CRANFIELD UNIVERSITY

J BABRAH

A STUDY OF FT-IR SPECTROSCOPY FOR THE IDENTIFICATION AND CLASSIFICATION OF HAEMATOLOGICAL MALIGNANCIES

CRANFIELD HEALTH

PhD. THESIS

CRANFIELD UNIVERSITY

CRANFIELD HEALTH

PhD. THESIS

Academic Year 2008-2009

J BABRAH

A Study of FT-IR spectroscopy for the identification and classification of haematological malignancies

Academic Supervisors: N. Stone, C. Bessant
Clinical Supervisors: A. Rye, R. Lush, K. McCarthy

June 2009

This thesis is submitted in partial fulfilment of the requirements for the Doctor of Philosophy

Acknowledgements

This research is dedicated in the memory of Robert George Dalton. Robert was the co-founder of Cheltenham's Leukaemia and Intensive Chemotherapy Charity (LINC) .LINC financially supported my research project. I would like to specially thank Dr Gill Rouse, the co-ordinator of LINC for Dr Rouse's support and commitment to my research.

I would like to also thank my supervisors for their contributions to this research namely Dr Adam Rye, Dr Richard Lush, Dr Keith McCarthy for their great interest and guidance, Dr Conrad Bessant for his support and experience and Dr Nick Stone for his valuable suggestions and assistance throughout the research.

I would like to further thank Professor Mel Greaves and Ms Sue Colman who are based at the Institute of Cancer Research in Sutton for their contribution to culturing the cell lines. I am also extremely grateful for the hard work of Karol Baxter and Rajvinder Dhillion in the Cytology department at Gloucestershire Royal Hospital for all their help with preparation of the slides. Special thanks go to the nurses and laboratory staff, especially David Ah-Moye at the Haematology Department for their help in the collection of blood samples and allowing me the use of the laboratory facilities. Martin Isabelle also deserves a special gratitude for all his contributions in developing in-house analysis programmes. I would also like to thank Martin for all his help and continuous support throughout the project

Finally special thanks go to Kamar my partner for his continuous support and understanding throughout my PhD. I would also like to thank my Mum, Granddad, Grandma, and my sisters Ruby and Kiran for their continued encouragement.

Abstract

The aim of the work presented in this thesis was to explore the use of FT-IR spectroscopy, as a complementary clinical tool for haematological laboratory analysis.

FT-IR spectra were measured from air-dried and frozen cell lines derived from lymphoma, lymphoid, myeloid leukaemia and normal and chronic lymphocytic leukaemia blood samples. Multivariate statistical analysis was used to extract important spectral information with the greatest discriminative power. Principal component fed linear discriminant spectral models have been tested with leave one out cross validation procedures. A preliminary unfiltered classification model using 50 frozen and air-dried samples correctly classified 54% of 18556 spectra. The performance improved with the three cell line group datasets, with 71% of 19903 spectra correctly classified. Furthermore, the use of the frozen spectra improved the performance of the three cell line group classification model considerably. Findings showed that 73.3% of 9920 spectra were correctly classified in the frozen datasets, whereas in the air-dried only 41.5% of 9983 spectra are correctly classified. Optimisation of the spectral models by selection of principal components, application of Savitsky-Golay filters and selecting spectra using standard deviation and absorption filter tool was investigated. Using the first 25 significant PCs, a 0^{th} derivative Savitsky-Golay filter and the absorbance filter tool on the frozen five cell line spectral dataset were shown to be the optimal parameters for constructing a classification model. When tested with leave one batch out cross validation 90% of the spectra were correctly classified for the five cell line model. Blood component classification models tested with leave one batch out cross validation performed well. The whole blood model correctly classified 70% of 1736 spectra, measured on 22 samples. The plasma model correctly classified 80.6% of 331 spectra and the buffy coat model correctly classified 99.5% of 1438 spectra. This demonstrated that the buffy coat (containing white blood cells) holds the key biochemical information for discrimination between the pathology of the blood samples. Partial least squares analysis has been demonstrated as a method to support whole blood count tests for real time prediction of cellular constituents. These findings demonstrate the potential of FT-IR spectroscopy as a clinical tool although more work is needed if it is to be applied in clinical practice.

Table of Content

List of Abbreviations

ALL:	Acute lymphocytic leukaemia
AML:	Acute myeloid leukaemia
ANN:	Artificial neural network
ANOVA:	Analysis of variance
ATR:	Attenuated total reflection
BCC:	Basal cell carcinoma
C:	Celsius
CaF_2:	Calcium Fluoride
CDD:	Charge-couple device
CLL:	Chronic lymphoblastic leukaemia
CML:	Chronic myeloid leukaemia
CT:	Computed tomography
DNA:	Deoxyribose nucleic acid
EDTA:	Ethylene diamine tetra-acetic acid
ESS:	Elastic scattering spectroscopy
FAB:	French-American-British classification system
FACS:	Fluorescence activated cell sorter
FAD:	Flavin adenine dinucleotide
FBC:	Full blood count
FBS:	Foetal bovine serum
FISH:	Fluorescence in situ hybridization
FMN:	Flavin mononucleotide
FNA:	Fine needle aspirate
FPA:	Focal plane array
FS:	Fluorescence spectroscopy
FT:	Fourier transform
FT-IR:	Fourier transform infrared spectroscopy
HCA:	Hierarchical cluster analysis
HDL:	High density lipoprotein
H&E:	Haematoxylin and Eosin
HL:	Hodgkin's lymphoma
HSCs:	Haematopoietic stem cells
I:	Incident power density or irradiance
I_0:	Incident light intensity
IR:	Infrared
KBr:	Potassium Bromide
LDA:	Linear discriminate analysis
LDL:	Low density lipoprotein
LOOCV:	Leave one out cross validation
LT-HSCs:	Long-term haematopoietic stem cells
LV:	Latent variables
MAbs:	Monoclonal antibodies
MCT:	Mercury cadmium telluride
MRI:	Magnetic resonance imaging
NaCl:	Sodium Chloride
NHL:	Non-Hodgkin's lymphoma

NIR:	Near infrared
NK:	Natural killer cell
OTRS:	Optical tweezers Raman spectroscopy
PC:	Principal component
PCA:	Principal component analysis
PCR:	Polymerase chain reaction
PET:	Positron emission tomography
PLS:	Partial least square
RBC:	Red blood cell
RMSE:	Root of the mean square error
RMSEC:	Root mean error of calibration
RMSECV:	Root mean error of leave-one-out cross validation
RNA:	Ribonucleic acid
RPMI:	Roswell Park Memorial Institute
rpm:	Revolutions per minute
RS:	Reed Sternberg cells
spp:	Scans per pixel
SERS:	Surface enhanced Raman spectroscopy
ST-HSCs:	Short-term haematopoietic stem cells
T:	Transmittance
UK:	United Kingdom
USA:	United States of America
UV:	Ultraviolet
WBC:	White blood cell
WHO:	World Health Organisation

List of Figures

List of Tables

Chapter 1: Introduction

This thesis presents the application of Fourier Transform Infrared (FT-IR) spectroscopy as a potential diagnostic tool in a clinical setting for the detection of leukaemia and lymphomas. Chapter 1 is structured to introduce haematological malignancies and provide a general overview of the clinical rationale, highlighting incidence and histological classifications of leukaemia and lymphomas. The stages involved in the development of leukaemia and lymphomas are also discussed. This is followed by a review of "gold standards" used by haematologists as well as the associated deficiencies. It also introduces the need for improved techniques in early detection of haematological malignancies. A number of competing optical diagnostic techniques are discussed prior to introducing FT-IR spectroscopy. Chapter 1 concludes with detailed aims and objectives of this research project and an outline of the work presented.

1.1 Background

The scope of this thesis applies vibrational spectroscopy, which has mainly served for analytical chemistry and physics, to an area of clinical relevance. Leukaemia and lymphoma are haematological malignancies that result from clonal expansion. At the time of diagnosis, the normal marrow cell population has usually been replaced by the leukaemia clone accompanied by haematogenous dissemination to a variety of extramedullary sites. Current immunological, cytopathological and molecular biology procedures identify variations in cellular morphology or granularity or use antibodies to detect abnormal leucocytes and cell-specific marker proteins. Although these methods are well established, they are invasive, requiring extensive sample preparation as well as highly trained personnel for interpretation. In addition, none of these diagnostic tools are able to yield biochemical information.

The interest in the potential use of FT-IR spectroscopy, which is based on mid-infrared absorption, arises from the fact that this spectroscopic technique can provide details of the biochemical composition and molecular structures of cells and tissues. In principle, disease and other pathological alterations lead to biochemical and structural changes at a

molecular level thus modifying the infrared spectrum, which becomes a phenotypic marker of the disease. As these spectral changes are very specific and unique, they are also refered to as fingerprints. The added advantages of the method are that it is non-destructive and does not require the use of extrinsic staining agents. Automated, repetitive analyses can therefore be carried out at a low cost with the added advantage of utilizing a small volume of the sample.

This research project aims to provide a practical, effective spectroscopic method that will compete on all levels with the present gold standards in diagnosis, analysis, and imaging of haematological malignancies. This will identify key biomarkers and potentially provide information to clinicians that are otherwise not available.

1.1.1 Incidence of leukemia and lymphoma

Cancer is a major cause of morbidity in the UK with more than a quarter of a million people newly diagnosed with cancer every year (Cancerstats 2009). In the UK, acute and chronic leukaemia and lymphomas combined represent a total of 6.5% of all cancers in both men and women (Cancerstats 2009). Furthermore, one in twenty five people have a risk of developing blood cancer in their lifetime. In 2005, 7,279 (2%) both acute and chronic leukaemia incidence cases, 1,564 (0.5%) Hodgkin's lymphoma (HL) and 10,310 (4%) Non-Hodgkin's lymphoma (NHL) cases were registered (Cancerstats 2009). Chronic leukaemia is the most common leukaemia in adults, accounting for approximately 40% of all leukaemias. The overall incidence is about 3/100,000 per year (Cancerstats 2009). Overall, one-year and five-year survival from acute and chronic leukaemia and lymphomas remains low, despite improvements in diagnosis and treatment (Figure 1-1).

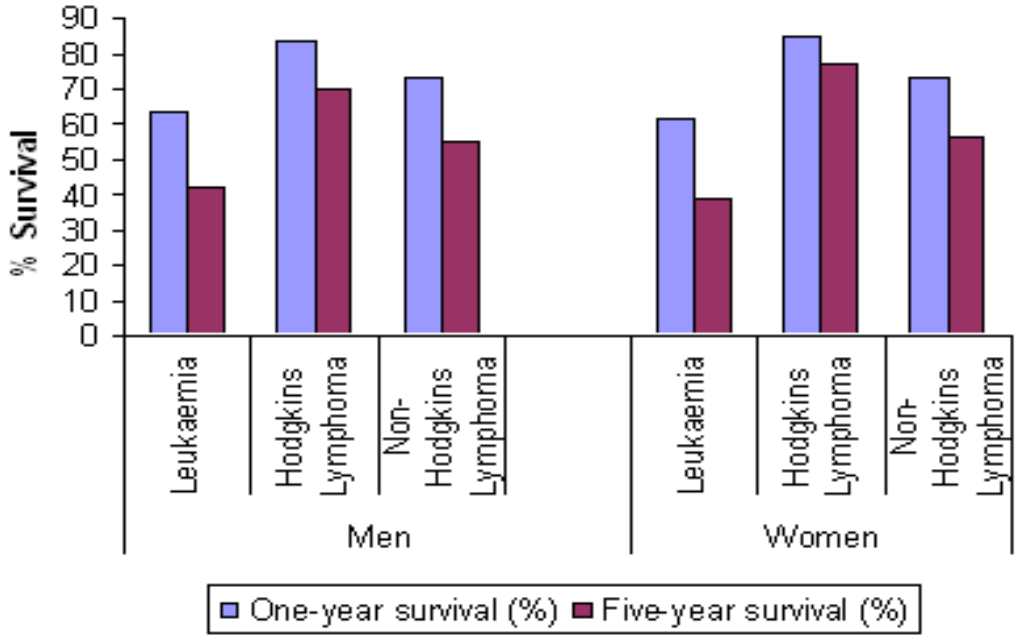


Figure 1-1: Survival rates for Leukaemia, Hodgkin's Lymphoma and Non-Hodgkins Lymphoma in men and women at one and five years during 1971-2001 (Cancerstats 2009)

The spectrum of haematological malignancies varies significantly between children and adults (Figure 1-2). In children (0-14 years), the majority of leukaemia are of the acute type (acute lymphocytic leukaemia (ALL)), whereas in adults (> 19 years), chronic leukaemia (chronic lymphoblastic leukaemia (CLL) and chronic myeloid leukaemia (CML)) are more common (Pui 1999; Downing 2002). Incidence of NHL is prominent in adults. Overall acute and chronic leukaemia and lymphoma are slightly more common in males than females, with a male to female incidence case ratio of 1.3:1 (Levi 2000). The risk of developing leukaemia and lymphoma increases with age characterised by a bimodal age distribution. The highest incidence in children is in the 0-4 age group. Rates then decline until the early 20's, and increase slowly from the early 30's to the early 50's. Incidence rates rise sharply and peak in the over 85's (Xie 2003).

This thesis concentrates on CLL, which is of particular concern in the Western world since it is the most common type of leukaemia. 35 out of every 100 people have a risk of developing CLL in their lifetime.

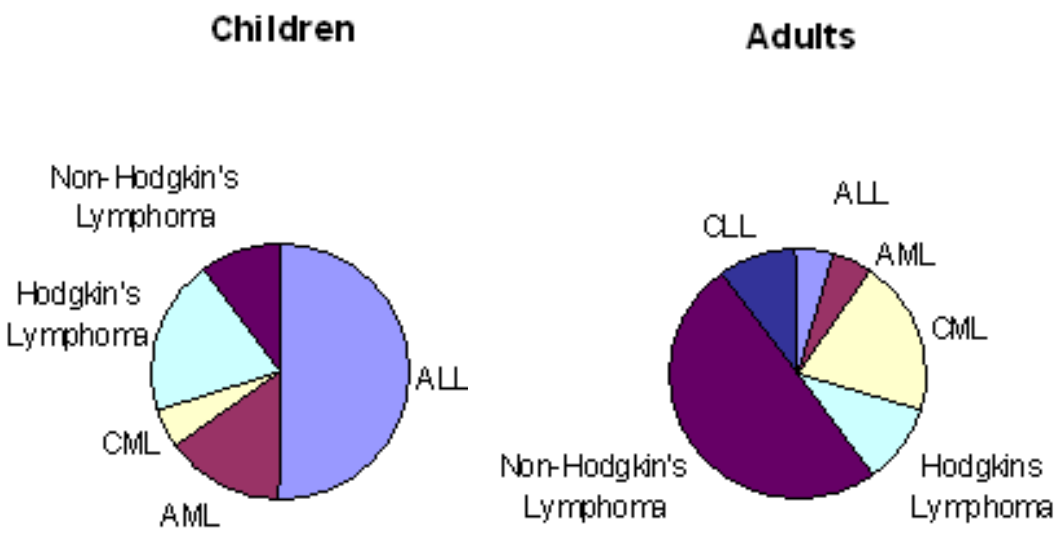


Figure 1-2: Frequency of the major subtypes of haematological malignancies in children and adults. Adapted from Downing 2002

Risk factors for CLL, apart from old age are not clearly defined. CLL is uncommon in childhood, rare in adolescents and becomes progressively more common from the age of 35 years (Cancerstats). Some differences in CLL incidence exist between geographical regions and ethnic groups, suggesting genetic or environmental influence (Alexander 1995). The incidence amongst Caucasians and non-Caucasians is approximately equal but the disease is rare in Asians (Groves 1995 and Xie 2003). Familial links have been observed, but the nature of possible hereditary factors is yet to be elucidated (Redaelli 2004).

1.1.2 Haematological Malignancies

Haematological malignancies derive from the haematopoietic (bone marrow) or lymphoid tissues and hence are targets for malignant transformations and clonal proliferation of lymphocytes. The biology and pathology of bone marrow and lymphoid tissues are detailed in Appendix A.

During normal haematopoiesis, blood cells arise from an immature haematopoietic stem cell (HSC) in the bone marrow. HSCs go through stages of development until they matures into a functioning cells. HSCs are programmed to differentiate into lymphoid or

myeloid progenitors. The lymphoid cells migrate to the thymus, lymph node and the spleen and mature into lymphocytes (T cells and B cells) and natural killer (NK) cells (Figure 1-3). The myeloid cell differentiates into several cell types: erythrocytes, basophils, megakaryocytes, neutrophils, monocytes, and esinophils (Kwong 1988). Following maturation these cells are released from the bone marrow into the blood stream. Haematopoiesis is controlled by cytokines, growth factors and interleukins as well as an array of genes that maintain the balance of cell growth, differentiation, and programmed cell death (apoptosis) in the bone marrow (Kwong 1988; Cantor 2001; Fisher 2002).

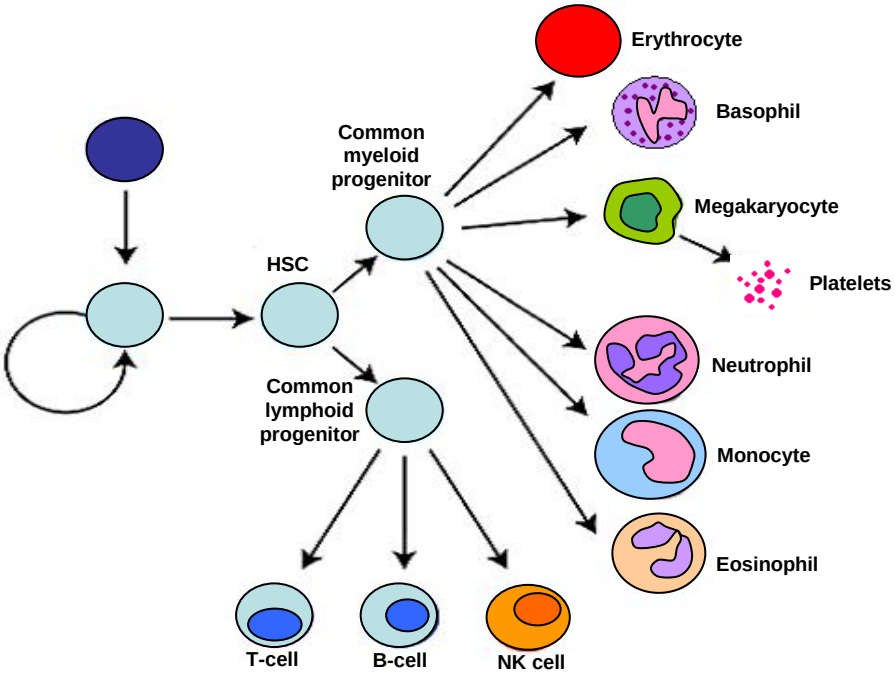


Figure 1-3: Illustration of Haematopoiesis. Haematopoiesis occurs in the bone marrow which gives rise to HSCs. These develop into an assortment of committed stem cells, which then differentiate into mature cells. Erythrocytes distribute oxygen and carbon dioxide between lungs and tissues. Megakaryocytes (platelets) co-ordinate coagulation that prevents blood loss in damaged vessels. Basophils, neutrophils, monocytes, eosinophils and lymphocytes are collectively known as leucocytes. Neutrophils and monocytes are known as phagocytes as they defend the body by engulfing and ingesting bacteria and viruses. Lymphocytes (B-cells and T-cells) react against invading bacteria or viruses by producing antibodies.

Haematological malignancies can arise at any stage during the process of differentiation and maturation within the different cellular lineages, leading to many different types of leukaemia and lymphoma (Bonnet 1997, Enver 1998 and Tenen 2003). Lymphoma and lymphocytic leukaemia develop from lymphocytes, whereas myeloid leukaemia stem from myeloid cells.

In leukaemia, early haematopoietic progenitors fail to respond to external stimuli within the bone marrow microenvironment and differentiate into discrete types of blood cells (Figure 1-4) (Fialkow 1981). These leukaemia cells tend to lose their capacity to terminally differentiate, resulting in the progressive accumulation of immature, non-functional cells with a relatively prolonged life span in the bone marrow and then spill into the peripheral blood. As leukaemia cells do not undergo apoptosis and fail to differentiate (Domen 2001) it makes them disorders of both haematopoietic cell differentiation and apoptosis (Tsiftsoglou 2003) (see Figure 1-4).

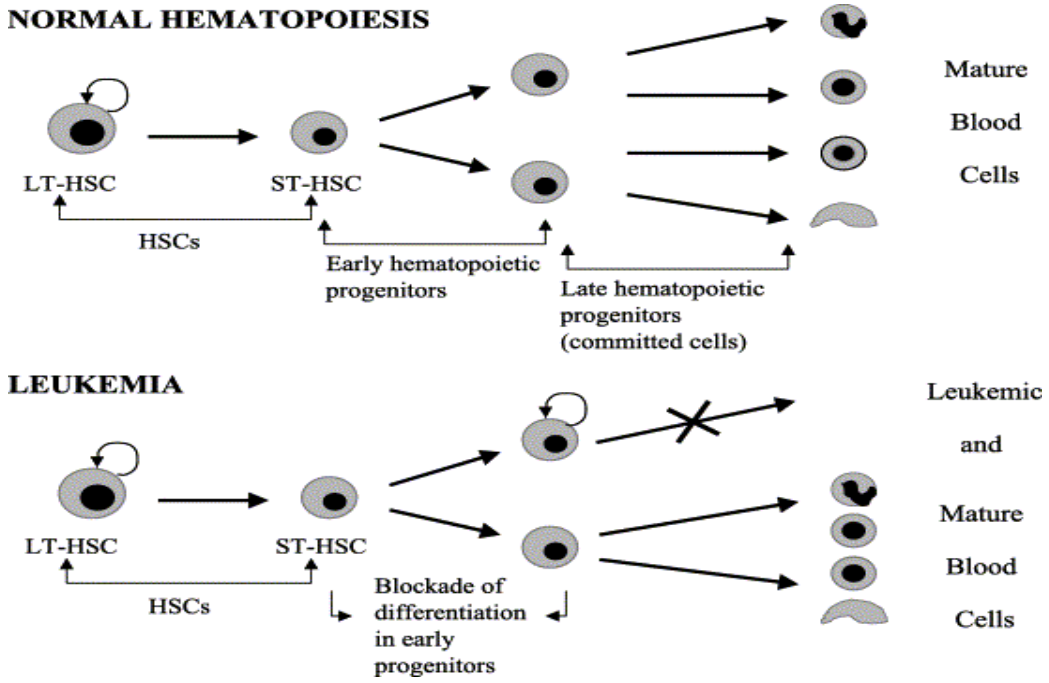


Figure 1-4: Diagrammatic presentation of normal haematopoiesis and leukemiogenesis. Normally, long-term HSCs (LT-HSCs) that self-renew also produce short-term HSCs (ST-HSCs) that subsequently differentiate into late progenitors (committed haematopoietic progenitors) and mature blood cells. In leukemic state, ST-HSCs regenerate early progenitors that are transformed and unable to respond to differentiation stimuli. Source: Tsiftsoglou 2003

A failure in the apoptotic pathway creates a permissive environment for genetic instabilities and accumulations of genetic mutations involving oncogenes (myc, ras, and bcl-2) and tumour suppressor genes (p53) (Debatin 2003). The activation of a proto-oncogene creates a pathological activity whereby the cell receives increased growth or proliferation signals (Figure 1-5). Inactivation or loss of a tumour suppressor gene results in loss of function and malignant cells do not undergo apoptosis (Debatin 2003). As a result, malignant cells tend to lose their capacity to terminally differentiate. This further cause deregulated expansion of malignant cells leading to tumour initiation and progression.

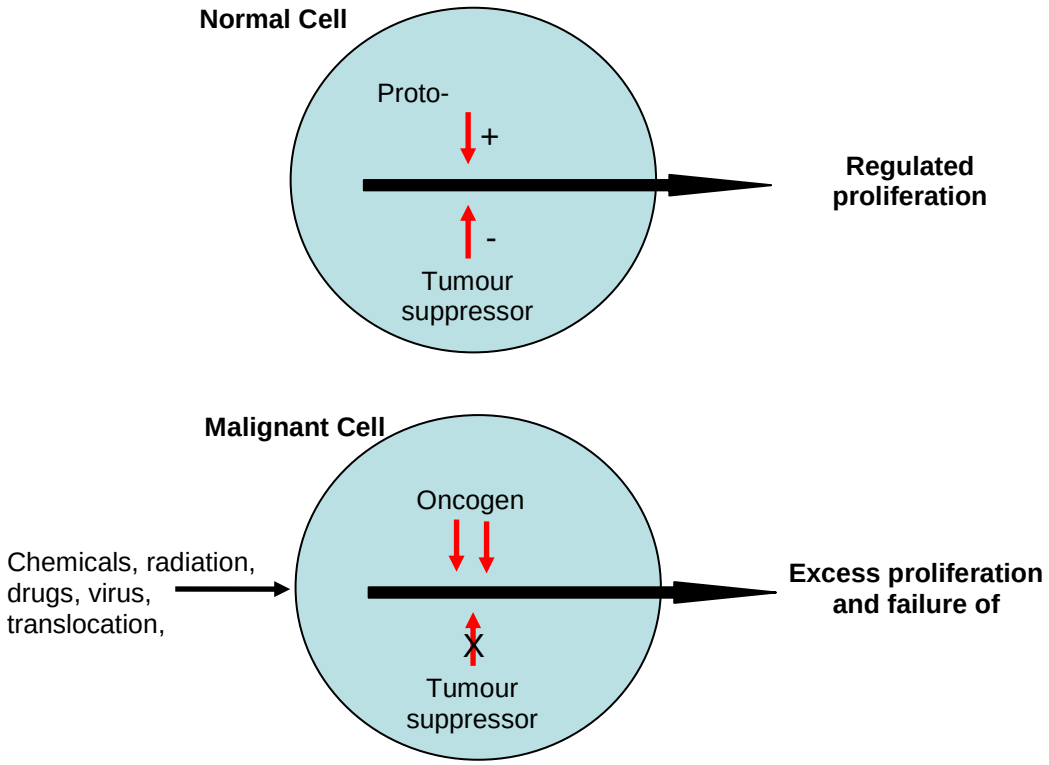


Figure 1-5: Cell balance. Proliferation of normal cell depends on a balance between the action of proto-oncogenes and tumour suppressor genes. In a malignant cell this balance is disturbed leading to uncontrolled cell proliferation. Adapted from Hoffbrand 2006

1.1.3 Histological Classification of leukaemia and lymphoma

Leukaemia ("white blood" in Greek): "When the colour of blood is fading" was first described by a British pathologist, John Hughes Bennett in 1845 (Tefferi 2008). Leukaemias are a group of heterogeneous disorders. They are characterised by the proliferation and accumulation of malignant (blast) white cells in the bone marrow and circulating blood. The main classification is into four types (Hoffbrand 2006): acute and chronic leukaemia, which are further divided into lymphoid or myeloid lineages. Acute leukaemia (ALL and AML) is aggressive in nature. Malignant transformation is caused by rapid proliferation and accumulation of immature, non-functioning HSCs (blast cells). Chronic leukaemia (CLL and CML) involves prolonged progression of mature, well differentiated leukocytes (Figure 1-6). The lymphoid lineage is further subdivided into either into B or T cell origin whist the myeloid lineage involves cells derived from the myeloid stem cells (that is cells from neutrophils, eosinophils, monocytes, basophils and megakaryocytes).

CLL are mostly derived from B-cells. T-cell CLL is rare. The scoring system is based on distinct morphology and presence of certain antigens (immunophenotyping) on the surface of malignant lymphocytes.

Lymphomas are divided into two types: "Hodgkin's lymphoma (HL)" (named after Dr.Thomas Hodgkin, who first recognised it in 1832) and non-Hodgkin's lymphoma (NHL). Classification of lymphoma is based on immunohistochemical staining as well as conventional morphology. Genetic parameters are also included since they have diagnostic value important for clinical decisions. The World Health Organisation (WHO) Classification released in 2001 is used by clinicians and pathologists to classify lymphomas (Jaffe 2001). NHL includes more than 20 subtypes based on B cell or T cell origin according to the WHO Classification. HL is classified into five subtypes according to the Rye classification system, based upon presence of Reed Sternberg (RS) cell morphology (Isaacson 2000).

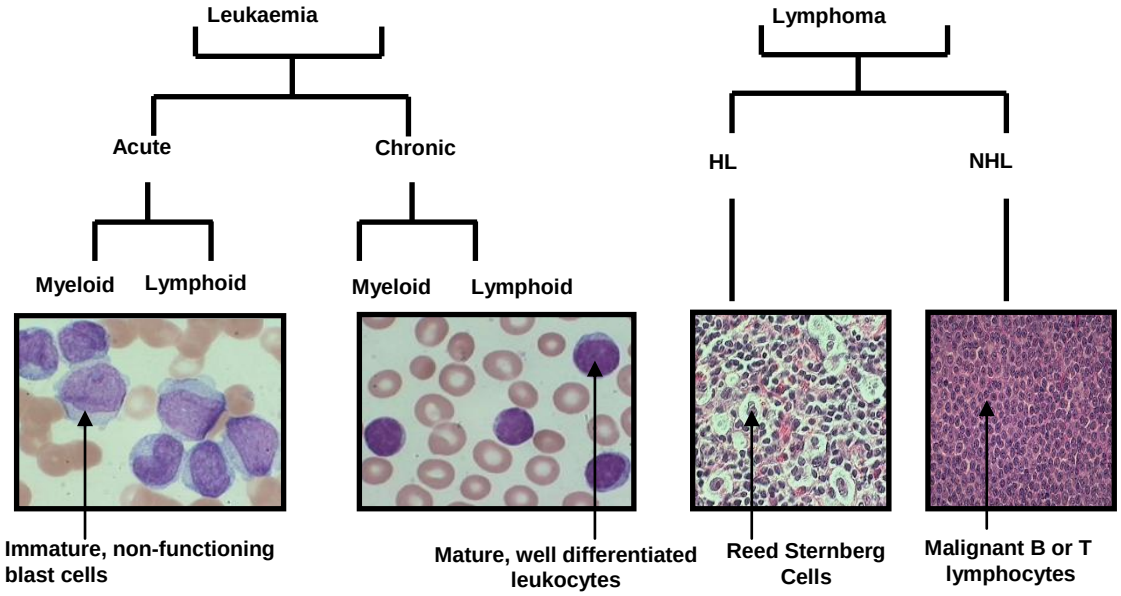


Figure 1-6: Classification of leukaemia and lymphoma and characteristics of malignant cells

1.1.4 Staging of CLL and lymphoma

Clinical staging of CLL and lymphoma is used to determine the extent of spread of the disease from its starting point. In CLL, period of activity can prolong and go undetected from several months up to 10-20 years. In many patients, CLL is diagnosed by chance when a full blood count is done for another reason (Hoffbrand 2007). Prognosis of CLL is estimated through the use of clinical staging systems developed by Rai 1975 and Binet 1981. The Rai staging system (Stage 0-IV) is used primarily in the USA while the Binet staging system (Stage A-C) is commonly used in the Europe. These systems define early (Rai 0, Binet A), intermediate (Rai I/II, Binet B) and advanced (Rai III/IV, Binet C) stages of the disease (Appendix B).

Clinical staging/grading of lymphomas (HL and NHL) depends on the number of lymph nodes affected, location and distribution of the affected lymph nodes throughout the body. Lymphomas of different grades/stages proliferate at different rates and respond to treatment differently. According to the Ann Arbour Staging system (Appendix C), lymphomas are divided into three main grades (clinical stages): Low-grade lymphomas are also referred to as indolent lymphomas; Intermediate- grade lymphomas as faster

growing lymphomas and High-grade lymphomas are also known as highly aggressive lymphomas which tend to grow and spread quickly.

1.1.5 Current diagnostic techniques and associated deficiencies

In haematological malignancies, normal bone marrow haematopoiesis is often disrupted due to infiltration of clonal tumour cell growth. This is the main reason for the common laboratory findings of anaemia, granulocytopaenia and thrombocytopaenia. The blood leukocyte count is often elevated due to clonal expansion of malignant cells. These changes in blood cell composition are reflected in the clinical symptoms of fatigue, tendency to bleed and bruise easily as well as an increased susceptibility to infections. Other common symptoms are malaise, lymph node enlargement, fever, weight loss and night sweats.

The malignancy further evolves by the accumulation of mutations within a clone resulting in genetic diversity. The nature of the clone determines the clinical outcome, which can often be time consuming. Confirmation results for the exact diagnosis can take about 5 days causing possible delays in prognosis and treatment. Delay in diagnosis increases the likelihood that the clone will have progressed to the point that additional mutations will have been acquired. Survival rates heavily depend on the disease stage and early diagnosis is a primary requirement for successful treatment. The role of clinicians in the management and the ideal strategy for early diagnosis of CLL and lymphoma is prompt attention to clinical manifestations, close follow-up for individuals with a known genetic abnormality and early identification of abnormal leucocytes. Currently in the UK, routine diagnostic procedures for CLL and lymphomas rely on presentation of clinical manifestations, morphology and cytochemistry. These are frequently combined with molecular methods, flow cytometric immunophenotyping and cytogenetic studies (Basso 2007). The results from these investigations serve as a mandatory prerequisite for individual treatment strategies and for the evaluation of treatment response.

1.1.5.1 Gold Standard for CLL diagnosis

The "gold standard" diagnostic criterion for CLL is morphology. This involves the evaluation of stained (Haematoxylin and Eosin (H&E) stain) blood specimens or bone marrow aspiration films under a microscope. The morphological examination determines the distinctive architectural features of CLL cells. Morphological investigations form the initial diagnosis for CLL but are not sufficient to identify the various clinical subsets within the main category of CLL. The main difficulty associated with the morphology involves targeting malignant cells due to variations in the fraction of circulating CLL cells and the distribution of cell deoxyribose nucleic acid (DNA) content. Interpreting results in heterogeneous samples or those with few malignant cells can therefore prove to be difficult. The problem is further complicated by sampling large numbers of viable cells as well as requiring numerous staining reagents per patient sample.

In most cases morphological assessment is combined with secondary immunophenotyping (flow cytometry) investigations to establish a precise classification (Richards 2003). Immunophotyping defines immunologic phenotype of malignant cells by using monoclonal antibodies (MAbs) reactive against lymphoid or myeloid associated antigens. Although being able to separate and evaluate single cells from mixed populations, this technique requires expensive equipment and reagents. In addition, there is no visualisation of cell architecture and interpretation of data in heterogeneous sample or those with few blasts is difficult.

In a minority of difficult cases (5–10%) further molecular investigations may be required for an accurate diagnosis (Szczepanski 2003). This involves the use of polymerase chain reaction (PCR) and fluorescence in situ hybridisation (FISH) assays for detection of genetic rearrangement and chromosomal translocations (e.g. trisomy 12). These methods target specific part of the DNA sequences but due to DNA contamination and degradation tend to be prone to false positive results. Despite all investigations, up to 5% of CLL cases defy lineage assessment (Szczepanski 2003).

1.1.5.2 Gold Standard for Lymphoma diagnosis

The gold standard for detection of HL and NHL involves histological examination of excised lymph nodes. This technique involves examining an H&E stained lymph node tissue section (one cell thick (<10 microns)) under a conventional light microscope. Changes in tissue morphology and presence of Reed Sternberg (RS) cells can therefore be identified in order to determine histological classification and stages of disease. Often, the patient presents several enlarged lymph nodes during diagnosis. In such cases the largest lymph node is chosen for biopsy as smaller lymph nodes, although easily accessible, may show nonspecific histological changes that may obscure the correct diagnosis. This results in delayed diagnosis or even erroneous results. Furthermore, surgical biopsy is affected by random sampling (only a small part of the tissue is removed and examined) therefore the extent of the disease can not always be quantified. Although fine needle aspirate (FNA) biopsy techniques have reduced the number of patients undergoing surgical biopsies for early detection of lymphomas, there is a delay in obtaining immediate results. FNA biopsy yields a high rate of conclusive cytological diagnoses in the assessment of high grade NHL (79–90%), but has lower performance in the assessment of low grade NHL with an average accuracy rate of 75-80% (Hehn 2004; Lioe 1999; Kocjan 1997; Young 1999 and Dong 2001).

The analysis of sectioned lymph node tissue is prone to high levels of inconsistencies between histopathologists. In one study, amongst expert haematopathologists, interobserve agreement ranged between 21-65% (Benard 1985). In complex cases, immunological marker studies (flow cytometric immunophenotyping) may be used to classify lymphoma with either T or B cell lineage. Although highly specific, the process requires a comprehensive and costly panel of antibodies and there are sometimes problems associated with suboptimal antibody selection and inadequate or insufficient fresh material (Katz 1997; Mourad 2003 and Olsen 2007). Despite the best efforts of morphology and immunophenotyping, some lymphomas are impossible to diagnose (Hans 2003; Landgen 2004). Genetic rearrangement studies are helpful in uncommon cases where immunophenotypic studies are unable to assign lineage (T versus B). Standard imaging tests such as x-rays, computed tomography (CT) scans, magnetic

resonance imaging (MRI) and positron emission tomography (PET) scanning may also be performed for advanced investigations to determine the degree of malignant lymphocyte invasion in other areas of the body. These imaging techniques can be expensive and involve patients being exposed to ionizing radiation.

1.1.6 Summary

Overall, the diagnostic strategy for haematological malignancies is not based on a single diagnostic procedure, but relies upon a diverse range of measurements and observations to determine the presence and nature of malignancy. The extensive sample preparation procedures required by these diagnostic methods rely upon the experience and skills of both the haematologists and histopathologists to achieve an accurate diagnosis. The complexity is further enhanced by the existence of approximately various subtypes of CLL and lymphoma, each with its own genetic signature requiring a different prognosis and treatment. As a result, testing for CLL and lymphoma is complicated, expensive and occasionally inaccurate.

Laboratory investigations rely on as many as 5-6 different technologies to provide diagnostic and prognostic information. Even so, a number of patients are misdiagnosed in regards regardless of the use of various diagnostic tests. Furthermore, a significant number of malignancies are either not detected or have advanced by the time they are diagnosed. Part of the problem stems from the steep cost of various technologies used to diagnose haematological malignancies. Furthermore, these methods comprise of large lab-to-lab variability with the resulting risk that the patient might not receive the right treatment or delayed in a situation where time is critical. The problem is compounded by significant inter-observer variability between histopathologists and haematologists.

1.2 Optical Diagnostic Techniques

1.2.1 Clinical Need

There is a clinical need for new technologies that can overcome inadequacies of current diagnostic methods and provide automated, accurate methods for detection of haematological malignancies at an early stage. The demand for a real-time technique is on the increase to fulfil clinical needs. The main reasons include: i) to determine prognosis and to aid in the selection of the best treatment for an individual; ii) to determine accurate pathological classification of leukaemia and lymphoma and iii) to advance knowledge on the biochemical changes underlying disease progression. A potential complementary detection test should be cost effective, easy to use with minimum sample processing. In, addition it needs to provide real-time detection and accurate classification of pathological changes enabling immediate diagnosis. In any diagnosis, clinicians look for deviations from the norm, especially changes in morphology and biochemical composition of the organs, tissues and fluids they affect; these differences are the basis of everyday clinical tests, tissue staining and medical imaging techniques.

Current diagnostic techniques provide limited information on the biochemical changes that occur during disease progression at a cellular level. As all malignancies are caused by changes in the cellular biochemistry, detection of biochemical changes can provide quantitative information which can prove vital in the early diagnosis of CLL and lymphoma. Qualitative analyses of biochemical changes can also provide important information in measuring the extent of the disease process as well as accurate classification. An easier and accurate diagnosis of haematological malignancies at an early stage is important in the identification of high risk patients and monitoring the efficiency of treatment.

The detection of biochemical constituents via optical diagnostic techniques has been proposed to comply with these clinical needs for detection and classification of haematological malignancies. Optical diagnostic techniques have become increasingly important in recent years and shown great potential to date. They have been proposed to assist in providing reliable, real-time and non-invasive diagnostic tool to suit the demand of the medical community. These methods rely upon the measurement of the interaction of light with the biochemical constitutes to study abnormalities is cells and tissue. The data obtained can provide evaluation of changes in morphology and altered concentration of sub-cellular attributes. This information can aid in the early detection and classification of leukaemia and lymphoma, providing possible benefits to both clinicians and patients.

The main techniques being explored for detection of biochemical changes are Fluorescence spectroscopy (FS), Elastic scattering spectroscopy (ESS), Vibrational spectroscopy including Raman spectroscopy and Fourier transform infrared (FT-IR) spectroscopy.

1.2.2 Properties of electromagnetic radiation

The spectrum of electromagnetic radiation is divided into different regions as shown on Figure 1-7. Electromagnetic radiation consists of an oscillating electrical fields and an oscillating magnetic fields perpendicular to each other. However in IR it is the electric field that is of importance. The infrared spectrum ranges from 1 μm to 1000 μm, and is surrounded by the visible and microwave regions (Figure 1-7). The IR region is further subdivided in the near infrared, the mid infrared and the far infrared regions.

