Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy- A new diagnostic tool for diagnosing pathogens and analytes in blood and sera

Supti Roy
BSc. (Honours), MSc.

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in 2019
School of Chemistry

Copyright notice

Abstract

The thesis focused on the diagnosis of blood-borne diseases include malaria and hepatitis employing Attenuated Total Reflection Fourier transform infrared (ATR-FTIR) spectroscopy, Raman spectroscopy, synchrotron FTIR spectroscopy and atomic force microscopy-based infrared spectroscopy (AFM-IR). Multivariate models were built for pathogens and analytes detection in blood and sera. Furthermore, a host-virus interaction was monitored in hepatoblastoma cell lines using a combination of synchrotron FTIR and Raman micro-spectroscopy.

Chapter 1 introduces the literature background of malaria and hepatitis along with the key concepts of FTIR spectroscopy and the application in terms of diagnosis.

Chapter 2 highlights the techniques that were employed to analyse sera and cell samples (infected and controls). This chapter overview details the working principles, related theory, instrumentation and data processing for all techniques and provides the reader with an insight into the capability of the analytical methods in characterizing biological samples.

Chapter 3 outlines a new approach for blood sample preparation on glass slide along with ATR-FTIR based quantification of *Plasmodium falciparum* parasites, glucose and urea simultaneously in blood. The Partial Least Squares Discriminant Analysis (PLS-DA) model was able to detect >0.5% parasitemia with sensitivity of 98% and specificity of 70%. The lower specificity was due to the lower number of control samples in the model. The Root Mean Square Error of Cross Validation for the Partial Least Squares Regression models were 0.58%, 22.70 mg/dL and 30.48 mg/dL for parasite detection (0-5%), urea (0-250 mg/dL) and glucose (0-400 mg/dL), respectively.

Chapter 4 focuses on the application of ATR-FTIR technology for developing point-of-care testing for hepatitis due to the infection of hepatitis B and hepatitis C. Using basic blood separation techniques and multivariate data analysis I developed powerful models to identify specific disease markers. These markers were found to be associated with the virus particles or due to the immune response. The prediction capability of PLS-DA models for HBV vs controls and HCV vs controls were >90% for the Area Under the Receiver Operator Curve with an error value of 11.20% and 11.30%, respectively. The separation was predominantly based on Immunoglobulin (Ig), glucose and lipid components. An AFM-IR confirmed band at 1093 cm^{-1} was assigned to the hepatitis B surface antigen (HBsAg).

Chapter 5 investigates the response of the human hepatoblastoma cell line to HBV infection using IR spectroscopy and Raman micro-spectroscopy. The aim of the study is to assess the cell's physiological behaviour to HBV infection and to determine the contribution of the intracellular viral particles to the spectrum. In the previous chapters I have shown the ability of the approach identify the presence of HBV infection using ATR-FTIR combined with multivariate data analysis in human serum samples (infected vs uninfected). The next step is to understand the mechanisms leading to the spectral changes, and gain a holistic understanding of the HBV infection (together with its consequences) and explore in detail subcellular level changes in response to the virus. Utilizing a combination of Raman micro-spectroscopy and synchrotron FTIR micro-spectroscopy it was possible to identify and locate lipid deposits in virally transfected and control cell lines. The results showed lower lipid accumulation in the HepG2.2.15 cell line compared to the control cell line upon transfection with HBV complete DNA.

Chapter 6 suggests the future scopes, which can be performed for in detailed investigation of malaria and hepatitis by applying other spectroscopic analysis.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: SUPTI ROY

Date:

Publications during enrolment

(1) Roy, S.; Perez-Guaita, D.; Andrew, D. W.; Richards, J. S.; McNaughton, D.; Heraud, P.; Wood, B. R. Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide. *Analytical Chemistry* **2017**, *89* (10), 5238–5245.

(2) Roy, S.; Perez-Guaita, D.; Bowden, S.; Heraud, P.; Wood, B. R. Spectroscopy Goes Viral: Diagnosis of Hepatitis B and C Virus Infection from Human Sera Using ATR-FTIR Spectroscopy. *Clinical Spectroscopy* **2020**, *1*, 100001.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes (*1)* original papers published in peer-reviewed journals and (*2)* unpublished publications. The core theme of the thesis is Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy- A new diagnostic tool for diagnosing pathogens and analytes in blood and sera. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the School of Chemistry under the supervision of Prof. Bayden R. Wood, Dr. Philip Heraud and Prof. Scott Bowden from Doherty Institute.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of *Chapter 3, 4 and 5* my contribution to the work involved the following:

Thesis Chapter	**Publication Title**	**Status** (published, in press, accepted or returned for revision)	**Nature and % of student contribution**	**Co-author name(s) Nature and % of Co-author's contribution***	**Co-author(s), Monash student Y/N***
3	Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide	Published	60%. Designed and carried out all experiments, collecting and analysing data and writing the first draft	1) David Perez-Guaita (Key idea, Data analysis, Experimental, editing draft20%) 2) Dean W. Andrew (Techniques 2%) 3) Jack S. Richards (Techniques 2%) 4) Don McNaughton (Editing draft 2%) 5) Philip Heraud (Key idea, editing draft 5%) 6) Bayden R. Wood (Key idea, editing draft 9%)	No No No No No No
4	Spectroscopy goes Viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy	Published	70%. Designed and carried out all experiments, collecting and analysing data and writing the first draft	1) David Perez-Guaita (Key idea, Experimental, editing draft10%) 2) Scott Bowden (Key idea, editing draft 5%) 3) Philip Heraud (Key idea, editing draft 5%) 4) Bayden R. Wood (Key idea, editing draft 10%)	No No No No
5	Monitoring the infection of a human hepatoma cell line by hepatitis B virus using synchrotron FTIR micro-spectroscopy and Raman spectroscopy	Submitted	70%. Designed and carried out all experiments, collecting and analysing data and writing the first draft	1) Kamila Kochan (Key idea, Experimental, editing draft10%) 2) Danni Colledge (Experimental, editing draft 8%) 3) Diana E. Bedolla (Experimental, editing draft 2%) 4) Scott Bowden (editing draft 1%) 5) Philip Heraud (editing draft 2%) 6) Bayden R. Wood (Key idea, editing draft 7%)	No No No No No No

I have/have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: **Date:**

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author, I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: **Date:**

Acknowledgements

I would like to express my deepest acknowledgement to Prof. Bayden Wood, Dr. Philip Heraud, Em. Prof. Donald McNaughton and Prof. Scott Bowden for their guidance and supervision I received throughout my Ph.D. journey. During my journey, their knowledgeable advice and expert opinions always helped me to drive into the next stage. The thesis was possible with their cordial and continuous support.

I am also grateful to Prof Alan Chaffee, Prof. Philip Marriott, Dr. David Turner for their valuable feedback and support during the panel meeting.

I am thankful to Dr. David Perez-Guaita for providing his expertise in data analysis and training me in the use of Matlab. I am grateful to Dr. Kamila Kochan for her tremendous support to train me with Raman spectroscopy and data analysis. I would like to thank Mr. Finlay Shanks, laser safety officer for his technical support with instruments. I am also thankful to all my past and present colleagues from Wood group, especially Dr. Aazam Khoshmanesh, Dr. Khansa Al-Jorani, Mrs. Diana E. Bedolla, Miguela Martin, Dale Christensen, John, Savi, Zack and Elizabeth for their wonderful support in many ways. I am thankful to Mrs. Danni Colledge from VIDRL, for her continuous support with the cell culture experiments.

I would also take the opportunity to thank Mrs. Anna Severin, Sarah Williams and Dr. David Turner for their prompt responses and valuable suggestions regarding administrative issues.

Lastly, I am grateful sincerely to my beloved husband Mr. Ponsubbiah S. Nadar and I believe it would not be so smooth to submit my dissertation without his support and care. In addition, I want to give thanks to my parents (Mr. and Mrs. Roy) and my brother (Uttam Kumar Roy) who were always nice and supportive of me.

Table of contents

Chapter 1. Introduction

1.1 Background

Blood borne diseases including viral, bacterial and parasitic infections have an enormous impact on mortality, and in some cases can be very difficult if not impossible to treat. Human Immunodeficiency Virus (HIV) is the causative agent for Acquired Immune Deficiency Syndrome (AIDS), whereas hepatitis B (HBV) and hepatitis C (HCV) viruses are responsible for causing hepatitis in humans. These are generally transmitted by direct contact with infected blood and body fluids through injured skin or splashes on mucosal surfaces. Also, unscreened blood transfusion, intravenous drug abuse and unprotected sexual intercourse are the main vectors for infection. Malaria, a parasitic disease, is transferred via a mosquito bite and contaminated blood transfusion [1–3]. Malaria alongside HBV and HCV are the most devastating diseases (cumulative mortality accounts over 1.7 million per year, as reported by WHO) to afflict humanity and consequently new, rapid, inexpensive and highly sensitive techniques are urgently required, especially in the developing world where these diseases cause most deaths.

This investigation aims to demonstrate potential for ATR-FTIR as a portable state-of-the-art technology with the capability to instantaneously diagnose blood borne infections including those caused by HBV, HCV and malaria at the point-of-care (POC). This will be achieved for a fraction of current costs and with unprecedented levels of sensitivity and speed, in remote field locations.

1.2 Infection and Impact

1.2.1 HBV

HBV belongs to the family of viruses called Hepadnaviridae. It is an enveloped virus with a relaxed genome consisting of a partially double-stranded DNA molecule of approximately 3.2 kb. In spite of being a DNA virus, HBV genome replicates through RNA intermediates[4,5]. Its compact genome, represents four overlapping ORFs (open reading frames) and encodes for seven viral proteins. Among those proteins, 3 envelope proteins are required to form the HBV surface antigen (HBsAg). Others are the core proteins HBcAg, the e antigen (HBeAg), theX protein and the viral reverse transcriptase, which also possess the activity of DNA polymerase and RNaseH[5].

Hepatocytes are the primary cellular location for HBV replication. After the entry of the virus into a host cell, it uncoats in the cytoplasm and transports its relaxed circular DNA to the nucleus. In the nucleus, both host and viral polymerase combine to repair the relaxed genome into an enclosed double-stranded cccDNA (covalently closed circular DNA). Transcription of all viral mRNA occurs from the cccDNA. There are also some pre-genomic viral RNAs, which can function as templates for reverse transcription and are responsible for synthesizing the viral core proteins and polymerase. The 3 subgenomic mRNA translates the viral envelope and X proteins during this time. In the cytoplasm, translation of viral proteins occurs after transportation of viral messenger RNA to the cytoplasm. The assembly of viral particles and replication also occurs at this stage[6,7].

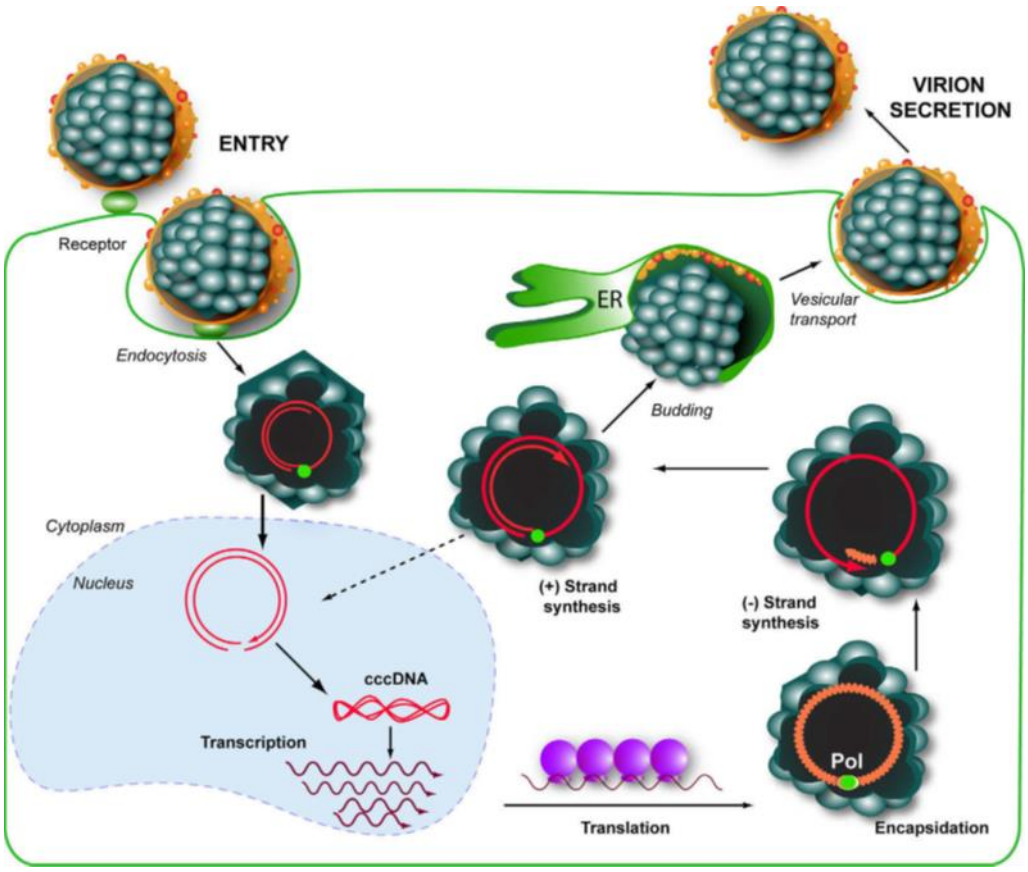


Figure 1: Schematic representation of HBV viral life cycle[8]

Approximately, 80% of hepatocellular carcinoma cases are directly related to chronic infection due to HBV and HCV[9]. Globally, hepatocellular carcinoma (HCC) or liver cancer is ranked sixth in prevalent cancer and is the second most common reason for deaths resulting from cancer (approximately 0.7 million deaths occur per annum). Worldwide, around 90% of primary liver cancer cases are HCC[10,11]. Liver cirrhosis is also an emerging global health burden with more than one million cases accounting for 2% of deaths in the world in 2010[12,13]. It has been reported that liver cirrhosis precedes the majority of HCC cases[14,15].

The association between chronic HBV infection and developing hepatocellular carcinoma was first identified in early 1970[16]. Later studies indicated that the majority (80%) of patients with hepatocellular carcinoma were serologically positive for HBsAg (HBV surface antigen) and more than 90% of patients with HCC carry the anti-HBc (HBV core antigen) in their serum[17]. In recent years, the global burden of viral infection by HBV has increased significantly. Worldwide, there are more than 240 million people who are living with HBV chronic infection[18], and yearly up to one million individuals die from HBV-associated consequences including liver cirrhosis or cancer[19]. Studies suggest that, between 1965 and 2013, Africa (prevalence rate estimated at 8.83%) and the Western Pacific Region (prevalence rate 5.26%) were the most prevalent areas which reported HBsAg due to HBV infection and the number of infected persons was 75.6 million and 95.3 million, respectively. Moreover, the Africa and Western Pacific regions contributed around 70% of the total (686000 deaths) global burden of infection in 2013 because of HBV infection[20,21].

In developing countries, childrens are prone to HBV infection via horizontal and perinatal transmission routes. While in developed countries, infections are generally common in young adults because of promiscuous sexual behavior and injection drug use[22–25]. Lavanchy *et al.*[26] identified that people who develop HBV infection in their childhood, have a 25% chance of developing cirrhosis or primary liver cancer in their adult lives, however a complete understanding of the development process is not known.

According to a report from China[27], the HBV vaccine works well in neonates and helps to decrease the viral load significantly. However, considering that the vast majority of the world population chronically infected with HBV and remains untreated, it appears there will be a massive increase in liver cirrhosis and hepatocellular carcinoma within the next 10 years[9]. Thus, there is an urgent need for action.

In 2005, the WHO of the Western Pacific Region, set a regional goal to combat HBV by 2012, in an effort to prevent the global spread of both HBV and HCV infection. In the 67th World Health Assembly organized by WHO in 2014, the resolution was passed that prevention of viral hepatitis through viral hepatitis diagnosis, control and treatment was an urgent priority[28].

1.2.2 HCV

HCV is a member of the *Hepacivirus* genus of the Flaviviridae family. It is also an enveloped virus that contains single-stranded (+)RNA genome of about ~9.6 kb in length. The genome possesses a large open reading frame (ORF) which is flanked by untranslated regions (UTRs) at both 5’ and 3’ ends of the RNA molecule[29]. The ORF encodes a polyprotein precursor, which

goes through several co and post-translational processing by both cellular and viral proteinases and is cleaved into individual viral proteins. The HCV proteins include structural proteins (core protein and E1, E2 envelope glycoproteins and p7) as well as non-structural proteins (NS2 protein, NS3-4A complex, NS4B protein, NS5A protein and NS5B protein)[30]. CD81, a host-cell surface protein, serves as HCV receptor for the attachment and viral entry. The soluble form of HCV envelope protein E2 is necessary for the identification of surface proteins like CD81 which are involved in the HCV attachment and entrance procedure[31].

A 5' non-translated region of HCV genome, internal ribosome entry site (IRES) mediates the translation of viral proteins in the cytoplasm and yields a polyprotein precursor[32]. Along with other viral proteins, NS5B-RNA dependent RNA-polymerase (RdRp) also produces a polyprotein precursor during the translational process. Just like other positive RNA viruses, HCV replication starts by using its own (+) RNA genome as a template to synthesize (-) RNA. Then the same NS5B RdRp catalyzes the 2nd step to yield positive stranded RNA from the negative one. This time, the new RNA replicons (positive RNA progeny) are 5 to 10 times more powerful in transcription level[30].

HCV virion comprises of an outer envelope (made up of a single lipid membrane along with some envelope proteins) and nucleocapsid. The virion assembly is assumed to initiate in the cytosolic part of the endoplasmic reticulum (ER) membrane during the structural core protein synthesis by the action of signal peptidase in the ER. When the core proteins are come into contact with the progeny genome, they form the viral capsid. After maturation in the ER, the viral particles are assembled and released as an individual progeny infectious virion on the luminal side of the ER membrane. The ER plays a vital role in HCV replication and packaging as a specialized site known as the membranous web, which is responsible for RNA replication of the HCV virus[33–36].

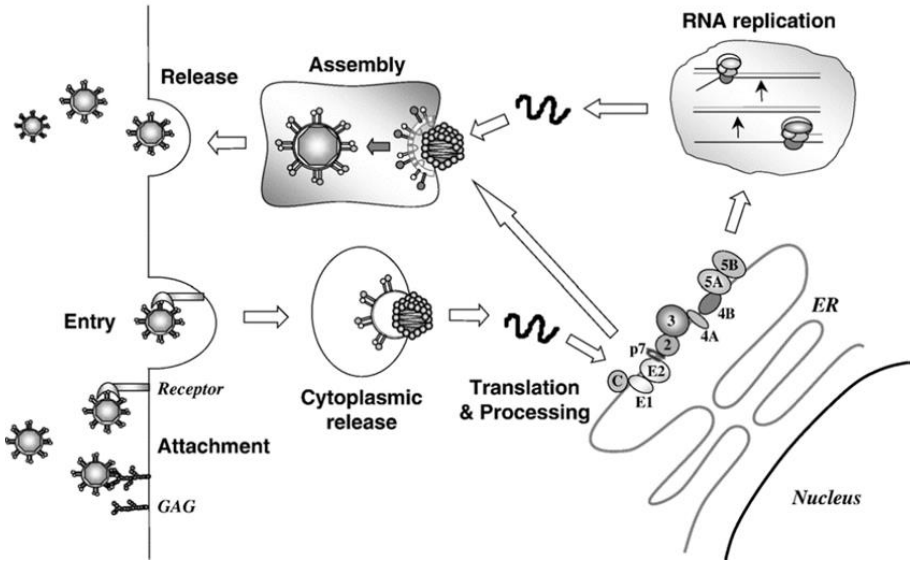


Figure 2: HCV virus attachment, replication, and release[30]

Hepatitis C virus has threatened many countries worldwide by causing over 500,000 deaths through HCV associated liver illness. Reports suggest that around 185 million patients are living with HCV chronic infection[37,38]. In developing HCC, persistent HCV infection has proven as a high-risk factor. The risk of HCC is 15 to 20 times higher for patients with HCV in their serum. Though, several individuals infected by HCV have not seen effects for over 25-30 years, studies have been forcasted that about 15-35% of these individuals may eventually develop liver cirrhosis and then HCC[39–41].

In the 1970's most of the blood transfusable diseases occured either from HAV or HBV. A new disease form designated as "non-A, non-B" hepatitis[39] was speculated thereafter in 1989. HCV was discovered as a virus with distinguishing features[42] compared to HBV and HBV but it was considered at that time to be limited to small numbers of people who were intravenous drug users or people that received a contaminated blood product. Twenty years later the scenario changed entirely and now HCV infection is considered as a leading public heath problem in the world[43].

HCV has become the major pathogenic virus for causing HCC. Using a mouse transgenic model Koike *et al.*[44] demonstrated that HCV can directly cause liver carcinogenesis leading to chronic inflammation, cell proliferation, and cell death. The core protein of the virus is responsible for the oncogenic characteristics[44]. Studies have also shown that HCV can induce active free radical production, which is an integral feature of the protein core of the virus[45].

HCV infection is not limited to liver disease and can cause extrahepatic clinical manifestations, which make the host-virus interaction more complex. Roughly 74% of individuals with HCV

infections also acquired extrahepatic manifestations[46]. These manifestations are mostly associated with the autoimmune disorders, which can be lymphomas (HCV-related lymphotropism), palpable purpura and immune-complex deposition[46,47]. Most recent studies have reported that extra-hepatic disorders are also involved in cardiovascular problems, renal tumor and disease related to the central nervous system[48].

1.2.3 Plasmodium

Five strains of a unicellular parasite - *Plasmodium* are the causal agents for malarial infection in humans. These include *Plasmodium* falciparum, *Plasmodium* vivax, *Plasmodium* ovale, *Plasmodium* malariae and *Plasmodium* knowlesi. 99.7% cases of malaria infection in sub-Saharan Africa were due to *Plasmodium* falciparum and 74% malarial infection cases in tropical region of South America is caused by *Plasmodium* vivax[49,50].

A schematic representation of Plasmodium spp. life cycle is presented in Figure 3. The *Plasmodium* parasites use female Anopheles mosquitoes as their vector for spreading malaria. These parasites develop into infectious and motile sporozoites within the mosquito and then transfer to human host via mosquito bite. The sporozoites travel through the bloodstream, invade the hepatocytes and mature into schizonts (asymptomatic), which eventually produce tens of thousands merozoites. A fraction of merozoites invades, matures and replicates in red blood cells. The stages include pathogenic-asexual ring stage, trophozoites and formation of new merozoites. The remaining fraction of the merozoites are differentiated into sexual male and female gametocytes, which are non-infectious but can transmit into a new vector and later on infect a new host.[50,51]

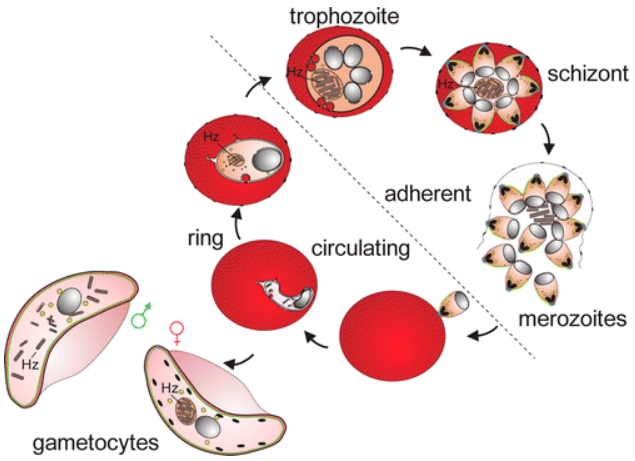


Figure 3: Schematic representation of Plasmodium spp. life cycle[51]

Malaria has infected over 200 million people worldwide and caused around half a million deaths in 2017, and is still a leading cause of mortality. Case studies represents that around 90% of the total infected population were from African sub-Saharan region [49].

1.3 Current Diagnosis

1.3.1 HBV and HCV

Current diagnostic techniques for HCV and HBV infections include serological assays, molecular assays, and rapid diagnostic tool[52]. Serological assays work on antigens and antibodies detections. Several types of serological assays which are commonly used in hepatitis diagnosis are enzyme immune assay (EIA)[53,54], Radioactive immune assay (RIA)[55,56], Enzyme-linked immunosorbent assay (ELISA)[57,58] and immunoblot assay[59]. Molecular assays detect the viral genome (DNA or RNA)[54]. Different types of molecular assays used in HBV and HCV diagnosis are Nucleic Acid Sequence-Based Amplification (NASBA), Transcription-Mediated Amplification (TMA), multiplex and Real-time PCR, reverse transcriptase PCR[60] and loop-mediated isothermal amplification (LAMP)[61,62]. The Rapid Diagnostic Tool (RDT) kit is another quick, cost-effective method for both HBV[63] and HCV[64] diagnosis, but due to low sensitivity (e.g. in HCV detection[64]) and sometimes giving only qualitative and false-positive results make the RDT's not very suitable all time. Moreover, the kits require proper refrigeration for sample storage, making them inappropriate for using at remote locations.

The antigenic part (HBsAg and HBeAg) of HBV virus and antibodies to HBV core antigen (IgG and IgM), surface antigens and envelope proteins which changes according to the stage of infection can be assessed by serological assays[65]. For serological tests for HCV, anti-HCV is the usual target. This allows a deeper understanding of the human response to infection and presents a scenario for treatment.

Acute hepatitis is diagnosed before the clinical manifestations by the presence of surface antigens (HBsAg, a serum marker) which appear in the blood. IgM antibody to hepatitis B core antigen can be detected as soon as the onset of symptoms. The persistent presence of HBsAg in blood for more than 6 months is an indicator of chronic infection[66,67]. Person having anti-HBsAg antibody in their serum after recovery from HBV infection shows a sign of developing immunity against the virus. Despite the presence of anti-HBs, some HBsAg mutant viruses have the ability to actively infect and can elude detection in the HBsAg assay[68].

The presence of HBeAg (another serum marker) in serum is directly related to the viral loading, virus replication and the infectivity status[69]. A strong positive correlation is observed between

HBV DNA titer and HBeAg presence. The DNA level is significantly higher when tested in HBeAg positives than HBeAg negatives[70]. However, there are some HBV mutants (which do not interfere with the replication of the infectious virus) with a pre-core regional mutation in their genome preventing the expression of HBeAg.

During antibody measurements (such as anti-HBc or anti-HCV), weakly positive samples are generally false positive. Further, there is a greater chance to have false-positive results because of non-specific binding. It then becomes necessary for a confirmatory test[59,71], where isolated viral antigens are used to confirm the identification of the virus. In HBsAg testing, the manufacturers to confirm the presence of HBsAg provide a neutralization test. Chen and Kaplan *et al.*[72] found that most of the weakly positive results become negative in the neutralization test by testing random samples. O'Brien *et al.*[73] recommended this test is only appropriate for weakly positive results. The Center for Disease Control also suggested the same for weakly positive results obtained from anti-HCV assays[74].

The infected persons may develop protective antibody (anti-HCV and anti-HBe) with infection. A detectable amount of antibody takes 5-6 weeks after infection. This stage is known as seroconversion. However, there are several limitations in interpreting results from serological assays during different stages, which includes i) indistinguishability of acute, chronic or past active infection, ii) time lag between the onset of symptoms and seroconversion of infected individuals (e.g. for HCV, 3 months is the maximum time for seroconversion), and iii) there is no significant correlation between the active replications and detection of antibody [75,76].

The recombinant immunoblot assay is often not a good choice for HCV detection because of the poor cost-effectiveness and the difficulty in performing the test, which results in a high number of indeterminant results[77]. Moreover, ELISA and EIA for HCV Ag detection are time-consuming and require highly trained operators. Although recently available 3rd generation tests are considerably more sensitive compared to the 1st and 2nd generation tests but tend to be less specific because of the high sensitivity giving a high number of false-positive results[78].

The improper development of anti-HCV can occur in immunocompromised people, and therefore, a qualitative HCV-RNA detection test is required for the verification of active viral infection[79]. This makes serological diagnosis for HCV more challenging because of the versatility of viral genome and antigen[80]. Further, detecting HCV RNA is a reliable method for HCV diagnosis, but, the method is time-consuming, expensive and requires considerable expertise. This method can also produce a high number of false-positive results because of cross-contamination[81].

Recent advancements have introduced a new highly sensitive automated chemiluminescent technique (CLIA) for HCV core antigen detection which can overcome the limitations of previous core antigen methods[82,83]. Modern serological laboratory tests are especially based on automated but care should be taken during handling the sample to avoid hemolysis as it may interfere with the colorimetric signal measurement and hampers the accuracy in marker detection[84].

When serological approaches are not absolute, molecular diagnostic assays are useful for the detection of the hepatitis viral genome. The intelligent and logical combinations of both serological and molecular assays as diagnostic tools are more reliable and applicable in modern days for viral hepatitis diagnosis but are expensive and time-consuming[85].

1.3.2 Malaria

Effective and accurate diagnosis of malaria is significantly important as misdiagnosis has a substantial impact on mortality, morbidity and drug resistance. Current methods for malaria diagnosis includes microscopy, rapid diagnostic test (RDT), serological and molecular-based techniques[86]. Of these, Microscopy is a 'gold standard' for *Plasmodium* parasite detection. However, expertise is required to distinguish the several stages of parasites in the red blood cells during microscopy detection. As reasoned above in HBV and HCV diagnosis, RDT, molecular assays and serological assays are less effective diagnosis methods, specifically when the detection needs a point of care diagnosis at remote place.

1.4 Infrared Spectroscopy in Blood-Borne Diseases

Vibrational spectroscopy provides the molecular information of functional groups and molecules for samples in all physical states. Infrared technology is gaining attention over the year in the stream of vibrational spectroscopy. Technological advancement in IR spectroscopy has broadened its application in a variety of fields, specifically in developing medicinal and biological sciences[87–89]. When a sample is probed with infrared radiation, the oscillating nuclei absorbs light at particular wavelengths and gets excited. Mostly, molecules or functional groups that have a change in dipole moment during the vibration tend to be good infrared absorbers.

The infrared region of the electromagnetic spectrum begins at 780 nm (just after visible red light) and ends at ~1 mm wavelength (the starting point of microwave spectra). The IR spectral region is further divided into three ranges, NIR (Near-Infrared Spectra), MIR (Mid-Infrared

Spectra) and FIR (Far-Infrared Spectra). NIR covers between 12500-4000 cm^{-1} range[90]. MIR represents the wavelength range 4000-400 cm^{-1}, and the rest of the IR region 400-10 cm^{-1} is FIR[91].

The mid-IR spectral region is the first choice for biological research particularly for medical diagnostics, as this range shows many peaks from fundamental vibrational modes that can be readily assigned. The fundamental vibrations contain 'fingerprint' characteristics, which are directly correlated with the functional groups of the biochemical substances[92]. For example, in biological samples, the stretching vibrations of CH_2 and CH_3 fatty acids range from 3050–2800 cm^{-1}. The region between 1750–1500 cm^{-1} is attributed to the amide I and II bands (stretching of C=O, C-N and bending of NH bond) of peptides, and between 1300-900 cm^{-1} wavenumber represent vibrations of C-O, C-O-C bonds from polysaccharides and phosphodieter stretches from nucleic acids[93,94].

The three main types of infrared based techniques are transmission, transflection and Attenuated Total Reflection (ATR) spectroscopy. In a transmission experiment, IR radiation is transmitted through the sample and substrate before reaching the detector, which is usually a liquid N2 Mercury Cadmium Telluride (MCT) detector or a room temperature Deuterated Triglycine Sulfate (DTGS) detector. One of the drawbacks with transmittance measurements is that the IR transparent substrates are very expensive. Furthermore, the transmission spectra can be affected by various types of physical effects during the sample measurements. While algorithms have been developed for correcting the light scattering effect[95], refraction and dispersion of light[96] and some other optical effects particularly related to thin film measurements[97], these are rather computationally demanding and the jury is still out as to whether these algorithms actually improve the diagnostic capability of the technique[98].

In transflection mode, detection of IR absorption occurs after transmission of radiation through the sample, which is then reflected from the substrate and transmitted again through the sample. This doubles the path length and therefore doubles the absorption (Beer-Lambert law) and increases the signal-to-noise ratio. According to recent studies, a standing wave can be generated when low-emissivity slides are used as the substrates. Due to this standing wave, the resulting spectra show significant differences across the spectrum depending on the sample thickness[99].

A more direct approach that does not necessarily require the use of substrate and avoids the physical effects observed in transmission and transflection measurements is Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectroscopy, which is based on the

principle of total internal reflection. This sampling mode is well suited to both dry films and bio-fluids[100] and has been used to study protein and lipid components within cellular membranes[101]. The ATR technique has a wide range of applications in the quantitative analysis of a variety of serum components of clinical interests[102].

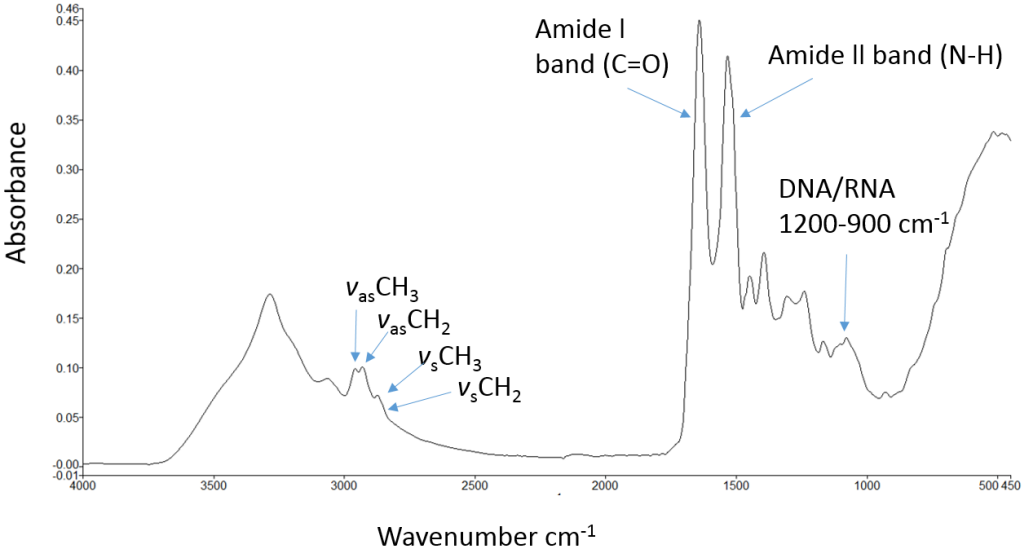


Figure 4: A typical ATR-FTIR spectrum of serum showing the major bands

ATR-FTIR instruments contain an internal reflection element (IRE), which is usually a germanium, diamond, silicon or zinc selenide crystal[103]. The sample should be placed in direct contact with the IRE so that the evanescent wave can penetrate into the sample. Sample thickness is not important because the IR wave usually penetrates to a depth of 1-3 µm depending on the wavelength, angle of incidence and refractive indiex of the material under investigation. Providing the sample is greater than 3 µm there will be no contributions from the background substrate[104].

