagnosis of sepsis/ SIRS. Clotted blood samples (4 mL) were centrifuged and extracted serum samples were stored at -80 °C until experiments were performed.

The diagnosis of sepsis was based on the presence of 2 conditions:

1) Two or more of the 4 SIRS criteria (from the previous definition of sepsis[21] as patients were recruited before 2016):

- Body temperature > 38°C or < 36°C

- Heart rate > 90/min after fluid resuscitation

- Respiratory rate > 20/min or partial pressure of carbon dioxide ($PaCO_2$) < 32 mmHg

- White blood cell count > 12,000 cells/ μL or < 4,000 cells/ μL or > 10% immature neutrophils

2) Presence of an infection which is defined by an invasion of normally sterile tissue, body cavity, or fluid by pathogenic or potentially pathogenic micro-organisms

Diagnoses of sepsis, SIRS and control patients in the different care centres taking part in the study were reviewed and confirmed by a clinical advisory panel.

Samples in this study include:

- For the sepsis group, sera (n=380) obtained before surgery, up to 3 days before sepsis diagnosis and on the day of diagnosis.

- For the control group, sera (n=353) from age/ sex/ procedure-matched patients who did not go on to develop sepsis over similar time frames post-surgery.

- For the SIRS group serum samples (n= 180) collected before surgery and on the day of SIRS diagnosis from age/ sex/ procedure-matched patients

Repartition of the samples included in the study is presented in the Table 6.1. Among sepsis patients, a proportion of 70% males was observed with a patient mean age of 64 year-old. Among the 100 patients with sepsis, the cause of infection was due to Gram-negative bacteria (G^-) in 37 cases, Gram-positive bacteria (G^+) for 18 patients, due to a mix of G^-/G^+ in 15 cases or a mix of G^-/G^+/yeasts for 10 patients, 4 cases with yeasts and 1 patient with yeasts/virus. No microorganism growth was observed in 10 cases and causes were not specified for 5 patients.

Table 6.1 *Serum sample repartition in the sepsis, control and SIRS groups. Day 0 represents the day of diagnosis.*

	Sepsis	Control	SIRS
Pre-surgery	100	99	95
Day -3	57	57	–
Day -2	72	72	–
Day -1	99	95	–
Day 0	52	30	85

Note: Samples were received and analysed at 3 different periods which will be referred at batch 1, 2 and 3 in the result section. Batch 1 is composed of serum samples collected

from 70 patients in each sepsis and control groups; from 70 SIRS patients in batch 2 and from 30 patients of each sepsis, control, SIRS groups in batch 3.

VI.4.2 Infrared spectroscopy

Based upon an established methodology previously described in chapters 2 to 5, all serum samples were 3-fold diluted with normal saline (NaCl 0.9%). Samples from the different groups were randomly analysed to avoid any experimental bias. Five microliters per replicate and 8 replicates per sample were deposited onto a silicon plate (384 wells). Drops were dried at room temperature before spectral acquisition (90 min). Spectra were collected with a high-throughput screening HTS-XT extension coupled to a Tensor 27 spectrometer (both from Bruker Optics GmbH, Ettlingen, Germany). Spectral acquisition was performed in the transmission mode via OPUS v6.5 software (Bruker Optics GmbH, Ettlingen, Germany) and collected at a spectral resolution of 4 cm^{-1}, co-adding 32 scans over the spectral range of 4000-400 cm^{-1}. A background spectrum was acquired before each sample measurement and automatically removed from sample spectra. One spectrum is obtained over the whole surface of a well (~3 mm diameter) of a silicon plate.

VI.4.3 Pre-processing and data analysis

A quality test as previously described[20,22] was performed on raw spectra to control absorbance intensity, signal-to-noise ratio and water vapour content (OPUS v6.5 software, Bruker Optics GmbH, Ettlingen, Germany). A sample was discarded when less than half replicates met the quality test. Four percent of samples were rejected because of a too high absorbance and notably more than half of them were pre-surgery samples. Data pre-processing and multivariate analysis were realised with Matlab software version 7.11.0 (The Math Works, Inc., USA). All spectra were cut to the spectral range of 4000-800 cm^{-1} before baseline correction and vector normalisation. An exploratory study has been conducted on the data collected using Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA)[23,24].

Principal Component (PC) removal based corrections and/or Extended Multiplicative Signal Correction (EMSC)[25] based on an interference constituent spectrum[26] were tested on serum sample spectra to correct their variation in water content. As previously described in chapter V, the mean spectrum of all normalised absorbance spectra was defined as the reference spectrum, and a spectrum of normal saline was used as the interference spectrum to be removed from the dataset.

A Random Forest (RF) analysis was performed for the classification of sepsis patients against controls. RF is a machine learning technique which consists in averaging previsions of independent models called decision trees to reduce the variance and thus, the prediction error thanks to a high number of decision trees (n=500 in this study). Each tree is based on the original data random sampling and associated features (sub-sampling) that explain the maximum variance between classes[25].

VI.5 Results

VI.5.1 Preliminary study

In a first approach, sepsis and control samples were analysed from batch 1 individually. All spectra from each sub-group previously described were averaged. The mean spectra (Figure 6.1) present spectral variations around 3500 cm^{-1}, in the water region as well as differences over the "fingerprint region", especially over the spectral range of 1200-800 cm^{-1} which corresponds to carbohydrates, glycogen, and nucleic acids. Spectra from the sepsis group can be differentiated from the control group spectra and from the pre-surgery samples. This observation has been confirmed by HCA (Figure 6.2). Indeed, the sepsis group and the control group are separated in different clusters. Pre-surgery samples can be also distinguished and are more similar to the control than the sepsis group.
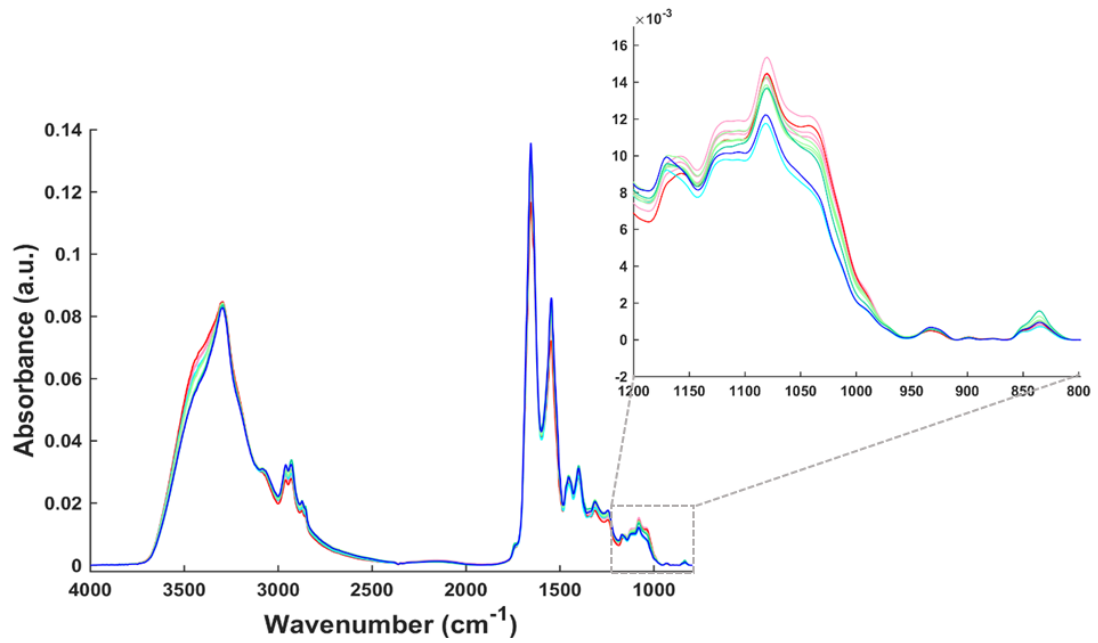
Figure 6.1 *HT-FTIR pre-processed mean serum spectra of patients from sepsis (pink and red) and control groups (light and dark green) as well as pre-surgery samples spectra (blue) over the spectral range of 4000-800 cm^{-1} and 1200-800 cm^{-1}.*
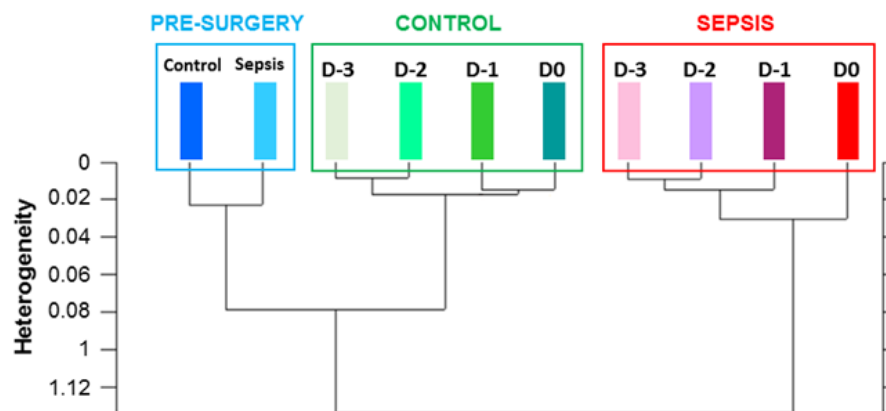
Figure 6.2 *HCA of HT-FTIR pre-processed mean serum spectra of patients from sepsis and control groups over the spectral range of 1800-900 cm^{-1}. D0 corresponds to the day of sepsis diagnosis or control equivalence.*

Then, PCA were performed on spectra from sera collected on the day of sepsis diagnosis (S0) compared to spectra of samples obtained up to 3 days before the diagnosis (S-1, S-2, S-3) and their matched control (C-1, C-2, C-3). S0 spectra are well separated from C-1 spectra, while S-1 spectra are overlapping with S0 and C-1 groups (Figure 6.3a). However, this tendency is not as clear for the comparison of S0/S-2/C-2 and S0/S-3/C-3 spectra (Figure 6.3c and e). Related PC1 loadings (Figure 6.3b, d, f) exhibit very similar features, mainly the water band around 3500 cm^{-1}, the amide I/II bands at 1658 a 1545 cm^{-1} and a peak at 1000 cm^{-1} corresponding to carbohydrates.
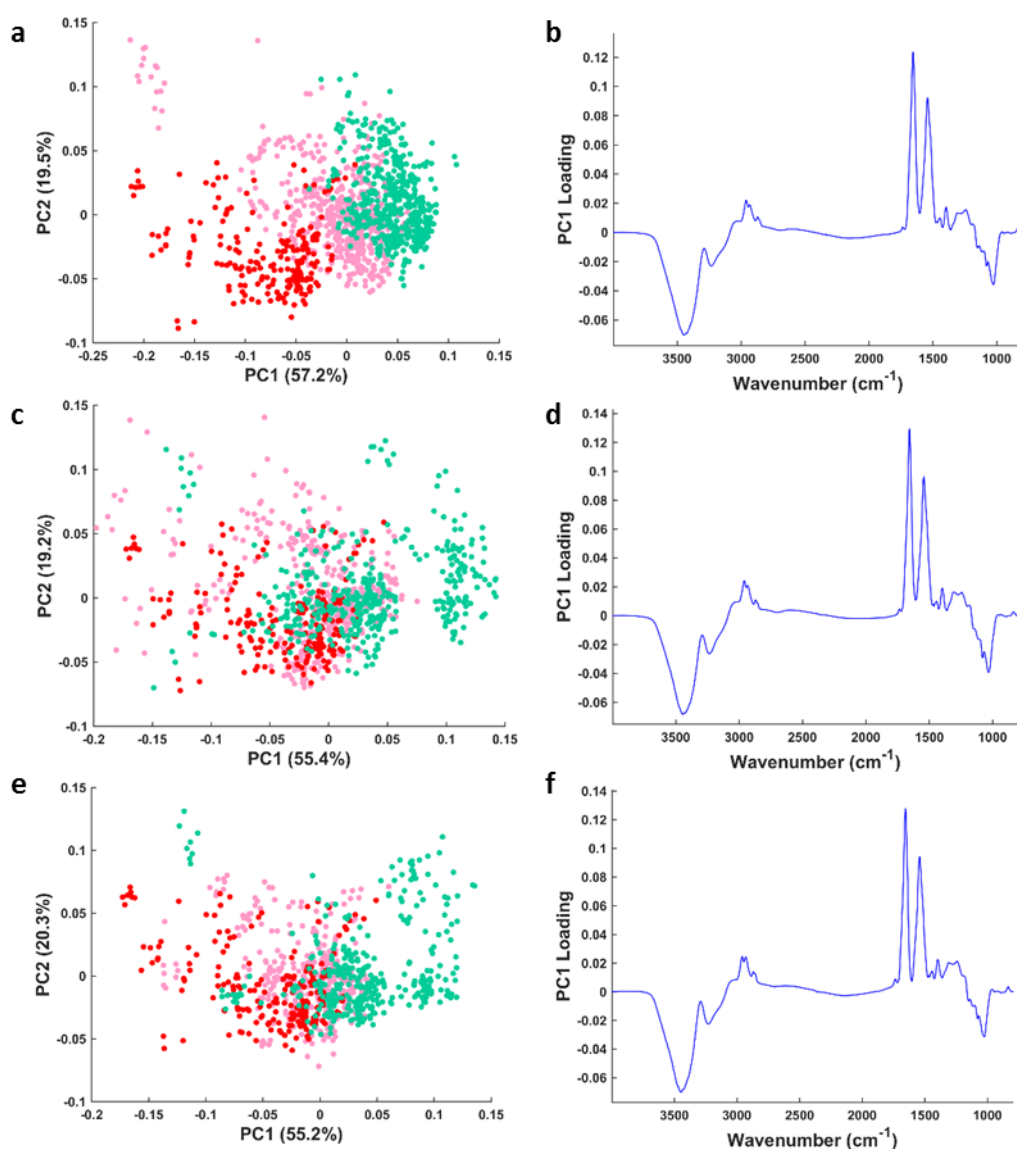
Figure 6.3 *PCA plots of HT-FTIR spectra of patients from sepsis and control groups from batch 1, S0/S-1/C-1 (**a**), S0/S-2/C-2 (**c**), S0/S-3/C-3 (**d**) over the spectral range of 4000-800 cm^{-1} and their respective PC1 loadings (**b, d, f**).*

A RF analysis, a more powerful method, was used for the classification of the sepsis group (including S0, S-1, S-2, S-3) *versus* the control group (C0, C-1, C-2, C-3). A sensitivity of 84%, specificity of 74% and accuracy of 79% were obtained. The feature importance is presented in Figure 6.4. The main features are attributed to the water band, lipids around 2960 cm[-1], proteins (amide I and II bands), nucleic acids peaks at 930 cm[-1] and around 830 cm[-1] [27,28].
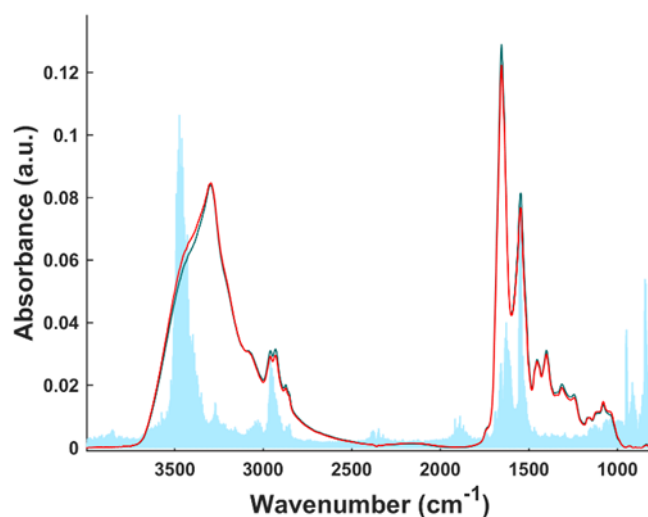
Figure 6.4 *RF feature importance (blue) of sepsis (S0/S-1/S-2/S-3) versus control samples (C0/C-1/C-2/C-3) from batch 1. The mean spectrum sepsis of (red) and control (green) group are presented for comparative purposes*

VI.5.2 Influence of water interferences on sample discrimination

The batch 1 sample set has been combined with samples from batch 2 and 3. All pre-processed spectra are presented in Figure 6.5a. Major spectral differences between serum samples are seen in lipid content (2955-2850 cm[-1] and 1737 cm[-1]), carbohydrates around 1000 cm[-1] and water content (~3500 cm[-1]). A PCA was applied as an exploratory method, the PCA plot (Figure 6.5b) shows that batch 1 spectra are separated from spectra of batch 2 and 3. It can be observed on the PC 1 loading (Figure 6.5c) that the water band is responsible for the data clustering observed rather than patient information or diagnosis.
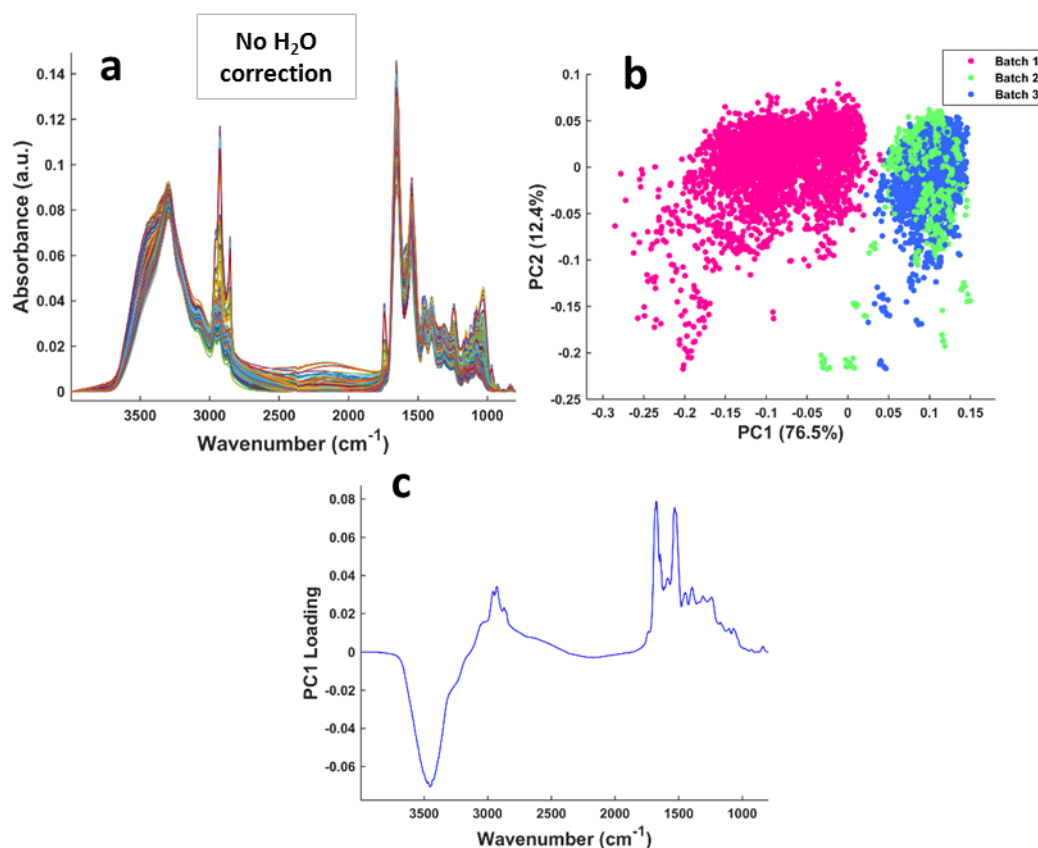
Figure 6.5 *Pre-processed spectra with no water correction from samples of all batches (**a**). Related PCA plot (**b**) and PC1 loading (**c**).*

To evaluate the impact of water on sample discrimination, different types of corrections were tested as previously described in chapter 5. The methods applied consisted of (1) removing the first PC (PC1r) responsible for the separation of the data based upon water, (2) using an EMSC with a constituent spectrum of normal saline (EMSC$_{NS}$) which is the solvent used to dilute the sample, (3) a combination of 1 and 2 (EMSC$_{NS}$ + PC1r) and (4) removing the first PC of each of the 12 sub-groups of samples individually before to apply a global PC1 removal (PC1r$_G$) to recombine all spectra in one dataset as the preliminary study performed on batch 1 suggested a difference in water also intra-batch (Figure 6.3 and 6.4). Spectra corrected with the 4 methods are presented in Figure 6.6. All corrections have reduced the variation of water content in the spectra. Similar results were obtained with the PC1r (Figure 6.6a) and PC1r$_G$ (Figure 6.6d). The EMSC$_{NS}$ approach has induced baseline distortions and seems to be the less efficient correction (Figure 6.6b) whereas the EMSC$_{NS}$ associated