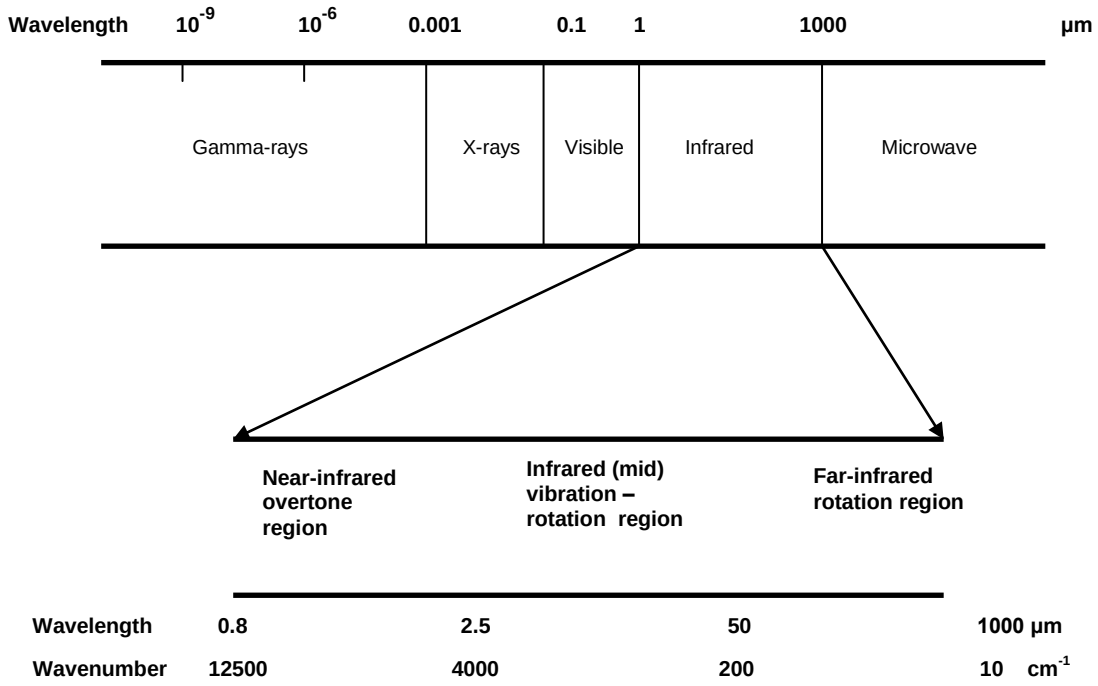


Figure 1-7: Infrared region of the electromagnetic spectrum

1.2.3 Interactions of light with matter

Light can interact with matter in various manners including, scattering (elastic scattering and inelastic scattering), fluorescence and absorption.

The major form of elastic light scattering is Rayleigh scattering. Inelastic scattering effects include Raman scattering. Light that is scattered due to vibrations in molecules or optical phonons is called Raman scattering. The Raman scattering is an inelastic scattering and depends upon the polarizability of the molecules.

1.2.3.1 Elastic Scattering Spectroscopy (ESS)

ESS uses the scattering of light that generates a wavelength dependant spectrum. It reflects structural and morphological change within subcellular components to differentiate between normal and abnormal tissue (Swinson 2006).

An elastic scattering event redirects incident photon without a change in wavelength or energy (Song 2005). The principle relies on the fact that tissues with a higher density

have a higher refractive index and therefore have a greater ability to scatter light. Higher densities that occur at cellular and sub-cellular levels result in a higher refractive index and are therefore responsible for the majority of the elastic scattering. The structures that induce elastic scattering are the nucleus, chromatin concentration, and sub-cellular organelles. Thus, ESS has been shown to be sensitive to nuclear size, chromatin content, nuclear/cytoplasmic ratio and cellular crowding which are all criteria that the histopathologist looks for when establishing of malignancy within a tissue (Wallace 2000).

ESS has been involved in various clinical fields. Main interest is diagnosis of melanocytic lesions (Scarisbrick 2004) premalignant/malignant lesions in oral tissues (Sharwani 2006), breast cancer (Bigio 2000), Barrett's oesophagus (Lovat 2006) and sentinel lymph node metastases (Chicken 2004). Classification of normal hyperplastic and dysplastic polyps and adenocarcinoma using principal component fed linear discriminant analysis (PCA-fed-LDA) and cross validation has also been reported recently with sensitivity and specificity ranging between 75-88% (Dhar 2006).

1.2.3.2 Inelastic scattering (Raman Effect)

The scattering of light by various media had long been studied by Rayleigh in 1871, Einstein in 1910. The theory of Raman Effect was discovered by Sir C. V. Raman (1928), together with K. S. Krishnan and independently by Grigory Landsberg and Leonid Mandelstam (Krishnan 1981). Raman won the Nobel Prize in Physics in 1930 for this discovery.

Raman scattering (Raman Effect) is the inelastic scattering of monochromatic light, in which the energy of the scattered photons either donate or receive energy to cause vibrations of molecular bonds (Song 2005). Molecules are excited to a virtual state on interaction with photons. Majority of these molecules de-excite back to the initial vibrational energy level, which results in release of photons of the same energy and wavelength (Rayleigh scattering). These are filtered out and the remaining light is dispersed onto a detector. A small proportion (approximately 1 in 10^7) of the scattered

radiation is shifted to a different wavelength. This change in the wavelength is known as the Raman shift which is specific to the species of the molecule responsible for the scattering. When all the shifted wavelengths from photons scattered by the different molecules present in a sample are combined they form a Raman spectrum (wavelength shift vs. photon count). The spectrum provides a biochemical fingerprint, revealing characteristic peaks caused by the molecular vibrations.

Raman shifted photons of light can be either of higher or lower energy, depending upon the vibrational state of the molecule. Stokes Raman scattering (molecule absorbs energy) occurs when the final energy level is higher than the initial level, while anti-stokes Raman scattering (molecule loses energy) occurs when the final energy level is lower than the starting level (Figure 1-8). Energy differences between the incident and scattered light is proportional to the vibrational energy of the scattered molecules. This change in vibration produces a change in polorizability, which is the change of the dipole moment with distance. Stokes scattering is more common than anti-stokes scattering because at any given time an electron in the most common temperature range is most likely to be in its lowest energy state, in accordance with the Boltzmann distribution. Thus the Stokes spectrum is more intense than the anti-Stokes spectrum and therefore is commonly used in spectroscopy.

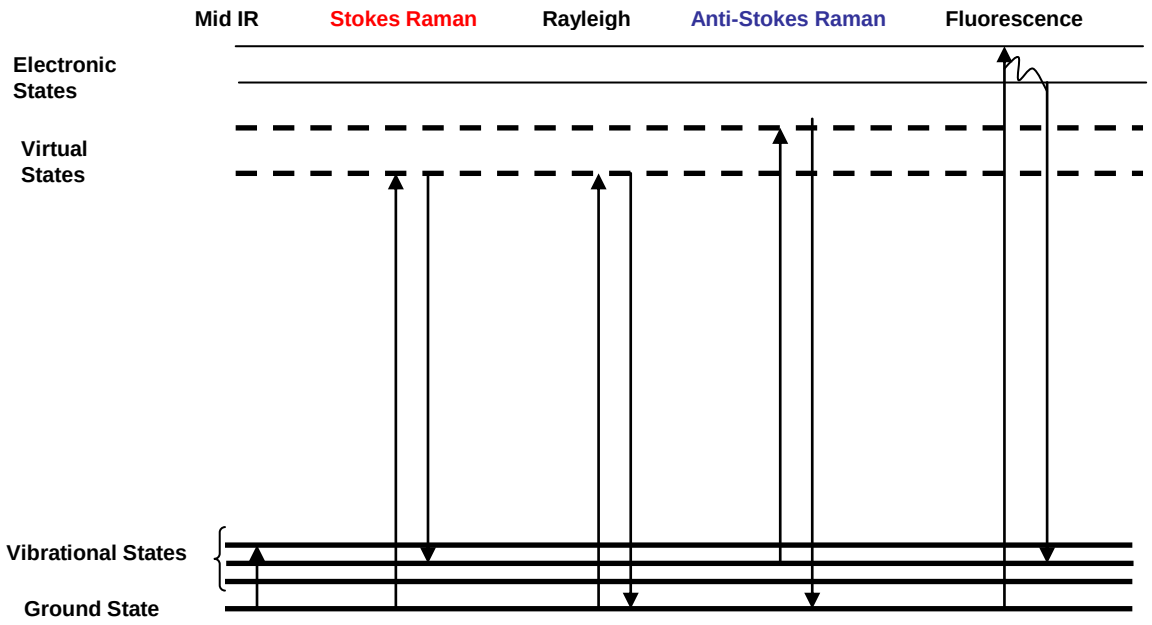


Figure 1-8: Energy-level diagrams of Rayleigh scattering, Strokes Raman scattering, and anti-stokes Raman scattering

Raman scattering is a relatively weak process (Song 2005). The number of photons scattered is quite small, especially when probing cell samples or body fluids. However, there are several processes which can be used to enhance the sensitivity and improve the spatial resolution of a Raman measurement as well as acquire specific information. These include:

a) Surface Enhanced Raman Spectroscopy (SERS), which uses silver or gold colloid substrate to enhance the intensity of Raman signal of biomolecules measured (by up to 10^{11}) within body fluid such as lactic acid (Hui 2005) and glucose (Haynes 2005). The feasibilty of SERS has also been investigated in live cells (Kneipp 2002) and detection of cancer gene (Vo-Dinh 2002).

b) Optical Tweezers Raman Spectroscopy (OTRS), which examines individual particles and biochemical processes in single cells trapped by optical tweezers. This technique has been applied for single-cell analysis of eukaryotic cells (Snook 2009), leukaemia cells (Chan 2008) and live prostate and bladder cancer cell lines (Harvey 2008).

As mentioned previously, the peaks within the Raman spectrum are characteristic of different molecular vibrations. Individual peaks are directly proportional to the concentration of varying biochemical constituents present in tissue and cells. Therefore the spectrum can be utilised to obtain quantitative information making Raman spectroscopy potentially a powerful tool for biomedical application. Although the Raman Effect was first observed in 1928, it was not until the early 1990s that Raman spectroscopy was applied in the successful analysis of cancerous and precancerous tissues (Mahadevan-Jensen 1996). Applications of Raman spectroscopy have been used to study various cancers and pre-cancers including cervix (Jess 2007; Lyng 2007 and Kanter 2009), oesophagus (Kendall 2003; Shetty 2006 and Song 2006), prostate (Crow 2003 and Crow 2005), bladder (Crow 2005 and Grimbergen 2009), breast (Haka 2006 and Rehman 2007), larynx (Stone 2000 and Lau 2005) and oral tissue (Malini 2006).

Raman spectroscopy has also advanced in several single cells studies including living cells (Notingher 2005), embryonic stem cells (Notingher 2004) and single red blood cells (Wood 2002). Other studies evaluate spectral features of cells and sub-cellular components of freeze dried and living cells in media (Krafft 2003). In a study by Krishna 2005, the efficiency of Raman spectroscopy was evaluated to identify a cell type in randomly distributed mixed cell populations of human promyelocytic leukaemia (HL60), human breast cancer (MCF7) and human uterine sarcoma (Mes-sa) cell lines. Raman spectroscopy has also been applied to discriminate normal and transformed human haematopoietic cells (Chan 2006), identify leukaemia cells (Chan 2008); differentiate between normal and gastrointestinal cancer cells (Yan 2005) and prostatic adenocarcinoma cells (Crow 2005). Identification of sub-cellular organelles such as nucleus, cytoplasm, endoplasmic reticulum, vesicles and cellular membrane have also been investigated in single human lung fibroblast cells by Raman microspectrocopic mapping (Kraft 2005). Furthermore, the Raman microspectrocopic approach has been applied to study the nuclei and cytoplasmic regions of single human granulocytes (Puppels 1991).

The analysis of body fluids is a recent advancement in application of Raman spectroscopy. Whole blood analysis, although proven to be difficult due to contributions

from fluorophores and haemaglobin, quantification of biochemical constiutents has been possible. Berger 1999 have reported quantification of glucose, cholestrol, triglycerides, urea, albulmin and total protiens in whole blood. Qualtification of haemaglobin and haematocrit have also been determined (Enejder 2002 and Wood 2002). To overcome some of the complexity asssociated with whole blood measurements, serum or plasma samples have been used for the quantification of biochemical constiutents such as glucose (Dou 1996), total proteins, cholestrol, high- and low- density lipoproteins, triglycerides, glucose, urea and uric acid (Rohleder 2004). Serum samples from breast cancer patients have also been examined by Raman spectroscopy to highlight changes in the biochemical composition (Pichardo-Molino 2007). The prediction of concentrations of all these biochemical constituents in the various body fluid analyses were determined by partial least square (PLS) analysis.

1.2.3.3 Fluorescence Spectroscopy (FS)

FS is a type of electromagnetic spectroscopy used for analyzing fluorescent spectra. It uses incident beam of UV light to excite electrons in molecules, which emit light of a lower energy and varying colours. This light is called fluorescence and is detected qualitaively by sensitive spectrometers (Swinson 2006). All tissues fluoresce due to the presence of fluorescent chromophores (fluorophores) e.g. NADH, collagen, elastin and flavins (FAD, FMN). FS can detect these substances and provide characteristic spectra that reflect biochemical and architectural changes occurring within the tissue. The resultant spectra not only detect the light that is fluoresced but also are sensitive to the structures that absorb light (Swinson 2006). Alfano 1984 first described the use of autofluoresence spectroscopy in vivo to differentiate between normal and malignant tissues. Since then fluorescence spectroscopy has been applied to investigate cervical cancer (Chang 2005).

1.2.3.4 Infrared (IR) Absorption Spectroscopy

IR spectroscopy is a powerful tool for determining structure of biochemical compounds within biological tissues and fluids. In addition, IR analysis may be used for quantification of biochemical components.

Infrared spectroscopy utilizes the mid-IR region of the light to obtain an IR spectrum. It involves absorption measurements of different IR frequencies by a sample positioned in the path of an IR beam. The absorption of IR light results in changes in the vibrational modes of molecules that are IR active. For a vibrational transition to be IR active there must be a dipole moment of the molecule during vibration. If the vibration affects the polorisability of the molecule then the vibrational mode will be Raman active. Thus, mid-IR absorption spectroscopy is complimentary to Raman since each is sensitive to different functional groups due to selection rules. Most biological molecules are IR active allowing proteins, nucleic acids and lipids to be examined. Changes in a spectrum are the result of subsequent structural changes in the configuration of these IR active biological molecules in single cells or tissues. These are key factors that differentiate abnormal from normal.

IR spectroscopy has experienced vast improvements with the introduction of interferometry and availability of new detection elements (charge-coupled device (CCD)) technology. With the development of rapid digital data acquisition and processing, the speed of FT-IR spectrometers has increased considerably. These implementations contribute to high reproducibility and low sampling noise. By combining the FT-IR spectrometer with a light microscope which includes a stepping motor an efficient IR imaging tool has become available. This provides structural detail of the sample along with information on the biochemical composition of the sample. Furthermore, the introduction of synchrotron IR source permits smaller regions to be measured without affecting the signal to noise ratio compared to conventional IR source. The development of a robust attenuated total reflection (ATR) probe also makes mid-IR spectroscopy a potential tool for monitoring biochemical processes at cellular level. In addition, new sampling techniques have been introduced, enabling measurements of samples in solvents, as well as the measurements of cell suspensions

and dried films. The incorporation of advanced computational and multivariate statistical methods has enabled the immense utilisation of FT-IR spectroscopy for research purposes in diagnosis, follow-up and prediction of various cancers.

Infrared spectroscopy can be used for mapping studies, providing imaging of tissue sections as well as of single cells (Diem 2000). As IR spectroscopy is the main technique utilised within this study further details of principles and theory of infrared spectroscopy as well as the instrumentation of IR spectroscopy is outlined in Chapter 2. Applications of FT-IR spectroscopy are reviewed in Chapter 3.

1.2.4 Summary

Future prospects of these optical spectroscopic techniques range from application in real-time diagnostics to an important complementary tool to improve detection and classification of haematological malignancies. Optical spectroscopic techniques mentioned above each have advantages and disadvantages:

- ESS only a point measurement technique. There is no potential for imaging analysis. It only provides information on morphology.

- Fluorescence spectroscopy is sensitive and can detect low quantities of the compound in a sample but requires expensive and sophisticated equipment.

- Raman can provide detailed biochemical information but Raman signal is weak, which leads to low sensitivity, making it difficult to measure low concentrations of biochemical substances. Vast majority of molecules are not Raman-active, producing weak interactions. However, recent technological advances have overcome such a limiting factor. Sample thickness is another limitation. Measurements from a thin dried film can prove to be difficult due to large contributions from CaF_2, cosmic rays, fluorescence and artefacts. Water is only a weak Raman scatter and adding little interference.

- Mid-IR spectroscopy is established for probing biochemical changes. The signal is greater than Raman scatter with a higher signal to noise ratio of spectra. Imaging methods are also more rapid than Raman mapping techniques. IR spectroscopy is limited by water absorption; however this is overcome by eliminating water and air-drying cell samples into thin layer films. As a result, acquiring IR spectra of thin monolayer cell preparations is feasible with less difficulty than Raman spectra in view of the higher sensitivity of IR spectroscopy.

Compared to other techniques discussed in this section, IR spectroscopy is the technique of choice as it holds several attractions for consideration in the haematology laboratory with advantages such as:

1) No chemical reagents or specific molecular probes are required for sample preparation- the IR spectral patterns of the species of interest can provide the basis for detection and quantitation.

2) The IR light is not damaging to the sample, thus subsequent analyses are readily feasible.

3) Only small amounts of the sample (of the order of microliters of fluids, or small number of cells (10^3)) are required, leaving ample material for other clinical tests.

4) Less time consuming, as the measuring process of the spectrum for one sample is completed within approximately few minutes, with little training required for the operator.

5) It is a computer-operated system where the analysis of a large number of samples is feasible by the continuously automated process.

6) Uses all IR energy simultaneously and therefore achieve a low noise level.

7) An internal laser calibrates the interference information, providing high wavenumber accuracy and reproducibility.

1.3 Aims and Objectives

The rapid developments in the field of infrared spectroscopy in the past decade have demonstrated a potential for disease diagnosis. Several studies have highlighted the advantages of using infrared spectroscopy in the mid-infrared region for diagnostic purposes at a clinical level. Most of the studies to date have focused their research on differentiation of normal and diseased tissues in various organs using biopsy material from solid tissues at specific sites. Very little work has been done on the identification and classification of liquid-phase haematological malignancies, which are on the increase. The ability to diagnose the early onset of haematological malignancies rapidly has multiple benefits. These include the early prognosis and intervention of therapeutic strategies leading to an increase chance of survival. In addition, the identification of some biomarkers of disease could also be used to monitor the progression of therapy.

The aim of this project was to carry out a detailed study of the application of Fourier transform infrared (FT-IR) spectroscopy as a potential diagnostic tool for clinical use for the detection of leukaemia and lymphomas. The study shows the application of FT-IR spectroscopy as a method to characterise biochemical differences that detect and distinguish leukaemia and lymphoma types based the use of diverse liquid-based clinical/biological samples. This is based on the FT-IR spectroscopic measurements and statistical analysis of leukaemia and lymphoma derived cell lines as well as normal and leukaemia blood samples. The study aims to address the following objectives:

1. Cell lines
    - Interpret and assess differences in spectral profile of lymphoid, myeloid and lymphoma derived cell lines.
    - Classify cell lines combining objective measurements and multivariate statistical techniques to assess feasibility and validation of the spectroscopic method

- Assess the effect of storage of cell lines at room temperature (air-dried) and -80ºC (frozen).

2. Normal and Chronic Lymphoid Leukaemia (CLL) blood specimens
    - Interpret and assess differences in spectral profile of normal whole blood and its components (buffy coat and plasma).
    - Interpret and assess differences in spectral profile of leukaemia (CLL) whole blood and its components (buffy coat and plasma).
    - Interpret and compare infrared spectra of normal and leukaemia blood (biochemical differences).
    - Classify normal and leukaemia blood using multivariate statistical techniques to assess feasibility and validation of the spectroscopic method.
    - Compare concentration levels of various cellular components within normal and leukaemia blood (lipids, proteins and nucleic acids etc) using chemometric techniques.
    - Quantitative analysis of cellular analytes within FT-IR spectra of normal and CLL samples through partial least square (PLS) analysis.

The overall aim of this research was to study the application of highly informative technique, FT-IR spectroscopy to detect spectral parameters to distinguish between leukaemia and lymphomas using a variety of specimens, all of which are liquid based.
The differences in the spectral profile between the various specimens may serve as biomarkers for the identification and detection of different pathological subgroups of leukaemia and lymphoma. The identification of leukaemia and lymphomas at various clinical stages/grades could indicate changes in the degree of maturation of malignant cells as well as in their biochemical composition. This may give insight into the biochemical changes and their impact on the blood and tissues, which is of great importance in understanding the pathological processes associated with leukaemia and lymphoma progression. Clinicians may arrive at a diagnosis and classification of specific subtypes of leukaemia and lymphomas at a much earlier stage and help to decide available potential treatments.

1.4 Thesis Structure

Chapter 2 provides an overview of the theory and principles of FT-IR spectroscopy as well as instrumentation. Chapter 3 reviews the literature highlighting applications of FT-IR spectroscopy in biomedicine. This will include detailed discussion of spectroscopic studies of various diseases in human tissues, single cells and bio-fluids. Chapter 4 gives a review of all the subsequent chemometric techniques used in this study. This is followed by an outline of the Materials and Methods used for the processing, measurement and classification of cell line and blood samples in Chapter 5. This chapter also covers system optimisation and data analysis. The Results are presented in Chapter 6 with discussion as the data was analysed. The Results chapter is divided into 3 sections, presenting and discussing results from a) cell line analysis, b) classification model optimisation and c) blood sample analysis. The Conclusions drawn from the work follow in Chapter 7 with suggestions for Future Work in Chapter 8.

Chapter 2: FT-IR Spectroscopy

2.1 FT-IR Instrumentation

In IR spectroscopy, the attenuation of the intensity of a radiation beam of polychromatic infrared light upon passing through a sample is measured. This attenuation is caused by the interaction of the light with the vibrational transitions of the sample molecules. This principle is described in further detail in Section 2.2. There are three basic spectrometer components in an FT-IR system that makes this course of action possible. These include: IR radiation source, Michelson interferometer, and detector (mercury cadmium telluride (MCT detector)). A simplified layout of a typical FT-IR spectrometer is illustrated in Figure 2-1.

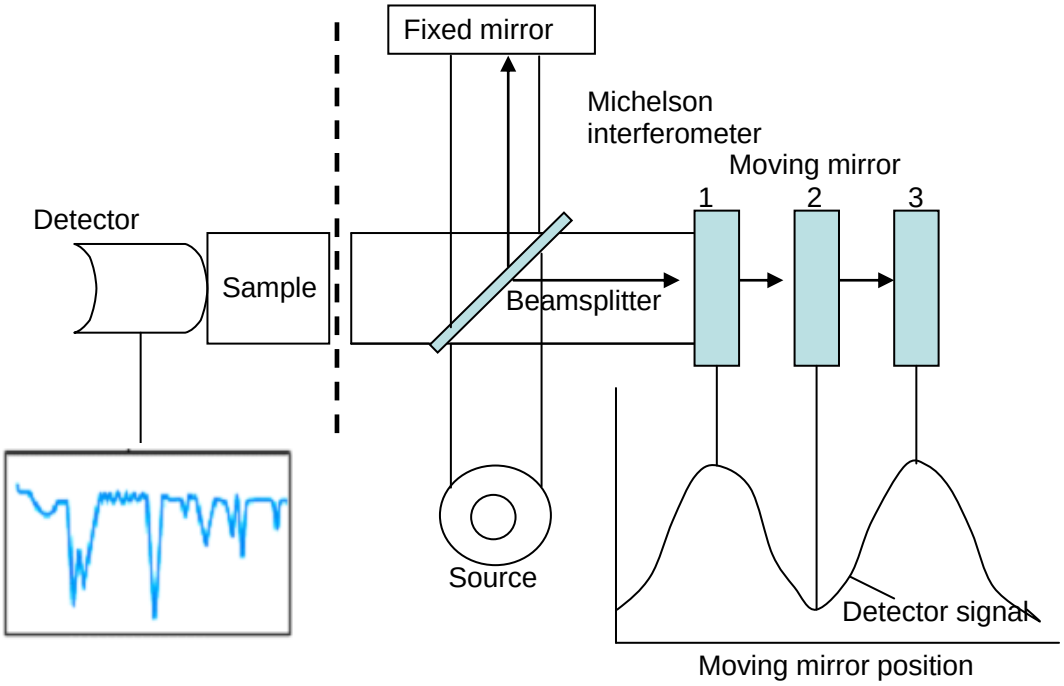


Figure 2-1: Simplified layout of a typical FT-IR spectrometer

2.1.1 Interferometer

The Michelson interferometer consists of three active components: a moving mirror, a fixed mirror oriented perpendicularly and a beam splitter (Figure 2-2). The interferometer splits the IR radiation into two beams by means of a semi transparent mirror (beam splitter) at right angles. The beam splitter is thinly coated with germanium sandwiched between two pieces of Potassium bromide (KBr), which split the IR beam. At the beam splitter, half the IR beam is transmitted to the fixed mirror and the other half is reflected to the moving mirror. These beams recombine at the beamsplitter, but the difference in path lengths creates constructive and destructive interference: an interferogram. The recombined beam passes through the sample, which absorbs all the different wavelengths characteristic of its spectrum, and this subtracts specific wavelengths from the interferogram. A mathematical operation known as a Fourier transformation converts the interferogram (a time domain spectrum displaying intensity verses time within the mirror scan) to the final IR spectrum which is the frequency domain spectrum showing intensity versus frequency.

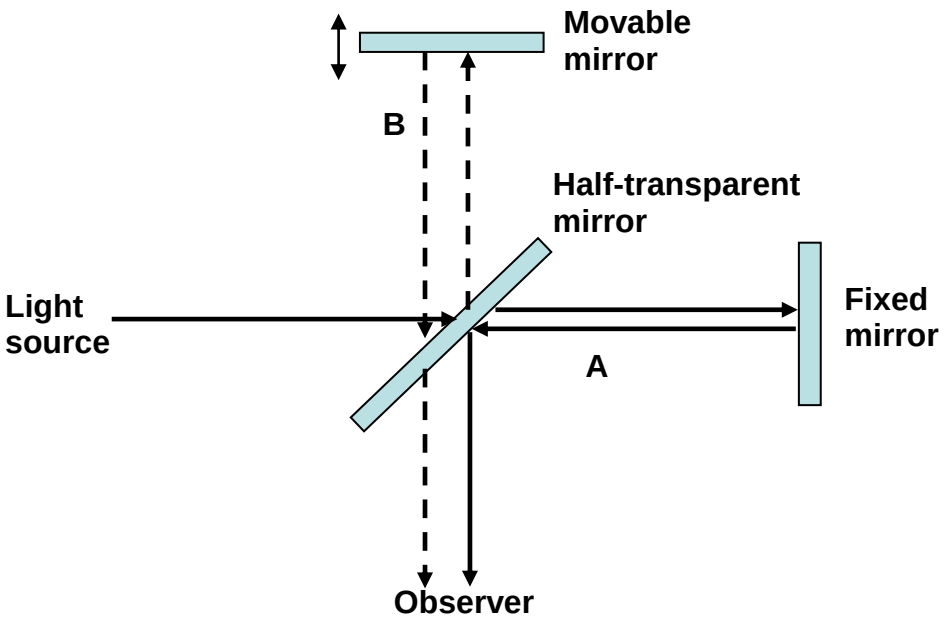


Figure 2-2: Michelson interferometer

2.1.2 HgCdTe (MCT) Detector

MCT (mercury cadmium tellurium) detector is a photon detector operating at liquid N_2 temperature. IR radiation excites the electrons to higher energy levels. This changes the electrical conductivity of the detectors. MCT detector is highly sensitive and used to obtain good quality spectra.

2.2 Theory of Infrared (IR) Absorption Spectroscopy

Photon energies associated with the infrared region of the electromagnetic spectrum are not large enough to excite electrons, but may induce vibrational excitation of covalently bonded molecules. At temperatures above absolute zero, all the atoms in molecules are in continuous vibration with respect to each other. When a beam of electromagnetic radiation is passed through a substance, it can be either absorbed or transmitted, depending upon its frequency and the structure of the molecule it encounters. Thus when a molecule absorbs radiation it gains energy as it undergoes a quantum transition from one energy state to another and excites molecular vibrations. The mode of excitation depends on the wavelength of the light. Electrons are promoted to higher orbitals by ultraviolet or visible light, molecular vibrations are excited by infrared light, and rotations are excited by microwaves.

The spectrum of a molecule depends on its energy level structure and therefore the absorption spectra are useful for identification of compounds. Measuring the concentration of an absorbing species in a sample is accomplished by applying the Beer-Lambert Law (Lasch 2002).

$$\mathbf{A = a\ b\ c} \qquad \text{Eq. 1}$$

Where **A** is the intensity (the amount of radiation absorbed) of the band; **a** is the absorbitivity or extinction coefficient which depends on the nature of the absorbing species, **b** is the sample thickness and **c** is the concentration of the component of interest.

IR absorption information is presented in the form of a spectrum with wavenumber at the x-axis and the absorption intensity or % transmittance the y-axis. Transmittance is the ratio of radiant power transmitted by the sample to the radiant power incident on the sample. Absorbance is the logarithm to the base 10 of the reciprocal of the transmittance.

Absorption of IR is restricted to compounds with small energy differences in the possible vibrational and rotational states. When the radiant energy matches the energy of the specific molecular vibration, absorption occurs. In order to absorb IR radiation the covalent bond of a molecule must undergo a net change in dipole moment as a consequence of its rotational or vibrational motion (Figure 2-3) (Siebert 2008). The vibrational wavenumber must also be identical to the wavenumber of the incident radiation. The dipole moment changes as the bonds expand and contracts. In contrast, if the vibration affects the polorisability of the molecule then the vibrational mode will be Raman active.

Figure 2-3: Change in the dipole moment of a molecule. The symmetric stretch in carbon dioxide is not IR active because there is not change in the dipole moment. The asymmetric stretch is IR active due to a change in dipole moment.

As the selection rules for Raman and IR are complementary, Raman can be used to detect vibrations missing in a compounds IR spectrum. Vibrations that are prominent in an infrared spectrum, those involving polar bond (e.g. C-H, C-Cl and C=O) and strong dipole moments are usually weak in a Raman spectrum. Likewise, non-polar vibrations that give very strong Raman bands usually result in weak infrared signals. For example, hydroxyl or amine stretching vibrations and the vibrations of carbonyl groups, are usually very strong in an FT-IR spectrum and usually very weak in a Raman spectrum. However, the stretching vibrations of carbon-carbon double or triple bonds and

symmetrical vibrations of aromatic groups are very strong in the Raman spectrum (McClain 2000).

2.3 Vibrational Modes

The absorption of IR radiation by a molecule is induced by changes in molecular dipoles associated with vibrations and rotations. Only a few compounds exhibit pure rotational bands and therefore are of little use. The vibrational absorption bands are of more interest. Vibrations can be subdivided into two types, depending on whether the bond length or angle is changing: stretching and bending. Stretching is a symmetric or antisymmetric rhythmical movement along the bond length. The bending vibration occurs when the bond angle between two atoms or a group of atoms change (Siebert 2008). These motions are known as scissoring, wagging, rocking, and twisting as illustrated in Figure 2-4.

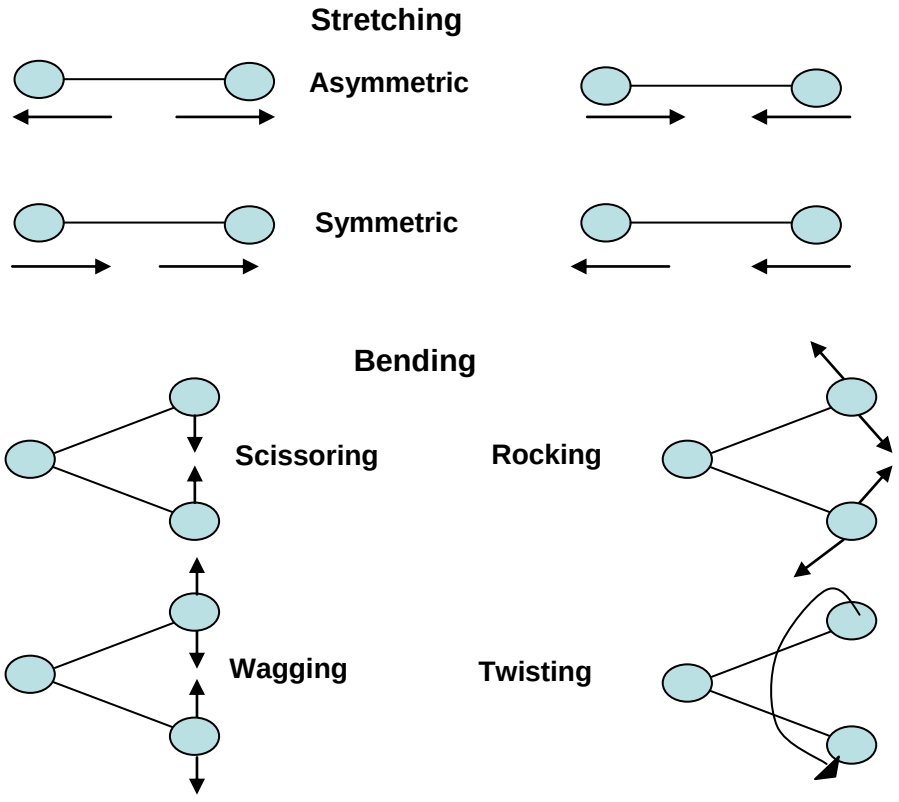
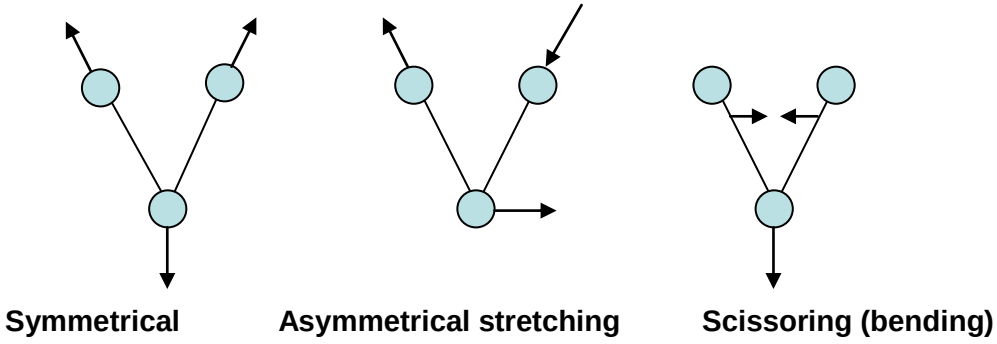


Figure 2-4: Vibrational modes

According to group theory, a molecule consisting of n atoms has 3 n degrees of freedom (Lasch 2002). In a nonlinear molecule, 3 of these degrees are rotational and 3 are translational and the remaining corresponds to fundamental vibrations; in a linear molecule, 2 degrees are rotational and 3 are translational. Water, which is nonlinear, has three fundamental vibrations (Figure 2-5A) and carbon dioxide, CO_2, is linear and hence has four fundamental vibrations (Figure 2-5B).

A



B

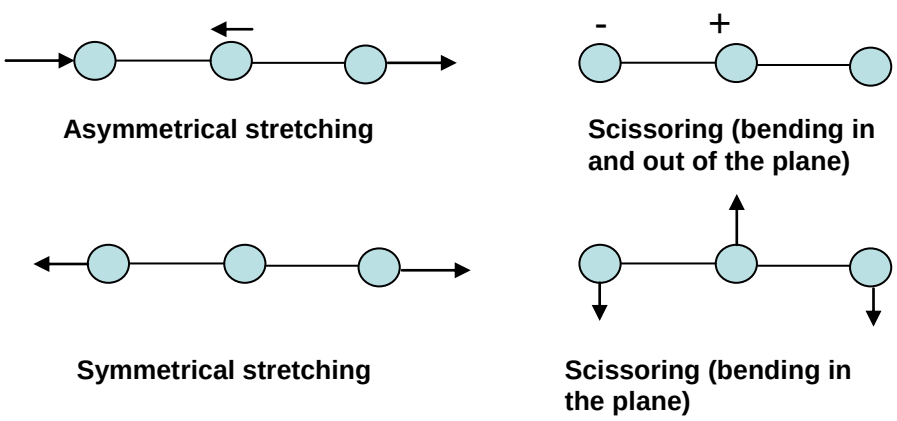


Figure 2-5: a) Stretching and bending vibrational modes for H_20 and b) Stretching and bending vibrational modes for $C0_2$

A symmetrical vibration produces no change in the centre and therefore is infrared inactive resulting in weaker vibrations. In contrast, the unsymmetrical stretching gives a charge separation and thus is capable of interacting with radiation of the same frequency. Stretching frequencies are higher than corresponding bending frequencies. It is easier to bend a bond than to stretch or compress it. Bonds to hydrogen have higher

stretching frequencies than those to heavier atoms. Triple bonds have higher stretching frequencies than corresponding double bonds, which in turn have higher frequencies than single bonds.

2.4 Infrared spectrum of bio-molecules

Molecular vibrations characterising different molecules can be found along the entire IR spectrum (Figure 2-6). The majority of analytical IR applications are confined to the IR region between 2000 and 650 cm^{-1}. This is called the fingerprint region of the spectrum. As most of the functional groups have relatively sharp absorption bands in the fingerprint region, quantification of the components can be determined using data handling techniques.

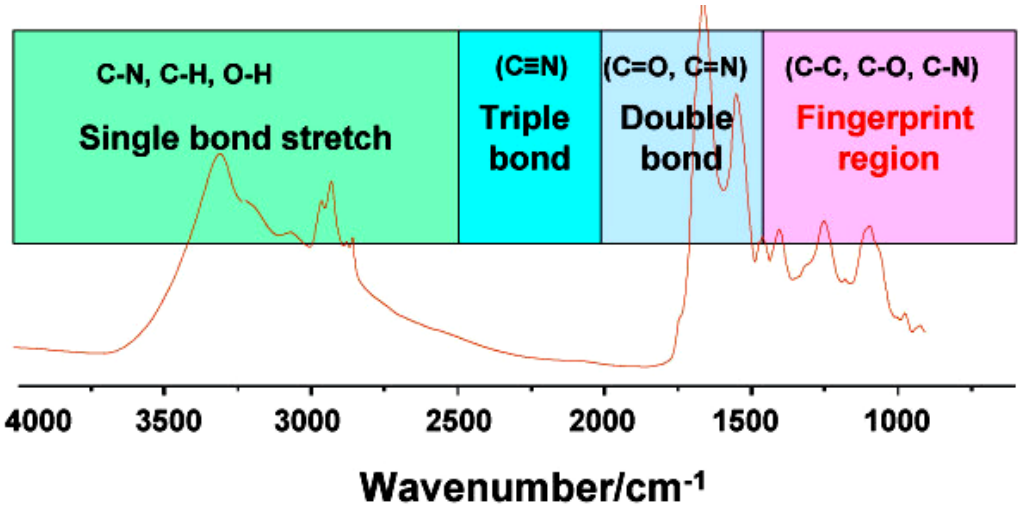


Figure 2-6: Correlation chart with functional groups in the IR fingerprint region from ranging between 2000–650 cm^{-1}. Adapted from www.biomedcentral.com

FT-IR spectroscopy probes the biochemical composition of a sample and determines the precise position and amplitude of infrared (IR) absorption by biochemical bonds. However, visual interpretation of IR spectra is very complex as the IR spectra results from various vibrational modes of molecules. Despite the complex molecular structure of cells, the major vibrational modes of molecules belong to lipids, proteins, nucleic acids and polysaccharides (carbohydrate).

2.4.1: Lipids

Lipids provide sources of energy for cells and play a major role as components in biological membranes. Spectral absorptions of lipids are found within the 3000-2800 cm^{-1} region. The bands at 2852, 2874 cm^{-1} are due to the symmetric CH_2 stretching of the methylene chains in membrane lipids (Lasch 2002); peaks at 2922 and 2960 cm^{-1} is due to the asymmetric CH_2 and CH_3 stretch (Mourant 2003; Lasch 2002). Other lipid bands mentioned in literature have absorption at 1099,1742, 1468, 1171 and 1058 cm^{-1} were assigned to the C = O (ester) stretch, CH_2 bend, C-O (ester) stretch and C-O stretch, respectively (Manoharan 1993). Lipids have long hydrocarbon chains and therefore their have peaks at 1352 cm^{-1} (CH_3 vibration) and 1740 cm^{-1} (C=O (ester) stretch) (Lasch 2002 and Erukhimovitch 2002).

2.4.2 Proteins

Proteins within single cells catalyze chemical reactions; provide structural rigidity, control membrane permeability and provide the means for gene expression. Proteins are composed of individual units called amino acids. In the spectra of a cell, intensive bands due to amide I (1620-1690 cm^{-1}) have been observed (Andrus 1998; Mourant 2003; Lasch 2002 and Gasparri 2003). These corresponds to the C=O stretching coupled to the N-H bending and the C-H stretching modes of peptide bonds (Sarver 1991; Chapman 1989 and Mourant 2003). Amide I band frequencies have been correlated with the presence of α-helices, β-sheets, β-turns and random coils (Byler 1986). It is well documented that absorbance in the range from 1645 to 1662 cm^{-1} is generally associated with the presence of α-helices (Byler 1988). Absorption bands between 1613-1637 cm^{-1}, 1662-1682 cm^{-1} and 1637-1645 cm^{-1} correspond to β-sheets, β-turns and random coils respectively. The amide II band between 1530-1570 cm^{-1} arises from the vibrational modes that involve the C-N-H bending and C-N stretching of peptidic bonds (Andrus 1998; Gasparri 2003 and Lasch 2002). The amide III band between 1240-1340 cm^{-1} is contributed by proteins arising from C-N stretching and N-H bending (Andrus 1998; Mourant 2003 and Gasparri 2003).