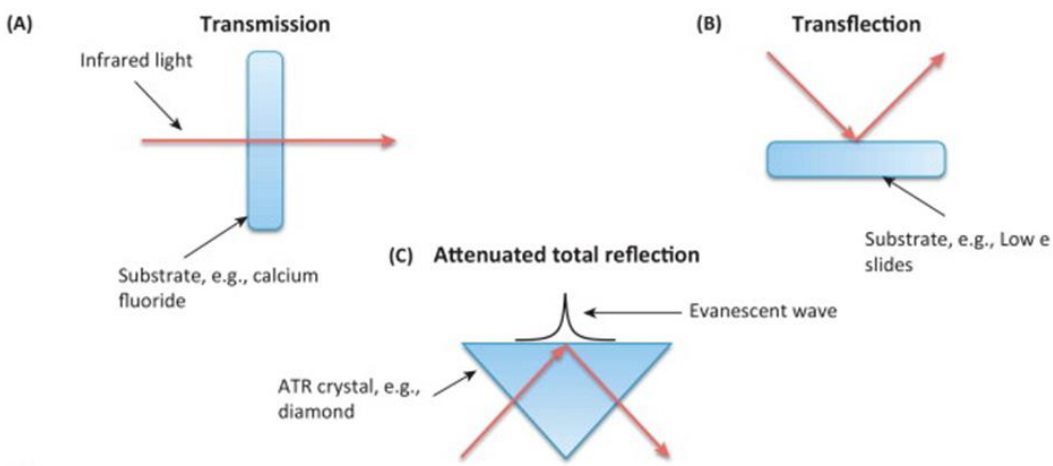


Figure 5: Diagram of the three main Infrared spectroscopy modes[105]

ATR has proven to be a powerful diagnostic tool that is a fast, cost-effective, high-throughput, non-destructive and reagent-less technique, which has a huge range of applications in bioscience. Furthermore, the sample can be directly analysed with minimal preparation resulting in low waste generation. ATR instruments are generally small and affordable and can be battery operated, which is an advantage in developing countries, giving the technique a number of advantages over current diagnostic tools. However, limited sensitivity and selectivity in complex biological samples during IR measurements can be a disadvantage compared to an enzyme-based colorimetric method. To further enhance the diagnostic capability of the technique robust and validated chemometric modelling is required. When analysing wet samples, the contribution to the spectrum from water is significant and water subtraction is required. Drying the sample before ATR measurements can remove this step but adds more time to the measurement[94].

The application of the FTIR technique for biofluid analysis, especially serum is very limited. In the medical field most studies have been performed on tissues and cells[92] with a few studies investigating the concentration of lipid, protein, and glucose molecules in serum[106–108]. ATR spectroscopy using serum has also been applied to investigate spongiform encephalopathy, acute pancreatitis and rheumatoid arthritis along with analyte detection[109–111].

A study applied FTIR spectroscopy to assess hepatic fibrosis in individuals with chronic HCV infection. The serum makes an ideal medium to detect disease markers because it is relatively homogenous compared to cells and tissues. Moreover, a number of sophisticated processing and preprocessing techniques have been developed to compare and classify spectra[112]. In summary, ATR-FTIR has been established as an outstanding spectroscopic tool for serum analysis due to its simplicity and rapidity in the clinical laboratory[105]. There is one paper in the literature on the detection of metabolites in HIV-infected serum samples[113].

This work entailed liquefying the frozen serum slowly at room temperature followed by 30 min incubation in a 56°C water bath, 5–10 μL serum was transferred onto separate glass slides (Kevley Technologies, Ohio, USA) and air-dried overnight to form homogeneous dried films. Serum samples were then scraped off onto the diamond crystal surface area and the spectra recorded. The findings indicated that ATR-FTIR has a potential to monitor the progression of HIV infection and evaluation of treatment by detecting metabolites associated with the infection but not the virus per se[113,114] . The time was taken to prepare

the samples, in this case, is over 24 hours whereas our approach will entail analyzing the serum directly in the liquid state.

In another study, a structural analysis of HIV protein (HIV-1-gp41) was performed using ATR-FTIR spectroscopy. This study indicated that ATR-FTIR may be applied to an immunological screening of HIV-related antigens[115]. A review[114] in 2016, depicts a picture of vibrational spectroscopic applications in biofluids (serum, plasma and bile) where serum was dried and the time taken to dry the sample and the effects of differential drying can lead reduced sensitivity.

1.5 Blood parameters associated with infection

Glucose and urea are the two important blood parameters in *Plasmodium* infected individuals and their simultaneous detection along with the *Plasmodium* parasite by ATR-FTIR can assist a clinical management of malaria patients. However, no previous studies using ATR-FTIR and other detection methods were reported for simultaneous detection.

No study has investigated the use of ATR-FTIR for detecting HBV and HCV clinical serum sample, therefore, limiting its suitability as a point of care diagnostic tool. Moreover, the literature lacks a comprehensive and robust model as a standard for detecting an infected sample.

As serum proteins typically reflect the proteins generated from host cells (infected/non-infected), by comparing the hepatocytes infected with HBV with the non-infected control cells, it may be possible to identify spectral bands associated with the virus. However, incomplete knowledge of such spectral bands and the implicit mechanism upon infection limits the development in HBV diagnosis.

Raman spectroscopy serves as a complementary technique for infrared spectroscopy and studying both will lead to a deeper understanding of the biochemical behaviour associated with HBV infection. Particularly, these powerful tools in combination show promise in monitoring changes at subcellular level. Due to lack of resources, infrastructure and clinical samples the subcellular biochemical changes have been under the shadow and should be investigated.

1.6 Specific objectives and motivation

1. To simultaneously quantify glucose, urea and *Plasmodium* parasites using ATR-FTIR from single pinprick, which can potentially be used as a point-of-care malaria diagnostic in remote regions.
2. To develop a robust multivariate model using ATR-FTIR spectra to differentiate HBV and HCV infected serum from uninfected controls from clinical samples.
3. To identify unique infrared marker bands for each type of infection and assign them to specific macromolecule functional groups.
4. To determine the absolute detection limit of the ATR-FTIR technique by spiking serum with HBV and HCV of known viral load.
5. To discriminate between infected and uninfected cells, based on infrared and Raman spectroscopic signatures as well as identify the changes related to the presence of virus.
6. To distinguish the HBV infected hepatoma cells from control by comparing the cluster analysis at subcellular level (cell membrane, cytoplasm and nucleus).

1.7 Thesis organization

This thesis is organized into six chapters:

Chapter one introduces the literature survey of the target diseases, scope of ATR-FTIR, research problems, objectives of the study and articulation of the thesis.

Chapter two describes the instrumentation and chemometrics analysis, which have been used to perform the study

Chapter three depicts the simultaneous detection of two essential blood analytes (glucose and urea) by using ATR-FTIR method along with the malaria parasites.

Chapter four discusses about the implementation of ATR-FTIR in diagnosing hepatitis B and hepatitis C viral infection in infected human serum.

In chapter five, hepatitis B viral infected and uninfected hepatocyte are monitored by using Raman micro-spectroscopy and synchrotron FTIR micro-spectroscopy.

Chapter six talks about future scopes of this study

1.8 References

(1) Ampofo, W.; Nii-trebi, N.; Ansah, J.; Naito, H.; Aidoo, S.; Nuvor, V.; Yamamoto, N.; Ofori-adjei, D.; Abe, K.; Brandful, J.; *et al*. Prevalence of Blood-Borne Infectious Diseases in Blood Donors in Ghana. *Journal of Clinical Microbiology* **2002**, *40* (9), 1–4.

(2) Pirozzolo, J. J.; LeMay, D. C. Blood-Borne Infections. *Clinics in Sports Medicine*. July 2007, pp 425–431.

(3) Kitchen, A. D.; Chiodini, P. L. Malaria and Blood Transfusion. *Vox Sanguinis* **2006**, *90* (2), 77–84.

(4) Summers, J.; Mason, W. S. Replication of the Genome of a Hepatitis B-like Virus by Reverse Transcription of an RNA Intermediate. *Cell* **1982**, *29* (2), 403–415.

(5) Soussan, P.; Garreau, F.; Zylberberg, H.; Ferray, C.; Brechot, C.; Kremsdorf, D. In Vivo Expression of a New Hepatitis B Virus Protein Encoded by a Spliced RNA. *The Journal of clinical investigation* **2000**, *105* (1), 55–60.

(6) Ganem, D.; Varmus, H. E. The Molecular Biology of the Hepatitis B Viruses. *Annual review of biochemistry* **1987**, *56* (1), 651–693.

(7) Seeger, C.; Mason, W. S. Hepatitis B Virus Biology. *Microbiology and Molecular Biology Reviews* **2000**, *64* (1), 51–68.

(8) Neuveut, C.; Wei, Y.; Buendia, M. A. Mechanisms of HBV-Related Hepatocarcinogenesis. *Journal of Hepatology* **2010**, *52* (4), 594–604.

(9) El-Serag, H. B. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology* **2012**, *142* (6), 1264–1273.

(10) Siegel, R.; Ma, J.; Zou, Z.; Jemal, A. Cancer Statistics, 2014. *CA Cancer J Clin* **2014**, *64* (1), 9–29.

(11) Bosch, F. X.; Ribes, J.; Díaz, M.; Cléries, R. Primary Liver Cancer: Worldwide Incidence and Trends. *Gastroenterology* **2004**, *127* (5), S5–S16.

(12) Murray, C. J. L.; Vos, T.; Lozano, R.; Naghavi, M.; Flaxman, A. D.; Michaud, C.; Ezzati, M.; Shibuya, K.; Salomon, J. A.; Abdalla, S. Disability-Adjusted Life Years (DALYs) for 291 Diseases and Injuries in 21 Regions, 1990–2010: A Systematic Analysis for the Global Burden of Disease Study 2010. *The lancet* **2013**, *380* (9859), 2197–2223.

(13) Lozano, R.; Naghavi, M.; Foreman, K.; Lim, S.; Shibuya, K.; Aboyans, V.; Abraham, J.; Adair, T.; Aggarwal, R.; Ahn, S. Y. Global and Regional Mortality from 235 Causes of Death for 20 Age Groups in 1990 and 2010: A Systematic Analysis for the Global Burden of Disease Study 2010. *The lancet* **2013**, *380* (9859), 2095–2128.

(14) Bialecki, E. S.; Di Bisceglie, A. M. Clinical Presentation and Natural Course of Hepatocellular Carcinoma. *European journal of gastroenterology & hepatology* **2005**, *17* (5), 485–489.

(15) Moradpour, D.; Blum, H. E. Pathogenesis of Hepatocellular Carcinoma. *European journal of gastroenterology & hepatology* **2005**, *17* (5), 477–483.

(16) Sherlock, S.; Niazi, S. P.; Fox, R. A.; Scheuer, P. J. Chronic Liver Disease and Primary Liver-Cell Cancer with Hepatitis-Associated (Australia) Antigen in Serum. *The lancet* **1970**, *295* (7659), 1243–1247.

(17) Beasley, R. P.; Lin, C.-C.; Hwang, L.-Y.; Chien, C.-S. Hepatocellular Carcinoma and Hepatitis B Virus: A Prospective Study of 22 707 Men in Taiwan. *The lancet* **1981**, *318* (8256), 1129–1133.

(18) Ott, J. J.; Stevens, G. A.; Groeger, J.; Wiersma, S. T. Global Epidemiology of Hepatitis B Virus Infection: New Estimates of Age-Specific HBsAg Seroprevalence and Endemicity. *Vaccine* **2012**, *30* (12), 2212–2219.

(19) Zeisel, M. B.; Lucifora, J.; Mason, W. S.; Sureau, C.; Beck, J.; Levrero, M.; Kann, M.; Knolle, P. A.; Benkirane, M.; Durantel, D. Towards an HBV Cure: State-of-the-Art and Unresolved Questions—report of the ANRS Workshop on HBV Cure. *Gut* **2015**, gutjnl-2014-308943.

(20) Global, Regional, and National Age-Sex Specific All-Cause and Cause-Specific Mortality for 240 Causes of Death, 1990–2013: A Systematic Analysis for the Global Burden of Disease Study 2013. *Lancet* **2015**, *385* (9963), 117–171.

(21) Schweitzer, A.; Horn, J.; Mikolajczyk, R. T.; Krause, G.; Ott, J. J. Estimations of Worldwide Prevalence of Chronic Hepatitis B Virus Infection: A Systematic Review of Data Published between 1965 and 2013. *The Lancet* **2015**, *386* (10003), 1546–1555.

(22) Edmunds, W. J.; Medley, G. F.; Nokes, D. J.; O'callaghan, C. J.; Whittle, H. C.; Hall, A. J. Epidemiological Patterns of Hepatitis B Virus (HBV) in Highly Endemic Areasr. *Epidemiology and infection* **1996**, *117* (2), 313–325.

(23) Kiire, C. F. The Epidemiology and Prophylaxis of Hepatitis B in Sub-Saharan Africa: A

View from Tropical and Subtropical Africa. *Gut* **1996**, *38* (Suppl 2), S5-12.

(24) Mast, E. E.; Weinbaum, C. M.; Fiore, A. E.; Alter, M. J.; Bell, B. P.; Finelli, L.; Rodewald, L. E.; Douglas Jr., J. M.; Janssen, R. S.; Ward, J. W. A Comprehensive Immunization Strategy to Eliminate Transmission of Hepatitis B Virus Infection in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP) Part II: Immunization of Adults. *MMWR Recomm Rep* **2006**, *55* (Rr-16), 1–4.

(25) Zhang, J.; Zou, S.; Giulivi, A. Epidemiology of Hepatitis B in Canada. *Canadian Journal of Infectious Diseases* **2001**, *12* (6), 345–350.

(26) Lavanchy, D. Hepatitis B Virus Epidemiology, Disease Burden, Treatment, and Current and Emerging Prevention and Control Measures. *Journal of viral hepatitis* **2004**, *11* (2), 97–107.

(27) Yang, S.; Wang, B.; Chen, P.; Yu, C.; Deng, M.; Yao, J.; Zhu, C.; Ren, J.; Wu, W.; Ju, B. Effectiveness of HBV Vaccination in Infants and Prediction of HBV Prevalence Trend under New Vaccination Plan: Findings of a Large-Scale Investigation. *PloS one* **2012**, *7* (10), e47808.

(28) WHO. Sixty-Seventh World Health Assembly. Agenda Item 12.3. May 24, 2014 . *Retrieved From http://apps.who.int/gb/ebwha/pdf_files/WHA67/A67_Jour1-en.pdf (Accessed April 20, 2015)* **2014**.

(29) Tellinghuisen, T. L.; Rice, C. M. Interaction between Hepatitis C Virus Proteins and Host Cell Factors. *Current opinion in microbiology* **2002**, *5* (4), 419–427.

(30) Suzuki, T.; Ishii, K.; Aizaki, H.; Wakita, T. Hepatitis C Viral Life Cycle. *Advanced Drug Delivery Reviews* **2007**, *59* (12), 1200–1212.

(31) Pileri, P.; Uematsu, Y.; Campagnoli, S.; Galli, G.; Falugi, F.; Petracca, R.; Weiner, A. J.; Houghton, M.; Rosa, D.; Grandi, G. Binding of Hepatitis C Virus to CD81. *Science* **1998**, *282* (5390), 938–941.

(32) Honda, M.; Beard, M. R.; Ping, L.-H.; Lemon, S. M. A Phylogenetically Conserved Stem-Loop Structure at the 5′ Border of the Internal Ribosome Entry Site of Hepatitis C Virus Is Required for Cap-Independent Viral Translation. *Journal of virology* **1999**, *73* (2), 1165–1174.

(33) Egger, D.; Wölk, B.; Gosert, R.; Bianchi, L.; Blum, H. E.; Moradpour, D.; Bienz, K. Expression of Hepatitis C Virus Proteins Induces Distinct Membrane Alterations

Including a Candidate Viral Replication Complex. *Journal of virology* **2002**, *76* (12), 5974–5984.

(34) Gosert, R.; Egger, D.; Lohmann, V.; Bartenschlager, R.; Blum, H. E.; Bienz, K.; Moradpour, D. Identification of the Hepatitis C Virus RNA Replication Complex in Huh-7 Cells Harboring Subgenomic Replicons. *Journal of virology* **2003**, *77* (9), 5487–5492.

(35) Appel, N.; Schaller, T.; Penin, F.; Bartenschlager, R. From Structure to Function: New Insights into Hepatitis C Virus RNA Replication. *Journal of Biological Chemistry* **2006**, *281* (15), 9833–9836.

(36) Jones, D. M.; McLauchlan, J. Hepatitis C Virus: Assembly and Release of Virus Particles. *Journal of Biological Chemistry* **2010**, *285* (30), 22733–22739.

(37) Mohd Hanafiah, K.; Groeger, J.; Flaxman, A. D.; Wiersma, S. T. Global Epidemiology of Hepatitis C Virus Infection: New Estimates of Age-specific Antibody to HCV Seroprevalence. *Hepatology* **2013**, *57* (4), 1333–1342.

(38) WHO. Hepatitis C. *Available: http://www.who.int/mediacentre/factsheets/fs164/en/* **2014**.

(39) Di Bisceglie, A. M. Natural History of Hepatitis C: Its Impact on Clinical Management. *Hepatology* **2000**, *31* (4), 1014–1018.

(40) Lauer, G. M.; Walker, B. D. Hepatitis C Virus Infection. *New England journal of medicine* **2001**, *345* (1), 41–52.

(41) Koike, K.; Tsutsumi, T.; Miyoshi, H.; Shinzawa, S.; Shintani, Y.; Fujie, H.; Yotsuyanagi, H.; Moriya, K. Molecular Basis for the Synergy between Alcohol and Hepatitis C Virus in Hepatocarcinogenesis. *Journal of gastroenterology and hepatology* **2008**, *23* (s1), S87–S91.

(42) Choo, Q.-L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M. Isolation of a cDNA Clone Derived from a Blood-Borne Non-A, Non-B Viral Hepatitis Genome. *Science* **1989**, *244* (4902), 359–362.

(43) Smith, D. B.; Mellor, J.; Jarvis, L. M.; Davidson, F.; Kolberg, J.; Urdea, M.; Yap, P. L.; Simmonds, P. Variation of the Hepatitis C Virus 5' Non-Coding Region: Implications for Secondary Structure, Virus Detection and Typing. The International HCV Collaborative Study Group. *J Gen Virol* **1995**, *76 (Pt 7)*, 1749–1761.

(44) Koike, K. Hepatitis C Virus Contributes to Hepatocarcinogenesis by Modulating Metabolic and Intracellular Signaling Pathways. *Journal of gastroenterology and hepatology* **2007**, *22* (s1), S108–S111.

(45) Farinati, F.; Cardin, R.; Bortolami, M.; Burra, P.; Russo, F. P.; Rugge, M.; Guido, M.; Sergio, A.; Naccarato, R. Hepatitis C Virus: From Oxygen Free Radicals to Hepatocellular Carcinoma. *Journal of viral hepatitis* **2007**, *14* (12), 821–829.

(46) Zignego, A. L.; Giannini, C.; Monti, M.; Gragnani, L. Hepatitis C Virus Lymphotropism: Lessons from a Decade of Studies. *Dig Liver Dis* **2007**, *39 Suppl 1*, S38-45.

(47) Ramos-Casals, M.; Font, J. Extrahepatic Manifestations in Patients with Chronic Hepatitis C Virus Infection. *Curr Opin Rheumatol* **2005**, *17* (4), 447–455.

(48) Cacoub, P.; Gragnani, L.; Comarmond, C.; Zignego, A. L. Extrahepatic Manifestations of Chronic Hepatitis C Virus Infection. *Digestive and Liver Disease* **2014**, *46*, S165–S173.

(49) WHO. *WHO Publishes World Malaria Report 2018*; World Health Organization, 2018.

(50) Greenwood, B. M.; Fidock, D. A.; Kyle, D. E.; Kappe, S. H. I.; Alonso, P. L.; Collins, F. H.; Duffy, P. E. Malaria: Progress, Perils, and Prospects for Eradication. *The Journal of clinical investigation* **2008**, *118* (4), 1266–1276.

(51) Khoshmanesh, A.; Dixon, M. W. A.; Kenny, S.; Tilley, L.; McNaughton, D.; Wood, B. R. Detection and Quantification of Early-Stage Malaria Parasites in Laboratory Infected Erythrocytes by Attenuated Total Reflectance Infrared Spectroscopy and Multivariate Analysis. *Analytical Chemistry* **2014**, *86* (9), 4379–4386.

(52) WHO. *Guidelines on Hepatitis B and C Testing*; World Health Organization, 2017.

(53) Ismail, N.; Fish, G. E.; Smith, M. B. Laboratory Evaluation of a Fully Automated Chemiluminescence Immunoassay for Rapid Detection of HBsAg, Antibodies to HBsAg, and Antibodies to Hepatitis C Virus. *Journal of clinical microbiology* **2004**, *42* (2), 610–617.

(54) Macalino, G. E.; Vlahov, D.; Sanford-Colby, S.; al., et. Prevalence and Incidence of HIV, Hepatitis B Virus, and Hepatitis C Virus Infections among Males in Rhode Island Prisons. *Am J Public Health* **2004**, *94* (7), 1218–1223.

(55) Purcell, R. H.; Wong, D. C.; Alter, H. J.; Holland, P. V. Microtiter Solid-Phase

Radioimmunoassay for Hepatitis B Antigen. *Applied microbiology* **1973**, *26* (4), 478–484.

(56) Fwu, C.-W.; Chien, Y.-C.; Nelson, K. E.; Kirk, G. D.; You, S.-L.; Kuo, H.-S.; Feinleib, M.; Chen, C.-J. Mortality after Chronic Hepatitis B Virus Infection: A Linkage Study Involving 2 Million Parous Women from Taiwan. *Journal of Infectious Diseases* **2010**, *201* (7), 1016–1023.

(57) Usuda, S.; Okamoto, H.; Iwanari, H.; Baba, K.; Tsuda, F.; Miyakawa, Y.; Mayumi, M. Serological Detection of Hepatitis B Virus Genotypes by ELISA with Monoclonal Antibodies to Type-Specific Epitopes in the preS2-Region Product. *Journal of virological methods* **1999**, *80* (1), 97–112.

(58) Asli, M.; Kandelouei, T.; Rahimyan, K.; Davoodbeglou, F.; Vaezjalali, M. Characterization of Occult Hepatitis B Infection Among Injecting Drug Users in Tehran, Iran. *Hepatitis Monthly* **2016**, *16* (3).

(59) Zhang, K.; Wang, L.; Lin, G.; Li, J. Is Anti-Hepatitis C Virus Antibody Level an Appropriate Marker to Preclude the Need for Supplemental Testing. *Intervirology* **2016**, *58* (5), 310–317.

(60) Roth, W. K.; Lee, J.-H.; Rüster, B.; Zeuzem, S. Comparison of Two Quantitative Hepatitis C Virus Reverse Transcriptase PCR Assays. *Journal of clinical microbiology* **1996**, *34* (2), 261–264.

(61) Cobo, F. Application of Molecular Diagnostic Techniques for Viral Testing. *The open virology journal* **2012**, *6* (1).

(62) Heiat, M.; Ranjbar, R.; Alavian, S. M. Classical and Modern Approaches Used for Viral Hepatitis Diagnosis. *Hepatitis Monthly* **2014**, *14* (4).

(63) Amini, A.; Varsaneux, O.; Kelly, H.; Tang, W.; Chen, W.; Boeras, D. I.; Falconer, J.; Tucker, J. D.; Chou, R.; Ishizaki, A.; *et al*. Diagnostic Accuracy of Tests to Detect Hepatitis B Surface Antigen: A Systematic Review of the Literature and Meta-Analysis. *BMC Infectious Diseases* **2017**, *17* (S1), 698.

(64) Hee Kim, M.; Young Kang, S.; In Lee, W. Evaluation of a New Rapid Test Kit to Detect Hepatitis C Virus Infection. *Journal of Virological Methods* **2013**, *193*, 379–382.

(65) Fattovich, G. Natural History of Hepatitis B. *Journal of hepatology* **2003**, *39*, 50–58.

(66) Beltrami, E. M.; Williams, I. T.; Shapiro, C. N.; Chamberland, E. Risk and Management

of Blood-Borne Infections in Health Care Workers Risk and Management of Blood-Borne Infections in Health Care Workers. **2000**, *13* (3), 385–407.

(67) Lin, S.; Sun, Q.; Mao, W.; Chen, Y. Serum Immunoglobulin A (IgA) Level Is a Potential Biomarker Indicating Cirrhosis during Chronic Hepatitis B Infection. *Gastroenterology Research and Practice* **2016**, *2016*, 1–6.

(68) Coleman, P. F. Detecting Hepatitis B Surface Antigen Mutants. *Emerging infectious diseases* **2006**, *12* (2), 198.

(69) Beltrami, E. M.; Williams, I. T.; Shapiro, C. N.; Chamberland, M. E. Risk and Management of Blood-Borne Infections in Health Care Workers. *Clinical microbiology reviews* **2000**, *13* (3), 385–407.

(70) Berger, A. Viral Genome Quantification as a Tool for Improving Patient Management: The Example of HIV, HBV, HCV and CMV. *Journal of Antimicrobial Chemotherapy* **2002**, *49* (5), 713–721.

(71) Marcellin, P.; Martinot-Peignoux, M.; Boyer, N.; Pouteau, M.; Aumont, P.; Erlinger, S.; Benhamou, J.-P. Second Generation (RIBA) Test in Diagnosis of Chronic Hepatitis C. *The lancet* **1991**, *337* (8740), 551–552.

(72) Chen, D.; Kaplan, L. A. Performance of a New-Generation Chemiluminescent Assay for Hepatitis B Surface Antigen. *Clinical chemistry* **2006**, *52* (8), 1592–1598.

(73) O'Brien, J. E. Hepatitis B Surface Antigen: Decreased Need for Confirmation of Reactive Results. *Clinical chemistry* **2000**, *46* (4), 582–588.

(74) Alter, M. J.; Kuhnert, W. L.; Finelli, L. *Guidelines for Laboratory Testing and Result Reporting of Antibody to Hepatitis C Virus*; Massachusetts Medical Society, 2003.

(75) Alter, M. J. Review of Serologic Testing for Hepatitis C Virus Infection and Risk of Posttransfusion Hepatitis C. *Archives of pathology & laboratory medicine* **1994**, *118* (4), 342–345.

(76) Yuen, M.-F.; Chen, D.-S.; Dusheiko, G. M.; Janssen, H. L. A.; Lau, D. T. Y.; Locarnini, S. A.; Peters, M. G.; Lai, C.-L. Hepatitis B Virus Infection. *Nature Reviews Disease Primers* **2018**, *4*, 18035.

(77) Chevaliez, S. Virological Tools to Diagnose and Monitor Hepatitis C Virus Infection. *Clinical microbiology and infection* **2011**, *17* (2), 116–121.

(78) Mansour, M. A.; Abdella, H. M.; Al Gouhary, E. A.; Helmy, A.; Saleh, S. A.

ImmunoComb II Test in Detection of HCV Antibodies in Oral Fluid. *Journal of Medical Science and Clinical Research* **2015**, *3* (6), 6042–6052.

(79) Richter, S. S. Laboratory Assays for Diagnosis and Management of Hepatitis C Virus Infection. *Journal of clinical microbiology* **2002**, *40* (12), 4407–4412.

(80) Ghany, M. G.; Strader, D. B.; Thomas, D. L.; Seeff, L. B. Diagnosis, Management, and Treatment of Hepatitis C: An Update. *Hepatology* **2009**, *49* (4), 1335–1374.

(81) Bouvier-Alias, M.; Patel, K.; Dahari, H.; Beaucourt, S.; Larderie, P.; Blatt, L.; Hezode, C.; Picchio, G.; Dhumeaux, D.; Neumann, A. U.; *et al*. Clinical Utility of Total HCV Core Antigen Quantification: A New Indirect Marker of HCV Replication. *Hepatology* **2002**, *36* (1), 211–218.

(82) Mederacke, I.; Wedemeyer, H.; Ciesek, S.; Steinmann, E.; Raupach, R.; Wursthorn, K.; Manns, M. P.; Tillmann, H. L. Performance and Clinical Utility of a Novel Fully Automated Quantitative HCV-Core Antigen Assay. *Journal of Clinical Virology* **2009**, *46* (3), 210–215.

(83) Tanaka, E.; Ohue, C.; Aoyagi, K.; Yamaguchi, K.; Yagi, S.; Kiyosawa, K.; Alter, H. J. Evaluation of a New Enzyme Immunoassay for Hepatitis C Virus (HCV) Core Antigen with Clinical Sensitivity Approximating that of Genomic Amplification of HCV RNA. *Hepatology* **2000**, *32* (2), 388–393.

(84) Mel Krajden, M. D.; ART, G. M. Bs. The Laboratory Diagnosis of Hepatitis B Virus. *Can J Infect Dis Med Microbiol* **2005**, *16* (2), 65.

(85) Poljak, M.; Lepej, S. Z.; Rode, O. D. [Recent developments in serologic and molecular diagnosis of hepatitis B and C]. *Acta Med Croatica* **2013**, *67* (4), 281–290.

(86) Tangpukdee, N.; Duangdee, C.; Wilairatana, P.; Krudsood, S. Malaria Diagnosis: A Brief Review. *The Korean journal of parasitology* **2009**, *47* (2), 93–102.

(87) Nafie, L. A.; Freddman, T. B. In Infrared and Raman Spectroscopy of Biological Materials; Gremlich, H.-U., Yan, B., Eds. Marcel Dekker, Inc.: New York 2001.

(88) Mahadevan-Jansen, A.; Puppels, G. J. Biomedical Spectroscopy: Vibrational Spectroscopy and Other Novel Techniques. In *Biomedical Spectroscopy: Vibrational Spectroscopy and Other Novel Techniques*; 2000; Vol. 3918.

(89) Mantsch, H. H.; Jackson, M. Infrared Spectroscopy: New Tool in Medicine. In *Infrared Spectroscopy: New Tool in Medicine*; 1998; Vol. 3257.

(90) Moros, J.; Armenta, S.; Garrigues, S.; de la Guardia, M. Univariate near Infrared Methods for Determination of Pesticides in Agrochemicals. *Analytica chimica acta* **2006**, *579* (1), 17–24.

(91) Goldstein, M.; Seeley, M. E.; Willis, H. A.; Zichy, V. J. I. The Far Infra-Red Spectrum (400-10 Cm− 1) of Isotactic Polypropylene. *Polymer* **1973**, *14* (11), 530–534.

(92) Ellis, D. I.; Goodacre, R. Metabolic Fingerprinting in Disease Diagnosis: Biomedical Applications of Infrared and Raman Spectroscopy. *The Analyst* **2006**, *131* (8), 875.

(93) Ellis, D. I.; Harrigan, G. G.; Goodacre, R. Metabolic Fingerprinting with Fourier Transform Infrared Spectroscopy. In *Metabolic Profiling: its role in biomarker discovery and gene function analysis*; Springer, 2003; pp 111–124.

(94) Perez-Guaita, D.; Garrigues, S.; de la, M. Infrared-Based Quantification of Clinical Parameters. *TrAC - Trends in Analytical Chemistry* **2014**, *62*, 93–105.

(95) Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Byrne, H. J.; Dumas, P.; Gazi, E.; Brown, M.; Clarke, N.; Gardner, P. Resonant Mie Scattering (RMieS) Correction of Infrared Spectra from Highly Scattering Biological Samples. *Analyst* **2010**, *135* (2), 268–277.

(96) Chan, K. L. A.; Kazarian, S. G. Correcting the Effect of Refraction and Dispersion of Light in FT-IR Spectroscopic Imaging in Transmission through Thick Infrared Windows. *Analytical chemistry* **2012**, *85* (2), 1029–1036.

(97) King, S. W.; Milosevic, M. A Method to Extract Absorption Coefficient of Thin Films from Transmission Spectra of the Films on Thick Substrates. *Journal of Applied Physics* **2012**, *111* (7), 73109.

(98) Kochan, K.; Heraud, P.; Kiupel, M.; Yuzbasiyan-Gurkan, V.; McNaughton, D.; Baranska, M.; Wood, B. R. Comparison of FTIR Transmission and Transfection Substrates for Canine Liver Cancer Detection. *Analyst* **2015**, *140* (7), 2402–2411.

(99) Filik, J.; Frogley, M. D.; Pijanka, J.; Wehbe, K.; Cinque, G. Optical Standing-Wave Artifacts in Reflection-Absorption FTIR Microspectroscopy of Biological Materials. In *Journal of Physics: Conference Series*; IOP Publishing, 2012; Vol. 359, p 12006.

(100) Diessel, E.; Willmann, S.; Kamphaus, P.; Kurte, R.; Damm, U.; Heise, H. M. Glucose Quantification in Dried-down Nanoliter Samples Using Mid-Infrared Attenuated Total Reflection Spectroscopy. *Applied spectroscopy* **2004**, *58* (4), 442–450.

(101) Goormaghtigh, E.; Raussens, V.; Ruysschaert, J.-M. Attenuated Total Reflection

Infrared Spectroscopy of Proteins and Lipids in Biological Membranes. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1999**, *1422* (2), 105–185.

(102) Shaw, R. A.; Mantsch, H. H. Infrared Spectroscopy in Clinical and Diagnostic Analysis. **2000**.

(103) Glassford, S. E.; Byrne, B.; Kazarian, S. G. Recent Applications of ATR FTIR Spectroscopy and Imaging to Proteins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2013**, *1834* (12), 2849–2858.

(104) Bassan, P.; Sachdeva, A.; Lee, J.; Gardner, P. Substrate Contributions in Micro-ATR of Thin Samples: Implications for Analysis of Cells, Tissue and Biological Fluids. *Analyst* **2013**, *138* (14), 4139–4146.

(105) Dorling, K. .; Baker, M. . Highlighting Attenuated Total Reflection Fourier Transform Infrared Spectroscopy for Rapid Serum Analysis. *Trends in biotechnology* **2013**, *31* (2008), 327–328.

(106) Shaw, R. A.; Kotowich, S.; Leroux, M.; Mantsch, H. H. Multianalyte Serum Analysis Using Mid-Infrared Spectroscopy. *Annals of Clinical Biochemistry: An international journal of biochemistry in medicine* **1998**, *35* (5), 624–632.

(107) Petibois, C.; Rigalleau, V.; Melin, A.-M.; Perromat, A.; Cazorla, G.; Gin, H.; Déléris, G. Determination of Glucose in Dried Serum Samples by Fourier-Transform Infrared Spectroscopy. *Clinical chemistry* **1999**, *45* (9), 1530–1535.

(108) Petibois, C.; Paiva, M.; Cazorla, G.; Deleris, G. Discriminant Serum Biochemical Parameters in Top Class Marathon Performances. *The Japanese journal of physiology* **2002**, *52* (2), 181–190.

(109) Schmitt, J.; Beekes, M.; Brauer, A.; Udelhoven, T.; Lasch, P.; Naumann, D. Identification of Scrapie Infection from Blood Serum by Fourier Transform Infrared Spectroscopy. *Analytical chemistry* **2002**, *74* (15), 3865–3868.