with PC1r is more efficient to decrease water in spectra and has slightly improved the baseline (Figure 6.6c). The efficacy of the 4 corrections was assessed by PCA (Figure 6.7). As suspected from spectra (Figure 6.6b), the $EMSC_{NS}$ correction alone is not efficient to reduce water presence in the spectra as the 3 different batches are still distinguishable on the PCA plot and water features remain in the PC1 loading (Figure 6.7b). No separation of the batch spectra is observed with the 3 other corrections proving their efficiency as well as the water band is not present in their related PC1 loadings which are very similar for the 3 types of corrections (Figure 6.7a, c, d). It can be noted that on the PCA plot, spectra corrected by $EMSC_{NS}$ + PC1r show more heterogeneity in their distribution (Figure 6.7c) than with PC1r and $PC1r_G$ corrections (Figure 6.7a, d).
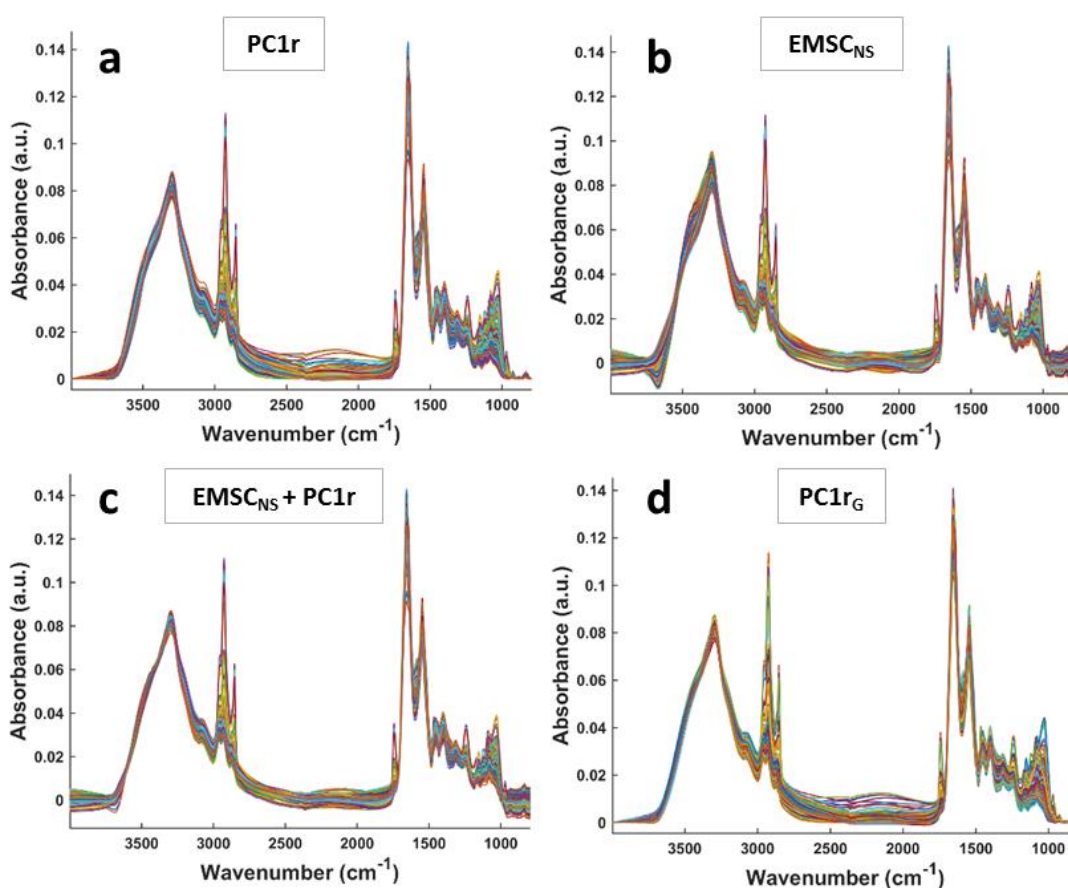
Figure 6.6 *Influence of the different water corrections on serum spectra. All pre-processed spectra corrected for water content by PC1r (**a**), EMSC_NS (**b**), EMSC_NS + PC1r (**c**) and PC1r_G (**d**) methods.*
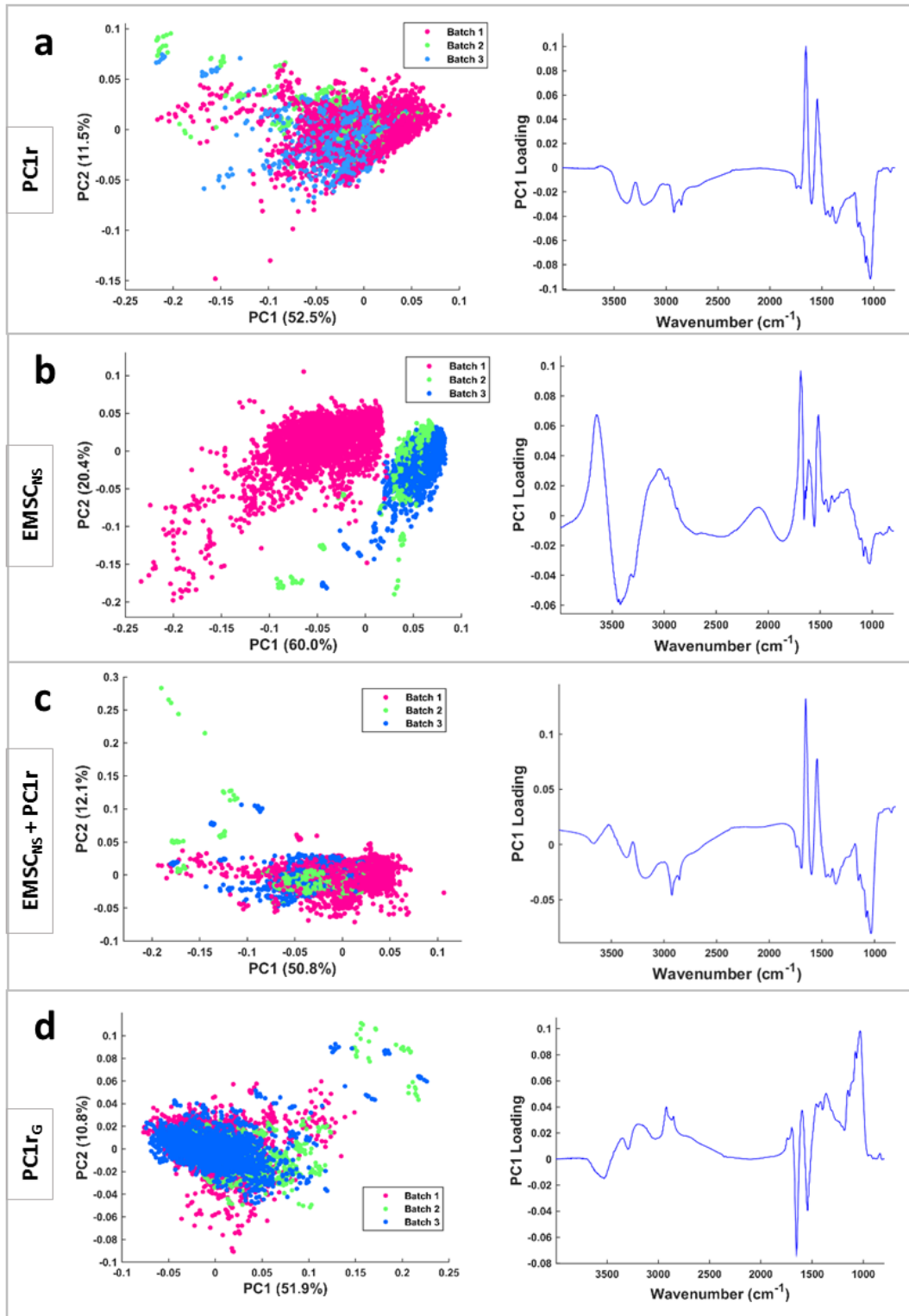
Figure 6.7 *Influence of the different water corrections on serum spectra assessed by PCA. PCA plots and PC1 loadings corresponding to all pre-processed spectra corrected for water content by PC1r (**a**), EMSC$_{NS}$ (**b**), EMSC$_{NS}$ + PC1r (**c**) and PC1r$_G$ (**d**) methods.*

As the EMSC$_{NS}$ method did not show any improvement for the water content correction, further investigations with the 3 other techniques were carried on sepsis day of diagnosis spectra (S0) and corresponding control spectra (C0) by PCA (Figure 8). On the PCA plot Figure 6.8a, two groups of sepsis/control spectra corresponding to 2 different batches (B1, B3) are separated based upon water (PC1 loading, Figure 6.8a). It can be seen that batch 1 (B1) sepsis sample spectra are well separated from control group spectra whereas both groups are mixed in batch 3 (B3). PCA plots Figure 6.8b and c, show that spectra of sepsis and controls samples corrected by the PC1r and EMSC$_{NS}$ +PC1r methods are mixed while a separation tendency is visible between sepsis and control data corrected with the PC1r$_G$ approach (PCA plot, Figure 6.8d). PC1 loadings (Figure 6.8b, c, d) are similar, exhibiting amide I/II bands and carbohydrate peak (1000 cm^{-1}). However, it can be noticed that the water band at ~3500 cm^{-1} present in the PC1 loading Figure 6.8d, is also contributing to the slight separation of the PC1r$_G$ corrected data.

Similar results were obtained while comparing spectra from sepsis and SIRS patient samples collected on their day of diagnosis by PCA (Figure 6.9). When data are not corrected, the spectra from the different batches (B1, B2/B3) are distinguishable based upon water (Figure 6.9a). After correction with the 3 different methods, spectra of both groups are mixed (Figure 6.9b, c, d). Spectra of sepsis and SIRS samples appear slightly less mixed with the EMSC$_{NS}$ + PC1r and PC1r$_G$ corrections, tendency which could be attributed to carbohydrates and lipids (PCA loadings, Figure 6.9c, d). However, these features could be resulting from the influence of a few patients with a highly different spectral profile as shown in Figure 6.10. For example, a high level of absorbance in the lipid region (2955-2850 cm^{-1}) can be observed on the mean spectra of 2 SIRS patients (spectrum 1 and 2: 0.08 and 0.06 a.u.) compared to the mean spectrum of the SIRS group (spectrum 0: >0.04 a.u.). Two other SIRS patient mean spectra exhibit a higher absorbance in the carbohydrate region (spectrum 3 and 4: 0.04 a.u.) than the SIRS group mean spectrum (spectrum 0: 0.02 a.u.).
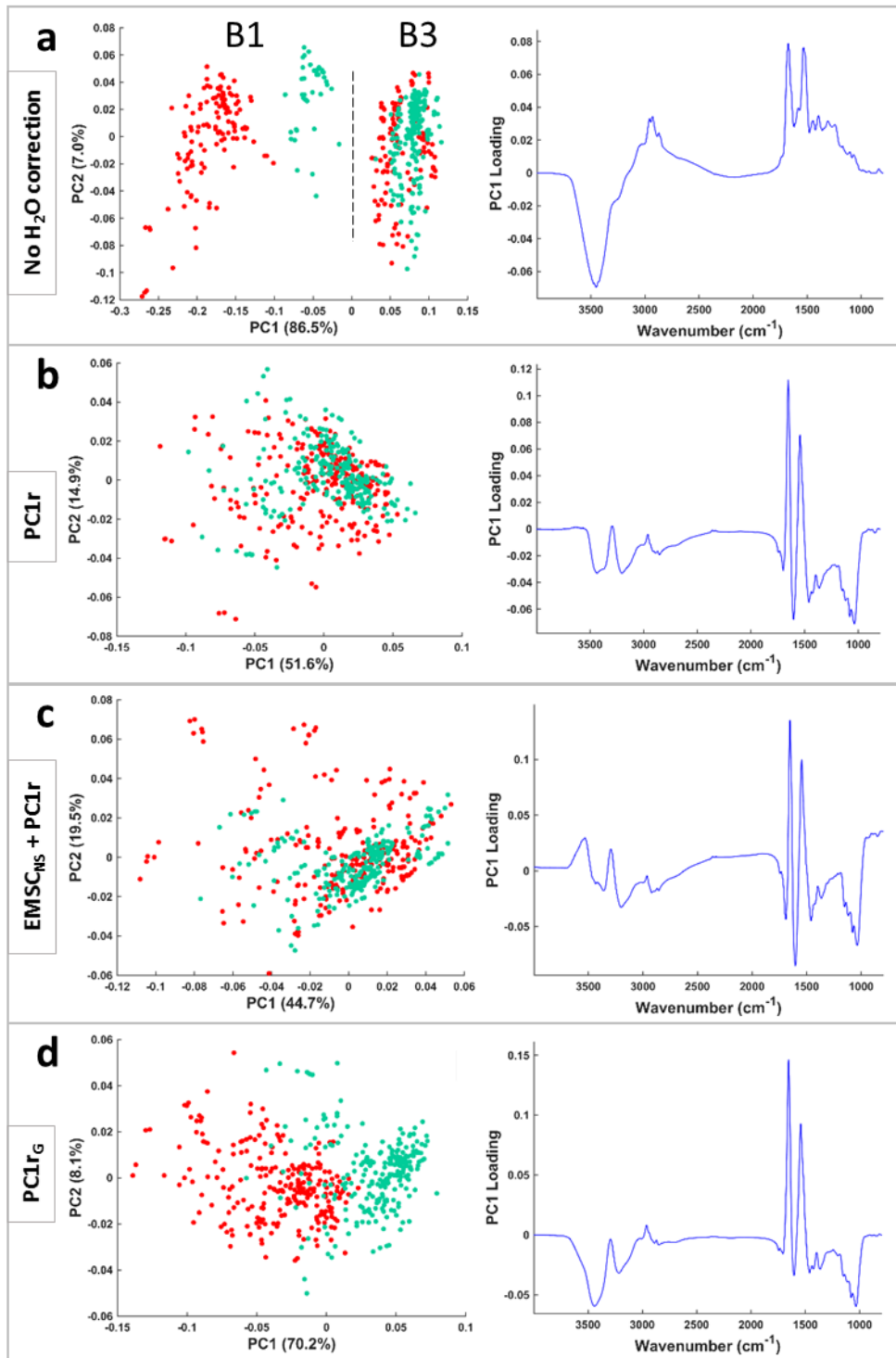
Figure 6.8 *Influence of the different water corrections on the discrimination of sepsis (S0, in red) and control (C0, in green) serum spectra assessed by PCA. PCA plots and PC1 loadings corresponding to all pre-processed spectra without water correction (**a**) or corrected for water content by PC1r (**b**), EMSC$_{NS}$ + PC1r (**c**) and PC1r$_G$ (**d**) methods. B1/3: batch 1/3.*
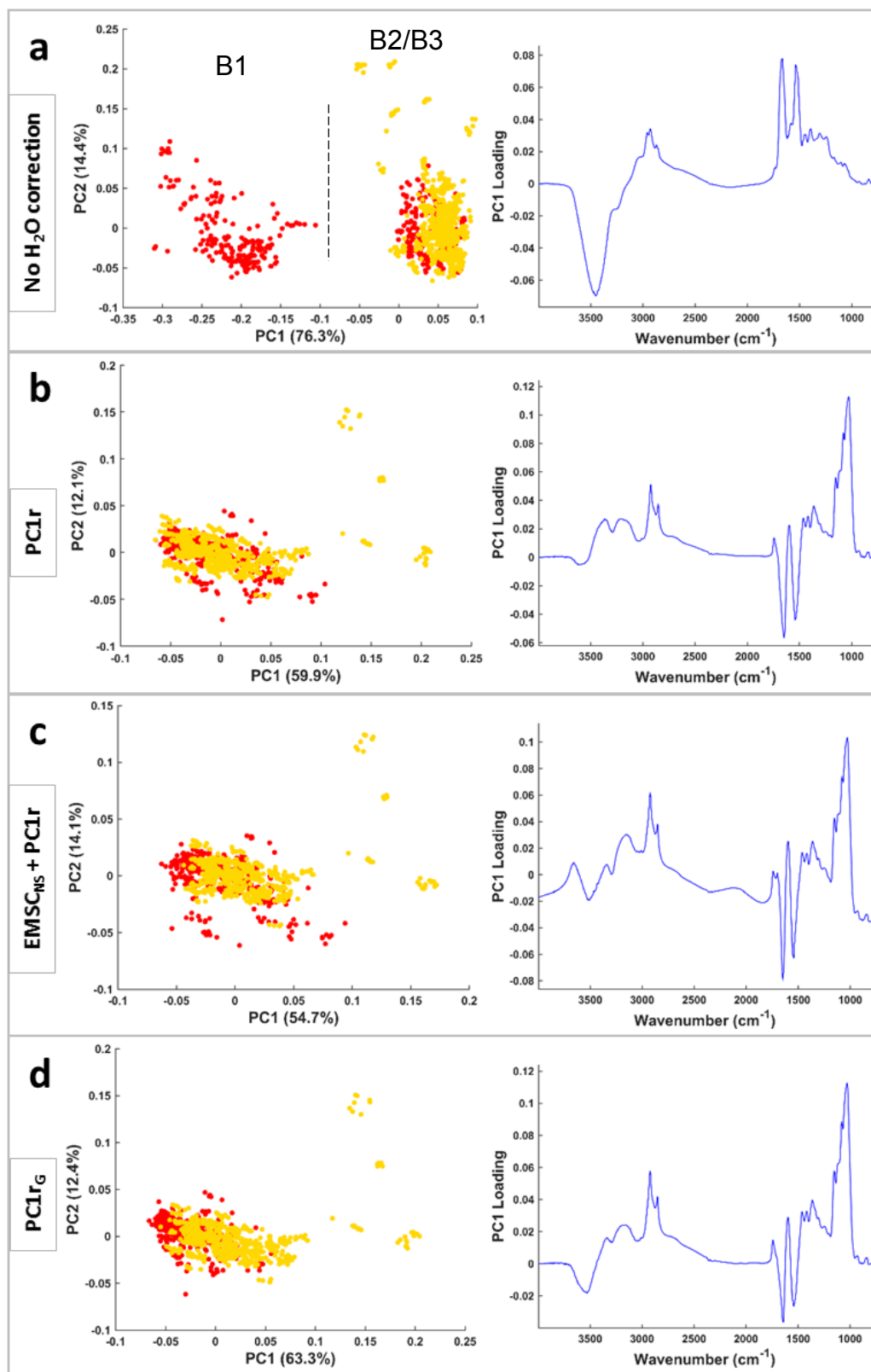
Figure 6.9 *Influence of the different water corrections on the discrimination of sepsis (S0 in red) and SIRS (orange) serum spectra assessed by PCA. PCA plots and PC1 loadings corresponding to all pre-processed spectra without water correction (**a**) or corrected for water content by PC1r (**b**), EMSC$_{NS}$ + PC1r (**c**) and PC1r$_G$ (**d**) methods. B1/2/3: batch 1/2/3.*

172
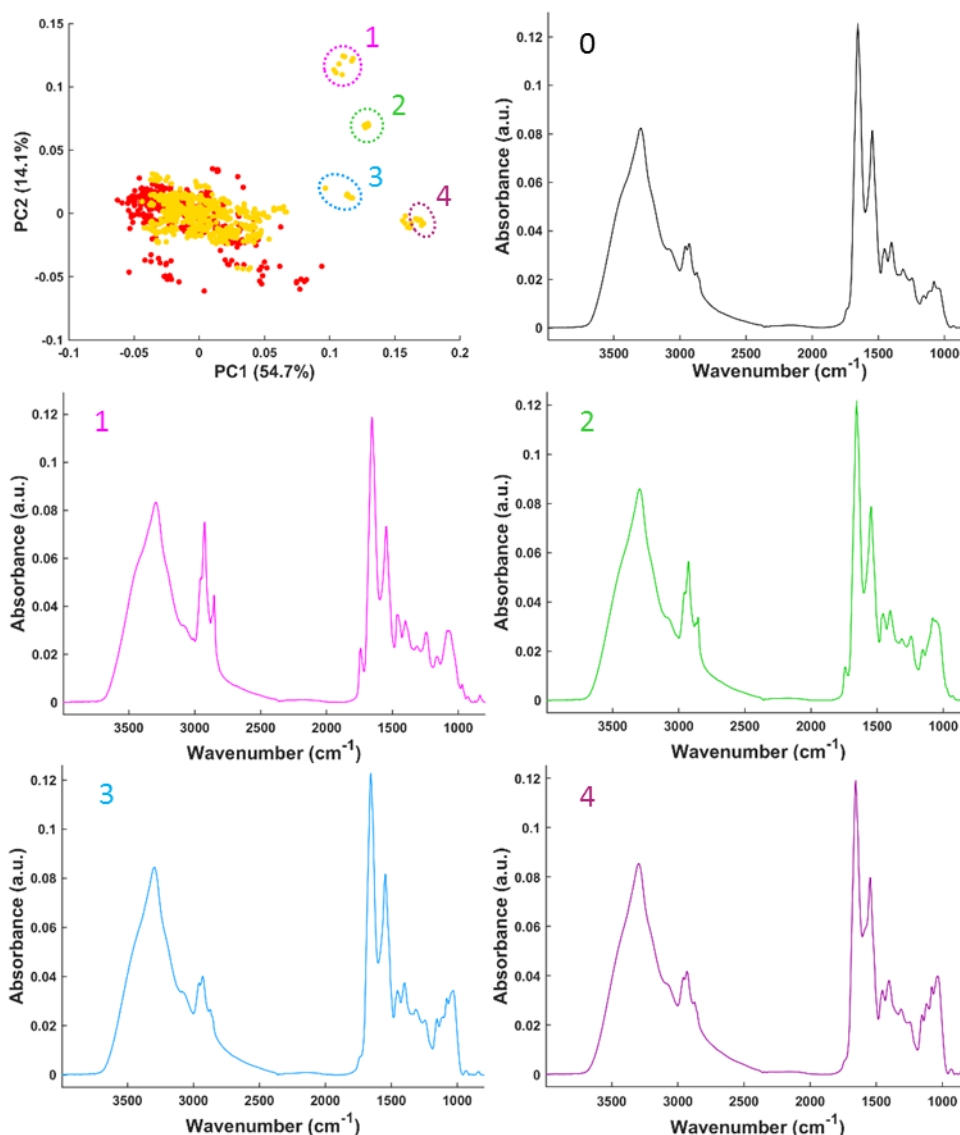
Figure 6.10 *Example of 4 SIRS patients with a highly different mean spectral profile (spectra 1, 2, 3, 4) compared to the mean spectrum of the SIRS group (spectrum 0) highlighted on a PCA plot from Figure 6.9.*

Then, the influence of the different corrections (except $EMSC_{NS}$ correction) on the discrimination of the sepsis group (S0, S-1, S-2, S-3) against the control group (C0, C-1, C-2, C-3) was evaluated by a RF analysis. Results are presented in Table 6.2. Over the 4000-800 cm^{-1} spectral range, similar sensitivities of 77%, 74% and specificities of 73%, 72% were obtained for the PC1r and $EMSC_{NS}$ + PC1r correction approaches respectively compared to the data not corrected for water content (79% sensitivity, 73% specificity) while a 92% sensitivity, 90% specificity was achieved with the PC1r$_G$ correction. Corresponding feature importance are presented in Figure

6.11. Compared to data not corrected for water content (Figure 6.11a), the importance of the water feature has been decreased for all 3 types of corrections especially for PC1r and PC1r$_G$ corrections (Figure 6.11b, c). Otherwise, it can be noted that similar features are found for the data corrected by PC1r and EMSC$_{NS}$ + PC1r compared to no correction, such as proteins (amide I and II bands), carbohydrates (~1000 cm^{-1}), nucleic acids (930-830 cm^{-1}) and lipids (2955-2850 cm^{-1} and 1737 cm^{-1}). However, for the corrected data the contribution of the amide I band features are increased while amide II and lipids (2955-2850 cm^{-1}) features are decreased. The overtone (~2000 cm^{-1}) and carbon dioxide (CO_2) at ~2360 cm^{-1} are highly increased with the PC1r correction (Figure 6.11b). Regarding the PC1r$_G$ correction (Figure 6.11d), only a high contribution of the overtone and the 930-800 cm^{-1} spectral region appear to be responsible for the data discrimination. A RF analysis on the "fingerprint region" (1800-800 cm^{-1}) was also performed with no discrimination improvement compared to an analysis performed on the full spectral range for all types of corrections (Table 6.2).

Table 6.2 *Sensitivities and specificities obtained for the discrimination of sepsis (S0, S-1, S-2, S-3) against control (C0, C-1, C-2, C-3) serum spectra without water correction or corrected for water content by PC1r, EMSC$_{NS}$ + PC1r and PC1r$_G$ methods.*

	Spectral zone	Sensitivity	Specificity	Accuracy
No correction	1	0.79	0.73	0.76 (± 0.03)
	2	0.74	0.74	0.74 (± 0.02)
PC1r	1	0.77	0.73	0.76 (± 0.01)
	2	0.74	0.73	0.73 (± 0.04)
EMSC$_{NS}$ + PC1r	1	0.74	0.72	0.73 (± 0.03)
	2	0.73	0.72	0.73 (± 0.04)
PCr$_G$	1	0.92	0.90	0.91 (± 0.03)
	2	0.91	0.91	0.91 (± 0.03)

Spectral zones: 4000-800 cm^{-1} (1) and 1800-900 cm^{-1} (2)
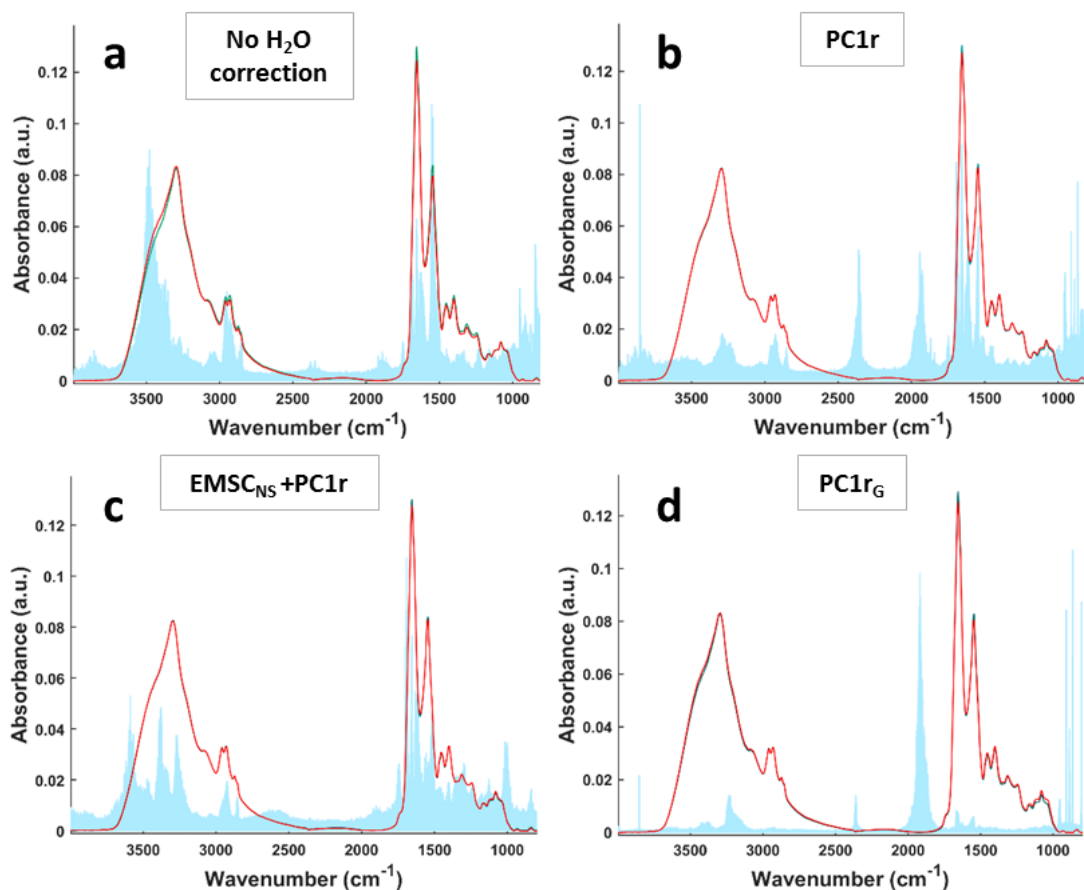
Figure 6.11 *Influence of the different water corrections on the discrimination of sepsis (S0, S-1, S-2, S-3) against control (C0, C-1, C-2, C-3) serum spectra assessed by RF. Feature importance (blue) responsible for the discrimination of sepsis and control pre-processed spectra without water correction (**a**) or corrected for water content by PC1r (**b**), EMSC$_{NS}$ + PC1r (**c**) and PC1r$_G$ (**d**) methods. The mean spectrum of sepsis (red) and control (green) group are presented for comparative purposes.*

A RF analysis was also tested to discriminate the sepsis group including only samples collected on the day of diagnosis (S0) and on the day before the diagnosis (S-1) *versus* matched controls (C0 and C-1). Indeed, as S-1 in term of time course is the closest to the diagnosis, a better discrimination could be expected. However, similar results were obtained on both 4000-800 cm^{-1} and 1800-800 cm^{-1} spectral regions (Table 6.3). Sensitivities of 77%, 71%, 96% and specificities of 81%, 72%, 97% for the PC1r, EMSC$_{NS}$ + PC1r and PC1r$_G$ corrections respectively compared to a sensitivity, specificity of 81% obtained for non-corrected data. Same features were also highlighted (Figure 6.12).