2.4.3 Nucleic Acids

A single cell consists of two types of nucleic acids. DNA is found mainly in the nucleus of the cell, while ribose nucleic acid (RNA) is found mainly in the cytoplasm of the cell although it is usually synthesized in the nucleus. The DNA contains the genetic codes to make RNA and the RNA contains the codes for the primary sequence of amino acids to synthesize proteins .Both nucleic acids have infrared active vibration in the region 900-1300 cm^{-1} (Benedetti 1997 and Omberg 2002). Nucleic acid signals contribute to the infrared spectra at 1091 cm^{-1} (sym PO_2^-), 1065 cm^{-1} (PO_2^- and ribose groups) and 968-972 cm^{-1} (phosphodiester group present in the DNA and RNA) (Lasch 2002 and Benedetti 1997).

2.4.4 Carbohydrates

Carbohydrates collectively include sugars, starches, saccharides, and polysaccharides. The name "carbohydrate" means a “hydrate of carbon”. Sugars are present in single cells as glucose, which provides source of energy. Polysaccharides are either found in free-state or bound to proteins known as glycoproteins. Major absorption peaks are observed at 1028 cm^{-1} (C-OH, C-O stretch), 1080 cm^{-1} (C-OH, C-O stretch), 1095 cm^{-1} (CH_2, C-C stretch) and 1455 cm^{-1} (CH_2, C-H bending) (Lasch 2002).

2.5 Infrared spectroscopy in cell and tissues

A large portfolio of research covering extensive FT-IR studies of malignant tissue has been established. A review of all tissue studies is beyond the scope of this thesis. However, important publications related to different disease sites are briefly summarised in Chapter 3. Advancement in infrared spectroscopy has permitted analysis at single cell level and extended the capabilities of this methodology to investigations based on cell morphology and supplementary biochemical information contained within the IR spectra. A review of single cell studies is also provided in Chapter 3.

2.6 Limitations of FT-IR spectroscopy

Limitations of FT-IR spectroscopy:

1) It cannot detect single atoms as they do not contain chemical bonds.

Consequently, they do not absorb infrared radiation and hence and do not possess any vibrational motion. Diatomic molecules, such as N2 and O2 also do not absorb infrared radiation and are also not detected.

2) The spectra acquired from samples are complex with overlapping bands and can be difficult to interpret. Peaks often correspond to more than one molecule.

3) Aqueous solutions have been proven to be difficult to measure. Water is a strong infrared absorber in specific wavenumber ranges (1270-2000 $cm^{-1}$and 3200-4000 cm^{-1}) and causes difficulty in infrared spectrum analysis. So FT-IR works best with solids, or films.

4) FT-IR spectrometers are a single beam technique. The samples and the background are measured at different times. In order to eliminate the instrumental and environmental contributions to the spectrum, the sample spectrum is ratioed against the background spectrum

5) The thickness of the sample has to be less than 10 microns for IR measurements.

Chapter 3: Biomedical Application of FT-IR Spectroscopy

IR spectroscopy is a powerful tool and is widely used for the study of bio-molecular structures. Being highly specific and sensitive, spectroscopy of molecular vibrations using mid-infrared has proven to be of great value in fields of biology and medicine for the analysis of biochemical compositions. This chapter reviews the application of infrared (IR) spectroscopy in various biomedical applications. The first Section (3.1) briefly summarises the diagnostic applications of in human malignant tissues. The spectroscopic study of biological cells is an active area of research. Recent publications on single cell analysis are reviewed in Section 3.2. The analysis of analytes in bio-fluids is discussed in Section 3.3 and finally, Section 3.4 reviews the application of FT-IR spectroscopy in the field of Haematology.

3.1 FT-IR spectroscopy in tissue diagnostics

Several research groups have established that IR spectroscopy can distinguish normal from diseased (malignant) tissue in a variety of organs. The majority of groups have focused their analysis on FT-IR imaging using an array detector that permits simultaneous imaging of tissue areas as an objective method. By collecting infrared spectra at each sampling point (pixel), it has been possible to acquire infrared spectral maps for tissue sections obtained from biopsy samples. In all cases, the spectral dataset acquired as a result of FT-IR spectroscopic imaging has been subjected to various data pre-processing modalities and multivariate analytical techniques in order to detect spectral differences.

FT-IR spectroscopic imaging technique has been applied to examine precancerous and cancerous tissue within biopsies from a variety of disease sites, such as cervical (Chiriboga 1998; Rigas 2000; Wood 2004); breast (Gao 1999; Dukor 1998; Eckel 2001; Fabian 2003); lung (Wang 1997; Yano 2000; Krafft 2008), oesophagus (Wang 2003; Wang 2007; Maziak 2007); Colon (Rigas 1990, Salman 2004; Lasch 2004), liver (Diem 2000), thyroid (Lui 2003) and prostate (Paluszkiewicz 2001; German 2006).

The main target of these studies is development of a clinical diagnostic method for early cancer detection. All studies have reported biochemical differences between normal and malignant tissues, which have been shown to correlate well with diagnoses provided by pathologists. Due to the heterogeneous nature of biological tissue, the infrared spectrum obtained from these sources consists of several absorption bands. Major peaks have been identified, corresponding to proteins, nucleic acids, lipids, and carbohydrates, which are the main constituents of biological tissues. RNA, DNA and protein levels were found to rise with proliferation and malignant grade while lipid levels remain constant with increasing malignant grade (Andrus 2006; Lasch 2002; Andrus 1998; Sahu 2004).

3.2 FT-IR spectroscopy in single cells

Spectroscopic studies of individual cells are more complicated than those of tissues. The initial work investigating the application of IR spectroscopy in detection of cancer cells was undertaken by Wong (1991). Several spectral differences were noted between normal and malignant cervical cells. However the results were based on visual inspection of spectra which is subject to bias. In addition, experimental method involving the use of pap smears (Papanicolaou test used to detect premalignant and malignant processes**)** is flawed, as pellets of cells were analysed rather than individual cells, contributing to variability in sample thickness and hence contributing to bias. Furthermore, pellets of cells generate grossly abnormal spectra. An extensive investigation into the potential confounding variables associated with cervical Pap smears was carried out by Wood (1998). Through analysis of different cell types and other variables and contaminants (mucous, blood and leukocytes) present in Pap smear by FT-IR spectroscopy it was found that visual inspection of spectra could potentially result in misdiagnosis.

With the goal of developing a more robust and sensitive spectroscopic diagnostic method, a multitude of single cell studies have been embraced over time. A thin preparation approach was developed to filter out debris and spread cells uniformly on slides (Cohenford 1998).

Given a number a potential confounding variables associated with cell analysis (Wood 1998), in most cases infrared spectroscopy has been coupled to multivariate analytical techniques to create effective and sensitive classification tools to determine spectral variance associated with cellular heterogeneity. Extensive studies using this combined approach have been conducted on cervical cancer cells (Wood 1996 and Cohenford 1997). In a more recent study cervical cell samples screened by FT-IR spectroscopy was compared with histological diagnosis (gold standard). FT-IR results versus histology showed sensitivity of 96.3% and specificity of 96.4 % (Sindhuphak 2003).

The characteristics of human cells have been studied by IR spectroscopy indicating significant differences between cellular cytosol and nuclear spectral signatures. Large differences are revealed in the spectral contributions of RNA, DNA and phospholipids in metabolically inactive cells (showing spectral signatures of proteins only) and dividing cells (showing spectral signatures of nucleic acids and phospholipids) (Diem 2002). Boydston-White (1999), revealed differences in FT-IR spectra between proliferating myeloid leukaemia (ML-1) cells in different phases of the cell cycle as being due to alterations in DNA content or amide I and II protein structure. Results show pronounced DNA detection in the S phase of cell cycle (DNA replication) and cytoplasmic RNA in the G1 a G2 phases of the cell cycle. Furthermore, cells in G2 phase synthesis more RNA than cells in G1 phase, resulting in boarding of RNA absorption bands. In an extended study of HeLa cells by microspectroscopy, changes in the spectrum of a single proliferating cell were also observed. This study reported spectral changes due to drastic biochemical and morphological changes occurring due to state of maturation, differentiation and development during cell proliferation (Boydston-White 2006). A similar study Mourant (2003) also assessed spectral differences in different proliferative states i.e. exponential (proliferating) or plateau (non-proliferating) phases of growth of fibroblast cells and reported that cell proliferation plays an important role in cancer initiation and progression. The ratios of RNA/lipid and protein/lipid increased when the cells were in the exponential phase compared to the plateau phase of growth. For these studies, homogeneous cell fractions obtained through centrifugal separation were utilised.

Research on single cancer cells is ongoing and subsequently a study by Matthaus 2006 reported the first IR microspectral imaging of mitotic human cells at different stages of mitosis. They report that variations in the intensities of protein and DNA spectral markers are dependent on chromatin density during the sub-phases of mitosis. It is well documented that distinct features differentiate between normal and malignant cells (Diem 2008). These include:

- Morphology and size of malignant cells are more variable compared to normal cells
- Malignant cells consists of larger and multiple nuclei than normal cells, therefore contain higher level of DNA and RNA
- Reduced glycogen level in malignant cells.

Single cell mapping at a sub-cellular level has been investigated on different types of cells. Studies have been conducted at a sub-cellular level revealing distinct variations between nuclear, heavy mitochondrial and light mitochondrial regions of a cell (Lasch 2002). IR microspectroscopy studies for individual normal and cervical cancer cells are also reported that allows a distinction between nuclear and cytoplasmic regions. Metabolically inactive cells have revealed spectral signatures of prominently proteins. In contract actively dividing cells also show spectral signatures of nucleic acids and phospholipids. Jamin 2003 investigated biochemical heterogeneity in cell death as a result of apoptosis and necrosis in single cells. Changes were observed at different stages in cell death between spectral regions of 1044-1234 cm^{-1}. The total absorbance intensity of nucleic acid is shown to be highly influenced by variations in the size of nucleus during necrosis and apoptosis.

Most work conducted on the distinction between normal and malignant tissue has focused on spectral differences in the spectral region of 600-1800 cm^{-1}. However, variations in the spectral region of 2800-3200 cm^{-1} have been reported (Sahu 2005). Results have shown alterations in lipid and fatty acid profile corresponding to cell membranes and organelles during multiplication of cancer cells.

Other references to single cell studies include investigations of oral mucosal cells (Romeo 2006), epithelial cells (Romeo 2006) and leukaemia cells (Boydston-White 1999 and Schultz 1996). As investigation of leukaemia cells is the main focus of this thesis, detailed review of publications in this area is included in Section 3.4.

3.3 FT-IR spectroscopy in biofluids

Besides the application of FT-IR to the tissue diagnostics and single cell analysis, its role in the diagnostic aspects involving body fluids is gaining importance. Mantsch group reported the successful diagnosis of rheumatoid arthritis based on near infrared (NIR) analysis of synovial fluid (Shaw 1995). Other authors revealed IR determination of the lecithin/sphingomyelin ratios in amniotic fluid (Lui 1997). The practical use of mid-IR spectroscopy in quantitative clinical applications has been previously demonstrated for serum analysis (Shaw 1998) and urine analysis (Shaw 2000) using partial least squares to calculate concentration of different analytes. For example, it was found that at least six serum analytes, namely albumin, total proteins, glucose triglycerides, urea, and cholesterol may be quantified on the mid-IR spectra of dried serum films (Shaw 2000; Petibios 1999). Measurements of the concentrations of glucose, total protein and cholesterol in human plasma have also been performed (Deleris 2003). The feasibility of FT-IR spectroscopy for the quantification of serum LDL-cholesterol and HDL-cholesterol concentration has also been reported (Lui 2002). Partial least squares (PLS) was utilised in most bio-fluid analysis as both the dependent and independent variables (absorbance and concentration) are used in the development of the calibration model. Recently attenuated total reflection (ATR) FT-IR spectroscopy has been developed for the quantitative analysis of blood plasma, whole blood and urine constituents. PLS was used to construct a prediction model which can calculate the concentration of albumin, cholesterol, glucose, total protein, urea and triglyceride in whole blood or blood plasma samples. For urine samples a calibration model was constructed to predict concentrations of urea, uric acid and phosphate and creatinine (Hosafci 2007).

Whole blood samples are also used for screening different diseases and cancer diagnostics as blood is more homogenous and requires very little sample preparation. In most cases, quatification of analytes in whole blood by infrared spectroscopy is performed on isolated cells that are dried into a thin film (Low-Ying 2002). However, Shen 2003 determined the concentration of glucose in aqueous whole blood samples using FT-IR spectroscopy and PLS analysis. Furthermore, Khanmohammadi 2007 distinguished between healthy subjects and those with basal cell carcinoma (BCC) by investigating the IR absorbance spectra from aqueous blood samples. Major spectral changes were observed in proteins and nucleic acids between the normal and BCC samples. Furthermore results revealed 96.7% of accuracy when compared to current clinical methods.

3.4 Applications of FT-IR in Haematology

It is only in recent years that IR spectroscopy has been adequately well-developed to enable spectroscopic studies of cells and molecules within blood. There are various contributions of IR spectroscopy to haematological malignancies, including studies of differentiation between normal leucocytes and leukaemic cells; classification and establishment of DNA modifications in leukaemia; chemotherapy assessments, determination of apoptosis in leukaemia cells and differentiation of clinico-pathological subdivisions of lymphoma.

3.4.1 Differentiation of leukaemic cells from normal leukocytes

An initial infrared study by Polli (1955) demonstrated no spectroscopic differences between the DNA isolated from normal and leukaemic human leukocytes. Following improvements in IR spectroscopic instrumentation and techniques, research carried out by Benedetti 1984 revealed various differences in the infrared spectra obtained from isolated normal and leukaemic cells. They used a group frequency concept and identified two DNA marker bands in the IR spectrum, at 966 cm^{-1} and 530 cm^{-1} characteristic for lymphoid leukaemia. In addition, they showed that the ratio between

two integrated spectral areas (the symmetric PO^{-2} stretching absorption band at 1080 cm^{-1} and the amide II absorption band at 1540 cm^{-1}) discriminates between normal lymphocytes and leukaemic cells. Total DNA content was increased in leukaemic cells isolated from either CLL or from ALL cases, compared to that in normal mononuclear cells.

Rather than relying on the classical group frequency approach to identify individual marker bands and interpret changes in the infrared spectra of leukaemic cells, Schultz (1996) alternated to pattern recognition algorithms that make better use of all the fingerprint-like characteristics of the infrared spectrum. Spectral characteristics of human lymphocytes isolated from the peripheral blood of normal and leukaemic (CLL) subjects were analysed. Multivariate statistical analysis has been used to classify spectra into intrinsic groups. Unsupervised multivariate analysis revealed that spectra from normal blood separated into two subgroups, due to variations in the proportion and degree of activation of lymphocytes and monocytes and differences in DNA and lipid content. In contrast, the CLL cells separated into three subgroups distinct from the normal cell grouping due to cytogenetical variations corresponding DNA content. Supervised multivariate analysis including LDA has been used to construct a classification model and to identify new samples. This method was successfully applied to discriminate between sensitive and multi-resistant human leukaemic K562 cells using cross validation (Gaigneaux 2002).

FT-IR spectroscopy has also been applied to the analysis of human plasma samples for monitoring and identification of patients with chronic lymphocytic leukaemia (Erukhimovitch 2006). Results showed reduction of carbohydrate, Amide III and amino acid peaks in CLL patient plasma samples compared to plasma samples obtained from healthy persons to due to higher metabolic activity and lower levels of nutrients in leukaemia cells.

These findings reveal that CLL cells can be differentiated from normal cells using FT-IR spectroscopy, based on qualitative and quantitative differences at the DNA, protein and lipid levels. For example, leukaemic cells have increased DNA content, conformational alterations in DNA, and reduced lipid content signifying that infrared

spectra can contain valuable information to allow differentiation of the cell state that can be used for diagnostic purposes.

3.4.2 Chemotherapy and sensitivity/resistance assessments

A study by Lui (1997) extended the previous studies by Schultz (1997) from diagnosis to prognosis. They assessed whether variations observed in FT-IR spectra of CLL cells can be used to predict the ex vivo sensitivity/resistance to treatment with two common antileukaemic drugs, chloroambucil and cladribine. The spectral changes indicated that the main cellular changes responsible for differences in the drug-sensitive or drug-resistant cells were: lipids, sugars/phosphates (DNA and RNA) and proteins.

A further study by Schultz (1997) showed a new aspect of infrared diagnosis of CLL – patient monitoring. They revealed that utilising a differential of cell alteration in CLL (indicative for DNA base pairing), it seems to be possible to monitor the disease progression of patients over time and after drug treatment. A relative intensity of a DNA band seems to be indicative of the severity of the disease and may predict the outcome of drug-treatment (Schultz 1997). These promising results may also provide the foundation for a future diagnostic tool of blood cancer in haematology.

In a recent publication, Ramesh (2002) employed FT-IR spectroscopy to follow-up chemotherapy treatment for B- and T-cell childhood leukaemia patients. Predominant spectral changes in the leukaemic cells occur in the region between 1000 and 1200 cm^{-1}, which corresponds to the cellular nucleic acids. Good conformity was noted between reduced cellular nucleic acid content and reduction of blasts, due to chemotherapy treatment, in both B- and T-cell patients. Decreased levels of total carbohydrates, cholesterol, phospholipids and proteins level were also observed.

A study by Sahu (2006) assessing the effectiveness of chemotherapy in leukaemia patients have also reported similar patterns in spectral difference in the WBC of an adult AML patient and a child ALL patient. Results reveal that phosphate absorbance from nucleic acids and the lipid-protein ratio in the WBC decreased immediately after

treatment (Sahu 2006). Understanding the molecular changes in cells following chemotherapy treatment could be used for monitoring of patients during chemotherapy.

3.4.3 Determination of apoptosis in leukaemia cells

The deregulation of apoptosis, a pre-programmed physiological mode of cell death, is of great importance in the pathogenesis and progression of cancer. As an active process, apoptosis is accompanied by profound biochemical changes to three essential cellular components, i.e. DNA, proteins and lipids simultaneously and may provide a new tool for monitoring the apoptotic process and its effect on the conventional anticancer drugs and irradiation therapy. Several research groups have demonstrated that IR spectroscopy can detect apoptosis induced by different reagents in various cell lines. Research by Lui 2001 applied IR spectroscopy to estimate the percentage of apoptotic leukaemic cells in two different cell lines (CEM and 562) induced with etoposide (VP-16). Apoptotic changes in DNA were observed in the spectral region between 900-1300 cm^{-1}. As the percentage of apoptosis increases, the protein structure of the treated cells shifts from ß-sheets to unordered (random) coil, the overall lipid content increases and the amount of detectable DNA decreases.

Similar findings were obtained in a recent study carried out by Gasparri 2003 who induced apoptosis in the human promyelocytic leukaemia cell line, HL60, with camptothecin, a cytotoxic drug over a period of time and analysed apoptotic cells by using a FT-IR approach based on an attenuated total reflection (ATR). A decrease in the overall spectral DNA in the apoptotic cells was observed.

3.4.4 Differentiation of lymphoma clinico-pathological grades

A study by Andrus (1998) has demonstrated that FT-IR can distinguish between clinically aggressive (high-grade) and clinically indolent (low-grade lymphomas) based on the nucleic acid absorbance ratio. They observed that absorbance ratio A1121/A1020 increased, along with the emergence of an absorbance peak at 1121cm^{-1}, with increasing clinico-pathological grade of malignant lymphoma. Furthermore, the absorbance profile

moves closer to that of RNA and further from that of DNA with increasing lymphoma clinico-pathological grade. This is because clinically aggressive neoplasms have more cycling cells in the G1 and S phases of the cell cycle and/or fewer apoptotic cells, and therefore greater absorbance at 1121cm^{-1} (RNA) relative to that at 1020cm^{-1} (DNA) when compared with indolent neoplasms. Absorbance attributable to collagen increased with lymphoma grade and was greater in benign inflammatory tumours than in low-grade lymphomas. The A1121/A1020 trend observed may form the basis of a universal cancer grading to assist with cancer treatment decisions. FT-IR micro-spectral imaging of non-Hodgkin lymphoma is also reported using Focal Plane Array detector (Burattini 2007). Spectral differences were reported between reactive and tumour lymph nodes. Major spectral differences between normal and malignant spectra were identified in the low frequency region. Spectral classification using principal component analysis (PCA) and hierarchical cluster analysis (HCA) between distinct compartments such as intrafollicular, interfollicular and mantle zones of healthy and tumoral lymph nodes were also reported.

3.5 Summary

This chapter has documented that a great deal of work has been done on various cancerous and precancerous tissue types. FT-IR spectroscopy has also shown to be feasible in single cells studies and bio-fluid analysis. However, due to the complexity of the leukaemia and lymphomas classification system and variations in the biochemical composition of different lineages of leukemic cells, there is a need for a more reliable classification algorithm to discover spectral patterns and characterise spectral differences among different cell progenitors.

To date various research publications have focused their attention on detection of only one subtype of leukaemia, mostly comparing spectral signatures of normal and chronic lymphoblastic leukaemia (CLL) cells. Previous FT-IR spectroscopic studies have used only spectral differences between normal and abnormal cells, PCA or HCA to demonstrate potential causes contributing to the variance in spectra rather than classifying the biochemical and spectral variations.

This aspect has been ignored in previous studies, but needs to be understood for diagnostic applications of infrared spectroscopy. This research project investigates whether FT-IR spectroscopy can be applied in the classification of multiple sub-types of leukaemia and lymphoma cells. The focus of this thesis is to construct a classification model to classify cells derived from human T-cell lymphoma, B- cell lymphoid, and myeloid leukaemia cell lines by using a combination of spectroscopic measurements and powerful multivariate (supervised and unsupervised) statistical analysis. The focus of the research has also expanded to include the investigation of normal and CLL blood samples as well as blood fractions (plasma and buffy coat) in order to obtain more information on cellular composition. To date, no one to our knowledge has compared the biochemical composition of whole blood, plasma and buffy coat blood fractions with FT-IR spectroscopy. Understanding differences in the biochemical composition of various fractions of blood by FT-IR spectroscopy is fundamental to understanding disease progression as well as diagnostics and screening of different cancers. In an attempt to use valuable information from full blood count through laboratory investigations, partial least squares was utilised to develop a predictive model to determine concentrations of neutrophil, lymphocyte, monocyte, eosinophil and basophil in buffycoat samples.

Chapter 4: Data Analysis

This chapter reviews the techniques used for pre-processing spectral data and subsequent chemometric techniques used for developing a classification model. The main emphasis is on both univariate and multivariate analytical techniques.

4.1 Data pre-processing

Raw data acquired through FT-IR spectroscopy requires spectral manipulation to extract qualitative and quantitative information. Spectra are subjected to pre-processing methods to effectively suppress or eliminate unnecessary artefacts and noise. Noise can obscure important features within the IR spectra such as peaks or peak widths, therefore calculations of peak areas and peak widths can prove difficult.

In this study, the methods of spectral manipulation employed include: smoothing, spectral derivative, normalization and mean-centring. The applications of either standard deviation or an absorption filter tool in order to further improve the quality of dataset were also investigated.

4.1.1 Smoothing

The Savitzky-Golay smoothing filter was applied using the "Savgol" function from the PLS toolbox in Matlab. The Savitzky-Golay smoothing filter was used to smooth a noisy signal (high frequency fluctuations) in spectra. The filter is defined as a weighted moving average with weighting given as a polynomial of a certain degree (Savitzky 1964). The filter uses any number of points for the weighted average. However, over smoothing using a large number of points can distort band shapes and therefore; a nine-point Savitzky-Golay zero-derivative filter was used in this study.

4.1.2 Spectral Derivative

Derivation is used to eliminate fluctuations in the baseline (first derivative) and slope (second derivative) of the absorption spectra. Calculating the derivative of a mean spectrum is a method of identifying peak frequencies of biochemical components and therefore provides qualitative and quantitative information. It also causes effective narrowing of spectral features, which may enhance the specificity of the method.

An infrared spectrum can be described as a mathematical function; therefore one can calculate and determine the slope of such function by taking the derivative (Ng 1999). The resulting derivative spectrum contains important features. In the second derivate spectrum, the downward pointing peaks pinpoint the exact wavenumber of the maximum absorbance of the peak. These wavenumbers can be used to estimate the number of overlapping bands and locate their peak positions (Ng 1999).

4.1.3 Normalization

Normalisation removes the variations in absolute signal intensity as a result of thickness variations. Normalisation can either be carried out to a well defined peak, although overlapping bands can be difficult, or with the total area under the curve.

4.1.4: Mean-centring

Mean centring is a common approach to standardise data to ensure that spectral intensities of each spectrum could be statistically compared to the mean of the dataset. Mean-centring is achieved by calculating the mean of the dataset and subtracting that mean from each spectrum.

4.1.5 Standard deviation

The standard deviation at each wavelength was calculated as a measure of the spread or variance of the spectra in a dataset about its mean. A low standard deviation indicates that the data points tend to be very close to the same value (the mean); while high standard deviation indicates that the data are "spread out" over a large range of values. Using standard deviation is an effective method of removing extreme spectra in the dataset. These may be likely to include spectra with low signal to noise levels and high signal levels indicating thin (or no sample) or excessively thick sample areas. The standard deviation used in this study is defined in the methodology Section 5.9.2.

4.1.6 Absorption filter tool

Absorbance filter tool filters out weak or low absorbance. All spectra outside a defined absorbance range are removed. The absorbance range used in this study is defined in the methodology Section 5.9.2. The absorbance filter tool in combination with other pre-processing techniques such as normalisation are used to enhance the spectral features that incorporate information regarding biochemical composition and effectively suppresses or eliminate unnecessary artefacts, noise, calcium fluoride, water and CO_2 signals as well as variations in sample thickness.

4.2 Chemometrics for spectral anlaysis

Univariate analytical methods are the simplest ways of displaying the biochemical information and determining peak positions. Although this is a simple approach of spectral analysis, it only utilises a small part of the biochemical data. As the intricacy of the IR spectrum increases within and between sample types, unambiguous assignment of the spectral bands to specific molecular groups becomes more difficult. Therefore in complex biological samples, multivariate statistical techniques are employed as they utilise the entire spectra and all of the biochemical information contained within the spectra. It is therefore likely to provide a more rigorous approach at separating spectra by cell type. Two generic strategies of multivariate analysis namely, supervised and

unsupervised classification methods, were used to retrieve analytical information from the IR spectra.

4.2.1 Univariate Analysis

Univariate methods used in this study include: assignments of peaks for identification of specific spectral features; calculating difference spectra and measurements of absolute peak intensity and peak intensity ratios.

4.2.1.1 Peak assignments

In order to identify spectral signatures a "group frequency approach' was used as an traditional univariate method of spectral analysis and assigning key peak positions to mean spectrum. This approach monitors vibrations due to pre-assigned specific functional groups (e.g. C=O, N-H, C-H, etc) and explores changes in these functional groups as potential disease biomarkers.

4.2.1.2 Peak Intensity Ratio

Absolute peak intensity of the mean spectrum was calculated to determine variation amongst spectra that are not identified visually and also to determine possible changes in peak positions and peak shifts. The intensities (height) of the most significant peaks in the mean IR spectra were measured relative to the baseline. The intensity values were then used to calculate peak ratios.

Peak intensity ratio calculations reflect changes in the content and structures of proteins, nucleic acid and fat in cells, thus the diagnosis of cancer cells could be made at molecular level. This method of has been used for detailed spectral analysis by many groups (Wang 2003; Mordechai 2005; Bonnier 2006). In this study, peak intensity ratios corresponding to proteins, lipids and nucleic acids were used.

4.2.1.3 Difference spectra

Spectra are often similar visually, with only subtle variations in peak height and peak position from one group to another. Calculating difference spectra emphasizes the variation between spectra. A difference spectrum was calculated by subtracting the mean normalized spectrum from one group from the mean normalized spectra from another group.

4.2.2 Multivariate Analysis

Combination of unsupervised and supervised multivariate techniques was employed in this study to construct a classification model based on variance between the spectra. Initially, an unsupervised classification method of principal component analysis (PCA) was employed. Supervised classification method, known as linear discriminant analyses (LDA) were used to construct a classification model. Many groups have used PCA-fed-LDA for separation of FT-IR and Raman spectral data (Stone 2000; Kendall 2003; Pichardo-Molina 2007 and Lyng 2007)

Furthermore cross validation was used to probe the quality of spectroscopic classification and performance of the classification model.

4.2.2.1 Principal Component Analysis (PCA)

PCA is an unsupervised classification method, where the spectra are separated solely on measure of their variance. It combines the advantages of both unequivocal classification and outlier detection (Bossart 2003). It is a well-established method ideally suited to distinguish and visualise small, reoccurring spectral variations in large datasets. These can include both spectral data and pseudo-colour PC score images. PCA reduces the number of parameters needed to represent the variance in the spectral dataset. It resolves a complete spectral dataset into a few key spectral components and can thus identify and isolate important trends within the dataset (Lasch 1998). The first principal component (PC) represents the highest variance in the spectral dataset; the second PC represents the

next highest variance. Noise is introduced in later PCs. ANOVA ("analysis of variance") was used to identify the most significant principal components reflecting biochemical trends and differences. The ANOVA test calculates an F-value from the sum of the squares of the variance. The larger the F-value results in more effective clustering of spectra within a group. Fcrit is also calculated during ANOVA which is the cut-off value for 99% significance.

The most significant PC scores for each individual spectrum in the dataset were plotted against one another to enable the natural clustering with cell line type or blood group to be visualised. This approach identifies natural clusters or grouping with no prior knowledge of their classification. Hence PCA is known as an unsupervised technique.

Generally, PCA is not ideal for the separation of the groups. Therefore a multivariate technique that involves an external classification scheme to separate spectra must be applied.

4.2.2.2 Linear Discriminant Analysis (LDA)

LDA is a supervised method, since the number of classification groups (group identity) is dictated rather than discovered during the analysis. It uses PC scores (describing the greatest variance in the dataset) to calculate linear discriminant functions that maximises the variance between the different groups and minimises the variance within each group (Kraft 2006). To prevent over-fitting of data it is important to use minimum PCs. LDA has been used widely for classification of gastric biopsies (Wu 2005), Thalassemia (Mantsch 2003); oesophagus (Stone 2002; Kendall 2003) and laryngeal malignancies (Stone 2000).

4.2.2.3 Leave-one-out Cross Validation (LOOCV)

Preferably, a classification model must be tested using an independent dataset. In this study, due to limitations with patient and sample numbers this was not possible. To overcome this, leave-one-out cross validation was used. Similar approaches for

validating classification models have been used in investigation of synovial fluids from joints affected by rheumatoid arthritis and osteoarthritis (Shaw 1995) and classification of gastric inflammation and malignancy on endoscopic biopsies (Li 2005).

Leave-one-out cross validation (LOOCV) is often used to test the accuracy of the predictions of the classification model. Classification accuracy is assessed whereby the class of spectra within a batch is predicted by holding out all of the spectra measured from a particular sample (the test set) and building an LDA model with the remaining data (the training set). This method is repeated so that each sample in the dataset is left out in turn and used once as the test data. Sensitivities and specificities were calculated as a measure of model accuracy and performance to evaluate the efficacy of classification systems. Sensitivity is defined as the ratio of the number of true positives to the sum of true positives and false negatives and the specificity is defined as the ratio of the number of true negatives to the sum of true negatives and false positives. See equations 2 and 3.

$$\text{Sensitivity} = \frac{\text{No. of True Positives}}{\text{No. of True Positives} + \text{No. of False Negatives}} \times 100\% \qquad \text{Eq.2}$$

$$\text{Specificity} = \frac{\text{No. of True Negatives}}{\text{No. of True Negatives} + \text{No. of False Positives}} \times 100\% \qquad \text{Eq.3}$$

4.2.2.4 Partial Least Squares (PLS)

PLS is useful to extract information from complex spectra and is quantitative spectral decomposition technique. It utilizes the spectral data and concentration information during the decomposition process. PLS takes advantage of the correlation relationship between the spectral data and the concentration of cellular analytes. Two sets of scores are produced as a result; one set for the spectral data and the other for the analyte concentration which enables a construction of calibration model.

When applied to spectra, PLS analysis finds a mathematical relationship between a set of independent variables, the X matrix ($N_{objects}$ x $K_{wavelenghts}$), and set of dependent variables, the Y matrix ($N_{objects}$ x $M_{measurements}$). The resulting model has the form:

$$Y = XB + E \qquad \text{Eq.4}$$

where B is the matrix of regression coefficients obtained from PLS analysis and E is the matrix of residuals (Osborne 1997). PLS is also a regression technique, which utilises latent variables (LV). These variables are independent variables. Latent variables, like PCs are calculated to determine the most of the variance in the dataset. The first LV describes most of the variance in the independent set, the second LV the next largest amount of variance and so on. The aim is to obtain most of the concentration information into the first few latent variables that account most of the variation to include in the model. Cross validation is performed in the construction of PLS model to estimate the number of LV required and estimate prediction error. The error is determined by fitting a PLS model to n-1 samples and making a prediction value for the omitted sample. This was carried for each sample in the dataset and a predictive sum of squares (PRESS) was calculated. The CUMPRESS was also calculated, which is the cumulative error.

There are two types of PLS algorithms: PLS 1 and PLS 2. PLS 2 calibrate for all cellular analytes simultaneously and as a result one set of scores are produced. In PLS 1 separate scores are calculated for each cellular analyte and therefore separate sets of

scores are calculated for each analyte. This gives a more accurate prediction. Since separate scores are generated for individual analytes, the speed of calculations is considerably increased. In this study PLS 1 was used.

PLS analysis has been used in various studies for the quantification of total proteins, cholesterol, high and low density lipoproteins, triglyceride, glucose, urea and uric acid (Rohleder 2005; Low-Ying 2002; Petiois 1999) in fluids, cells and tissues (Petiois 2006).

4.3 Summary

The main analytical techniques used in this study for data analysis have been briefly reviewed in this section. It is important that univariate and multivariate analytical techniques to create a statistical method to extract the relevant biochemical information buried in the spectrum and convert this experimental data to useful diagnostic information. Application of these statistical methods has been discussed in Chapter 5.

Chapter 5: Materials and Methods

The first Section (5.1) in this chapter begins by defining the pathological classification of cell lines used for the preliminary phase of the study. Detailed description regarding cell line origin and culture conditions are also discussed in this section. Section 5.1.1 describes the procedures involved in preparation of cell line samples. Section 5.2 describes the procedures involved in collection of normal blood specimens from healthy subjects and leukemic blood specimens from patients with chronic lymphocytic leukaemia (CLL). Isolation of whole blood, plasma and buffycoat and procedures of sample preparation are discussed in Section 5.2.3.

Section 5.3 covers the key components of the FT-IR spectrometer, followed by sampling techniques in Section 5.4. Sections 5.5 and 5.6 focus on the principles of infrared microspectroscopy and FT-IR imaging respectively. Optimised parameters for spectral measurements are discussed in Section 5.8. Technical issues such as the effects of aperture size and number of co-added scans on the signal-to-noise ratio of spectral measurements have been investigated prior to conducting spectral measurements on both cell line and blood samples. The final Section 5.9 in this chapter looks in greater detail at the spectral measurements as well as the order of statistical and analytical methods used in this study.

5.1 Cell lines

The preliminary studies carried out as part of this research project, rely on measuring FT-IR spectra from cell line samples. Cell lines as a rich resource of abundant, accessible and manipulable living cells were used to better understand the pathophysiology of haematopoietic malignancies and identify potential biomarkers. The major advantages of using cell lines are the unlimited supply and availability of identical (homogenous) cell material unlike real tissues. It is important to use homogenous cell material in order to evaluate the potential of spectroscopic methods before implementing them on complex cellular material and blood specimens. Another

advantage of using cell lines is the possibility to control growth rate and the infinite viable storability in liquid nitrogen.

Leukaemia and lymphoma cell lines used in this study are characterized generally by: monoclonal origin i.e. derived from one malignant precursor cell; differentiation arrest at a discrete stage during maturation; sustained autonomous proliferation of the cultured cells. Leukaemia and lymphoma cell lines originate from different cell lineages based on their immunophenotype, genotype and functional features and grow as single or cultured cells in suspension.

For the purpose of this study five different cell lines were supplied by the Institute of Cancer Research, Sutton. These included: Karpas 299 (Human T-cell lymphoma cell line) cells, REH (Human acute lymphoid cell line) cells, RCH-ACV (Human acute lymphoid cell line) cells, HL60 (Human acute myeloid leukaemia cell line) cells and Meg01 (Human chronic myeloid cell line) cells. A total of 5 sets of the same cell types were supplied as separate batches (25 cell samples in total). The cells were grown as a suspension in 80-90% RPMI 1640 medium containing 10-20% fetal bovine serum (FBS). Cells were maintained at 37ºC in a humidified atmosphere with 5% $CO_{2.}$ Table 5-1 gives detailed information regarding cell line origin and culture conditions. This information was provided by the Institute of Cancer Research.

Cell Line	Cell Type	Origin	Medium	Incubation
REH	Human B-cell precursor leukaemia	Established from the peripheral blood of a 15-year-old North African girl with ALL in 1973.	90% RPMI 1640 + 10% FBS	at 37 °C with 5% CO_2
RCH-ACV	Human B-cell precursor leukaemia	Established from the bone marrow cells taken at relapse of common ALL, 7 months after diagnosis, from an 8-year-old girl treated with chemotherapy.	80-90% RPMI 1640 + 10-20% FBS	at 37 °C with 5% CO_2
HL60	Human acute myeloid leukaemia	Established from the peripheral blood of a 35-year-old woman with AML FAB M2 in 1976;	90% RPMI 1640 + 10% FBS	at 37 °C with 5% CO_2
Meg 01	Human chronic myeloid leukaemia in megakaryocytic blast crisis	Established from the bone marrow of a 55-year-old man with CML in megakaryocytic blast crisis in 1983	90% RPMI 1640 + 10% FBS	at 37 °C with 5% CO_2
Karpas 299	Human T cell lymphoma	Established from the peripheral blood of a 25-year-old man with T cell non-Hodgkin lymphoma in 1986; now classified as "CD30+ anaplastic large cell lymphoma"; cells carry the NPM-ALK fusion gene	90% RPMI 1640 + 10% FBS	at 37 °C with 5% CO_2

Table 5-1: Cell line characteristics and culture conditions. Source: Institute of Cancer Research, London

5.1.1 Sample Preparation

Prior to spectroscopic measurements the cells were washed and diluted (1:5) in cold, sterile saline (0.9% NaCl) solution. The samples were pelleted by centrifugation at 3000rpm for 3 mins to concentrate the cells. The pellet was re-suspended in 0.9% phosphate buffer saline solution and the washing cycle was repeated twice more. Typically 200µL of the final diluted cell samples was transferred onto CaF_2 slides by

Shandon Cytospin® Cytocentrifuge (1500 rpm for 5 min), a method formerly developed for cytology as described below. The cytospin produces a sparse monolayer of cells so that cells are separated. Deposition of cellular material using a cytospin has been previously used by other groups for FT-IR spectral measurements (Romeo 2006; Romeo 2006 and Crow 2005).

Cells were transferred onto two sets of slides for each sample. Both sets of slides were then allowed to air-dry at room temperature under airflow for 30 min as thin circular disk films (diameter of 5 mm). The air-drying approach overcomes the difficulties associated with strong water absorptions by simply eliminating water from the specimen. This also eliminates the spectral inference of water and provides better spectral resolution by reducing the water/solute interactions. One of the two sets of slides was frozen at -80°C. Prior to FT-IR measurements, the frozen slides were allowed to defrost at room temperature. Infrared spectroscopic measurements on dried films have been reported by various groups (Boydston-White 1999; Ramesh 2001 and Cohen 1998).

The Shandon Cytospin® Cytocentrifuge is a thin-layer cell preparation system. It consists of Cytofunnel® and a layer of filter paper with a 5 mm diameter hole, which is clamped against a CaF_2 slide using a Cytoclip™ slide clip (Figure 5-1). The entire assembly is placed into a special rotor of a centrifuge, and spun. The centrifugal force separates and deposits a thin-layer of cells onto a clearly defined area of the CaF_2 slide while maintaining cellular integrity. The residual fluid is absorbed into the sample chamber's filter card.

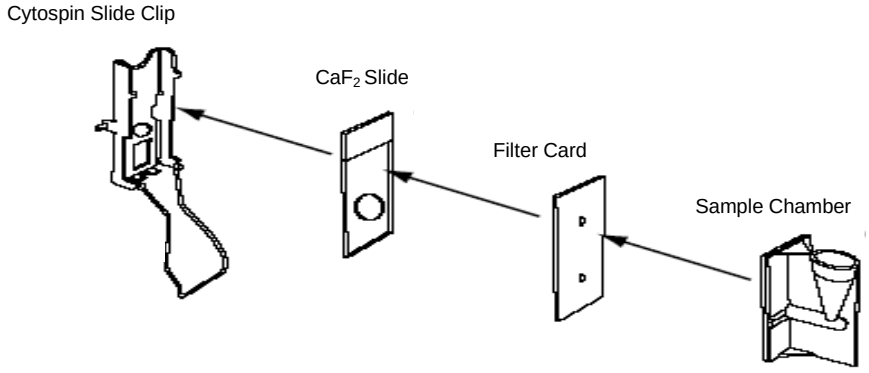


Figure 5-1: Assembly of the Cytofunnel and Cytoclip slide clip. Adapted from www.thermo.com

5.2 Normal and CLL Blood Samples

5.2.1 Ethical Approval

All experimental procedures of the study were assessed and approved by the Gloucestershire Local Research and Ethics Committee prior to commencement. This process included review of study design, protocols, patient information and consent forms. All study subjects gave informed consent prior to participation. The project REC reference number for the study is 06/Q2005/17.