(110) Martin, T. C.; Moecks, J.; Belooussov, A.; Cawthraw, S.; Dolenko, B.; Eiden, M.; Von Frese, J.; Köhler, W.; Schmitt, J.; Somorjai, R. Classification of Signatures of Bovine Spongiform Encephalopathy in Serum Using Infrared Spectroscopy. *Analyst* **2004**, *129* (10), 897–901.

(111) Staib, A.; Dolenko, B.; Fink, D. J.; Früh, J.; Nikulin, A. E.; Otto, M.; Pessin-Minsley, M. S.; Quarder, O.; Somorjai, R.; Thienel, U. Disease Pattern Recognition Testing for Rheumatoid Arthritis Using Infrared Spectra of Human Serum. *Clinica chimica acta*

2001, *308* (1), 79–89.

(112) Scaglia, E.; Sockalingum, G. D.; Schmitt, J.; Gobinet, C.; Schneider, N.; Manfait, M.; Thiéfin, G. Noninvasive Assessment of Hepatic Fibrosis in Patients with Chronic Hepatitis C Using Serum Fourier Transform Infrared Spectroscopy. *Analytical and Bioanalytical Chemistry* **2011**, *401* (9), 2919–2925.

(113) Sitole, L.; Steffens, F.; Krüger, T. P. J.; Meyer, D. Mid-ATR-FTIR Spectroscopic Profiling of HIV/AIDS Sera for Novel Systems Diagnostics in Global Health. *Omics : a journal of integrative biology* **2014**, *18* (8), 1–11.

(114) Baker, M. J.; Hussain, S. R.; Lovergne, L.; Untereiner, V.; Hughes, C.; Lukaszewski, R. A.; Thiefin, G.; Sockaalingum, G. D. Developing and Understanding Biofluid Vibrational Spectroscopy-a Critical Review.pdf. *Chemical Society Reviews*. 2016, pp 1803–1818.

(115) Agopian, A.; Ronzon, F.; Sauzeat, E.; Sodoyer, R.; El Habib, R.; Buchet, R.; Chevalier, M. Secondary Structure Analysis of HIV-1-gp41 in Solution and Adsorbed to Aluminum Hydroxide by Fourier Transform Infrared Spectroscopy. *Biochim Biophys Acta* **2007**, *1774* (3), 351–358.

Chapter 2. Instrumentation and Chemometrics

2.1 Fourier Transform Infrared (FTIR) Spectroscopy

2.1.1 Michelson Interferometer

Fourier Transform Infrared (FT-IR) spectroscopy is a technique that utilises an infrared beam to generate an interferogram that interacts with the sample before impinging onto a detector. Fourier Transformation is performed on the interferogram to generate the final spectrum. Modern FTIR spectral analysis is almost entirely based on an apparatus known as Michelson Interferometer. The heart of the Michelson Interferometer is the beam splitter along with a fixed and moving mirror. The moving mirror moves back and forth within a specific distance that determines the resolution. At the beam splitter 50 % of the wavetrain is directed to the moving mirror and the other 50% is directed to a fixed mirror. Both the mirrors reflect the wavetrains back to the beamsplitter where they recombine. The Optical Path Difference (OPD) between the reflected wavetrains enables the wavetrains to constructively and destructively interfere when recombined at the beam splitter leading the generation of individual frequencies which impinge on the sample and then the detector simultaneously in a process known as the Multiplex Advantage.[1]

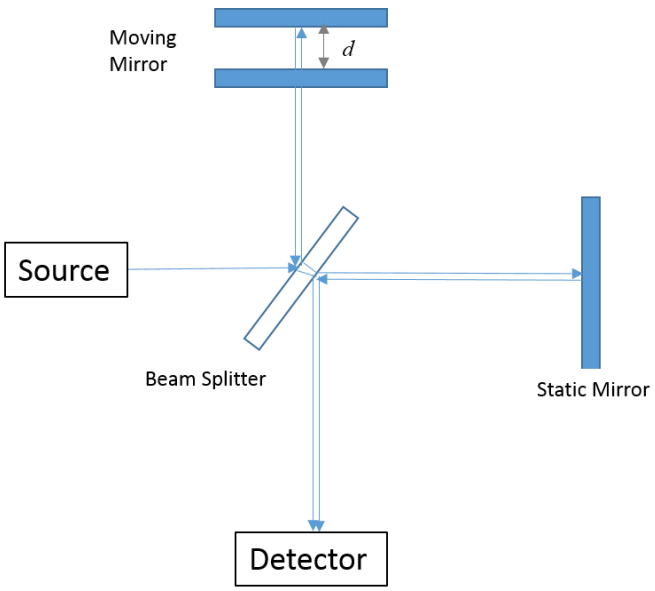


Figure 1: A schematic diagram of the Michelson Interferometer system

2.1.2 Spectral Resolution

In a Michelson interferometer, the spectral resolution is equivalent to 1/OPD. Therefore, if the distance is 0.25 cm then the resolution will be 4 cm^{-1}.

2.1.3 Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) Spectroscopy

The ATR-FTIR system consists of a Michelson Interferometer, infrared detector and an internal reflecting element (IRE). The IRE can be a prism, crystal or similar structural material that enables internal reflection of the source IR radiation. The material usually has a high refractive index such as germanium, zinc selenide or a diamond crystal. The IR beam is directed to the IRE surface and interacts with the sample. The evanescent wave penetrates only about 3 micron into the sample depending on the wavelength, angle of incidence, and the refractive index of the IRE. This small penetration depth gives ATR a major advantage over transmission or transflection measurements for analysing cells and biofluids because the water spectrum does not totally saturate the sample spectrum and can be easily subtracted yielding a high quality spectrum of the biological material. The small amount of light penetration by the evanescent mitigates detector saturation. It is important to maintain good contact between the sample and the crystal at the time of data acquisition to achieve spectra with a good signal-to-noise ratio and a clamp is often employed to apply pressure to the sample to improve the contact. The IR radiation source is usually a hotwire or globar that generates a blackbody of infrared frequencies. The detector is usually a Deuterated Triglycine Sulfate (DTGS), which can operate at room temperature or a Mercury Cadmium Telluride (MCT) semiconductor device, which requires cooling with liquid nitrogen.[2] Another advantage of ATR technology is its size and portability, which enables the technology to be moved from site to site and can easily fit into a biosafety cabinet. This was a critical necessity in this project because the measurements were recorded at the Victorian Infectious Diseases Reference Laboratories (VIDRL) where no samples are allowed outside the building, and must be handled within the biosafety cabinets.

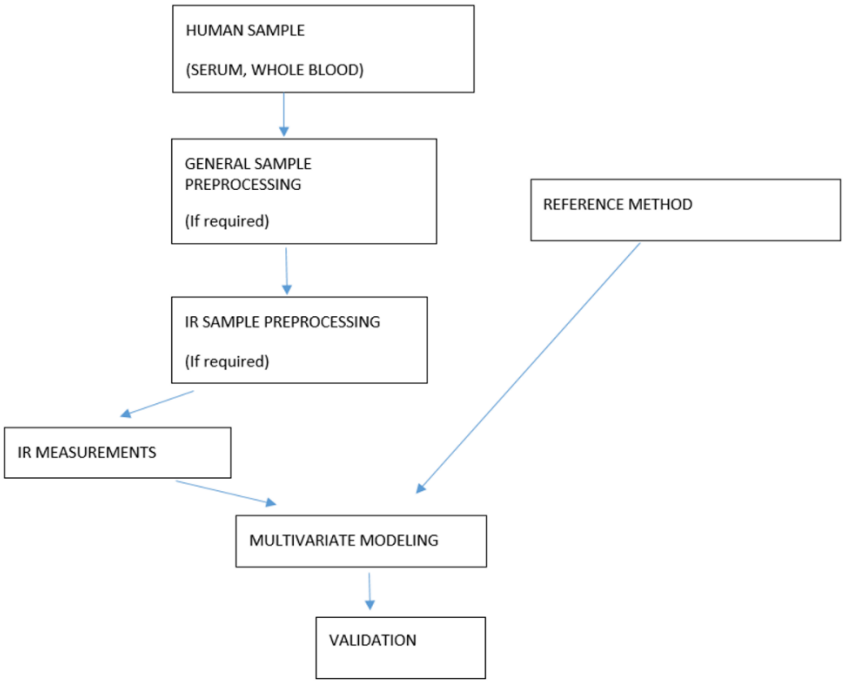


Figure 2: A general sequential process for performing infrared spectroscopy (IR) -based analysis

2.2 Raman spectroscopy

2.2.1 Basic theory

Raman spectroscopy proposed by Sir C.V Raman in 1928[3], is the technique, which uses scattering effect to characterize molecular and crystalline structure of a sample. A single frequency monochromatic laser beam is used as the incident light to induce a dipole moment in the sample molecule. When the monochromatic laser hits the sample, it distorts the electron cloud of the sample molecule, which creates an unstable virtual state and causes polarization. The photon reradiates quickly the original scattered light. Most of the scattered radiation is in the form of elastic incident light called Rayleigh scattering. In this process, the electron is excited to the virtual state and falls back to the ground level without changing the energy state and hence the scattered photon is the same as the incident photon.

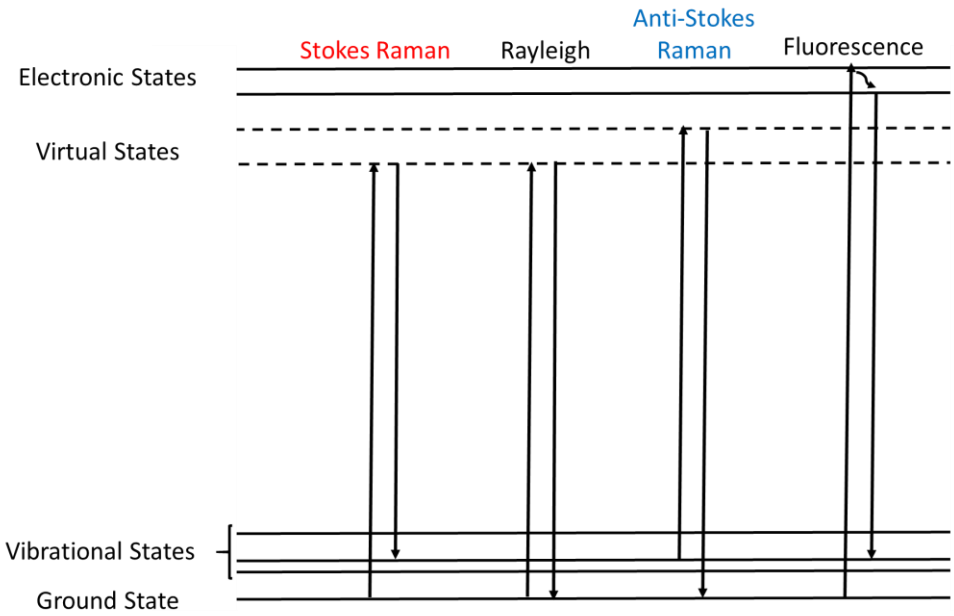


Figure 3: The energy diagram of Raman and Rayleigh scattering process

Only a small portion (one photon per 10^6-10^8 photons) is scattered as inelastic radiation known as Raman scattering. The change in frequency, which is detectable, originates from the induced nuclear motion (dipole moment) that changes during the photon scattering process. A Raman spectrum arises from a change in polarizability due to the molecular vibration of the sample molecule. The sample molecule either absorbs energy from incident photon or transfers the energy to the scattered photon. Stokes scattering results when the scattered photon is at a lower energy than the incident photon, while anti-Stokes scattering occurs when the scattered energy is at a higher energy than the incident photon. Thus, anti-Stokes scattering has more energy and is shifted towards the blue end or shorter wavelength of the spectrum compared to the incident photon. On the other hand, Stokes scattering has less energy and is red shifted towards the longer wavelengths compared to the incident light[4,5].

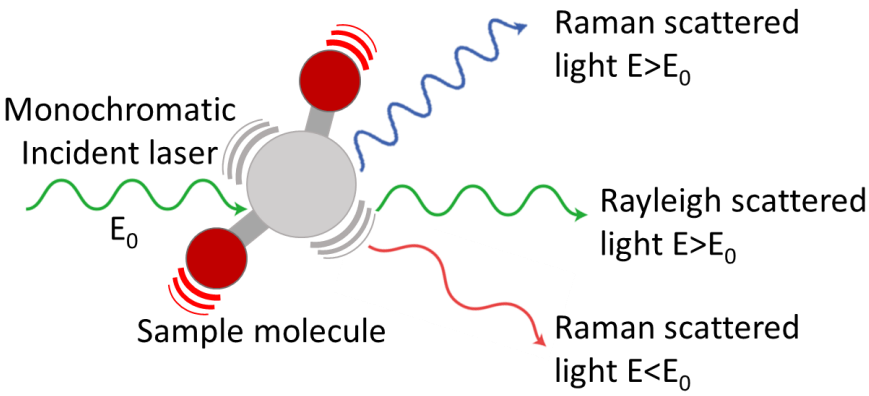


Figure 4: A schematic representation of the scattering of light by sample molecule

Conventional Raman spectroscopy uses Stoke's-scattering rather than anti-stokes, because Stokes scattering is more intense due to the Boltzmann distribution, which states more molecules are in the ground vibrational state at room temperature compared to the excited vibrational states. Because the total number of molecules positioned at the higher excited vibrational state is always less than the number present in lower ground state at room temperature. However, anti-Stoke's scattering can be useful to avoid interference from fluorescence that can occur in Stoke's shifted spectra. Due to increased temperature the band intensity ratio of anti-Stoke's and Stoke's scattering increases, which can be useful for sample temperature measurements[5].

2.2.2 Major components of a Raman spectrometer

Modern Raman spectrometers are generally of two types, dispersive and non-dispersive. The dispersive spectrometer uses a prism or a grating for dispersion and a multichannel CCD detector (charge-coupled-device) for detection of the spectrum. The non-dispersive type is an interferometer-based system such as Michelson interferometer and uses Fourier transformation to produce the Raman spectrum. In Raman spectroscopy, a wide range of laser sources are used for excitation of the sample molecules. For example, Krypton ion laser (647.1 and 530.9 nm), Argon ion laser (515.5 and 488 nm), Helium-Neon laser (632.8 nm), NIR (near infrared) diode lasers (830 and 785 nm), frequency doubled Neodymium–Yttrium Ortho-Vanadate (Nd:YVO_4) and Neodymium–Yttrium Aluminium Garnet (Nd:YAG) diode lasers (532 nm) and Nd:YAG and Nd:YVO_4 lasers (1064 nm)[6].

Coupling with an optical microscope allows the Raman spectrometer to have both visible and spectroscopic analysis providing mapping and/or imaging capability. A highly sensitive confocal (same focus) technique is used in Raman micro-spectroscopy to reduce the intense fluorescence background. Confocality results when light is at the same focal plane of the

illuminating point source and the pinhole. This technique allows the optical microscope to selectively pass the light in the focal plane, therefore, the light emitted from any position above or below the focal plane is restricted from passing through the detector, thus, not contributing to the image[7].

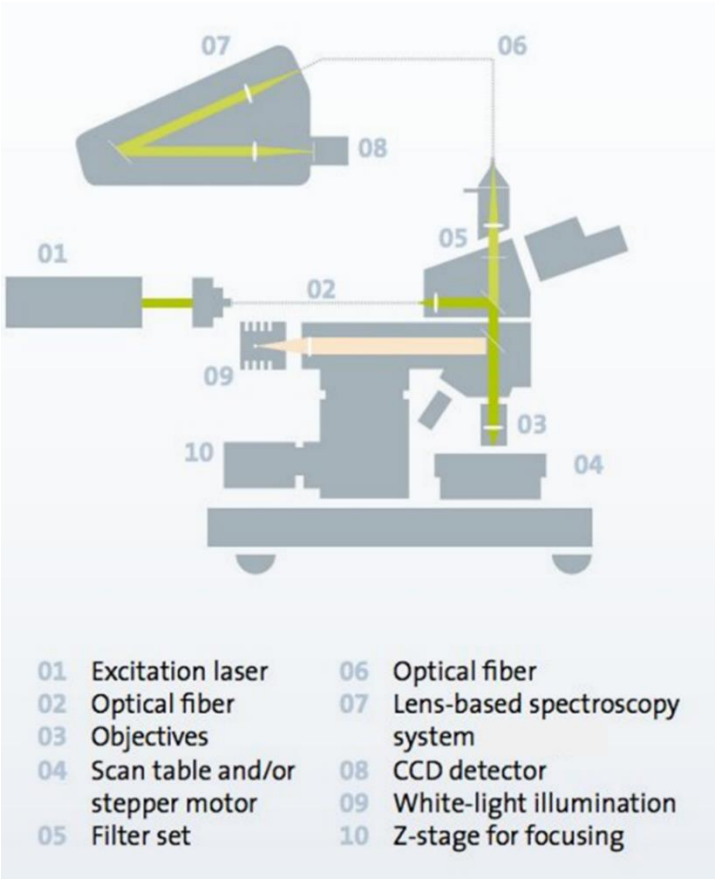


Figure 5: A typical Raman microscope system consisting of an excitation laser, a microscope, and a spectrometer with CCD detector[7]

2.3 Infrared Micro-spectroscopy using Synchrotron radiation

2.3.1 Basic theory

A synchrotron is a circular, large, Giga volt facility using high energy electrons, forcing them to travel in an extremely strong electromagnetic circular orbit (a storage ring), which has been 'synchronized' in a way that the created electron beam travels at the speed of about 299792 $kmsec^{-1}$(just a bit below than the speed of light), known as synchrotron light. The light is so intense that, comparing to the sunlight intensity, the synchrotron light is even more than a million times brighter. After a filtering process, the generated intense synchrotron light can be used to probe the molecular structure of a material including biological samples at a sub-microscopic scale. Previously the application of synchrotron light was limited to particle

physics, nowadays this advanced technology is applied to a vast range of fields including agriculture biomedicine, advanced materials, defence science, food and food technology, environmental sustainability, cultural heritage, forensic science, energy industry, mining, electronics, planetary science and many more.[8,9]

The 1st generation synchrotron beamlines were based on available facilities designed for studying particle physics. Second generation synchrotrons was mainly focusing on the synchrotron radiation production and by implementation of electron storage rings. In current 3rd generation synchrotron facilities long straight magnets (undulators and wigglers) are incorporated as 'insertion devices' for optimizing the synchrotron light in the storage ring. For creating broad intense incoherent radiation, wiggler magnets are used whereas undulators are used for creating narrower beams of coherent light, which are significantly more intense. The wavelength of the coherent radiation can be selected (known as 'harmonics') and 'tuned' when the magnetic field is manipulated in the device. For the establishment of a 4th generation synchrotron sources, the technical challenges are there, and will be generally be used in hard X-ray FEL (free electron lasers).[9]

2.3.2 Australian Synchrotron

The Australian synchrotron beamline facility is a 3rd generation synchrotron technology at the advanced level. Three types of light sources (bending magnets, unduators and multipole wigglers) are used in the Australian synchrotron to perform an extensive range of advanced research experiments.

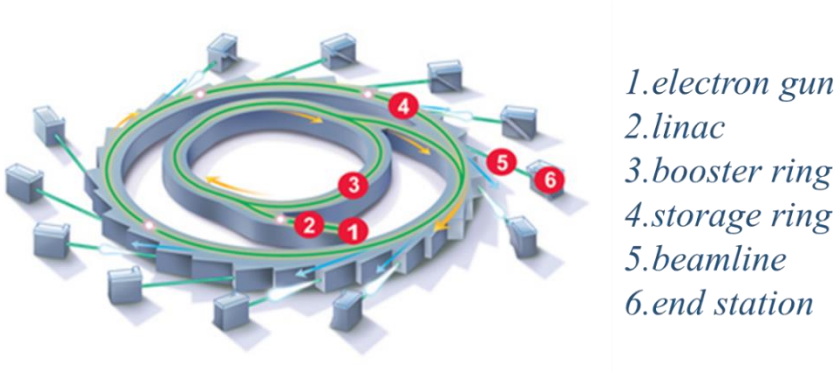


Figure 6: Schematic diagram of the Synchrotron beamline source

Electron guns produce electrons at the centre of the facility and a linear accelerator or 'linac' accelerates them at the speed of about 299792 $kmsec^{-1}$ (99.9997% percentage of light speed). The generated electrons travel to the booster ring in a circular trajectory and within a half of a second, increase the energy level up to 3000 MeV from 100 MeV. A storage ring with 216

meters circumference is in action for storing these high energetic electrons. A series of magnets (bending magnets, unduators and multipole wigglers) are incorporated in the storage ring, separated by long straight sections so that the created magnetic field can cause the electron deflection. Therefore, the electrons release electromagnetic radiation and each bending magnet produces a beam of synchrotron light. The released electromagnetic radiation is emitted in the forward direction in narrow cone maintaining the tangent of electron's orbit. Infrared and x-ray radiation created by synchrotron are channelled down as 'beamlines' and ending up to the work-stations.[9]

2.3.3 Synchrotron IR source

Synchrotron infrared radiation is 100 to 1000 times intense than the conventional globular sources, when the aperture size is less than 20 µm. By providing high spectral resolution, in mid IR range even up to 3 micron, the highly collimated and polarised infrared beam offers the scope to analyse the sample components in microscopic level, including the production of chemical images through mapping of diverse materials such as biological tissues, surface coatings and composite materials.[8]

2.3.4 Synchrotron IR micro-spectroscopy

The combination of a Bruker VERTEX 80v FTIR spectrometer with Hyperion 3000 IR microscope allows a collimated beamline of synchrotron IR source and increases the signal to noise ratio at diffraction limited spatial resolutions in the range of 3-8 µm. Both the collimated synchrotron IR and conventional thermal IR lights follow the same beam path. Therefore, only a flat mirror needs to be incorporated in a commercially available microscope to shift between the synchrotron-light from the traditional globar source. Hence, the collimated synchrotron IR beam at first enters the FTIR spectrometer and then directed towards the microscope. The detectors used in the infrared microscopes are usually highly responsive narrow- band or broad-band mercury cadmium telluride (MCT) detectors, cooled by using liquid nitrogen. In the Australian synchrotron IR facility, the online Hyperion 3000 IR microscope is equipped with a single-point MCT detector.[10]

2.4 Atomic Force Microscope Infrared-spectroscopy (AFM-IR)

2.4.1 Basic theory

Atomic force microscopy is an advanced analytical instrument that functions by interaction of the cantilever tip (probe) with the sample. The cantilever provides a force sensor and a force

actuator. The sample-probe interaction is monitored by a laser beam reflected over the cantilever, thereby, allowing visualization of the sample's morphology when the cantilever scan the surface. A photo detector is present in most instrument to record the vertical and lateral motion of the probe and typically allows calibration of <20 nm. A number of methods that include tunneling current measurement, optical deflection technique, fiber interferometry, and piezo-resistive methods are used to achieve nanometer scale accuracy.[11]

A much more enhanced spatial resolution can be achieved by combining AFM's cantilever oscillation with infrared beam. On introduction of IR laser on the sample, an induced photo-thermal expansion is observed which enables the cantilever probe to act as a detector, as the oscillation is proportional to the IR absorption.[12,13]

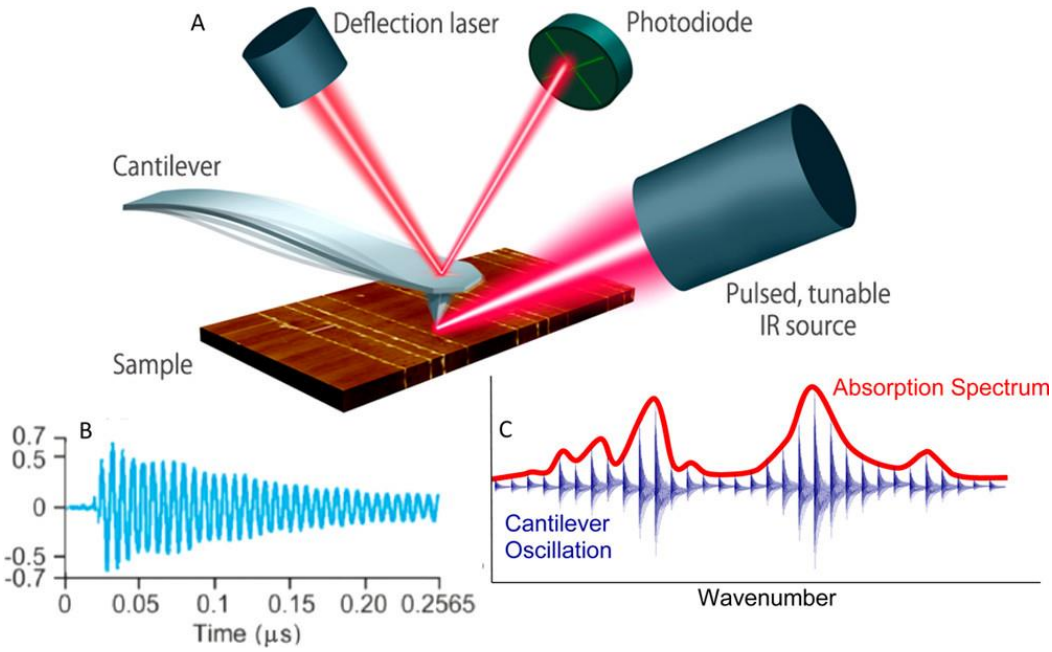


Figure 7: Schematic diagram of AFR-IR A) and the cantilever oscillation B) is proportional to IR absorption spectrum C)[12]

Atomic force microscopy can sense the mechanical interaction, which interacts at atomic level. When the AFM tip is located in a close proximity (0.1 -100nm) of the sample surface, the existing atomic scale forces are called attractive and repulsive forces, and can be detected by the AFM tip. The attractive forces are characterized as electrostatic interaction, Van der Waal's force and chemical force. The repulsive forces are usually short-range forces, and can be categorized as Coulomb interaction, Pauli-exclusion interaction and hard sphere repulsion.[14]

2.4.2 Major components of the AFM-IR spectrometer

An Atomic force microscope- infrared spectroscopy (AFM-IR) system has four major components, a microscope, a cantilever probe, a pulsed tunable infrared (IR) source and a photodetector.

During data acquisition, a sharp AFM-tip is faced towards the surface of the sample, and the tip is attached with a long, thin, flexible micro-fabricated Silicon structure, named as cantilever probe. This flexible probe allows AFM to generate a topographic image by scanning back and forth over the sample surface area. After introduction of the tuned IR-source, the sample excitement leads to a molecular absorption and a rapid thermal expansion of that sample area. That expansion is detectable by the cantilever probe and the resulted cantilever oscillation is correlated to the conventional FTIR-spectra. The photodetector detects the cantilever deflection, when a red laser beam coming from a laser diode (photodiode) hits the back of the reflective cantilever and bounces back up to the detector. The primary signal generates by AFM-IR is usually called the 'deflection signal'. This high-resolution measurement (deflection signal) can be interrupted by loud noises, strong air currents, vibrations or by other effects, as it operates for relatively longer period of time related to the relative positions of the tip over the sample. The AFM-IR microscope also has an attached IR-source, which covers the mid-IR region from 3600-900 cm^{-1}. Some optics are active to direct the IR-radiation towards the probe position in the AFM system. Along with the optics, the electronics and an infrared power meter are also functional during the microscope operation. A bright field optical microscope is used to locate the probe on to the sample.

2.4.3 Modes of AFM-IR operation

Contact and tapping modes are two basic modes for operating atomic force micro-spectroscopy. During the contact mode, probes contact with the sample surface and scanning back and forth to generate the imaging signal. The deflection signal varies with the bending of cantilever into various directions. As it bends upward, an increase of the signal can be detected by the photodetector. So, a pre-set increased signal can be achieved by pressing the probe with enough pressure into the sample surface. With the moving probe over the sample, the cantilever bends according to the topography of the surface. If the cantilever's height (Z position) is constant, the deflection (cantilever's angle) will be changing continuously and with an adjusted cantilever's height, the Deflection signal will be at "the set point". The force applied between the probe and the sample is directly proportional to the set point, which means more force should be applied to achieve a larger set point.

Another AFM height-imaging mode is tapping mode. In this mode, the tip and the sample usually uses smaller forces compared to the contact mode, where a mechanical oscillation of the probe creates an AC Deflection signal categorized by its amplitude. When the tip is tapping on the surface, if the cantilever is closer the amplitude will be smaller because of the less swing, and if the cantilever is far enough from the sample surface, the amplitude will be larger because of more swing. It changes according to the sample surface, where tall surface generates smaller amplitude and low surface generates larger amplitude.

2.5 Multivariate Data Analysis

2.5.1 Principal Component Analysis

Principal Component Analysis or PCA is an unsupervised method which primarily decomposes a data matrix into signal and noise parts. A data matrix (X) is comprised of objects (n) and variables (p), where 'objects' belong to samples, observations or a number of experiments and 'variables' are simply the measurements obtained from objects. The key purpose of all multivariate data analysis is to find "hidden phenomena" and it is a fundamental assumption that structure portion of a dataset is correlated with the phenomena. By using a PCA model, the original dataset can be decomposed into orthogonal components known as principal components (PCs), latent variables or factors. The maximum number of components that can be decomposed is N-1, where N represents the total number of objects or spectra. The 1st PC is derived along the direction in the data where maximum variance lies. "Maximum variance" is a term which defines the largest variance of objects within a dataset. The 2nd PC will account for the second largest variance and so forth.

A PCA model is interpreted by examining two plots, score plot and loading plot.

2.5.1.1 Scores plot

The purpose of a scores plot is to map the samples according to their similarities and dissimilarities. The scores plot is obtained typically by plotting two pairs of score vectors against each other, where score vectors act as a snapshot of the objects and projected through the PCs. In a multivariate approach, the most common plot is PC1 vs PC2. Grouping the samples, identifying the outliers and trends, finding out similarities and dissimilarities are the goals for using a scores plot.

2.5.1.2 Loadings Plot

In terms of spectral data, loadings represent the spectral features responsible for the sample clustering. It is the mapping of the variables onto the PCs and shows the contribution of each variable to construct each PC.

2.5.1.3 PCA modeling

A PCA model is basically comprised of four components, which are the dataset, scores, loadings and residuals. The X- data matrix can be decomposed in a PCA model as follow:

$X=TP^T + E$

TP^T is the signal part of X data and E is the noise part. T simply describes the score matrix and P^T describes the loadings in the transpose form. E is the "residual matrix" and is not explained by the model.[15]

2.5.2 Partial Least Squares Regression (PLS-R)

PLS-R is a linear regression technique for predicting the dependent or response variables from a huge set of independent variables or predictors. The prediction depends on a range of orthogonal factors known as "latent variables (LVs)". LVs are extracted from the dataset (predictors) and contain the best prediction power. Multivariate methods have adopted PLS regression as a tool for analyzing experimental and non-experimental dataset in a supervised way. The purpose of this model is to predict Y (dependent variable) from X (independent variables) and extracting their common features. In contrast, PCA only decomposes the X and extract the components (PCs) best describing the X, PLS-R searches for PLS-components (latent vectors) from X which can predict Y the best. These components are capable of decomposing both X and Y in a simultaneous manner. The goal of PLS-R is to use the predicted values for new observations derived from the same population used to construct the model. X has an influence on Y variables and PLS helps to understand that influence by developing a model for doing prediction.

2.5.2.1 PLS-R modelling

If *N* is the large number of objects for X and Y matrices, *K* and *M* are the moderate number of variables for X and Y respectively PLS-R can be modeled as follows:

$Y=X\,[B]+[F]$,

Where B is the regression coefficient matrix of size $(K \times M)$ and F is the residual matrix of size $(N \times M)$. The regression coefficient is expressed as,

$B=W(P^TW)^{-1}C^T$

Here W is the X weights, P is the X loadings and C is the Y weights.[16–18]

2.5.3 Partial Least Square Discriminant Analysis (PLS-DA)

Partial Square Least Discriminant Analysis (PLS-DA) is a multivariate data analysis discrimination tool usually employed to separate two groups. In PLS-DA, PLS is used for discrimination where PLS (Partial Least Squares) and LDA (Linear Discriminant Analysis) are correlated to each other. It is a supervised method for developing a model to classify samples. If a data matrix contains only two groups of samples it is easy to predict the group by setting the y variable 1 or 0. If 1, has been set up for positive samples, 1 will represent the positive group and 0 for negatives.

A threshold is calculated called the Bayesian threshold, which minimizes the error (false positive or false negative) and assuming that they value prediction will be similar for the future observations. Another method can be used to evaluate the PLS-DA models known as Receiver Operating Characteristics Curve or ROC curve. It is a graphical representation of sensitivity and specificity values of the model. In ROC curve X-axis denotes for 1-specificity (false positive rate, FPR) and Y-axis represents sensitivity (true positive rate, TPR). The TPR can be obtained by the number of true positive results achieved by the model divided by the number of actually true cases, similarly, the number of false positive results divided the number of positive cases states the false positive rate. The obtained values for TPR and FPR can be 0 to 1, where a value closer to 1 for TPR and a value closer to 0 for FPR indicates a better model with a better prediction power. The assessment of a ROC curve also depends on area under the ROC curve (AUROC). By analogy with the TPR and FPR values, AUROC value can be between 0 to 1, where a value closer to 1 indicates more area has been covered by the curve means better accuracy of the model.[19,20]

2.6 References

1 P. W. Atkins and J. De Paula, *Atkins' Physical chemistry*, Oxford University Press, Oxford,UK, 9th edn., 2010.

2 M. M. Mossoba, J. K. G. Kramer, H. Azizian, J. Kraft, P. Delmonte, A. R. F. Kia, F. J. Bueso, J. I. Rader and J. K. Lee, *JAOCS, Journal of the American Oil Chemists' Society*, 2012, **89**, 419–429.

3 C. V. Raman and K. S. Krishnan, *Nature*, 1928, **121**, 501–502.

4 G. S. Bumbrah and R. M. Sharma, *Egyptian Journal of Forensic Sciences*, 2016, 6, 209–215.

5 E. Smith and G. Dent, *Modern Raman Spectroscopy-A Practical Approach*, John Wiley & Sons, Ltd, Chichester,United Kingdom, 2005, vol. 1.

6 J. M. Chalmers, H. G. M. Edwards and M. D. Hargreaves, *Infrared and Raman spectroscopy in forensic science*, John Wiley & Sons Ltd., Chichester,United Kingdom, 1st edn., 2012.

7 T. Dieing, O. Hollricher and J. Toporski, Eds., *Confocal Raman Microscopy*, Springer, Berlin, Heidelberg, 2011, vol. 158.

8 L. M. Miller and P. Dumas, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2006, **1758**, 846–857.

9 G. LeBlanch, *The synchrotron light source*, 2008.

10 Synchrotron | Infrared microspectroscopy | ANSTO, https://www.ansto.gov.au/user-access/instruments/australian-synchrotron-beamlines/infrared-microspectroscopy, (accessed November 17, 2018).

11 Y. Seo and W. Jhe, *Reports on Progress in Physics*, 2008, **71**, 23.

12 A. Dazzi and C. B. Prater, *Chemical Reviews*, 2017, **117**, 5146–5173.

13 C. Marcott, J. F. Rabolt, C. B. Prater, A. Dazzi, Q. Hu and D. B. Chase, *Applied Spectroscopy*, 2012, **66**, 1365–1384.