Table 6.3 *Sensitivities and specificities obtained for the discrimination of sepsis (S0, S-1) against control (C0, C-1) serum spectra without water correction or corrected for water content by PC1r, EMSC$_{NS}$ + PC1r and PC1r$_G$ methods.*

	Spectral zone	Sensitivity	Specificity	Accuracy
No correction	1	0.81	0.81	0.81 (± 0.03)
	2	0.75	0.78	0.77 (± 0.05)
PC1r	1	0.77	0.81	0.79 (± 0.09)
	2	0.74	0.81	0.77 (± 0.04)
EMSC$_{NS}$ + PC1r	1	0.71	0.72	0.71 (± 0.04)
	2	0.75	0.71	0.73 (± 0.07)
PCr$_G$	1	0.96	0.97	0.96 (± 0.02)
	2	0.90	0.89	0.89 (± 0.02)

Spectral zones: 4000-800 cm^{-1} (1) and 1800-900 cm^{-1} (2)
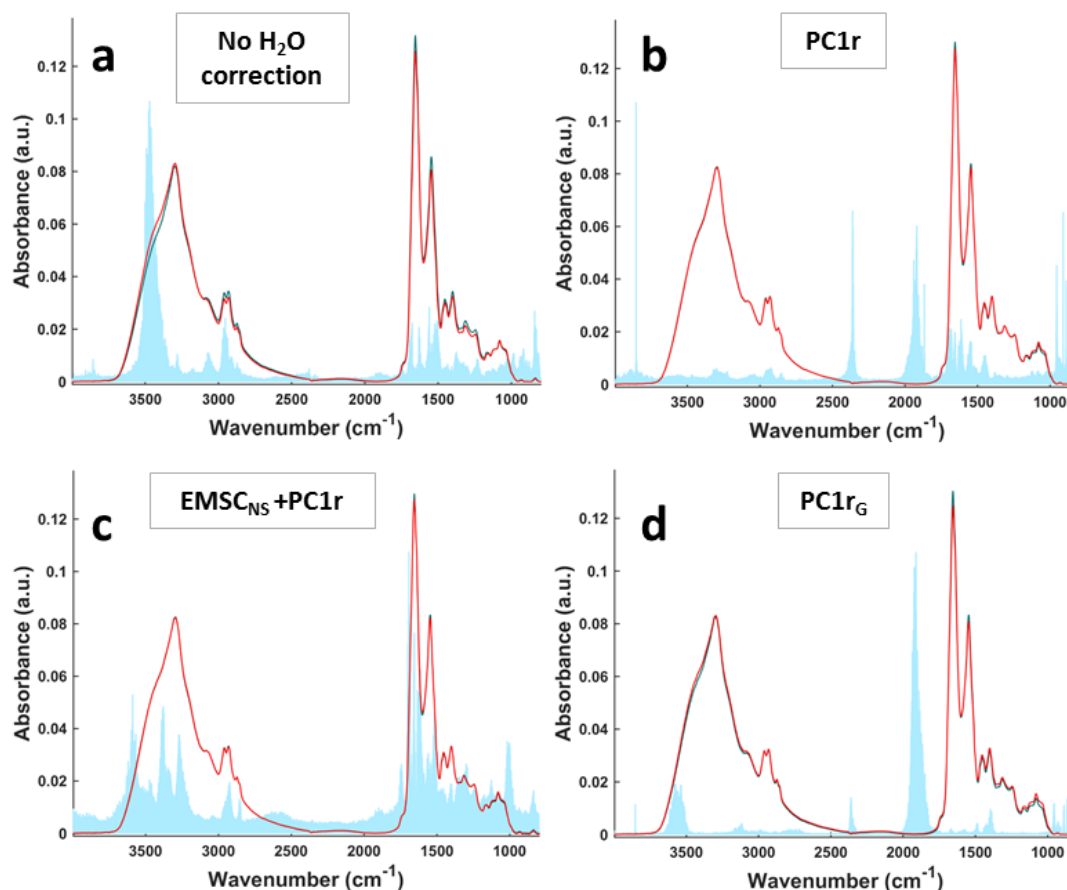
Figure 6.12 *Influence of the different water corrections on the discrimination of sepsis (S0, S-1) against control (C0, C-1) serum spectra assessed by RF. Feature importance (blue) responsible for the discrimination of sepsis and control pre-processed spectra without water correction (**a**) or corrected for water content by PC1r (**b**), EMSC$_{NS}$ + PC1r (**c**) and PC1r$_G$ (**d**) methods. The mean spectrum of sepsis (red) and control (green) group are presented for comparative purposes.*

VI.6 Conclusions and discussion

It has been demonstrated that the difference in water content in serum samples after drying, due to variations in relative humidity in the air while performing experiments, has a strong influence on the spectral data acquisition and analysis. This study highlights the importance of a robust experimental design and standardisation to avoid any bias that can lead to misinterpretations of spectral data and erroneous overall results.

Regarding the different types of water corrections tested, it can be assumed that the correction by EMSC based on a normal saline constituent spectrum was inefficient (Figure 6.7b) as it allowed to correct only free-bound water molecules in the samples whereas it is known that water interacts with proteins[29,30]. Thus, corrections based on a PC were able to overcome the difference of water content in samples as they reflected the presence of water molecules bound to proteins. However, the almost complete decrease of every feature in the RF analysis related to the correction based on PC1 removal group by group (Figure 6.11d) might be due to a loss of patient spectral information included in PCs and thus, the high sensitivity/specificity achieved are questionable and highlight the limitation of PC based corrections. Moreover, the water content of a serum sample can be directly linked to its composition and be useful for discrimination. The best way to confirm the previous hypotheses would be to perform again the experiments in a controlled environment in order to dry serum samples and collect spectra in the same conditions.

The low sensitivities and specificities obtained could also be explained by the fact that besides sepsis, patients recruited have various concomitant health conditions which can interfere with the discrimination such as different types of cancers, chronic obstructive pulmonary disease, inflammatory bowel disease, ischaemic heart disease, liver cirrhosis, chronic renal failure. An example in Figure 6.10 of 4 patients with highly different spectral profiles compared to patients of the same group, illustrates the influence of a patient general health conditions.

Furthermore, it can be mentioned that procalcitonin (PCT), widely considered as one of the most useful biomarkers for the diagnosis of sepsis, has a specificity and

sensitivity below 90%[4]. It has even been reported that PCT has a specificity and sensitivity close to 70% for the differentiation of sepsis from sterile SIRS[31].

Time to sepsis diagnosis post-surgery varies among patients as some of them developed the disease (or were diagnosed) more rapidly than others. As a result, a maximum of 4 samples (S-3 to S0) and a minimum of 2 samples (S-1, S0) of a same patient have been included in the data analyses, which can introduce a bias in the discrimination. Other classification methods in a non-binary way of application should be tested such as Partial Least-Square Discriminant Analysis (PLS-DA).

In this study, some spectral features such as the amide I/II bands, carbohydrates and nucleic acids have been highlighted several times, for the discrimination of sepsis and control patients with sensitivities and specificities superior to 70%. Furthermore, based on the same spectral features, data including spectra of samples collected up to 3 days before and on the day of diagnosis have shown similar sensitivities and specificities to distinguish sepsis from control patients (Table 6.2) compared to the dataset composed of only samples obtained on one day before and on the day of sepsis diagnosis (Table 6.3). This suggests that these spectral features could be potential biomarkers for the early diagnosis of sepsis and should be further investigated in future work.

VI.7 References

1   M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. Vincent and D. C. Angus, *JAMA*, 2016, **315**, 801–810.

2   C. Fleischmann, A. Scherag, N. K. J. Adhikari, C. S. Hartog, T. Tsaganos, P. Schlattmann, D. C. Angus and K. Reinhart, *Am. J. Respir. Crit. Care Med.*, 2016, **193**, 259–272.

3   O. Opota, A. Croxatto, G. Prod'hom and G. Greub, *Clin. Microbiol. Infect.*, 2015, **21**, 313–322.

4   C. Pierrakos and J.-L. Vincent, *Crit. Care*, 2010, **14**, 1–18.

5   K. N. Iskander, M. F. Osuchowski, D. J. Stearns-Kurosawa, S. Kurosawa, D. Stepien, C. Valentine and D. G. Remick, *Physiol. Rev.*, 2013, **93**, 1247–1288.

6   M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nat. Protoc.*, 2014, **9**, 1771–1791.

7   Angus, *N. Engl. J. Med.*, 2013, **369**, 840–851.

8   D. Rittirsch, M. A. Flierl and P. A. Ward, *Nat. Rev. Immunol.*, 2008, **8**, 776–787.

9   C. M. Robertson and C. M. Coopersmith, *Microbes Infect.*, 2006, **8**, 1382–1389.

10   A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar and M. Cheang, *Crit. Care Med.*, 2006, **34**, 1589–1596.

11   C. W. Seymour, F. Gesten, H. C. Prescott, M. E. Friedrich, T. J. Iwashyna, G. S. Phillips, S. Lemeshow, T. Osborn, K. M. Terry and M. M. Levy, *N. Engl. J. Med.*, 2017, **376**, 2235–2244.

12   R. Ferrer, I. Martin-Loeches, G. Phillips, T. M. Osborn, S. Townsend, P. R. Dellinger, A. Artigas, C. Schorr and M. M. Levy, *Crit. Care Med.*, 2014, **42**, 1749–1755.

13   A. Rhodes, L. E. Evans, W. Alhazzani, M. M. Levy, M. Antonelli, R. Ferrer, A. Kumar, J. E. Sevransky, C. L. Sprung, M. E. Nunnally, B. Rochwerg, G. D. Rubenfeld, D. C. Angus, D. Annane, R. J. Beale, G. J. Bellinghan, G. R. Bernard, J.-D. Chiche, C. Coopersmith, D. P. De Backer, C. J. French, S. Fujishima, H. Gerlach, J. L. Hidalgo, S. M. Hollenberg, A. E. Jones, D. R. Karnad, R. M. Kleinpell, Y. Koh, T. C. Lisboa, F. R. Machado, J. J. Marini, J. C. Marshall, J. E. Mazuski, L. A. McIntyre, A. S. McLean, S. Mehta, R. P. Moreno, J. Myburgh, P. Navalesi, O. Nishida, T. M. Osborn, A. Perner, C. M. Plunkett, M. Ranieri, C. A. Schorr, M. A. Seckel, C. W. Seymour, L. Shieh, K. A. Shukri, S. Q. Simpson, M. Singer, B. T. Thompson, S. R. Townsend, T. Van der Poll, J.-L. Vincent, J. W. Wiersinga, J. L. Zimmerman and P. R. Dellinger, *Crit. Care Med.*, 2017, **45**, 486–552.

14   T. J. Iwashyna, C. R. Cooke, H. Wunsch and J. M. Kahn, *J. Am. Geriatr. Soc.*, 2012, **60**, 1070–1077.

15   P. P. Pandharipande, T. D. Girard, J. C. Jackson, A. Morandi, J. L. Thompson, B. T. Pun, N. E. Brummel, C. G. Hughes, E. E. Vasilevskis, A. K. Shintani, K. G. Moons, S. K. Geevarghese, A. Canonico, R. O. Hopkins, G. R. Bernard, R. S. Dittus and E. W. Ely, *N. Engl. J. Med.*, 2013, **369**, 1306–1316.

16   N. Hex, J. Retzler, C. Bartlett and M. Arber, *YHEC*, 2017, 1–46.

17   M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszewski, G. Thiéfin and G. D. Sockalingum, *Chem. Soc. Rev.*, 2016, **45**, 1803–1818.

18   U. Neugebauer, S. Trenkmann, T. Bocklitz, D. Schmerler, M. Kiehntopf and J. Popp, *J. Biophotonics*, 2014, **7**, 232–240.

19   L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D.

Sockalingum and M. J. Baker, *Anal. Methods*, 2015, **7**, 7140–7149.

20 L. Lovergne, P. Bouzy, V. Untereiner, R. Garnotel, M. J. Baker, G. Thiéfin and G. D. Sockalingum, *Faraday Discuss.*, 2016, **187**, 521–537.

21 M. M. Levy, M. P. Fink, J. C. Marshall, E. Abraham, D. Angus, D. Cook, J. Cohen, S. M. Opal, J.-L. Vincent and G. Ramsay, *Intensive Care Med.*, 2003, **29**, 530–538.

22 X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon and G. D. Sockalingum, *Transl. Res.*, 2013, **162**, 279–286.

23 J. H. Ward, *J. Am. Stat. Assoc.*, 1963, **58**, 236–244.

24 H. Abdi and L. J. Williams, *Wiley Interdiscip. Rev. Comput. Stat.*, 2010, **2**, 433–459.

25 H. Martens and E. Stark, *J. Pharm. Biomed. Anal.*, 1991, **9**, 625–635.

26 N. K. Afseth and A. Kohler, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 92–99.

27 D. Naumann, *Appl. Spectrosc. Rev.*, 2001, **36**, 239–298.

28 C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, *Analyst*, 2015, **140**, 2280–2286.

29 A. Barth and P. I. Haris, *Biological and Biomedical Infrared*, IOS Press, 2009.

30 A. Barth, *Biochim. Biophys. Acta*, 2007, **1767**, 1073–1101.

31 G. Bernard, J.-L. Vincent, P.-F. Laterre, S. P. Larosa, J.-F. Dhainaut, A. Lopez-Rodriguez, J. S. Steingrub, G. E. Garber, J. D. Helterbrand, W. E. Ely and C. J. Fisher, *N. Eng. J. Med.*, 2001, **344**, 699–709.

Chapter VII. General conclusions and discussion

VII.1 Conclusions générales et discussion

VII.1.1 Développement d'une méthodologie pour l'application de la spectroscopie au sérum

L'un des principaux objectifs de cette recherche était de développer et d'évaluer le potentiel de la spectroscopie vibrationnelle appliquée au sérum humain pour le diagnostic rapide et pré-symptomatique du sepsis. La spectroscopie infrarouge à transformée de Fourier et à haut débit a démontré son potentiel pour l'analyse d'échantillons de sérum avec une perspective d'application clinique, comparée à d'autres approches incluant la microimagerie, le mode de réflectance totale atténuée, l'utilisation d'une source Synchrotron ou la spectroscopie Raman.

La spectroscopie infrarouge est sensible à l'eau, par conséquent les échantillons de sérum sont séchés pour les analyses. Cependant, une goutte séchée de sérum expose un profil inhomogène qui peut affecter les données spectrales collectées et introduire de potentiels biais dans leurs interprétations. Ainsi, la spectroscopie infrarouge à haut débit présente l'avantage d'être une macroanalyse qui permet d'assurer la reproductibilité des acquisitions spectrales. Autre aspect positif, cette technique étant automatisée, elle permet une acquisition rapide de spectres d'un grand nombre d'échantillons en une seule analyse.

Cependant, deux points négatifs ont été mis en évidence lors de nos investigations. Les échantillons de sérum doivent être dilués pour une analyse par spectroscopie infrarouge en mode transmission pour éviter la saturation des bandes amide I et II. Cette étape supplémentaire peut être une source potentielle de contamination alors que le mode de réflectance totale atténuée permet d'analyser directement les échantillons. Ainsi, la conception d'un appareillage qui permettrait d'utiliser ce mode de manière automatisée et à haut débit serait d'une grande utilité pour une application aux biofluides. Le second désavantage observé est le fait que les échantillons de sérum contiennent différentes quantités d'eau après séchage, à cause des variations de l'humidité relative de l'air, interférant avec la collecte et l'analyse des données spectrales. Par conséquent, il est recommandé de sécher et collecter les spectres dans un environnement contrôlé dans le but de réaliser les expériences dans les mêmes

conditions. Concernant la correction des données pour la variation de contenu en eau, des recherches plus poussées sont nécessaires pour comprendre la contribution de l'eau dans les spectres, en particulier les molécules d'eau liées aux protéines.

Une autre possibilité serait idéalement de pouvoir analyser les échantillons de sérums dans leur forme liquide par exemple par spectroscopie infrarouge en réflectance totale atténuée ou Raman comme démontré dans une récente étude réalisée sur des échantillons liquides de plasmas. Une sensibilité de 96.5% et spécificité de 95% ont été obtenues dans la discrimination de patients atteints d'un cancer de la prostate et de volontaires sains[2].

VII.1.2 Diagnostic du sepsis par spectroscopie infrarouge

Dans ce projet, des profils spectraux associés aux protéines, glucides et acides nucléiques ont permis de distinguer des patients septicémiques de patients contrôles, incluant des échantillons collectés jusqu'à 3 jours avant le diagnostic d'un sepsis, avec des sensibilités et spécificités supérieures à 70%. A cause d'interférences liées à des variations de contenu en eau des sérums après séchage, il serait nécessaire d'analyser à nouveau les échantillons dans un environnement contrôlé dans le but de confirmer l'utilité de ces profils spectraux mis en évidence pour le diagnostic pré-symptomatique de patients atteints d'un sepsis.

Un travail futur inclura la comparaison des profils spectraux avec les informations patients enregistrées telles que le niveau de protéine C réactive ou de lactate qui sont couramment utilisés dans le diagnostic du sepsis malgré leur manque de sensibilité/spécificité[3–5]. D'autres recherches d'un intérêt majeur pour le traitement du patient[6,7] seront menées pour déterminer s'il est possible d'identifier des patients avec un sepsis causé par des bactéries Gram positives, Gram négatives ou d'autres pathogènes. Il sera également évalué la possibilité de distinguer ou non les cas de sepsis, des cas de syndrome de réponse inflammatoire systémique (SRIS).

VII.2 Serum spectroscopy methodology development

The main objective of this research was to develop and evaluate the potential of vibrational spectroscopy applied on human serum for the rapid pre-symptomatic diagnosis of sepsis. High-Throughput Fourier Transform Infrared (HT-FTIR) spectroscopy has demonstrated its potential for the analysis of serum samples with a perspective of clinical application when compared with different IR techniques including microimaging, Attenuated Total Reflectance (ATR)-FITR and Synchrotron-FTIR spectroscopy as well as Raman spectroscopy.

IR spectroscopy is sensitive to water, as a consequence serum samples were dried for analysis. However, a dried serum drop exhibits an inhomogeneous pattern that could affect the spectral data collected in a single point spectrum mode and introduce potential bias in their interpretation, as seen in chapter II when using transmission FTIR and Raman microspectroscopies. Moreover, spectra acquired by transmission FTIR microspectroscopy in the periphery of dried serum drops are saturated in the amide I and II bands, even when samples are diluted. Raman spectroscopy is also prone to fluorescence interferences, especially in the case of serum samples with a higher concentration in bilirubin (yellow pigment), requiring an additional step to correct spectra. As such, HT-FTIR spectroscopy in transmission mode presents the advantage to be a macro-analysis approach which ensures the reproducibility of spectral acquisitions, one spectrum representing the whole sample. Another positive aspect is the fact that it is an automated technique which allows to quickly collect spectra of a large number of samples in one analysis. Spectra of a hundred samples can be easily collected in less than a day whereas weeks of acquisitions would be necessary to analyse the same number of samples with the other techniques tested.

However, two negative aspects have been highlighted during our investigations. Serum samples have to be diluted to perform the IR analysis in the transmission mode to avoid saturation of the amide I/II bands. This additional sample preparation step can be a source of potential contaminations whereas in ATR-FTIR spectroscopy serum samples can be directly analysed onto a crystal without dilution. This technique allows

to analyse only a zone of a hundred microns of a serum dried drop but its single bounce way of measurement enable to acquire reproducible spectra Therefore, an ATR-FTIR device designed for HT analysis is of great promise for biofluid application.

The second drawback observed was the fact that serum samples contained different amounts of water after drying due to variation of relative humidity in air, interfering with the spectra collection and data analysis. As a consequence, it is recommended to dry serum samples and collect spectra in a controlled atmosphere in order to perform experiments in the same conditions. Regarding the data corrections for the variation in sample water content, further work is needed to better understand the contribution of water in the serum spectra, especially water molecules bound to proteins.

Another possibility would be ideally to analyse serum samples in their liquid form for example by ATR-FTIR spectroscopy[1] or Raman spectroscopy as demonstrated in a recent study performed on liquid plasma samples. Sensitivity of 96.5% and specificity of 95% were achieved for the discrimination of patients with prostate cancer from healthy volunteers[2].

VII.3 Diagnosis of sepsis by serum infrared spectroscopy

In this project, spectral features associated with proteins, carbohydrates and nucleic acids were responsible for the discrimination of sepsis and control patients including patient serum samples collected up to 3 days before sepsis diagnosis with sensitivities and specificities superior to 70%. Due to water interferences, it would be necessary to recollect spectra from serum samples dried in a controlled environment in order to confirm the utility of the features highlighted for the pre-symptomatic diagnosis of patients with sepsis.

Future work will include the comparison of spectral profiles with patient data recorded such as for example the level of C-reactive protein (CRP) and lactate which are commonly used for sepsis diagnosis despite their lack of sensitivity/specificity[3-5].

Further investigations will be carried out to see whether it is possible to distinguish between patients with sepsis caused by Gram-positive, Gram-negative bacteria or other pathogens, and patients with sterile Systemic Inflammatory Response Syndrome (SIRS) which is of major interest for patient treatment[6,7].

It could be also interesting to investigate the potential of ultra-filtration of serum samples prior to perform spectral analysis to improve patient discrimination. The predominant presence of high molecular weight proteins such as Human Serum Albumin (HSA) can mask the presence of other constituents of diagnostic interest[8,9].

VII.4 Perspectives

This project has highlighted the potential of HT-FITR spectroscopy applied to serum samples as a future clinical tool for the early diagnosis of sepsis. This technique is simple to operate, does not require heavy sample preparation and thus associated with multivariate analysis can provide a diagnosis within minutes, reducing time to treatment initiation. Sepsis and sterile SIRS patients are currently not easily differentiated[10], as such IR spectroscopy could be also a useful technique for their rapid triage in intensive care units.