5.2.2 Sample Collection

Blood samples were collected from 10 healthy subjects through voluntary participitation. 12 chronic lymphocytic leukaemia (CLL) blood samples were collected for this study during routine check-up procedures and redundant samples were used for research purposes. Full Blood Count (FBC) was carried on both normal and CLL blood samples prior to further sample preparation. The FBC reports were reviewed by the Consultant Haematologists involved in the research. Subjects with normal FBC (healthy volunteers) and those with a high FBC (CLL patients) were included in the study in accordance with the study protocol.

5.2.3 Isolation of whole blood, plasma and buffycoat

Blood samples from healthy volunteers and CLL patients were collected in 9 ml vials containing ethylene diamine tetra-acetic acid (EDTA) an anticoagulant agent at 25ºC. Approximately 3ml was transferred into another EDTA vials as a representative whole blood sample. 3 ml of blood was loaded over the 3 ml of histopaque in the Accuspin™ Histopaque®-1077 tubes (purchased from Sigma-Aldrich, UK) and centrifuged (3000 rpm for 10 min) at 23ºC. Histopaque tubes were used to isolate lymphocytes and mononuclear cells as well as plasma from whole blood with the aid of a separation medium, HISTOPAQUE-1077. This is an aqueous solution of a high molecular weight polysaccharide and sodium diatrizoate with density of 1.077 g/ml. The separation of blood components using the HISTOPAQUE-1077 medium avoids the possibility of red blood cell contamination.

On centrifugation of blood, three fractions were generated: fraction A, the red blood cell layer (bottom layer), which descends through the "frit" (porous high density polyethylene barrier); fraction B, a clear and yellow plasma fraction (top layer) and fraction C, the so-called "buffy-coat" (middle layer) mainly consisting lymphocytes and mononuclear cells as shown in Figure 5-2. The dense buffycoat layer and the plasma fraction were extracted using a pipette into separate glass tubes. For each blood sample, 3 glass tubes were labelled 1) whole blood, 2) plasma and 3) buffycoat.

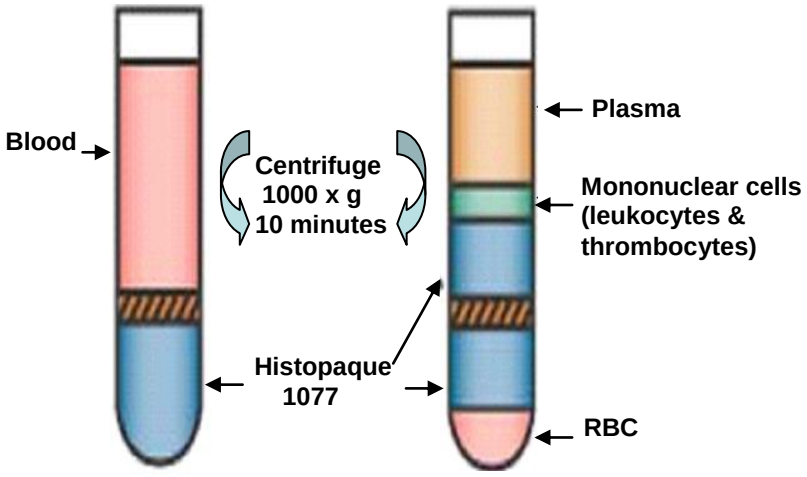


Figure 5-2: Separation of blood components using HISTOPAQUE-1077

5.2.4 Sample Preparation

Approximately 200µl of whole blood, buffycoat and plasma samples were transferred onto CaF_2 slides by Cytospin® Cytocentrifuge (1500 rpm for 5 min) as described in Section 5.1.1 without any sample preparation.

In order to isolate cells from the buffy coat samples, cells were centrifuged (1000rpm for 5 min) to concentrate the cells into a pellet. The pellet was washed twice with physiological sterile saline (0.9% NaCl) and re-suspended in 100 µL saline to remove any adhering media and organic compounds. Buffy coat samples were then deposited onto CaF_2 slides. Only one slide was prepared per sample which was subsequently frozen at -80°C. Prior to FT-IR measurements, the frozen slide was allowed to defrost at room temperature. Small volumes of EDTA and Histopaque were also transferred onto CaF_2 slides to assess any biochemical contribution from these agents on the spectra obtained from a blood sample.

5.3 Components of the Spectrum Spotlight FT-IR Imaging System

The infrared spectral data presented in this study were collected using a Perkin-Elmer Spectrum One/Spotlight FT-IR Imaging 300 System as shown in Figure 5-3. This instrument is incorporated with a computerised microscope x-y-z stage, visual image capture software and a highly sensitive, liquid nitrogen-cooled 1 x 16 element HgCdTe (MCT) array detector with a step scanning interferometer.

The Spectrum Spotlight FT-IR Imaging System enables the collection of images or spectra from small areas of samples either in transmission or reflectance mode. The imager includes a camera and viewing system that magnifies the visible light image of the sample so that a position of interest can be isolated. The motorised stage, controlled by a joystick and the Spotlight software, enables the identification of points of interest on the sample and focus the imager.

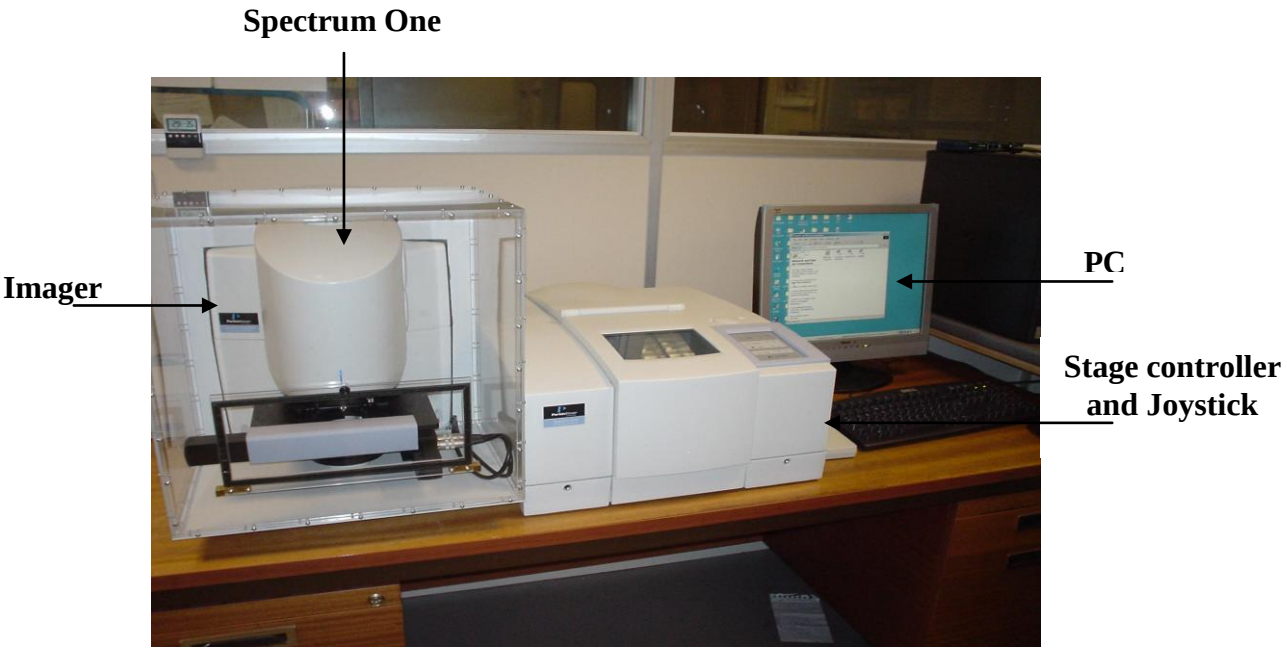


Figure 5-3: PerkinElmer Spectrum Spotlight FT-IR Imaging 300 System

5.4 Sampling techniques

The analysis of cell lines and blood specimens involved a combination of point mode measurements and IR image mode studies. The two main techniques generally employed for FT-IR spectroscopy are transmission and reflectance. In this study the transmission technique was applied, in which the IR radiation passes through the sample. This technique generates spectra of high signal-to-noise and can be used for solid and liquid samples. However, it is unfeasible to measure samples thicker than 20 µm as they display a no signal. Samples that are thinner than 1 µm are also difficult to measure due to weak absorbance signals.

Prior to any spectral measurements, a background spectrum must be obtained from a region of no sample. The background spectrum is used to ratio out the instrument's inbuilt response from sample spectra (Figure 5-4). The transmittance T can be calculated as:

$T = I/I_O$ Eq.5

where T is the transmittance, I_O is the intensity entering the sample, and I the intensity passed through the sample at each wavenumber.

The absorbance spectrum can be calculated from the transmittance spectrum using the following equation.

$A = -\log_{10} T$ Eq.6

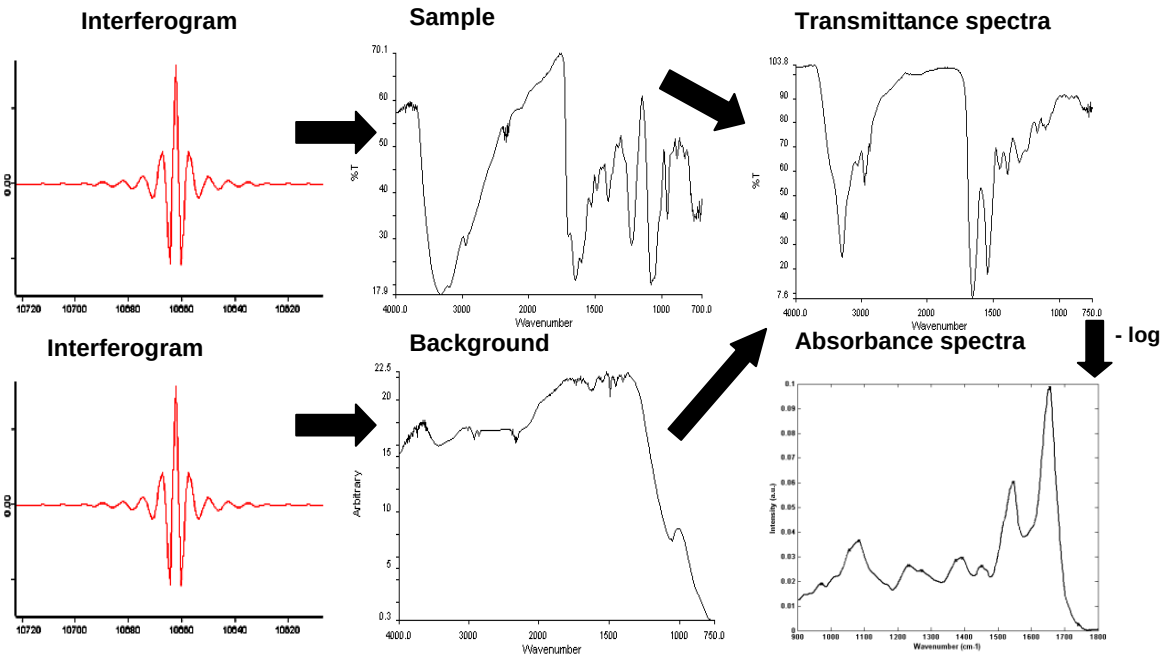


Figure 5-4: The FT-IR measuring sequence. The final absorbance spectra is truncated to highlight fingerprint region between 900-1800cm^{-1}

5.5 Infrared Microspectroscopy

In microspectroscopy, an IR microscope is coupled to infrared spectrometer. It permits both visual and infrared examination of different parts of the sample. The infrared microscope works in visible transmission, visual reflection, infrared transmission and infrared reflection. In this study cell line and blood samples mounted on the transmitting CaF_2 slide were examined in visible mode. Visual images were collected via a CDD camera under white light LED illumination and quilted together to produce an image of the entire sample deposition area. Various cellular regions were selected from this sample image for each cell line and blood sample.

Subsequent to visual examination using visible light, IR point spectra was then obtained from cell line and blood samples in transmission mode. Prior to acquiring IR point spectra for a sample background spectra was obtained using the same number of scans and spectral resolution. The background spectrum was ratioed against the sample spectrum.

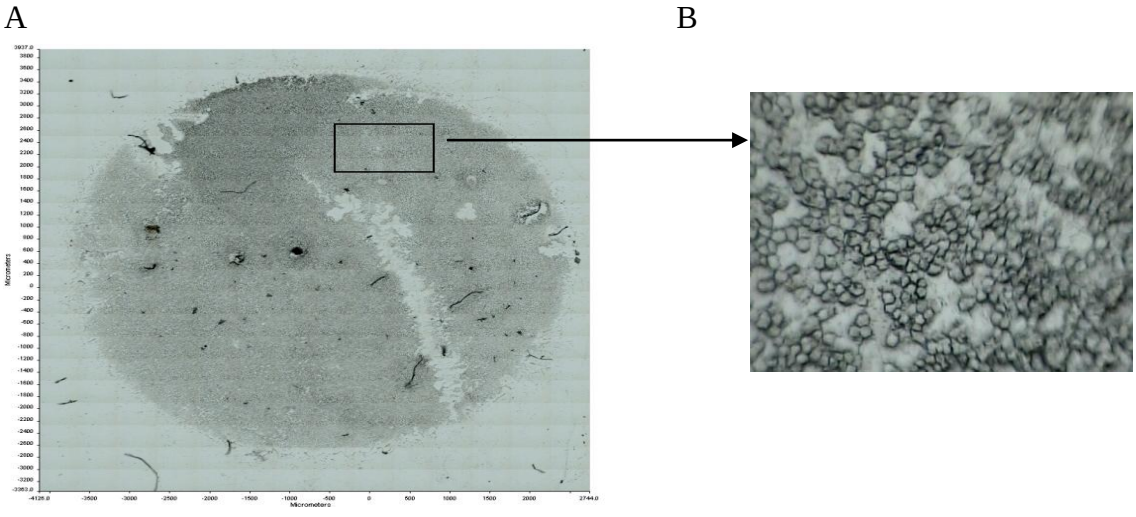


Figure 5-5: A) An example of visual light image obtained from a cell line sample and B) zoomed area of cell line sample to enable visualisation of individual cells

5.6 FT-IR Imaging

FT-IR spectral image data can be obtained using an FT-IR spectrometer that incorporates a step scan transform interferometer, a microscope and a multi element Focal Plane Array (FPA) detector. In this study, a 16 x 1 linear array which allowed 16 spectra to be measured simultaneously. The stage is then moved and 16 more pixels are measured and so on until all pixels are measured in the defined area. FT-IR imaging data are obtained in the format of a spectral hypercube as shown in Figure 5-6. This is a 4-dimensional data array consisting of two spatial dimensions, i.e. X-spatial dimension (pixels) and Y-spatial dimension (pixels) (Shaw 2000). For each pixel point an interferogram is acquired which is then Fourier transformed to obtain an absorption spectrum consisting of thousands of data points, depending on the resolution and size of the area probed. Infrared micro-spectroscopic imaging has previously been employed in various studies including; analysis of lymph node tissue (Romeo 2005); liver tissue (Diem 2000); benign breast tumour tissue (Fabian 2003) and aortic tissue (Bonnier 2006) as well as single cell analysis (Romeo 2006 and Lasch 2002).

High spectral (4 cm^{-1}) and spatial (6.25 µm) resolution was used in the study as detailed in Section 5.8.3. Resolution is defined as the ability to separate two adjacent features. In the case of spectral resolution, this refers to the specific wavelength intervals in the electromagnetic spectrum. Narrower bands with minimum intervals within FT-IR spectra along the x axis (wavenumber) require a higher spectral resolution. This is often difficult in overlapping peaks.

Spatial resolution is the ability to sharply and clearly define the features within an image. The selection of pixel size (6.25 or 25µm) is an important consideration. Decreasing the pixel size on an image increases the spatial resolution. Therefore using a 6.25µm is better than 25µm. Spatial resolution has been investigated for spectral imaging applications for FT-IR spectroscopy (Wood 2000; Collet 2004; Lasch 2006; Kuimova 2009).

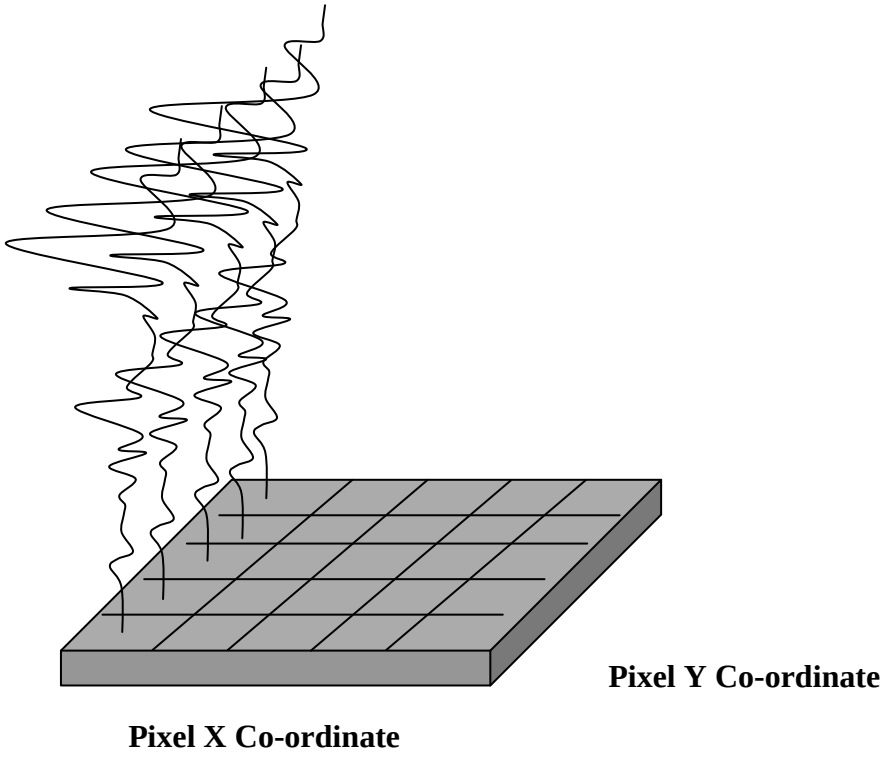


Figure 5-6: Spectral hypercube displaying intensity values at a given wavenumber. Raw spectral data in which the dimension for each pixel point is pixel coordinate X, the wavenumber axis and the pixel coordinate Y, the intensity axis. The intensity value of each spectrum is assigned a colour and displayed against the X and Y coordinates of the spectral element. The intensities at different spectral elements show variations in the biochemical composition for each pixel in the hypercube.

5.7 Software

Conversion of spectral hypercube into a useable format and conversion of raw FSM files into ASCII format was performed using Matlab version 7.4.0 (The MathWorks, USA) and the PLS Toolbox (Eigenvector, USA). All spectral pre-processing and data analysis was also performed in Matlab. In-house developed tools were used to pre-process the spectral data and conduct univariate and multivariate analysis.

5.8 Optimisation of spectral measurement parameters

5.8.1 Optimisation: Choice of aperture size

Optimum aperture size was selected by collecting 20 point spectra from pure glucose compound over the 720-4000 cm^{-1} spectral region. Figure 5-7 shows the averaged spectra obtained with decreasing aperture size from 50 x 50µm to 5 x 5µm. Aperture size at 5 x 5µm and 10 x10µm demonstrate a low signal to noise ratio. Increasing the size of aperture to 30 x 30µm and 50 x 50µm improves the signal to noise ratio. Aperture size of 30 x 30µm was selected for this study as an optimum parameter for collecting point mode spectra from all cell line samples and blood specimens. This aperture gives better quality spectra with the best signal to noise ratio.

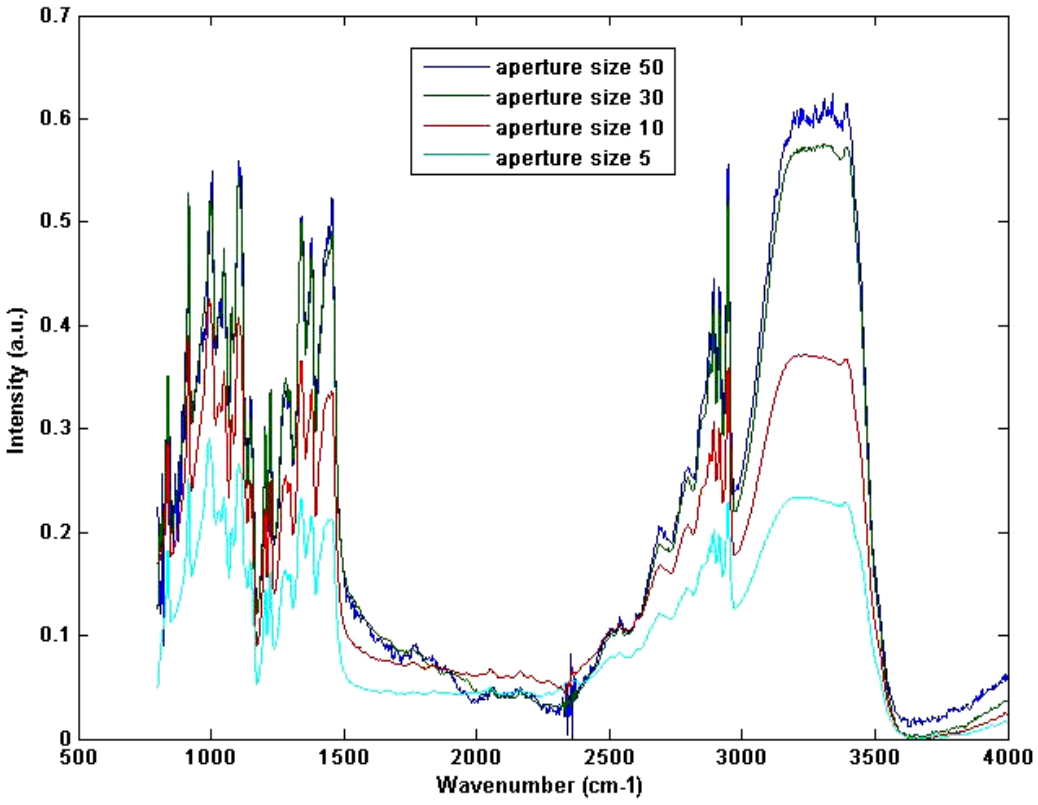


Figure 5-7: Effects of increasing aperture size on the signal-to-noise ratio of glucose spectral measurements

5.8.2 Optimisation: Choice of co-added scans

The optimum number of co-added scans was selected by collecting 20 point spectra from pure glucose compound over the 720-4000 cm^{-1} spectral region. Each spectrum was taken as an average of increasing number of co-added scans. Figure 5-8 shows the averaged spectra obtained with increasing the number of co-added scans from 16 to 128. Spectra taken as an average of 64 co-added scans was selected for this study as an optimum parameter for collecting point mode spectra from all cell line samples and blood specimens due to high signal to noise ratio.

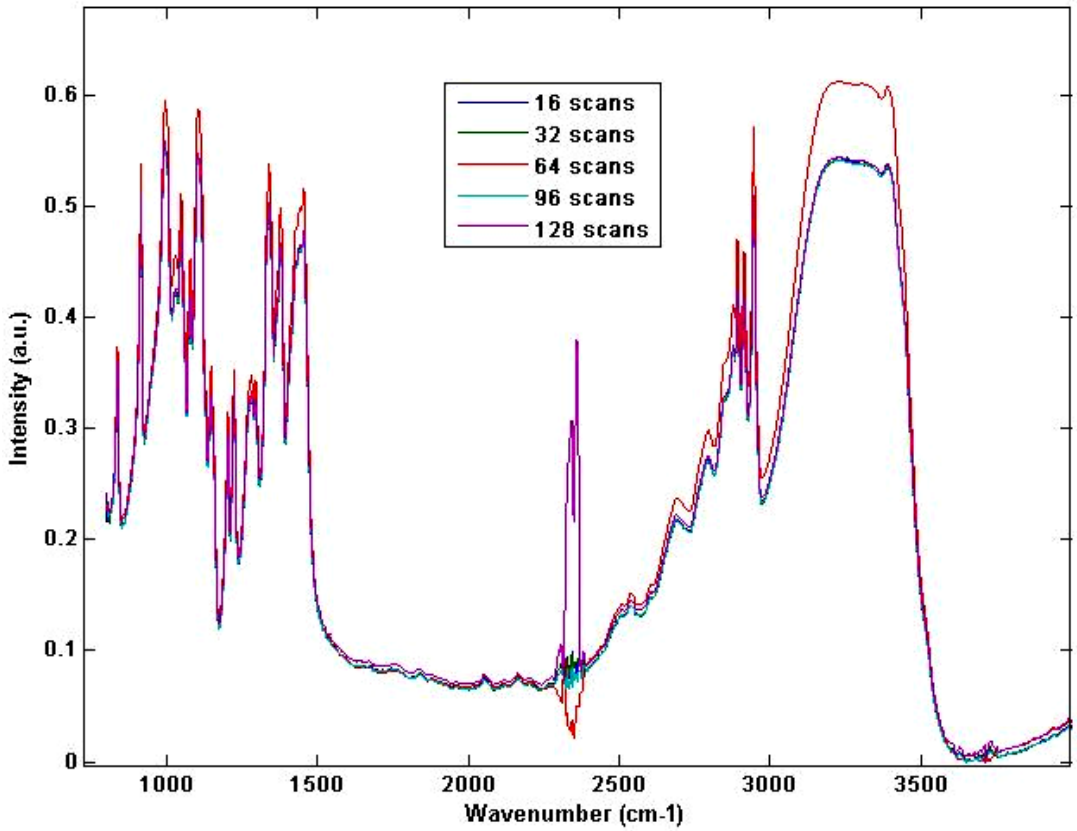


Figure 5-8: Effects of increasing co-added scans on the signal-to-noise ratio of glucose spectral measurements

5.8.3 Optimisation: Choice of spectral resolution

Spectral resolution refers to the width of the spectral band at each wavenumber. It determines the accuracy of separating different wavelength of radiation in the absorption spectrum. Applying spectral resolution ranging from 4 to 64 cm^{-1} has significant visual impact on the quality of spectra as demonstrated in Figure 5.9. However, it is important to use high spectral resolution in order to identify the precise position of narrow absorbance bands and retain all data points. Spectral resolution of 4 cm^{-1} was selected for this study as an optimum parameter for collecting point mode spectra from all cell line samples and blood specimens as it will identify overlapping peaks.

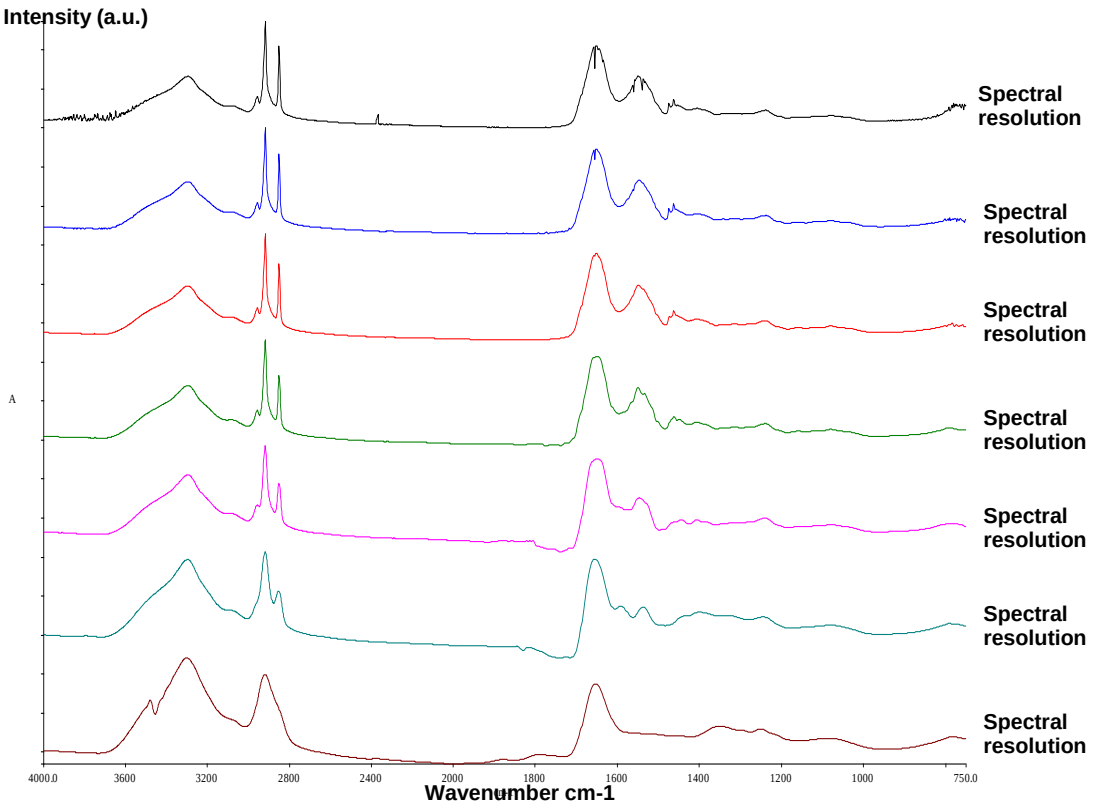


Figure 5-9: Effects of increasing co-added scans on the signal-to-noise ratio of glucose spectral measurements

5.9 Spectral Measurements

In this study, the type of spectral measurements obtained varies with the type of biological and clinical samples analysed. There are three main objectives outlined for investigation in this thesis as discussed in Section 1.3. The means of collecting and analysing spectral data vary amongst the three sections. Combinations of point and image mode spectral measurements were obtained. The parameters for acquiring spectroscopic data have been divided into 3 main sections for explanatory purposes:
Method 1: Classification of B-cell lymphoid leukaemia, myeloid leukaemia and T-cell lymphoma derived cell lines
Method 2: Optimisation of the PCA-fed-LDA classification model
Method 3: Classification of normal and CLL blood specimens.

5.9.1 Method 1: Classification of leukaemia and lymphoma derived cell lines

Spectral data was obtained for all leukaemia (RCH-ACV, REH, HL60 and Meg01) and lymphoma (Karpas) derived cell lines using both point mode and image mode measurements. Measurements of each cell type were repeated five times with subsequent batch sets (1-5) prepared at different times to test for reproducibility. In total 50 samples were measured, out of which 25 samples were frozen and the other 25 were air-dried samples.

Prior to point mode measurements of air-dried and frozen cell line samples, a background spectrum was obtained from a blank area of the CaF_2 slide and ratioed against the sample spectrum each time. The optimised point measurement parameters of 64 scans per pixel (spp), with a spectral resolution of 4 cm^{-1} using an aperture of 30 µm over the wavenumber range 1800 to 800 cm^{-1} were used to measure FT-IR spectra of leukaemia and lymphoma cell line samples. Point spectra obtained in transmission mode from cell line samples at different regions were averaged. Univariate analytical methods were used to compare the individual mean spectra representing leukaemia and lymphoma cell lines and comparing spectra from air-dried and frozen cell line samples.

This involved assignments of peaks to identify specific spectral features and calculating difference spectra.

IR imaging data was also obtained for both air-dried and frozen cell line samples in the format of a spectral hypercube as described in Section 5.6. Regions of high cellular content from the visual image map were selected for all cell line samples. The average size of the area probed was approximately 200-400µm. Image mode spectral data was collected by scanning the computer controlled microscope stage in an x-y direction in increments corresponding to the 6.25µm step size, using a spectral resolution of 4 cm^{-1} and 64 scans per pixel (spp) to increase the signal-to-noise ratio. Individual spectra from each pixel point were extracted from the spectral hypercube. Background scans were obtained from a region of no sample and rationed against the sample spectrum. Absorption Spectra extracted in the wavenumber range of 4000 to 720 cm^{-1} were converted to ASCII files using PS convert (Perkin Elmer). Similar imaging techniques have been employed by other groups in investigation of aortic tissue (Bonnier 2006), liver tissue (Yee 2000) as well as single cells (Lasch 2002)

Raw spectral data contained within the spectral hypercube were compared with published spectral assignments to provide information on the biochemical composition across the sample in visual form. Data analysis was carried out using Matlab. Principal components (PCs) were calculated and pseudo-colour PC score maps were generated to identify major biochemical differences. Three regions (approximately 20µm x 10µm) of high cellular content were selected from the chosen PC score image. The spectra from the selected regions of the PC score map were then combined and subjected to data pre-processing methods described in chapter 4. The pre-processed spectral dataset was used to construct a PCA-fed-LDA classification model and further unsupervised and supervised classification algorithms were used to discover spectral patterns and grouping spectra that share common features. The first 25 PC scores (describing the greatest variance in the dataset) were used to calculate linear discriminant functions that maximises the variance between the different groups and minimises the variance within each group. These are techniques are discussed in Sections 4.2.2.1 and 4.2.2.2 respectively. The prediction accuracy of the classification model was tested using cross

validation procedure described in Section 4.2.2.3. The proposed sequence for spectral measurements and data analysis is shown in Figure 5-10.

Image maps were collected for all five cell lines from five subsequent batches (25 air-dried and 25 frozen samples) to obtain a large spectral dataset for multivariate analysis as point spectra are not sufficient. The number and quality of spectra in the dataset have a considerable impact on the reliability of multivariate analysis. The performance of multivariate techniques improve with increased dataset size.

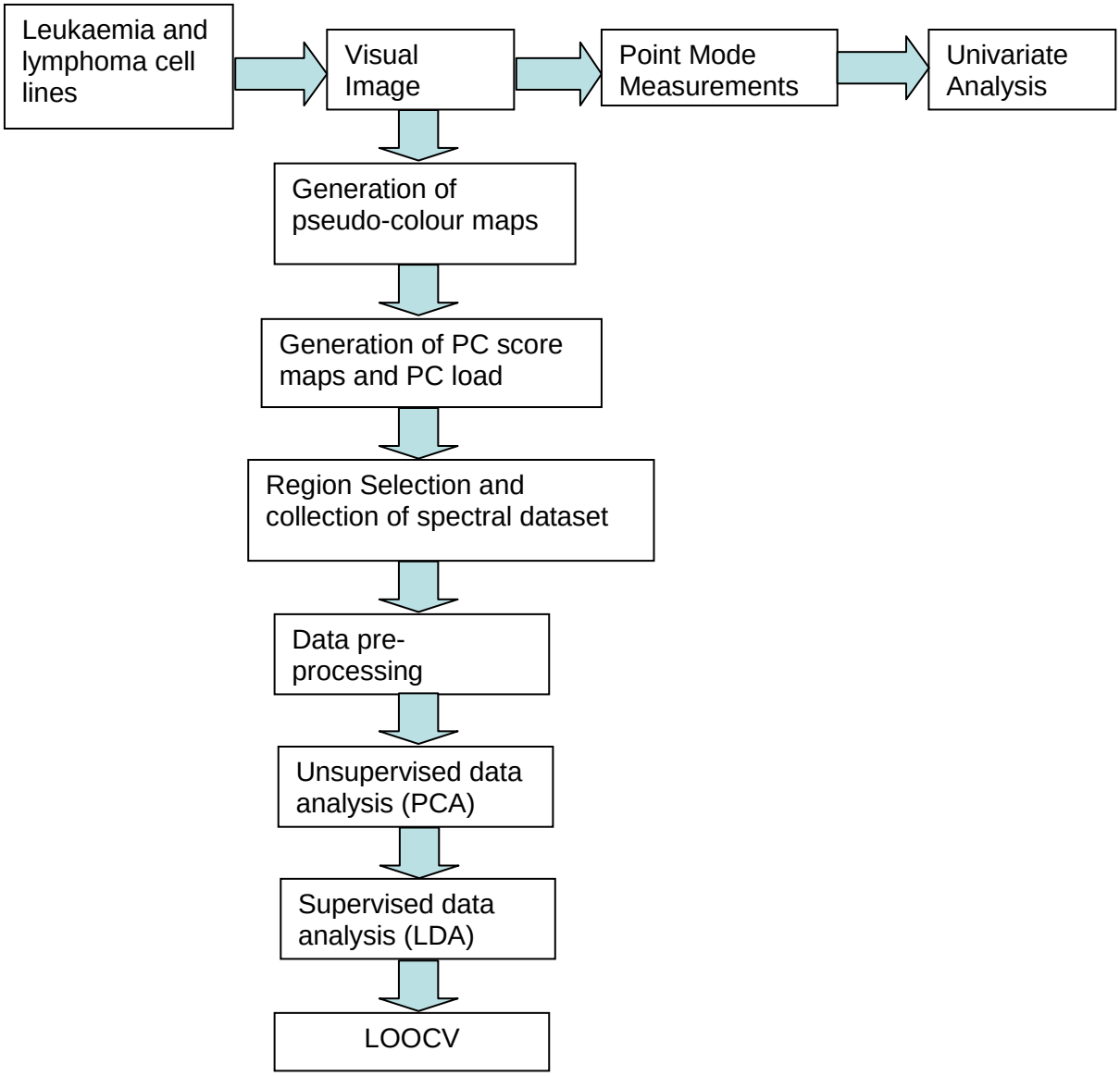


Figure 5-10: Overview of spectral measurements and data analysis process

5.9.2 Method 2: Optimisation of the PCA-fed-LDA classification model

Raw imaging spectral data collected for the five air-dried and frozen leukaemia and lymphoma cell lines was used to optimise the PCA-fed LDA classification model. Firstly, the effects of increasing the PCs from 5 to 25 input into the PCA-fed-LDA classification model were investigated. Savitsky-Golay derivative was tested as a pre-processing tool on the PCA-fed-LDA classification model to obtain the optimal performance. Spectral dataset was subjected to a range of 0^{th}, 1^{st} and 2^{nd} order Savitsky-Golay derivative calculated using 9-points. The application of standard deviations and absorption filter tool on the spectral dataset were also assessed. These were only used to filter out poor quality spectra from a dataset separately and not in combination.

Standard deviation filtering was only applied to cell line spectra obtained from the selection of different cellular regions on the PC score maps for all 5 cell samples from 5 subsequent batches. The combined spectral dataset was then normalised and smoothed. In this study a measure of 1.5 standard deviations as an optimal limit was used to identify any outliers. In this case, outliers are referred to as spectra that are more than 1.5 standard deviations either side of the mean of the dataset (Figure 5-11). Any outlier spectra beyond the standard deviation range were removed to improve the quality of spectral dataset and obtain a stable model. The remaining spectra were collated to create a new improved dataset.

The absorption filter tool was only applied to raw FSM (binary) files for all 5 cell samples from 5 subsequent batches. Prior to application of the absorption filter tool, the raw FSM file was imported into Matlab using a combination of in-house functions and a PerkinElmer data plugin script which converted and truncated (1800–900 cm^{-1}) the raw binary data into an ASCII format which can then be processed by Matlab. This loading process also converted the data from transmittance into absorbance values.

Prior to the application of the in-house absorbance spectral filter tool, the dataset was normalised by calculating the area under the curve. The in-house absorbance spectral filter tool was applied to the loaded hyperspectral data cube filters out weak or low

absorbance. All spectra outside the absorbance range of 0.2 (upper absorbance limit) and 0.15 (lower absorbance limits) are removed. In this study, this absorbance range was chosen to optimise the stability of the classification model.

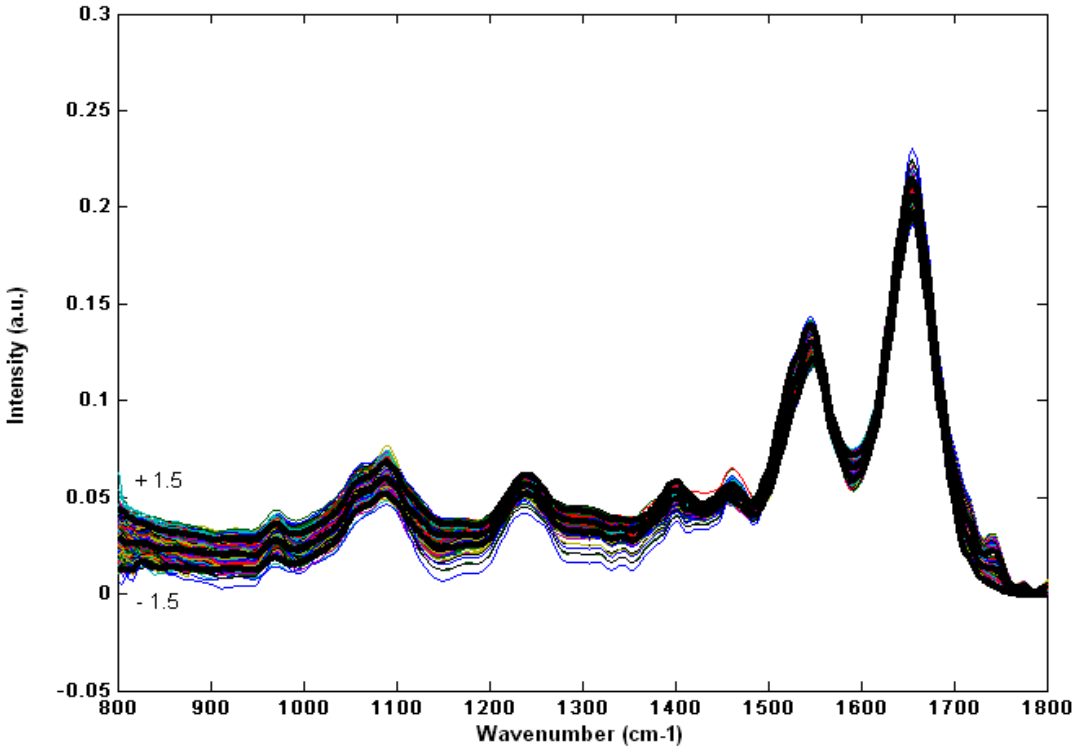


Figure 5-11: Spectral dataset with mean, + and – 1.5 standard deviation highlighted in the bold black lines

5.9.3 Method 3: Classification of normal and CLL blood specimens

Spectral measurements for normal and CLL blood specimens were acquired in point mode only from different regions of the sample. Prior to spectral measurements of blood components, a background spectrum were obtained from a blank area of the CaF_2 slide and ratioed against the sample spectrum each time. Spectra for EDTA and Histopaque were also obtained to assess any spectral contributions. The optimised standard point measurement parameters of 64 scans per pixel (spp), with a spectral resolution of 4 cm^{-1} using an aperture of 30 µm over the wavenumber range 1800 to 800 cm^{-1} were used to measure FT-IR spectra of the blood component samples as well as

EDTA and Histopaque. Figure 5-12 shows that there is relatively low contribution from EDTA and Histopaque in the overall spectra obtained from blood samples.