14 F. J. Giessibl, *Reviews of Modern Physics*, 2003, **75**, 949–983.

15 L. P. H. Kim H. Esbensen, Dominique Guyot, Frank Westad, *Multivariate data analysis : in practice : an introduction to multivariate Data Analysis and Experimental Design*, CAMO, Oslo,Norway, 5th edn., 2002.

16 F. Lindgren, P. Geladi and S. Wold, *Journal of Chemometrics*, 1993, **7**, 45–59.

17 R. Manne, *Chemometrics and Intelligent Laboratory Systems*, 1987, **2**, 187–197.

18 H. Abdi, *Wiley Interdisciplinary Reviews: Computational …*, 2010, **2**, 97–106.

19 S. H. Park, J. M. Goo and C. H. Jo, *Korean Journal of Radiology*, 2004, **5**, 11–18.

20 M. Barker and W. Rayens, *Journal of Chemometrics*, 2003, **17**, 166–173.

Chapter 3. Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide

analytical chemistry

Article

pubs.acs.org/ac

Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide

Supti Roy,[†] David Perez-Guaita,[*,†] Dean W. Andrew,[†] Jack S. Richards,[‡,§,∥] Don McNaughton, Philip Heraud,[*,†] and Bayden R. Wood[*,†]

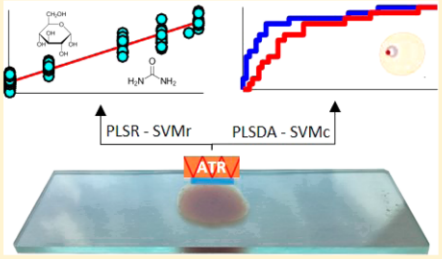
[†]Centre for Biospectroscopy, Monash University, Clayton, 3800, Victoria, Australia

[‡]Centre for Biomedical Research, Burnet Institute, Melbourne, 3004, Victoria, Australia

[§]Department of Medicine, University of Melbourne, Parkville, 3052, Victoria, Australia

[∥]Infectious Disease Department, Monash University, Melbourne, 3004, Victoria, Australia

ABSTRACT: New diagnostic tools that can detect malaria parasites in conjunction with other diagnostic parameters are urgently required. In this study, Attenuated Total Reflection Fourier transform infrared (ATR-FTIR) spectroscopy in combination with Partial Least Square Discriminant Analysis (PLS-DA) and Partial Least Square Regression (PLS-R) have been applied as a point-of-care test for identifying malaria parasites, blood glucose, and urea levels in whole blood samples from thick blood films on glass slides. The specificity for the PLS-DA was found to be 98% for parasitemia levels >0.5%, but a rather low sensitivity of 70% was achieved because of the small number of negative samples in the model. In PLS-R the Root Mean Square Error of Cross Validation (RMSECV) for parasite concentration (0−5%) was 0.58%. Similarly, for glucose (0−400 mg/dL) and urea (0−250 mg/dL) spiked samples, relative RMSECVs were 16% and 17%, respectively. The method reported here is the first example of multianalyte/disease diagnosis using ATR-FTIR spectroscopy, which in this case, enabled the simultaneous quantification of glucose and urea analytes along with malaria parasitemia quantification using one spectrum obtained from a single drop of blood on a glass microscope slide.



INTRODUCTION

Malaria is one of the most deadly vector-borne diseases afflicting humanity. It is caused by *Plasmodium* spp. parasites, which are transmitted via the bite of female *Anopheles* mosquitoes. *Plasmodium falciparum* and *Plasmodium vivax* are the major species afflicting humans, resulting in approximately 200 million infections annually and up to 1 million deaths.[1] In 2015, 90% of malaria related deaths occurred in sub-Saharan Africa and most were children under 5 years of age with *Plasmodium falciparum* infection being the main infectious agent.[2]

Hypoglycemia is defined as an abnormally low concentration of blood sugar, and the current threshold level determined by WHO is less than 2.5 mmol/L (45 mg/dL) for a properly nourished infant, 2.2 mmol/L (40 mg/dL) for neonates, and <3 mmol/L (54 mg/dL) for children having severe malnutrition.[3] Hypoglycemia is a well-described metabolic complication of "severe malaria", especially with *Plasmodium falciparum* infection. Children under the age of 5 and pregnant women are the most vulnerable to hypoglycaemia following *Plasmodium falciparum* infection.[4,5] The risk increases when more than 2% of red blood cells (RBCs) become infected.[6]

Malaria-associated hypoglycemia is more common in children compared to adults[6] and it seems to be an independent risk factor for child mortality due to the disease.[7,8]

The complication can impose adverse consequences like neurological sequelae or brain damage and, in children having severe malaria, it has been correlated with a high death rate in Gambia (33%) and Malawi (36.8%).[9,10] Glucose metabolism differs according to age and in severe conditions of malaria, hypoglycemia is a common feature in children prior to receiving any treatment.[6] The cause of hypoglycemia in children is multifactorial but may include: increased metabolism, higher glucose clearance rate[11] and low glucose intake due to malnutrition and loss of appetite.[12] In adults, low blood glucose is due to increased peripheral glucose uptake.[13] Hypoglycemia can also occur in patients who are not undergoing drug treatment. With the progression of malaria, glucose demand can rise by around 50% and glucose production by the host can become inadequate to fulfill the increasing demand. Glucose clearance rate has also been observed to increase 40%−70% above normal levels in severe malaria.[13−15] During pregnancy, the pathophysiology of hypoglycemia is poorly understood. *Plasmodium falciparum*

Received: November 20, 2016
Accepted: April 14, 2017
Published: April 14, 2017

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

infected (nonsevere malaria) pregnant women produce more glucose compared to healthy controls, suggesting that the hypoglycemic condition is a consequence of quinine treatment or prolonged fasting.[5,16]

Malaria interferes with cerebral, hepatic and renal function.[17] In *Plasmodium falciparum* associated severe malaria, hyperparasitemia is a contributory factor for the pathogenesis of renal dysfunction.[1] A high load of parasites is a strong determinant for causing hepatic and renal abnormalities.[18] Some studies[19−21] have reported significantly high levels of urea in serum from malaria infected individuals compared to healthy controls. Higher than normal reference values of both serum creatinine and urea are proven indicators of renal impairment.[22,23] The higher urea concentration in serum can be attributed to the excessive destruction of RBCs by the protozoan parasites during their intraerythrocytic life cycle, during which the digestion of globin protein derived from hemoglobin ultimately leads to the production of excess urea.[20]

The application of ATR-FTIR spectroscopy for blood analysis shows great potential for the diagnosis of several diseases[24] and the prediction of clinical parameters,[25] which can be used to assess the metabolic state of the patient. During the parasitic life-cycle, *Plasmodium* spp. go through several sexual and asexual stages. From a diagnostic point of view, it is necessary to detect the parasites at the early phases of the erythrocytic life cycle. Using a combination of chemometrics and ATR-FTIR spectroscopy our group has unequivocally diagnosed and quantified early ring-stage and gametocyte-stage parasites in spiked blood samples.[26,27] In a laboratory study, early ring-stage parasites were cultured in a normal RBC suspension and the absolute sensitivity was found to be less than one parasite per microliter of blood ($P < 0.008$),[27] whereas in a recent pilot study (workshop held on October 15−16, 2015, Robert Koch Institute, Berlin, Germany) conducted in Thailand by our group, we have analyzed over 300 patient samples with an extremely high sensitivity and specificity (>95%).[26]

Previous reports indicate that glucose and urea are in sufficient enough concentrations to enable them to be recognized components in high-quality ATR spectra from blood.[25] Both analytes have been quantified in plasma,[28] whole blood,[29] and serum[30] using ATR-FTIR spectroscopy of wet samples. Although fast and simple, this approach requires thorough cleaning of the liquid ATR cell and subtraction of the liquid water spectral component from the infrared spectra. Transmission measurement of the dry samples on transparent IR substrates has been also proposed.[31] This method eliminates the water contribution and preconcentrates the proteins and metabolites of the serum. Drying the sample onto an ATR crystal is also impractical because of the extensive time required to dry each sample to a consistent level, record the measurement and then clean the crystal. In this paper we used a different approach; 3 μL of whole blood was dried on a commercial glass slide as a thick film and was then measured using ATR-FTIR spectroscopy by inverting the slide onto the crystal. This method entails drying several samples at once providing rapid spectral acquisition time with minimal contamination of the ATR crystal.

From a theoretical point of view, the IR spectrum of blood contains information about a wide range of blood analytes as well as biomolecules from the cells themselves. Metabolic changes produced by the *Plasmodium* spp. could interfere in the prediction of clinical parameters and vice versa. The aim of this study was to investigate the ability of ATR-FTIR spectroscopy to simultaneously diagnose malaria, glucose, and urea from a single spectrum recorded from a drop of blood on a glass microscope slide. This is a pilot study on spiked whole blood samples, but it may lead to a simple point-of-care test that would greatly assist in the clinical management of patients presenting with malaria.

■ METHODOLOGY

This work was granted ethical approval by the Human Low Risk Review Committee at Monash University.

Parasite Cultures. *Plasmodium falciparum* D10 strain was cultured in human red blood cells provided by the Australian Red Cross. Infected blood was grown at a 5% hematocrit in RPMI 1640 media supplemented with AlbuMAX II (Gibco). Parasites were synchronized during the ring stage with 5% D-sorbitol and cultured until a 10% parasitemia was achieved.

Sample Spiking. Whole blood was obtained from different malaria-uninfected volunteers ($N = 6$) in lithium heparin anticoagulant tubes. For each biological sample, 12 aliquots were obtained and spiked with varying concentrations of *Plasmodium falciparum* (0−5%), giving a total of 72 samples including controls. A total of 1% of parasitemia corresponds to 50000 parasites per microliter. Some of the samples spiked with parasites ($N = 30$) were divided into three aliquots (60 new samples), and each aliquot was subsequently spiked with a different value of glucose (0−400 mg/dL) and urea (0−250 mg/dl) using D-glucose anhydrase (Ajax Chemicals) and urea ACS reagent (Sigma-Aldrich) as the standard stock solutions. Concentration levels were randomly assigned to different aliquots of each blood sample. The initial concentration of the analytes in the original samples were unknown. In all cases the volume added was less than 5% of the sample volume. The resulting sample set contained 132 samples with different amounts of *Plasmodium falciparum*, glucose, and urea.

Sample Measurement. Measurements were performed on a Spectrum 2 spectrophotometer from PerkinElmer (Waltham, U.S.A.) using 4 cm^{-1} spectral resolution with 16 coadded interferograms. Whole blood samples were lysed by freeze-thawing without the addition of water or any lysing agents. A total of 3 μL of blood was deposited onto a glass microscope slide, with a thick smear formed by using a pipet tip to spread the blood onto the glass slide. After allowing the blood to dry for 5 min, the thick smear was placed upside-down onto the diamond ATR-crystal, the clamp was actioned and the pressure was adjusted to 100 units according to the pressure sensor of the ATR. This provided homogeneous pressure on the ATR crystal before each measurement. In less than 5% of the samples we observed some contribution from the glass slide in the spectrum, which were consequently rejected. Each spectrum was obtained in approximately 1 min. A background of the empty clean ATR crystal was obtained using 16 coadded interferograms. Each thick smear was analyzed in triplicate and the spectra were averaged.

Data Analysis. ATR-FTIR spectra were preprocessed and analzed by MATLAB 8.6 Release2015b from Mathworks (Natick, U.S.A.) using the functions available in the PLS toolbox from eigenvector (Manson, U.S.A.). Several models were created using different preprocessing (Savitzky-Golay derivatives, standard normal variate, mean centering, and Pareto scaling) and various combinations of spectral regions. The optimal model was selected based on the minimal cross validation error obtained from a maximum of 10 latent variables

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

(LVs). Partial Least Squares-Discriminant Analysis (PLS-DA) was performed to calculate the presence of malaria parasitemia, and PLS regression (PLS-R) was used for quantifying the parasite percentage along with the concentration of glucose and urea in the samples.

■ RESULTS AND DISCUSSION

Spectra of Whole Blood Thick Smears on Glass Slides. Although ATR-FTIR spectroscopy measurement of a whole blood thick smear on a glass slide is a fast and simple method and greatly facilitates the transport and storage of blood samples spectral contributions from silicon−oxygen bonds from the glass can interfere with the diagnostic performance when the blood film is too thin. Figure 1 shows the IR spectra

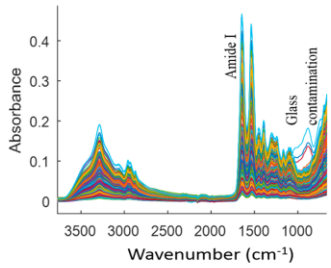


Figure 1. Representative spectra recorded of thick smears of whole blood on glass slides showing how glass contamination can manifest in the spectra if the blood film is too thin. The spectra show a large range of absorbance values indicating significant variability in the blood film thickness.

of the thick smears acquired in the study. It can be seen that a spectral contribution from the glass was detected only in three cases from the 400 measurements (132 samples, 3 replicates) shown, which manifested as a broad band between 1040 and 870 cm^{-1}. This indicates that, although only 3 μL of blood was used, the thickness of the smear exceeded the penetration depth of the infrared beam on the ATR crystal in the majority of samples. Figure 1 also shows the variability of absorbance between the spectra. Absorbance values of amide I at 1650 cm^{-1} ranged from 0.02 to 0.45. This wide variation is probably due to the irreproducibility of the drying process, which could lead to an irregular smear with air bubbles inside, indicating that normalization is a prerequisite prior to modeling. The large intensity of the amide II mode at ∼1550 cm^{-1} is due to contributions of urea superimposed onto the amide II mode.

Overlaid Mean Spectra. Second derivative, vector normalized spectra of whole blood dried on a glass slide from control and infected (ring stage *Plasmodium falciparum*) spiked samples are shown in Figure 2A. Figure 2A shows the C−H stretching region (3100−2800 cm^{-1}), amide I, amide II, and fingerprint region. The C−H vibrations from lipid moieties are more intense in infected samples in agreement with earlier studies.[27,32] Bands at 2934 and 2851 cm^{-1} are assigned to symmetric/asymmetric stretches of − CH_2 acyl chains lipid.[32,33] In contrast, bands at 2959 and 2871 are assigned to CH stretching vibrations, which contain strong contributions from proteins and are similar intensity for controls and high parasitemia samples. The observation of lipid signatures in the spectra of infected-samples suggests the formation of fatty acids during the parasite development. Average spectra from infected-samples show the amide I and II bands to be less intense, indicative of hemoglobin digestion by the malaria parasites. There is a distinct band at 1043 cm^{-1} observed in the spectra of positive cells that is not correlated with any other band and, therefore, at this stage, remains unassigned but is most likely a C−O stretch from a sugar moiety. The presence of this distinct band in the fingerprint region is a potential marker for early ring stage parasites.

Second derivative spectra of glucose and urea spiked samples and the corresponding standard deviation spectra are shown in Figure 2B. The bands coming from glucose (1031, 1014, 1080, and 988 cm^{-1}) and urea (1161 and 783 cm^{-1}) can be seen in the vector normalized mean spectra of spiked samples (with high concentrations of glucose and urea). The band at 1591 cm^{-1} is assigned primarily to $\delta(NH_2)$ based on spectra recorded of O-alkyl derivatives, which do not have a carbonyl group but do have a band between 1610 and 1600 cm^{-1}.[34] In α-D-glucose characteristic valence vibrations from CC and CO bonds can be observed in the 1070−970 cm^{-1} region.[35] In urea, the band at 1152 cm^{-1} (1161 cm^{-1} in the average spectra of spiked samples) corresponds to the NH_2 rocking vibration.[36,37] A band located at 783 cm^{-1} can be assigned to the bending vibration due to the N−C−N deformation mode in urea.[37]

Discrimination between Control and Highly Infected Samples. Partial Least Squares Discriminant Analysis (PLS-DA) is a multivariate analysis tool for developing a model to discriminate between two or more groups. The model was constructed by recording spectra of *Plasmodium falciparum* spiked whole blood samples with more than 0.5% of parasitemia as positive (N = 59) and the samples without parasitemia as negative (N = 15). The data was preprocessed by taking the first derivative (11 smoothing points), and preprocessing with the standard normal variate, Pareto scaling, and mean centering (Table 1) and the selected model was based on nine latent variables, using the 2999−2770, 1784−1589, 1469−1355, and 1185−1006 cm^{-1} regions of the spectra.

For a proper evaluation of the prediction capability of the model a Receiver Operation Characteristic (ROC) curve was generated (See Figure 3A). This curve is a way of plotting the sensitivity and specificity of a diagnostic model where different values are selected for the threshold. In a ROC curve, the x-axis represents the false positive rates (1−specificity) and y-axis contains the value of a true positive rate (sensitivity). True positive rate (TPR) can be calculated by the number of true positive decisions obtained by the model divided by the number of true actual cases. Similarly, the number of false positives divided by the number of positive cases defines the false positive rate (FPR). It is well understood that, as true positivity increases and false positivity decreases, the reliability of the model improves. Due to the biological variability of the diseased state, it is impractical to define the performance of a diagnostic approach simply using sensitivity and specificity values alone. In many diagnostic cases, it is impossible to fit the test results within only two obvious outcomes, either positive or negative. To overcome these limitations, the ROC curve has become a powerful tool to analyze such data. Another important factor related to ROC curve is the area under the ROC curve (AUROC), which is used to summarize the overall performance of the test method. AUROC values can be 0 to 1, where a value closer to 1 is defined as better overall accuracy of the test. It is possible to compare the diagnostic capability among different techniques by comparing the corresponding values of AUROCs.[38] From the leave-one-out cross validation

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

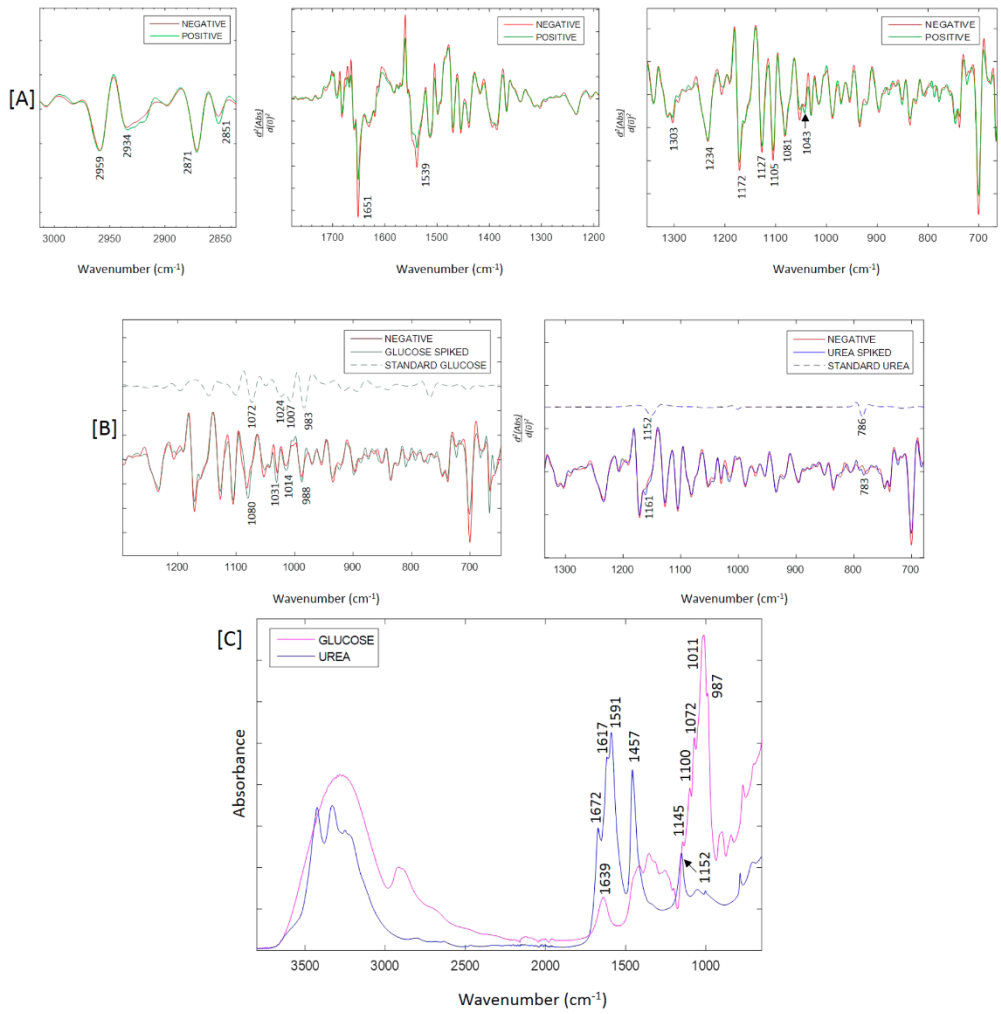


Figure 2. (A) Overlaid mean normalized second derivative ATR-FTIR spectra obtained from control and spiked infected (with 5% parasite concentration) samples. (B) Normalized average second derivative spectrum of spiked sample with high concentration of glucose and urea along with the standard. (C) Standard absorbance spectra of urea and glucose.

in our case the AUROC was found to be 0.89, indicating a good prediction capability. Furthermore, the error, specificity, sensitivity for the optimum threshold was found to be 9.5%, 98%, and 70%, respectively. The sensitivity of the model is rather low because of the low number of negatives used to build the model.

The regression plot, represents the contribution of each spectral band to the classification between the positive and negative malaria samples. The data was preprocessed using the first derivative, and hence, the regression vector bands also present as a derivative shape, making the band assignments difficult. In Figure 3B we show the integration of the regression vector, calculated as the cumulative sum using the *cumsum* function from Matlab. This transformation allows a straightforward interpretation of the regression vector bands, with bands assigned to the analyte pointing to positive values and bands associated with the absence of the analyte pointing to negative values. In the PLS-DA model the major minima bands are associated with the infection (Y values were set to 1 for the controls in the PLS-DA). The assignments of the bands responsible for the discrimination are summarized in Table 2. The bands in the C−H stretching region located at 2935 and 2883 cm^{-1} are attributed to *Plasmodium falciparum* neutral lipids that are synthesized during the parasite's development and deposit in the digestive vacuole, ultimately changing the lipid layer composition of infected RBCs. The sensitivity of

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

Table 1. Main Parameters and Numerical Results for the Models Developed in This Study[a]

analyte	model	preprocessing	regions (cm^{-1})	LV	RMSECV	RPD	CV classification error (%)	specificity (%)	selectivity (%)	AUROC
plasmodium	PLS-DA >0.5%	FD (11)	2770−2999	9			9.5	98	70	0.89
		SNV	1589−1784							
		PS	1355−1469							
		MC	1006−1185							
parasitemia	PLS-R	FD (11)	2770−2999	9	0.589%	2.77				
		SNV								
		MC	1006−1185							
urea	PLS-R	FD (15)	1355−1588	10	22.70 mg/dL	4.07				
		SNV								
		MC	1185−1006							
glucose	PLS-R	FD (15)	700−1354	10	30.48 mg/dL	4.51				
		SNV								

[a]FD: first derivative (smoothing points); MC: mean centering; SNV: standard normal variate; PS: pareto scaling; LV: latent variables; RMSECV: root mean square error of cross validation; RPD: ratio prediction to deviation; AUROC: area under the receiver operating characteristic curve.

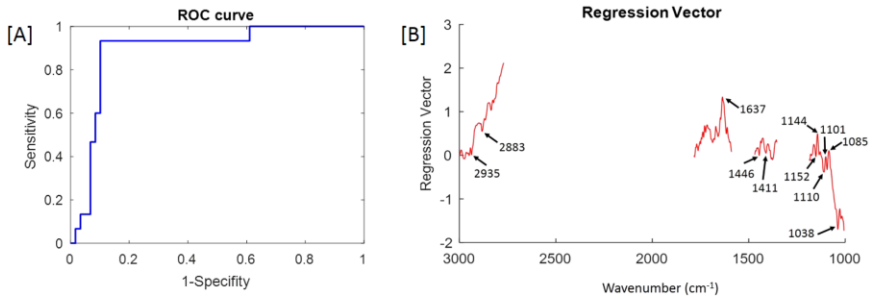


Figure 3. Graphical results of PLS-DA for malaria diagnosis. Left panel (A) shows the receiver operation curve for the cross validation for samples with low or high parasitemia, and right panel (B) shows the cumulative sum (integration) of the regression vector of the discrimination of the control samples (bands associated with the parasite pointing down).

Table 2. Major Band Positions Observed from Regression Vectors along with Their Assignments

ATR band positions;[33,35,37] wavenumber cm^{-1}	assignments
2977	$\nu_{as}CH_3$ acyl chains lipid
2935	$\nu_{as}CH_2$ acyl chains lipid
2883	$\nu_{s}CH_3$ acyl chains lipid
1637	amide I; proteins, β-pleated sheet
1446	proteins, lipids, δCH_2
1411	νCOO_2^- of fatty acids and amino acid side chains
1110	RNA, ribose ν(C=O)
1038	RNA, ribose ν(C=O)
1144, 1101, 1085	valence vibration of CC and CO in glucose
1591	bending NH_2 vibration in urea
1152	NH_2 rocking vibration in urea
783	N−C−N deformation

ATR-FTIR spectroscopy is very high for lipid molecules, nucleic acids, proteins, and carbohydrate vibrations. The complex spectral region from 1250 to 800 cm^{-1} is very informative for DNA/RNA vibrations and absorption of sugar molecules.[39] The presence of minima bands at 1152, 1110, and 1038 cm^{-1} in the regression vector (Figure 3B) suggests the presence of parasite-related RNA vibrations. During the developmental phases, the protozoan parasites produce structurally and functionally unique stage specific ribosomes. Ribosomes start to proliferate along with protein synthesis after invasion of parasites into the host erythrocytes. Most of the ribosomes form during the early ring-stage of development.[40,41] Strong positive bands in the carbohydrate region indicate the existence of a large quantity of glucose in the normal erythrocytes, which is consistent with other studies because *Plasmodium* infected RBCs have a 30−100× higher rate of glucose consumption compared to normal cells.[6,42] The intense amide I band located at 1637 cm^{-1} associated with the control group indicates that this group has a higher hemoglobin content than the infected group, which of course is to be expected given that the majority of hemoglobin is catabolized by the parasites in the infected samples. Bands at 1446 and 1411 cm^{-1} are assigned to bending vibrations of methylene group (lipids and proteins) and carboxylate stretching from fatty acid/amino acid side chains,[33] are assigned to parasitic lipids and proteins.

PLS-R Models. Partial Least Squares Regression (PLS-R) was applied to assess the prediction capability of ATR-FTIR spectroscopic analysis of thick smears for parasitemia, glucose and urea in whole blood samples. PLS-R is a weighted linear regression method where the modeling is based on finding linear relationships between observed and predicted variables. For spiked concentrations of parasitemia (0−5%), glucose (0−400 mg/dL), and urea (0−250 mg/dL), different PLS-R models were created using combinations of regions and

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

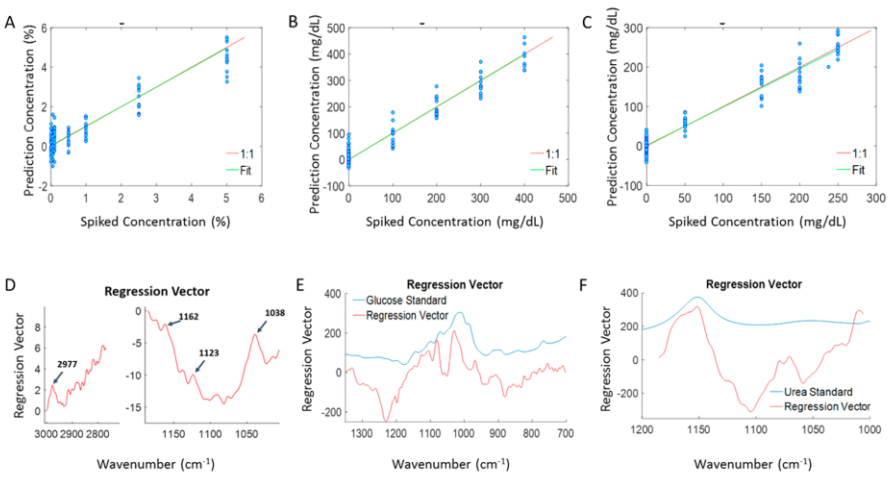


Figure 4. PLS-R results for modeling the spiked concentration. The upper panels A, B, and C indicate the spiked vs predicted concentration for each sample for parasitemia (0−5%), glucose (0−400 mg/dL), and urea (0−250 mg/dL), respectively. Lower panels D, E, and F show the corresponding regression vectors for parasitemia, glucose, and urea, respectively. In the case of glucose and urea, the regression vector is compared with spectra of glucose and urea standards recorded under the same conditions. The regression vectors have been integrated using *cumsum* (cumulative summation) function in MATLAB.

preprocessing and validated using venetian blinds (seven splits) cross validation.[30,43] The best models were selected according to the best CV error in a reasonable number of Latent Variables (LVs). The spectral range was selected in the same way as we selected the parameters for the PLS-R of the quantifications. Several models were generated by combining different preprocessing algorithms and spectral regions. Then the models were sorted according to the CV classification error. We selected the model that provided both low error with a reasonable a number of LVs. Table 1 summarizes the regions and preprocessing steps selected for each analyte. In all cases standard normal variate (SNV) normalization, first derivative processing and mean centring were used for obtaining the best results. To quantify malaria parastemia with PLS-R, the regions of 2999−2777 cm^{-1} and 1006−1185 cm^{-1} were used in the decomposition, which have strong contributions from the lipid (CH) and DNA (PO_2^-) molecules. In the case of glucose modeling, the region selected was the C−O stretching region 1300−1000 cm^{-1} assigned to glucose hydroxyl groups. The urea model was based on the 1355−1588, 1185−1006 cm^{-1} regions where the CN and NH groups of urea show strong absorbance bands (see spectrum of Figure 2C).

The three models show the ability to predict the analytes under investigation. The corresponding values of the Root Mean Square Error of Cross Validation (RMSECV) and Ratio Prediction to Deviation (RPD) of the models for prediction of parasitemia are 0.58% and 2.77; for glucose are 22.73 mg/dL and 4.07; for urea are 30.48 mg/dL and 4.51. The relative RMSECV (RRMSECV = RMSECV/avg concentration of the data set × 100) for glucose is 16% and for urea is 15%. Root Mean Square Error value is a measurement of the prediction error of a model based on the validation process. In this study, cross validation has been used. In summary, it calculates the deviation between actual and prediction values of the model and is estimated in root mean squared format.[44] Another indicator for measuring the prediction ability of a model is the RPD value. It is the ratio of the standard deviation (SD) obtained from reference values to the RMSECV (RPD = SD/RMSECV). A good-fit model shows a higher value of the RPD, which indicates the error is lower than the standard deviation. The value >2 is good for performing calibration and values greater than 3 mean the model can be used for analysis of a new data set.[45] Figure 4A−C represents the actual versus predicted values for the models. It can be seen that the fit is consistent for a large concentration range.

Regression plots for each of the corresponding models are depicted in Figure 4. Again, models were generated based on derivatives, so the regression vector obtained is a derivative shape. We used the *cumsum* function from MATLAB for integrating the regression vector and facilitating the interpretation of the bands by producing bands in the regression vector plot that emulate absorbance features. The major bands highlighted in Figure 4D play a key role in generating the linearity and are associated with mainly lipid moieties and nucleic acid vibrations form the parasites, which suggests that the model is built on the basis of real changes of spectra and not on other factors like noise or baseline offsets. Within the lipid region a similar value (2977 cm^{-1}) is obtained with the same parasite concentration (0−5%) observed from our earlier study,[27] which demonstrates the importance of the C−H stretching region in *Plasmodium falciparum* detection. Parasite-associated RNA bands (1123 and 1038 cm^{-1}) are also prominent due to presence of the parasites. In Figure 4E,F standard spectra of glucose and urea are overlaid by comparing them with their corresponding regression vectors demonstrating that the good CV error values are based on the detection of glucose and urea within the samples.

Comparison with Other ATR Based Studies. In relation to the detection of quantification of *Plasmodium falciparum*, the method under investigation showed a prediction capability lower than that of the ATR analysis of packed red blood cells fixed in methanol, which has been reported to have a LOD of

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

0.00001% parasitemia.[27] In that study the Root Mean Square Error of Validation (RMSEV) for the 0−5% range was 0.32%, slightly better that the error obtained in this study. This can be explained because the current study was carried out using whole blood, which implies that there is a contribution from serum and other blood cells to the spectra. The metabolites of the serum could act as a confounding variable, especially considering that we added large amounts of glucose and urea to the model, which share spectral regions with those of the parasite, especially in the case of DNA and RNA. Also, rather than methanol fixation unfixed RBC-lysed blood samples were used in this study. Because the models were built with samples spiked with *Plasmodium falciparum*, the model performance obtained only applies for the detection of this species. It is unknown if the methodology will be able to detect and quantify other species such as *Plasmodium vivax* with similar sensitivity but given the molecular phenotype is very similar this should indeed be possible. In the regression vector, IR bands responsible for the detection and quantification were assigned to lipids and nucleic acids, which are compounds present in all of the *Plasmodium* species. Nevertheless, we plan to increase the database to include different *Plasmodium spp.* from clinical samples located in Thailand and Laos.

Enzymatic methods are considered as laboratory standards for measuring the blood parameters (glucose and urea). Among the enzymatic techniques, hexokinase and glucose-oxidase methods are used almost exclusively for glucose measurement, whereas the urease-glucose dehydrogenase coupled enzyme is commonly used for blood urea analysis. In the point-of-care field, the methods have been miniaturized in strips, which allow the measurement within minutes. In previous reports where infrared spectroscopy was used as a point-of-care testing for diagnostic application in terms of measurement blood parameters, it has been found relative RMSEP values for the quantification of glucose and urea in whole blood[29] (5% for both glucose and urea) or serum[30] (14% for glucose and 16% for urea) were lower than the ones obtained in this study. This can be caused by the evaporation of the water, which preconcentrates all of the analytes so the final concentrations being measured are, of course, different to those in the initial stock solutions. In those previous studies, validation was performed by using an independent set of samples, whereas cross-validation has been applied in the present study, which also can be a reason for higher relative RMSECV values. Nevertheless, the reference data is expressed as mg/dL of whole blood (including water) and no internal standard was used to correct for the final concentration values. In addition, the reference data would have contributions from the initial glucose and urea in the blood prior to the spiking, albeit very minor. In general, the models obtained from untreated blood dried on glass performed less accurately than models applied to wet blood or red blood cells fixed in MeOH,

■ CONCLUSION

The ATR-FTIR method has a number of advantages over conventional testing using glucose and urea strips, which are based on an enzymatic test that has been miniaturized into strips. The ATR-FTIR approach requires only a microvolume of whole blood to enable the simultaneously detection and quantification of urea, glucose and parasitemia. The ATR spectrum from a pinprick of blood gives information both on the presence of the illness and also the metabolic state of the patient. This is achieved without the need of any reagent or strip, just a hand-held battery powered spectrophotometer. The method proposed is more suitable in situations where rapid medical response is needed in remote locations where only 200 μL of whole blood can be collected from a single pinprick with a lancet. The low volume of whole blood makes the technique more amenable to infants who are the major malaria victims. The portability and battery operation of the technology enables it to be carried into remote conditions where power and road transport is an issue. The diagnosis and quantification of parasitemia along with the quantification of two important chemical parameters associated with malaria took less than 10 min (from blood extraction to acquisition of the results). The next stage is to apply the model to predict parasitemia, glucose and urea from clinical samples and a clinical field will be undertaken in Papua New Guinea paving the way for a multianalytical tool for malaria diagnosis and analyte quantification.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: bayden.wood@monash.edu.
*E-mail: david.perez.guaita@monash.edu.
*E-mail: phil.heraud@monash.edu.