An extrapolation of the IR technique application would be for the patient follow-up as it would provide an easy way to assess the patient response to treatment.

Sepsis is a leading cause of death in critically ill children[11,12]. It has been acknowledged that the recent update of the sepsis definition (2016) does not refer to children[13]. Time consuming blood cultures remain, as well as for adults, the principal method for the causative agent identification. Moreover, blood culture sensitivity is highly dependent on the volume of samples collected. Despite specific blood bottles designed for pediatric patients, up to 3 mL of blood are required for one culture[14]. As a result, the lack of sufficient blood volume collected is a common cause of negative blood culture results[15]. In this research project, it has been demonstrated that only a few microliters of serum are necessary for spectral acquisitions. Therefore, HT-FTIR

spectroscopy appears as a technique of choice for the diagnosis of children and neonates with sepsis.

VII.5 References

1   F. Bonnier, F. Petitjean, M. J. Baker and H. J. Byrne, *J. Biophotonics*, 2014, **7**, 167–179.

2   D. K. R. Medipally, A. Maguire, J. Bryant, J. Armstrong, M. Dunne, M. Finn, F. M. Lyng and A. D. Meade, *Analyst*, 2016, **142**, 1216–1226.

3   C. W. Seymour, V. X. Liu, T. J. Iwashyna, F. M. Brunkhorst, T. D. Rea, A. Scherag, G. D. Rubenfeld, J. M. Kahn, M. Shankar-Hari, M. Singer, C. S. Deutschman, G. J. Escobar and D. C. Angus, *JAMA*, 2016, **315**, 762–774.

4   M. Shankar-Hari, G. S. Phillips, M. L. Levy, W. Christopher, V. X. Liu, C. S. Deutschman and D. C. Angus, *JAMA*, 2016, **315**, 775–787.

5   C. Pierrakos and J.-L. Vincent, *Crit. Care*, 2010, **14**, 1–18.

6   A. Rhodes, L. E. Evans, W. Alhazzani, M. M. Levy, M. Antonelli, R. Ferrer, A. Kumar, J. E. Sevransky, C. L. Sprung, M. E. Nunnally, B. Rochwerg, G. D. Rubenfeld, D. C. Angus, D. Annane, R. J. Beale, G. J. Bellinghan, G. R. Bernard, J.-D. Chiche, C. Coopersmith, D. P. De Backer, C. J. French, S. Fujishima, H. Gerlach, J. L. Hidalgo, S. M. Hollenberg, A. E. Jones, D. R. Karnad, R. M. Kleinpell, Y. Koh, T. C. Lisboa, F. R. Machado, J. J. Marini, J. C. Marshall, J. E. Mazuski, L. A. McIntyre, A. S. McLean, S. Mehta, R. P. Moreno, J. Myburgh, P. Navalesi, O. Nishida, T. M. Osborn, A. Perner, C. M. Plunkett, M. Ranieri, C. A. Schorr, M. A. Seckel, C. W. Seymour, L. Shieh, K. A. Shukri, S. Q. Simpson, M. Singer, B. T. Thompson, S. R. Townsend, T. Van der Poll, J.-L. Vincent, J. W. Wiersinga, J. L. Zimmerman and P. R. Dellinger, *Crit. Care Med.*, 2017, **45**, 486–552.

7   A. Kumar, P. Ellis, Y. Arabi, D. Roberts, B. Light, J. E. Parrillo, P. Dodek, G. Wood, A. Kumar, D. Simon, C. Peters, M. Ahsan and D. Chateau, *Chest*, 2009, **136**, 1237–1248.

8   R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads and T. D. Veenstra, *Mol. Cell. Proteomics*, 2003, **2**, 1096–1103.

9   J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D. Jenkinson, R. W. Lea, C. Walker and M. J. Baker, *J. Biophotonics*, 2014, **7**, 189–199.

10   C. M. Robertson and C. M. Coopersmith, *Microbes Infect.*, 2006, **8**, 1382–1389.

11   S. L. Weiss, J. C. Fitzgerald, J. Pappachan, D. Wheeler, J. C. Jaramillo-Bustamante, A. Salloo, S. C. Singhi, S. Erickson, J. A. Roy, J. L. Bush, V. M. Nadkarni and N. J. Thomas, *Am. J. Respir. Crit. Care Med.*, 2015, **191**, 1147–1157.

12   A. G. Randolph and R. J. McCulloh, *Virulence*, 2014, **5**, 179–189.

13   M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. Vincent and D. C. Angus, *JAMA*, 2016, **315**, 801–810.

14   O. Opota, A. Croxatto, G. Prod'hom and G. Greub, *Clin. Microbiol. Infect.*, 2015, **21**, 313–322.

15   N. Mancini, S. Carletti, N. Ghidoli, P. Cichero, R. Burioni and M. Clementi, *Clin. Microbiol. Rev.*, 2010, **23**, 235–251.

Appendix I. Supplementary figures of chapter IV
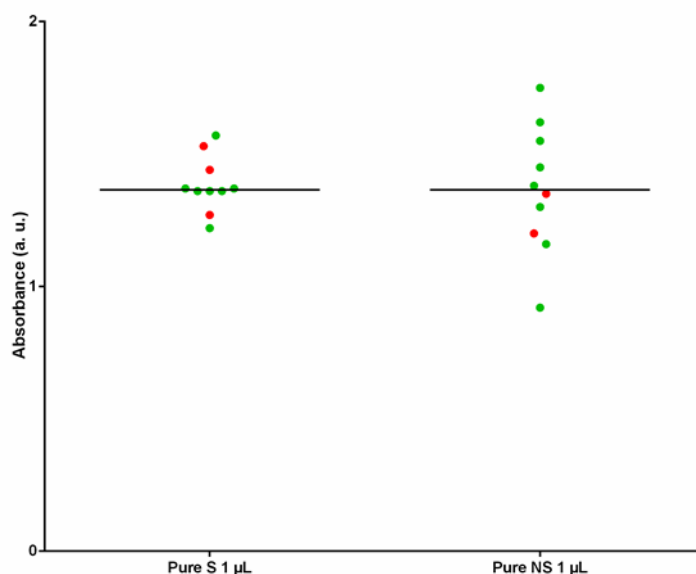
AI.1 *Scatter plot based on the amide I absorbance value of serum replicate spectra from spread and non-spread pure dried drops (1 µL deposits onto 3 mm diameter well).Validated and discarded spectra after the quality test are represented by green and red dots respectively. The median absorbance value is represented by the black bar. S: spread dried drop, NS: non-spread dried drop.*
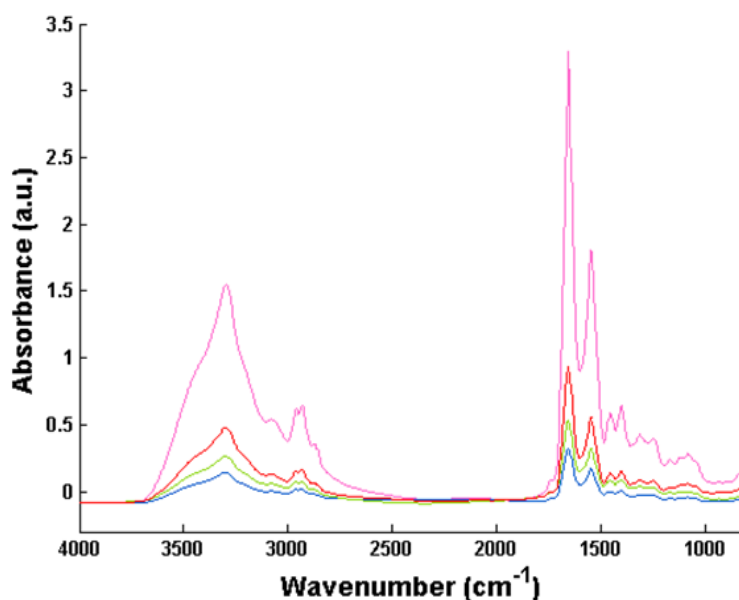
AI.2 *FTIR raw median spectra of serum samples (10 µL deposits onto 7 mm diameter well) at different dilutions 2-fold (red), 3-fold (green), 4-fold (blue) dilutions and without dilution (pink).*
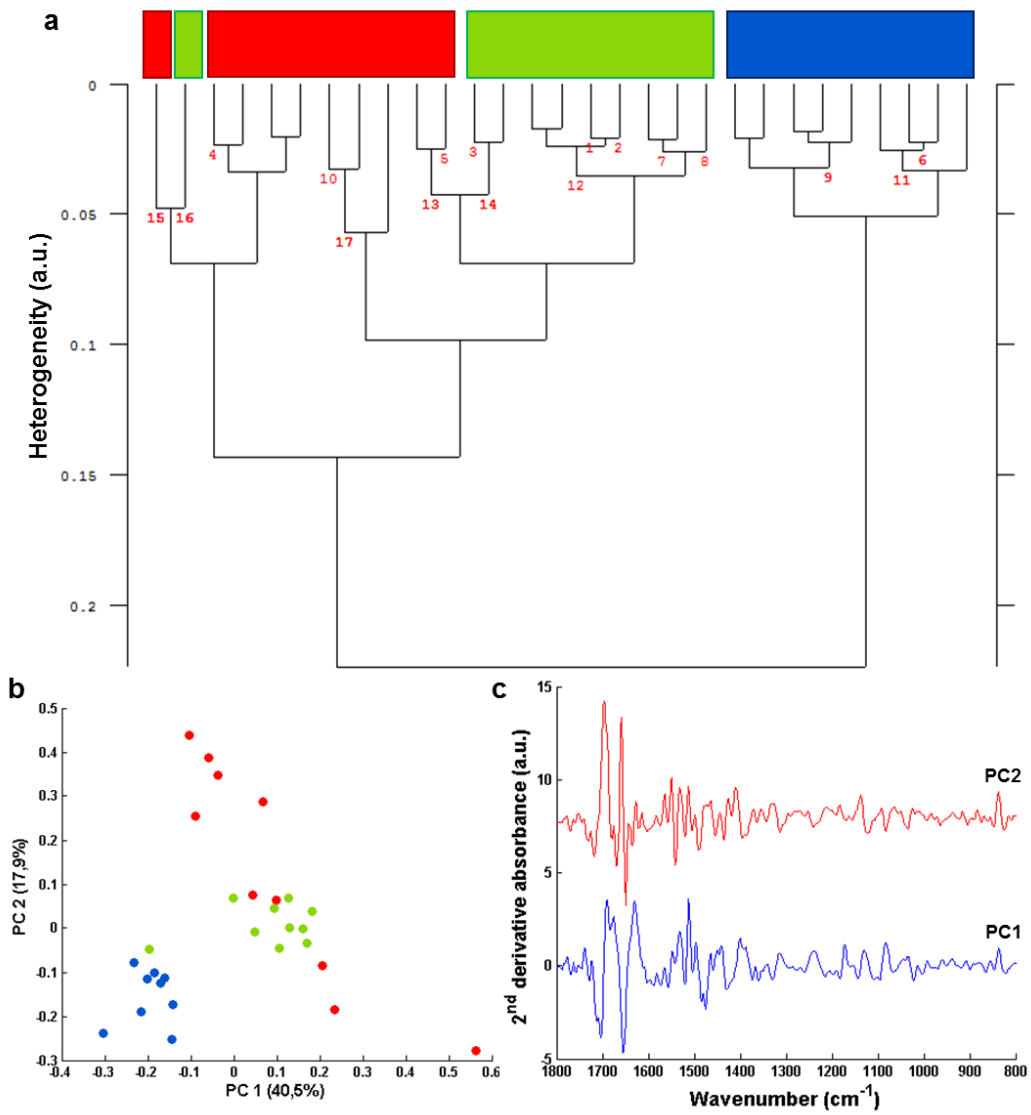
AI.3 (a) *Hierarchical cluster analysis of FTIR pre-processed spectra from 2-fold (red), 3-fold (green) and 4-fold (blue) diluted serum (5 µL) deposited onto 7 mm diameter wells.* **(b)** *PCA scatter plot of the spectral dataset from (a),* **(c)** *PC1 and PC2 loadings from (b).*
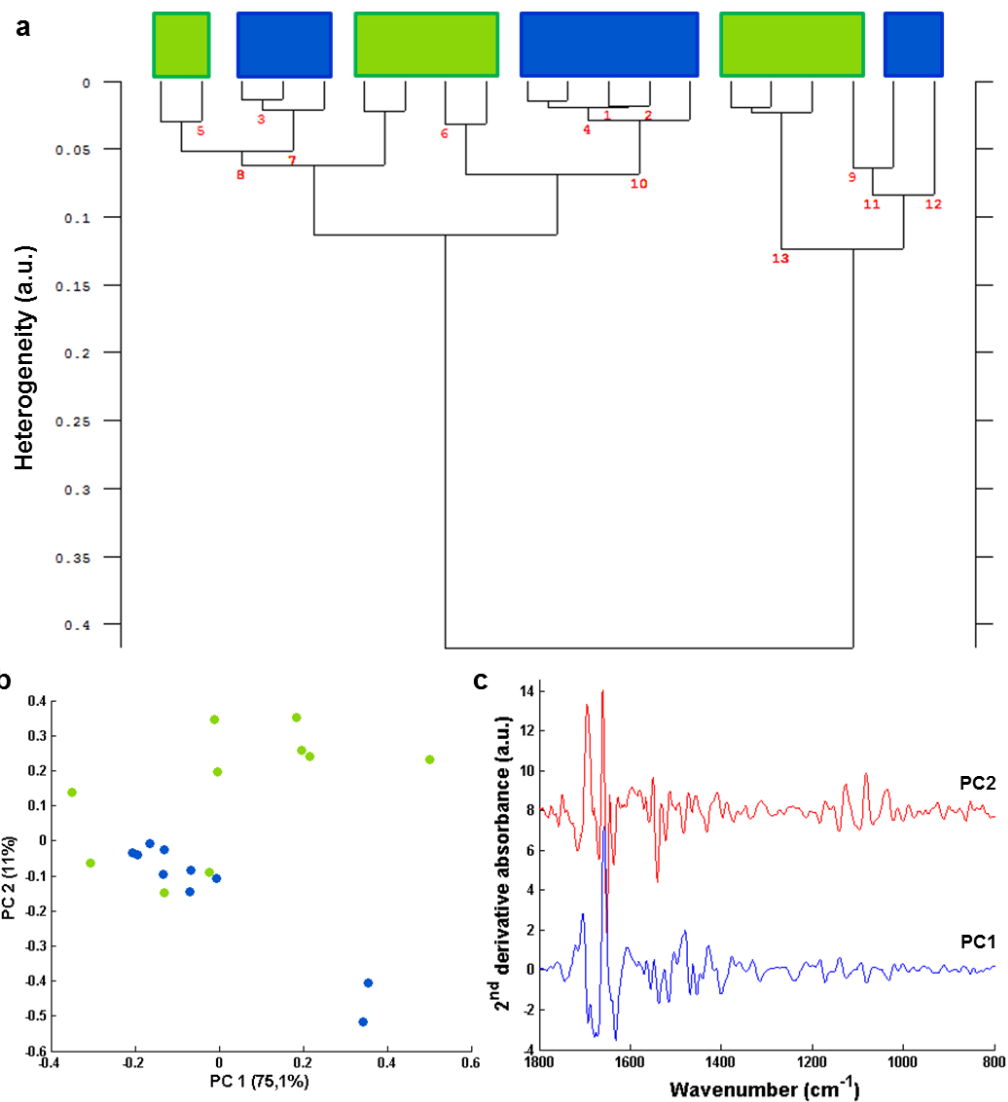
AI.4 (a) *Hierarchical cluster analysis of FTIR pre-processed spectra from 3-fold (green) and 4-fold (blue) diluted serum (5 µL) deposited onto 3 mm diameter wells.* ***(b)*** *PCA scatter plot of the spectral dataset from (a),* ***(c)*** *PC1 and PC2 loadings from (b).*
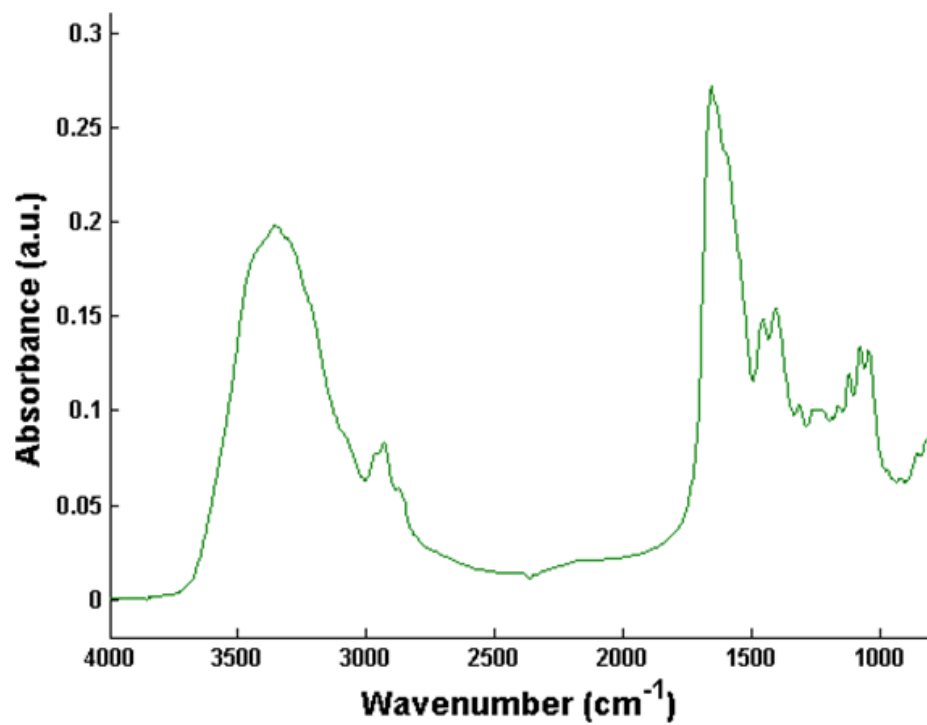
AI.5 *FTIR raw median spectrum of bile samples (5 µL deposits onto 3 mm diameter well) without dilution.*
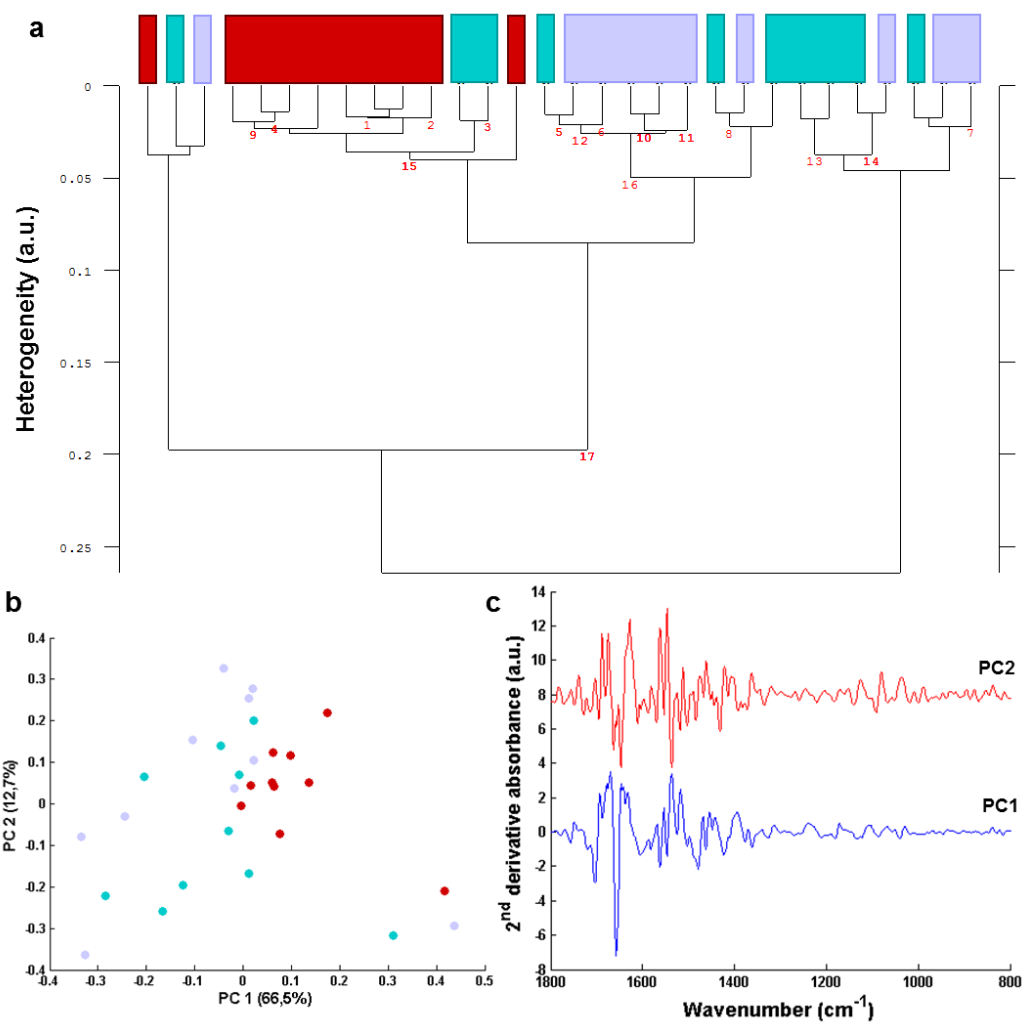
AI.6 (a) *Hierarchical cluster analysis of FTIR pre-processed spectra from one 3-fold diluted serum sample spread (5 μL) over the surface of 3 mm diameter wells by three different operators (red, light blue and purple).* ***(b)*** *PCA scatter plot of the spectral dataset from (a),* ***(c)*** *PC1 and PC2 loadings from (b).*