Univariate analytical methods were used to compare the individual mean spectra between:

- Normal whole blood, plasma and buffy coat samples
- CLL whole blood, plasma and buffy coat samples
- Normal and CLL whole blood, plasma and buffy coat samples

Univariate methods involved assignments of peaks to identify specific spectral features, calculating difference spectra and measurements of absolute peak intensity and peak intensity ratios.

Further analysis of spectra from different blood components was carried out by multivariate analysis using an in-house Matlab programme. Initially, PCA is used to generate data for linear discriminant analysis. Three PCA-fed-LDA classification models were constructed using 1) normal and CLL whole blood spectra, 2) normal and CLL plasma spectra and 3) normal and CLL buffy coat spectra. The performance of these classification models was tested using a cross validation procedure. PLS was performed on buffycoat spectra from normal and CLL blood samples. It was used to construct a linear predictive model for each cellular analyte amount based on the spectrum obtained from normal and CLL buffycoat samples.

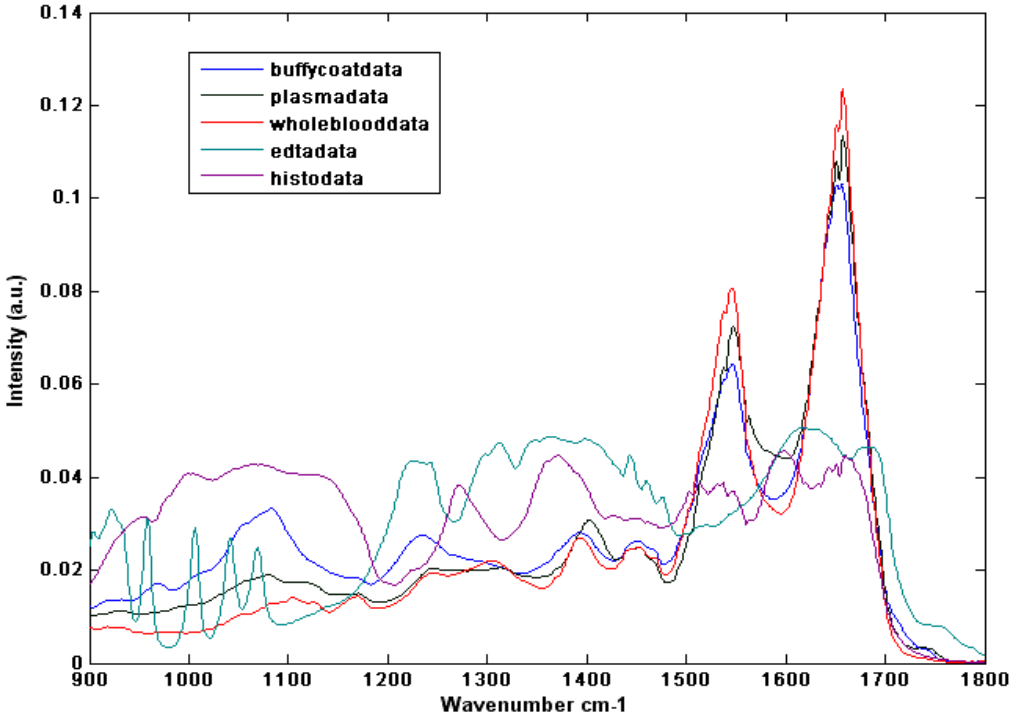


Figure 5-12: Spectra of blood components, EDTA and Histopaque

5.10 Summary

The section summaries the procedures involved in preparation of cell line samples and leukemic blood specimens from normal healthy volunteers and patients with chronic lymphocytic leukaemia (CLL). It also provides details on sampling techniques using FT-IR spectroscopy. Through various experimental measurments, the optimised parameters for spectral measurements used in this study are comprised of 64 scans per pixel (spp), with a spectral resolution of 4 cm^{-1} using an aperture of 30 µm over the wavenumber range 1800 to 800 cm^{-1} for both cell line and blood sample mesurements.

Chapter 6: Results and Discussion

This chapter has been compiled to outline the results of FT-IR spectral measurements from various biological and clinical samples and has been divided into 3 main sections. The results from a preliminary study, which reviews the spectra obtained from B-cell lymphoid leukaemia, myeloid leukaemia and T-cell lymphoma derived cell lines are discussed in the first Section (6.1). The effects of air-drying and freezing cell samples are also evaluated in this section. Section 6.1.2 determines the efficiency and sensitivity of the PCA-fed-LDA classification model to distinguish cell line groups using multivariate techniques. The second Section (6.2) evaluates methods of optimising the PCA-fed-LDA classification model for distinguishing different cell lines. The second phase of the study (Section 6.3) investigates the clinical application of FT-IR spectroscopy on blood specimens. Section 6.3.1 reviews the spectra obtained from blood samples from normal healthy subjects. Results obtained through detailed empirical analysis of the different components of normal blood including whole blood, plasma and buffy-coat are discussed in Sections 6.3.1.1 and 6.3.1.2. Section 6.3.2 reviews the spectra obtained from blood samples from patients with chronic lymphocytic leukaemia (CLL) with detailed empirical analysis of the different components of CLL blood in Sections 6.3.2.1-6.3.2.3. Detailed comparison of normal vs CLL spectra through empirical analysis are discussed in Section 6.3.3. The efficiency and sensitivity of the PCA-fed-LDA classification model to classify normal and CLL spectra using multivariate techniques are discussed in Section 6.3.4. Finally section briefly investigates the application of Partial Least Square (PLS) analysis to construct a prediction model to calculate the concentration of cellular analytes in buffy coat samples.

6.1 Empirical and Multivariate Analysis for Classification of leukaemia and lymphoma cell lines

6.1.1 Empirical analysis of leukaemia and lymphoma cell lines

Point spectra obtained in transmission mode from cell line samples at different regions were averaged using an in-house Matlab programme. The optimised standard point measurement parameters are described in Section 5.8 and 5.9.1. The mean spectra were plotted as intensity (arbitrary units) against wavenumber (cm^{-1}). A plot of the mean FT-IR spectra of the five cell line groups is shown in Figure 6-1 for comparison. The spectra have a degree of similarity. However, subtle differences between the spectra from the different cell line groups are observed due to slight variations in peak positions and intensities. The peak positions of the major vibrational modes are labelled. The full list of peak assignments are summarised in Appendix D.

Despite the minor spectral differences, all five cell types were found to display prominent characteristic spectral features. The most distinctive peaks in the IR absorption spectra are assigned to the absorption modes of proteins: the most intense is the amide I band, centred near 1655 cm^{-1}, which corresponds to the C=O stretching coupled to the N-H bending and the C-H stretching modes of peptide bonds (Cecilia 2008 and Ami 2003). It has been reported previously that the shape of the amide I band is influenced by the overall secondary structure of cellular proteins (Liu 2001). In particular, the peaks assigned to α-helices, β-sheets, β-turns and random coils have absorption maxima between 1645-1662 cm^{-1}, 1613-1637 cm^{-1}, 1662-1682 cm^{-1} and 1637-1645 cm^{-1} respectively. The amide II band at 1549 cm^{-1} arises from the vibrational modes that involve the C-N-H bending and C-N stretching of peptidic bonds (Gasparri 2003). The amide III at 1242 cm^{-1} is contributed by proteins arising from C-N stretching and N-H bending (Liu 1997 and Gasparri 2003).

The peaks at 1400-1460 cm^{-1} are attributed to CH_2 and CH_3 deformation vibrations primarily due to lipid contribution. Nucleic acid signals contribute to the infrared spectra at 1091 cm^{-1} (sym PO_2^-), 1065 cm^{-1} (PO_2^- and ribose groups) and 972 cm^{-1}

(phosphodiester group present in the DNA and RNA) (Ami 2003). Lipids have long hydrocarbon chains and therefore have their peaks at 1352 cm^{-1} (CH_3 vibration). The peak at 1740 cm^{-1} (C=O bending) corresponds to ester carbonyl (Bonnier 2006), although this peak at 1787 cm^{-1} emerges somewhat as a shoulder (Figure 6-1). Phosphate bands of phospholipids are also found between 1080-1230 cm^{-1}.

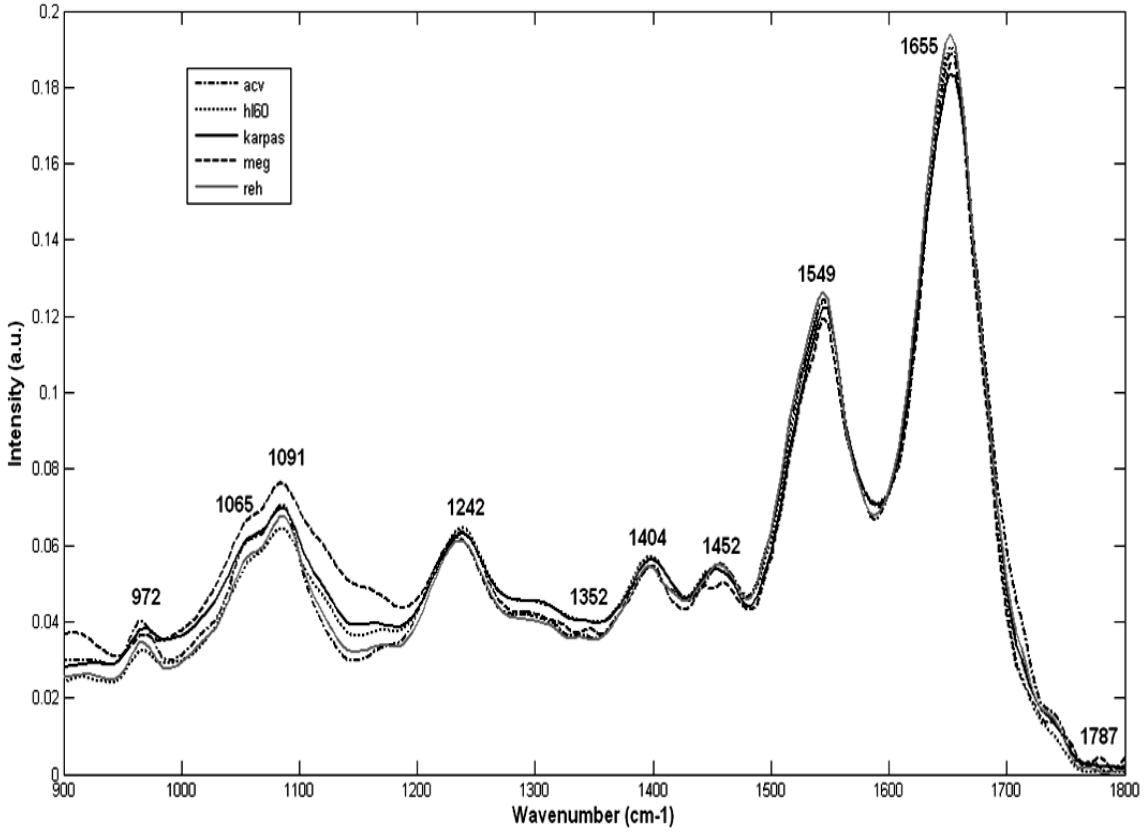


Figure 6-1: Mean FT-IR spectra for B-cell lymphoid (REH and RCH-ACV), myeloid leukaemia (HL60 and Meg01) and T-cell lymphoma (Karpas 299) in the range 800-1800cm^{-1}

6.1.1.1 Difference Spectra

To further emphasise and extract the spectral variations between the cell line groups, difference spectra, were calculated by subtracting the mean spectrum from one cell line group from another as described in Section 4.2.1.3. The difference spectrum was calculated using an in-house Matlab programme. In order to calculate difference spectra, all spectra belonging to REH and RCH-ACV cell line groups were combined and the mean was calculated to obtain a mean representative spectrum for the B-cell lymphoid cell line group. This was done subsequently for all spectra belonging to HL60 and Meg01 cell line group to obtain a mean representative spectrum for the myeloid cell line group. Mean representative spectrum for T-cell lymphoma (Karpas 299) was also calculated. A plot of the mean FT-IR spectra of the three cell line groups normalised to the intensity at 1655 cm^{-1} is shown in Figure 6-2 for comparison.

Figures 6-3 – 6-5 shows the resulting spectra obtained from taking the difference between mean normalised spectra from each cell line group in turn. The positive peaks in the difference spectra plot with an intensity greater than zero represent biochemical constituents in abundance in that specific cell line group. In contrast, negative peaks represent a lower concentration of that biochemical constituent comparative to that present in the spectrum of the subtracted cell line group. Variations in the concentration of biochemical constituents between the lymphoid, myeloid and lymphoma are seen throughout the fingerprint region (Figure 6-1). The complicated band patterns in the difference spectra as shown in Figures 6-3 – 6-5 are composed of peaks arising from changes of vibrational modes, to changes of the protein structures (Amide I, II and III), and to changes of specific nucleic acids, amino acid side chains and fatty acids. Table 6-1 lists the biochemical constituents in high concentrations as positive peaks in certain cell line groups and those with lower concentrations as negative peaks in other cell line groups.

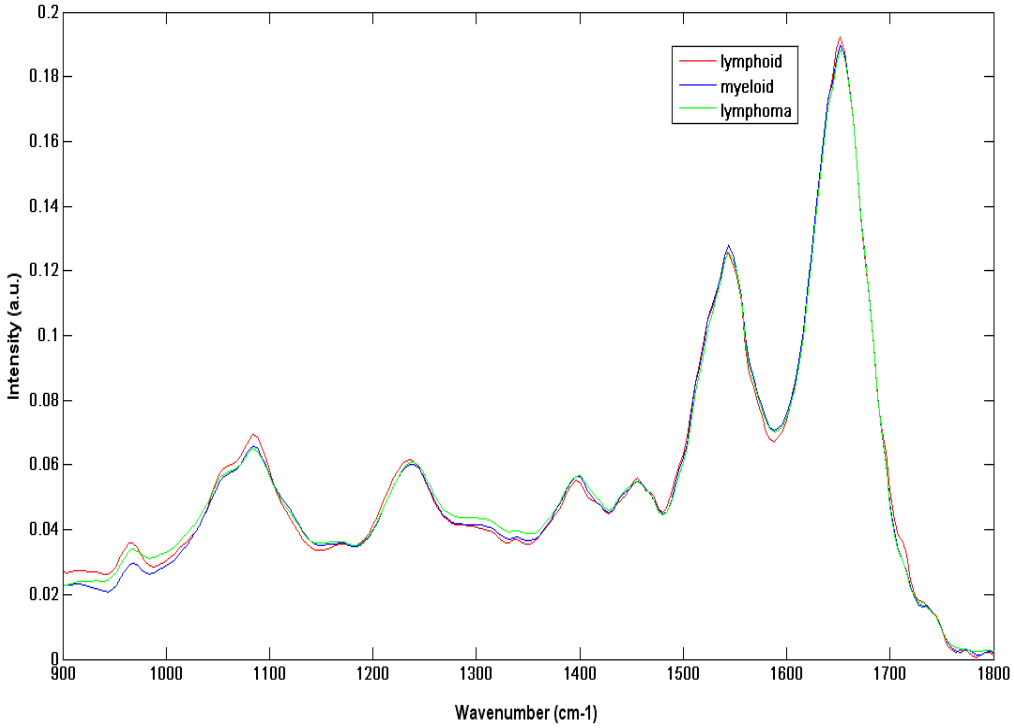


Figure 6-2: Mean representative spectrum for T-cell lymphoma (Karpas 299), B-cell lymphoid (RCH-ACV and REH) and myeloid (HL60 and Meg01) cell lines

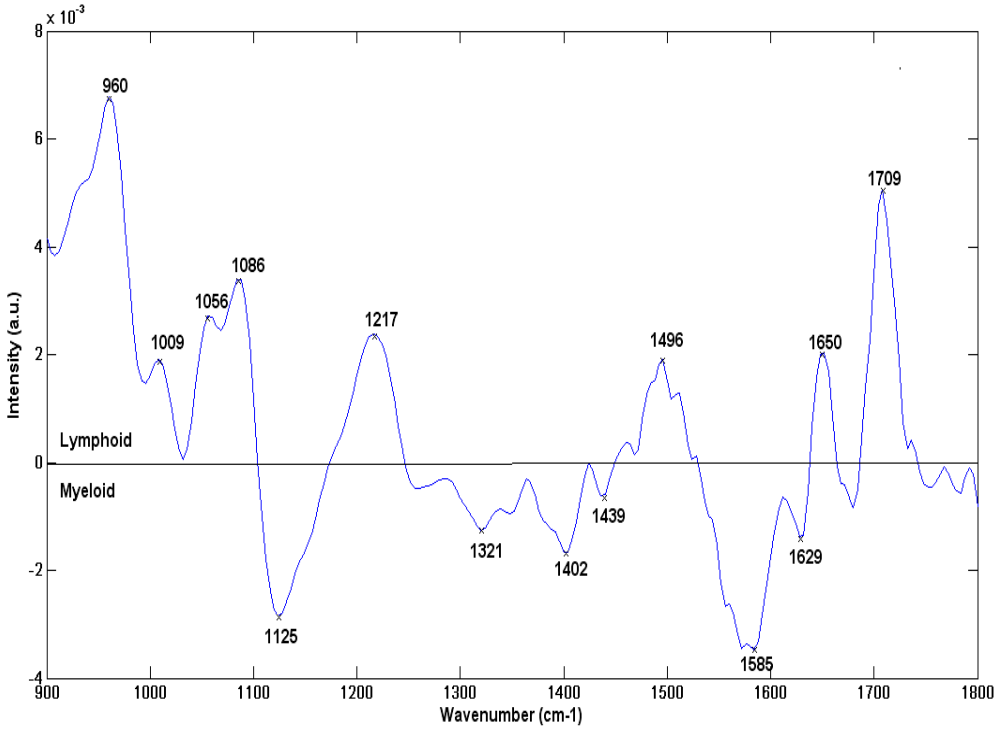


Figure 6-3: Difference spectra obtained by subtraction of B-cell lymphoid cell line spectrum from the spectrum obtained from myeloid cell line

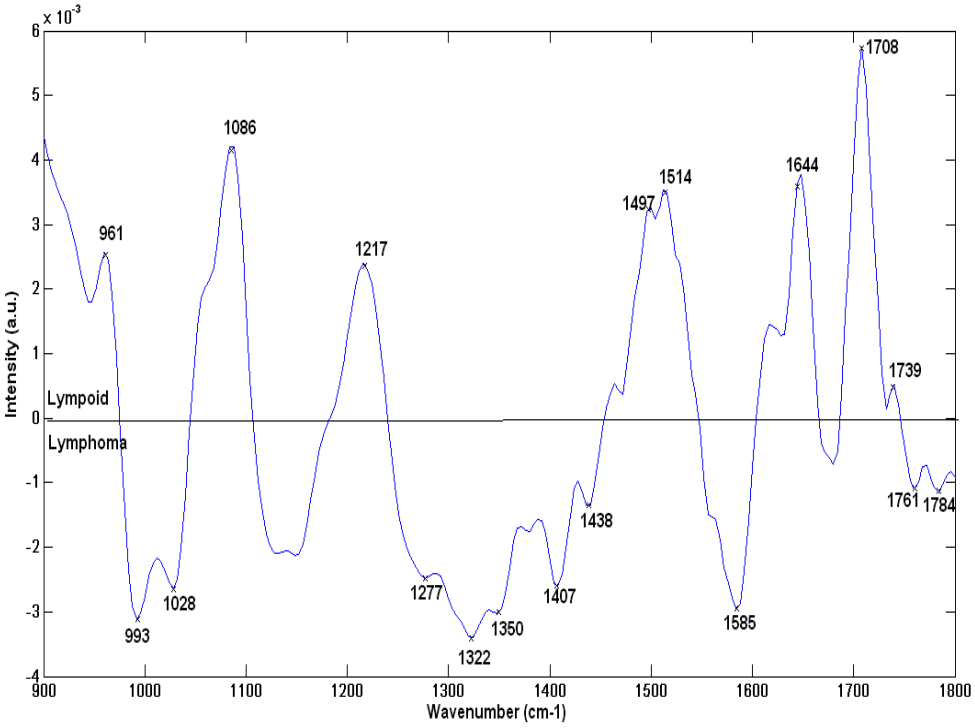


Figure 6-4: Difference spectra obtained by subtraction of B-cell lymphoid cell line spectrum from the spectrum obtained from T-cell lymphoma cell line

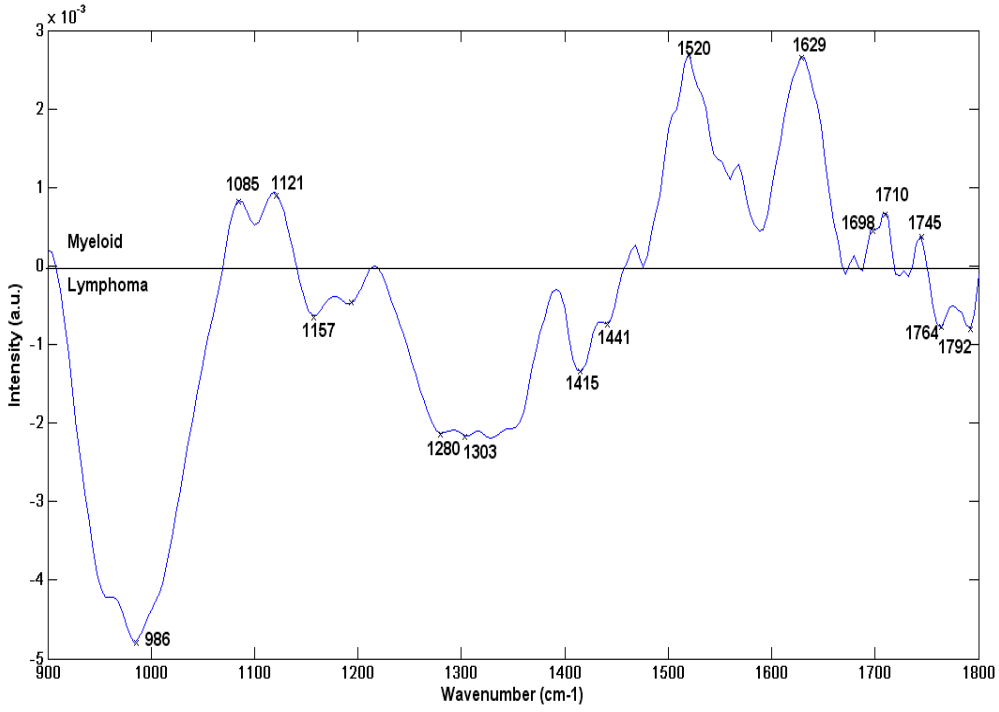


Figure 6.5: Difference spectra obtained by subtraction of myeloid cell line spectrum from the spectrum obtained from T-cell lymphoma cell line

	Lymphoid - Myeloid		Lymphoid - Lymphoma		Myeloid - Lymphoma	
	Peak	**Assignment**	**Peak**	**Assignment**	**Peak**	**Assignment**
Positive Peaks	**960** **1009** **1056** **1086** **1217** **1496** **1650** **1709**	DNA Nucleic acids Nucleic acids, -CO-O-C str P=O str of $>PO_2^-$ PO_2^- str CH_2 bending Amide I C=O in nucleic acids	**961** **1086** **1217** **1497** **1514** **1644** **1708** **1738**	DNA P=O str of $>PO_2^-$ PO_2^- str CH_2 bending Tyrosine Amide I C=O in nucleic acids Lipid, amino acid side chains, C=O str in esters	**1085** **1121** **1520** **1629-1698** **1710** **1745**	P=O str of $>PO_2^-$ Carbohydrates, C-O, C-OH str Tyrosine Amide I C=O in nucleic acids Lipid, amino acid side chains, C=O str in esters
Negative Peaks	**1125** **1321** **1402** **1439** **1585** **1629**	Carbohydrates, C-O, C-OH str Amino acid, CH vib Lipid, CH_2 wagging vib, C-O str in carbohydrates Lipid (CH_3 sym bending vib) Amide II, NH vib, CN str NH_3^+, def vib Amide I, C=O str	**993** **1028** **1277** **1322-1350** **1407** **1438** **1585** **1761-1784**	Phosphorylated proteins, RNA C-OH carbohydrates, Amide III Amino acids, COO^- Lipid (CH_3 sym bending vib) Lipid (CH_3 sym bending vib) Amide II, NH vib, CN str Liids, esters, C=O str	**986** **1157** **1280** **1415-1441** **1767-1790**	RNA Glycogen Amide III Fatty acids, phospholipids and triglycerides Lipids, esters, C=O str

Table 6-1: Lists positive and negative peaks obtained from the difference spectra between B-cell lymphoid, myeloid and T-cell lymphoma cell lines with peak assignment

6.1.1.2 FT-IR spectral comparison of air-dried and frozen cell lines

The optimised standard point measurement parameters (described in Section 5.8) were used to collect 20 FT-IR spectra each from both air-dried and frozen cell line samples at different sites. Figures 6-6 illustrate the comparison of mean air-dried and frozen spectra taken as an average of 20 different measurements at various sites of the leukaemia (RCH-ACV, REH, HL60 and Meg01) and lymphoma (Karpas 299s) cell lines sample.

Table 6-2 compares the effects of air-drying and freezing cell line samples on the intensity of major peaks in the fingerprint region. Subtle changes are observed in peaks corresponding to DNA and lipids; 972 cm^{-1}, 1092 cm^{-1}, 1240 cm^{-1}, 1400 $cm^{-1}$and 1456 cm^{-1}. The intensity of these peaks generally tends to decline in frozen spectra compared to that of air-dried spectra. Significant changes in intensity are observed in the major protein peaks at 1548 cm^{-1} and 1655 cm^{-1} corresponding to amide II and amide I proteins. The intensity of these peaks generally tends to intensify in the spectra belonging to frozen samples compared to the spectra belonging to the air-dried samples. A recent study coincides with these finding and reveals that proteins (amide I and II) are relative stable during freezing with an abrupt decrease in α-helical structures. Freezing tend to have greater effect on the behaviour of liposomes and biological membrane lipids of cells by changing the hydrophobic and hydrophilic interactions (Wolkers 2007). Study by Mourant (2003), also observed a reduced intensity of the Amide II peak and a change in position for the spectra of dried M1 fibroblast cells. Alterations in the secondary structure of proteins are suggested to cause this change in position when the cells are dried, introducing artefacts. Changes in amplitude and shifts in absorption bands corresponding to proteins and nucleic acids as a result of hydration (air-drying) have also been reported due to changes in conformation (Pevsner 2003). Major changes in the peak intensity of amide I and amide II bands and the shape of the amide III band have been noted.

Table 6-3 further support these results, as area under the curve for the major peaks enlarge in all frozen spectra compared to that of air-dried spectra. These findings

suggest that freezing cell line samples preserves cellular composition and reduces rate of cellular damage. As a result, stable and high quality spectra can be obtained.

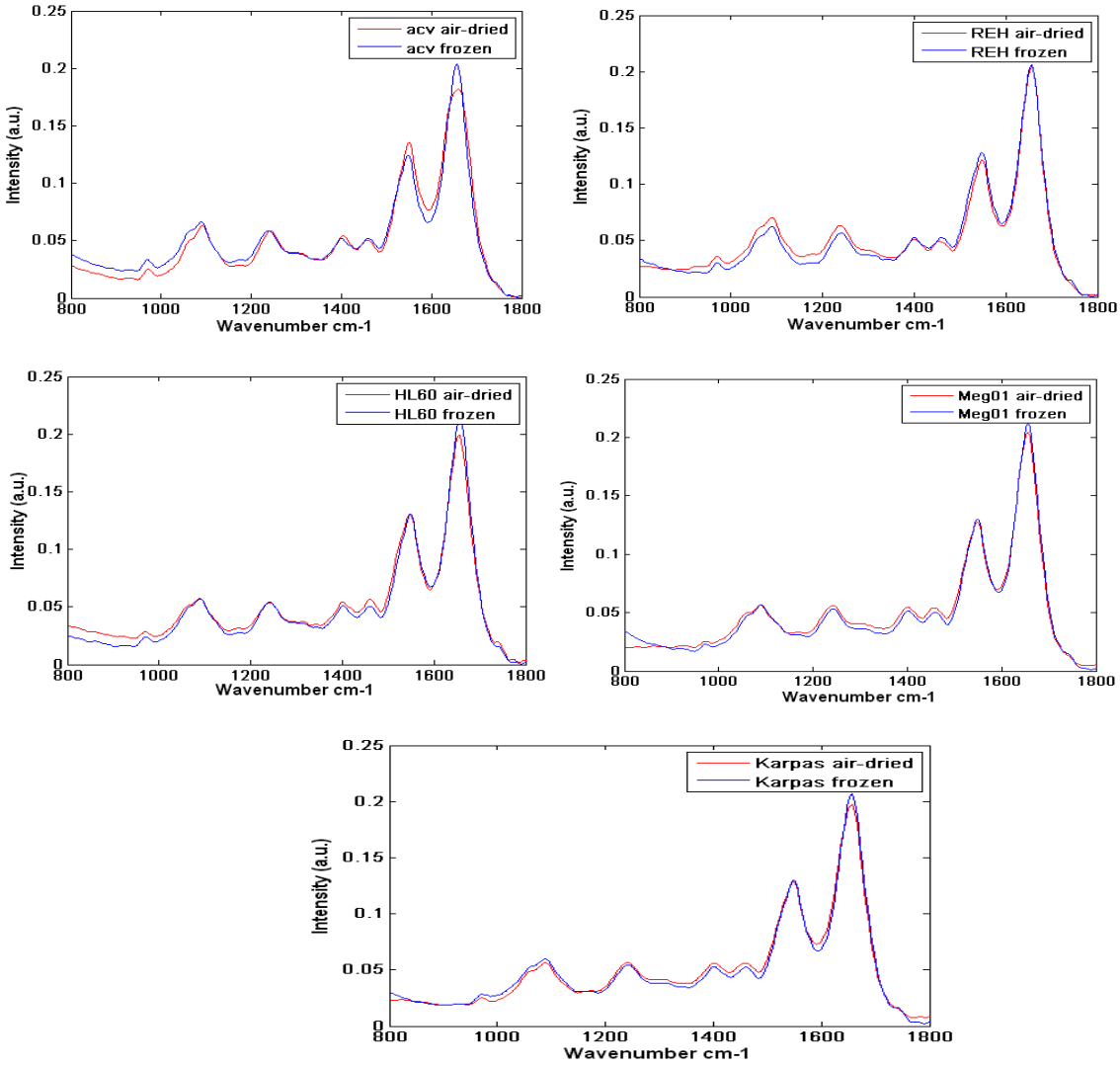


Figure 6-6: Comparison of mean air-dried and frozen cell line spectra

	Intensity of paticular peaks (cm-1)						
	972	**1092**	**1240**	**1400**	**1456**	**1548**	**1655**
Mean ACV dry spectra	0.024	0.061	0.059	0.054	0.05	0.123	0.181
Mean ACV frozen spectra	0.032	0.065	0.059	0.051	0.05	0.133	0.203
Mean REH dry spectra	0.035	0.069	0.062	0.05	0.048	0.12	0.203
Mean REH frozen spectra	0.029	0.062	0.056	0.052	0.052	0.128	0.206
Mean HL60 dry spectra	0.027	0.056	0.053	0.054	0.056	0.129	0.198
Mean HL60 frozen spectra	0.023	0.056	0.053	0.05	0.05	0.129	0.213
Mean Meg01 dry spectra	0.025	0.055	0.056	0.054	0.053	0.126	0.203
Mean Meg01 frozen spectra	0.023	0.055	0.052	0.051	0.05	0.129	0.211
Mean Karpas dry spectra	0.024	0.056	0.056	0.055	0.055	0.128	0.196
Mean Karpas frozen spectra	0.027	0.059	0.053	0.053	0.052	0.128	0.206

Table 6-2: Comparison of peak intensities between air-dried and frozen spectra for B-cell lymphoid (RCH-ACV and REH), myeloid leukaemia (HL60 and Meg01) and T-cell lymphoma (Karpas 299) cell lines

	Area under the curve						
	972	**1092**	**1240**	**1400**	**1456**	**1548**	**1655**
Mean ACV dry spectra	0.21	3.02	1.14	0.51	0.28	3.96	11
Mean ACV frozen spectra	0.69	1.37	4.3	2.23	1.19	15.4	40.1
Mean REH dry spectra	0.16	2.49	1.08	0.43	0.22	3.39	9.47
Mean REH frozen spectra	0.19	2.98	1.25	0.51	0.27	3.72	10.5
Mean HL60 dry spectra	0.07	1.84	0.73	0.4	0.25	3.35	8
Mean HL60 frozen spectra	0.12	2.99	1.29	0.63	0.35	4.78	12.5
Mean Meg01 dry spectra	0.07	2.74	1.2	0.59	0.35	4.34	12.1
Mean Meg01 frozen spectra	0.13	4.08	1.68	0.91	0.55	7.02	18.7
Mean Karpas dry spectra	0.12	2.71	1.07	0.55	0.32	4.09	10.6
Mean Karpas frozen spectra	0.13	2.68	1.04	0.55	0.3	4.27	10.9

Table 6-3: Comparison of area under the curve between air-dried and frozen spectra for B-cell lymphoid (RCH-ACV and REH), myeloid leukaemia (HL60 and Meg01) and T-cell lymphoma (Karpas 299) cell lines

6.1.2 Multivariate Analysis

To further identify spectral variations and develop a classification model to distinguish cell line groups with a view to diagnostic spectral analysis, multivariate analysis in form of principal component analysis (PCA), linear discriminant analysis (LDA) and leave-one-out cross validation (LOOCV) was employed. PCA and LDA are described in Sections 4.2.2.1 and 4.2.2.2 respectively. LOOCV is explained in Section 4.2.2.3. All multivariate analysis was performed using in-house Matlab scripts.

Many different cell line datasets were constructed in order to obtain optimal datasets and have been presented in the thesis. The original dataset consists of both dry and frozen spectra from five different cell line groups taken from five subsequent batches, with a total of 50 samples. The five cell line group datasets consists of individual spectral datasets for RCH-ACV, REH, HL60, Meg01 and Karpas 299. Dry and frozen spectra obtained from 50 cell line maps were further combined into three clinically significant subsets to obtain the second, three cell line group dataset. Spectra belonging to REH and RCH-ACV cell lines were combined to obtain a B-cell lymphoid group. This was done subsequently for all spectra belonging to HL60 and Meg01 cell lines to obtain a myeloid group. All Karpas 299 spectra represented the T-cell lymphoma group. For explanatory purposes, the five cell line model is defined as dataset 1a and the three cell line model as dataset 2a throughout the subsequent results and discussion Sections. These datasets were further sub-divided into either air-dried dataset or frozen dataset.

- Five cell line group dataset comprised of only frozen spectra with 25 samples in and 9929 spectra total. This dataset is defined as spectral dataset 1c.
- Three cell line group dataset comprised of both air-dried and frozen spectra with 50 samples and 19690 spectra in total. This dataset is defined as spectral dataset 2a.
- Three cell line group dataset comprised of only air-dried spectra with 25 samples in and 9929 spectra total. This dataset is defined as spectral dataset 2b.
- Three cell line group dataset comprised of only frozen spectra with 25 samples in and 9929 spectra total. This dataset is defined as spectral dataset 2c.

6.1.2.1 Principal component analysis

This initial section evaluates the results of PCA, LDA and LOOCV on the following datasets:

- Five cell line group dataset comprised of both air-dried and frozen spectra with 50 samples and 19690 spectra in total. For explanatory purposes, this dataset is defined as spectral dataset 1a throughout the subsequent result sections.
- Five cell line group dataset comprised of only air-dried spectra with 25 samples and 9929 spectra in total. This dataset is defined as spectral dataset 1b.

White light images as well as their corresponding pseudo-colour maps, of total absorbance at each point, in transmission mode were obtained from 50 dry and frozen cell line samples (dataset 1a and 2a). From the visual image, an area of high cellular content was selected for spectral imaging. The average size of the area probed was approximately 200-400µm for each cell line sample. Pseudo-colour map show absorbencies of varying intensities, where the blue indicates low and red indicates high absorbance values. For explanatory and visual purposes, Meg01 frozen cell line obtained from batch 5 is used as an example. The white light image and pseudo-colour map for Meg01 cell line is shown in Figure 6-7.

A B

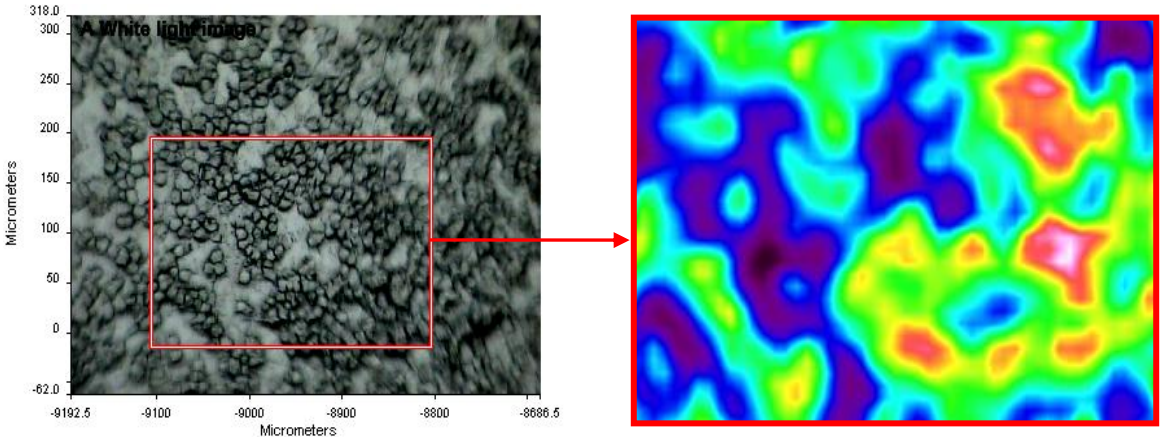


Figure 6-7: A) White light image of Meg01 cell line, B) pseudo-colour map of frozen Meg01 cell line (batch 5) in transmission mode. Red box highlights the region of interest

To establish biochemical changes in the spectral dataset across the pseudo-colour map, PCA was employed. PCA reduces a complete spectral dataset into a few spectral components and reveals maximal biochemical variance over the whole spectral dataset. As a result of PCA, 25 PC scores and PC loads are produced in order of variance. Figures 6-8 and 6-9 show the PC scores and loads of the first 10 principal components for the frozen Meg01 cell line sample. PC loads are the spectral components and PC scores relates to the abundance of that spectral component in the sample. The PCs were used as the input into a linear discriminant analysis (LDA) model as described in Section 6.1.24. Analysis of variance (ANOVA) was performed on the PCs to identify the most significant PCs, which were then used to develop LDA models. Leave-one-out cross validation (LOOCV) procedure was applied to indicate the accuracy of the predictions of the LDA classification model as described in Section 6.1.2.5.