ORCID

Bayden R. Wood: 0000-0003-3581-447X

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) White, N. J.; Pukrittayakamee, S.; Hien, T. T.; Faiz, M. A.; Mokuolu, O. A.; Dondorp, A. M. *Lancet* **2014**, *383*, 723−735.
(2) WHO. http://www.who.int/mediacentre/factsheets/fs094/en/; 2016.
(3) WHO. *Pocket book of hospital care for children: guidelines for the management of common childhood illnesses*; World Health Organization: Geneva, 2013.
(4) Bassat, Q.; Guinovart, C.; Sigaúque, B.; Aide, P.; Sacarlal, J.; Nhampossa, T.; Bardají, A.; Nhacolo, A.; Macete, E.; Mandomando, I. *Malar. J.* **2008**, *7*, 1.
(5) Looareesuwan, S.; White, N.; Karbwang, J.; Turner, R.; Phillips, R.; Kietinun, S.; Rackow, C.; Warrell, D. *Lancet* **1985**, *326*, 4−8.
(6) WHO. *Trans. R. Soc. Trop. Med. Hyg.* **2000**, *94*, 1−90.
(7) Nadjm, B.; Mtove, G.; Amos, B.; Hildenwall, H.; Najjuka, A.; Mtei, F.; Todd, J.; Reyburn, H. *Am. J. Trop. Med. Hyg.* **2013**, *89*, 232−237.
(8) Kendjo, E.; Agbenyega, T.; Bojang, K.; Newton, C. R.; Bouyou-Akotet, M.; Pedross, F.; Kombila, M.; Helbok, R.; Kremsner, P. G. *PLoS One* **2013**, *8*, e58686.
(9) Taylor, T. E.; Molyneux, M. E.; Wirima, J. J.; Fletcher, K. A.; Morris, K. *N. Engl. J. Med.* **1988**, *319*, 1040−1047.
(10) White, N.; Marsh, K.; Turner, R.; Miller, K.; Berry, C.; Williamson, D.; Brown, J. *Lancet* **1987**, *329*, 708−711.
(11) Agbenyega, T.; Angus, B. J.; Bedu-Addo, G.; Baffoe-Bonnie, B.; Guyton, T.; Stacpoole, P. W.; Krishna, S. *J. Clin. Endocrinol. Metab.* **2000**, *85*, 1569−1576.
(12) Zijlmans, W.; Van Kempen, A.; Ackermans, M.; De Metz, J.; Kager, P.; Sauerwein, H. *Am. J. Trop. Med. Hyg.* **2008**, *79*, 605−612.
(13) Davis, T.; Looareesuwan, S.; Pukrittayakamee, S.; Levy, J.; Nagachinta, B.; White, N. *Metab., Clin. Exp.* **1993**, *42*, 334−340.
(14) Urscher, M.; Alisch, R.; Deponte, M. *Semin. Cell Dev. Biol.* **2011**, *22*, 262−270.
(15) Binh, T.; Davis, T.; Johnston, W.; Thu, L.; Boston, R.; Danh, P.; Anh, T. *Metab., Clin. Exp.* **1997**, *46*, 1435−1440.
(16) van Thien, H.; Ackermans, M.; Weverling, G.; Chien, V. T.; Endert, E.; Kager, P.; Sauerwein, H. *Clin. Nutr.* **2004**, *23*, 59−67.

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

(17) Edington, G. *Br. Med. J.* **1967**, *1*, 715−718.
(18) Sharma, M.; Nand, N.; Kumar, H.; Suman, L. *JIMSA* **2012**, *25*, 229−230.
(19) Raphael, E. C.; Benjamin, A. U. *Res. J. Med. Sci.* **2010**, *4*, 208−212.
(20) Ndukaku, O. Y.; Emmamuel, E. U.; Loveth, U. E.; OluchiI, O. I. *Rom. J. Biochem.* **2015**, *52*, 39−49.
(21) Ogbadoyi, E. O.; Gabi, B. *Af. J. Infect. Dis.* **2007**, *1*, 57−64.
(22) Narayanan, S.; Appleton, H. *Clin. Chem.* **1980**, *26*, 1119−1126.
(23) Whelton, A.; Watson, A.; Rock, R. In *Tietz Textbook of Clinical Chemistry*, 2nd ed.; Burtis, C. A., Ashwood, E. R., Eds.; WB Saunders Company: Philadelphia, 1994; pp 1513−1575.
(24) Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A. *Nat. Protoc.* **2014**, *9*, 1771−1791.
(25) Perez-Guaita, D.; Garrigues, S. *TrAC, Trends Anal. Chem.* **2014**, *62*, 93−105.
(26) Wood, B. R.; Heraud, P.; Perez-Guita, D.; Khoshmanesh, A.; Jearanaikoon, P.; Dixon, M. W. A.; Tilley, L.; McNaughton, D. *Workshop: FTIR Spectroscopy in Microbiological and Medical Diagnostics*; Robert Koch-Institute, Berlin, October, 2015.
(27) Khoshmanesh, A.; Dixon, M. W.; Kenny, S.; Tilley, L.; McNaughton, D.; Wood, B. R. *Anal. Chem.* **2014**, *86*, 4379−4386.
(28) Janatsch, G.; Kruse-Jarres, J.; Marbach, R.; Heise, H. *Anal. Chem.* **1989**, *61*, 2016−2023.
(29) Hoşafçı, G.; Klein, O.; Oremek, G.; Mäntele, W. *Anal. Bioanal. Chem.* **2007**, *387*, 1815−1822.
(30) Perez-Guaita, D.; Ventura-Gayete, J.; Pérez-Rambla, C.; Sancho-Andreu, M.; Garrigues, S.; de la Guardia, M. *Microchem. J.* **2013**, *106*, 202−211.
(31) Rohleder, D.; Kocherscheidt, G.; Gerber, K.; Kiefer, W.; Ko, W.; Mo, J.; Petrich, W. J. *J. Biomed. Opt.* **2005**, *10*, 031108−03110810.
(32) Webster, G. T.; De Villiers, K. A.; Egan, T. J.; Deed, S.; Tilley, L.; Tobin, M. J.; Bambery, K. R.; McNaughton, D.; Wood, B. R. *Anal. Chem.* **2009**, *81*, 2516−2524.
(33) Wood, B. R. *Chem. Soc. Rev.* **2016**, *45*, 1980−1998.
(34) Piasek, Z.; Urbanski, T. B. *Bull. Pol. Acad. Sci-Tech X* **1962**, 113−120.
(35) Korolevich, M.; Zhbankov, R.; Sivchik. *J. Mol. Struct.* **1990**, *220*, 301−313.
(36) Keuleers, R.; Desseyn, H.; Rousseau, B.; Van Alsenoy, C. *J. Phys. Chem. A* **1999**, *103*, 4621−4630.
(37) Fischer, P.; McDowell, C. *Can. J. Chem.* **1960**, *38*, 187−193.
(38) Park, S. H.; Goo, J. M.; Jo, C.-H. *Korean J. Radiol.* **2004**, *5*, 11−18.
(39) Ami, D.; Mereghetti, P.; Doglia, S. M. *Multivariate Analysis for Fourier Transform Infrared Spectra of Complex Biological Systems and Processes*; INTECH Open Access Publisher, 2013.
(40) Gunderson, J.; Sogin, M.; Wollett, G.; Hollingdale, M.; De La Cruz, V.; Waters, A.; McCutchan, T. *Science* **1987**, *238*, 933−937.
(41) Kozicki, M.; Creek, D.; Sexton, A.; Morahan, B.; Wesełucha-Birczyńska, A.; Wood, B. *Analyst* **2015**, *140*, 2236−2246.
(42) Roth, E., Jr. *Blood Cells* **1989**, *16*, 453−460.
(43) Ballabio, D.; Consonni, V. *Anal. Methods* **2013**, *5*, 3790−3798.
(44) Esbensen, K. H.; Guyot, D.; Westad, F.; Houmoller, L. P. *Multivariate Data Analysis-in Practice: An Introduction to Multivariate Data Analysis and Experimental Design*, 5th ed.; CAMO Process AS: Oslo, Norway, 2001.
(45) Kamruzzaman, M.; ElMasry, G.; Sun, D.-W.; Allen, P. *Innovative Food Sci. Emerging Technol.* **2012**, *16*, 218−226.

DOI: 10.1021/acs.analchem.6b04578
Anal. Chem. 2017, 89, 5238−5245

Chapter 4. Spectroscopy goes Viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy

Clinical Spectroscopy 1 (2019) 100001

Contents lists available at ScienceDirect

Clinical Spectroscopy

journal homepage: www.elsevier.com/locate/clispe

Spectroscopy goes viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy

Supti Roy [a], David Perez-Guaita [a], Scott Bowden [b], Philip Heraud [a,c,**], Bayden R. Wood [a,*]

[a] Centre for Biospectroscopy, Monash University, Clayton, 3800, Victoria, Australia
[b] Victorian Infectious Disease Reference Laboratory, Melbourne, Victoria, Australia
[c] Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, Victoria, 3800, Australia

ARTICLE INFO

Keywords:
Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR-FTIR)
Hepatitis B (HBV)
Hepatitis C (HCV)
Immune response
IgG
Hepatitis B surface antigen (HBsAg)
Support Vector Machine (SVM)
Partial Least Squares Discriminant Analysis (PLS-DA)

ABSTRACT

The development of a new fast, portable and reagent-free diagnostic technique for hepatitis B (HBV) and hepatitis C (HCV) viruses would be an enormous benefit to society. Here, we evalulate the ability of Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectroscopy combined with multivariate data analysis to classify human serum samples based on the presence of HBV and HCV infection. Sera samples were prepared using three different methodologies: i) Sera depsoited onto glass cover slips, airdried and placed onto the ATR crystal. ii) Whole serum dried directly onto the ATR crystal. iii) Serum separated into high and low molecular weight compounds using a filtration approach and the high molecular weight fraction placed directly onto the ATR-FTIR diamond window and dried. For methodology i) the Partial Least Squares Discriminate Analysis (PLS-DA) calibration set included 313 (70 %) samples and the validation set 93 (30 %) samples. For HBV vs control the sensitivity and specificity was found to be 69.4 % and 73.7 % (10 latent variables (LV)), respectively. For HCV vs control the sensitivity and specificity was 51.3 % and 90.9 % (LV 11), respectively. In the second set of experiments the serum samples were dried directly onto the ATR diamond. PLS-DA models were constructed using 144 (70 %) samples for the calibration set and tested using an independent test set containing 62 (30 %) samples. For HBV versus control the sensitivity and the specificity was 84.4 % and 93.1 %, respectively (LV 8). For HCV versus control the sensitivity and specificity was 80.0 % and 97.2 %, respectively (LV 9). For HBV versus HCV the sensitivity and the specificity was 77.4 % and 83.3 %, respectively (LV 5). To increase the sensitivity and specificity serum sample was fractionated into high and low molecular weight components. In PLS-DA cross validated model (LV 8) the sensitivity and specificity was 87.5 % and 94.9 %, respectively for HBV vs control (high molecular concentrate). The PLS-DA cross-validated model (LV 8) for HCV vs control high molecular fraction produced a sensitivity and specificity of 81.6 % and 89.6 %, respectively. No linear correlation was observed for sera samples spiked with known viral loads using Partial Least Squares Regression (PLS-R) modelling.

Spectra of positive serum (HBV and HCV) showed a strong band observed at 1631 cm^{-1}, which was absent in the spectra of controls and assigned to the β-pleated sheet protein marker of immunoglobulin (Ig). A band at 1093 cm^{-1}, observed in spectra of HBV infected sera, was assigned to C−C and CO−modes of polysaccharide N-glycan from hepatitis B surface antigen (HBsAg). The assignment was confirmed by atomic force microsocpy infrared (AFM-IR) spectroscopy of the isolated protein. This band represents a unique marker for HBV infection. In summary, ATR-FTIR spectroscopy is a powerful tool to study blood composition and identify potential disease markers but care must be taken to ensure that the modelling is not biased by inflammation markers, which may confound diagnosis.

Introduction

Bloodborne viral infections can be difficult to treat and may have a tremendous impact on morbidity and mortality [1–6]. These infectious agents are transmitted by direct contact with infected blood and body fluids [7]. According to the Global Burden of Disease study from 2013, viral hepatitis from infection with hepatitis B virus (HBV), from the family *Hepadnaviridae* and hepatitis C virus (HCV), from the *Flaviviridae* family were the seventh leading cause of mortality worldwide with 1.4 million deaths [8].

Diagnosis of HCV and HBV infection are currently based on the detection of antibodies or antigens using serological assays [9–11], and

* Corresponding author.
** Corresponding author at: Centre for Biospectroscopy, Monash University, Clayton, 3800, Victoria, Australia.
E-mail addresses: phil.heruad@monash.edu (P. Heraud), bayden.wood@monash.edu (B.R. Wood).

http://doi.org/10.1016/j.clispe.2020.100001
Received 20 October 2019; Received in revised form 19 December 2019; Accepted 13 January 2020

molecular polymerase chain reaction (PCR) based assays for the detection of viral nucleic acids [12,13]. These techniques require bulky equipment, are expensive, and time consuming [14]. Rapid Diagnostic Tests (RDTs) are relatively quick to perform (15−20 min), and although they perform well for Hepatitis B [15], commercially available RDTs for Hepatitis C have demonstrated less than perfect sensitivity (78.8 %) [16]. Furthermore, many RDTs provide only qualitative information, and require refrigeration making them difficult to use in remote environments. A new test that utilizes portable based technology that does not require special conditions for preservation of samples nor serological or molecular reagents would have significant social and economic benefits in diagnosing these infections. Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) is an established tool for analyzing and identifying different blood chemistry parameters (like lipids, glucose, proteins, urea, etc.) [17–19] and infectious agents in blood [20,21]. The technique is non-destructive and reagent-free, with little or no pre-processing of the sample required [22,23]. The detectable chemical compounds in a biological sample are normally expressed in the infrared spectrum as a set of overlapping bands, which can be related to the presence of proteins, lipids, carbohydrates and DNA. If a disease agent modifies the composition of blood, either by the presence of a pathogen or by inducing a phenotypic response in the host (e.g. inflammation) that is above the limit of detection of ATR-FTIR then a multivariate model can be used in a predictive (PLS-DA) or quantitative (PLS-R) capacity. PLS-DA is a supervised modelling approach whereby the data in the X-matrix (spectral data) is projected onto a Y-Matrix, which contains the known outcome from the disease state as a dummy variable. PLS-R is used as a quantitative tool where by a series of standards (known viral load) is used to develop a calibration model based on infrared spectra. The concentration of an unknown can then be predicted by the model.

Serum, being a relatively homogenous fluid in comparison to the cells and tissues, is possibly the best medium to detect infectious agents by spectroscopy. Raman spectroscopy has been applied to detect and quantify HCV, [24] while near-infrared spectroscopy has been applied to diagnose human immunodeficiency virus (HIV-1) in plasma [25]. ATR-FTIR spectroscopy has also been applied to detect HIV/HCV co-infection [26]. In addition to these studies, the application of ATR-FTIR spectroscopy to serum analysis is gaining momentum [21,27] due to its ease of use and portability. Hepatitis viruses target the hepatocytes and replicate within them, which further spreads the virus into the bloodstream. Life cycle studies of these virus infections include chronic and acute phases of infection, which have different effects on the blood chemistry parameters. Consequently, it is important to understand the components of serum and the changes that accompany infection. Studies [28–30] on some of the effects of HCV and HBV on blood chemistry are given in Table 1. Given the large changes in the concentration of these analytes, it was hypothesized that ATR-FTIR spectroscopy in combination with multivariate data analysis could discriminate between virus-infected and non-infected controls from blood sera samples. This study focuses on the detection of certain specific biomarkers related to chronic infection from HBV and HCV, which can be used as a method to discriminate infected serum from the controls. With this aim, a dataset of control and infected serum samples was collected and using the IR spectra as predictor variables PLS-DA classification models were created to classify: i) HBV versus controls, ii) HCV versus controls, iii) infected (HCV and HBV) versus controls, iv) HBV vs HCV. Furthermore, we determined whether the classification is based on immune markers or due to the direct detection of virus particles. To this end, a spiking experiment using a wide range of viral loads was performed, and PLS-R was employed to determine if there was a linear relationship between IR bands and viral load.

Table 1 highlights the increase and decrease of various blood chemistry parameters in response to infection.

Methods and materials

Sample preparation

This work was granted ethical approval (CF16/311-2016000139) by the Human Low Risk Review Committee at Monash University. Serum samples with different viral loads (10^2-10^9 IU/mL) of HBV and HCV were obtained from Victorian Infectious Disease Reference Laboratory (VIDRL). Whole blood was collected into serum separation tubes and the serum isolated after centrifugation for 10 min at 1000 g. Before the centrifugation step, samples were kept at room temperature for at least one hour to allow them to clot. In total, 497 samples (placing the sample on glass cover slip) and 361 samples (placing the sample directly onto the ATR-crystal) were analyzed in this study and the negative samples were tested in VIDRL for Hepatitis B, C, and HIV, using enzyme-linked immunosorbent assay (ELISA) and multiplex PCR assays.

Analysis using ATR-FTIR spectroscopy was performed on clinical samples in three ways:

i) *By placement of the sample onto a glass cover slip generating a thick film.* A minimum volume of 10 μL of each serum sample was deposited onto a glass cover slip (0.13-0.17 mm thickness, 22 × 22 mm length) using a pipette tip to generate a thick film, which was left to fully dry (10 min). The glass cover slip was then placed upside down with the dried serum sample in direct contact with the ATR crystal and constant pressure applied using the instrument's pressure clamp as demonstrated in a previous study by our group [18]. For each sample, 3 replicate spectra from three deposits were acquired and the average spectrum calculated for spectral comparison and modelling. In total, 191 negatives, 142 from HBV infected and 164 of HCV infected serum samples were measured using this method.

ii) *By direct placement of the sample onto the ATR crystal.* A 4 μL volume of serum is placed directly onto the ATR crystal and dried immediately using a gentle stream of air at room temperature inside a Biological Safety

Table 1
Effect of hepatitis B and C virus infection on various blood components (mean ± SE) or medians and ranges [28–33].

Parameters	Molecular weight	Controls	Hepatitis B	Control vs. HBV	Hepatitis C	Control vs. HCV
Cholesterol(mg/dL)	386.654 Da	172.44 ± 3.03	127.56 ± 1.70[c]	Decrease	124.24 ± 1.77[c]	Decrease
HDL(mg/dL)	180-360 kDa	30.42 ± 0.57	27.96 ± 0.36[b]	Decrease	27.44 ± 0.385[c]	Decrease
LDL(mg/dL)	2750 kDa	112.34 ± 2.98	97.88 ± 1.87[c]	Decrease	95.16 ± 1.81[c]	Decrease
Triglycerides(mg/dL)	885 Da	96.00 ± 4.33	95.30 ± 3.76[a]	Decrease	84.12 ± 1.51[b]	Decrease
Albumin (g/dL)	65 kDa	3.68 ± 0.083	3.20 ± 0.079[c]	Decrease	3.07 ± 0.090[c]	Decrease
Total protein (g/dL)	~ 20-200 kDa	8.40 ± 1.23	6.84 ± 0.80[b]	Decrease	6.74 ± 0.82[c]	Decrease
IgG (g/L)	160 kDa	11.44 ± 0.44	16.83 (8.86–30.76)[d]	Increase	21.8 ± 5.2[d]	Increase
IgA (g/L)	160 kDa	2.27 ± 0.14	2.50 ± 0.24[d]	Increase	3.3 (1.3–4.3)[d]	Increase
IgM (g/L)	1000 kDa	0.92 (0.54–1.77)	1.38 (0.75–3.47)[d]	Increase	1.6 ± 0.8[d]	Increase

[a] $P < 0.05$.
[b] $P < 0.01$.
[c] $P < 0.001$.
[d] P = not determined vs controls, HDL (High Density Lipoprotein), LDL (Low Density Lipoprotein).

Cabinet Class II (BSC II) for 2–3 min. In total, 114 negatives, 117 HBV infected and 130 HCV infected serum samples were measured using this approach. The positive and negative samples were randomized to minimize other influences, such as moisture and instrumental variability. No technical replicates for the same deposit were taken because of the extremely high signal-to-noise ratio and the spectra were highly reproducible (relative standard deviation was close to 5 % within the spectral range of 1762-906 cm^{-1}).

iii) *Separating the serum sample into high molecular and low molecular weight components.* Amicon ultra centrifugal filter devices Millipore Corporation, Bedford, MA, U.S.A were used for separating the high molecular weight components concentrate and low molecular metabolites filtrate from clinical serum samples. The nominal molecular weight limit NMWL for the filter is 10 kDa. Filters were pre-rinsed at 14,000 g for 4 min by using 500 μL of 0.9% NaCl to remove the trace amount of glycerine from the filter device. A 200 μL volume of sample was added immediately after the rinsing procedure and centrifuged at 14,000 *g* for 15 min. According to the manufacturer's instructions 10 kDa filters require a 15 min centrifuge time to achieve a retention percentage of proteins in the concentrate of more than 95 %. To recover the concentrated solute, the filters were placed upside down in a new micro-centrifuge tube and centrifuged for 2 min at 1000 g.

For the low molecular weight fraction, a 4 μL volume was deposited onto the ATR crystal. For the high molecular weight fraction, a 2.5 μL volume was deposited onto the ATR crystal. The reason for the different volumes was due to the different viscosities of the high and low weight molecular fractions. In both cases the serum was dried immediately using a gentle stream of air for 2–3 min inside a BSC II cabinet prior to spectral acquisition. A total subset of 40 samples from each group (HCV, HBV, and control) were measured using this method. The positive and negative samples were randomized to minimize other influences, such as moisture and instrumental variability. No technical replicate spectra were acquired because of the extremely low inter-sample variability as mentioned in the above section.

Spiking experiments

Negative clinical samples were spiked with high viral load of HCV and HBV clinical samples. Four high viral load HBV samples with viral load of 10^9 IU/mL were serial diluted to the range of 8×10^8 IU/mL to 10 IU/mL. While for HCV, three high viral load (10^7 IU/mL), were serial diluted to the range of 8×10^6 IU/mL to 10 IU/mL. Each time different clinically negatives were used to avoid the influence of other serum components during the modelling. The sample preparation for these experiments is described in the methods section: i) *By placement of the sample onto a glass cover slip generating a thick film*.

ATR-FTIR spectroscopy

ATR-FTIR spectra were acquired using an Agilent 4500 ATR FTIR spectrometer. The spectra were collected between 4200–600 cm^{-1}, with 64 co-added interferograms (background 128 co-added interferograms) and a spectral resolution of 8 cm^{-1}. A background spectrum of air at room temperature (approximately 20 °C) was acquired prior to each sample measurement. The portable instrument was placed inside the BSC II. For the direct drying method, the signal-to-noise ratio of the ATR-FTIR spectra was also calculated along with the recording of preparation and measurement times.

Data analysis and chemometrics

PLS-DA modelling

ATR-FTIR spectra were analyzed using MATLAB 9.1 Release 2016b from Mathworks (Natick, U.S.A) with the PLS toolbox supplied from Eigenvector (Manson, U.S.A). Several models were constructed using a combination of spectral regions from the dataset using in-house developed software. The best model was chosen based on the optimum sensitivity, specificity, and minimal cross-validation error, by modifying the preprocessing steps including Savitzky-Golay smoothing, the type of derivative (first or second), standard normal variate and mean centering. For the current analysis, several spectral regions were considered for each model and the data was preprocessed by performing Savitzky-Golay smoothing (9 points) followed by baseline correction with second order polynomial fitting, normalization using standard normal variate and mean centering. Partial Least Squares Discriminant Analysis (PLS-DA) was applied to classify control, HBV and HCV spectra and to investigate the influence of other blood parameters.

Development of PLS-DA models for sera on glass cover slip. As an initial approach, PLS-DA models were constructed using the spectral dataset obtained from thick films of sera on glass cover slips. A total of 497 samples were analyzed including: uninfected controls (n = 191); HBV samples with viral loads of 10^2-10^9 IU/ mL (n = 142); and HCV with viral loads 10^2-10^7 IU/ mL (n = 162) (IU = approx. 5 DNA/RNA copies). Of these HBV (n = 132), HCV (n = 152) and controls (n = 181) were considered for modelling, with the remaining outlying spectra excluded following identification based on having large residual errors and Hoteling's T-squared (T^2) values.

The Kennard-Stone algorithm was used to select the 70 % of samples for the calibration subset, which was used to build and optimize the models, leaving the remaining (30 %) of samples in the independent test set. The number of latent variables (LV) were selected based on the minima value in the latent variable versus cross-validation error plots. For cross validation Venetian Blind algorithm was used. The preprocessing steps, spectral range and latent variables selection are summarized in the Table 6: S1.

Development of PLS-DA models for whole sera from clinical patients. The PLS-DA models were generated from the spectra acquired by drying the sample directly onto the diamond ATR crystal. A total of 361 samples were analyzed including: uninfected controls (n = 114); HBV samples with viral loads of 10^5-10^9 IU/ mL (n = 117); and HCV with viral loads 10^5-10^7 IU/ mL (n = 130) (IU = approx. 5 DNA/RNA copies. Of these HBV (n = 102), HCV (n = 101) and controls (n = 104) were considered for modelling, with the remaining outlying spectra excluded following identification based on having large residual errors and Hoteling's T-squared (T^2) values.

The Kennard-Stone algorithm was used to select the 70 % of samples for the calibration subset, which was used to build and optimize the models, leaving the remaining (30 %) of samples in the independent test set. The number of latent variables (LV) were selected based on the minima value in the latent variable versus cross-validation error plots. Four PLS-DA models were constructed from the sample spectra namely: **1**. HBV & HCV vs negatives (LV = 8; spectral region: 1758-869 cm^{-1}), **2**. HBV vs negatives (LV = 8; spectral region: 1758-906 cm^{-1}), **3**. HCV vs negative (LV = 9; spectral region: 1762-906 cm^{-1}) and **4**. HBV vs HCV (LV = 5; spectral region: 1360-900 cm^{-1}). Cross-Validation was performed using the Venetian Blind algorithm from Eigenvector Research, Inc. (USA). The plots for both cross-validated and predicted models with the corresponding regression vector, receiver operating characteristic or ROC curve are shown in Figs. 2,4–6.

Development of PLS-DA models for dataset obtained from high molecular and low molecular weight sera samples. PLS-DA modelling was performed on the ATR-FTIR spectral dataset obtained from the Amicon ultra centrifugal filter separation of clinical serum into high and low molecular weight samples. A subset of 40 samples were included in the models from each group (HBV, HCV and controls). Four cross-validated models were generated using the same spectral region (1791–895 cm^{-1}) and different LVs, which are summarized in Table 4. Fig. 7 shows the plots of PLS-DA

models with the dataset of high molecular weight components with the corresponding regression vectors and ROC curves. The Venetian Blind algorithm from Eigenvector Research, Inc. (USA) was used for cross validation.

PLS-R modelling

Partial least Squares Regression (PLS-R) modelling was performed for different viral loads of HBV (max. 10^9 IU/mL) and HCV (max. 10^7 IU/mL) from the clinical samples. One positive sample was diluted using several negative serum samples so that other variable parameters (proteins, glucose, urea, etc.) could not bias the model. The measurements were performed in triplicate for each sample and each replicate was included in the PLS-R model. The spectral region 1769–900 cm^{-1} was chosen and cross-validation performed using the Venetian Blind algorithm. In all cases, Savitzky-Golay smoothing (smoothing point 9), baseline correction (2nd order polynomial fitting), with standard normal variate normalization and mean centering were used for preprocessing.

Results and discussion

Visual inspection

In one set of experiments, whole serum was placed directly onto the ATR crystal and dried immediately using a stream of air at room temperature in a BSC II cabinet. Spectra were acquired using the ATR-FTIR spectrometer. For this method, 114 negative samples, 117 HBV DNA positive samples and 130 HCV RNA positive samples were analyzed without any technical replicates. The second derivative spectroscopic data for positive versus negative control were normalized and compared.

After removing the outliers, Fig. 1A shows the second derivative normalized spectra for the HBV and negative samples, while Fig. 1B shows the analogous spectra for HCV and negatives. The dominant bands

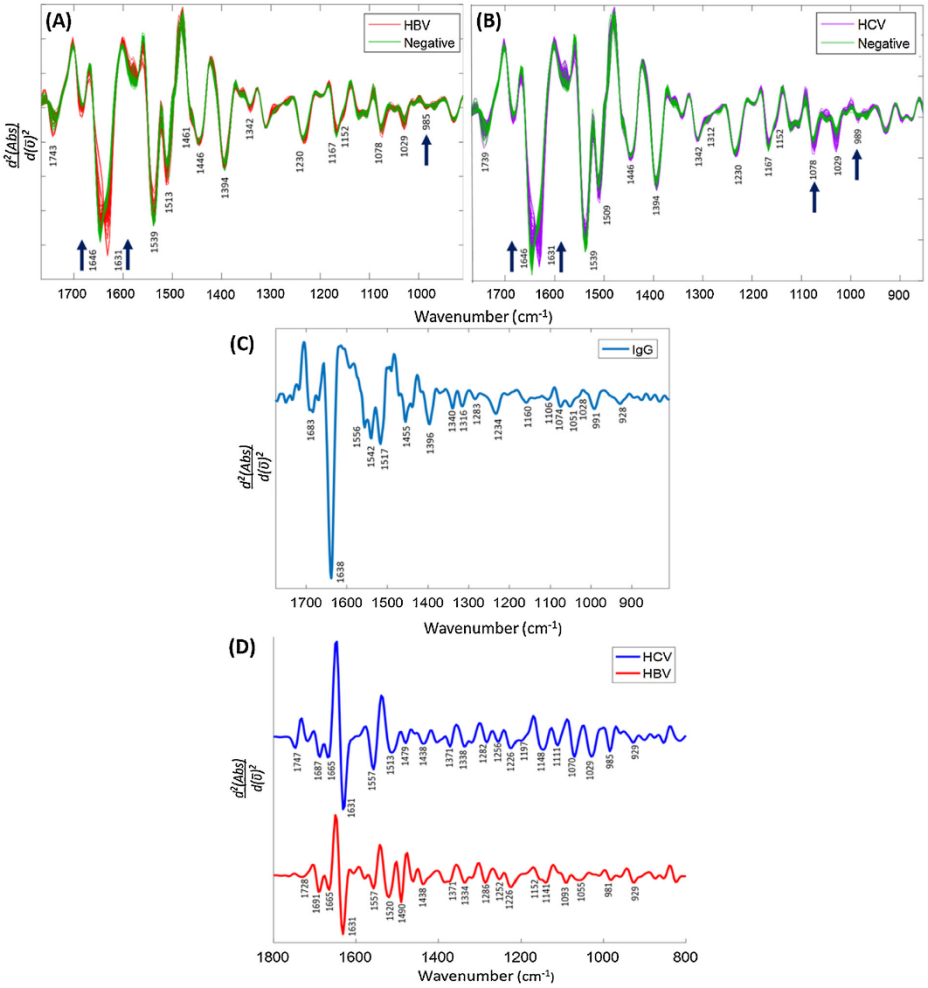


Fig. 1. The Second derivative normalized spectra of A). HBV versus negative samples, B). HCV versus negative samples, C). IgG D). Mean control (negative for viral infection) subtracted from mean of HCV (blue) and HBV (red), respectively.

that change between controls and infected samples are highlighted by arrows. The band assignments for vibrational modes are summarized in Table 2.

The spectra are dominated by bands assigned to peptide moieties from proteins (1539 cm^{-1}, 1631 cm^{-1}, 1646 cm^{-1}). A major difference observed between infected and control sera samples was a more intense minima band at 1631 cm^{-1} in infected samples, which is characteristic of a β-pleated sheet protein and is assigned to the inflammation marker immunoglobulin (Ig) [19]. As shown in Table 1 the IgG, IgA, and IgM levels increase in HBV infected patients [29]. IgG and IgA levels are also very high for patients with chronic HCV infection and severe hepatic fibrosis [30]. In Fig. 1C, a subclass Immunoglobulin G (IgG) is shown as a reference spectrum of Ig. The band is shifted slightly in the pure extract compared to the cells by 7 cm^{-1}, which is not surprising given that the Ig in the sera is in a protein/analyte matrix whereas the purified IgG is in a lyophilized form. The moderately intense band at 1230 cm^{-1} is slightly red-shifted in both HBV and HCV sample spectra compared to the control spectra. It is tempting to assign this band to the asymmetric vibration of the phosphodiester group from viral nucleic acids. However, a moderately intense band at 1234 cm^{-1} also appears in the spectrum of IgG and it would appear more likely this band is from immunoglobulin class or other proteins with high beta-pleated sheet component as opposed to viral nucleic acids and is tentatively assigned to the Amide III mode. The same can also be said for the 1078 cm^{-1} band, which can be assigned to the symmetric phosphodiester stretch of viral nucleic acids but this also appears quite strongly in the IgG spectrum at 1074 cm^{-1} along with the band at 991 cm^{-1}. The 991 cm^{-1} band is slightly red-shifted and appears in both infected HBV (985 cm^{-1}) and HCV (989 cm^{-1}) spectra, and also in the IgG [34] sample spectrum. All of these bands are more intense in the case of HCV positive samples compared to the HBV positive samples, and negative controls. Fig. 1D shows the second derivative mean spectra of controls were subtracted from both second derivative mean positive (HBV/HCV) spectra.

Development of models

PLS DA models obtained from spectra recorded from a thick film of serum deposited onto a glass cover slip

An initial set of experiments focused on developing an approach using glass cover slips as the substrates for ATR measurements as previously reported by our group for malaria and analyte quantification using whole blood [18]. A total of 497 samples were measured by generating a thick film of serum on glass cover slips. These samples included 191 negatives, 142 HBV and 164 HCV infected samples. However, the sensitivity and specificity was comparatively low and the classification errors were more than two times higher (supplementary information, (Table 6: S1) in comparison with the direct sampling approach discussed below. In the case of the glass cover slips the contact between the sample and the crystal is not as good compared to drying the serum sample directly onto the ATR crystal. Hence, while the glass cover slip method is more practical in terms of high throughput the lower signal-to-noise ratio reduces the sensitivity and specificity.