Appendix II. Supplementary figures of chapter V
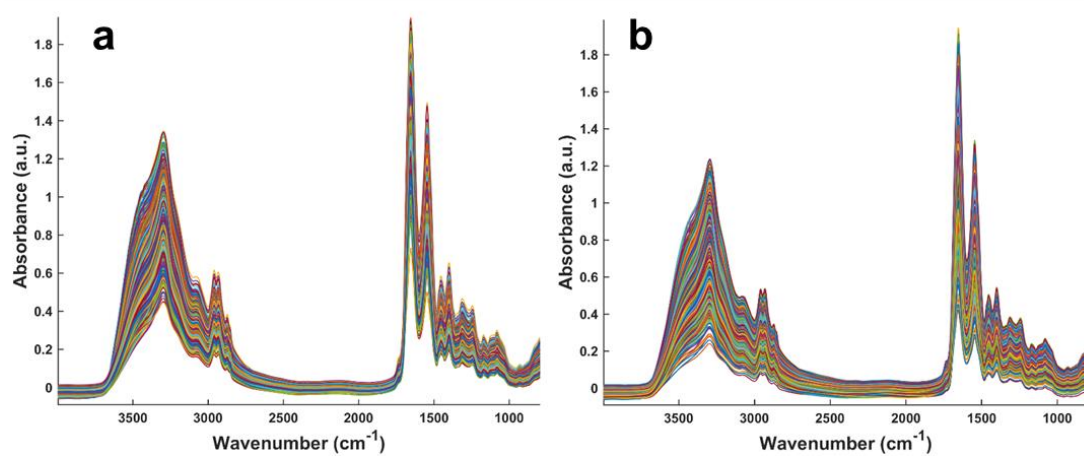
AII.1 *All positive quality tested serum (**a**) and plasma (**b**) raw spectra.*
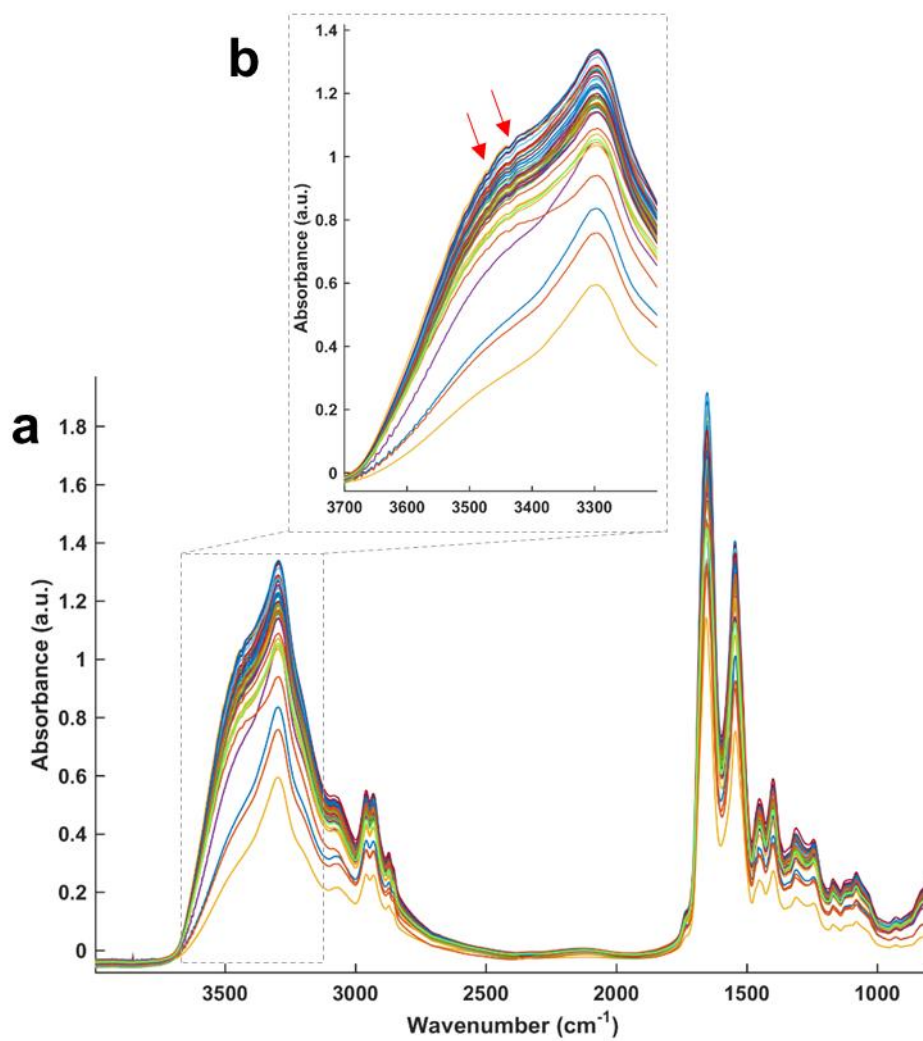
*AII.2 (**a**) Spectra removed from the study. (**b**) Zoom in the spectral region 3700-3200 cm⁻¹, red arrows indicate an abnormal decrease of the absorbance.*
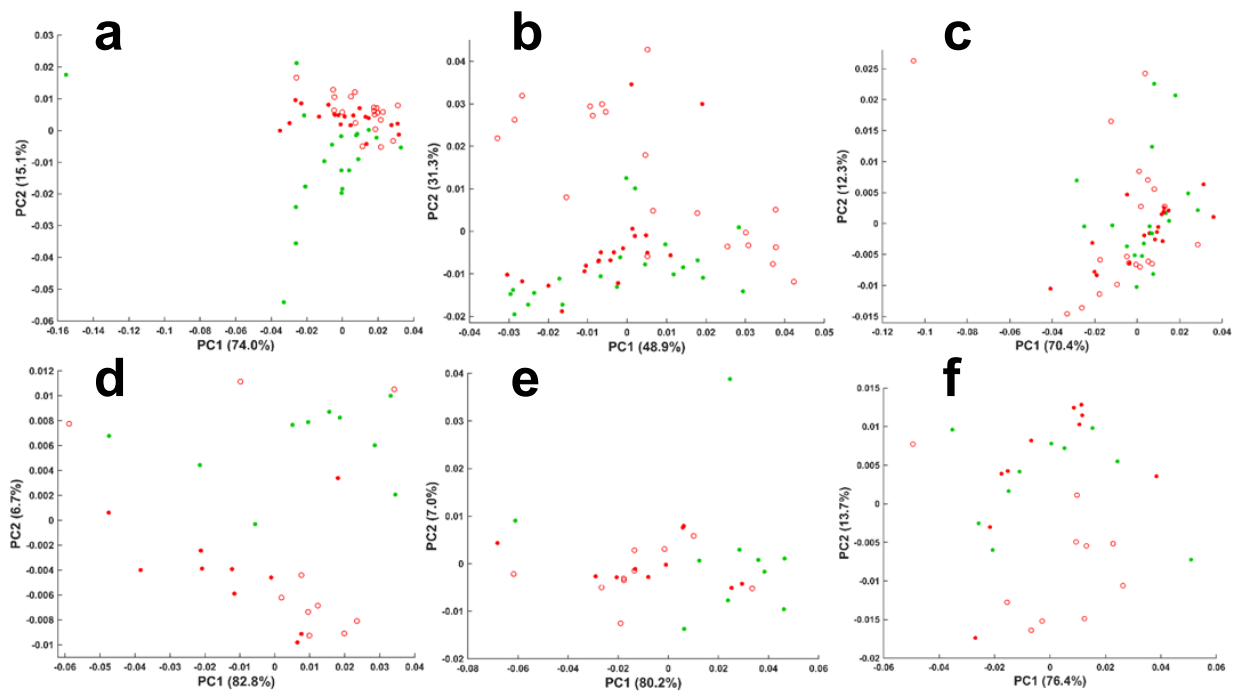
AII.3 *Example of the volume and tube impact assessment by PCA of plasma 1 (**a, d**), 2 (**b, e**), 3 (**c, f**), on the first (a-c) and last experiment (d-f) performed. Spectra of 50 µL of plasma stored in 0.2 mL capacity tubes are represented by green dots. Spectra of 50 µL or 100 µL of plasma stored in 1.5 mL capacity tubes are represented by red dots and circles respectively.*
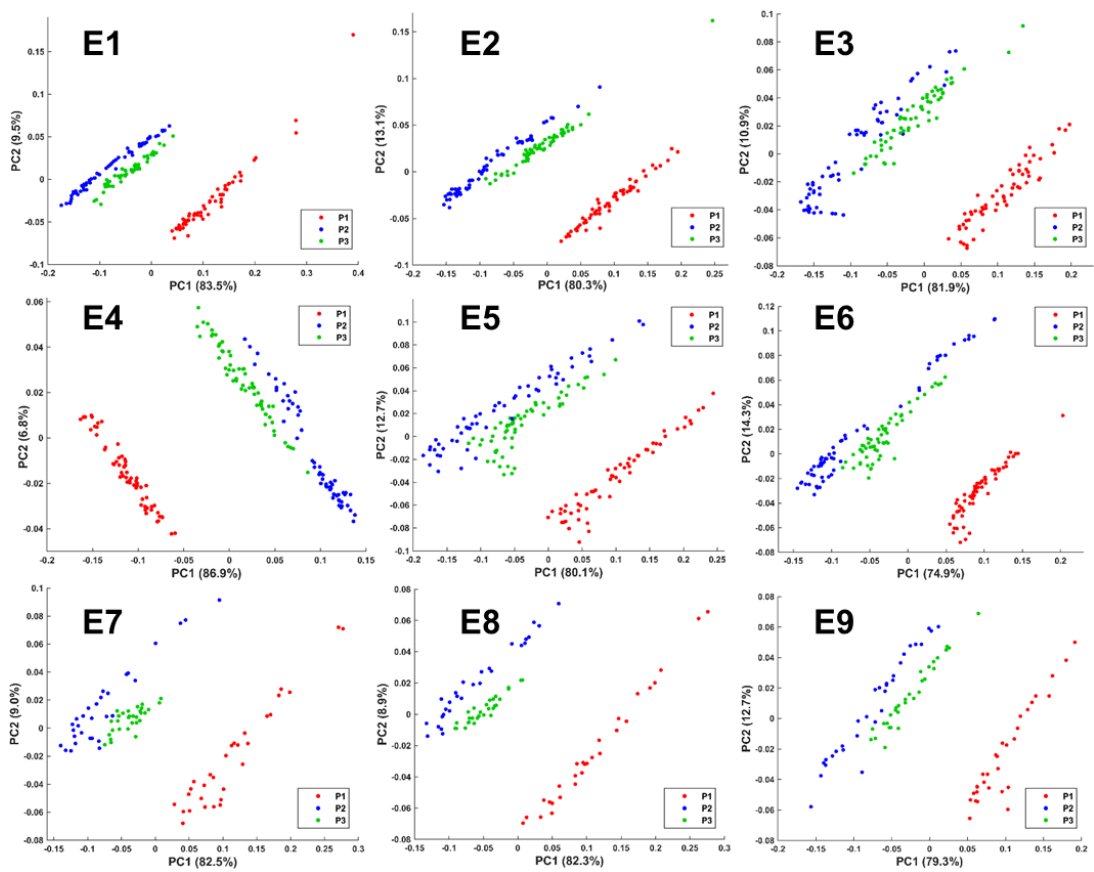
AII.4 *PCA performed on 3 subject plasma spectra (P1, P2, P3) in the spectral range 4000-800 cm^{-1} over a 9 month experiments (**E1-E9**) to assess the impact of a long-term storage.*
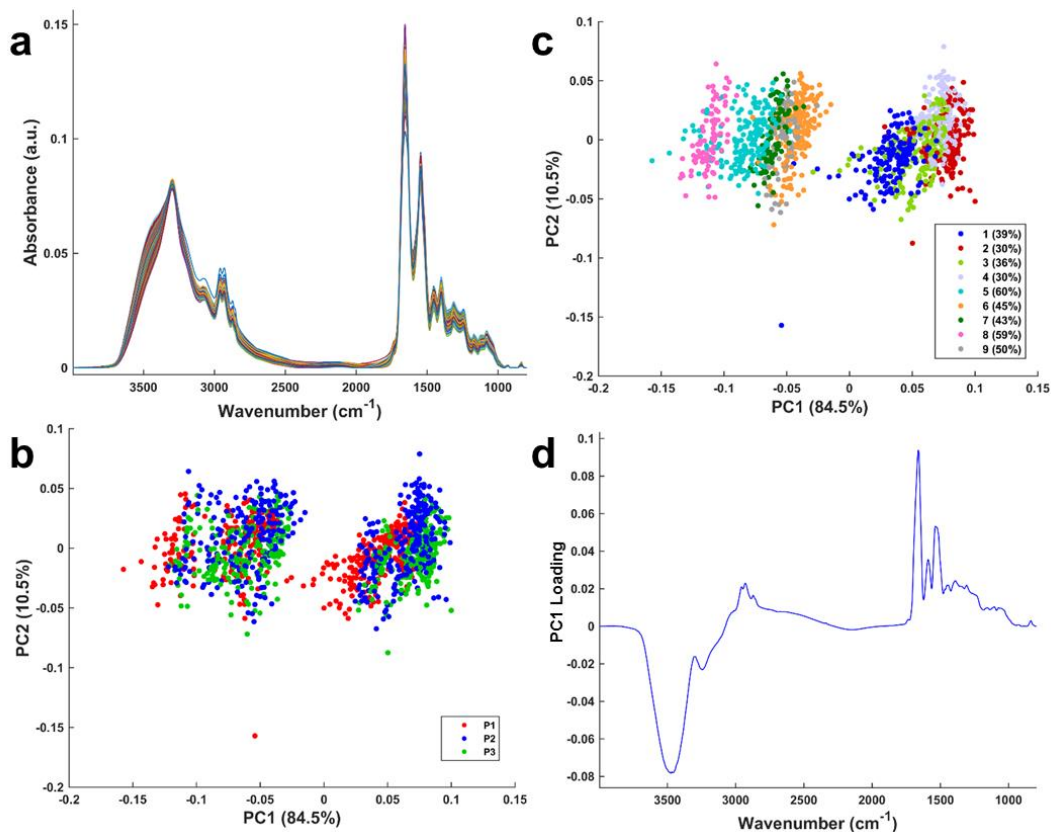
AII.5 (a) *All normalised absorbance plasma spectra of the 3 subjects.* ***(b)*** *PCA scatter plot of the 3 subject plasma spectra.* ***(c)*** *PCA scatter plot of the same plasma spectra highlighting the 9 experiments performed with associated relative humidity (%).* ***(d)*** *PC1 loading from (b) and (c).*
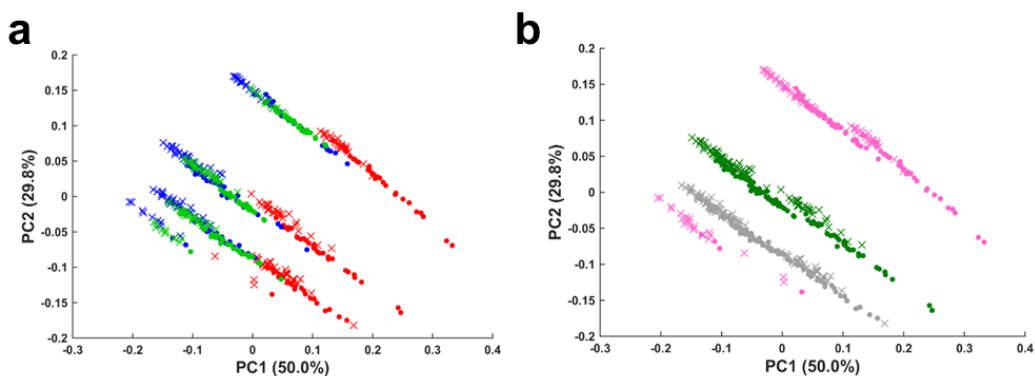
AII.6 *Plasma spectral collection experimental correction by sample vacuum drying.* ***(a)*** *PCA scatter plot of the 3 plasma spectra (red, blue, green) obtained after samples dried at room temperature (dots) or under vacuumed conditions (crosses).* ***(b)*** *PCA scatter plot of the same plasma spectra highlighting the 3 experiments at a RH of 43% (green), 50% (grey) and 59% (pink).*
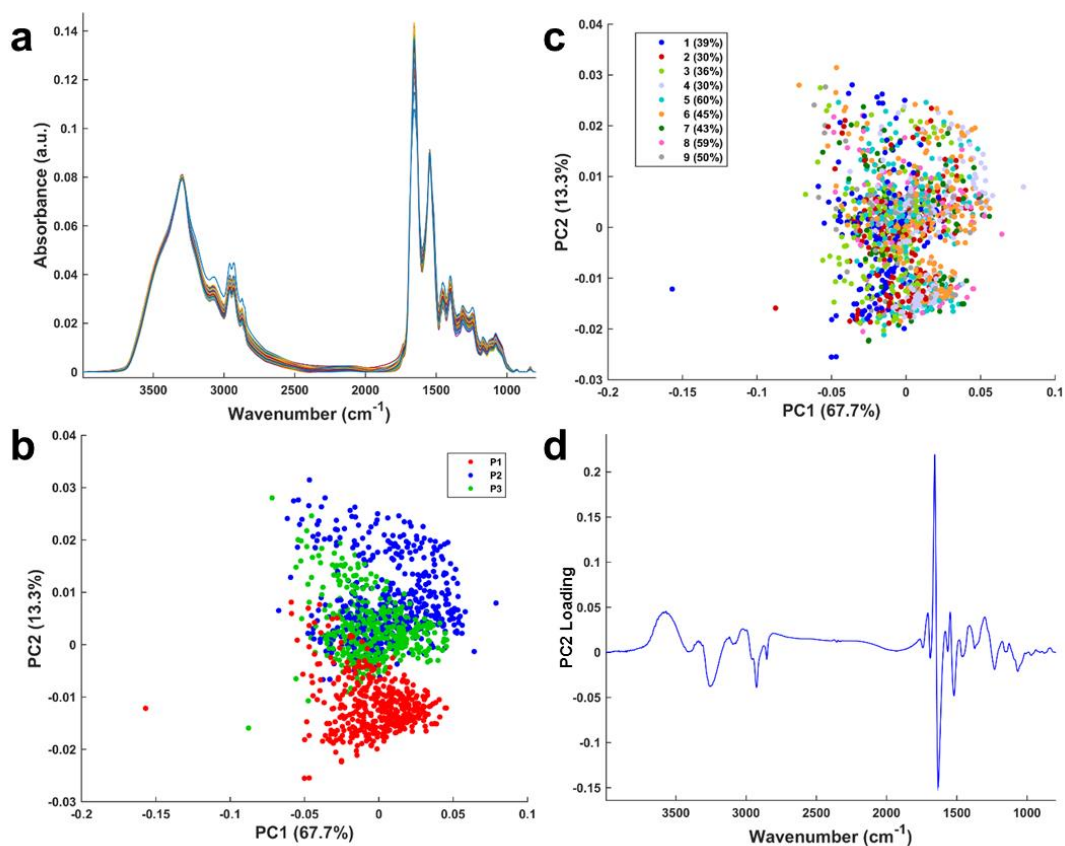
AII.7 *Spectral data water correction by PC removal. (**a**) All normalised absorbance plasma spectra of the 3 subjects reconstructed without the PC1 from AII.5d. (**b**) PCA scatter plot of the 3 subject plasma spectra. (**c**) PCA scatter plot of the same plasma spectra highlighting the 9 experiments performed with associated relative humidity (%). (**d**) PC2 loading from (b) and (c).*
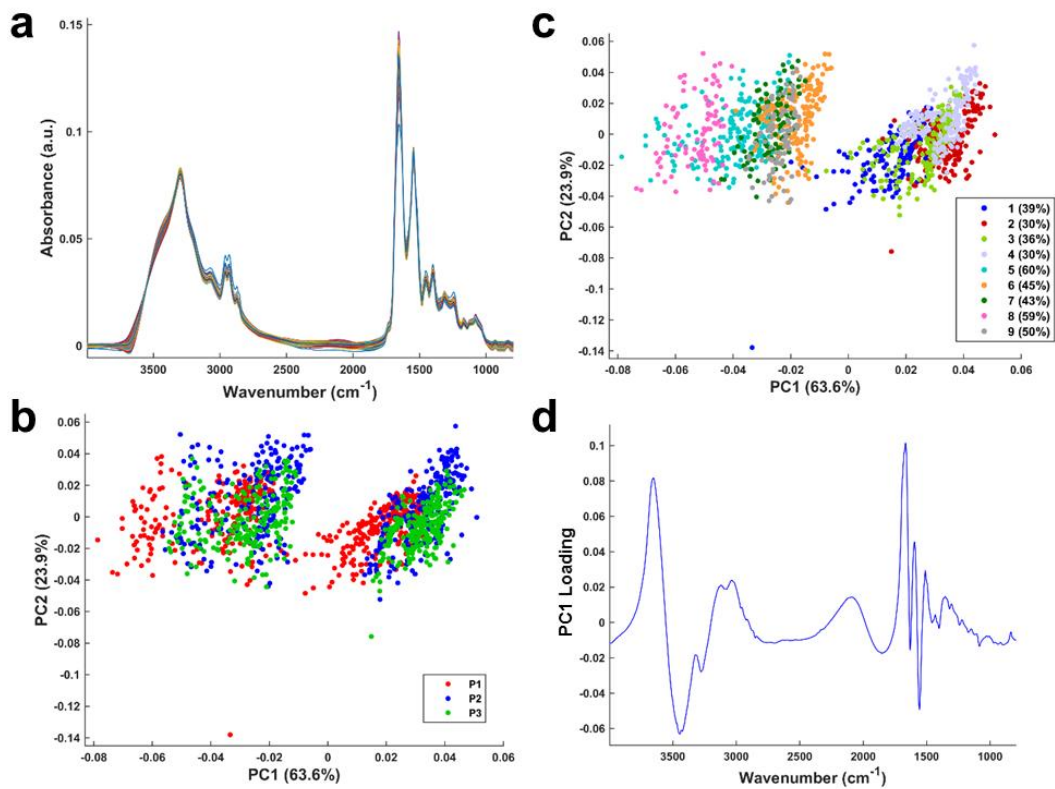
AII.8 *Spectral data correction by EMSC and a constituent spectrum. (**a**) All normalised absorbance plasma spectra after an EMSC with a normal saline constituent spectrum. (**b**) PCA scatter plot of the 3 subject plasma spectra. (**c**) PCA scatter plot of the same plasma spectra highlighting the 9 experiments performed with associated relative humidity (%). (**d**) PC1 loading from (b) and (c).*
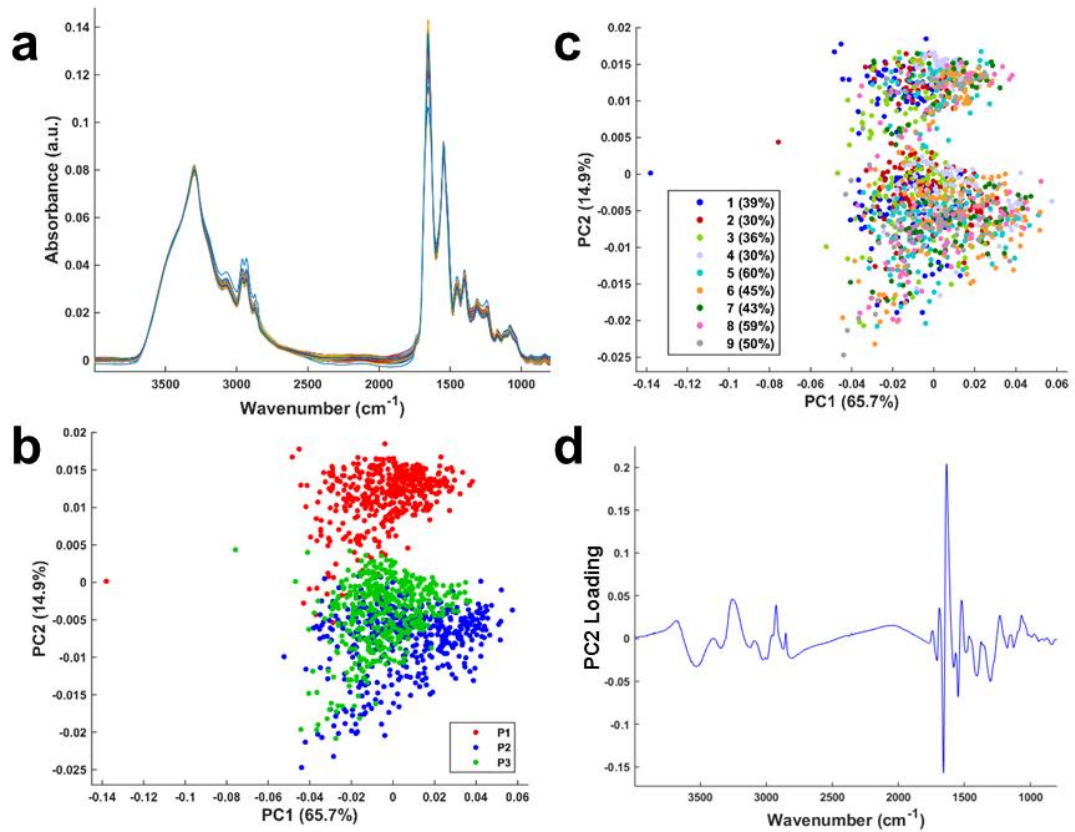
AII.9 *Spectral data correction by a constituent spectrum based EMSC and PC removal. (**a**) All normalised absorbance plasma spectra after an EMSC with a normal saline constituent spectrum and without the PC1 from AII.8d. (**b**) PCA scatter plot of the 3 subject plasma spectra. (**c**) PCA scatter plot of the same plasma spectra highlighting the 9 experiments performed with associated relative humidity (%). (**d**) PC2 loading from (b) and (c).*

AII.10 *Mean standard deviation (STD) calculated on 3 spectral regions according to different plasma spectral water correction methods.*