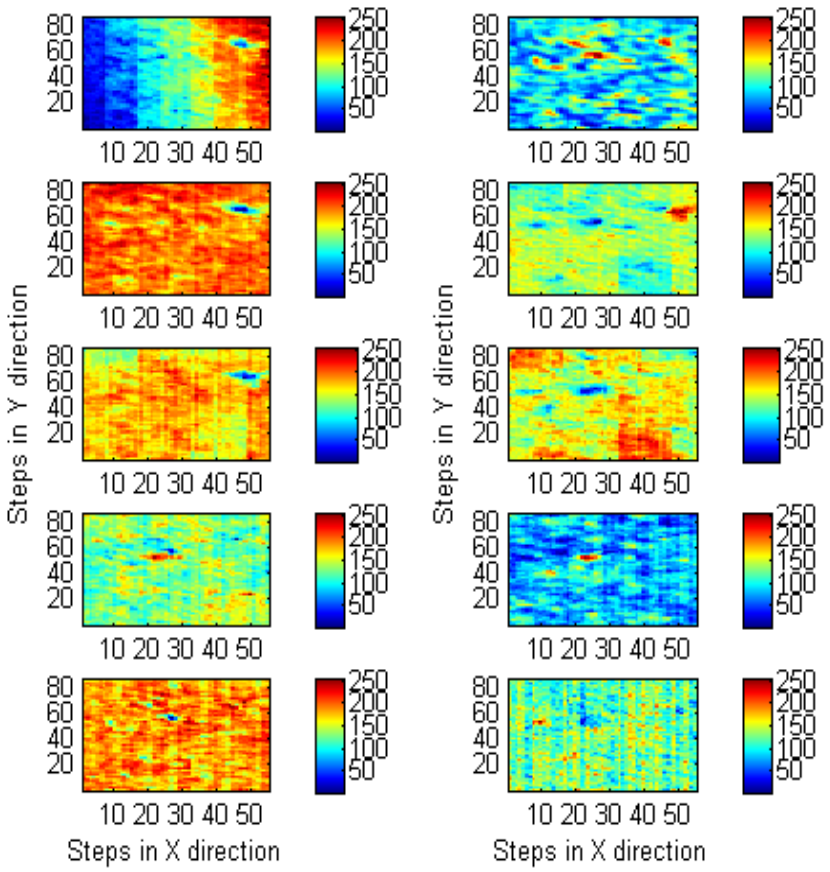


Figure 6-8: PC scores of the first 10 PCs for the Meg01 cell line (batch 5). Top left (1) to bottom right (10)

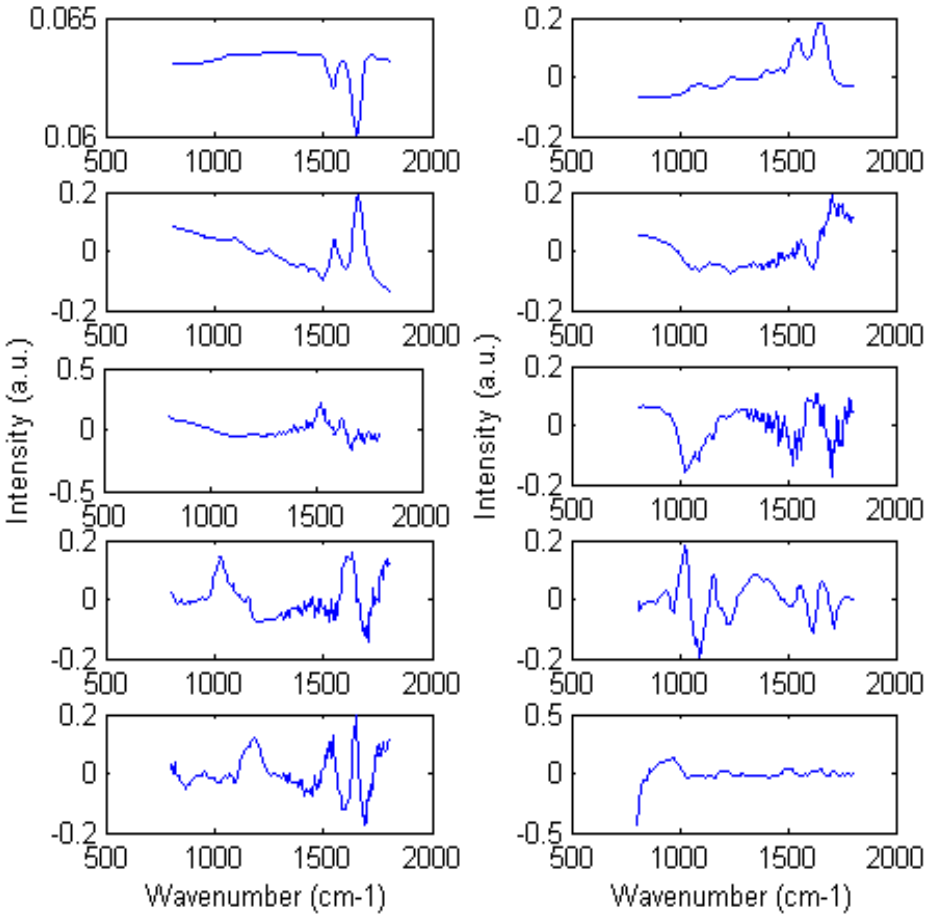


Figure 6-9: PC loads of the first 10 PCs. Top left (1) to bottom right (10)

The first few PC loads account for most of the spectral variance. Noise is introduced and more subtle variations are observed as the number of PC loads increase. Peaks within the same PC loads can be represented as positive or negative. Therefore, labelling PC loads can prove to be difficult. Major peaks have been identified, using assignments from the literature as listed in Appendix D to determine the key changes in the biochemical constituents across the sample.

Positive peaks indicate high contribution of a particular biochemical constituent when the PC score is high in a spectrum. In contrast, negative peaks indicate low contribution of a particular biochemical constituent when the PC score is high in a spectrum. PC 1 is shown inverted, indicating high levels of proteins, with protein peaks assigned at 1654 and 1549 cm^{-1} (Figure 6-10). PC2 and PC3 also describe protein features, which account

for the highest variance in the dataset (Figure 6-11 and 6-12). Two sharp positive peaks at 1661 cm^{-1} and 1552 cm^{-1} assigned to Amide I and Amide II proteins are prominent in both PC2 and PC3. PC8 has positive peaks that are common to carbohydrates at 1025 and 1159 cm^{-1} along with contribution from Amide II at 1556 cm^{-1} (Figure 6-13). PC 8 has negative peaks at 969, 1072, 1096, 1231, 1617 and 1715 cm^{-1}. Peaks at 969, 1072 and 1093 are contributed by nucleic acids (P=O stretch of $>PO_2^-$). 1715 cm^{-1} is also close to assignment for nucleic acids (C=O stretch). Negative peaks at 1231 cm^{-1} is assigned for Amide III protein. The other PCs are statistically less significant as they account for less of the variance in the data and therefore are not shown.

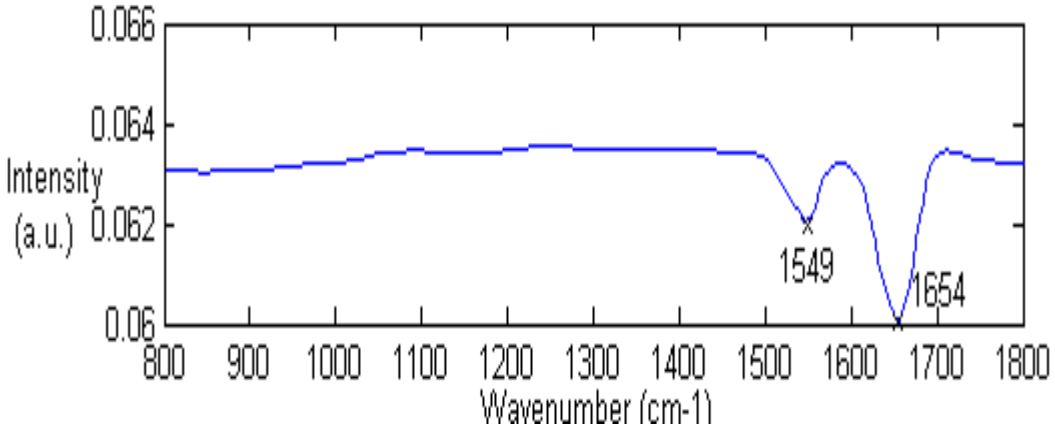


Figure 6-10: Principal component 1

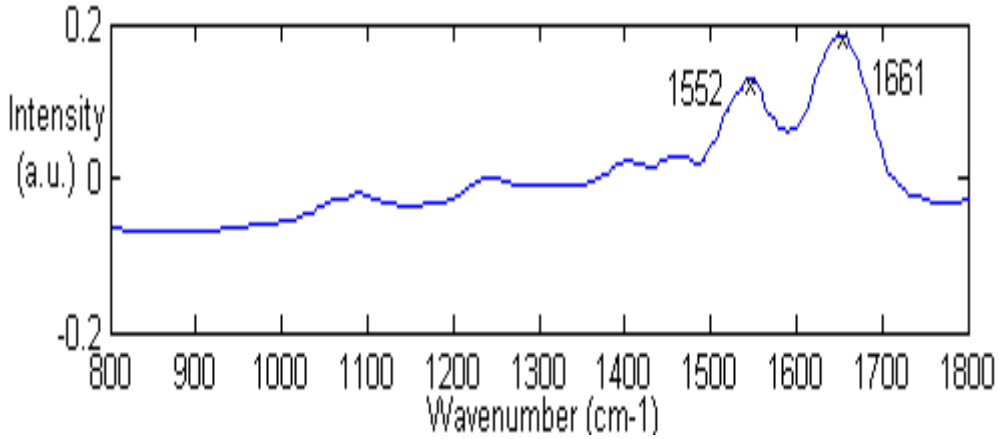


Figure 6-11: Principal component 2

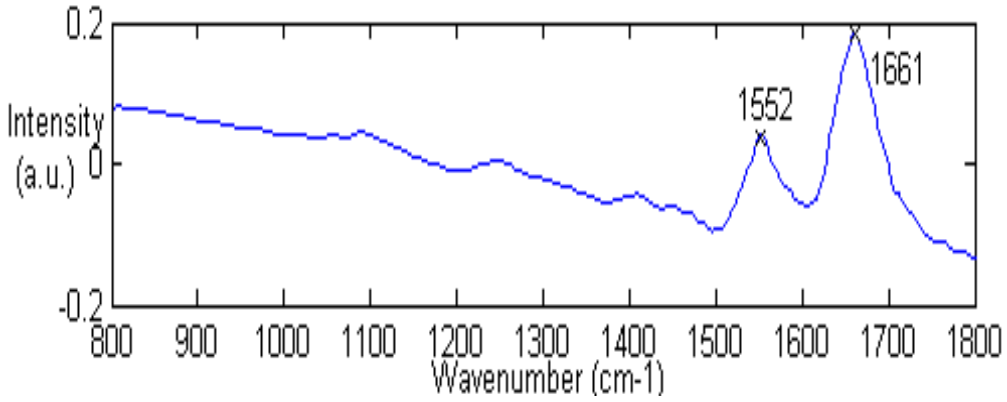


Figure 6-12: Principal component 3

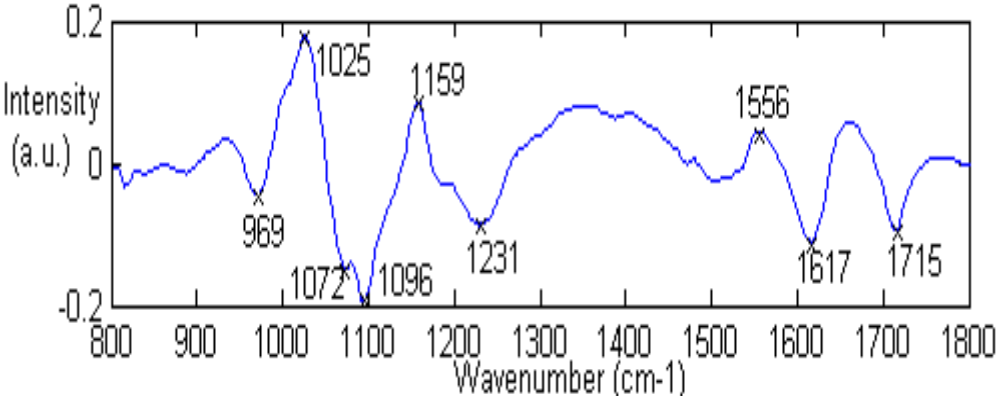


Figure 6-13: Principal component 8

6.1.2.2 Region selection and data collection

PC scores were used to identify major biochemical differences across the map and to select spectral data for the five and three cell line group classification models. Three regions (approximately 20µm x 10µm) of high cellular content were selected from each PC score for each of the 50 cell line samples (dataset 1a and 2a). The spectra from the selected regions of the PC score map were then combined and normalised to remove intensity variations. The normalised dataset was then used to construct and test classification models to enable the biochemical changes to be classified according to the defined cell line group. An example of selection of spectra for the classification model is shown in Figure 6-14. An in-house Matlab programme was used to select regions of interest. Averaged spectra from the three regions of interest were plotted for comparison and identification of spectral differences as shown in Figure 6-15.

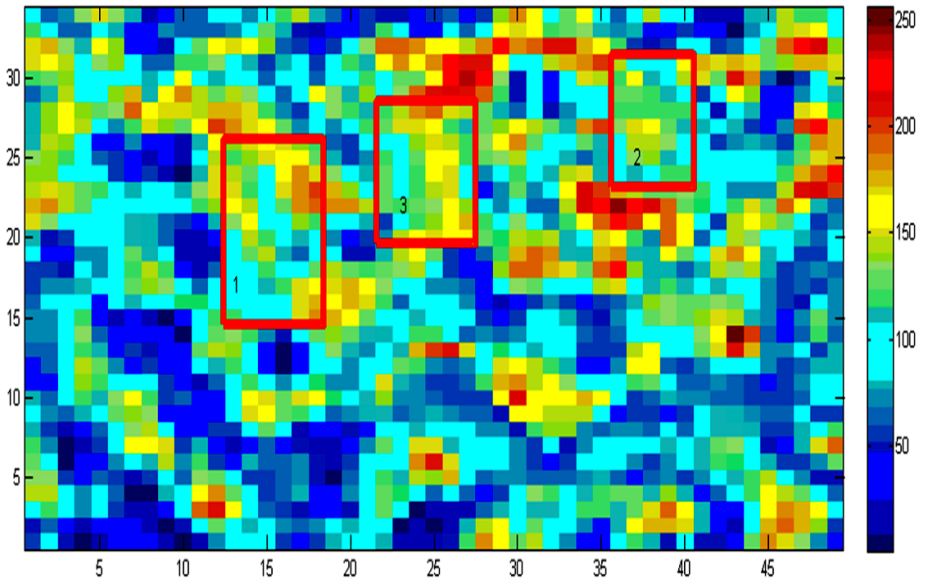


Figure 6-14: PC 2 score map for Meg01cell line. Red boxes represent three selected regions of interest. Each pixel on the pseudo-colour map represents the spectrum for that point

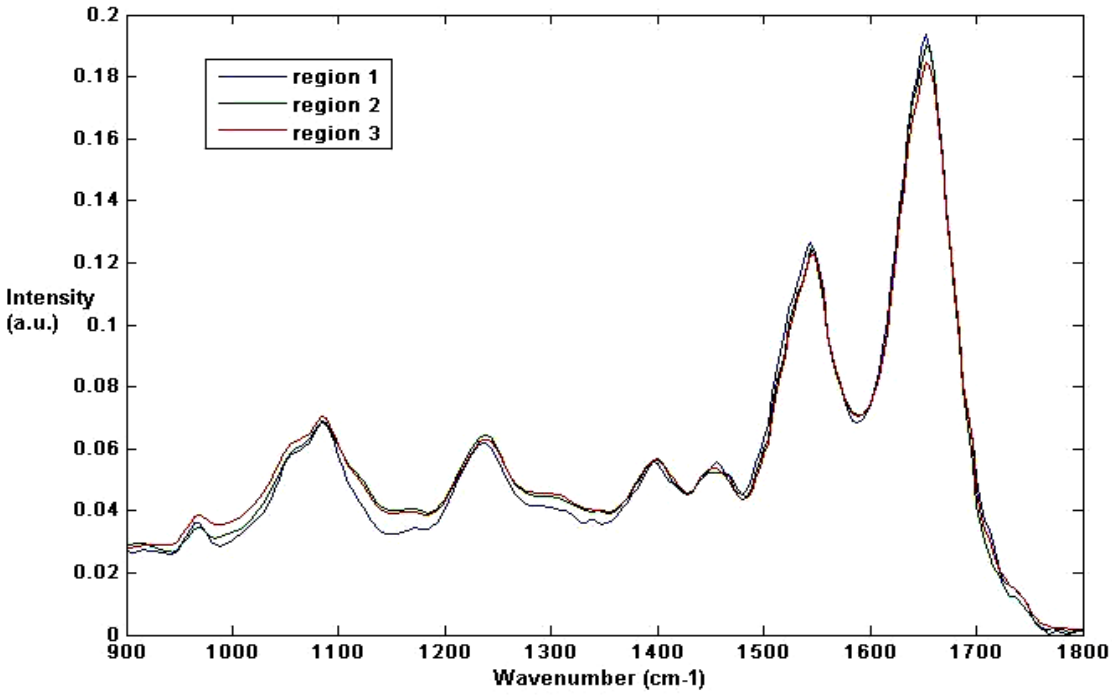


Figure 6-15: Mean normalized IR spectra from three selected areas, region 1 (blue line), region 2 (green line) and region 3 (red line) for Meg01cell line in the range 900-1800 cm^{-1}

6.1.2.3 ANOVA and scatter plots for PC scores

To interpret biochemical variations within the PC scores, ANOVA was performed on the dataset to demonstrate and assist in the selection of important principal components with maximal variance. The AVOVA F-values corresponding to each PC describes the ability of that PC to differentiate between the spectra obtained from the different cell line groups. With higher F-value; the better the principal component is at distinguishing the different cell line groups. The ANOVA F-value for the first 25 PCs for dataset 1a, 1b, 1c, 2a, 2b and 2c are tabulated in Table 6-5. The most significant PCs with the highest F-values are highlighted in red. The F_{crit} for spectral dataset 1a, 1b, 1c, 2a, 2b and 2c are tabulated in Table 6-4. The F_{crit} is the "threshold" value at the level of significance to classify a difference among group.

Dataset	Fcrit Value
1a	3.34
1b	3.20
1c	3.18
2a	4.25
2b	4.57
2c	4.54

Table 6-4: F_{crit} for spectral datasets 1a, 1b, 1c, 2a, 2b and 2c

The percentage of variance corresponding to the 25 PCs for datasets 1a, 1b, 1c, 2a, 2b and 2c are shown in Table 6-6. The first 11-12 PCs were seen to have the greatest variation ranging from 67% to 99% for the separation of cell line groups. The percentage variance does not exceed beyond 99% from 12 to 25 PCs. This may suggest that not all 25 PCs are necessary as an input into the LDA classification model. The effects of changing the numbers of PCs fed into the LDA classification model are discussed in Section 6.2.1.

The most significant principal components were plotted against one another in a two-dimensional scatter plot to visualise the natural clustering of cell line populations. The position of each spectrum is marked on the scatter plot and assigned a different colour

and symbol depending on cell line type and whether the spectrum belongs to frozen or air-dried sample. Direct comparison between the B-cell lymphoid and myeloid leukaemia and T-cell lymphoma cell lines for datasets 1a, 1b, 1c, 2a, 2b and 2c are presented in Figures (6-16 – 6-21).

Principal Components	F-values					
	Dataset 1a	Dataset 1b	Dataset 1c	Dataset 2a	Dataset 2b	Dataset 2c
1	333.4367	426.0863	**945.3188**	282.2076	34.887	778.9605
2	**1265.421**	**1890.23**	218.9316	**1284.78**	1076.793	270.6211
3	178.787	209.0607	886.359	234.1255	21.7052	766.5517
4	125.0908	839.7597	441.5297	182.7099	476.1455	**816.7263**
5	132.5464	699.0236	247.6401	64.3828	**1291.844**	326.1484
6	496.684	191.6305	52.7868	12.4013	330.0801	47.1793
7	684.0582	87.9959	779.8272	800.4704	102.0099	**841.4965**
8	**4563.345**	**970.1663**	132.6321	**4319.519**	**1508.761**	233.0032
9	103.0531	699.3829	409.3716	581.7895	1246.058	799.5467
10	202.2829	296.2329	327.3016	365.1057	6.0466	207.1826
11	669.7159	713.7848	528.3819	1044.695	42.7936	151.9631
12	278.2888	364.625	**907.4002**	321.1413	270.8251	423.8928
13	575.6955	133.9266	173.2946	519.5837	75.0516	341.8803
14	846.5202	388.3044	284.2646	803.6753	494.2436	553.5807
15	345.0524	75.614	219.4749	181.0117	75.3162	203.2445
16	1013.595	225.5205	46.3369	582.854	425.7424	29.897
17	394.5588	135.2361	100.0457	412.1973	76.722	24.6149
18	278.8043	99.3136	96.5434	126.9333	60.7738	685.9887
19	603.4616	57.7737	289.6301	101.0662	135.7848	374.7447
20	358.5668	38.6438	223.394	866.908	15.4326	30.6186
21	150.7639	83.3247	149.6214	95.5624	169.6881	244.592
22	46.9908	143.2041	259.0123	10.4237	50.2406	192.958
23	416.9666	85.6689	96.5434	779.0291	148.1598	32.7953
24	32.5906	226.433	84.158	126.9333	184.4157	87.6579
25	399.9253	53.4608	84.1769	286.1001	82.174	89.5393

Table 6-5: ANOVA F-value for the first 25 PCs for datasets 1a, 1b, 1c, 2a, 2b and 2c

Principal Components	% Variance					
	Dataset 1a	Dataset 1b	Dataset 1c	Dataset 2a	Dataset 2b	Dataset 2c
1	66.0239	66.4409	70.3406	66.3645	66.4409	70.3406
2	79.2583	78.7431	81.1373	79.6075	78.7431	81.1373
3	87.2333	87.6249	88.8913	87.4908	87.6249	88.8913
4	90.445	91.0243	91.8063	90.6503	91.0243	91.8063
5	93.0744	94.1781	93.9453	93.2816	94.1781	93.9453
6	94.8718	95.6509	95.9391	95.0769	95.6509	95.9391
7	95.9941	96.9023	97.0282	96.1969	96.9023	97.0282
8	96.9008	97.592	97.879	97.0668	97.592	97.879
9	97.6839	98.166	98.4396	97.8484	98.166	98.4396
10	98.2604	98.6362	98.8858	98.4148	98.6362	98.8858
11	98.6383	99.0441	99.1444	98.7851	99.0441	99.1444
12	98.9265	99.2744	99.3254	99.0587	99.2744	99.3254
13	99.167	99.4429	99.4714	99.2657	99.4429	99.4714
14	99.3111	99.5614	99.5661	99.4057	99.5614	99.5661
15	99.4376	99.6585	99.6497	99.5189	99.6585	99.6497
16	99.5541	99.7406	99.7195	99.6207	99.7406	99.7195
17	99.6365	99.7944	99.7691	99.6987	99.7944	99.7691
18	99.7092	99.8269	99.8117	99.7521	99.8269	99.8117
19	99.7498	99.8562	99.8414	99.7911	99.8562	99.8414
20	99.7879	99.8775	99.8682	99.8266	99.8775	99.8682
21	99.8165	99.8941	99.8884	99.8539	99.8941	99.8884
22	99.8405	99.9094	99.9026	99.8746	99.9094	99.9026
23	99.8591	99.9207	99.9133	99.891	99.9207	99.9133
24	99.8756	99.931	99.9222	99.9059	99.931	99.9222
25	99.8883	99.9384	99.9301	99.9168	99.9384	99.9301

Table 6-6: The percentage of variance corresponding to the 25 PCs for datasets 1a, 1b, 1c, 2a, 2b and 2c

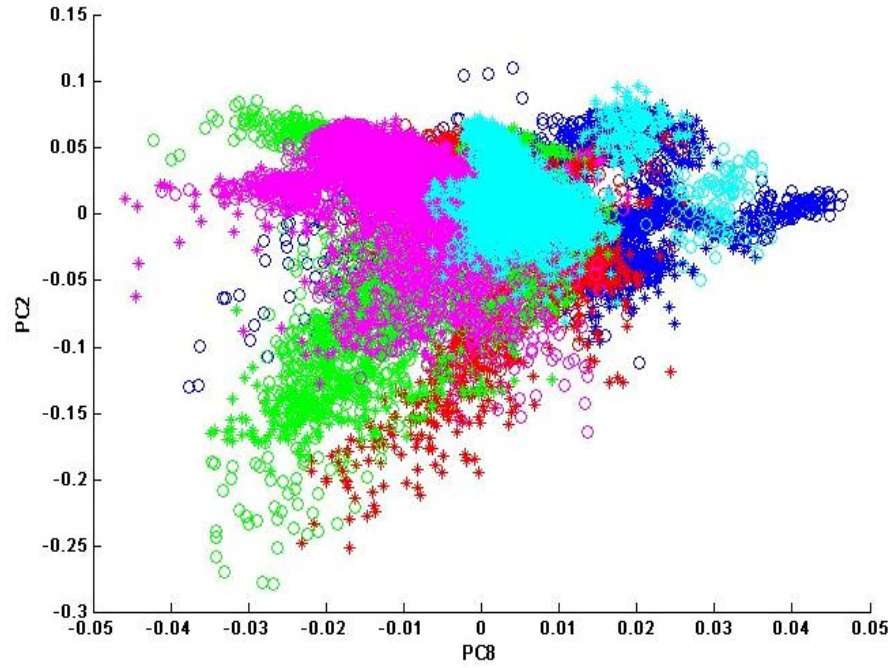


Figure 6-16: PC scores scatter plot (PC2 vs. PC8) represents clustering of cell lines for dataset 1a. Each symbol represents one spectrum. RCH-ACV dry= blue stars, RCH-ACV frozen= blue circles, REH dry= green stars, REH frozen= green circles, HL60 dry= red stars, HL60 frozen= red circles, Meg01 dry= magenta stars, Meg01 frozen= magenta circles and Karpas 299 dry= cyan stars, Karpas 299 frozen= cyan circles

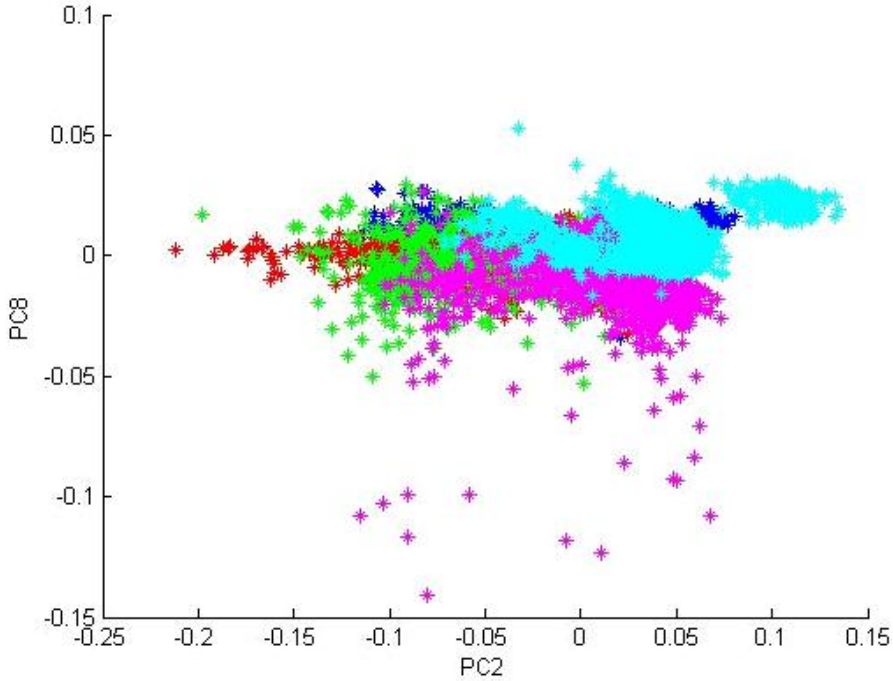


Figure 6-17: PC scores scatter plot (PC2 vs. PC8) represents clustering of cell lines for dataset 1b. RCH-ACV dry= blue stars, REH dry= green stars, HL60 dry= red stars, Meg01 dry= magenta stars, and Karpas 299 dry= cyan stars

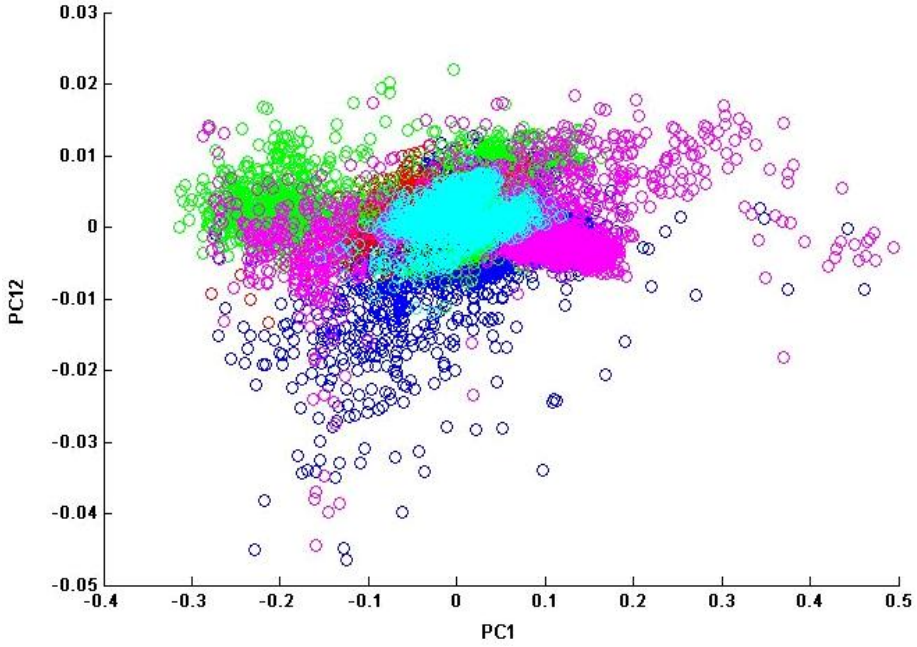


Figure 6-18: PC scores scatter plot (PC1 vs. PC12) represents clustering of cell lines for dataset 1c. RCH-ACV frozen= blue circles, REH frozen= green circles, HL60 frozen= red circles, Meg01 frozen= magenta circles and Karpas 299 frozen= cyan circles

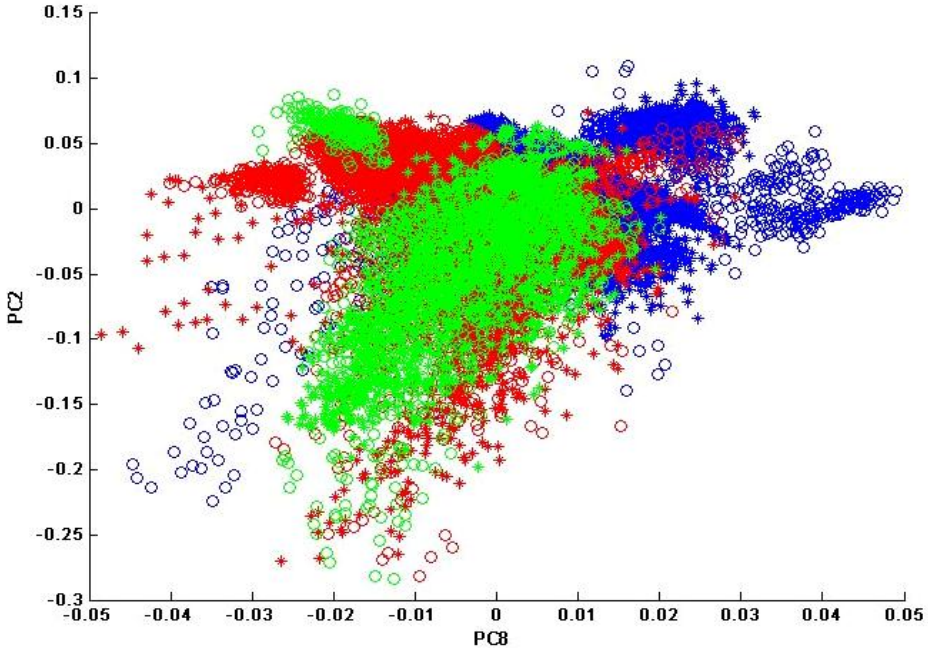


Figure 6-19: PC scores scatter plot (PC2 vs. PC8) represents clustering of cell lines for dataset 2a. Lymphoid group (RCH-ACV and REH) dry= blue stars; lymphoid group (RCH-ACV and REH) frozen= blue circles; myeloid group (HL60 and Meg01) dry= red stars; myeloid group (HL60 and Meg01) frozen= red circles; lymphoma (Karpas 299) dry= green stars and lymphoma (Karpas 299) frozen= green circles

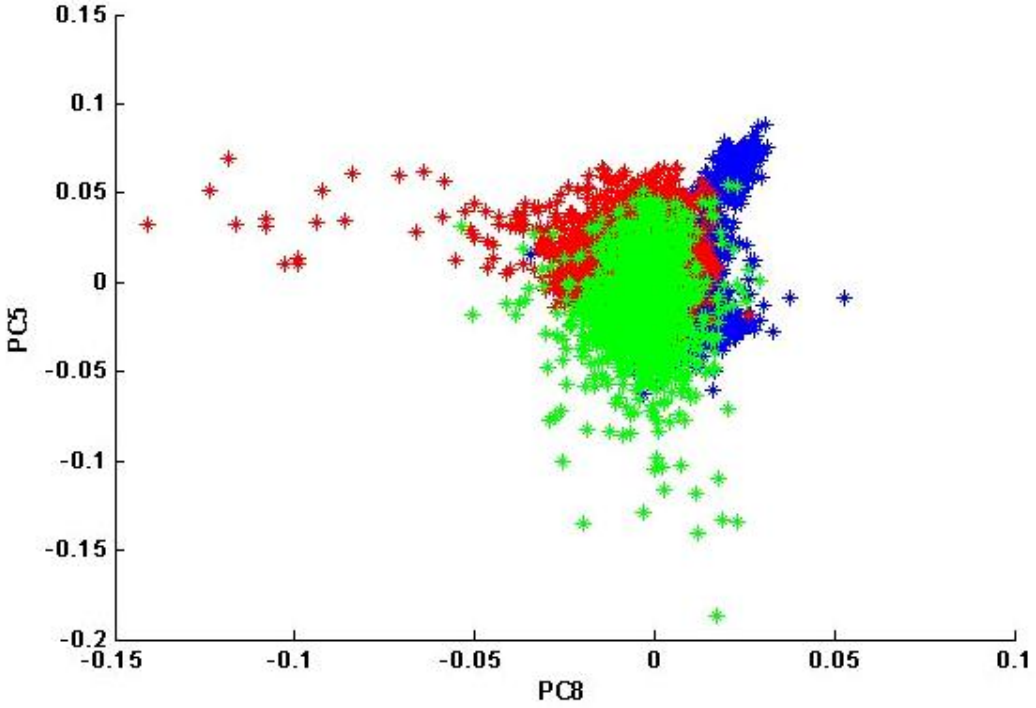


Figure 6-20: PC scores scatter plot (PC5 vs. PC8) represents clustering of cell lines for dataset 2b. Lymphoid group (RCH-ACV and REH) dry= blue stars; myeloid group (HL60 and Meg01) dry= red stars and lymphoma (Karpas 299) dry= green stars

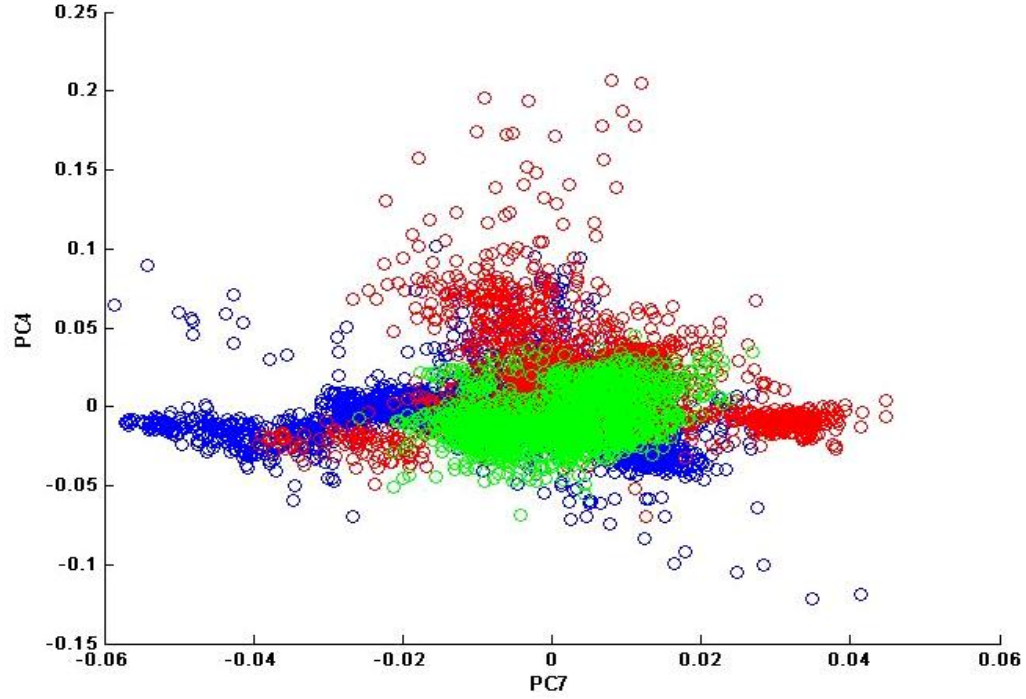


Figure 6-21: PC scores scatter plot (PC4 vs. PC7) represents clustering of cell lines for dataset 2c. Lymphoid group (RCH-ACV and REH) frozen= blue circles; myeloid group (HL60 and Meg01) frozen= red circles and lymphoma (Karpas 299) frozen= green circles

6.1.2.4 Linear Discriminant Analysis (LDA)

All 25 of the calculated PCs were used as input into an LDA classification model to maximise the variance between the cell line groups and minimise the variance within the groups. This required four linear discriminant functions for optimal separation of the cell line groups in datasets 1a, 1b and 1c. Two linear discriminant functions were required for optimal separation of the cell line groups in datasets 2a, 2b and 2c. A two-dimensional scatter plot of the linear discriminant scores from dataset 1a, 1b, 1c, 2a, 2b and 2c are displayed in Figures 6-22 – 6-27. Figure 6-24A – 6-26L shows subplots of linear discriminant scores in various combinations only for dataset 1c for visual purposes. Each dataset produced various combinations of LD subplots as a result of LDA and due to the sheer volume of data not all have been included in the thesis.