Hepatitis B virus vs Controls by applying serum directly onto the ATR crystal

Fig. 2 shows results of the HBV vs negatives PLS-DA modelling using the direct approach of placing the serum onto the ATR-FTIR diamond window. Fig. 2A, B show the Y predicted values and ROC curve for the cross validation, respectively, while Fig. 2C, D show the predicted Y values plot and corresponding ROC curve for the independent test, respectively. Error values (See Table 3) indicated that the model was able to predict accurately the presence of HBV, with specificity and sensitivity values above 84 % and Area Under the Receiver Operating Characteristic (AUROC) curve values above 90 % for both independent testing and cross validation.

Fig. 2E shows the corresponding regression vector for LV 8. Maxima (positive) bands are associated with HBV infected sample spectra, while minima (negative) bands are associated with control spectra. The positive regression vector bands (Fig. 2E) from the infected serum samples indicate the presence of β-pleated sheet from antibodies (1631 cm^{-1}, amide I band), presumably from of an increase in Ig, corroborating what is observed with the averaged spectra (Fig. 1). Negative bands including the α-helical amide I (1650 cm^{-1}), amide II (1505 cm^{-1}, 1591 cm^{-1}), and amide III (1308 cm^{-1}) are indicative of a decrease in α-helical protein due to hepatitis, which is observed in both HBV and HCV infected samples. The positively loaded band at ~1420 cm^{-1} was analysed in detail as it does not correspond to exact maxima bands in raw spectra or minima bands in second derivative spectra. It was concluded after visual examination of the entire population of spectra that this loading resulted from changes in the curvature of the spectra at this spectral position due to increases in absorbance of the adjacent antisymmetric (1446 cm^{-1}) and symmetric (1394 cm^{-1}) bands from methyl and methylene bending modes in infected samples. A study on hepatitis patients showed a significant decrease of albumin (more than 67 % α-helical) during the infection stage [28] providing a plausible explanation for our results. The regression vector for LV 8 (Fig. 2E) also shows positive bands at 1226 cm^{-1}, 1249 cm^{-1} and 1286 cm^{-1} assigned to amide III protein and phosphodiester DNA modes.

The surface antigen of hepatitis B virus (HBsAg) is a complex structure of glycoprotein and lipids. The virion or the infectious Dane particle of the

Table 2
Major Band position observed in the regression vector of the PLS-DA model and PLS-R model with their assignment.

ATR band positions; wavenumber cm^{-1}	Assignment	Reference value
966 (Infected)	A-DNA backbone ν(C–C)	965 [50]
1015, 1018 (Infected)	RNA, ν(C-O) ribose	1015 [48]
985, 988, 989, 1074 (Infected)	Proteins (Ig)	991, 1074 (reference spectrum Fig. 1C)
1078 (Infected)	DNA, RNA ν_s (PO_2^-) or phospolipid ν_s (PO_2^-)	1080 [48]
1067 (Infected)	RNA, ν(C-O) ribose	1060 [48]
1093, 1096, 1100 (Infected)	valence vibration of CC and CO in glucose from polysaccharide	1101 [18]
1119 (Infected)	RNA, ν(C = O) ribose	1120 [48]
1037 (Infected)	RNA, ν(C = O) ribose	1138 [48]
1145, 1148 (Infected)	ν_{as} (CO-O-C) of glycan and nucleic acids (DNA and RNA)	1155 [43]
1167,1174, 1182, 1186 (Uninfected)	ester C-O-C asymmetric stretching from phospholipid, triglyceride and cholesterol esters	1173 [42] or 1172 [43]
1223, 1226, 1234, 1249, 1286 (Infected)	A-DNA, ν_{as} (PO_2^-) and RNA, ν_{as} (PO_2^-) or	1240 and 1244 [48] 1240 [50]
	Phospholipid phosphodiester bond (lipid bilayer) or Protein (IgG), Amide III	1242, [42]
1379 (Uninfected)	Proteins, lipids, symmetric δCH_3	1378 [42]
1501, 1542, 1591 (Uninfected)	Proteins, amide II	1514, 1550 [43]
1308 (Uninfected)	Proteins, amide III	1308 [51]
1635, 1631 (Infected)	Amide I, protein, β pleated sheet	1635 [48]
1650, 1654 (Uninfected)	Amide I, protein, α-helical or DNA (C = O and C = N)	1650 [48];
1691 (Infected)	ν_a(C_2=O) vibration respectively in RNA	1690 [48]
1732,1747 (Infected, uninfected)	Lipids, ν(C = O) ester carbonyl	1740 [48]

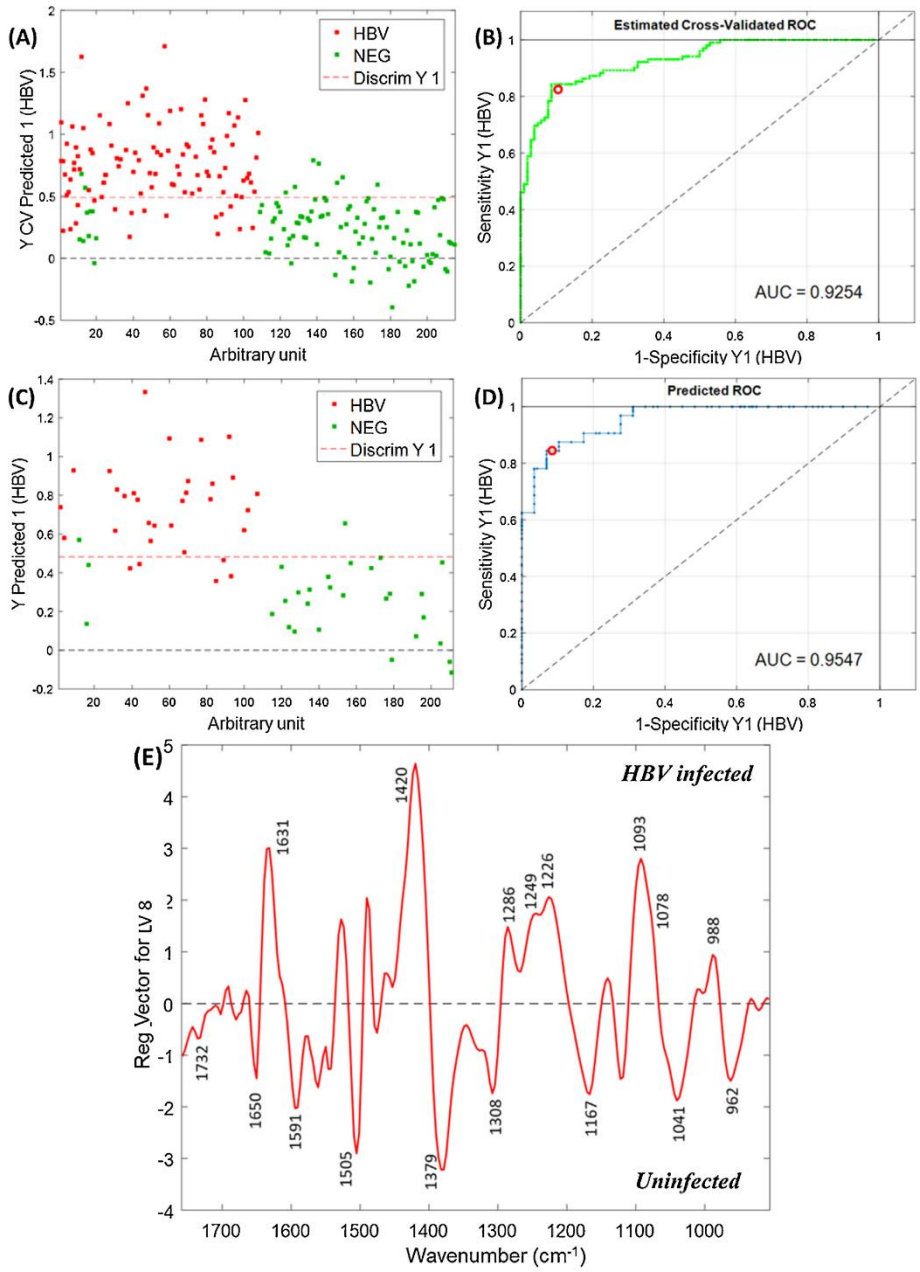


Fig. 2. HBV vs Negatives PLS-DA models, spectra acquired from whole serum. A). Cross-validated Y values plot. B). Estimated ROC for Cross-validation. C). Predicted Y values plot. D). Estimated ROC for prediction. E). Regression vector of PLS-DA models (CV & Predicted) for the dataset HBV vs Negatives, where minima bands are associated with controls and maxima bands are associated with HBV spectra.

Table 3
Modelling parameters, and classification performance parameters estimated by cross validation and independent testing of the models considered with whole serum placing on ATR diamond.

Analyte	Preprocessing	Region (cm^{-1})	(Cross Validation)					(Prediction)				
			LV	Classification error (%)	Sensitivity (%)	Specificity (%)	AUROC	LV	Classification error (%)	Sensitivity (%)	Specificity (%)	AUROC
Positives (HBV & HCV) vs Controls	Smoothing (9) Baseline correction SNV MC	1758-869	8	18 %	77.3 %	85.10 %	87.5 %	8	21.5 %	75. 0 %	81.8 %	85.05 %
HBV vs Controls	Smoothing (9) Baseline correction SNV MC	1758-906	8	12.10 %	84.3 %	91.3 %	92.5 %	8	11.2 %	84.4 %	93.10 %	95.4 %
HCV vs Controls	Smoothing (9) Baseline correction SNV MC	1762-906	9	15.2%	77.2 %	92.30 %	92.01 %	9	11.3 %	80.0 %	97.20 %	91.5 %
HBV vs HCV	Smoothing (9) Baseline correction SNV MC	1360-900	5	22.0 %	79.6 %	76.2 %	85.7 %	5	19.6 %	77.4 %	83.3 %	87.3 %

virus is embedded within the host derived lipid bilayers [35–37]. Bands in this region (1300–1200 cm^{-1}) also appear in the spectrum of IgG (Fig. 1C) and are tentatively assigned to phospholipid phosphodiester bonds and amide III modes from proteins. As stated above the band at 1078 cm^{-1} also appears in IgG spectra and is therefore likely to be assigned to this protein as opposed to the symmetric phosphodiester stretch from nucleic acids.

The HBV genome is composed of partially double-stranded DNA. The virion or virus particle is approximately 42 nm in diameter and consists of a 27 nm icosahedral core, surrounded by a viral envelope, composed of protein, lipid, and carbohydrate [38]. Exposed on the surface of the envelope is a mosaic of glycoproteins, known collectively as HBsAg [35–37]. During infection, an excess of hepatitis B surface antigen-virus like particles (>1000 HBsAg VLPs:1HBV virus particle) are secreted into the blood, where they circulate as 22 nm diameter spherical and filamentous subviral particles [39–41]. The spectrum of HBsAg VLPs [36,38,41] acquired using Atomic Force Microscope Infrared spectroscopy (AFM-IR) is shown in Fig. 3A. The strong band at 1096 cm^{-1} in the AFM-IR spectrum correlates with the maxima band at 1093 cm^{-1} that is observed in the regression vector plot (Fig. 2E) related to HBV infected sample spectra. This band is assigned to C–C and CO– bands from polysaccharide N-glycan found in the hepatitis surface antigen. Further validation of this was achieved by comparing the spectra of HBV monomeric DNA and vector plasmid DNA (Fig. 3B) because the analyzed VLPs extract may have these possible contaminants.

The minima band associated with control spectra at 1732 cm^{-1} can be assigned unambiguously to ester carbonyl stretching from lipids. This is corroborated by negative bands in the lower wavenumber region including bands assigned to symmetric bending modes of methyl groups (1379 cm^{-1}) and ester C–O–C asymmetric stretching from phospholipids, triglyceride and ester cholesterol (1167 cm^{-1}) (Fig. 2E; Table 2) [42,43]. The observed decreased in lipid levels for HBV infected samples is supported by several reports that show low levels of cholesterol, triglycerides, low-density lipoprotein (LDL) cholesterol and high-density lipoprotein (HDL) cholesterol in patients with chronic HBV and HCV infections (Table 1) [28,44,45]. The liver plays a key role in maintaining lipid metabolism and lipoprotein circulation. HBV and HCV infection disturbs the assembly of lipoproteins and modulates the pathway, as a result of an impeded biosynthetic mechanism. Viral infections are also the reason for enhanced degradation, impaired secretion, and beta-oxidation of fat molecules.

Hepatitis C virus vs Controls by applying serum directly onto the ATR crystal

HCV is an enveloped single-stranded RNA virus with a genome of 9.6 kilobases in length. The genome is enclosed by a core, surrounded by a lipid bilayer containing two viral glycoproteins (E1 and E2) to form the

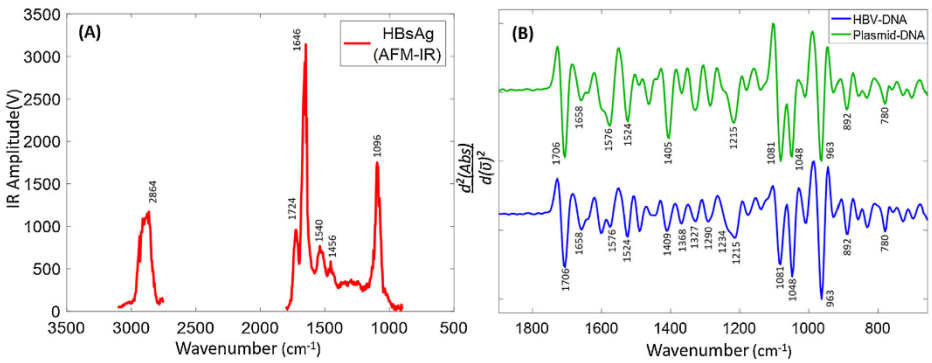


Fig. 3. Absorbance spectra of HBsAg. A). AFM-IR. B). 2nd derivative spectra of HBV-DNA and vector plasmid DNA.

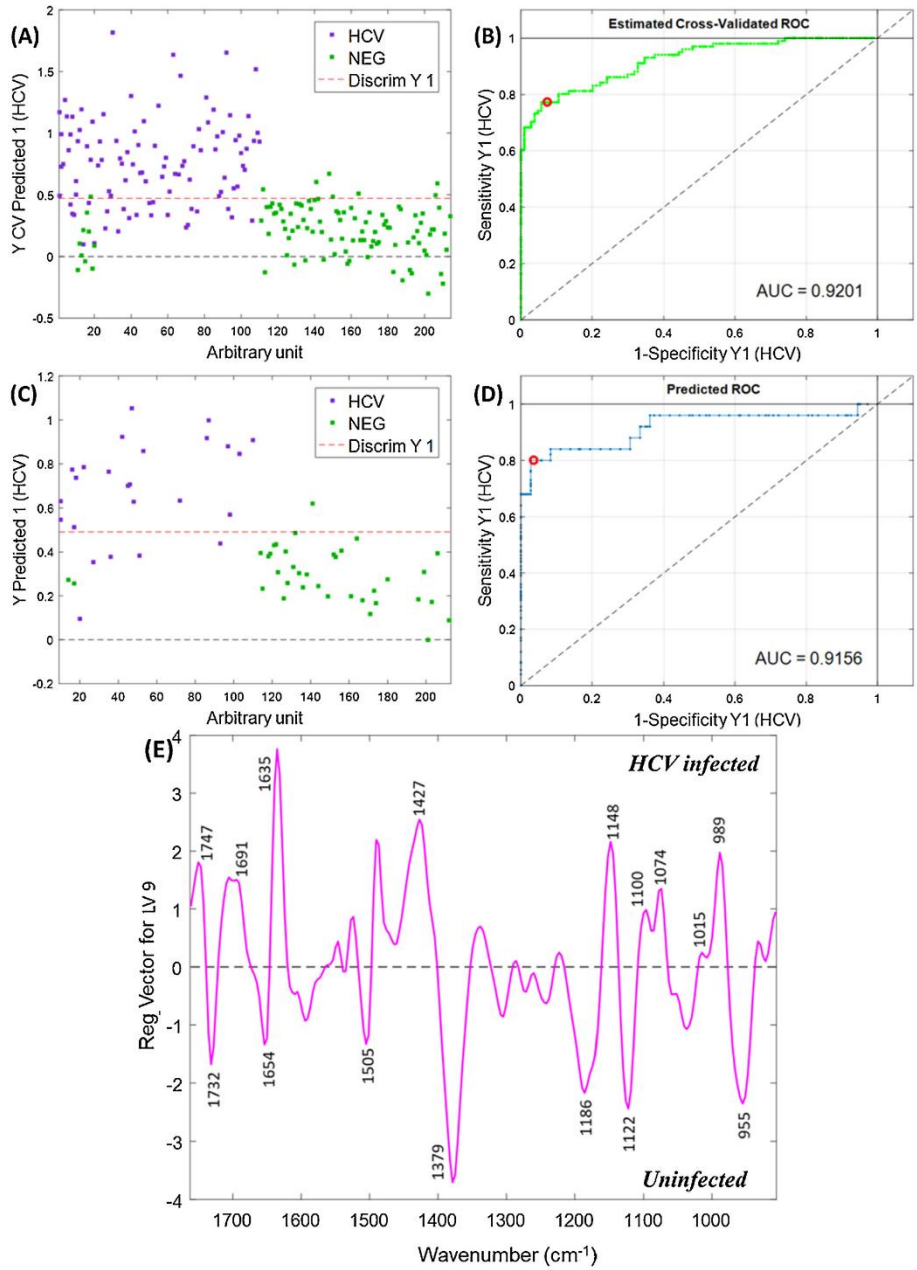


Fig. 4. HCV vs Negatives PLS-DA models, spectra acquired from whole serum. A). Cross-validated Y values plot. B). Estimated ROC for Cross-validation. C). Predicted Y values plot. D). Estimated ROC for prediction. E). Regression vector of PLS-DA models (CV & Predicted) for the dataset HCV vs Negatives, where minima bands are associated with controls and maxima bands are associated with HCV spectra.

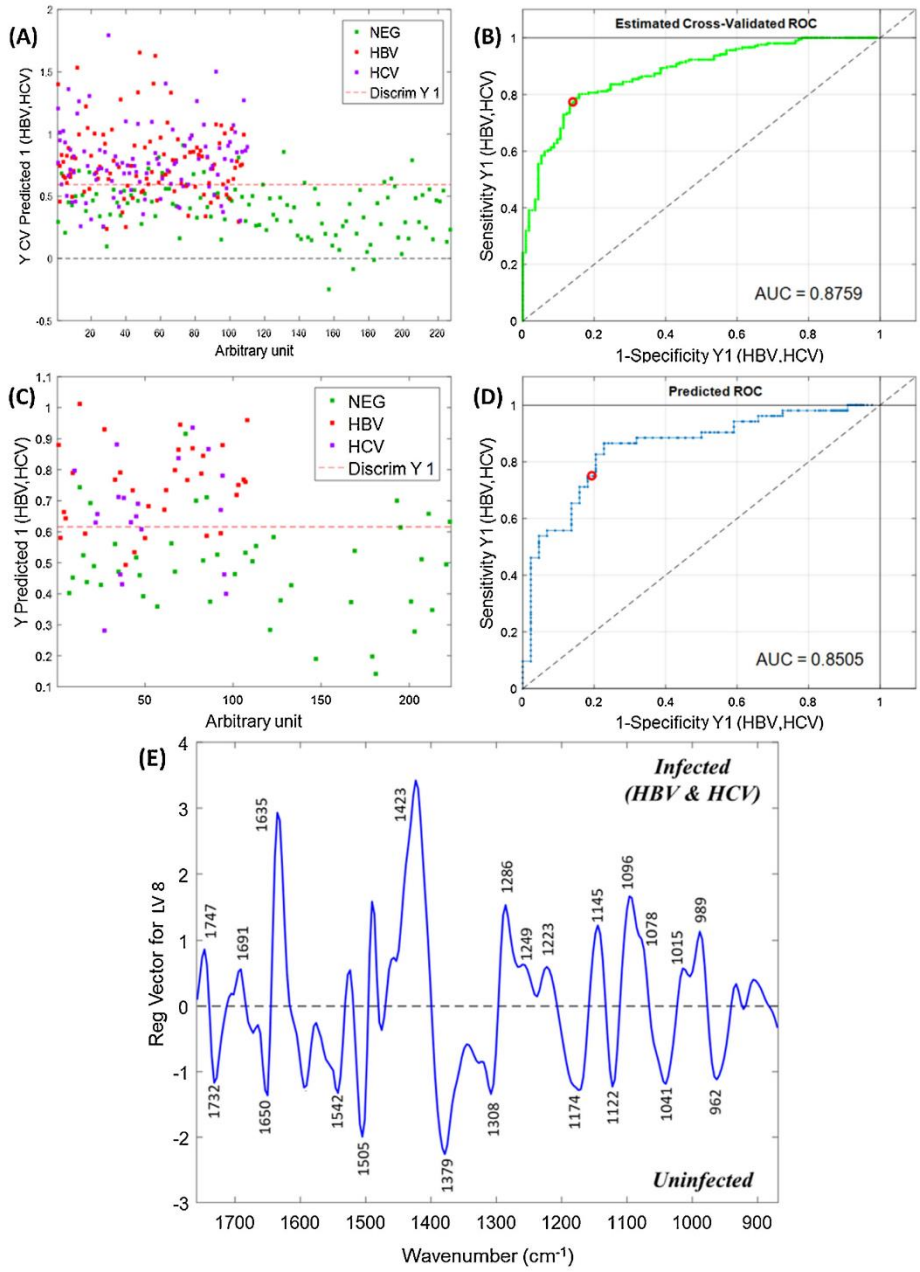


Fig. 5. Positives vs Negatives PLS-DA models, spectra acquired from whole serum. A). Cross-validated Y values plot. B). Estimated ROC for Cross-Validation. C). Predicted Y values plot. D). Estimated ROC for prediction. E). Regression vector of PLS-DA models (CV & Predicted) for the dataset Positives vs Negatives, where minima bands are associated with controls and maxima bands are associated with infected (HBV & HCV) spectra.

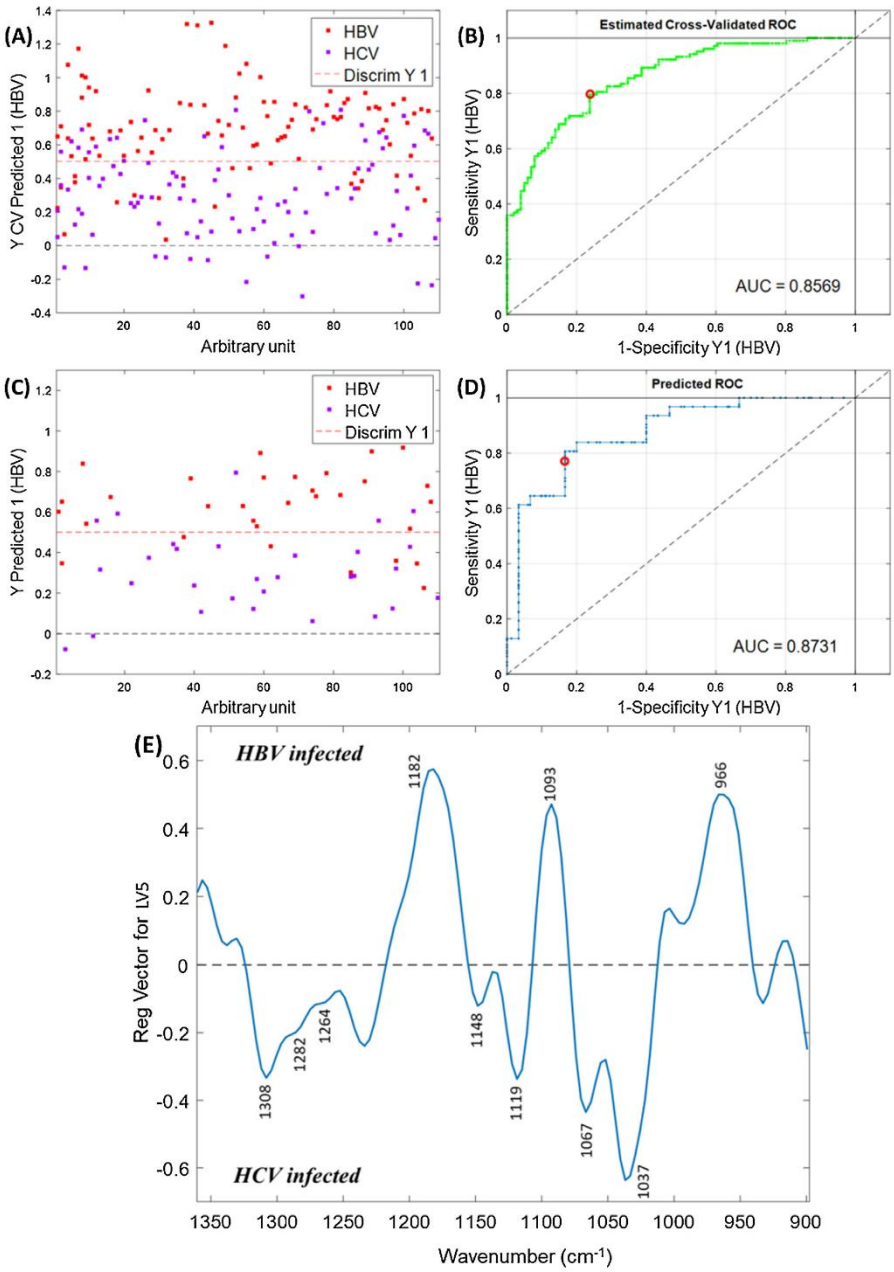


Fig. 6. HBV vs HCV PLS-DA models, spectra acquired from whole serum. A). Cross-validated Y values plot. B). Estimated ROC for Cross-Validation. C). Predicted Y values plot. D). Estimated ROC for prediction. E). Regression vector of PLS-DA models (CV & Predicted) for the dataset HBV vs HCV, where minima bands are associated with HCV spectra and maxima bands are associated with HBV spectra.

virion. The genome contains two highly conserved untranslated regions (UTR) 5'-UTR and 3'-UTR, which are flanking a single open reading frame encoding the viral proteins. In spite of the single-stranded RNA genomic structure, the 5' and 3' UTR contains numerous numbers of stable stem-loop or pseudo-DNA knots. These structures are double-stranded in nature [46]. Fig. 4 shows the classification models for HCV and control samples. Fig. 4A and B show the estimated Y values and ROC curve, respectively, for the cross-validation, while Fig. 4C and D show the estimated Y values plot and ROC curve, respectively, for the independent testing of the models. For HCV, classification accuracy was slightly less than for HBV, showing high values of specificity (>90 %) but sensitivity values around 80 %, with AUROC values of more than 90 % for both independent testing and cross validation.

Fig. 4E shows the corresponding regression vector for LV 9. The maxima bands at 1148 cm^{-1} and 1100 cm^{-1} in the regression vector shown in Fig. 4E are associated with HCV infected sample spectra and assigned to glycogen and sugar molecules [18,43,47]. From the reference spectra of IgG (Fig. 1.C) the intense bands at 1076 cm^{-1} and 989 cm^{-1} can be assigned to Ig in the infected serum. The positive bands at 1691 cm^{-1} and 1015 cm^{-1} are assigned to RNA modes [48]. The presence of lipid-associated bands at 1379 cm^{-1} and 1186 cm^{-1} in the negative regression vector indicates a decrease in total lipid in the infected samples as mentioned previously. A shift of ester carbonyl band from positive maxima (1747 cm^{-1}) to negative minima (1732 cm^{-1}) may indicate a change of lipid composition between infected and control samples. Lipid bands are expected in the positive regression vector from triglycerides in HCV lipo-viro-particles (LVP). [49] The positive bands, shown in the regression vector, are associated with spectra from HCV patients. The positive band at 1635 cm^{-1} is assigned to the β-pleated sheet amide I mode, while bands at 1654 cm^{-1} and 1505 cm^{-1} are observed as minima bands indicating a decrease in α-helical protein during hepatitis. This is corroborated by Raman and ATR-FTIR studies [24,26] with HCV infected plasma, where spectral markers associated with lipid and proteins moieties are the indication of their alterations occurred upon HCV infection.

Positives vs Controls by applying serum directly onto the ATR crystal

Fig. 5 A and B show the estimated Y values plot and ROC curve, respectively, for the cross-validation. Fig. 5C and B show the predicted Y values plot and ROC curve, respectively, for the independent validation of positives (HBV & HCV samples combined) vs control PLS-DA models. Modelling of both infections combined yielded higher classification errors than when the virally infected sera was directly compared to the control sera with AUROC values of 87.6 % and 85 % for cross validation and prediction, respectively.

Fig. 5E shows the corresponding regression vector for LV 8. Bands similar to the HBV vs controls and HCV vs controls PLS-DA regression vector were found in the regression vector, as discussed above in the previous two sections on modelling. All the band assignments are shown in Table 2. It can be concluded that the major differences can be attributed to the amount of Ig between infected and uninfected samples.

HBV vs HCV by applying serum directly onto the ATR crystal

Fig. 6 A and B show the estimated Y values and ROC curve for the cross-validation, while Fig. 6C and D show the predicted Y values and ROC curve for the independent validation, respectively for HBV vs HCV PLS-DA models. Classification error values (Table 3) are higher compared to models using controls and virally infected sera with AUROC values 85 % for both cross-validated and predicted dataset.

Fig. 6E shows the corresponding regression vectors for LV 5. In the regression vector, positive bands are associated with HBV infection and negative bands are associated with HCV infection. The 1182 cm^{-1} band can be assigned to cholesterol esters, while the band at 966 cm^{-1} is from the DNA structure of HBV. The band at 1093 cm^{-1} is due to the presence of polysaccharide N-glycan as discussed above. The minima bands at 1119 cm^{-1}, 1037 cm^{-1}, and 1067 cm^{-1} are attributed to RNA modes of the HCV virus.

PLS-DA cross-validated modelling on the dataset of high and low molecular weight samples

A study on vibrational spectroscopic analysis of serum focussed on improving the results by purifying the serum proteins using centrifugal filtration. [52] ATR-FTIR spectra were recorded from the serum samples separated using Amicon ultra centrifugal filter device. The results obtained from high molecular weight component dataset had a lower classification error (approximately 10 %–15 %) compared to the dataset of low molecular weight metabolites, suggesting that the whole serum classification might rely largely on information from high molecular weight proteins, glycoproteins, and lipoproteins. By comparing with the PLS-DA models (HBV vs controls and HCV vs controls) obtained from whole serum, the classification was improved by 1–3 % when using the high molecular weight serum fraction. The PLS-DA cross-validated models from the high molecular weight fraction are summarized in Table 4. The PLS-DA estimated Y values plots and corresponding ROC curves for high molecular weight sera vs controls are presented in Fig. 7 along with the corresponding regression vectors of positives vs controls (high molecular fraction). In both figures (Fig. 7C and F), similar maxima bands to those observed with the whole serum modelling (Figs. 2 and 4) associated with infection could be associated with immune response and the presence of Ig. The cross-validated models yielded from the high molecular weight concentrates gave better classification results compared to low molecular weight fractions and whole serum, with the AUROC values of 92.7 % and 93.3 %, respectively for HBV vs negatives and HCV vs negatives dataset.

Table 4
Summary of PLS-DA model for the ATR-FTIR data with cross-validation, spectra acquired on the samples containing high molecular weight (HMW) and low molecular weight (LMW) components.

Analyte	Preprocessing	Region (cm^{-1})	(Cross Validation)				
			LV	Classification Error (%)	Sensitivity (%)	Specificity (%)	AUROC
HBV-HMW vs Control-HMW	Smoothing (9) Baseline correction SNV MC	1791-895	8	8.8 %	87.5 %	94.9 %	92.7 %
HCV-HMW vs Control-HMW	Smoothing (9) Baseline correction SNV MC	1791-895	8	14.3 %	81.6 %	89.7 %	93.3 %
HBV-LMW vs Control-LMW	Smoothing (9) Baseline correction SNV MC	1776-895	12	23.6 %	69.4 %	83.3 %	84.5 %
HCV-LMW vs Control-LMW	Smoothing (9) Baseline correction SNV MC	1776-895	11	25.7 %	68.4 %	80.3 %	77.7 %

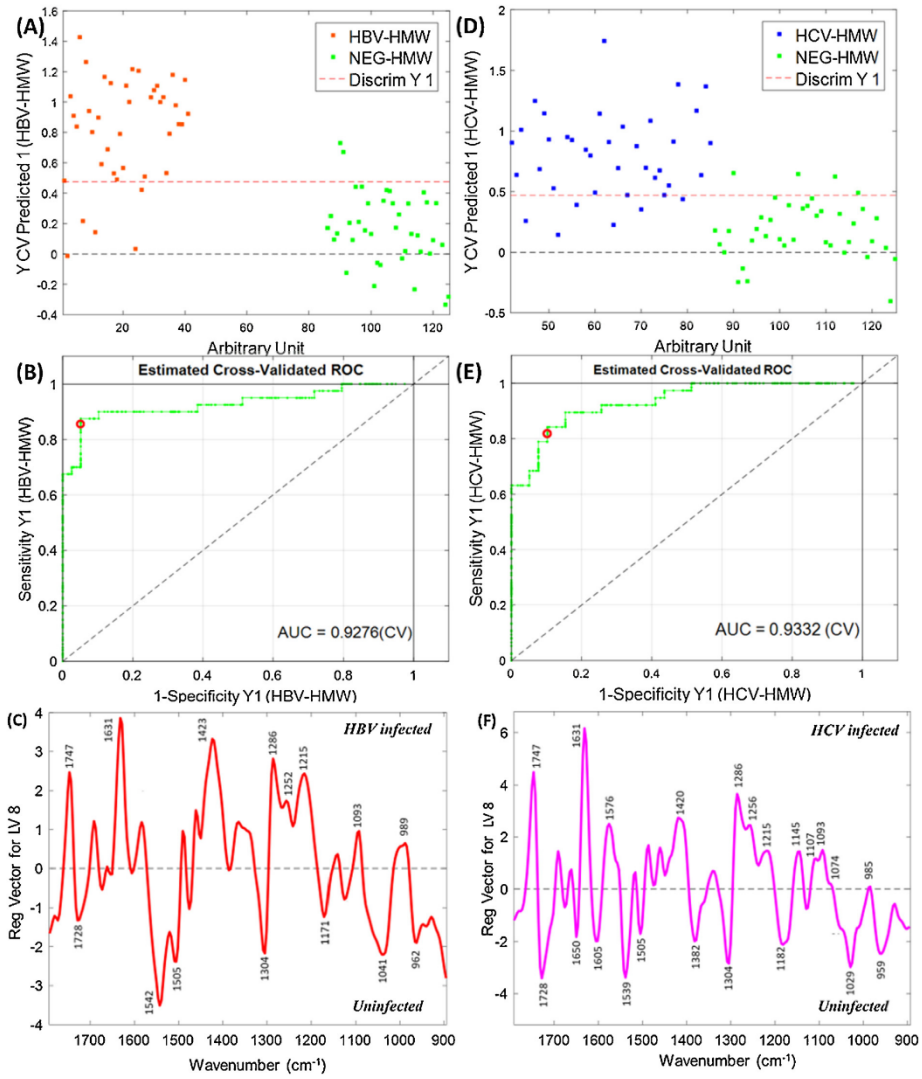


Fig. 7. PLS-DA cross-validated models with the dataset of high molecular weight components A). Estimated Y values of HBV vs Controls. B). Corresponding ROC. C). HBV vs Controls regression vector (8 latent varibles). D). Estimated Y values of HCV vs Controls. E). ROC for HCV vs uninfected controls. F). Regression vector for HCV invected vs unininfected controls (8 latent varaibles).