Correction	Mean STD		
	$3600\text{-}1320\ cm^{-1}$	$1720\text{-}1600\ cm^{-1}$	$1600\text{-}1500\ cm^{-1}$
No correction	$4.13e^{-03}$	$3.46e^{-03}$	$2.79e^{-03}$
PC removal	$4.76e^{-04}$	$1.59e^{-03}$	$9.87e^{-04}$
$EMSC_{NS}$	$8.27e^{-04}$	$2.40e^{-03}$	$1.31e^{-03}$
$EMSC_{NS}$ + PC removal	$2.96e^{-04}$	$1.50e^{-03}$	$8.52e^{-04}$

NS: Normal Saline

Appendix III. Developing and understanding biofluid vibrational spectroscopy: a critical review

Chem Soc Rev
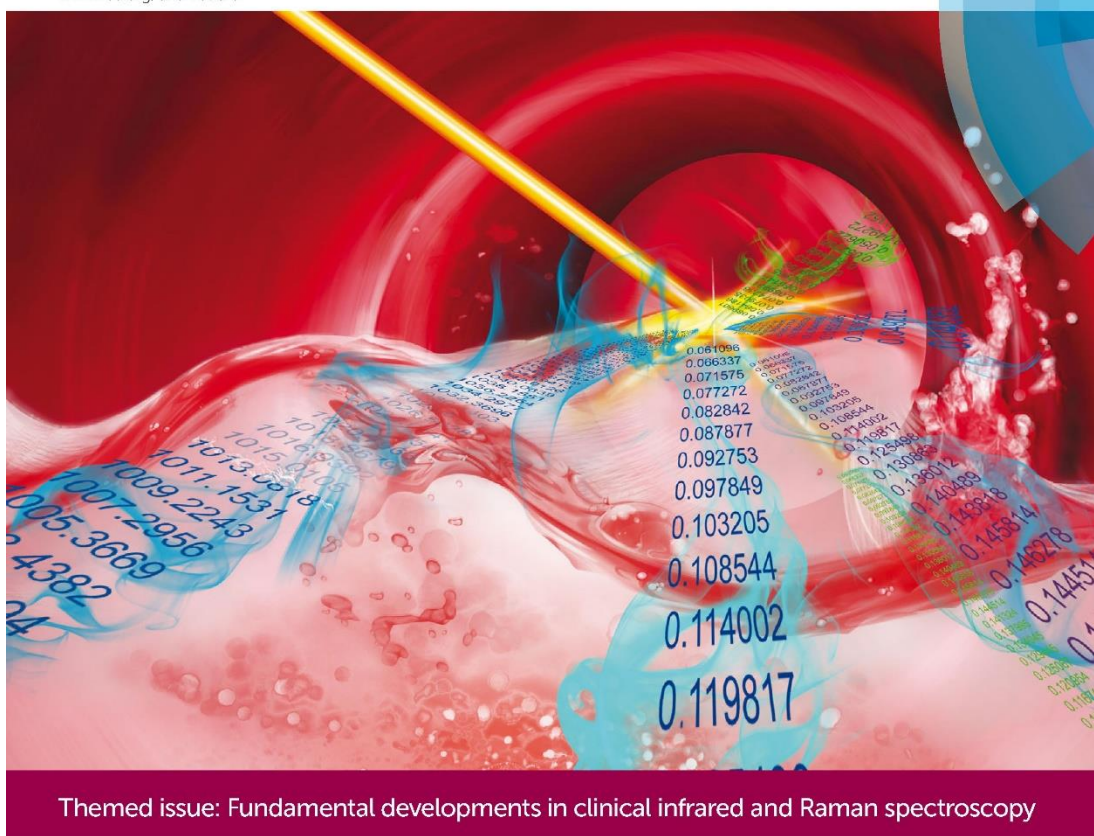
Themed issue: Fundamental developments in clinical infrared and Raman spectroscopy

ROYAL SOCIETY OF CHEMISTRY

175 YEARS

REVIEW ARTICLE

Developing and understanding biofluid vibrational spectroscopy: a critical review

Matthew J. Baker,†*[a] Shawn R. Hussain,‡[bc] Lila Lovergne,‡[ao] Valérie Untereiner,[bo] Caryn Hughes,[e] Roman A. Lukaszewski,[f] Gérard Thiéfin[og] and Ganesh D. Sockalingum†*[b]

Vibrational spectroscopy can provide rapid, label-free, and objective analysis for the clinical domain. Spectroscopic analysis of biofluids such as blood components (*e.g.* serum and plasma) and others in the proximity of the diseased tissue or cell (*e.g.* bile, urine, and sputum) offers non-invasive diagnostic/monitoring possibilities for future healthcare that are capable of rapid diagnosis of diseases *via* specific spectral markers or signatures. Biofluids offer an ideal diagnostic medium due to their ease and low cost of collection and daily use in clinical biology. Due to the low risk and invasiveness of their collection they are widely welcomed by patients as a diagnostic medium. This review underscores recent research within the field of biofluid spectroscopy and its use in myriad pathologies such as cancer and infectious diseases. It highlights current progresses, advents, and pitfalls within the field and discusses future spectroscopic clinical potentials for diagnostics. The requirements and issues surrounding clinical translation are also considered.

Introduction

Biophotonic techniques are widely used in research for developing new modalities with the aim to improve patient healthcare *via* better diagnosis, prognosis, and surveillance. Vibrational spectroscopy holds such promises because the "molecular fingerprint" that it provides represents a snapshot of the sample biomolecular composition and variations therein can be exploited to identify different pathologies.[1,2] Its sensitivity to such variations makes it possible to probe pathophysiological processes in cells and tissues as demonstrated by many reports for more than a decade,[3] leading to the concepts of "spectral cytopathology" and

[a] *WESTChem, Department of Pure and Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, 99 George Street, Glasgow, G1 1RD, UK. E-mail: matthew.baker@strath.ac.uk; Web: www.twitter.com/ChemistryBaker*

[b] *Equipe MéDIAN-Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, CNRS UMR 7369-MEDyC, UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims Cedex, France. E-mail: ganesh.sockalingum@univ-reims.fr*

[c] *Centre for Materials Science, Division of Chemistry, JB Firth Building, University of Central Lancashire, Preston, PR1 2HE, UK*

[d] *Plateforme en Imagerie Cellulaire et Tissulaire, Université de Reims Champagne-Ardenne, 51 rue Cognacq-Jay, 51096 Reims Cedex, France*

[e] *Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK*

[f] *Chemical Biological Radiological Division, DSTL Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK*

[g] *Service d'Hépato-Gastroentérologie, CHU de Reims, Hôpital Robert Debré, 51092 Reims Cedex, France*

† Both authors contributed to the work and project supervision equally.

‡ Both authors contributed equally to this work.

"spectral histopathology".[4–6] With the advances in spectroscopic/ imaging technologies and data processing techniques, cells and tissues can be analysed rapidly and non-invasively to identify disease-related abnormalities. Indeed, some promising studies have reported the added value of vibrational spectroscopy to deliver an objective diagnosis but they were performed on a limited number of patients.[7] In spite of these advances in cell and tissue spectroscopy, the technique has not yet been able to overcome the experimental research phase in order to be transferable from bench to bedside. This is mainly due to the lack of standardisation and validation in large clinical trials and multicentre actions. Access to large sample sets with ethical approval is also a limitation. We believe that spectroscopic diagnosis/prognosis *via* biofluids represents an interesting alternative to cells and tissues. Presently, there is limited research representing high-powered clinical studies for biofluid spectroscopy, yet through the use of animal systems several studies from Naumann's group have set the precedent for studies involving large sample numbers; instilling confidence in the high sensitivity and specificity model outcomes by using several hundred animals per study.[8–10]

The quest for disease markers through "liquid biopsies" is a fast emerging field and has only been recently explored by spectroscopic approaches. Blood components like serum and plasma are routinely used for blood testing as they contain biomarkers that are useful for disease diagnostics. For example, in diseases like cancer, they are known to be a rich source of information and represent readouts of the ongoing cellular and extracellular events.[11] Further, they are easily accessible and

minimally invasive for patients making large studies feasible. Other organ-specific biofluids in the proximity of the diseased cells or tissues like bile, urine, sputum, and cerebrospinal fluid are also of interest for diagnostic purposes. Recent trends tend to indicate that the use of single or few biomarkers has fallen out in favour of multiple biomarkers[12] and in this context the role of vibrational spectroscopic methods can be determinant as the information provided contains data on global sample biomolecular composition providing a chemical 'fingerprint' or 'signature' of the sample. We will focus on the ability of vibrational spectroscopic analysis to illuminate these disease signatures (disease pattern recognition) for diagnostic purposes as opposed to the quantitative determination of specific macromolecules within the biofluid.[13–15]

This critical review, from both the spectroscopic and clinical points of view, considers the issues encountered during translational research aimed at assessing the potentials of infrared and Raman approaches as rapid and label-free diagnostic methods for biological fluids. In addition, the techniques can be adapted to a variety of diseases and therefore represent a cost-effective investment for healthcare systems. This approach could provide a dynamic diagnostic environment that will enable rapid diagnostics leading to earlier treatment. In addition, the ability to accurately and rapidly monitor disease will allow for closer patient follow-up and earlier change in treatment if needed. This would enable patients to access treatment earlier with reductions in mortality and morbidity.

Vibrational spectroscopy

Vibrational spectroscopy relates to specific optical techniques of infrared (IR) and Raman spectroscopy. These techniques probe intramolecular vibrations and rotations of the sample when irradiated with light.[16] The light–matter relationship is underpinned by the electromagnetic theory postulated by Maxwell.[17] Vibrational spectroscopy has been used for analysing a myriad of samples in chemical, physical and biological applications.

The Raman effect constitutes the spontaneous inelastic light scattering process of photons following the interaction of a monochromatic radiation (e.g., laser source) with the sample. During this interaction both elastic and inelastic scattering processes take place. A high proportion of the photons are elastically scattered with no change in energy (so no molecular information), known as Rayleigh scattering.[17] When photons transfer energy to the molecules as vibrational energy, the energy loss of the scattered photons corresponds to the vibrational energy levels of the molecules. This is known as the Raman-Stokes scattering. The incident photons can in turn receive energy from vibrating molecules, and therefore their frequencies increase, described as the Raman anti-Stokes scattering. Fig. 1 shows the transitions involved during these three processes. In spontaneous Raman, the Stokes scattering is generally used due to its higher sensitivity.

Infrared spectroscopy (IR) is broadly defined as the study of absorption characteristics arising from the molecular motion of materials due to atomic displacement[4] upon intimate interaction with an infrared source.[18] Depending on the modality of choice, the radiation can be either transmitted, internally reflected, reflected, or transflected (a combination of transmission and reflectance). During the light–matter interaction, infrared light causes a molecule to enter a higher vibrational state due to the transfer of 'quanta' or 'packets' of energy at certain wavelengths dependent upon the composition of the matter under analysis. Fig. 1 illustrates the energy level transition involved in the IR
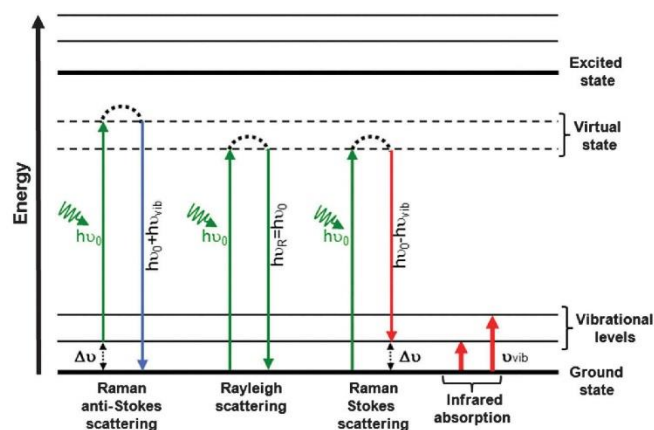
Fig. 1 Energy diagram showing transitions involved during infrared absorption, Rayleigh, Raman Stokes and anti-Stokes scattering. This Jablonski diagram shows that the same vibrational states of a given molecule can be probed via two different routes; one directly measures the absolute frequency (IR absorption) and the other measures the relative frequency or Raman shift (Stokes and anti-Stokes). $h\nu_o$ = incident laser energy, $h\nu_{vib}$ = vibrational energy, $\Delta\nu$ = Raman shift, ν_{vib} = vibrational frequencies.

1804 | Chem. Soc. Rev., 2016, 45, 1803–1818

This journal is © The Royal Society of Chemistry 2016

204

absorption process compared to Raman scattering showing that vibrational energy levels can be probed with both techniques using different physical processes. These transitions result in a spectrum constituted of peaks/bands that can be interpreted qualitatively (peak position) and quantitatively (peak intensity/area, relative intensity). For IR spectroscopy the bands arise from a change in the electric dipole moment of the molecules, whereas Raman is associated with a change in the molecular polarisability.

Constituent chemical molecular bonds present many forms of vibrations which occur at different energies corresponding to different allowed transitions. IR and Raman spectroscopies are complementary and provide a "fingerprint" or "signature" of the molecules contained within the sample depending on whether their bonds exhibit Raman or IR activities. Certain vibrations that are allowed in Raman may be forbidden in IR and *vice versa*. For a full treatise of fundamental spectroscopy studies, the authors direct the reader to two reviews by Barth and Haris on IR spectroscopy[19] and Long on Raman spectroscopy.[20]

Biological and biomedical vibrational spectroscopy

There is a continuing effort devoted to the exploration of new technologies that can detect early signs of diseases and therefore significantly reduce mortality and morbidity. This depends on the ability to detect biochemical/morphological changes at an early stage of the disease or before the disease becomes symptomatic. Detection of biomarkers plays an important role in this exploration, and in the case of cancer for example, they cover a broad range of biochemical entities, such as nucleic acids, proteins, carbohydrates, lipids, small metabolites, and cytogenetic and cytokinetic parameters, as well as entire circulating tumour cells found in body fluids. They can be used for risk assessment, diagnosis, prognosis, and for the prediction of treatment efficacy and toxicity and disease recurrence.

Over the last 20 years, there has been an exponential increase in the number of studies dedicated to identification of new cancer (Fig. 2a) and infectious disease (Fig. 2b) biomarkers, mainly because of the tremendous development of high throughput molecular technologies and associated bioinformatics. However, among the huge amount of candidate biomarkers, only a limited number have been validated for use in medical practice. A recent paper states that in DNA and proteomic research, out of 1000 biomarkers discovered less than 100 have been validated for routine clinical practice.[21]

Vibrational spectroscopy can contribute in bringing a new way for searching biomarkers, namely "spectral signatures" or "spectral biomarkers", which reflect the total biochemical composition of the studied sample as it has been employed for cell and tissue analysis since the pioneering work by Mantsch, Naumann and Diem, to list just a few.

Biological samples are frequently analysed *via* the transmission mode in the mid-IR region, where most molecules absorb
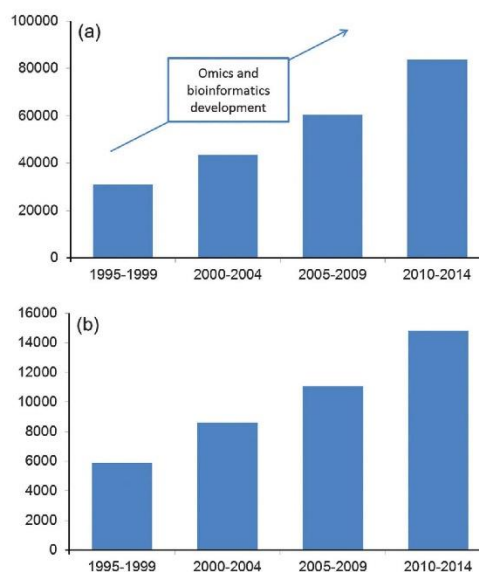
Fig. 2 Number of publications returned in PubMed when inputting the terms "cancer biomarker" (a) and "infection biomarkers" (b).

and the molecular absorbance is proportional to concentration, obeying Beer–Lambert's law for non-scattering samples. Mid-IR absorption features between approximately 4000 and 400 cm^{-1} (2.5 to 25 μm). Fig. 3 illustrates an example of an FTIR spectrum of a breast tissue with the assignment of some important biomolecules. The spectrum can be divided into four regions where the main macromolecules absorb: –CH2 and –CH3 groups of fatty acids and proteins (3050–2800 cm^{-1}); C—O stretching vibrations mainly from lipid esters (1800–1700 cm^{-1}); C—O, N–H, and C–N modes from Amide I and II protein bands (1700–1500 cm^{-1}); phosphate vibrations from nucleic acids (1225 and 1080 cm^{-1}); and carbohydrate absorptions (1200–900 cm^{-1}). Libraries housing spectra from biological and biochemical samples have been collected over the years.

Over the years, variants of IR spectroscopic technologies have been tested. A recent review highlights the use of IR techniques to probe the functionality of biological and biomimetic systems.[22] Their applications to study biological and biomedical specimens have continuously increased.[23,24] When used to analyse biofluids, the mid-IR or near-IR spectroscopies would be performed on drying samples to negate the overwhelming water band from obscuring spectra and to increase automation.[25]

Another method of obtaining an IR spectrum is when the sample is placed on a highly reflecting surface, typically aluminium/teflon coated substrates or a glass slide with tin oxide-based silver reflective coating called low e-slides (*e.g.*, MirrIR). In this case the process is termed transflection because the IR beam passing through the sample is reflected off the slide and passes

This journal is © The Royal Society of Chemistry 2016

Chem. Soc. Rev., 2016, **45**, 1803–1818 | **1805**

Fig. 3 FTIR biological spectrum showing frequent biomolecular band assignments from 3000–800 cm^{-1}, where ν = stretching vibrations, δ = bending vibrations, s = symmetric vibrations and as = asymmetric vibrations. Illustration taken from transmission spectra on human breast ductal carcinoma, prepared on 1 mm thick BaF$_2$ slides.[4] The 3000–2800 cm^{-1} region originates mostly from lipids (CH, CH$_2$ and CH$_3$ stretching modes), but protein absorption of the same modes also contribute to these absorption bands.

again through the sample before detection. These substrates have very low cost but recently they have been shown to cause significant spectral intensity variations, due to an electric free standing wave artefact (EFSW)[26,27] which could be misinterpreted as composition variations while it is the sample thickness variation that is questionable. The fundamental question when using low e-slides is whether the spectral variations observed due to the EFSW impact on the discriminant spectral differences. In case of thin samples such as air-dried cellular monolayers, recent research by Cao *et al.* has shown that the same classification was obtained when performing transmission and transflection measurements.[28]

Attenuated total reflectance (ATR) FTIR spectroscopy is a promising modality for biological sample analysis. The guided IR beam propagates through a high refractive index crystal surface producing an evanescent standing wave that penetrates the sample by a few microns. However for appropriate use, several issues need to be considered, such as contact between the ATR crystal and the sample, the beam penetration depth and image distortion due to high refractive indices.[25,29]

Despite its molecular specificity, FTIR spectroscopy suffers from some shortcomings which limit its application to the measurement of biological samples and their dynamic behaviour. An important one is sensitivity, in particular in thin samples as a result of the Beer–Lambert's law. Signal amplification can be achieved by the plasmonic resonances of nano-scale metallic particles,[30] resulting in the phenomena of surface-enhanced infrared absorption (SEIRA),[31] in analogy with surface-enhanced Raman scattering (SERS).[32] Early SEIRA studies utilised metal island films[22,31] and dried samples, but today plasmonic

chip-based technology enables the *in situ* monitoring of protein and nanoparticle interactions in aqueous media, at high sensitivity in real time.[33]

One method of choice for cell and tissue analysis has been IR microspectroscopy. The coupling of an FTIR spectrometer with a microscope has helped to perform microanalysis and gain in spatial resolution ~15–20 μm with a thermal source and ~5–10 μm with a synchrotron source using single element detectors. These systems are limited by low sensitivities and time-consuming experiments (several hours) as they remain a point by point acquisition. In the 1990s, the advent of imaging devices with multi-element detectors combined with aperture less microscopes, high-tech automation and faster computers, have d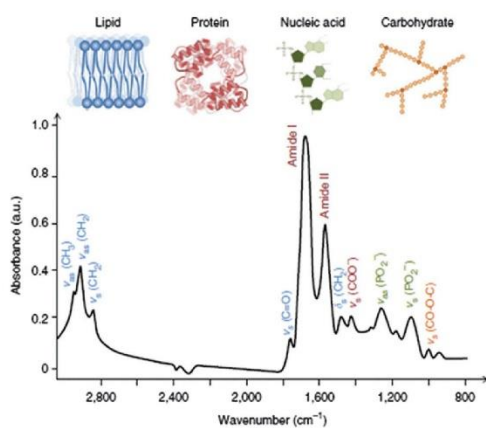rastically reduced the data acquisition times (few hours) with resolution going down to ~2 μm per pixel with liquid nitrogen cooled focal plane array detectors. Many research groups have demonstrated the efficacy of employing this to a clinical setting on biopsy samples; minimising subjectivity and increasing diagnostic accuracy.[4] In spite of these progresses, such instruments remain research machines and are not adapted to be used as benchtop techniques for routine analysis in a clinical setting.

The launching of new IR imaging devices incorporating high-intensity tunable quantum cascade lasers (QCL) could revolutionise the way clinical IR images are acquired.[34] High-throughput IR chemical imaging is now in its early days, and needs to be tested and validated. However, a gain of three orders of magnitude in acquisition time has recently been reported for large samples by Bhargava's group.[35] Combining signal enhancement from SEIRA and fast imaging using a QCL source with small bandwidths, a recent study claimed a ~200 fold gain in imaging time.[36]

The Raman shift covers the range between 0 and 4000 cm^{-1}. Raman spectroscopy can be used in the confocal mode and with the resonance and surface-enhanced modalities. Applications of Raman microspectroscopy for probing biological systems have been continuously expanding over the years along with IR spectroscopy.[37] Its high spatial resolution (~0.5 μm with green lasers), compatibility with aqueous environment,[38,39] and *in vivo* amenability[40–43] makes it a good candidate for biological and biomedical research. Akin to FTIR, it also provides high content biomolecular information. Microspectroscopy with immersion measurements can be used to enhance the signal to background ratio enabling higher quality data acquisition as demonstrated by Bonnier *et al.*[44]

Due to its advantages, label-free, high spectral specificity, limited water signal, and the fact that most biological molecules are Raman active, Raman has been deployed to *in vitro* cell and tissue studies, but now significant developments of *in vivo* work due to the compatibility with fibres has enabled Raman endoscopy in a label-free manner and *in vivo* Raman probes/endoscopes have made direct tissue analysis possible.[45]

New fields of measurement and implementation possibilities have multiplied due to recent hardware developments, improved sampling methods, and advances in the design of Raman technology alongside developments and advances in

1806 | *Chem. Soc. Rev.*, 2016, **45**, 1803–1818

This journal is © The Royal Society of Chemistry 2016

206

multivariate data analysis. It has been possible to uncover subtle disease-related spectral changes and exploit them in classification models. However, an important drawback of Raman spectroscopy is that the effect is inherently weak as a very small proportion of incident photons are scattered (~ 1 in 10^8) with a corresponding change in frequency.[17] This together with the fact that to date most of the commercial systems use dispersive configurations adds another limitation compared to fast IR imaging systems, and makes Raman imaging of biological specimens a slower process. These limitations can be partly circumvented with other Raman modalities based on Resonant Raman Scattering (RRS) and Surface-Enhanced Raman Scattering (SERS) to enable gains in detection sensitivity.[37] In SERS technology, the use of functionalised metal nanosurfaces has allowed optimising the enhancement to several orders of magnitude depending on the metal substrate. Metal nanoparticle arrays and single nanoparticles have been utilized for high-throughput detection.[46] SERS has been applied in different areas in the chemical and biological fields[47] and its very high sensitivity has allowed single molecule detection.[48] Until recently, SERS was not widely applied to biomedical research because of issues linked to complexity of the biological medium, biocompatibility, reproducibility, and short shelf life. However, using silver and gold colloids as SERS substrates, Bonifacio *et al.* recently showed that repeatable spectra could be obtained from protein-free blood serum and plasma.[49]

Furthermore, non-linear Raman spectroscopy has been developed to be applied to biomedical analysis like Stimulated Raman Scattering (SRS) and Coherent Anti-Stokes Raman Scattering (CARS), for rapid image acquisition (one Raman band at a time) with higher sensitivities than spontaneous Raman.[50–52] For non-linear Raman, it is important to know which marker band(s) are useful, in analogy to the application of Discrete Frequency-IR (DF-IR) as enabled by the use of QCL sources.