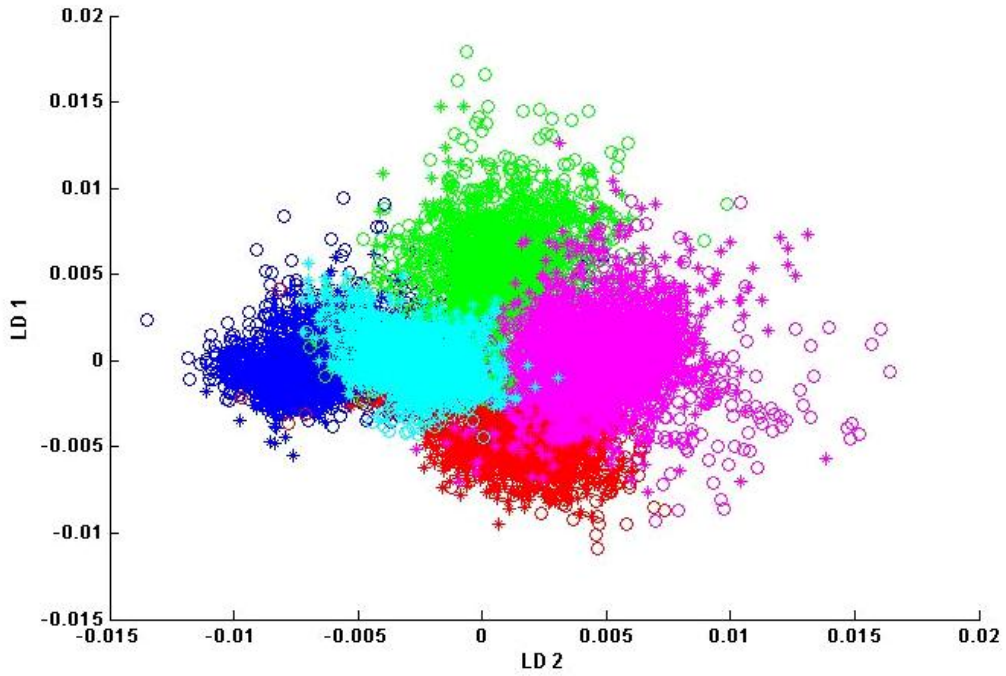


Figure 6-22: LDA scatter plot (LD1 vs. LD2) represents clustering of cell lines for model 1a. Each symbol represents one spectrum. RCH-ACV dry= blue stars, RCH-ACV frozen= blue circles, REH dry= green stars, REH frozen= green circles, HL60 dry= red stars, HL60 frozen= red circles, Meg01 dry= magenta stars, Meg01 frozen= magenta circles and Karpas 299 dry= cyan stars, Karpas 299 frozen= cyan circles

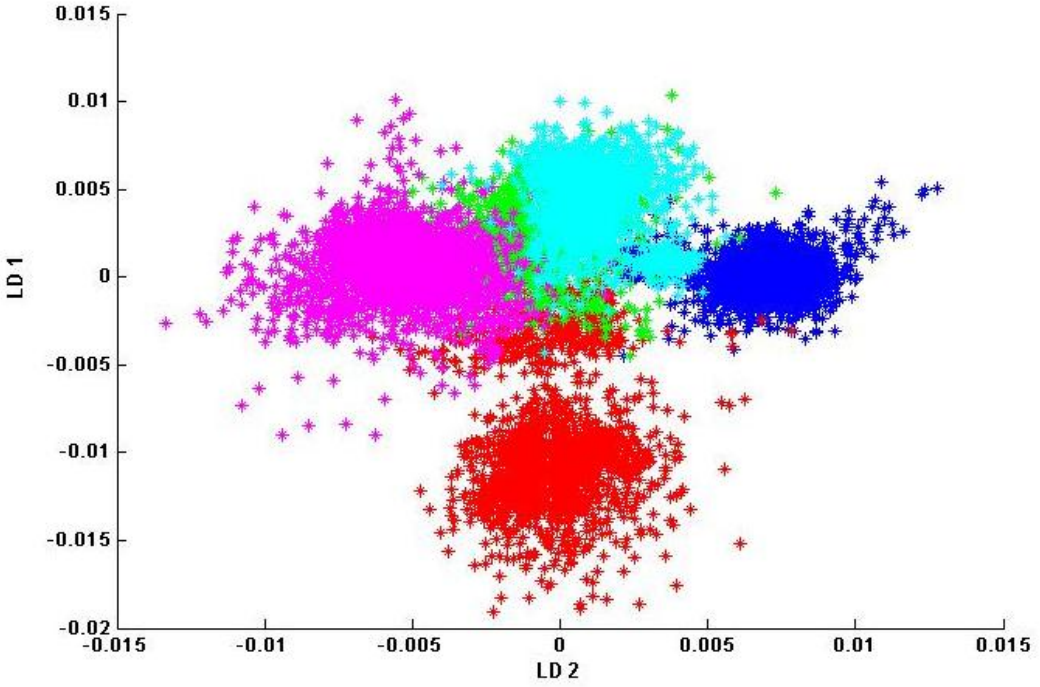
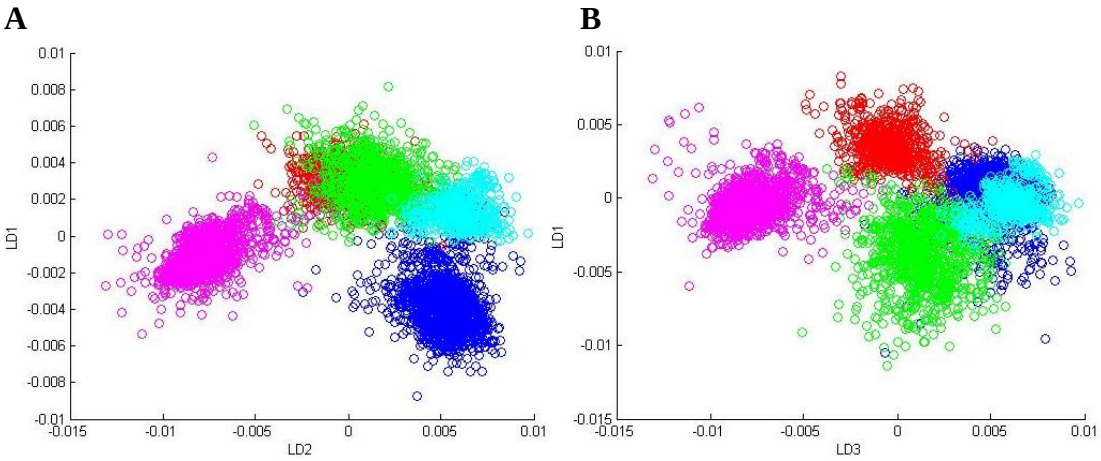
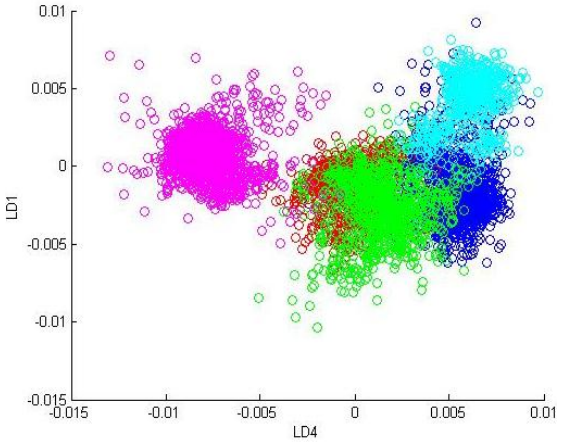


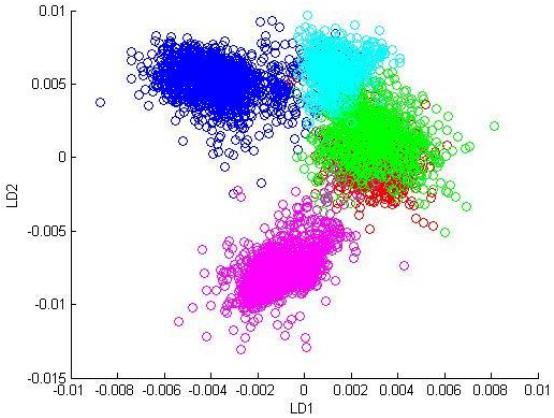
Figure 6-23: LDA scatter plot (LD1 vs. LD2) represents clustering of cell lines for dataset 1b. RCH-ACV dry= blue stars, REH dry= green stars, HL60 dry= red stars, Meg01 dry= magenta stars, and Karpas 299 dry= cyan stars



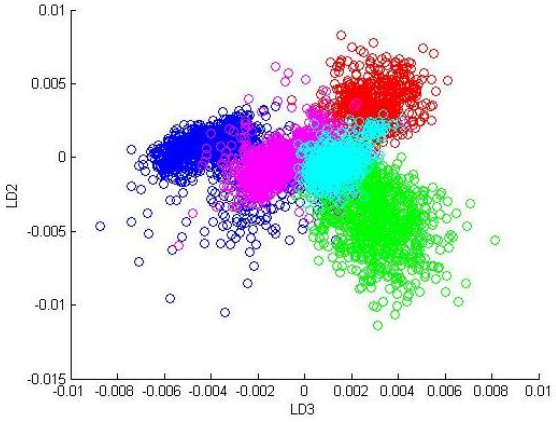
C



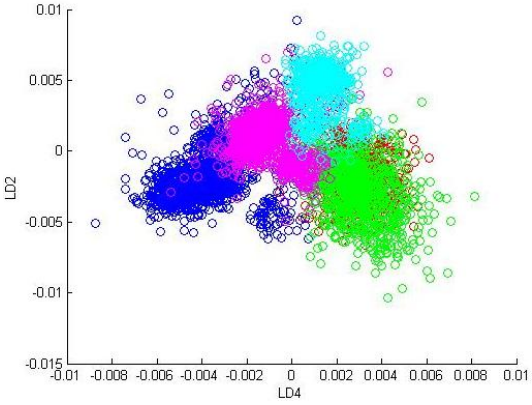
D



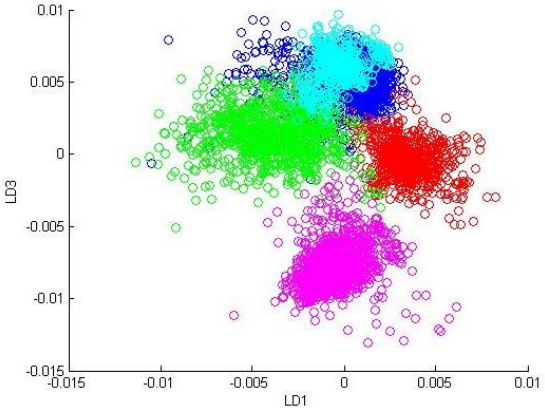
E



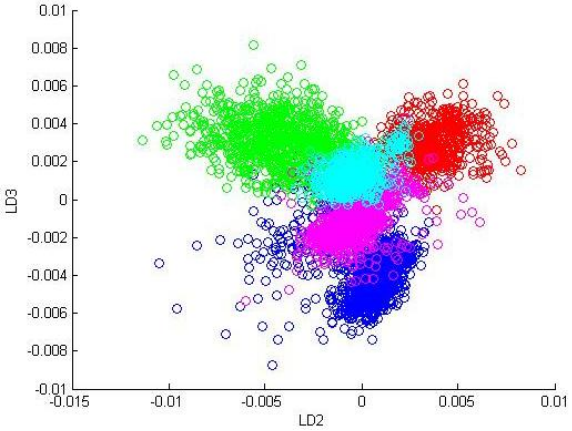
F



G



H



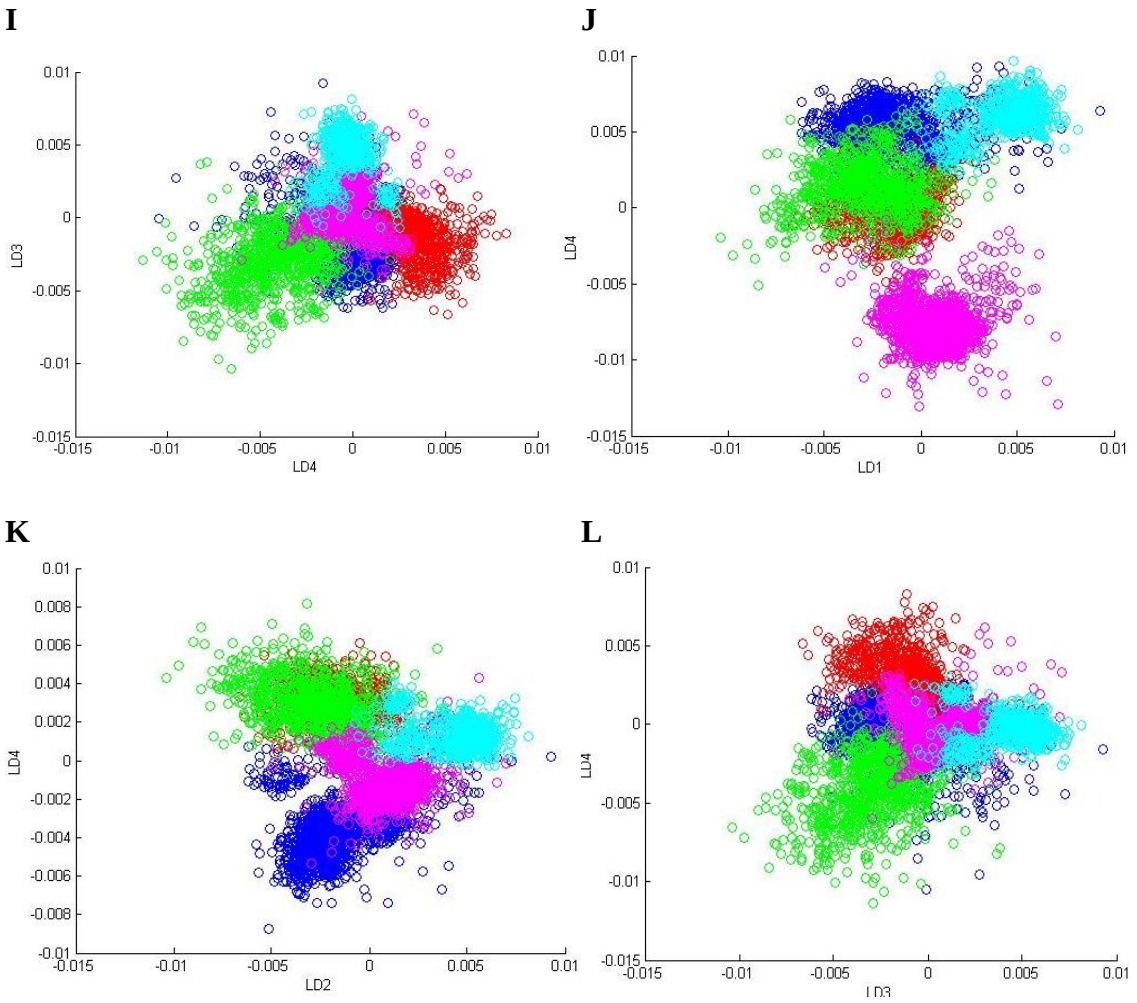


Figure 6-24: LDA scatter plots for dataset 1c. A) LDA scatter plot (LD1 vs. LD2) represents clustering of cell lines. RCH-ACV frozen= blue circles, REH frozen= green circles, HL60 frozen= red circles, Meg01 frozen= magenta circles and Karpas 299 frozen= cyan circles. B) LD1 vs. LD3, C) LD1 vs. LD4, D) LD2 vs. LD1, E) LD2 vs. LD3, F) LD2 vs. LD4, G) LD3 vs. LD1, H) LD3 vs. LD2, I) LD3 vs. LD4, J) LD4 vs. LD1, K) LD4 vs. LD2, L) LD4 vs. LD3

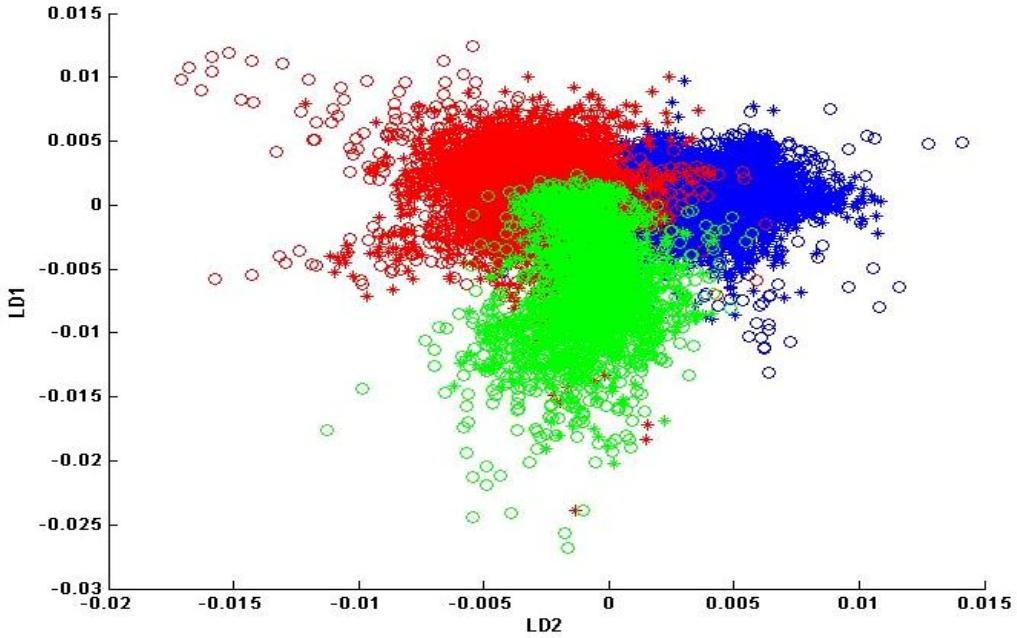


Figure 6-25: LDA scatter plot (LD1 vs. LD2) represents clustering of cell lines for dataset 2a. Lymphoid group (RCH-ACV and REH) dry= blue stars; lymphoid group (RCH-ACV and REH) frozen= blue circles; myeloid group (HL60 and Meg01) dry= red stars; myeloid group (HL60 and Meg01) frozen= red circles; lymphoma (Karpas 299) dry= green stars and lymphoma (Karpas 299) frozen= green circles

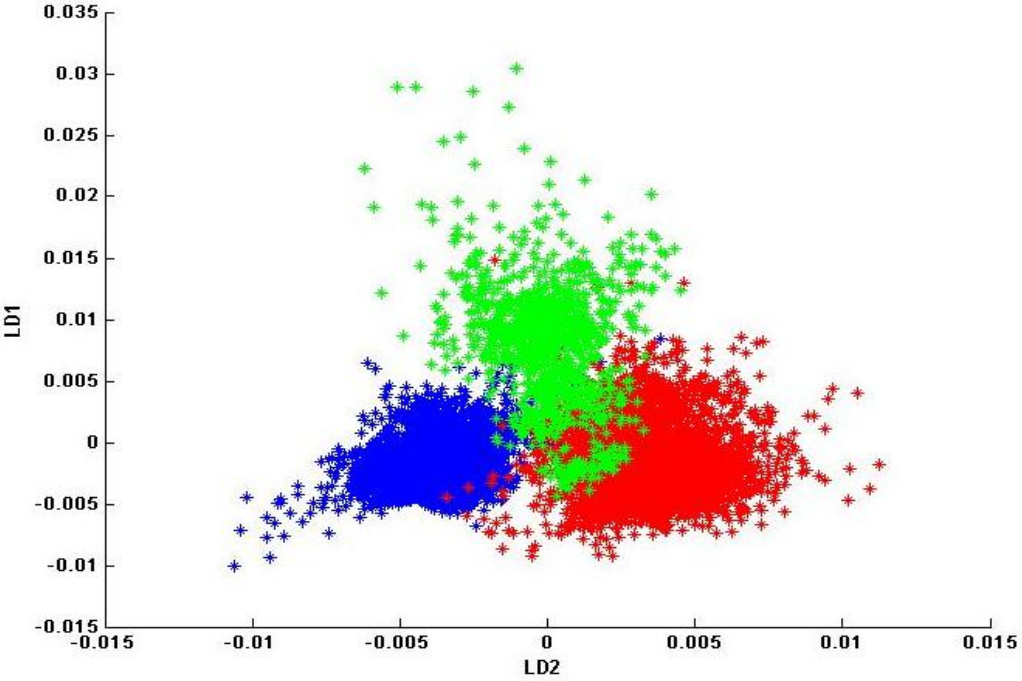


Figure 6-26: LDA scatter plot (LD1 vs. LD2) represents clustering of cell lines for dataset 2b. Lymphoid group (RCH-ACV and REH) dry= blue stars; myeloid group (HL60 and Meg01) dry= red stars and lymphoma (Karpas 299) dry= green stars

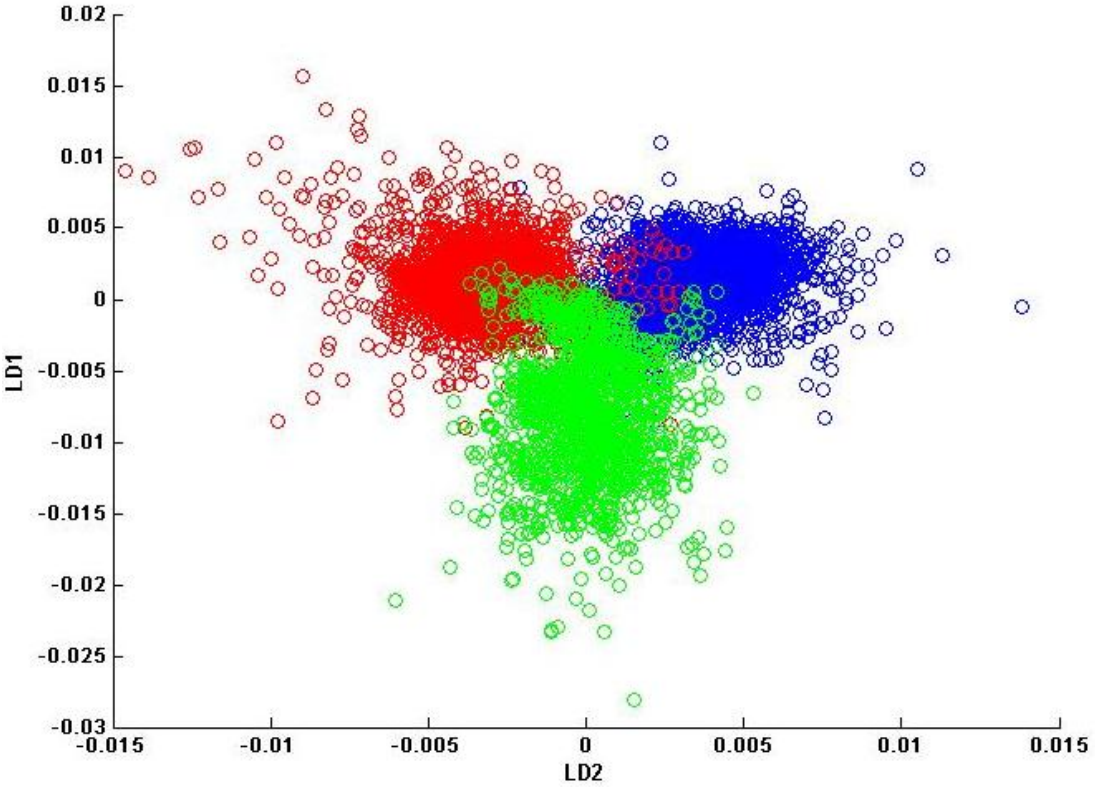


Figure 6-27: LDA scatter plot (LD1 vs. LD2) represents clustering of cell lines for dataset 2c. Lymphoid group (RCH-ACV and REH) frozen= blue circles; myeloid group (HL60 and Meg01) frozen= red circles and lymphoma (Karpas 299) frozen= green circles

6.1.2.5 Leave one out cross validation (LOOCV)

The efficiency of the PCA-fed LDA classification model was evaluated using leave-one-sample-out cross validation as described in Section 4.2.2.3. This method is ideal to determine the initial accuracy of the predictions of the PCA-fed LDA classification model if independent data is not available.

The PCA-fed LDA classification model was constructed with 25 PCs as previously mentioned calculated with 18,556 spectra for dataset 1a and 19,903 spectra for dataset 2a. For datasets 1b and 2b the PCA-fed LDA classification model was constructed with 25 PCs calculated collected from 25 dry cell line samples with 9168 spectra for dataset 1b and 9983 spectra for dataset 2b. For datasets 1c and 2c, the PCA-fed LDA

classification model was also constructed with 25 PCs calculated collected from 25 frozen cell line samples with 9398 spectra for dataset1c and 9920 spectra for dataset 2c.

The classification results, reflecting the number of spectra correctly classified for each cell line group in datasets 1a, 1b, 1c, 2a, 2b and 2c as well as their sensitivity and specificity are presented in Tables 6-7 – 6-12 . The results show that the percentage of spectra correctly predicted in the three cell line group (datasets 2a and 2c) is considerably higher than the five cell line group model.

The percentage of correctly classified spectra reflects a poor measure of the PCA-fed LDA classification model performance. The possible factors that contribute to the large number of spectra misclassified may be associated with:

- Unbalanced number of spectra contributed by each cell line group.
- Inhomogeneties in cellular sample as may contain cells at different stages of the cell cycle and variation in degree of cell differentiation and maturation
- Inclusion of saturated spectra.
- Inclusion of poor quality spectra introduced through air-drying or freezing of cell line samples.
- Intersample spectral variability within the same cell line group.

Despite a consistent method of spectral imaging, the number of spectra collected per map may have varied due to small inconsistencies in the size of the area probed during region selection. Ideally the number of spectra per cell line samples in the model should be equal for the leave one out cross validation procedure (LOOCV) in order to avoid poor prediction results. In addition, contributions from calcium fluoride spectra during region selection may also influence the quality of the spectral dataset. Methods of improving the PCA-fed LDA classification model are evaluated in Section 6.2.

Cell line group	Cell type classification					
	RCH-ACV	HL60	Karpas 299	Meg01	REH	Total No.
RCH-ACV	**1367**	253	628	4	1448	**3700**
Hl60	39	**2256**	482	42	399	**3218**
Karpas 299	35	294	**2169**	47	382	**2927**
Meg01	0	612	611	**2821**	845	**4889**
REH	306	112	1007	900	**1497**	**3822**
Number correct	**1367**	**2256**	**2169**	**2821**	**1497**	
						18556
						Average
Sensitivity (%)	**78.2**	**63.9**	**44.3**	**73.9**	**32.7**	**58.6**
Specificity (%)	**84.2**	**86.3**	**78.7**	**80.5**	**68.4**	**79.6**

Table 6-7: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1a. 54% spectra are correctly classified

Cell line group	Cell type classification					
	RCH-ACV	HL60	Karpas 299	Meg01	REH	Total No.
RCH-ACV	**424**	284	875	0	25	**1608**
Hl60	32	**807**	147	465	99	**1550**
Karpas 299	7	0	**1256**	13	264	**1540**
Meg01	0	33	888	**1109**	303	**2333**
REH	100	57	1060	895	**25**	**2137**
Number correct	**424**	**807**	**1256**	**1109**	**25**	
						9168
						Average
Sensitivity (%)	**75.3**	**68.3**	**29.7**	**44.6**	**3.5**	**44.3**
Specificity (%)	**84.8**	**86.6**	**58.8**	**67.7**	**69.3**	**73.4**

Table 6-8: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1b. 39.5% spectra are correctly classified

Cell line group	Cell type classification					
	RCH-ACV	HL60	Karpas 299	Meg01	REH	Total No.
RCH-ACV	**1095**	66	214	11	714	**2100**
Hl60	6	**1256**	370	0	34	**1666**
Karpas 299	16	381	**674**	4	312	**1387**
Meg01	0	201	130	**1765**	464	**2560**
REH	0	3	213	0	**1469**	**1685**
Number correct	**1095**	**1256**	**674**	**1765**	**1469**	
						9398
						Average
Sensitivity (%)	**98.0**	**65.8**	**42.1**	**99.2**	**49.1**	**70.8**
Specificity (%)	**87.6**	**86.9**	**81.2**	**89.4**	**78.1**	**84.6**

Table 6-9: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1c. 66.5% spectra are correctly classified

Cell line group	Cell type classification			
	Lymphoid	Myeloid	Lymphoma	Total No.
Lymphoid	**6308**	908	935	**8151**
Myeloid	1164	**6911**	568	**8643**
Lymphoma	262	723	**2124**	**3109**
Number correct	**6308**	**6911**	**2124**	
				19903
				Average
Sensitivity (%)	**77.5**	**79.6**	**68.4**	**75.3**
Specificity (%)	**87.9**	**85.6**	**91.4**	**88.2**

Table 6-10: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2a. 77.1% spectra are correctly classified

6.1.3 Summary

The raw FT-IR spectra acquired from T-cell lymphoma, B- cell lymphoid, and myeloid leukaemia cell lines reveal that peak shapes and positions are similar, although the peak intensities tend to vary between air-dried and frozen samples. The initial imaging experiments of the B-cell lymphoid, myeloid leukaemia and T-cell lymphoma cell lines have demonstrated problems associated with poor region selection and contributions

from CaF_2 which possibly result in poor quality spectral datasets. The preliminary five cell line group classification model using 50 frozen and air-dried samples correctly classified only 50% of the spectra. Furthermore, results show that the use of only frozen spectra improves the performance of classification model considerably.

Cell line group	Cell type classification			
	Lymphoid	Myeloid	Lymphoma	Total No.
Lymphoid	**1299**	954	1901	**4154**
Myeloid	1108	**1715**	1384	**4207**
Lymphoma	124	372	**1126**	**1622**
Number correct	**1299**	**1715**	**1126**	
				9983
				Average
Sensitivity (%)	**31.3**	**40.8**	**69.4**	**47.1**
Specificity (%)	**78.9**	**77.0**	**60.7**	**72.2**

Table 6-11: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2b. 41.5% spectra are correctly classified

Cell line group	Cell type classification			
	Lymphoid	Myeloid	Lymphoma	Total No.
Lymphoid	**3411**	160	426	**3997**
Myeloid	626	**3055**	757	**4438**
Lymphoma	266	418	**801**	**1485**
Number correct	**3411**	**3055**	**801**	
				9920
				Average
Sensitivity (%)	**85.3**	**68.8**	**53.9**	**69.3**
Specificity (%)	**84.9**	**89.4**	**85.9**	**86.8**

Table 6-12: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2c. 73.3% spectra are correctly classified

6.2 PCA-fed LDA classification model optimisation

In order to optimise the PCA-fed LDA classification model, firstly, the effects of increasing the principal components used for the PCA-fed-LDA classification model was investigated on the five cell line group model using datasets 1a, 1b and 1c. The cell line groups in these datasets are RCH-ACV, REH, HL60, Meg01 and Karpas299. PCA-fed-LDA was also preformed with variations of an additional factor to assess impact on results. Savitsky-Golay derivatives were tested as a pre-processing tool on the PCA-fed-LDA classification model to obtain the optimal performance. Futher details on methodology are described in Section 5.9.2. The results from these investigations are shown in Sections 6.2.1 and 6.2.2 respectively.

Other data pre-processing methods used prior to multivariate analysis (PCA-fed-LDA) were considered to achieve an optimal and stable classification model. Two distinct methods of data pre-processing were employed to enhance spectral features and effectively suppress or eliminate unnecessary artefacts and noise. The initial method exploits standard deviation filtering as a measure to identify outliers and determine spectral variability. The second method utilizes the application of an absorption filter tool. These data pre-processing methods are described in Sections 4.1.5 and 4.1.6.The outcome of using standard deviation filter and absorption filter tool on the above mentioned four datasets are discussed in Sections 6.2.3 and 6.2.4 respectively.

6.2.1 Effect of increasing PCs input for LDA

The number of PCs fed into the PCA-fed-LDA classification model was varied to exploit its ability to separate the five cell line groups as well as to improve on the prediction results. The number of PCs fed into the classification model were increased from 5 PCs to 25 PCs. PCs 5, 10, 15 and 25 have a percentage variance of 93.3%, 98.4%, 99.5% and 99.9% respectively as shown in Table 6-6. The percentage of spectra correctly classified for each of the cell line groups using spectral datasets 1a, 1b and 1c

are presented in Figures 6-28 – 6-30. The overall performances of the classification models developed with varying numbers of PCs are shown in Table 6-13.

The correct prediction of spectra is dependent of the number of PCs used in the analysis. Fewer PCs input into the PCA-fed-LDA classification model are shown to reduce the prediction achieved. This is seems to be the case particularly for spectra from spectral dataset 1a (dry and frozen spectra) and spectral dataset 1b (only dry spectra). Using cross validation on spectral dataset 1b (only frozen spectral dataset) more spectra are correctly classified (Figure 6-29). The percentage of spectra correctly classified increases when data from higher of PCs is entered into LDA as demonstrated in Figure 6-28 and 6-30. Although this is not the case for some cell line groups (RCH-ACV, Meg01 and REH) in spectral dataset 1b.

Overall, the highest percentage correct classification is seen when the data from the first 25 PCs entered using spectral dataset 1c. This may indicate lesser degree of inter-sample spectral variation within the same cell line group in the frozen spectral dataset. Sensitivity and specificity achieved by a PCA fed LDA model developed with 5, 10, 15 and 25 PCs are tabulated in Table 6-13.

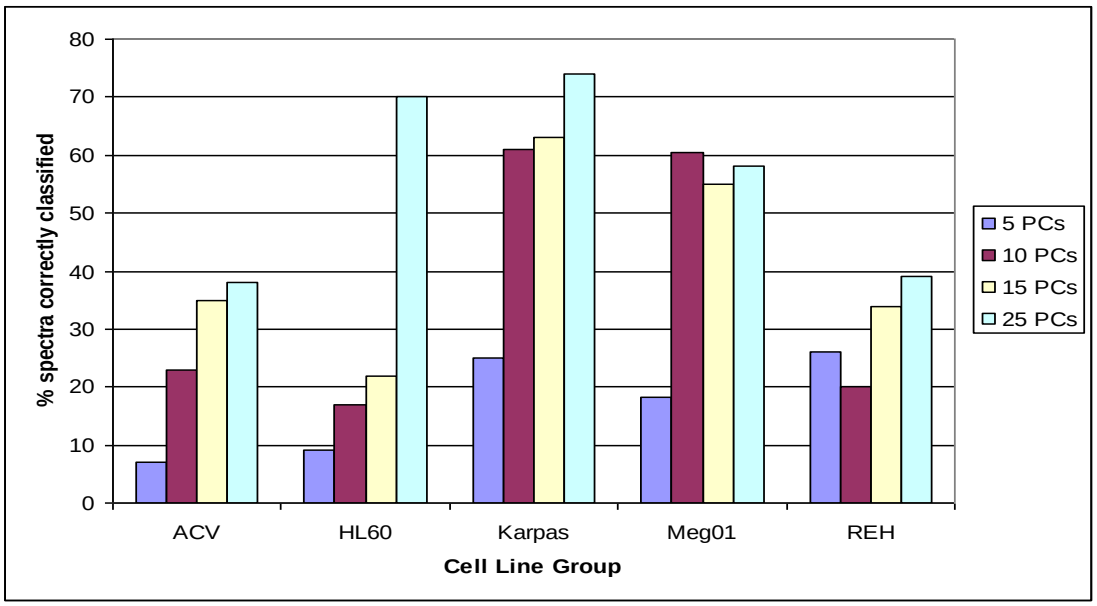


Figure 6-28: percentage of spectra correctly classified by leave-one-sample-out cross validation model developed with increasing numbers of PC for the separation of the five cell line groups in spectral dataset 1a

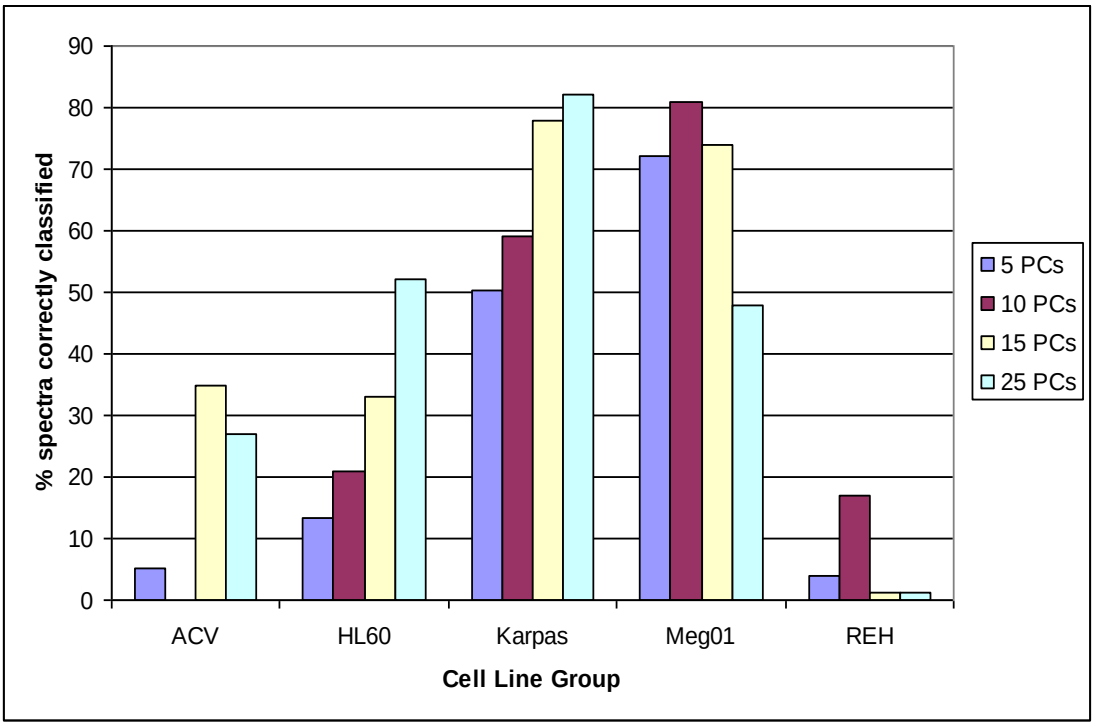


Figure 6-29: percentage of spectra correctly classified by leave-one-sample-out cross validation model developed with increasing numbers of PC for the separation of the five cell line groups in spectral dataset 1b

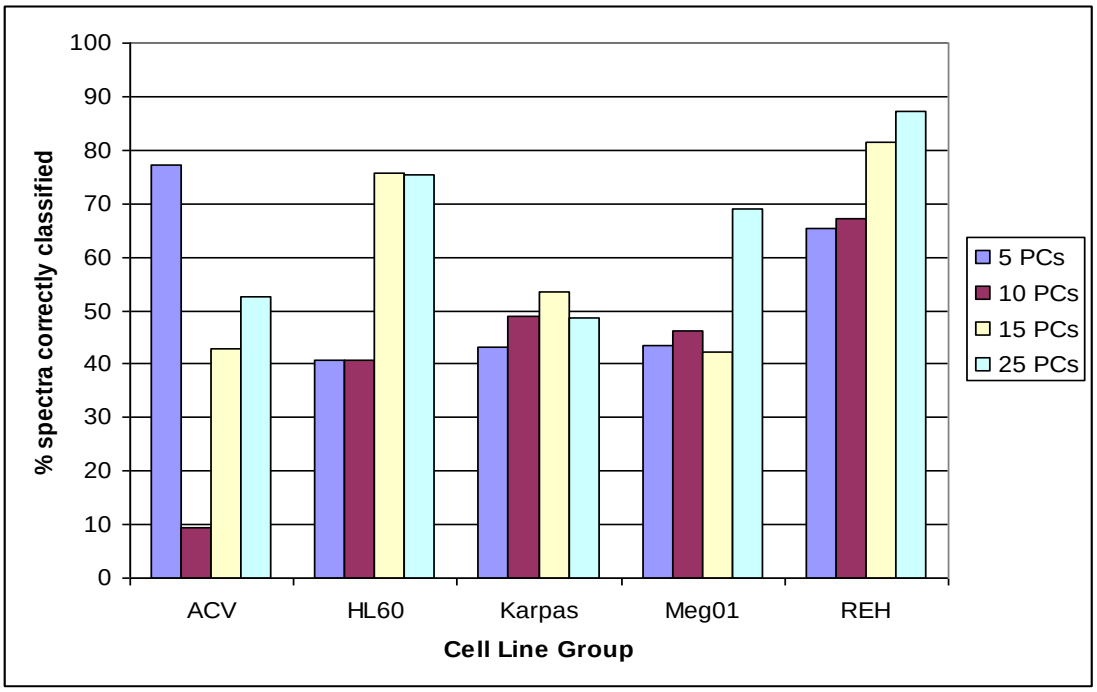


Figure 6-30: percentage of spectra correctly classified for each of the cell line groups developed with increasing numbers of PC for spectral dataset 1c

5 PCs						
%		RCH-ACV	HL60	Karpas 299	Meg01	REH
Dataset 1a	Sensitivity	7.4	10.9	17.2	25.9	21.1
	Specificity	63.4	70.9	68.0	62.9	70
Dataset 1b	Sensitivity	22.6	18.8	22.1	45.4	34.0
	Specificity	80.1	75.1	58.3	64.3	74.4
Dataset 1c	Sensitivity	32.3	32.8	49.5	91.4	37.5
	Specificity	65.9	78.3	84.1	81.3	70.9
10 PCs						
Dataset 1a	Sensitivity	31.9	21.3	37.9	69.6	16.9
	Specificity	73.8	73.9	75.8	79.3	62.1
Dataset 1b	Sensitivity	0	20.6	33.3	57.6	34.6
	Specificity	76.6	71.9	70.2	74.8	72.1
Dataset 1c	Sensitivity	28.4	40.1	32.7	91.8	30.9
	Specificity	73.9	77.1	75.8	81.9	62.8
15 PCs						
Dataset 1a	Sensitivity	61.0	26.8	38.8	57.7	30.2
	Specificity	82.3	74.9	76.1	73.7	67.9
Dataset 1b	Sensitivity	33.5	48.9	51.0	52.9	3.2
	Specificity	74.9	81.8	81.2	71.2	67.9
Dataset 1c	Sensitivity	94.1	49.2	45.1	96.6	44.1
	Specificity	85.2	79.1	82.1	81.8	74.4
25 PCs						
Dataset 1a	Sensitivity	78.2	63.9	44.2	73.9	32.7
	Specificity	84.2	86.3	78.7	80.5	68.4
Dataset 1b	Sensitivity	75.3	68.3	29.7	44.7	3.5
	Specificity	84.9	86.6	58.9	67.8	69.3
Dataset 1c	Sensitivity	98	65.9	42.1	99.2	49.1
	Specificity	87.6	86.9	81.2	89.4	78.1

Table 6-13: Sensitivity and specificity achieved by a PCA fed LDA model developed with 5, 10,15 and 25 PCs

6.2.2 Effects of Savitsky-Golay filters on PCA-fed-LDA classification model

In an attempt to achieve an optimal performance of the 5 cell line group classification model, spectral dataset was subjected to a range of 0^{th}, 1^{st} and 2^{nd} order Savitsky-Golay derivative calculated using a 9-points. The derivative Savitsky-Golay filters were tested on PCA-fed-LDA model developed with 25 PCs, calculated with 9398 spectra from spectral dataset 1c. Spectral dataset 1c was chosen, as it has been proven to contain a better quality spectral dataset as demonstrated in Section 6.2.1.

The application of derivative Savitsky-Golay filters on a spectral dataset has significant impact on the performance of the PCA-fed-LDA classification model as shown in Figure 6-31. 1st derivative and in particular 2nd derivative reduce the percentage of data correctly classified. Higher percentages of the spectra were correctly classified using 0th derivative. This significantly differs from findings in the literature, as 1st order Savitsky-Golay derivative is often used as a pre-processing procedure. Table 6-14 demonstrates the overall performance of the classification model with varying Savitsky-Golay derivatives. The PCA-fed-LDA classification model achieves the highest sensitivity and specificity using 0th order Savitsky-Golay derivative on the spectral dataset.

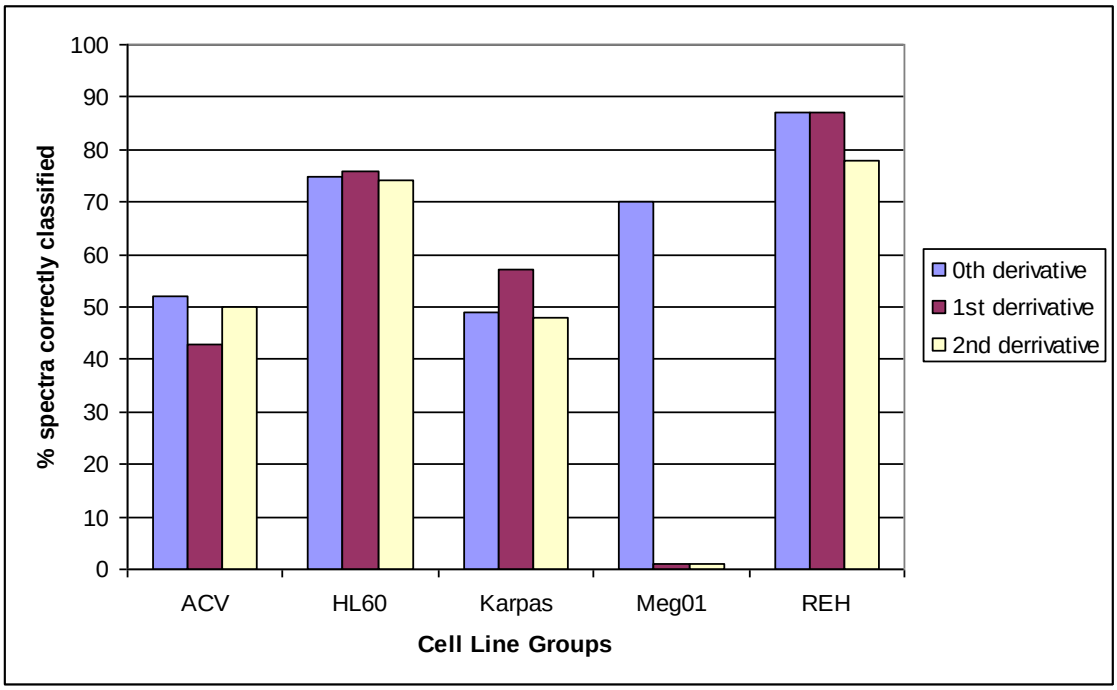


Figure 6-31: Percentage of spectra correctly classified for each of the cell line groups achieved with 0th, 1st and 2nd derivative Savitsky-Golay filters for spectral dataset 1c.