Partial Least Square Regression (PLS-R) of viral load using sera samples dried onto glass cover slips

PLS-R modelling is based on finding linear relationships between observed and predicted variables. The viral load ranges 10-10^2 IU/mL, 10^2-10^3 IU/mL, 10^3-10^4 IU/mL, 10^4-10^5 IU/mL, 10^5-10^6 IU/mL, 10^6-10^7 IU/mL, 10^7-10^8 IU/mL and 10^8-8×10^8 IU/mL for HBV sera samples dried onto glass cover slips were taken to construct the PLS-R models. The regression models work best if the modelling is performed over smaller ranges. If the data was modelled throughout the entire range 10 to 10^9 IU/mL, the cross-validation error would increase and hence the accuracy of the model would decrease. The best models were selected based on the most accepted CV error using a reasonable number of latent variables (LVs). Table 5, shows the latent variable number with corresponding R^2 values. However, there was no linearity found in the model for HBV infected samples within these ranges of viral loads. This implies that modelling for the amount of whole virus, even at higher concentration is

Table 5
R^2 value for the PLSR model for a various range of hepatitis B and C viral load.

HBV			HCV		
Selected concentration (IU/mL)	R^2 value (CV)	LV	Selected concentration (IU/mL)	R^2 value (CV)	LV
$10\text{-}10^2$	0.2145	12	$10\text{-}10^2$	0.215	10
$10^2\text{-}10^3$	0.3713	12	$10^2\text{-}10^3$	0.458	11
$10^3\text{-}10^4$	0.255	10	$10^3\text{-}10^4$	0.222	8
$10^4\text{-}10^5$	0.356	11	$10^4\text{-}10^5$	0.402	14
$10^5\text{-}10^6$	0.382	12	$10^5\text{-}10^6$	0.464	11
$10^6\text{-}10^7$	0.124	14	$10^6\text{-}8 \times 10^6$	0.72	12
$10^6\text{-}10^7$	0.059	10			
$10^8\text{-}8 \times 10^8$	0.485	9			

not achievable by ATR-FTIR spectroscopy and therefore putative spectral markers of the virus itself should be viewed with caution. It should be noted that the signal-to-noise using the glass cover slip approach is less compared to placing serum directly onto the ATR crystal, which could also affect the R^2 values.

Model construction for HCV infected samples with a viral load ranging from $10\text{-}10^2$ IU/mL, $10^2\text{-}10^3$ IU/mL, $10^3\text{-}10^4$ IU/mL, $10^4\text{-}10^5$ IU/mL, $10^5\text{-}10^6$ IU/mL and $10^6\text{-}8 \times 10^6$ IU/mL were performed but similar to the case with HBV, no linearity was observed.

Comparison with existing diagnostic technologies. The ATR-FTIR technique can be compared with existing serological-based assays, molecular-based assays, and RDT techniques based on sensitivity and specificity of the model and its applicability to point-of-care diagnosis. While molecular-based and serological-based assays perform extremely well for HBV and HCV detections (achieving about 98 %–100 % of sensitivity and specificity) [11–13], they are less suitable as a point of care diagnostic tool for reasons stated above. The RDT method can diagnose HBV with a sensitivity of 90 % and a specificity of 99.5 % [15] and HCV with a sensitivity of 78.8 % and a specificity of 100 % [16], which are comparable with the results from ATR-FTIR analysis (with 2–6 % difference). The ATR-FTIR technique also shows potential in the simultaneous detection of health markers and disease markers similar to other approaches [18]. Thus, upon considering the improved longevity and robustness of the ATR-FTIR technique, it seems to be a suitable candidate for HBV and HCV intial point-of-care screening for deployment in remote undeveloped communities where other genetic based or antibody tests are impractical.

Conclusion

ATR-FTIR spectroscopy was performed on negative serum controls and HBV and HCV infected serum samples by: i) drying the serum onto a glass cover slip and inverting the slide so the serum sample is in direct contact with the ATR crystal. ii) Drying the whole serum sample onto the ATR crystal; and iii), by separating the serum into high molecular weight and low molecular weight concentrates, with subsequent deposition on the ATR crystal. The fractionation approach provided the best sensitivity and specificity, although the time taken to prepare the samples is greater than directly placing the serum onto the ATR crystal. The glass cover slip approach is eminently suitable for high throughput but did not produce the same sensitivity and specificity as drying the sera directly onto the ATR crystal. The major changes observed in the spectra relate to changes in immunoglobulin (Ig), lipid and glucose levels. A marker band at 1093 cm^{-1} was observed in the case of HBV infection, which was assigned to hepatitis B surface antigen virus like particles (HBsAg VLPs).

The qualitative validations were performed using PLS-DA and the classification error for the modelling of HBV vs control samples are 12.10 % (whole serum) and 8.8 % (separated concentrate) for the cross-validation modelling. Data prediction was performed on whole serum from a randomly selected independent data set with a classification error of 11.20 %. For HCV infected samples, the classification error was 15.20 % (whole serum) and 14.30 % (high molecular weight concentrate) in cross-validation. For whole serum from a randomly selected independent data set the error was 11.30 %. In order to improve the classification error nonlinear methodologies such as Support Vector Machine or Random Forest feature selection will be trailed in the future.

In both cases, HBV vs control and HCV vs control models generated using the low molecular weight dataset, gave the highest classification error, which suggests that the PLS-DA models are dependent on bands related to high molecular weight components present in the serum. A PLS-R model for HBV and HCV infected samples with different spiked amounts of virion particles was performed and no linearity was found for several ranges indicating the major changes observed in the spectra were related to immune response or compounds produced by the virus as opposed to direct detection of the virion particles.

ATR-FTIR shows promise as an initial screening approach for initial detection of underlying infection and appears to be able to discriminate HBV and HCV infected samples. However, since the approach appears to detect the response to infection or compounds synthesized by the virus rather than the presence of the virus itself, more work is required to ascertain how specific the response is to different viruses and other infectious agents, before a viral diagnostic based on ATR-FTIR spectroscopy can be realized.

Declaration of Competing Interest

BRW, PH and DPG declare they have a US patent "Spectroscopic systems and methods for the identification and quantification of pathogens" US Patent App. 16/326,686. BRW and PH are Directors on the boards of two spin-off companies Total Blood Profile and Biotech Resources (BTR) which aims to commercialise the invention described in this patent.

Acknowledgements

This work was finacially supported by an Australin Research Discovery grant DP180104068 and an Australian Research Council Future Fellowship (FT120100926) supported BRW for the initial part of this this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.clispe.2020.100001.

References

[1] P.E. Milhiet, P. Dosset, C. Godefroy, C. Le Grimellec, J.M. Guigner, E. Larquet, F. Ronzon, C. Manin, Biochimie 93 (2011) 254–259.
[2] F.M. Averhoff, N. Glass, D. Holtzman, Clin. Infect. Dis. 55 (2012) S10–S15.

[3] A. Kishore, B. Nagpal, M. N. Sumana and U. Hegde, 2016, 6, 150–153.
[4] T. Millard, J. Elliott, S. Girdler, AIDS Patient Care STDS 27 (2013) 103–113.
[5] M. Kozicki, J. Czepiel, G. Biesiada, P. Nowak, A. Garlicki, A. Wesołucha-Birczyńska, Analyst 140 (2015) 8007–8016.
[6] V. Di Martino, T. Thevenot, J.-F. Colin, N. Boyer, M. Martinot, F. Degos, J.-P. Coulaud, J.-L. Vilde, F. Vachon, C. Degott, D. Valla, P. Marcellin, Gastroenterology 123 (2002) 1812–1822.
[7] A.M. Di Bisceglie, Hepatology 49 (2009) S56–S60.
[8] J.D. Stanaway, A.D. Flaxman, M. Naghavi, C. Fitzmaurice, T. Vos, I. Abubakar, L.J. Abu-Raddad, R. Assadi, N. Bhala, B. Cowie, M.H. Forouzanfour, J. Groeger, K.M. Hanafiah, K.H. Jacobsen, S.L. James, J. MacLachlan, R. Malekzadeh, N.K. Martin, A.A. Mokdad, A.H. Mokdad, C.J.L. Murray, D. Plass, S. Rana, D.B. Rein, J.H. Richardus, J. Sanabria, M. Saylan, S. Shahraz, S. So, V.V. Vlassov, E. Weiderpass, S.T. Wiersma, M. Younis, C. Yu, M. El Sayed Zaki, G.S. Cooke, Lancet 388 (2016) 1081–1088.
[9] M. Homs, M. Buti, J. Quer, R. Jardí, M. Schaper, D. Tabernero, I. Ortega, A. Sanchez, R. Esteban, F. Rodriguez-Frias, Nucleic Acids Res. 39 (2011) 8457–8471.
[10] M. Helal, R. Ranjbar and S. M. Alavian, Hepatitis Monthly, DOI:10.5812/hepatmon.17632.
[11] C. Colin, D. Lanoir, S. Touzet, L. Meyaud-Kraemer, F. Bailly, C. Trepo, J. Viral Hepat. 8 (2001) 87–95.
[12] S.A. Galel, T.L. Simon, P.C. Williamson, J.P. AuBuchon, D.A. Waxman, Y. Erickson, R. Bertuzis, J.R. Duncan, K. Malhotra, J. Vaks, N. Huynh, L.L. Pate, Transfusion 58 (2018) 649–659.
[13] J.D. Scott, D.R. Gretch, JAMA 297 (2007) 724.
[14] A. Sakudo, Y. Suganuma, T. Kobayashi, T. Onodera, K. Ikuta, Biochem. Biophys. Res. Commun. 341 (2006) 279–284.
[15] A. Amini, O. Varsaneux, H. Kelly, W. Tang, W. Chen, D.I. Boeras, J. Falconer, J.D. Tucker, R. Chou, A. Ishizaki, P. Easterbrook, R.W. Peeling, BMC Infect. Dis. 17 (2017) 698.
[16] M. Hee Kim, S. Young Kang, W. In Lee, J. Virol. Methods 193 (2013) 379–382.
[17] D. Perez-Guaita, S. Garrigues, M. de la, TrAC - Trends in Analytical Chemistry 62 (2014) 93–105.
[18] S. Roy, D. Perez-Guaita, D.W. Andrew, J.S. Richards, D. McNaughton, P. Heraud, B.R. Wood, Anal. Chem. 89 (2017) 5238–5245.
[19] K. Spalding, F. Bonnier, C. Bruno, H. Blasco, R. Board, I. Benz-de Bretagne, H.J. Byrne, H.J. Butler, I. Chourpa, P. Radhakrishnan, M.J. Baker, Vib. Spectrosc. 99 (2018) 50–58.
[20] A. Khoshmanesh, M.W.A. Dixon, S. Kenny, L. Tilley, D. McNaughton, B.R. Wood, Anal. Chem. 86 (2014) 4379–4386.
[21] M.C.D. Santos, Y.M. Nascimento, J.M.G. Araújo, K.M.G. Lima, RSC Adv. 7 (2017) 25640–25649.
[22] P. Larkin, Infrared and Raman Spectroscopy (2011) 27–54.
[23] D.I. Ellis, R. Goodacre, Analyst 131 (2006) 875.
[24] H. Nawaz, N. Rashid, M. Saleem, M. Asif Hanif, M. Irfan Majeed, I. Amin, M. Iqbal, M. Rahman, O. Ibrahim, S.M. Baig, M. Ahmed, F. Bonnierg, H.J. Byrnee, J. Raman Spectrosc. 48 (2017) 697–704.
[25] A. Sakudo, R. Tsenkova, T. Onozuka, K. Morita, S. Li, J. Warachit, Y. Iwabu, G. Li, T. Onodera, K. Ikuta, Microbiol. Immunol. 49 (2005) 695–701.
[26] C. Pizarro, I. Esteban-Díez, I. Arenzana-Rámila, J.M. González-Sáiz, J. Biophotonics 11 (2018) e201700035.
[27] L. Sitole, F. Steffens, T.P.J. Krüger, D. Meyer, OMICS 18 (2014) 1–11.
[28] A.A. Mastoi, World J. Gastroenterol. 16 (2010) 603.
[29] G. Zhang, Q. Zhao, C. Xie, L. Peng, T. Zhang, Z. Gao, J. Immunol. Res. 2019 (2019) 1–10.
[30] Z. Ortarık, A. Toyran, S. Sen, S.Z. Mart Kömürcü, E. Güvener, Mikrobiyol. Bul. 45 (2011) 296–305.
[31] N.W. Tietz, C.A. Burtis, E.R. Ashwood, D.E. Bruns, *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*, 5th edn., Elsevier Saunders, Missouri, 2006.
[32] M.A. Crook, Clinical Biochemistry & Metabolic Medicine, 8th edn., CRC Press, London, 2013.
[33] Q. R. Ahmad, D. H. Nguyen, M. A. Wingerd, G. M. Church and M. A. Steffen, Proteome Science, DOI:10.1186/1477-5956-3-6.
[34] C. Petibois, G. Cazorla, A. Cassaigne, G. Déléris, Clin. Chem. 47 (2001) 730–738.
[35] D.L. Peterson, J. Biol. Chem. 256 (1981) 6975–6983.
[36] R. Oropesa, J.R. Ramos, V. Falcón, A. Felipe, Adv. Nat. Sci. Nanosci. Nanotechnol. 4 (2013) 25007.
[37] W.S. Tan, K.L. Ho, World J. Gastroenterol. 20 (2014) 11650.
[38] D.S. Dane, C.H. Cameron, M. Briggs, Lancet 295 (1970) 695–698.
[39] A. Mehta, X. Lu, T.M. Block, B.S. Blumberg, R.A. Dwek, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 1822–1827.
[40] M. Werr, R. Prange, J. Virol. 72 (1998) 778–782.
[41] J. Hu, K. Liu, Viruses 9 (2017) 1–17.
[42] K.Z. Liu, R.A. Shaw, A. Man, T.C. Dembinski, H.H. Mantsh, Clin. Chem. 48 (2002) 499–506.
[43] J. Vongsvivut, P. Heraud, A. Gupta, M. Puri, D. McNaughton, C.J. Barrow, Analyst 138 (2013) 6016.
[44] A. Agbecha, C. Usoro, M. Etukudo, Chrismed J. Health Res. 4 (2017) 81.
[45] D. Siagris, M. Christofidou, G.J. Theocharis, N. Pagoni, C. Papadimitriou, A. Lekkou, K. Thomopoulos, I. Starakis, A.C. Tsamandas, C. Labropoulou-Karatza, J. Viral Hepat. 13 (2006) 56–61.
[46] S. Chevaliez, J.-M. Pawlotsky, Hepatitis C Viruses: Genomes and Molecular Biology (2006) .
[47] F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L.F.C.S. Carvalho, D. Bertrand, M.J. Baker, H.J. Byrne, I. Chourpa, Analyst 142 (2017) 1285–1298.
[48] B.R. Wood, Chem. Soc. Rev. 45 (2016) 1980–1998.
[49] S. Agaugué, L. Perrin-Cocon, P. André, V. Lotteau, PLoS One 2 (2007) e330.
[50] D.R. Whelan, K.R. Bambery, P. Heraud, M.J. Tobin, M. Diem, D. McNaughton, B.R. Wood, Nucleic Acids Res. 39 (2011) 5439–5448.
[51] K. Kaiden, T. Matsui, S. Tanaka, Appl. Spectrosc. 41 (1987) 180–184.
[52] F. Bonnier, M.J. Baker, H.J. Byrne, Anal. Methods 6 (2014) 5155–5160.

Spectroscopy goes Viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy

Supti Roy[1], David Perez-Guaita[1], Scott Bowden[2], Philip Heraud[1,3*], Bayden R. Wood[1*]

[1] *Centre for Biospectroscopy, Monash University, Clayton, 3800, Victoria, Australia*

[2] *Victorian Infectious Disease Reference Laboratory, Melbourne, Victoria, Australia*

[3] *Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia*

Electronic Supplementary information (ESI):

Spectroscopy goes Viral: Diagnosis of hepatitis B and C virus infection from human sera using ATR-FTIR spectroscopy

Supti Roy[1], David Perez-Guaita[1], Scott Bowden[2], Philip Heraud[1,3*], Bayden R. Wood[1*]

[1] *Centre for Biospectroscopy, Monash University, Clayton, 3800, Victoria, Australia*

[2] *Victorian Infectious Disease Reference Laboratory, Melbourne, Victoria, Australia*

[3] *Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia*

Table 1: S1 Comparison of sensitivity and specificity in PLS-DA models obtained from thick film on glass and direct placement of serum onto the crystal slide. Preprocessing steps considered smoothing (9 points), baseline correction, SNV, and mean centering.

Analyte	**Region (cm^{-1})**	**(Cross Validation)**				**(Prediction)**			
		LV	**Classification error (%)**	**Sensitivity (%)**	**Specificity (%)**	**LV**	**Classification error (%)**	**Sensitivity (%)**	**Specificity (%)**
HBV vs Controls (Glass slide)	1762-906	10	21.3%	76.6%	80.6%	10	28.4%	69.4%	73.7%
HBV vs Controls (ATR-crystal)	1762-906	8	12.10%	84.3%	91.3%	8	11.2%	84.4%	93.10%
HCV vs Controls (Glass slide)	1762-906	11	19.6%	78.0%	82.8%	11	28.9%	51.3%	90.9%
HCV vs Controls (ATR-crystal)	1762-906	9	15.2%	77.2%	92.30%	9	11.3%	80.0%	97.20%

Table 7: S2: LV vs cross validation error for whole sera directly placed on the crystal and analyzed with ATR-FTIR. Models were selected based on the lowest value of CV error

Positives vs Negatives		HBV vs Negatives		HCV vs Negatives		HBV vs HCV	
LV	CV Error	LV	CV Error	LV	CV Error	LV	CV Error
1	0.315	1	0.327	1	0.226	1	0.359
2	0.277	2	0.278	2	0.187	2	0.271
3	0.269	3	0.229	3	0.186	3	0.246
4	0.244	4	0.234	4	0.172	4	0.265
5	0.231	5	0.19	5	0.162	5	0.221
6	0.214	6	0.171	6	0.182	6	0.216
7	0.207	7	0.17	7	0.147	7	0.231
8	0.188	8	0.122	8	0.162	8	0.226
9	0.195	9	0.151	9	0.152	9	0.221
10	0.195	10	0.146	10	0.157	10	0.186
11	0.194	11	0.131	11	0.152	11	0.187
12	0.204	12	0.122	12	0.147	12	0.187
13	0.207	13	0.131	13	0.162	13	0.182
14	0.209	14	0.141	14	0.147	14	0.186
15	0.208	15	0.112	15	0.152	15	0.186

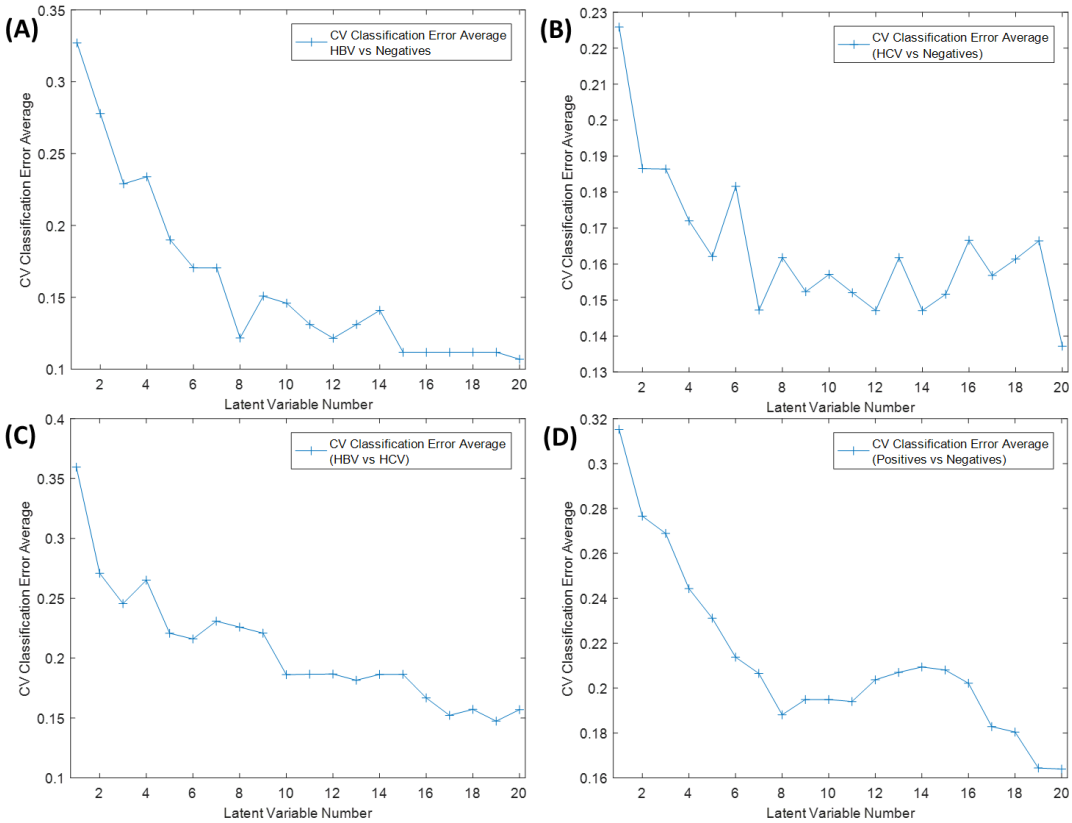


Figure 8-S1: Latent variable vs Cross validation error plots A).HBV vs Negatives B)HCV vs Negatives C)HBV vs HCV D)Positives vs Negatives using the dataset of whole sera directly placed on the ATR-crystal.

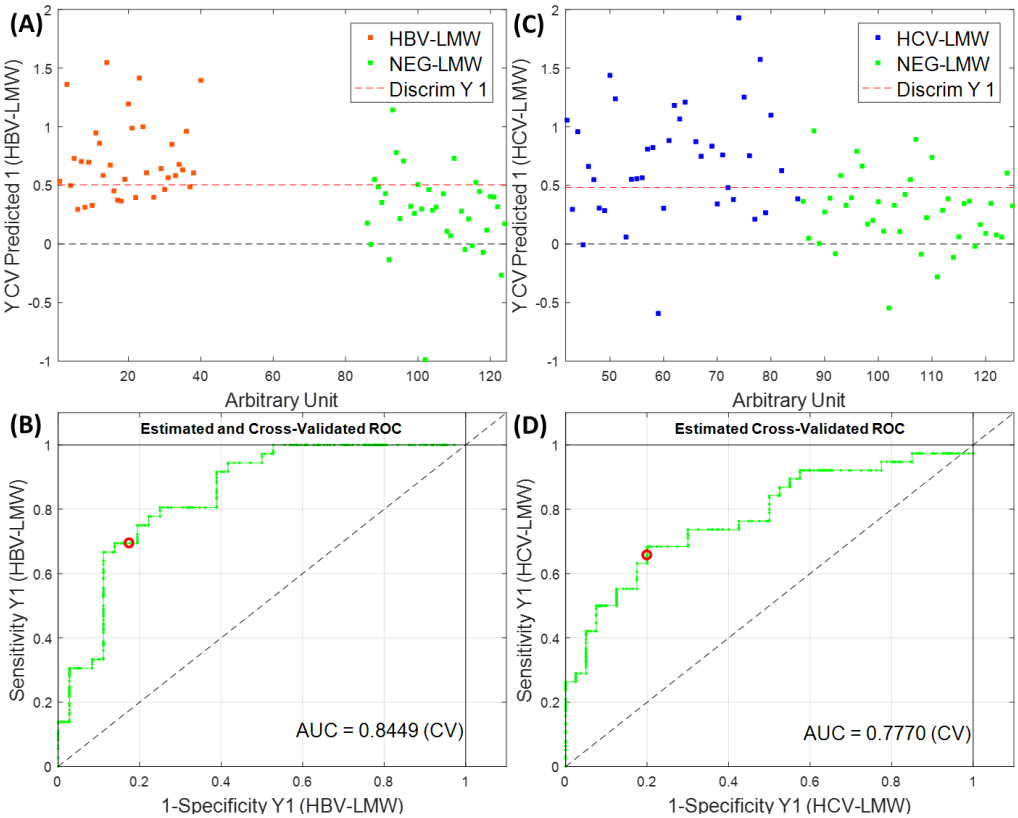


Figure 9-S2: PLS-DA cross-validated models with the dataset of low molecular weight components A) Estimated Y values of HBV vs Controls and B) corresponding ROC D) Estimated Y values of HCV vs Controls E) corresponding ROC.

Chapter 5. Monitoring the infection of a human hepatoblastoma cell line by hepatitis B virus using synchrotron FTIR micro-spectroscopy and Raman micro-spectroscopy

COMMUNICATION

Monitoring the infection of a human hepatoblastoma cell line by hepatitis B virus using synchrotron FTIR micro-spectroscopy and Raman micro-spectroscopy

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Supti Roy,[a] Kamila Kochan,[a,*] Danni Colledge,[b] Diana E. Bedolla,[a] Scott Bowden,[b] Philip Heraud,*[a,c] and Bayden R. Wood[a]

A complementary study applying synchrotron FTIR and Raman micro-spectroscopy revealed a decreased in lipid accumulation for HepG2.2.15 cells upon hepatitis B transfection. The results from Unsupervised Hierarchical Cluster Analysis and Principal Component Analysis of Raman data showed intense bands (2849 cm^{-1} and 3007 cm^{-1}) for control HepG2 cells which confirm the presence of lipid components in the cells.

The World Health Organization (WHO) estimates that two billion people have been infected with hepatitis B virus (HBV), a member of the Hepadnaviridae family. Despite the availability of an efficacious vaccine, there are currently 250 million people, predominantly in Asia, suffering from chronic or persistent HBV infection. Of these chronically infected individuals, approximately 25% will develop liver cancer, which is the fifth most common cancer worldwide, ranking third as a cause of cancer mortality. The Global Burden of Disease Study 2010 has estimated that the total number of deaths attributable to HBV infection was 786,000 and when combined with 499,000 deaths from HCV infection, viral hepatitis ranks as one of the highest causes of human mortality.[1,2] Several studies have attempted to delineate host-virus interaction to devise biomarkers that will be beneficial for rapid diagnosis and treatment.[3–7] However, there is a lack of understanding due to complex expression of virus within the host.

The liver is the largest solid organ in the body and is responsible for a wide variety of critical metabolic functions, including the metabolism of carbohydrates, proteins and lipids. Hepatocytes, the liver cells, are the target for HBV and chronic infection is associated with fatty liver and progressive liver disease, which may lead to liver cirrhosis and hepatocellular carcinoma upon infection. Kim *et al*[8] and Yun-li Wu *et al*[9] found that the HBV X protein (HBx) is responsible for upregulation of multiple lipogenic genes, which promotes intracellular lipid accumulation. The gene expression includes fatty acid binding protein 1 (FABP1), sterol regulatory element-binding protein 1 (SREBP1) and peroxisome proliferator-activated receptor (PPARγ). Furthermore, a comparative lipid profile study on cell culture suggested an increase in both saturated and unsaturated fatty acids with HBV infection.[10] Contrary to this, a negative association of HBV infection with fatty liver was also reported by *in vivo* studies.[11,12] A decline of lipid bodies (LBs) due to dysfunction of CIDE protein in the cultured cells with HBV infection was reported by Yasumoto *et al.*[13] These studies demonstrate that lipid accumulation in hepatocytes is a result of HBV infection, a topic of great debate in the scientific community.

Human hepatoblastoma cell lines have been studied extensively for understanding the chemistry between HBV virus and host cells.[10,13–15] The HepG2 human hepatoblastoma cell line was stably transfected with HBV genomes to derive the HepG2.2.15 cell line. Such a stable transfection is found to allow expression of viral genes with active replication and generation of viral progeny.[14]

Raman and synchrotron FTIR spectroscopy are capable of detecting cellular lipids and discriminate biological samples and infections with high signal to noise ratio.[16–19] The single cell measurements can provide the greatest chance to discriminate between infected and uninfected cells. Based on the spectroscopic signatures, it is possible to identify the changes associated with the presence of a virus. Furthermore, these changes can be localised within single cells, due to the high spatial resolution enables of both techniques.

In this study, the infected (HepG2.2.15) and the control (HepG2) cell lines were characterized using Raman and synchrotron FTIR microspectroscopy to understand the interaction between cell lines upon HBV infection at the subcellular level. Due to the size of HBV virus (20-42 nm in diameter) it is not possible to detect the virus directly by these two techniques, but the effects on cellular metabolism, leading to changes in the composition of cytoplasm with respect to e.g. lipid/protein ratio or relative content of proteins with

[a] *Centre for Biospectroscopy, Monash University, Clayton, 3800, Victoria, Australia.*
[b] *Victorian Infectious Disease Reference Laboratory, the Doherty Institute, Melbourne, Victoria, Australia.*
[c] *Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia.*
Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

different secondary structures, might be detected. The single cell spectral data was analysed using Principal Component Analysis (PCA) to resolve the ambiguity in the literature related to the lipid content of HBV infected cells. The results provide an insight to the complex processes occurring in response to HBV infection at the single cell level, contributing towards our understanding of the changes that occur upon HBV infection in the liver.

Experimental

Cell culture and transfection

Stably transfected HepG2.2.15 cell line with the HBV DNA and control HepG2 cell line were received from Victorian Infectious Disease Reference Laboratory (792 Elizabeth Street, Melbourne, 3000). These cells were cultured and maintained in the Minimum Essential Media (Gibco by Life Technologies, USA), supplemented with 10% foetal bovine serum (Bovogen, Interpath, Australia), incubated under 5% CO_2 at 37°C. After 7 days, the cells were harvested by trypsinization at 70%- 80% confluence, pelleted by centrifugation, washed 3 times with 0.9% sterile saline, fixed with freshly prepared 4% formalin solution and stored at 4 °C. The concentration of the control cell suspension was approx. 7×10^6 cells/mL in 4% formalin solution.

Cell attachment on CaF_2 windows by Cytospin method

The control and infectious cell suspensions were diluted (20-25 times) with 0.9% saline. An aliquot of 200 µL of the diluted cell suspension was centrifuged for 5 minutes at 2,000 r.p.m. in a sample chamber of a Cytospin 4 cytocentrifuge (Thermo Scientific, USA) containing the CaF_2 Raman grade slide, in order to obtain a monolayer of single cells deposited on the CaF_2 window. The slides were kept at 4°C until commencement of the synchrotron FTIR and Raman measurements.

Raman micro-spectroscopy

Three independent cell culture (3 days) experiments were performed for control (HepG2) and infected (HepG2.2.15) cell lines. From each culture, two CaF_2 slides were prepared from each cell line as technical replicates (discussed in the "Cytospin" section). The measurements were performed within 2 weeks of the cell preparation. In total, 111 cells (48- HepG2, 63- HepG2.2.15) were mapped using WITech Confocal Raman microscope (WITech CRM alpha 300 R, Ulm, Germany). This instrument was equipped with a 532 nm excitation wavelength laser in an air cooled solid state along with a Charged Coupled Device (CCD) detector. An optical fibre with a diameter of 50 µm was used to couple the laser with the instrument. All spectra were acquired by using 100x (NA 0.9) and 50x (NA 0.8) Olympus dry objectives. Before starting measurements, each time a calibration was done with Raman scattering line produced by a silicon plate (520.5 cm^{-1}). Cell imaging was performed with a sampling density of 0.3 µm. An average time required to collect each map was 30-45 minutes, depending upon the cell size. Spectra were collected in the range 0 – 3705 cm^{-1} with the resolution of 3 cm^{-1}. The laser power was kept at 12 mW. The integration time of 0.6 s was used for a single spectrum.

Synchrotron FTIR micro-spectroscopy

In total 422 single point (192- HepG2 and 230- HepG2.2.15) measurements were acquired in 3 days of synchrotron slot. For each cell line, three technical replicates (cells attached on CaF_2 substrate) were prepared from a single day culture, as discussed in the previous section. The single cell measurements were performed in the synchrotron FTIR microspectroscopy facility at the Australian Synchrotron. Measurements were recorded on a Bruker Vertex 80V FTIR spectrometer coupled with Bruker Hyperion 3000 FTIR microscope and a liquid nitrogen cooled mercury cadmium telluride (MCT) detector. The spectra were collected in the mid IR range 4000–800 cm^{-1}, using the aperture size at 12 µm × 12 µm with the spectral resolution of 4 cm^{-1}. 128 and 64 co-added scans were collected for measuring background and sample respectively.

Data analysis

Raman spectroscopy

An average Raman spectrum of each cell map was extracted for comparison between control and infected spectra. Raman imaging data was analysed with k-means cluster (KMC) analysis in the spectral range of 3122-600 cm^{-1} without preprocessing. During the clustering, cosmic ray spikes were removed manually. Spectra with similar spectral profiles were merged together, and ultimately the whole cell map was clustered into four groups. For clarity, these groups were labelled as 1. Cell edge (the outer cluster), 2. Cytoplasm, 3. Nucleus (identified through the presence of intense band at 785-790 cm^{-1}) and 4. Lipids (intense bands C-H stretching region). Not every mapped cell demonstrated the presence of a lipid body. In some cases, the KMC analysis revealed the presence of a fifth class, labelled as 5. Cytochrome (intense band at 745 cm^{-1}). An average spectrum was extracted from each cluster. The initial analysis was performed with WITec Project Plus™ software.

Synchrotron FTIR data

A single point spectrum was collected from each cell. Data was extracted after eliminating the spectral contribution due to water vapour using the OPUS 8.0 algorithm.

Data pre-processing and Principal Component Analysis (PCA)

Data preprocessing was done by using PLS toolbox from Eigenvector in MATLAB 9.1 from Mathworks (Natick, U.S.A). For achieving, overlaid mean spectra from whole cell, Savitzky-Golay 2nd derivatives (25 smoothing points) and standard normal variate (SNV) preprocessing was performed on Raman data extracted from each cell using WITech Project Plus software (average spectra of classes, obtained by KMC). PCA analysis was performed within the region 3097-2698 cm^{-1} and combining 1800-600 cm^{-1} and 3137-2666 cm^{-1} by adding an extra preprocessing step 'mean centering'. The venetian blinds algorithm was used for cross-validation of the dataset. Of 111 cell spectra, 105 (control – 45, infected – 60) were included in the PCA model. The outliers (5.4% spectra lying outside 95% confidence interval) were excluded based on hoteling and T^2 values. For synchrotron FTIR data, pre-processing included

Savitzky-Golay 2nd derivatives (15 smoothing points) and Standard Normal Variate (SNV) and was performed on extracted FTIR data by OPUS software (average spectra of control and infected cells). Water vapour correction was performe before FTIR data extraction.