Other areas of current interest for Raman spectroscopy are exploring the sampling depth and location of spectral information. For instance, seminal research conducted by Stone, Matousek and collaborators demonstrated the principle of spatially offset Raman spectroscopy (SORS) for subsurface analysis towards *in vivo* breast cancer[53,54] and deep Raman measurements using liquid tissue phantoms to mimic non-invasive cancer screening applications *in vivo*.[55] Through-tissue sensitivity was increased *via* SESORS measurements at several millimetres depth, *i.e.*, combining SORS with nano-tagged SERS particles.[56–59]

Building on the research described above, the field of biospectroscopy has continuously progressed and expanded to complex biological systems such as biofluids[60] with a major focus on the development of a potential diagnostic/prognostic tool with remarkable scope and future clinical promises.

With the global disease burden set to rise, a more rapid, non-invasive, label-free, non-destructive, automatic and cost effective diagnostic technique like vibrational spectroscopy would revolutionise the clinical environment. Its utility as a biofluid diagnostic tool is heavily reliant on the principle that cellular and tissue dysfunction or irregularities affect the biochemical make-up of biofluids, manifesting as protein, carbohydrate, lipid, and nucleic acid subtle differences.[16]

Over the last decade, developments in this field have been ongoing in order to fulfil these objectives and ultimately leading to better diagnostics and time to results to improve patient outcomes, offer more efficient public services, and reduce health costs.

Biomarkers in body fluids

According to the National Institutes of Health definition, a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention".[61] In line with this definition, there is a large range of clinical situations where the biomarkers are of paramount importance for the patient's management: screening of patients at risk of the disease or with the disease at an early stage, differential diagnosis of disease with other conditions, prognosis of the disease independently of the treatment, prediction of the response to treatment, and monitoring of disease evolution (Fig. 4).

Molecular biomarkers may be detectable in tissues and biofluids. Fig. 5 illustrates the case for cancer where tissue biomarkers can be used for cytological or pathological assessment of the disease or for molecular or spectral imaging techniques. The tumour is vascularised and markers are shed into the bloodstream. Another health issue is the early detection of biomarkers for the diagnosis of infectious diseases coming either from the host or from the pathogen. From the initial interaction onwards the majority of biomarkers available to measure are derived from the host since pathogen numbers are very low and the host is able to utilize components of both the innate and adaptive host response to drive an appropriate response. In serious infection, when pathogens are able to overcome the early host response to their presence, their numbers increase at an exponential rate resulting in significant mortality rates. In such cases, the relative concentration of microbial biomarkers increase over time whilst biomarkers associated with the ongoing, yet ineffective,
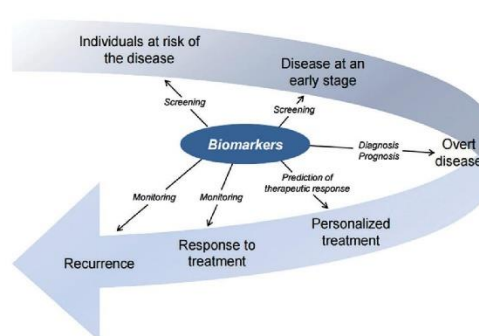

Fig. 4 Schematic of biomarker use in clinical practice.

This journal is © The Royal Society of Chemistry 2016

Chem. Soc. Rev., 2016, **45**, 1803–1818 | **1807**

207
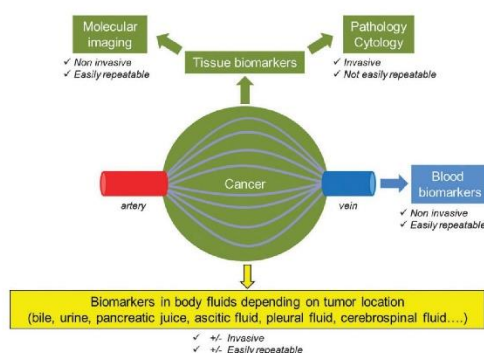
Fig. 5   Example of potential tumour-site related biomarkers.

host response are still readily detectable (Fig. 6). Preliminary evidence has been produced which indicates that it is possible to identify the presence of an infectious organism through analysis of host biomarker signatures before patients become symptomatic.[62] Thus, the concept of searching for such signatures in host biofluids pre-symptomatically appears as a promising avenue for exploration in order to enable early therapeutic intervention.

Regarding biofluids, blood and its constituents appear the most convenient for biomarker/biosignature detection given its ease of availability and the possibility to repeat the test as often as necessary to monitor disease progression or response to treatment.
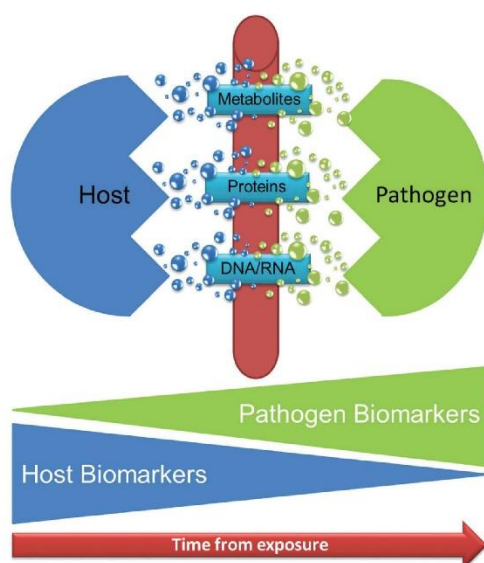
Fig. 6   The relative contribution of host and microbial derived biomarkers to enable diagnosis of infection.

Blood serum houses more than 20 000 different proteins. It perfuses all body organs meaning it contains a large range of proteomes from surrounding tissues and cells, making it the most complex biofluid.[63] The low molecular weight fraction serum component of blood, known as the "peptidome" is information rich for diagnostic purposes.[11] Other biofluids (bile, urine, sputum, pancreatic juice, and ascitic, pleural, cerebrospinal fluids), in direct contact with the diseased tissue, are of great interest as media to detect biomarkers/biosignatures that are secreted or shedded locally. These are expected to be present in higher concentration in these fluids than in the blood. In addition, their identification may be facilitated by a less complex molecular composition of local biofluids compared with blood. Although some biofluids such as urine share with blood samples ease of availability and repeatability, analysis of other biofluids requires an invasive procedure, which limits their repeated use in the clinical setting. An example is cerebrospinal fluid which requires a lumbar puncture for collection.

Whilst biomedical vibrational spectroscopy has been developed initially mainly for cell and tissue analysis, it has been also applied more recently to biofluids for biomarker discovery, generating a number of pilot studies with promising results as presented below. The challenge is now to translate the results of these exploratory studies to the routine clinical practice.

Biofluid spectroscopy

The search for disease markers in biofluids *via* photonic approaches is a fast emerging field and has only been recently explored by vibrational spectroscopic approaches. Biofluids are easily accessible and minimally invasive for patients making large studies feasible. Like cells and tissues, biofluids exhibit vibrational spectra that have characteristic bands reflecting their biomolecular composition. Fig. 7 compares the FTIR spectra of some dried biofluids (serum, plasma, and bile) obtained with a high-throughput module in the transmission mode. IR spectra of serum and plasma present very close profiles with subtle differences that are difficult to depict visibly. This is explained by the fact that serum is essentially plasma with the clotting factors of blood removed. The assignment of the main bands is provided in Table 1.

The bile spectrum differs through a higher lipid and carbohydrate content and by relative intensity changes of the protein amide I/amide II bands.[64]

Raman spectroscopy gives complementary information to IR. Besides the main macromolecules like proteins, lipids, and carbohydrates, other modes originating from amino acids for example are active. The assignment of the main bands is indicated in Fig. 8 showing an example of a typical Raman serum spectrum taken from a dried drop.

Serum and plasma

At present, the majority of the biofluid spectroscopy research has focused on serum and plasma. This is most likely due to

Fig. 7 Comparison between HT-FTIR spectra of different biofluids: serum (red curve), plasma (blue curve), and bile (green curve). Spectra are background corrected and normalised. Note: serum and bile were collected in dry tubes while for plasma samples lithium heparin tubes were used.

Table 1 Assignment of the major absorption bands of a plasma FT-IR spectrum[89]

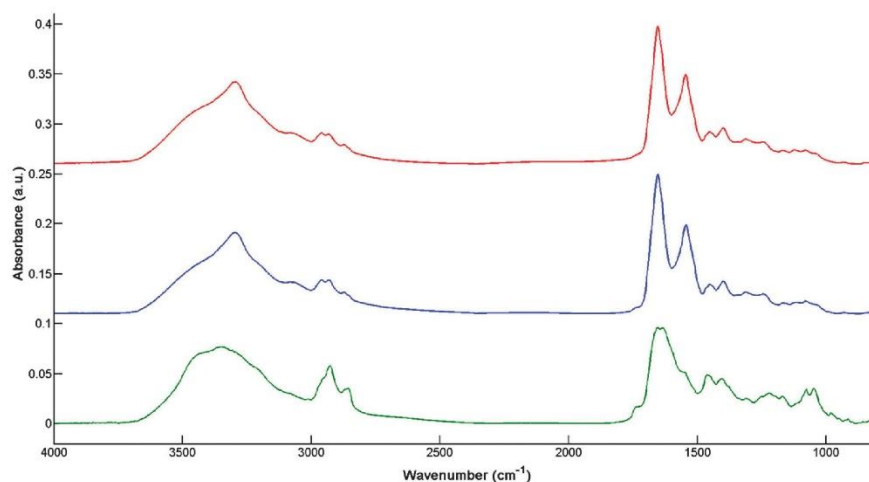| Bands (cm^{-1})	Major assignments for plasma contents
3300	ν(N–H) of proteins (amide A band)
3055–3090	ν(=CH) of lipids
2950–2960	ν_{as}(CH$_3$) of lipids
2920–2930	ν_{as}(CH$_2$) of lipids
2865–2880	ν_s(CH$_3$) of lipids
2840–2860	ν_s(CH$_2$) of lipids
1730–1760	ν(C—O) of fatty acids
1660	ν(C=O) of proteins (amide I band)
1550	δ(N–H) of proteins (amide II band)
1400	ν(COO$^-$) of amino acids
1240	ν_{as}(P=O) of nucleic acids
1170–1120	ν(C-O) and ν(C-O-C) of carbohydrates

ν: stretching vibrations, δ: bending vibrations, s: symmetric, as: asymmetric. Taken from Lacombe et al., Analyst, 2015, 140, 2280.

the prevalence of these types of samples within current biobank stocks or the fact that ethics are already established to collect these samples and all that is required is an addendum stating a separate use of the material.

Malignant diseases. Currently, in the field of oncology, most investigations are proof-of-principle studies showing the potentials of FTIR/Raman spectroscopy to identify different types of cancer from serum samples with high degrees of accuracy. HT-FTIR spectroscopy in transmission mode was used to discriminate urinary bladder cancer patients from patients with urinary tract infection with linear discriminant analysis (LDA) or random forest (RF) classifiers.[65] Using blood serum, Backhaus et al. distinguished between breast cancer and controls with a very high sensitivity and specificity.[66] Chemometrics combining support vector machine (SVM) classification and leave-one-out cross validation was employed by Zhang et al. to separate cirrhotic patients with or without hepatocellular carcinoma.[67]

Equally important is the possibility to identify liver fibrosis stages prior to the development of hepatocellular carcinoma, which are crucial for the clinical management. A study by Scaglia et al. revealed that patients with extensive fibrosis (F3/F4 stages) could be distinguished from those with no fibrosis (F0 stage) on the basis of their FTIR serum spectra using a combination of discriminant wavenumbers.[68] Studies using ATR-FTIR spectroscopy coupled with classification machine discriminated ovarian[69] and endometrial cancers.[70] It also allowed differentiating glioblastoma multiforme (GBM) from healthy control and low grade gliomas and GBM versus healthy control.[71,72]

Applications of Raman spectroscopy to the study of various biofluids from cancer patients are in continuous progress. Sahu et al. analysed serum samples and could differentiate oral cancer patients from controls.[73] More recently, they reported that Raman serum spectroscopy was capable to predict the probability of recurrence in this cancer.[74] Other studies have shown the potential of Raman spectroscopy for differentiating normal subjects from patients with breast,[75] colorectal,[76] or cervical[77] cancers. A proof-of-concept study using micro-Raman spectroscopy applied to the sera of 71 cirrhotic patients showed that it could be an alternative method for discriminating cirrhotic patients with and without hepatocellular carcinoma.[78] On the other hand, SERS of serum or plasma has also been shown as a promising tool for the diagnosis of various types of cancer such as nasopharyngeal,[79–81] digestive,[80,82–84] and prostate cancers.[85]

Non-malignant diseases. Serum and plasma have been also employed to diagnose other diseases using biospectroscopy. For example, Raman serum data allowed to differentiate Alzheimer's disease from other dementia[86] and Carmona et al. used plasma Raman spectral data to grade mild, moderate, and severe Alzheimer cases.[87] Via FTIR spectroscopy of plasma,

209
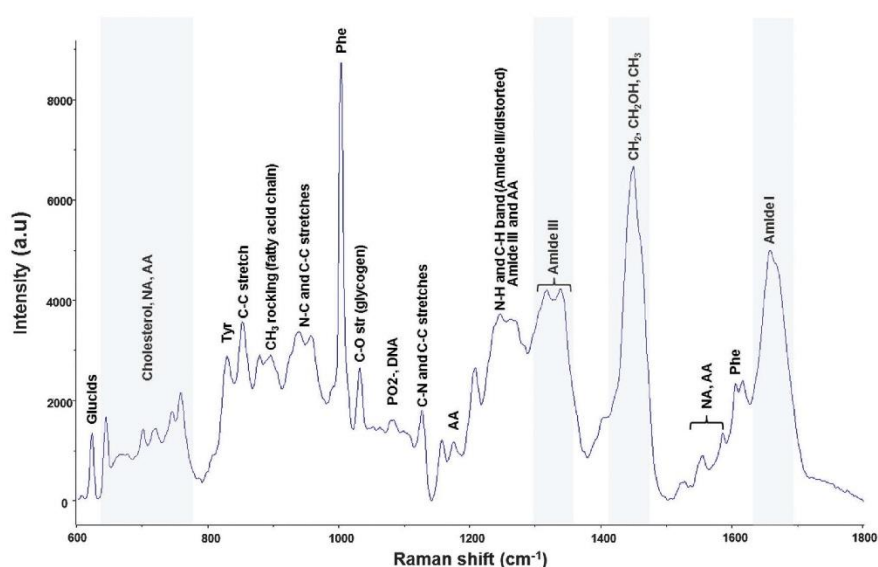
Fig. 8 Typical Raman spectrum of dried serum drop with spectral assignments. Spectrum was measured on a calcium fluoride window with a 785 nm laser excitation with an acquisition time of 2 × 30 seconds.

Peuchant *et al.* have shown that patients with Alzheimer's disease could be well delineated from normal ageing subjects used as controls.[88]

Recent plasma data published by Lacombe *et al.* clearly showed that HT-FTIR spectroscopy could be an interesting alternative technique in neonatal screening of rare diseases such as classic galactosemia. Promising results indicated that healthy/diabetic, healthy/galactosemic, and diabetic/galactosemic patients could be discriminated with good sensitivity and specificity.[89]

Few large studies have been reported. An example is the study led by Petrich's group showing the potential of mid-infrared spectroscopy in the triage of patients with acute chest pain.[90] This study included 1429 serum samples from 389 patients reporting to two US hospitals (Massachusetts General and Latter Day Saints, Utah) consisted of 104 suffering from acute myocardial infarction (AMI), 136 from unstable angina pectoris, and 149 from chest pain of other sources. FTIR measurements were performed in the transflection mode. Using a threshold value generated from a robust linear discriminant analysis, they achieved high sensitivity and specificity enabling triage of patients with AMI, those most at need within the accident and emergency setting, compared to the other sources of chest pain. They hypothesise on the involvement of carbohydrates as discriminant features, possibly a glycation reaction. Interestingly, their results were comparable to the performance of routine cardiac laboratory markers within the same study population. They conclude on the potential of FTIR to aid the diagnostic procedure as early as within the first 6 hours after the onset of chest pain.

Blood plasma from patients has been investigated with Raman spectroscopy as dried drops to identify a reliable biomarker that can differentiate sepsis patients from those with non-infectious systemic inflammatory response syndrome. Neugebauer *et al.* reported on the high sensitivity and specificity that can be achieved.[91] The possibility of separating the two groups of patients is crucial because a stratification of at risk patients can be established for a rapid delivery of appropriate treatment.

Finally, following the results obtained in a model of infected cultured cells, SERS appears as a promising approach for malaria parasite detection from whole blood.[92]

Other biofluids

Other biofluids non-invasively accessible (urine, saliva, sputum, and tears) and invasively accessible (bile, synovial fluid, cerebro-spinal fluid, and amniotic fluid) have been investigated by vibrational spectroscopy for diagnostic purposes.

Somorjai *et al.* were able to distinguish urine samples from normal renal transplants and rejected allografts, applying IR spectroscopy and a three-stage classification strategy.[93] A Raman spectroscopic analysis combined with PCA and quadratic discriminant analysis (QDA) performed on urine, has allowed identification of spectral biomarkers predictive of complications and kidney failure in the urine of diabetic and hypertensive patients.[94] Finally, in the field of oncology, Del Mistro *et al.* reported that SERS using Au nanoparticle substrates had the potential to detect in urine spectral biomarkers of prostate cancer.[95]

Another approach by FTIR spectroscopy associated with LDA on saliva has reported the correct classification of diabetic

210

patients from healthy control.[96] SERS of saliva showed the ability to predict lung cancer by monitoring the decrease of proteins and nucleic acids with 80%, 78%, and 83% accuracy, sensitivity, and specificity respectively.[97] A preliminary study using SERS on saliva suggested the possibility of a quick detection of AIDS but these results obtained on a small number of patients deserve to be confirmed on a larger population.[98]

An exploratory study has shown that FTIR spectroscopy applied to sputum could be a useful approach for the diagnostic of the chronic obstructive pulmonary disease.[99] Investigating the potential of human tears for the diagnosis of ocular diseases, Travo et al. have shown the discrimination of patients with keratoconus (degenerative disorder affecting the cornea) from healthy control and also between patients at an early or advanced stage of disease by HT-FTIR and PCA.[100] Additionally, Choi et al. report that SERS can be used for diagnosis of adenoviral conjunctivitis from tears.[101]

Using HT-FTIR spectroscopy in association with support vector machine (SVM) classification and leave-one-out cross validation (LOOCV), Untereiner et al. have shown that bile samples of patients with malignant biliary strictures were differentiated from those with benign biliary diseases.[64]

Eysel et al. using FTIR spectroscopy and LDA with LOOCV on synovial fluid were able to differentiate samples from joints affected by rheumatoid arthritis, osteoarthritis, spondyloarthropathies, and meniscal injuries.[102] Also from synovial fluid samples, a Raman spectroscopic study associated with a k-means analysis has shown discrimination between patients with osteoarthritis of low or high severity.[103]

Liu et al. have investigated the amniotic fluid potential for fetal lung development assessments by IR spectroscopy. The lecithin/sphingomyelin (lung surfactants) and lung surfactant/albumin ratio measurements by IR spectroscopy were quantitatively and qualitatively correlated to those obtained by thin-layer chromatography and fluorescence depolarization, two clinical methods used to determine fetal lung surfactant maturity in amniotic fluid.[104] Prenatal disorders from amniotic fluids have also been investigated by ATR-FTIR spectroscopy revealing spectral profile changes between amniotic fluids from pregnancies with fetal malformations, preterm delivery and healthy term pregnancies.[105]

Griebe et al. were able by FTIR spectroscopy to distinguish patients with Alzheimer's disease from healthy controls using cerebrospinal fluid.[106]

Translation

With a few exceptions, all the mentioned proof-of-concept studies have been carried out on rather small populations and have shown promises for clinical utility and highlight the potential of vibrational spectroscopy for spectral diagnostics. To our knowledge, two major programmes for large scale clinical trials in remote settings are ongoing using hand-held FTIR modalities. The first campaign led by Wood et al. concerns the screening of population in Thailand for malarial diagnosis (http://monash.edu/news/show/infrared-light-puts-malaria-to-the-test).

A similar approach is being taken in the UK with the establishment of Glyconics Ltd. Glyconics is using sputum to diagnose Chronic Obstructive Pulmonary Disorder and are moving towards clinical validation of handheld ATR-FTIR on a subset of the UK population (http://www.glyconics.com/technology.asp).

These steps towards actual clinical environment testing is pushing the field to the forefront of the application and will illuminate the utility of these techniques as well as barriers to clinical implementation that need to be overcome.

Multivariate analysis

It is becoming more and more evident that vibrational spectroscopy represents an interesting approach to explore the diagnostic potentials of circulating biomarkers/biosignatures in various body fluids.[60] Along with the technological development, the front-end sample preparation challenges and approaches, and the data acquisition procedures, the pre-processing and post-processing of spectral data are equally important for the deployment of various biofluid classes into diagnostics development. Vibrational spectroscopic data are inherently multivariate by nature and their pre- and post-processing require multivariate data analysis approaches.

Different instruments from different manufacturers have different responses and spectral distortions and backgrounds have to be taken into account via pre-processing algorithms in order to compare data from different studies for example. The pre-processing should therefore be able to give accurate, robust and reliable data. These considerations should also include how the sample is prepared and conditioned, the optical substrate used, and the acquisition mode used in order to post-process reliable data. The way the sample is dried or acquired (e.g. transmission or reflection) will also pre-empt the pre-processing procedures. For example, rapid drying of serum can produce a granulating effect which then causes more scattering/dispersion artefacts and a specific correction has to be implemented. It is clear that pre-processing is not the same for infrared and Raman spectra of biofluids because the physical phenomena involved are, respectively, absorption and scattering. In FT-IR spectroscopy the use of an interferometer ensures an excellent intensity and wavenumber calibration. In addition, a background signal is regularly recorded and automatically subtracted to obtain the sample spectrum. For Raman, a day-to-day calibration procedure needs to be implemented to correct for instrument response, and to calibrate the wavenumber and intensity axes. Other experimental considerations include the need to subtract substrate contributions and other physical phenomena such as fluorescence and heating. Biofluid vibrational spectra are therefore corrected, derived (or not), then normalized. As a general rule, it is also important to include prior to the pre-processing steps, a quality test to remove spectra with a poor signal/noise ratio (threshold to be defined depending on the sample nature) and a validated outlier removal routine before post-processing.

The post-processing step includes data mining and the construction of classifiers. Very often, the spectral differences

This journal is © The Royal Society of Chemistry 2016

Chem. Soc. Rev., 2016, 45, 1803–1818 | 1811

211

between normal and pathological states are very subtle and the next step is to perform data mining, *i.e.*, a process used to extract the salient information from the spectral data. By using specific algorithms, patterns can be found in large batches of data. Thus, such feature selection procedures can help to identify discriminant spectral features to discriminate between patient groups.[67] However, it is important to note that data mining depends on effective data collection, the size of the datasets, and as well as their pre-processing.

To build classification models, several multivariate approaches have been used and as of today there is no general consensus on which method is the best. In other research fields, numerous linear and non-linear supervised algorithms have been evaluated and a combination of methods like SVM and PLS-DA has been shown to enhance the sensitivity and specificity of the classifiers.[107] Generally, building the classifier should include a calibration phase (training phase), an internal validation phase, and an external validation phase (blind testing phase). One of the important issues encountered is the size of the data-sets used as a small dataset that does not accurately describe the patient population can lead to under- or over-fitting and impact the classifier outcome. For a classifier to be robust, it is important to have a large number of class-representative patient samples. In addition, the external validation requires a dataset that has not been used in the two previous steps of calibration and internal validation (based upon patient spectra and not replicate spectra from the same patient *i.e.* a spectrum from the same patient should not be in the calibration/internal validation and external validation phases). The leave-one-out cross validation method is often used for these models. It is important to note that all spectra from a given patient must be removed in this process in order to enable a valid outcome. Considering all individual spectra, mean spectra or median spectra as input datasets of the classifier should also be taken into consideration although it has been found that when spectra are highly reproducible and after applying a quality control test plus an appropriate outlier removal, the results are comparable.[64]

The workflow in Fig. 9 illustrates different steps, for both IR and Raman spectroscopies, starting from sample preparation to data pre- and post-processing and the building of classifiers for diagnostics. The issues dealing with pre-processing and post-processing procedures generally used are described in a more detailed manner in a dedicated review elsewhere in this special issue.

Requirements for clinical implementation

Over the last 20 years, the number of studies dedicated to identification of new biomarkers has increased exponentially, mainly because of the tremendous development of high-throughput molecular technologies and associated bioinformatics. However, among the huge amount of candidate biomarkers, only a limited number have been validated for use in medical practice.[108]
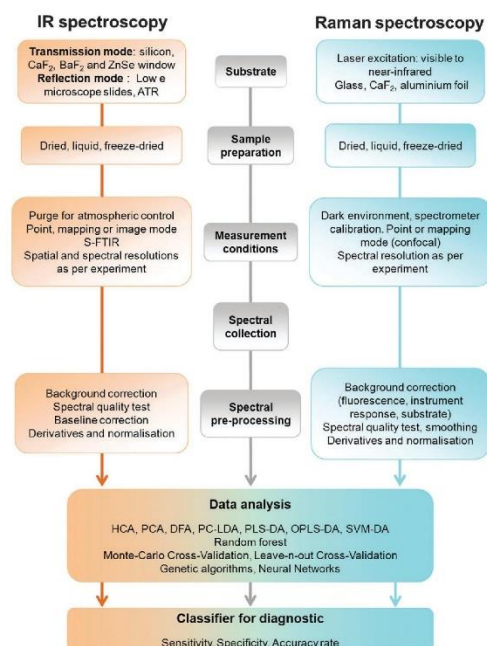
Fig. 9 Workflow of biofluid spectroscopy from substrate choice through sample preparation to spectral measurements and data analysis with diagnostic classifiers.

The origin of this discrepancy has been extensively analyzed in the field of proteomics and genomics. Methodological flaws have been identified in the process of their identification and/or clinical validation and recommendations have been set forth to overcome these inadequacies.[109–111] Studies based on vibrational spectroscopy are subject to the same problems. As for other high throughput technologies, the huge amount of data generated by spectroscopic analysis exposes this analysis to a significant risk of false positive findings. This risk should be minimized by rigorously controlling sample and patient related factors in the exploratory phase and by standardizing the conditions of spectral acquisition, processing and analysis (preanalytic/analytic validity). Subsequently, the findings from pilot studies need to be confirmed in independent large cohort of samples (clinical validity) and finally the benefit of using the biomarker/biosignature in the clinical decision-making setting should be clearly demonstrated as well as its favourable medico-economic profile. Only after this process, a newly discovered biomarker can pretend to reach the routine clinical use.[109–113]

Preanalytic/analytic validity

In the preanalytic step, attention should be paid to validate sample-related factors and patient-related factors. Standardiza-tion of specimen collection and storage is crucial to reach

1812 | Chem. Soc. Rev., 2016, 45, 1803–1818

This journal is © The Royal Society of Chemistry 2016

experimental reproducibility not only in an individual laboratory but also between different laboratories. In addition, investigators should be aware of the risks of contamination during sample handling. In a recent SERS study,[49] EDTA, citrate and Li-Heparin used as anticoagulants for plasma collection have been shown to exhibit confounding peaks. When using filtered plasma (with a 3 kDa cut-off), contrary to EDTA and citrate, Li-Heparin was filtered out and no longer interfered with the spectral information. FTIR studies have shown that EDTA and citrate spectral contributions can be circumvented using dialysed plasma. In contrast, no interference in FTIR spectra was observed when directly analysing plasma from Li-Heparin tubes.[114] Due to these limitations, serum is often preferred to plasma in spectroscopic analysis. Factors related to patients are of paramount importance to limit the risk of false positives. Inappropriate selection of case patients and control subjects is a common pitfall in spectroscopic studies as widely reported in other high throughput technologies.[108,115] When comparison groups are not matched for example for age, sex and physical conditions such as hormonal status or pathologies other than the disease of interest, the results may be biased and differences identified between groups may be linked to these confounding factors rather than to the disease of interest.[109]

Analytic validity includes the technical aspects of the biomarker assessment. In the field of vibrational spectroscopy, the interaction of light with biological molecules is subject to a certain number of drawbacks which should be overcome to meet the criteria of accuracy, reproducibility and robustness.

The most common protocol for spectral analysis of biofluids is the drying of drop deposits. A shortcoming of this method is the heterogeneous drop deposition characterized by the well-known coffee-ring effect, due to the migration of macromolecules towards the periphery of the drop.[116–118] In order to clarify the dynamics of such deposition, Esmonde-White et al. used both imaging and Raman spectroscopy to demonstrate that substrate and fluid concentration have a profound effect on dried drop morphology. They showed that the substrate did not affect the chemical composition within the outer ring of the drop whereas the macromolecular concentration has an impact on the spatial distribution of proteins.[119] Using HT-FTIR, Lovergne et al. have recently confirmed the impact of serum dilution on the deposition pattern as illustrated in Fig. 10.[120] Without dilution, serum spectra were saturated, due to the acquisition in the transmission sampling mode. The 3-fold dilution was shown to be the most suitable for spectral analysis with a good reproducibility and absorbance intensity. The signal/noise ratio was degraded with higher fold dilutions which precludes the analysis of molecules present at a low concentration in the serum. The heterogeneous deposition of macromolecules in the outer ring should be taken into account when using mode point spectroscopic assessment. It has been reported that this issue can be overcome by averaging spectra taken at different points of the outer ring.[121] Another possibility to avoid the coffee-ring effect is to perform an analysis on a film composed of an array of reduced-size dry drops each formed from 200 pL of serum.[122] The strict control of experimental
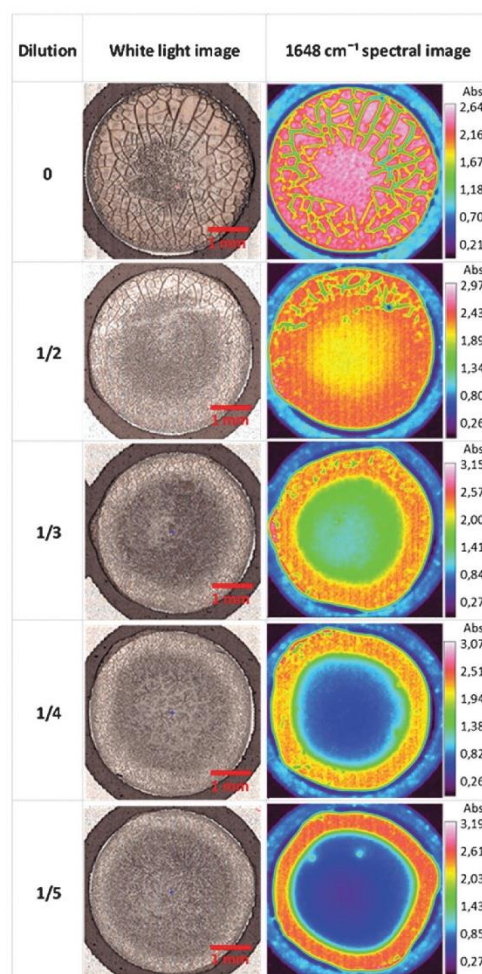
Fig. 10 Analysis of dried serum drops and coffee ring effect with different dilutions: white light images (left) and chemical images constructed on the amide I protein band (right).

parameters of drop deposition appears as a major prerequisite to obtain reproducible results.[119] This may be obtained at best by an automated sampling approach as described by Ollesch et al. Using this approach, these authors have reported a higher reproducibility of spectral data compared to a non-automatic sampling.[122]

ATR-FTIR spectroscopy has been shown to be an interesting approach for the analysis of biofluids as samples can be directly applied onto the ATR crystal without any dilution. However, currently there is no automated device available so that spectral acquisition is time consuming, about 9 times longer than with automated HT-FTIR spectroscopy.[120] The lack of automation is

This journal is © The Royal Society of Chemistry 2016

Chem. Soc. Rev., 2016, 45, 1803–1818 | 1813

213

a limiting factor for the transposition of ATR-FTIR spectroscopy into a high-throughput clinical application.[123] This may also be possible when using a high throughput source such as a QCL during a DF-IR approach. However, for limited patient cases, in a hand-held mode it offers advantages of ease of use and ease of sample preparation with no modification/adulteration of the sample. Identically, Raman spectroscopy is also of great interest for biofluid spectroscopy particularly due to developments in hand held technology and immersion Raman which could enable hand held analysis of "wet" serum, negating the need for a drying step.

The technical standardization of spectral acquisition makes sense if reproducible results can be obtained in different laboratories. This external validation is essential on the way towards clinical validity. The inter-instrument transferability is also a challenge that needs to be faced. Finally, the need for automated instruments underline the necessity of a close collaboration between research scientists, clinical practitioners and industrial partners in order to optimize currently available products according to a specific biomedical purpose.[1]

Beside the need of standardized spectral acquisition, there is also a need to validate the design of pilot studies including the chemometric analysis. Proof-of-concept studies raise the question of appropriate selection of case patients and controls as discussed below and also the question of sample size. In contrast with classical statistics, there is no simple method to calculate sample size in biospectroscopic studies. However, Beleites *et al.* have proposed in a recent report to use learning curves to determine the appropriate sample size needed to build good classifiers with specified performances.[124] When the number of patients is too limited to divide the population in one training set and one independent validation set, cross validation methods should be used to avoid the high risk of overfitting.[125]

Clinical validity

The next step after the phase of pre-analytic/analytic validation is to confirm the diagnostic performance of the biomarker on an independent population of a large number of patients. This means large multicenter randomized control trials where the sensitivity and the specificity of the putative biomarker may be evaluated against the gold standard diagnostic/screening procedure. These studies, particularly the criteria to include case patients and controls, should be carefully designed to demonstrate whether the biomarker is applicable to its specific purpose which may be screening, differential diagnosis, prognosis, treatment response prediction or monitoring of a disease (Fig. 4).

A common mistake is to validate a marker in the diagnostic setting of a disease and then to extrapolate its performance to the screening context. Candidate biomarkers are tested in pilot studies performed in small numbers of patients with patent disease already diagnosed using golden standard methods. It is crucial to validate the value of these markers in the screening context *i.e.* for early diagnosis in large populations of patients at risk of the disease. The biomarker sensitivity and specificity in the screening target population are usually much lower than in patients with patent disease. In the context of population screening, high specificity is of paramount importance to avoid false positive results, which means patients will be subject to additional diagnostic procedures, potentially invasive and costly for the society. This underlines the necessity of selecting case patients and control subjects according to the clinical setting where the biomarker is intended to be used.[109]

A methodology to avoid patient selection bias in screening studies has been proposed by Pepe *et al.*[109,111,126] In the so-called PRoBE study design, samples are collected prospectively in a cohort of patients before the knowledge of the final diagnosis. Once the outcome data becomes available and the diagnosis established, the sample cohort can be used retrospectively by randomly selecting cases and controls. This methodology is promoted by the research consortium "Early Detection Research Network" from the National Cancer Institute to establish specimen reference sets. It has proved efficient for rapid evaluation of potential biomarkers.[110]

Clinical utility

A crucial point in the process of biomarker validation before its adoption in routine clinical practice is to demonstrate its clinical decision-making usefulness at an acceptable cost for the society.[109] This means that the positive and negative predictive values of the biomarker should be evaluated in the "real life" patient population since these indicators are dependent on the prevalence of the disease of interest. The difference between clinical validity and clinical utility is illustrated by the debate about the usefulness of Prostatic Specific Antigen (PSA)-based screening program. It is well established that PSA-based screening programs significantly increase the detection of prostate cancer at an early stage.[127] However, there is also evidence that PSA-based screening carries a high risk of over-diagnosis leading to overtreatment in a significant number of men with early cancer that will never become symptomatic during their life time.[128] Whether the benefits of early detection of asymptomatic prostate cancer outweigh the harms related to over-diagnosis and over-treatment is highly controversial. There is no consensus regarding the clinical relevance of a PSA-based screening program.[129] This emphasizes that, in addition to its diagnostic performance, the biomarker clinical utility has to be demonstrated before its clinical implementation. The clinical utility refers to the balance of benefits to harms and the medicoeconomic evaluation. For this purpose, a validation study should be performed in a large number of unselected patients with clinical endpoints clearly defined to demonstrate the benefit of using a biomarker including quality of life for the patient and socioeconomic aspects for the society.[109]

Conclusion

The difficulty in translating biomedical spectroscopy to the clinic is fundamentally based on the fact that after over more than two decades of research, not enough has been done to

214

fully understand the accuracy of these tests with appropriate considerations applied to control groups and limitations of the clinical environment. In addition there is a need to perform large-scale studies to evaluate the spectroscopic tests' efficacy within the clinic. These approaches would enable this technology to be acceptable to the medical community through a "hearts and minds" approach. The particular requirements of a clinical spectrometer should be implemented for different clinical settings. Its instrumental requirements (*e.g.* detector sensitivity and source throughput) and how accurately it can diagnose disease or perform treatment monitoring must be validated.

This review has highlighted the potential of biomedical vibrational spectroscopy to analyse biofluids. However, care should be taken for biofluid spectroscopy not to suffer from the identified pitfalls. As the field of biofluid spectroscopy is further researched, a lot of commitment from different stakeholders (researchers, clinicians, and instrument manufacturers) will be necessary to demonstrate its real potential as a rapid, novel, and robust technology to pinpoint "spectral biomarkers/ signatures" that can be useful for diagnostic purposes and to predict clinical outcomes, with the promise that the test can be done periodically at low cost for monitoring care.

The initiatives *via* current networks like the EPSRC CLIRSPEC (http://clirspec.org/), the Raman4Clinics European COST action (http://www.raman4clinics.eu/raman4clinics-a-european-cost-action/) and the 1st International Society for Clinical Spectroscopy (CLIRSPEC) are currently gearing research, facilities and communities in the clinical spectroscopy arena to achieve these objectives.

6 K. Papamarkakis, B. Bird, J. M. Schubert, M. Miljković, R. Wein, K. Bedrossian, N. Laver and M. Diem, *Lab. Invest.*, 2010, **90**, 589–598.
7 C. Kendall, N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett and H. Barr, *J. Pathol.*, 2003, **200**, 602–609.
8 P. Lasch, M. Beekes, J. Schmitt and D. Naumann, *Anal. Bioanal. Chem.*, 2007, **387**, 1791–1800.
9 P. Lasch, J. Schmitt, M. Beekes, T. Udelhoven, M. Eiden, H. Fabian, W. Petrich and D. Naumann, *Anal. Chem.*, 2003, **75**, 6673–6678.
10 J. Schmitt, M. Beekes, A. Brauer, T. Udelhoven, P. Lasch and D. Naumann, *Anal. Chem.*, 2002, **74**, 3865–3868.
11 E. F. Petricoin, C. Belluco, R. P. Araujo and L. A. Liotta, *Nat. Rev. Cancer*, 2006, **6**, 961–967.
12 C. Pierrakos and J.-L. Vincent, *Crit. Care*, 2010, **14**, R15.
13 D. Qi and A. J. Berger, *Appl. Opt.*, 2007, **46**, 1726–1734.
14 J. M. Reyes-Goddard, H. Barr and N. Stone, *Photodiagn. Photodyn. Ther.*, 2005, **2**, 223–233.
15 D. Rohleder, W. Kiefer and W. Petrich, *Analyst*, 2004, **129**, 906–911.
16 R. A. Shaw, S. Low-Ying, A. Man, K.-Z. Liu, C. Mansfield, C. B. Rileg and M. Vijarnsorn, *Biomedical Vibrational Spectroscopy*, John Wiley and Sons, Inc, Hoboken, NJ, 2008, pp. 79–103.
17 C. N. Banwell and E. M. McCash, *Fundamentals of molecular spectroscopy*, McGraw-Hill, London, 1983.
18 P. Dumas, G. D. Sockalingum and J. Sule-Suso, *Trends Biotechnol.*, 2007, **25**, 40–44.
19 A. Barth and P. I. Haris, *Biological and biomedical infrared spectroscopy*, IOS

34 G. Clemens, B. Bird, M. Weida, J. Rowletteb and M. J. Bakera, *Spectrosc. Eur.*, 2014, **26**, 14–19.

35 K. Yeh, S. Kenkel, J.-N. Liu and R. Bhargava, *Anal. Chem.*,

37 F. S. Parker, in *Applications of infrared, Raman, and reso-*

59 N. Stone, R. Baker, K. Rogers, A. W. Parker and P. Matousek, *Analyst*, 2007, **132**, 899–905.

60 M. J. Baker, *Special Issue: Photonic Biofluid Diagnostics*,

D. F. Hoth, J. A. Oates, C. C. Peck, R. T. Schooley,

B. A. Spilker and J. Woodcock, *Clin. Pharmacol. The*

123 C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N. W. Clarke and P. Gardner, *J. Biophotonics*, 2014, 7, 180–188.

124 C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft and J. Popp, *Anal. Chim. Acta*, 2013, **760**, 25–33.

125 D. Pérez-Guaita, J. Kuligowski, S. Garrigues, G. Quintás and B. R. Wood, *Analyst*, 2014, **140**, 2422.

126 M. S. Pepe, Z. Feng, H. Janes, P. M. Bossuyt and J. D. Potter, *J. Natl. Cancer Inst.*, 2008, **100**, 1432–1438.

127 F. H. Schröder, J. Hugosson, M. J. Roobol, T. L. Tammela, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja and M. Zappa, *N. Engl. J. Med.*, 2009, **360**, 1320–1328.

128 V. A. Moyer, *Ann. Intern. Med.*, 2012, **157**, 120–134.

129 J. Cuzick, M. A. Thorat, G. Andriole, O. W. Brawley, P. H. Brown, Z. Culig, R. A. Eeles, L. G. Ford, F. C. Hamdy and L. Holmberg, *Lancet Oncol.*, 2014, **15**, e484–e492.

218

Appendix IV. List of publications

- Investigating Optimum Sample Preparation for Spectroscopic Serum Diagnostics, Lila Lovergne, Graeme Clemens, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, and Matthew J. Baker, *Analytical Methods,* 2015, **7**, 7140-7149, (inside front cover).

- Biofluid infrared spectro-diagnostics: pre-analytical considerations for clinical applications, L. Lovergne, P. Bouzy, V. Untereiner, R. Garnotel, M.J. Baker, G. Thiéfin and G.D. Sockalingum, *Faraday Discussions*, 2016, **187**, 521-537.

- Developing and Understanding Biofluid Vibrational Spectroscopy: A Critical Review, Matthew J. Baker, Shawn R. Hussain*, Lila Lovergne*, Valérie Untereiner, Caryn Hughes, Roman A. Lukaszewski, Gérard Thiéfin, and Ganesh D. Sockalingum, *Chemical Society Reviews*, 2016, **45**, 1803-1818, *equal contribution, (front cover).

Manuscripts in preparation

- Impact of the "coffee ring" effect on serum vibrational spectroscopy, Lila Lovergne, Ganesh D Sockalingum and Matthew J Baker.

- Investigating pre-analytical requirements for infrared spectroscopy of blood-derived products, Lila Lovergne, Jean Lovergne, Pascaline Bouzy, Valérie Untereiner, Marc Offroy, Roselyne Garnotel, Gérard Thiéfin, Matthew J. Baker, and Ganesh D. Sockalingum.

Appendix V. List of presentations

- Rapid Spectroscopic and Spectrometric Pre-symptomatic Diagnosis of Sepsis, oral communication at the DGA PhD Inauguration Day, Paris, France, 30th January 2014.

- Rapid Spectroscopic and Spectrometric Pre-symptomatic Diagnosis of Sepsis, Lila Lovergne, Graeme Clemens, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, poster presentation at the 2nd DGA/DSTL PhD conference, Bagneux, France, 3rd July 2014.

- Rapid Pre-symptomatic Diagnosis of Sepsis: A Feasibility Study by Infrared Spectroscopy; Lila Lovergne, Graeme Clemens, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, poster presentation, SciX conference, Reno, Nevada, USA, 30th September 2014.

- Infrared Spectroscopy: A Novel Approach for the Rapid Pre-symptomatic Diagnosis of Sepsis, Lila Lovergne, Graeme Clemens, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, poster presentation, CLIRSPEC conference, Exeter, UK, 20th April 2015.

- Diagnostic rapide et pré-symptomatique de la septicémie par spectroscopie infrarouge, Lila Lovergne, Graeme Clemens, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, poster presentation, GFSV conference, Reims, France, 17th-18th June 2015.

- Novel Healthcare Technologies: Rapid Pre-symptomatic Diagnosis of Sepsis, Lila Lovergne, Graeme Clemens, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, oral communication and poster presentation at the 3rd DGA/DSTL PhD conference, Salisbury, Wiltshire, UK, 1st July 2015.

- Rapid Pre-symptomatic Diagnosis of Sepsis by Infrared Spectroscopy, Lila Lovergne, Graeme Clemens, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, poster presentation (3rd poster prize), CLIRSPEC Summer School, Windermere, Cumbria, UK, 7th July 2015.

- Developing Serum Based Infrared Spectroscopic Diagnostics: Optimising Sample Preparation and Sampling Mode, Lila Lovergne, Valérie Untereiner,

Gianfelice Cinque, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, poster presentation, workshop "FT-IR Spectroscopy in Microbiological and Medical Diagnostics", Berlin, Germany, 15th-16th October 2015.

- Novel Healthcare Technologies: Rapid Pre-symptomatic Diagnosis of Sepsis, Lila Lovergne, Valérie Untereiner, Roman A. Lukaszewski, Ganesh D. Sockalingum, Matthew J. Baker, oral communication and poster presentation at the 4th DGA/DSTL PhD conference in Paris, France, 21st June 2016.

- Early Diagnosis of Sepsis by Infrared Spectroscopy, L. Lovergne, G. D. Sockalingum, V. Untereiner, R. A. Lukaszewski, M. J. Baker, poster presentation, SPEC conference, Montreal, Canada, 26th-30th June 2016.

- Sepsis: Serum Spectroscopic Sensing, L. Lovergne, G. D. Sockalingum, V. Untereiner, R. A. Lukaszewski, M. J. Baker, poster presentation, Maritime Enterprise Innovation Conference Scotland 2016, Glasgow, UK, 8th December 2016.

- Combining Infrared Spectroscopy and Chemometrics for Developing Serum Based Diagnostic for Sepsis, L. Lovergne, G. D. Sockalingum, V. Untereiner, R. A. Lukaszewski, M. J. Baker, poster presentation, Chimiométrie XVIII conference, Paris, France, 30-31 January & 1st February 2017.

- Developing a Serum Based Infrared Spectro-Diagnostic for Sepsis, L. Lovergne, G. D. Sockalingum, V. Untereiner, R. A. Lukaszewski, M. J. Baker, poster presentation, CLIRCON 2017, Manchester, UK, 2nd-5th April 2017.

- Developing a Serum Based Infrared Spectro-Diagnostic for Sepsis, L. Lovergne, G. D. Sockalingum, V. Untereiner, R. A. Lukaszewski, M. J. Baker, oral communication, CAMO event, Glasgow, UK, 31st May-1st June 2017.

- Developing a Serum Based Infrared Spectro-Diagnostic for Sepsis, L. Lovergne, G. D. Sockalingum, V. Untereiner, R. A. Lukaszewski, M. J. Baker, poster presentation, at the 5th DGA/DSTL PhD conference, Portsmouth, UK, 6th July 2017.