Results and Discussion

Comparison of cell lines

The average Raman spectra (3122-2750 cm^{-1} and 1800-600 cm^{-1}) and synchrotron FTIR spectra (3050-2800 cm^{-1} and 1800-900 cm^{-1}) are shown in the Figure 1A-B and 1C-D, respectively, for both the cell lines. Intense bands are present in the high wavenumber (C-H stretching) region of both the Raman and FTIR spectra. In the Raman spectra (Figure 1A), the HepG2 line exhibits more intense symmetric and asymmetric CH_2 stretching bands at 2851 cm^{-1} and 2889 cm^{-1} (shoulder) compared to the HepG2.2.15 cells. Furthermore, the 2935 cm^{-1} and 2976 cm^{-1} bands corresponding to the symmetric and asymmetric CH_3 stretching vibrations from lipid moieties are also more intense in the HepG2 cells compared to the HepG2.2.15 cells. A less intense shoulder band is present at 2986 cm^{-1} only for HepG2.2.15 in the average Raman spectrum. The Raman bands at 2889 cm^{-1}, 2935 cm^{-1}, 2976 cm^{-1} (intense in HepG2) and 2986 cm^{-1} (shoulder in HepG2.2.15) are also associated with C-H stretching vibrations from protein within the cell.[20] Additionally, the presence of the band at 3007 cm^{-1} in the Raman spectrum (Figure 1A) is assigned to the methine stretching vibration and is marker for unsaturated lipids. This band is present in both cell lines, but slightly more intense in the HepG2 cell line. The bands located at ~ 2850 cm^{-1} and ~ 3009 cm^{-1} are confirmatory for lipid C-H stretches. These are unique identifiers for lipid in the Raman spectra and are not dominated by protein C-H stretches. Similar intense C-H aliphatic stretching region for control cells can be observed in the synchrotron FTIR spectra (Figure 1C) along with the ester carbonyl band at 1742 cm^{-1} (Figure 1D). However, the band at 2873 cm^{-1} (Figure 1C) showed greater intensity for the HepG2.2.15 transfected cell line. In Figure 1B, the fingerprint region of Raman spectra shows intense bands at 750 cm^{-1}, 1127 cm^{-1}, 1311 cm^{-1} and 1583 cm^{-1} for HepG2.2.15, which are mostly associated with heme-containing molecules (cytochrome/proteins) of cytoplasm.[21] The fingerprint region of synchrotron FTIR average spectra (Figure 1D) also showed relative intense amide bands (1652 cm^{-1}, 1634 cm^{-1} and 1512 cm^{-1}) for HepG2.2.15 cell line along with the intense DNA associated peaks (1716 cm^{-1}, 1084 cm^{-1}, 1056 cm^{-1}, 961 cm^{-1}). The band at 961 cm^{-1} is blue shifted to 966 cm^{-1} in the control cell line, possibly indicative of DNA conformational changes occurring inside the cells.[22]

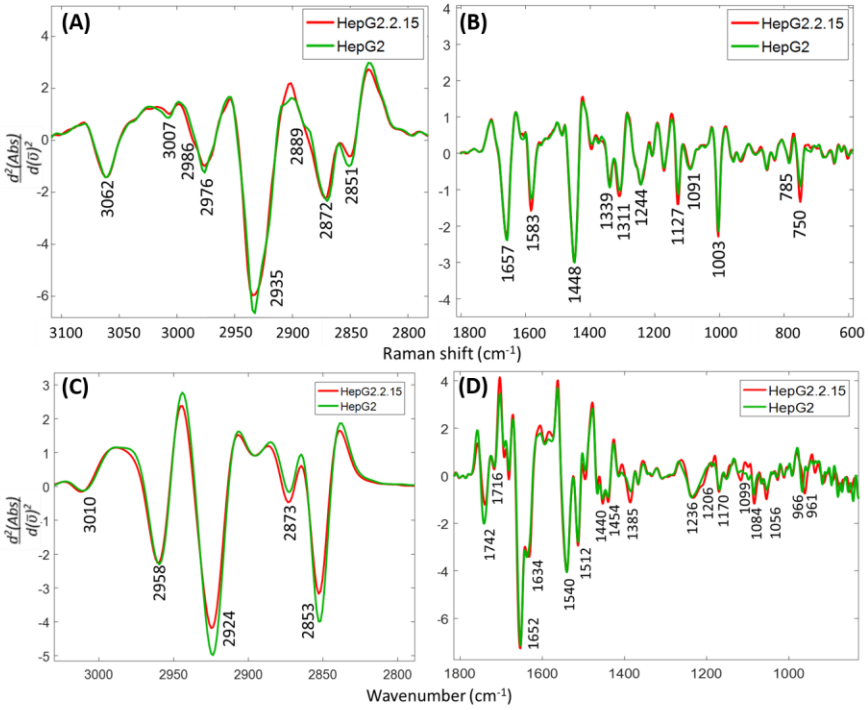


Figure 1: Normalised Second derivatives of average (**A**, **B**) Raman and (**C**, **D**) Synchrotron FTIR spectra of cell lines HepG2 (control, in green) and HepG2.2.15 (transfected, in red) in the spectral ranges of (A, C): 3050 – 2800 cm^{-1} and (B, D) 1800 – 600 cm^{-1}. Major band assignments are detailed in Table 1.

Formation of intracellular lipid bodies

Several studies have reported that HBV infection is associated with lipid accumulation in hepatocytes,[9,23] including cultured cells.[10] However, many other *in vivo* and *in vitro* studies report no such accumulation.[11–13] One of these studies, in particular ones with HepG2/HepG2.2.15 cell lines[13] reported that lipid storage was significantly higher in HepG2 compared with HepG2.2.15. Moreover, the average size of the LBs in both cell lines were measured by transmission electron microscopy and reported to be greater than the HepG2.2.15 cells, compared to HepG2 cells. In the current study, with Raman mapping, some of the transfected and control cells showed lipid accumulation in the cells. The vertical spatial resolution of the Raman setup is around 0.8 µm. This penetration depth is unable to cover the entire cell (thickness: 10-12 µm). Figure 2A and 2B represent the cell cluster imaging for a single HepG2 and HepG2.2.15 cell, respectively. In Figure 2A, the blue cluster corresponds to the lipid bodies of HepG2 cell and in Figure 2B, the blue cluster corresponds to the lipid body in HepG2.2.15 cell. Several lipid bodies are clearly visible in the HepG2 cells, whereas in the HepG2.2.15 cells only a single lipid body can be identified. In addition, the size of the LBs (~1 µm^2) are in agreement to those observed with transmission electron microscopy study.

Spectral characterization of cell lines

For a detailed investigation of the HepG2 and HepG2.2.15 cells, PCA was performed on second derivative, normalised Raman spectra of whole cells in the region 3097 cm^{-1} - 2786 cm^{-1} and combination of 3115 - 2781 cm^{-1} and 1800-600 cm^{-1}. In Figure 3A, the scores plot shows the positive scores are correlated with the HepG2.2.15 cell line and the negative scores with HepG2 cells. The loadings plots (Figure 3B&C) show the important variables responsible for the separation observed in the PC1 versus PC2 scores plot. In the PC1 vs PC2 scores (Figure 3A) the separation occurs along PC1. The PC1 positive loadings are correlated with the negative scores (marked as purple in Fig. 3A) and the negative loadings are associated with positive scores (marked as green in Fig 3A) because the second derivative was used for the analysis. Each point in the scores plot is representative of a single averaged spectrum from a Raman map of separate cells. In the loadings plot in C-H aliphatic stretching region (Figure 3B), strong positively loaded bands can be found at 2851 cm^{-1}, 2903 cm^{-1}, 2920 cm^{-1}, 2935 cm^{-1} and 3009 cm^{-1}. Of these, 2851 cm^{-1} and 3009 cm^{-1} are associated with CH_2 symmetric stretching vibration and the =C-H stretching vibration from lipids, respectively. This confirms the initial observation based on average Raman spectra, demonstrating a higher lipid content in the control HepG2 cells. Other bands correspond to the C-H stretching mode are either from lipids or from proteins in the cells. Strong negative loadings are observed at 2830 cm^{-1}, 2885 cm^{-1}, 2912 cm^{-1}, 2926 cm^{-1}, 2941cm^{-1} and

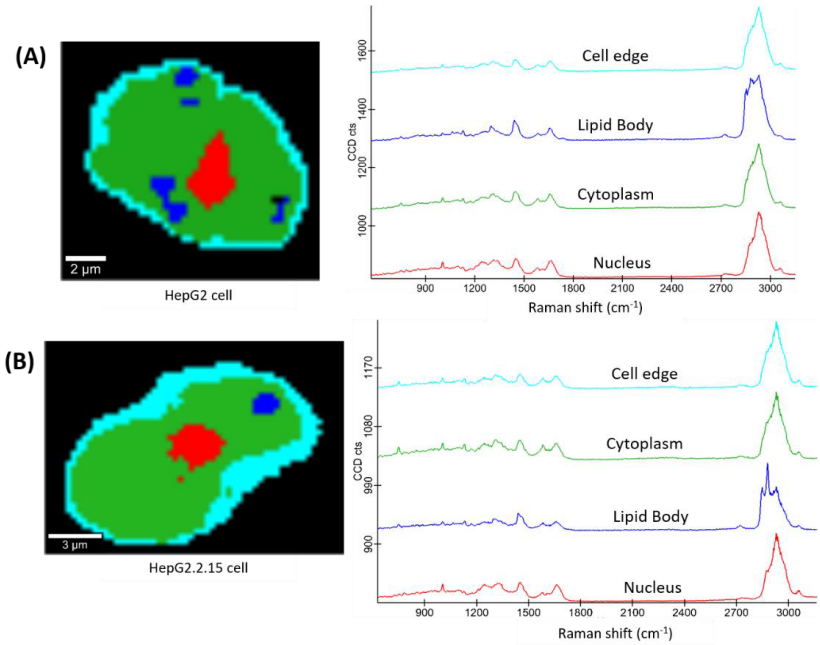


Figure 2: The results of KMC analysis of Raman ma for **(A)** HepG2 control cell and **(B)** HepG2.2.15 infected cell. Each panel presents the distribution of classes together with corresponding cluster spectra.

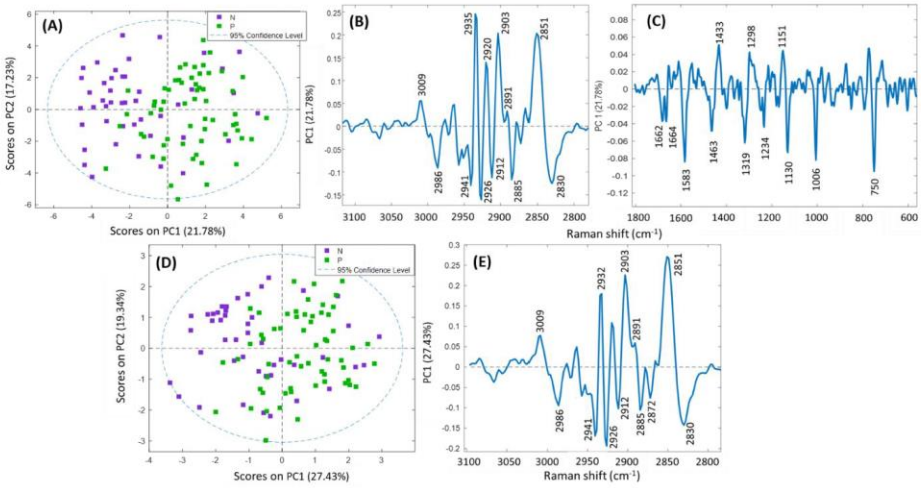


Figure 3: Upper panel shows: **(A)** PCA score plot combining the region 3115 cm^{-1}- 2781 cm^{-1} and 1800 cm^{-1}-600 cm^{-1} and corresponding loadings of PC1 (**B** & **C**). The lower panel shows (**D**) PCA score plot with the region 3097 -2786 cm^{-1} together with corresponding (**E**) PC1 loading.

2986 cm^{-1}, assigned to C-H stretching modes of proteins. The presence and discriminatory influence of several protein bands in this region shown in Figure 3B (in both, positive and negative loadings) is a clear indication of difference in protein composition in both cell lines. The intense bands explaining the separation of spectral clusters in the scores plot, are the negatively loaded bands (Figure 3C) at 750 cm^{-1}, 1006 cm^{-1}, 1130 cm^{-1}, 1319 cm^{-1} and 1583 cm^{-1} corroborating differences observed with the average Raman spectra shown in Figure 1B. The difference in thickness of the two cell lines may be the reason for these spectral differences. A slight difference in cell size is noticeable[24] between the HepG2 and HepG2.2.15 cell lines, with the latter being slightly larger.

Table 1: Major band assignments (FTIR and Raman) of HepG2 and HepG2.2.15 cell lines.

FT-IR band positions (cm-1)	Assignments	Raman band positions	Assignments	HepG2	HepG2.2.15
-	-	3009	ν(=CH), lipids	Increase	Decrease
-	-	2986	ν(C-H), proteins	Decrease	Increase
-	-	2976	$\nu_{as}(CH_3)$, lipids or ν(C-H), proteins	Increase	Decrease
2924	$\nu_{as}(CH_2)$, lipids	2935	$\nu_s(CH_3)$, lipids or ν(C-H), proteins	Increase	Decrease
2873	$\nu_s(CH_3)$, lipids	2889	$\nu_{as}(CH_2)$, lipids or ν(C-H), proteins	Increase	Decrease
2853	$\nu_s(CH_2)$, lipids	2851	$\nu_s(CH_2)$, lipids	Increase	Decrease
1652	Amide I, proteins, α- helical	1583, 1311, 1127, 750	Cytochromes/ Proteins	Decrease	Increase
1634	Amide I, proteins, β pleated sheet	-	-	Decrease	Increase
1512	Amide II, proteins	-	-	Decrease	Increase
1716	DNA, (νC=O & νC=N)	-	-	Decrease	Increase
1084	DNA, $\nu_s PO_2^-$	-	-	Decrease	Increase
1056	DNA, ν(C-O)	-	-	Decrease	Increase
961	DNA, ribose-phosphate skeletal motions	-	-	Blue shifted	-

Flow cytometry of both the cell lines revealed that G1/S cell cycle phase is arrested in HepG2.2.15 cells, which may lead to the size difference between these cell lines.[15]. The PCA performed using only the higher wavenumber region (3097-2698 cm^{-1}) shows a similar result (Figure 3D and E) compared with the combination of the upper and lower wavenumber regions, indicating that separation is mostly based on the high wavenumber region in Raman spectra.

Changes in intracellular composition due to the presence of virus

Figure 4A, 4B and 4C are representative of the average Raman spectra for each cluster characterizing the cytoplasm, lipid bodies and nucleus, respectively, for both the HepG2 and HepG2.2.15 cell lines. The cytoplasm average spectra show similar bands with similar intensities compared to the average spectra of the whole cell. Average spectra from LBs showed a significant increase in the intensity of bands at 2849 cm^{-1} and 3007 cm^{-1} for HepG2 cell, where the intensity of the symmetric CH_2 stretching band (2849 cm^{-1}) is significantly higher compared to asymmetric CH_2 stretching band (2883 cm^{-1}). On the other hand, a similar intensity can be observed for both symmetric (2849 cm^{-1}) and asymmetric (2883 cm^{-1} and a shoulder at 2873 cm^{-1}) CH_2 stretching bands in LB spectra of HepG2.2.15 transfected cells. The Raman bands at 2872 cm^{-1} and 2889 cm^{-1} in the HepG2 whole cell and cytoplasm cluster are assigned to the asymmetric CH_2 of proteins rather than lipids. The symmetric CH_2 stretching band (2849 cm^{-1}) is a confirmatory marker for lipids. The observation is in agreement with the observation from the cell cluster maps (Figure 3A and 3B) and Yasumoto *et al*,[13] where HepG2 cells were reported to show an increase in the accumulation of lipids, leading to a formation of larger LBs. It is expected that even though the lipid content increased, is not yet sufficient enough to form large LBs. Although it was not possible to record spectra of the entire volume of the whole cell, from the cytoplasm cluster average spectra we can observe a trend, clearly indicating the changes in the relative amount of lipid molecules in the cells.

The clinical relationship between HBV infection and lipid accumulation in the hepatocytes is still unclear. The mechanisms responsible for the lipid accumulation has been reported as upregulation of lipogenic and adipogenic genes due to the infection.[8,9] In addition, the presence of virus can inhibit

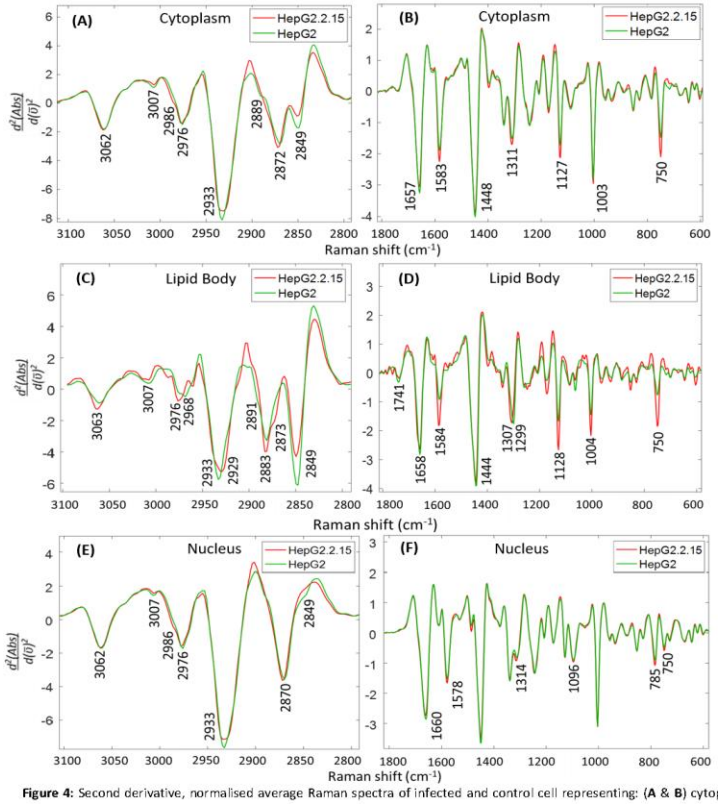


Figure 4: Second derivative, normalised average Raman spectra of infected and control cell representing: (**A** & **B**) cytoplasm, (**C** & **D**) lipid body and (**E** & **F**) nucleus cluster.

the secretion of apolipoprotein B from cells, ultimately leading to the accumulation of TG and cholesterol.[25] Nonetheless, several clinical studies reported that lipid accumulation is not associated with HBV infection. Studies have found an inverse association with fatty liver and hepatic steatosis due to HBV viral infection.[11,12,26,27] In the findings of Yasumoto *et al*[13] the authors suggested that inhibition of CIDE protein expression (a protein that aids in lipid storage in hepatocytes) during infection is responsible for reduced LB size in the HepG2.2.15 transfected cell lines. Furthermore, several studies have reported that the HBV level is inversely correlated with the mRNA level and protein expression of apolipoprotein within the HepG2.2.15 cells.[28–30] Our findings are in line with the outcome of these reports, which show an inverse relation between HBV infection and storage of lipid inside the cells.

The nucleic acid associated bands (785 cm^{-1}, 1096 cm^{-1}, 1486 cm^{-1} and 1578 cm^{-1}) are more prominent for HepG2.2.15 cells in the average spectra of nucleus cluster (Figure 4C). This could indicate more packed DNA inside the nucleus of the cells. Additionally, HepG2.2.15 cells have extra full length DNA from the hepatitis B virus itself incorporated inside the host cell genome.[14] However, it is important to state that slight differences in the Raman confocal plane between cells can result in different penetration depths along with the thickness of the cells as discussed above.

Conclusions

In this study, we applied Raman and synchrotron FTIR spectroscopy to investigate the effect of the HBV infection in the control HepG2 and transfected HepG2.2.15 cell lines. Our aim was to characterise the spectral features related to host-virus interaction. Raman and FTIR are complementary to each other and can assist in confirmatory analysis to shed new light in understanding the cellular biochemistry during infection. In this study, we compared Raman and synchrotron FTIR data, which allowed monitoring the changes in lipid, protein and nucleic acids at sub-cellular level. The application of PCA on Raman spectral dataset demonstrated the separation between spectral clusters from the different cell lines occurring along PC1 in PCA scores plots. The corresponding PC1 loadings plot revealed that the separation was predominantly based on bands from C-H stretching region associated with a lower relative lipid content in the HBV transfected cells. In addition to this, the analysis of cytoplasm clusters (identified *via* KMC) showed an increased amount of both, saturated and unsaturated lipids in HepG2 cytoplasm. This may be associated with early stage of lipid accumulation, preceding the formation of lipid bodies (lipid content in cytoplasm increased, but not yet sufficient to form a LB). The Raman map of a single cell revealed that HBV infection led to formation of LBs of smaller size, consequently translating to shortage of lipid storing capability. The synchrotron FTIR data supports the Raman data showing control HepG2 cells contains more lipid. More work is required to ascertain the specific type of lipids that decrease during HBV transfection, which is the subject of future GC-MS studies. This study highlights the significant effects HBV transfection has on hepatocytes and the important role lipids play in response to transfection.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal and S. Y. Ahn, *The lancet*, 2013, **380**, 2095–2128.

2 J. H. MacLachlan and B. C. Cowie, *Cold Spring Harbor Perspectives in Medicine*, 2015, **5**, a021410.

3 C. Pizarro, I. Esteban-Díez, I. Arenzana-Rámila and J. M. González-Sáiz, *Journal of Biophotonics*, 2018, **11**, e201700035.

4 A. A. Mastoi, *World Journal of Gastroenterology*, 2010, **16**, 603–607.

5 Z. Ortarık, A. Toyran, S. Sen, S. Z. Mart Kömürcü and E. Güvener, *Mikrobiyoloji bulteni*, 2011, **45**, 296–305.

6 A. Agbecha, C. Usoro and M. Etukudo, *CHRISMED Journal of Health and Research*, 2017, **4**, 81–86.

7 S. Roy, D. Perez-Guaita, S. Bowden, P. Heraud and B. R. Wood, *Clinical Spectroscopy*, 2020, **1**, 100001.

8 K. H. Kim, H. J. Shin, K. Kim, H. M. Choi, S. H. Rhee, H. B. Moon, H. H. Kim, U. S. Yang, D. Y. Yu and J. Cheong, *Gastroenterology*, 2007, **132**, 1955–1967.

9 Y. Wu, X. Peng, Y. Zhu, X. Yan, W. Chen and X. Lin, *Journal of Virology*, 2016, **90**, 1729–1740.

10 H. Li, W. Zhu, L. Zhang, H. Lei, X. Wu, L. Guo, X. Chen, Y. Wang and H. Tang, *Scientific Reports*, 2015, **5**, 8421.

11 V. W. S. Wong, G. L. H. Wong, W. C. W. Chu, A. M. L. Chim, A. Ong, D. K. W. Yeung, K. K. L. Yiu, S. H. T. Chu, H. Y. Chan, J. Woo, F. K. L. Chan and H. L. Y. Chan, *Journal of Hepatology*, 2012, **56**, 533–540.

12 Y. L. Cheng, Y. J. Wang, W. Y. Kao, P. H. Chen, T. I. Huo, Y. H. Huang, K. H. Lan, C. W. Su, W. L. Chan, H. C. Lin, F. Y. Lee and J. C. Wu, *PLoS ONE*, 2013, **8**, e72049.

13 J. Yasumoto, H. Kasai, K. Yoshimura, T. Otoguro, K. Watashi, T. Wakita, A. Yamashita, T. Tanaka, S. Takeda and K. Moriishi, *Journal of General Virology*, 2017, **98**, 1762–1773.

14 M. a Sells, M. L. Chen and G. Acs, *Proceedings of the National Academy of Sciences of the United States of America*, 1987, **84**, 1005–1009.

15 R. Zhao, T.-Z. Wang, D. Kong, L. Zhang, H.-X. Meng, Y. Jiang, Y.-Q. Wu, Z.-X. Yu and X.-M. Jin, *World journal of gastroenterology*, 2011, **17**, 1152–1159.

16 K. Kochan, E. Kus, E. Szafraniec, A. Wislocka, S. Chlopicki and M. Baranska, *The Analyst*, 2017, **142**, 3948–3958.

17 G. T. Webster, K. A. De Villiers, T. J. Egan, S. Deed, L. Tilley, M. J. Tobin, K. R. Bambery, D. McNaughton and B. R. Wood, *Analytical Chemistry*, 2009, **81**, 2516–2524.

18 D. McNaughton and B. R. Wood, *Australian Journal of Chemistry*, 2012, **65**, 218–228.

19 K. Kochan, E. Kus, A. Filipek, K. Szafrańska, S. Chlopicki and M. Baranska, *The Analyst*, 2017, **142**, 1308–1319.

20 N. K. Howell, G. Arteaga, S. Nakai and E. C. Y. Li-Chan, *Journal of Agricultural and Food Chemistry*, 1999, **47**, 924–933.

21 A. Chaichi, A. Prasad and M. Ranjan Gartia, *Biosensors*, 2018, **8**, 107.

22 B. R. Wood, *Chem. Soc. Rev.*, 2016, **45**, 1980–1998.

23 M. Hajjou, R. Norel, R. Carver, P. Marion, J. Cullen, L. E. Rogler and C. E. Rogler, *Journal of Medical Virology*, 2005, **77**, 57–65.

24 X. Zhu, C. Xie, Y. M. Li, Z. L. Huang, Q. Y. Zhao, Z. X. Hu, P. P. Wang, Y. R. Gu, Z. L. Gao and L. Peng, *Cell Death and Disease*, 2016, **7**, 1–12.

25 S. K. Kang, T. W. Chung, J. Y. Lee, Y. C. Lee, R. E. Morton and C. H. Kim, *Journal of Biological Chemistry*, 2004, **279**, 28106–28112.

26 M. V. Machado, A. G. Oliveira and H. Cortez-Pinto, *Journal of Gastroenterology and Hepatology (Australia)*, 2011, **26**, 1361–1367.

27 T. H. Chung, M. C. Kim and C. S. Kim, *Korean Journal of Family Medicine*, 2014, **35**, 81–89.

28 P. A. Norton, Q. Gong, A. S. Mehta, X. Lu and T. M. Block, *Journal of Virology*, 2003, **77**, 5503–5506.

29 C. Zhu, G. Gao, H. Song, F. Xu, K. Wu and X. Liu, *Lipids in Health and Disease*, 2016, **15**, 1–6.

30 C. L. Zhu, Y. Li, G. S. Gao, P. A. Zhang and X. S. Wei, *Chinese Journal of Microbiology and Immunology (China)*, 2012, **32**, 504–507.

Chapter 6. Conclusion and Future Work

6.1 Conclusion

The overall aim of the research project was to develop a point-of-care diagnostic tool using Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy to detect blood pathogens and analytes in host blood and serum samples. The target pathogens were Hepatitis B (HBV), hepatitis C (HCV) viruses (causal agents responsible for human hepatitis) and *Plasmodium falciparum* parasites, the causal agent for malaria infection. The hypotheses were two-fold: 1) Malaria infection in whole blood and associated analytes can be detected and quantified using ATR-FTIR spectroscopy. 2) Pathogenic and immune infrared markers for HBV and HCV viral antigens can be detected and used as diagnostic indicators to build multivariate models for point-of-care viral detection. Attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy, synchrotron infrared micro-spectroscopy, Raman micro-spectroscopy and atomic force microscope infrared-spectroscopy (AFM-IR) were applied to investigate the pathogenic response.

Chapter 1 introduces the target diseases malaria and hepatitis, literature survey, background about spectroscopic techniques and scopes for ATR-FTIR. It also highlights the objectives of this thesis.

Chapter 2 details the instrumentation used in the project, which includes ATR-FTIR spectroscopy, Raman micro-spectroscopy, synchrotron IR micro-spectroscopy, AFM-IR spectroscopy. Chapter 2 also explains the chemometric tools that were employed for data analysis.

Chapter 3 describes the application of ATR-FTIR spectroscopic method to detect and quantify *P. falciparum* infection and two associated metabolites namely urea and glucose simultaneously in whole blood. Portable ATR-FTIR spectroscopy has several advantages compared to enzymatic diagnosis for glucose and urea. First, it can generate results within a very short time (10 minutes including sample preparation and measuring time). The technique only requires a pinprick of blood to obtain a spectrum, making it ideal for infants who are the primary victims in malaria epidemic regions. Moreover, the dried thick smear preparation of whole blood samples on a glass slide is well suited for the remote regions where facility of preservation and transportation is a big issue. However, as discussed in Chapter 4, it does not have the same contact on the ATR crystal compared to directly drying the blood on the crystal

and hence the signal-to-noise ratio is smaller. In combination with multivariate analysis, the PLS-DA model using this approach resulted in a 98% specificity but a low sensitivity (70%) due to the lower number of control samples. It should be noted that only higher parasitemia could be detected using whole blood (>0.5%) compared to isolated red blood cells, where the parasitemia detection was found to be 0.0001 %.[1] The prediction capability using PLS-R models based on the root mean square cross validation error (RMSECV) values for malaria parasites (0-5%), glucose (0-400 mg/dL) and urea (0-250 mg/dL) concentrations were 0.58%, 30.48 mg/dL and 22.70 mg/dL, respectively. There is a still scope for improving the limit of detection using a digital filtration approach, recently developed by our group, by calculating the difference spectrum between control and infected samples and then squaring the difference spectrum and adding the standard deviation to generate a threshold function. This will be applied in future studies and would reduce the limit of detection.

Chapter 4 focuses on the detection of hepatitis viruses (HBV and HCV) in human clinical sera samples and spectroscopic biomarkers to detect viral antigens and host-virus interactions. Initially we performed the PLS-DA models generated from the whole serum using the glass slide method as previously reported.[2] The classification error for the predicted models were extremely high, 28.4% for HBV vs controls and 28.9% for HCV vs control, compared to the models obtained by placing the whole serum directly onto the ATR crystal due to insufficient contact between the sample and the ATR-crystal.

To further, improve the modelling PLS-DA was performed on clinical serum samples and separated into high and low molecular weight fractions. The prediction error of the PLS-DA model of HBV vs controls was 11.20%, whereas for HCV vs controls was 11.30% using whole serum. The error values were significantly lower for high molecular weight molecules compared to the low weight components, proving that the discrimination between the infected and uninfected serum samples were largely dependent on the high molecular weight molecules. The differences in spectral features were observed in the region responsible for antibodies (immunoglobulins), lipid and glucose. A band observed at 1093 cm^{-1} was assigned to the HBsAg (surface antigen for HBV) and confirmed using AFM-IR of isolated HbsAg VLP. The quantitative PLS-R regression analysis showed no linear relationship with the spiked viral samples, indicating the spectral signals are mostly associated with the other analytes in the sample (due to the host immune responses) rather than the viral particles. Overall, the ATR-FTIR has proven to be a promising diagnostic tool for HBV and HCV as an initial screening approach. However, judicious choice of the spectral region for modelling is important in discriminating between the immune response and direct detection of viral components.

Chapter 5 focusses on monitoring the human hepatoblastoma cell lines with and without transfected HBV full length DNA. The aim of this study was to determine what are the specific marker bands of HBV and HCV viral infections. Raman and synchrotron FTIR micro-spectroscopy were employed to analyse cells at the single cell level. The average Raman and synchrotron spectra from singles cells demonstrated more intense C-H stretching in the HepG2 controls. Furthermore, Unsupervised Hierarchical Cluster Analysis (UHCA) was used to discriminate and lipid bodies in the cytoplasm and Principal Component Analysis (PCA) performed on Raman data set of the two cell lines showed lipid confirmatory bands (2849 cm^{-1} and 3007 cm^{-1}) showing an increase in HepG2 control compared to the HBV transfected cell lines. The results indicate that, the storing capability of lipid molecules decreased with the HBV infection in the infected HepG2.2.15 cells. There is debate in the literature regarding the lipid accumulation in HBV infected cells. Our work is in agreement with a few other studies[3–5] that reported lower lipid accumulation in HepG2.2.15 cells compared to controls.

6.2 Future work

ATR-FTIR spectroscopy has been proven to be an efficient tool for the diagnosis of HBV and HCV, malaria parasites and blood analytes. However, more work is required to make it a robust and applicable tool for point-of-care diagnostic tool. The proposed future work includes:

1. To truly quantify malaria parasites, glucose and urea, the PLA-DA and PLS-R model needs to be applied in clinical *Plasmodium* infected blood samples.
2. To increase the sensitivity of the PLD-A model the number of negative samples required needs to be increased.
3. To diagnose hepatitis, the PLS-DA models of clinical serum samples (HBV vs negative and HCV vs negatives) requires more samples to be included in an independent test set from different geographical populations, the number in total should be at least 1000 samples to be applied in real field.
4. Coinfection with both HBV and HCV should also be modelled using ATR-FTIR analysis.
5. To investigate the spectral signatures related to the increase in immunoglobulin (Ig) upon hepatitis infection, the experiment should be performed with other viruses (e.g. influenza, polyomavirus) infected serum to confirm whether the response is specific for hepatitis viruses (HBV and HCV) or generic for other viruses.
6. Spiking experiments with known immunoglobulin (Ig) concentrations should be performed to quantify the Ig response.

7. Culturing of viruses is the next stage to investigate the spectral features in clinical samples. The culturing requires a biosafety level 3 facility and it was not possible to conduct these experiments at VIDRL. Spiking experiments with the cultured virus particles can be performed for ATR-FTIR analysis and generate PLS-R model to quantify viral particles.
8. The isolated viral antigens hepatitis B core antigens (HBcAg), hepatitis B envelope antigens (HBeAg), HCV core antigen (HCVcAg) can be characterised using ATR-FTIR spectroscopy and analysing the clinical samples based on the antigenic spectral signature.
9. Performing analysis using other spectroscopic techniques including using a portable Raman spectrometer or a hand held near-IR spectrometer to detect hepatitis in clinical samples. Similar to ATR-FTIR, Near-IR spectroscopy offers the potential for a fast, reagent-free, cost-effective and minimally invasive diagnosis of HBV and HCV in clinical samples. This technique is currently being trailed in our lab for malaria infection. Moreover, the water signal is relatively weak and narrow (1404-1454 nm^{-1}) and it doesn't interfere with the other spectral region in the Near-IR compared to mid-IR region and could be applied to aqueous blood/serum, which would expedite the diagnostic method by removing the drying step.

6.3 References

1 A. Khoshmanesh, M. W. A. Dixon, S. Kenny, L. Tilley, D. McNaughton and B. R. Wood, *Analytical Chemistry*, 2014, **86**, 4379–4386.

2 S. Roy, D. Perez-Guaita, D. W. Andrew, J. S. Richards, D. McNaughton, P. Heraud and B. R. Wood, *Analytical Chemistry*, 2017, **89**, 5238–5245.

3 J. Yasumoto, H. Kasai, K. Yoshimura, T. Otoguro, K. Watashi, T. Wakita, A. Yamashita, T. Tanaka, S. Takeda and K. Moriishi, *Journal of General Virology*, 2017, **98**, 1762–1773.

4 P. A. Norton, Q. Gong, A. S. Mehta, X. Lu and T. M. Block, *Journal of Virology*, 2003, **77**, 5503–5506.

5 C. Zhu, G. Gao, H. Song, F. Xu, K. Wu and X. Liu, *Lipids in Health and Disease*, 2016, **15**, 1–6.