%		RCH-ACV	HL60	Karpas 299	Meg01	REH
0th Derv	**Sensitivity**	93.0	65.8	42.0	99.2	49.1
	Specificity	87.6	86.9	81.2	89.4	78.1
1st Derv	**Sensitivity**	80.7	78.5	21.6	80.9	44.7
	Specificity	83.4	90.8	59.9	72.8	75.1
2nd Derv	**Sensitivity**	67.7	82.5	19.4	35.7	45.8
	Specificity	81.4	91.6	59.9	72.6	76.2

Table 6-14: Sensitivity and specificity achieved by testing 0th, 1st and 2nd order Savitsky-Golay derivative on the PCA fed LDA classification model

6.2.3 Application of standard deviation filtering on spectral dataset

Standard deviation filtering was used as a measure to identify outliers (poor quality spectra) and applied to the spectral datasets of all individual five cell line groups. A range of standard deviation multiples were applied to the raw dataset prior to PCA-fed-LDA. Mean spectra were calculated for all cell line groups and the number of standard deviation was reduced from +/- 5 to +/- 1.5 to identify outliers. Outliers are referred to as spectra that are more than the defined number of standard deviation either side of the mean of the dataset. Any outlier spectra beyond the standard deviation range were removed to improve the quality of spectral dataset and obtain a stable model. The remaining spectra for all cell line groups were collated to create a new, improved dataset and input into the PCA-fed-LDA classification model. Figure 6-32 shows the superimposed mean +/- 1.5 standard deviations of the infrared absorbance of the different cell line groups.

As before, for explanatory purposes, standard deviation filtering was applied to spectral dataset 1c as it has been proven to contain a better quality spectral dataset as demonstrated in previous sections. The application of standard deviation filtering on a spectral dataset has significant impact on the performance of the PCA-fed-LDA classification model (using 25 PCs and 0th derivative) as shown in Table 6-15. Higher percentages of the spectra were correctly classified using a standard deviation of +/-1.5

for all cell line groups compared to that of when no standard deviation filter is applied. The PCA-fed-LDA classification model achieves the highest sensitivity and specificity using standard deviation of +/- 1.5 on the spectral dataset (Table 6-16).

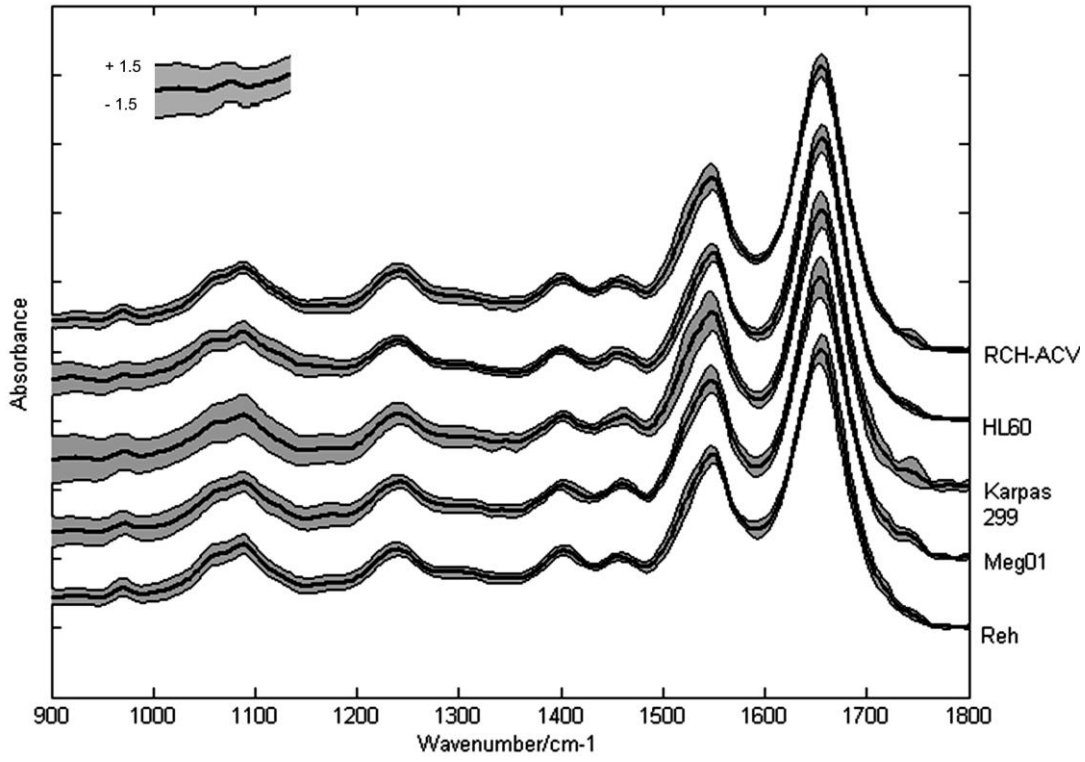


Figure 6-32: Shows the superimposed +/- 1.5 standard deviations of the infrared absorbance of the different cell types

	Percentage of spectra correctly classified					
Standard deviation	**RCH-ACV**	**HL60**	**Karpas 299**	**Meg01**	**REH**	**Total No. of spectra**
None applied	52	75	49	79	87	9398
+/- 5	53	75	49	73	88	9285
+/- 4	52	75	48	73	90	9143
+/- 3	53	71	42	28	88	8387
+/- 2	69	62	47	75	89	7132
+/- 1.5	**88**	**86**	**62**	**73**	**96**	**4037**

Table 6-15: Percentage of spectra correctly classified for each of the cell line groups achieved using a range of standard deviations for spectral dataset 1b. Those highlighted in red show that using standard deviation of +/- 1.5 has the most significant impact on the prediction results

%		RCH-ACV	HL60	Karpas 299	Meg01	REH
No std	**Sensitivity**	93.0	65.8	42.0	99.2	49.1
	Specificity	87.6	86.9	81.2	89.4	78.1
+/- 5	**Sensitivity**	98.2	66.6	41.6	99.6	51.6
	Specificity	87.9	87.2	80.6	90.7	79.6
+/- 4	**Sensitivity**	97.9	68.1	38.8	99.7	53.2
	Specificity	87.8	87.6	79.6	90.9	80.5
+/- 3	**Sensitivity**	96.7	67.6	34.2	99	51.7
	Specificity	87.5	87.2	78.0	91.8	79.6
+/- 2	**Sensitivity**	51.2	67.8	34.1	99.8	54.8
	Specificity	74.4	87	78.2	76.6	83.5
+/- 1.5	**Sensitivity**	71.7	80.8	65.2	99.7	72.9
	Specificity	82.5	94.5	94.6	85.7	93.9

Table 6-16: Sensitivity and specificity achieved by applying +/- 5, +/- 4, +/- 3, +/- 2 and +/- 1.5 standard deviations on the performance of the PCA fed LDA classification model compared to when no standard deviation is applied

Although using a standard deviation of +/- 1.5, has improved the quality of the spectral dataset with outliers removed, it has a huge impact on the number of spectra eliminated. In this case, 5361 of the spectra has been removed from the original spectral dataset comprised of 9398 spectra. This is more than half of the spectra being eliminated which account for 57% as demonstrated in Table 6-17. This suggests that there is large inter-sample variability within the spectral dataset. This is significantly a large number of spectra being removed from the original dataset and is not statistically valid. Although standard deviation is an effective method of eliminating poor quality spectra and retaining those of high-quality with little variability, detecting outliers using this approach could be flawed. An alternative method to improve the quality of spectral dataset is discussed in Section 6.2.4.

Standard deviation (std)	Total No. of spectra after std	No. of spectra eliminated	% of spectra eliminated
None applied	Original dataset (9398)	0	0
+/- 5	9285	113	1.2
+/- 4	9143	255	2.7
+/- 3	8387	1011	10.8
+/- 2	7132	2266	24.1
+/- 1.5	4037	5361	57

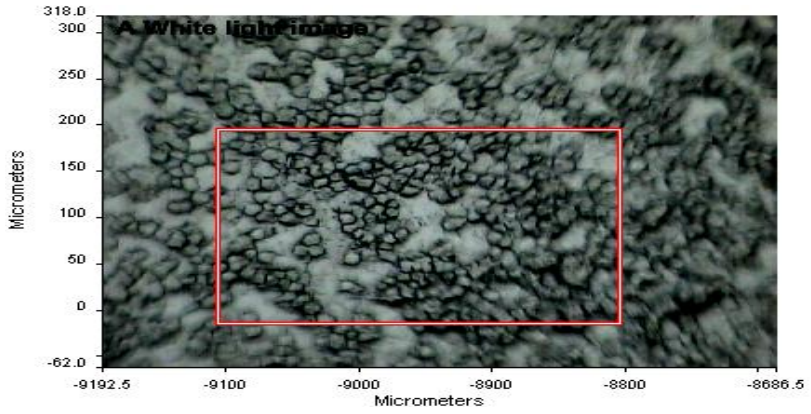
Table 6-17: Total number of spectra and percentage of spectra eliminated from the original dataset after applying a range of standard deviations

6.2.4 Application of an absorption filter tool on spectral dataset

White light images as well as their corresponding average absorbance image were obtained from 25 dry and 25 frozen cell line samples. An in-house Matlab programme was used to convert the data from transmittance into absorbance values from the captured regions as explained in Section 5.9.2. The average absorbance image shows absorbencies of varying intensities, where blue indicates low and red indicates high absorbance values. For explanatory and visual purposes, the Meg01 frozen cell line obtained from batch 5 is used as an example. The white light image and the corresponding average absorbance image for Meg01 cell line is shown in Figure 6-33.

An in-house absorbance spectral filter tool was applied to the loaded hyperspectral data cube to filter out weak or low absorbance. All spectra outside the absorbance range of 0.2 (upper absorbance limit) and 0.15 (lower absorbance limits) were removed. Figure 6-34 represents the average absorbance image after the absorbance spectral filter tool was applied to filter out weak and saturated spectra. As demonstrated in Figure 6-34, blue regions indicating low absorbance values corresponding to regions of calcium fluoride have been eliminated. The spectral dataset in Figure 6-35 represents the original unfiltered dataset and the remaining good quality filtered spectra.

A



B

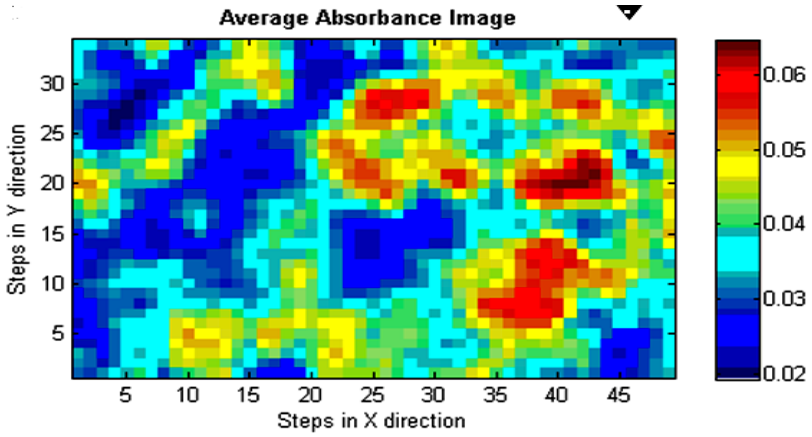


Figure 6-33: A) White light image of Meg01 cell line, B) average absorbance image of frozen Meg01 cell line (batch 5). Red box (200x300µm) highlights the region of interest

A

B

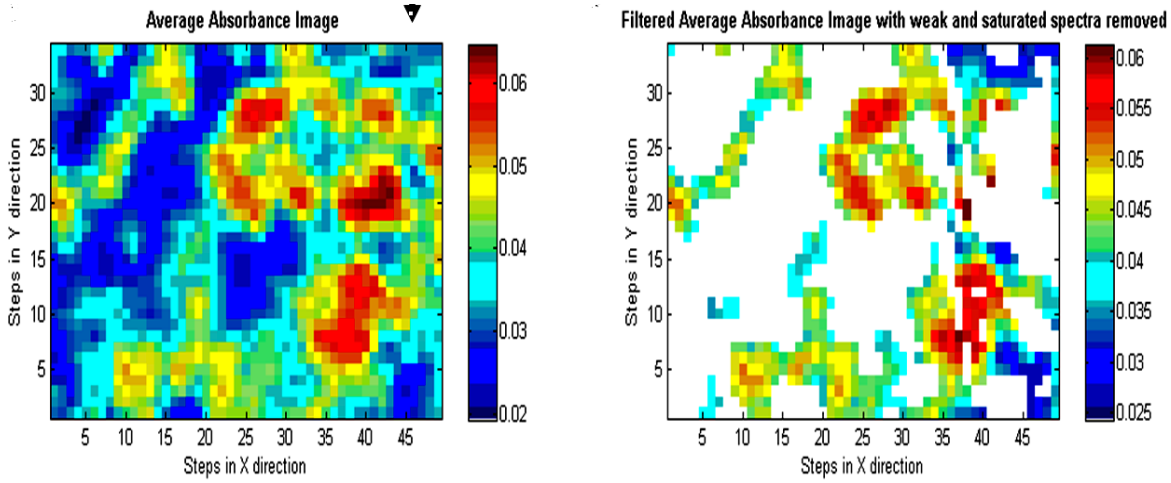


Figure 6-34: A) Average absorbance image of frozen Meg01 cell line (batch 5), B) Filtered average absorbance image. White areas represent weak or saturated spectra that have been filtered out

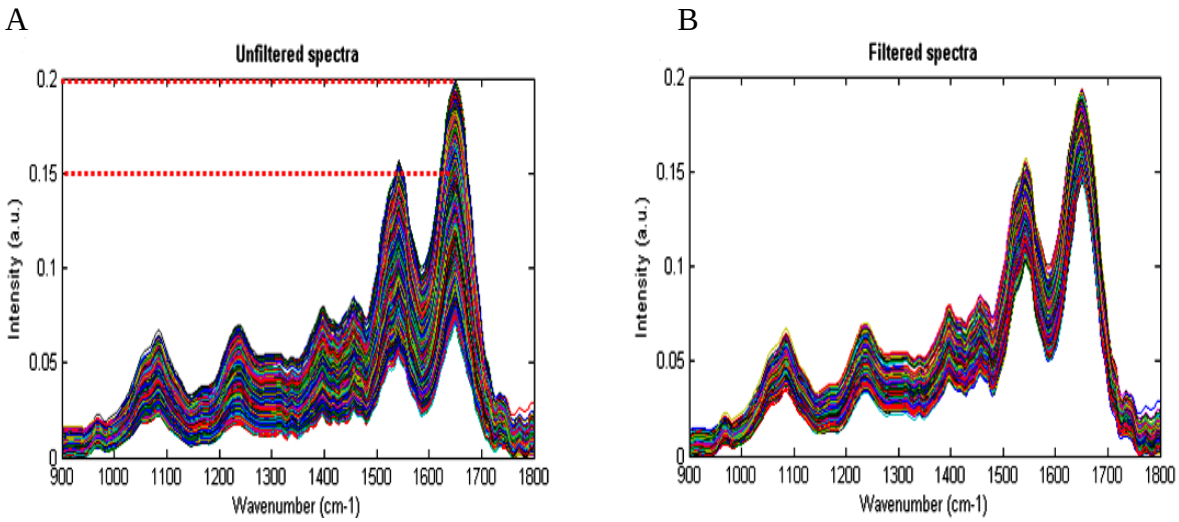


Figure 6-35: A) Unfiltered normalised spectral dataset and B) Remaining good quality normalised filtered spectra. All spectra outside the absorbance range of 0.2 (upper absorbance limit) and 0.15 (lower absorbance limits) were removed

The new filtered dataset with most weak absorbance and saturated spectra removed was divided into separate datasets:

- Five cell line group dataset comprised of both air-dried and frozen spectra from 50 samples. This dataset is defined as spectral dataset 1a.
- Five cell line group dataset comprised of only air-dried spectra from 25 samples. This dataset is defined as spectral dataset 1a.
- Five cell line group dataset comprised of only frozen spectra from 25 samples. This dataset is defined as spectral dataset 1c.
- Three cell line group dataset comprised of both air-dried and frozen spectra from 50 samples. This dataset is defined as spectral dataset 2a.
- Three cell line group dataset comprised of only air-dried spectra from 25 samples. This dataset is defined as spectral dataset 2b.
- Three cell line group dataset comprised of only frozen spectra with 25 samples. This dataset is defined as spectral dataset 2c.

The improved dataset was inputted into PCA-fed-LDA classification model to assess its ability to distinguish between the cell line groups. The application of absorbance filter tool on the various spectral datasets has significant impact on the performance of the PCA-fed-LDA classification model (using 25 PCs and 0^{th} derivative). The classification

results, reflecting the number of spectra correctly classified for each cell line group in datasets 1a, 1b, 1c, 2a, 2b and 2c as well as their sensitivity and specificity are presented in Tables 6-18 – 6-23.

Cell line group	Cell type classification					
	RCH-ACV	**HL60**	**Karpas 299**	**Meg01**	**REH**	Total No.
RCH-ACV	**618**	55	127	19	225	1044
Hl60	0	**499**	0	14	91	604
Karpas 299	0	188	**555**	75	69	887
Meg01	0	38	1	**105**	58	202
REH	180	281	72	58	**926**	1517
Number correct	**618**	**499**	**555**	**105**	**926**	4254
						Average
Sensitivity (%)	59.2	82.6	62.6	51.9	61.0	**63.5**
Specificity (%)	94.4	84.6	94.0	95.9	83.8	**90.5**

Table 6-18: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1a. 63.5% spectra are correctly classified

Cell line group	Cell type classification					
	RCH-ACV	**HL60**	**Karpas 299**	**Meg01**	**REH**	Total No.
RCH-ACV	**373**	0	0	0	25	398
Hl60	0	**94**	16	61	0	171
Karpas 299	0	1	**91**	4	1	97
Meg01	0	0	0	**166**	5	171
REH	311	18	196	94	**33**	652
Number correct	**373**	**94**	**91**	**166**	**33**	1489
						Average
Sensitivity (%)	93.7	54.9	93.8	97.0	5.06	**68.9**
Specificity (%)	71.5	98.5	84.7	87.9	96.2	**87.8**

Table 6-19: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1b. 50.1% spectra are correctly classified

Cell line group	Cell type classification					
	RCH-ACV	**HL60**	**Karpas 299**	**Meg01**	**REH**	Total No.
RCH-ACV	545	0	23	0	36	**604**
Hl60	0	658	1	0	7	**666**
Karpas 299	0	74	543	0	8	**625**
Meg01	0	0	0	646	0	**646**
REH	3	83	81	0	475	**642**
Number correct	**545**	**658**	**543**	**646**	**475**	**3183**
						Average
Sensitivity (%)	90.2	98.8	86.9	100	73.9	**89.9%**
Specificity (%)	99.9	93.8	95.9	100	97.9	**97.5%**

Table 6-20: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1c. 90.0% spectra are correctly classified

Cell line group	Cell type classification			
	Lymphoid	**Myeloid**	**Lymphoma**	**Total No.**
Lymphoid	**1730**	360	265	2355
Myeloid	1	**1253**	85	1339
Lymphoma	3	251	**898**	1152
Number correct	**1730**	**1253**	**898**	**4846**
				Average
Sensitivity (%)	**77.5**	**93.6**	**77.9**	**81.7**
Specificity (%)	**99.8**	**82.6**	**90.5**	**90.9**

Table 6-21: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2a. 80.1% spectra are correctly classified

Cell line group	Cell type classification			
	Lymphoid	Myeloid	Lymphoma	Total No.
Lymphoid	**774**	12	85	871
Myeloid	0	**432**	83	515
Lymphoma	2	4	**191**	197
Number correct	**774**	**432**	**191**	**1583**
				Average
Sensitivity (%)	**88.9**	**83.9**	**96.9**	**89.9**
Specificity (%)	**99.7**	**98.5**	**87.9**	**95.7**

Table 6-22: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2b. 83.6% spectra are correctly classified

Cell line group	Cell type classification			
	Lymphoid	Myeloid	Lymphoma	Total No.
Lymphoid	**808**	94	128	1030
Myeloid	15	**954**	139	1108
Lymphoma	20	91	**723**	834
Number correct	**808**	**954**	**723**	**2972**
				Average
Sensitivity (%)	**78.4**	**86.1**	**86.7**	**83.7**
Specificity (%)	**98.2**	**90.1**	**87.5**	**91.9**

Table 6-23: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2c. 88.1% spectra are correctly classified

6.2.5 Summary

Increasing the numbers of PCs input into the PCA-fed-LDA classification model from 5 PCs to 25 PCs is shown to increase the prediction accuracy achieved. Furthermore, the application of 1^{st} and 2^{nd} derivative reduces the percentage of spectra correctly classified in the frozen spectral dataset. 0^{th} derivative has a significant impact on the performance of the PCA-fed LDA classification model with higher percentages of the spectra correctly classified. Using a standard deviation of +/- 1.5 improved the performance of classification model resulting with highest sensitivity and specificity values. The use of an absorbance spectral filter tool to filter out weak/ low absorbance or saturated spectra is a better approach of identifying poor quality spectra. The spectra retained after the application of the absorbance spectral filter tool is of better quality and more stable.

6.3 Empirical and Multivariate Analysis for Classification of Normal and CLL Blood Specimens

6.3.1 Empirical analysis: Normal Blood

10 blood samples were collected from healthy subjects. Blood components such as plasma and buffy-coat (containing white blood cells) were isolated using standard procedures as described in Sections 5.2.3 and 5.2.4. White light images and spectral measurements were obtained for whole blood (containing red blood cells), plasma and buffycoat. An example of a white light image obtained from a buffy-coat sample is shown in Figure 6-36. Several regions were probed (marked in red) for their IR signature to evaluate spectral variations as shown in Figure 6-36. Point spectra obtained in transmission mode were obtained from whole blood, plasma and buffy-coat samples at different regions were averaged using an in-house Matlab programme. Due to lack of viscosity, the transfer of some whole blood, buffy-coat and plasma samples onto CaF_2 slides proved to be difficult. Hence, there are unbalanced numbers of spectra contributed by each blood component. The total number of point spectra collected for whole blood, buffy-coat and plasma are 829, 437 and 78 respectively.

A plot of the mean FT-IR spectra of the 3 separate blood components is shown in Figure 6-37 for comparison. Significant spectral differences were noted between the three blood components. These were identified using a standard empirical method of assigning key peak positions. Spectrum for individual blood component is sited in Appendix E. The peak positions of the major vibrational modes are labelled. The full list of peak assignments are summarised in Appendix D.

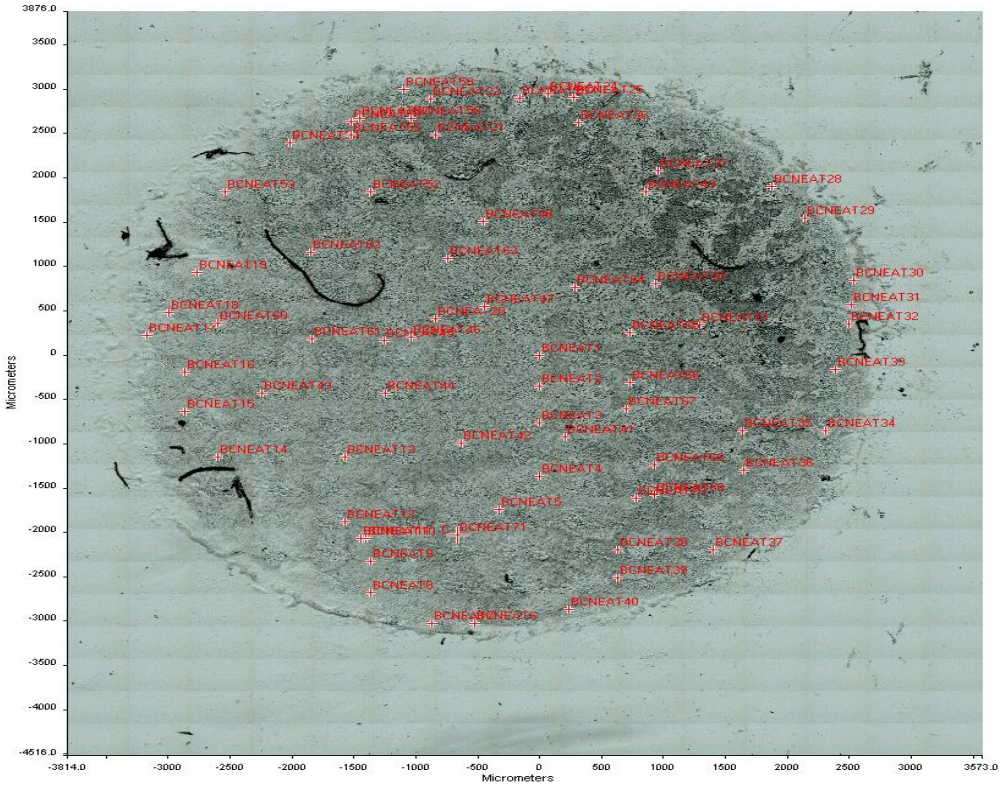


Figure 6-36: White light image from a buffy-coat sample. Point measurements were taken from areas marked in red

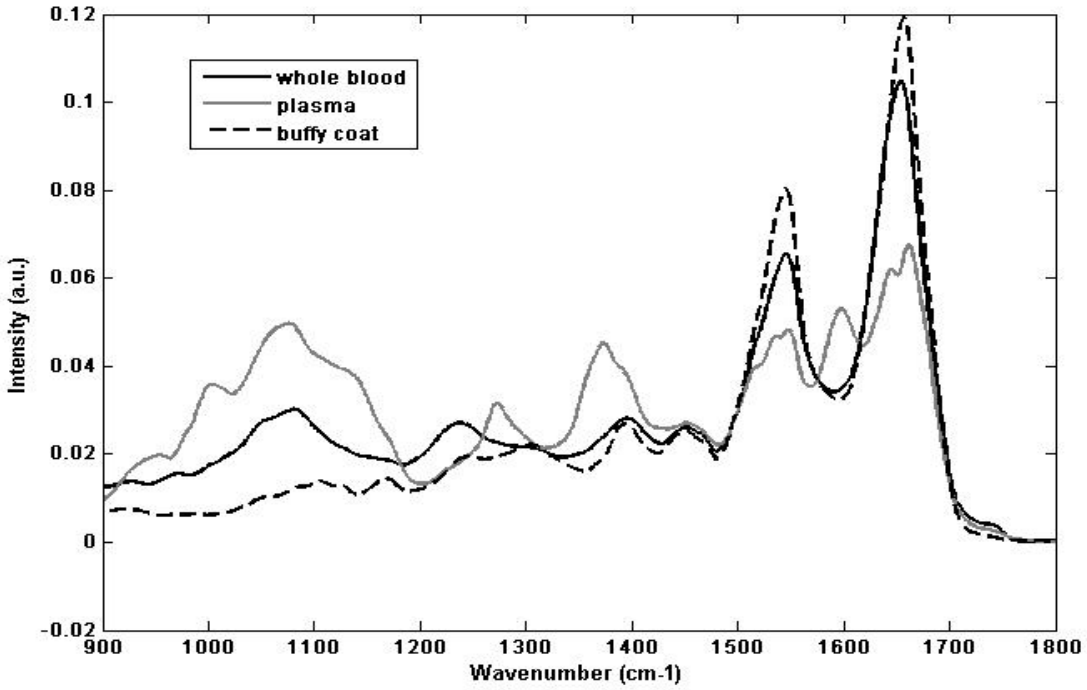


Figure 6-37: Represents the average IR spectra for whole blood, plasma and buffy-coat in the range 900-1800 cm^{-1} for normal blood

The representative mean spectra for whole blood and buffy-coat have a degree of variations in peak positions and intensities as shown in Figure 6-37. The most distinctive peaks in the IR absorption spectra for whole blood and buffy coat are assigned to the absorption modes of proteins the most intense is the amide I band, centred near 1658 cm^{-1} for whole blood spectra and 1651 cm^{-1} for buffy-coat spectra. The amide I corresponds to the C=O stretching coupled to the N-H bending and the C-H stretching modes of peptide bonds (Cecilia 2008 and Ami 2003). The amide II band at 1546 cm^{-1} for both whole blood and buffy coat spectra arises from the vibrational modes that involve the C-N-H bending and C-N stretching of peptidic bonds (Gasparri 2003). The amide III peak at 1241 cm^{-1} for whole blood and 1235 cm^{-1} for buffycoat spectra is contributed by proteins arising from C-N stretching and N-H bending (Gasparri 2003 and Liu 1997).

The peaks at 1453-1455 cm^{-1} are attributed to CH_2 and CH_3 deformation vibrations primarily due to lipid contribution in both whole blood and buffycoat spectra. Peaks at 1393-1396 cm^{-1} (CH_3 vibration) also correspond to lipids. A small peak at 1742 cm^{-1} was observed in the case of buffy-coat spectra, which may correspond to C=O stretching vibrations from phospholipids in membranes. This peak is absent in the spectra for whole blood.

Symmetric and anti-symmetric stretching vibrations from the phosphate group are seen in the FT-IR spectra of whole blood and buffy-coat in the region between 950 and 1280 cm^{-1}. Apart from phospholipids and phosphorylated proteins, nucleic acids are the main contributors of these bands in cells. For buffy-coat spectra, nucleic acid signals contribute at 1083 cm^{-1} (sym PO_2^-), 1051cm^{-1} (PO_2^- and ribose groups) and 969 cm^{-1} (phosphodiester group present in the DNA and RNA) (Ami 2003). For whole blood spectra, nucleic acid signals contribute at 1106 and 1172 cm^{-1}. The peak at 969 cm^{-1} is absent in the spectra for whole blood.

The representative mean FT-IR spectrum for plasma differs considerably, with significant variations in peak positions and vibrational modes as shown in Figure 6-37. The spectrum for plasma is dominated by peaks mainly corresponding to lipids, carbohydrates, amino acids and lactate. Main peaks arise at 1003 cm^{-1} (C-OH of

carbohydrates, C-O-P of lipids), 1372 cm^{-1} (CH_3) and 1455 cm^{-1} (CH_2 and CH_3 deformation vibrations). These vibrations in the low molecular weight biomolecular components are more prevalent in plasma compared to whole blood and buffy-coat. There are also minor peaks that contribute to Amide I (1660 cm^{-1} with a shoulder at 1641 cm^{-1}) and Amide II (1546 with a shoulder at 1538 cm^{-1}).

6.3.1.1 Difference Spectra: Normal Blood

To further emphasise and extract the spectral variations between the blood components, difference spectra, was calculated by subtracting the mean spectrum from blood component from another. The difference spectrum was calculated using an in-house Matlab programme. Figure 6-38 shows the resulting spectra obtained from taking the difference between mean whole blood from plasma. Figure 6-39 shows the resulting spectra obtained from subtracting the mean whole blood spectra from mean buffy-coat spectra. The difference spectra obtained subtracting the mean plasma spectra from mean buffy-coat spectra is represented in figure 6-40. The complicated band patterns in the difference spectra are composed of peaks arising from changes of vibrational modes, to changes of the protein structures and to changes of specific nucleic acids, amino acid side chains and fatty acids. Table 6-24 lists the biochemical constituents in high concentrations as positive peaks in a certain blood component and those with lower concentrations as negative peaks in other blood component.

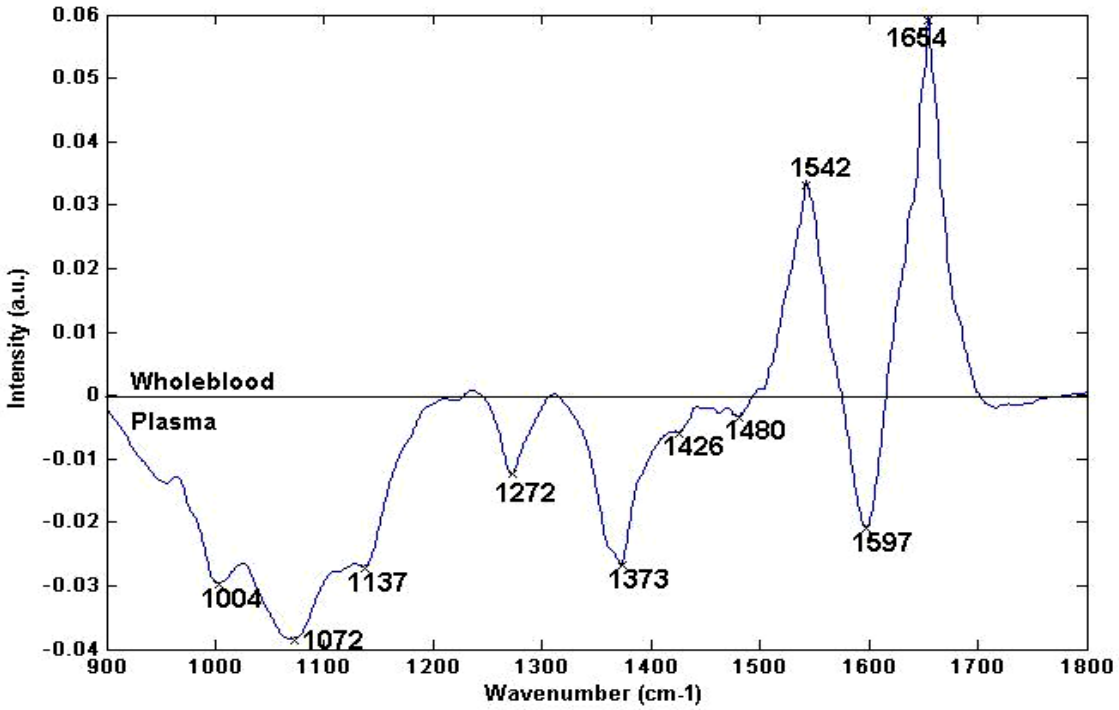


Figure 6-38: Difference spectra obtained by subtraction of whole blood spectrum from the spectrum obtained from plasma

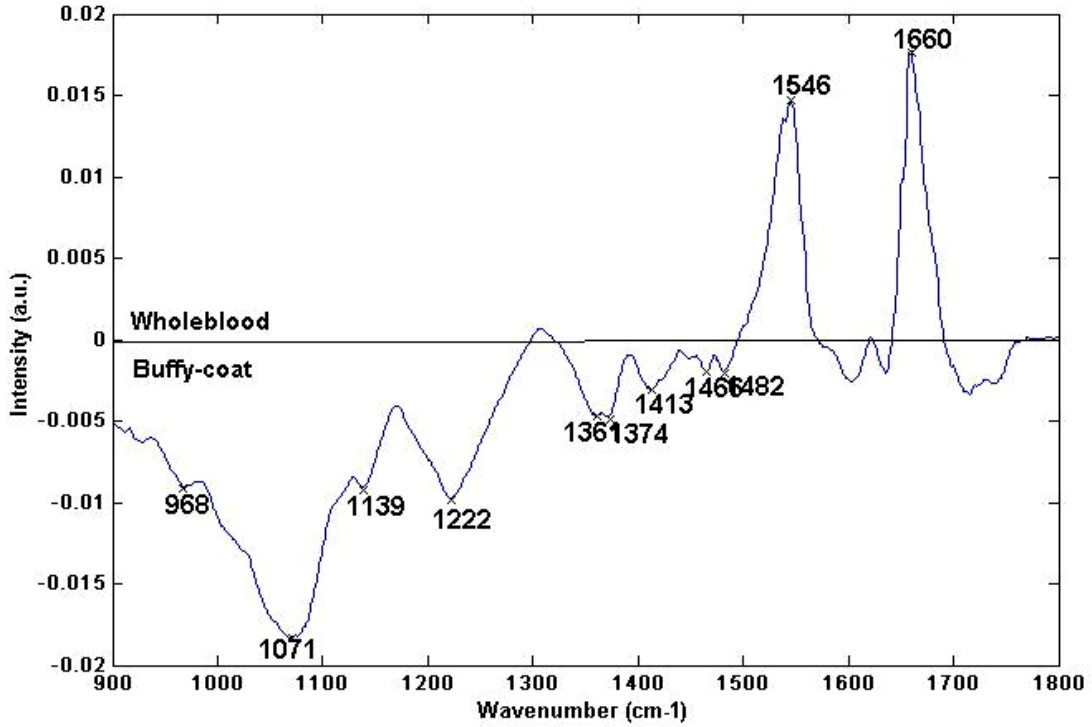


Figure 6-39: Difference spectra obtained by subtraction of whole blood spectrum from the spectrum obtained from buffy-coat

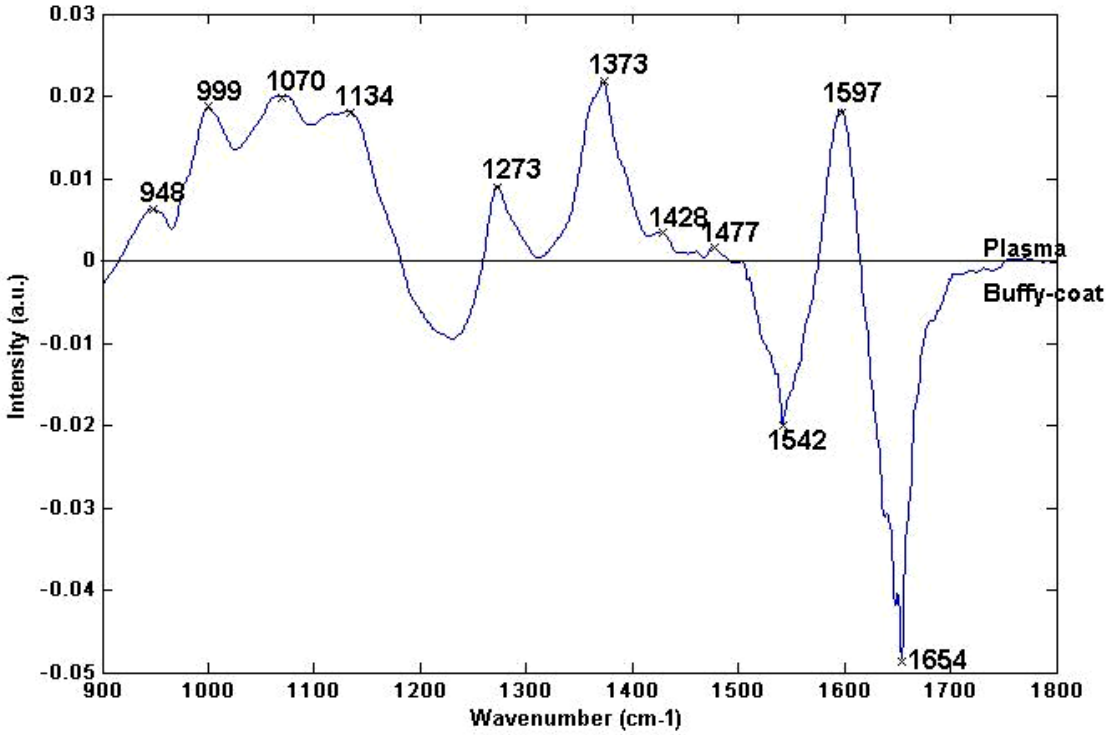


Figure 6-40: Difference spectra obtained by subtraction of plasma spectrum from the spectrum obtained from buffycoat

	Whole blood - Plasma		Whole blood - Buffycoat		Plasma - Buffycoat	
	Peak	**Assignment**	**Peak**	**Assignment**	**Peak**	**Assignment**
Positive Peaks	**1654** **1542**	Amide I Amide II	**1660** **1546**	Amide I Amide II	**948** **999** **1070** **1134** **1273** **1373** **1428-1477** **1597**	??? Phosphorylated proteins, RNA P=O str of $>PO_2^-$ Carbohydrates, Carbohydrates, C-O, C-OH str Carbohydrates, ??? Lipids (CH_3) Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) Amino acids
Negative Peaks	**1003** **1072** **1173** **1272** **1373** **1426-1480** **1597**	Carbohydrates, C-O, C-OH str P=O str of $>PO_2$ Carbohydrates, C-O, C-OH, C-C, C-O-C ??? Lipids (CH_3) Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) Amino acids	**968** **1071** **1139** **1222** **1361** **1374** **1413-1482**	Phosphorylated proteins, RNA P=O str of $>PO_2$ Carbohydrates, C-O, C-OH str Carbohydrates, C-O, C-OH, C-C, C-O-C Amino acids, Lipid (COO, CH_3) Lipids, CH_3 Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates)	**1654** **1542**	Amide I Amide II

Table 6-24: Lists positive and negative peaks obtained from the difference spectra between whole blood, plasma and buffy-coat samples with peak assignment

6.3.1.2 Peak Intensity Ratio: Normal Blood

In order to analyse the distribution of biochemical constituents in whole blood, plasma and buffy-coat spectra and consider variations in overall spectral intensities, peak intensity ratios were calculated in the range 900-1800 cm^{-1}. Specific peaks assigned to proteins, lipids, carbohydrates and nucleic acids have been used to calculate peak intensity ratio. Figure 6-41 shows, the relative peak intensity ratio of the bands assigned to proteins (amide I, II and III) and lipids (1454, 1390 and 1372 cm^{-1}) in various combinations. It has been found that the ratio of peak intensity corresponding to all proteins and lipids peaks differ greatly between whole blood, plasma and buffy-coat spectra. All combinations of protein/lipid peak intensity ratios are very similar and indicate a strong correlation with higher protein distribution in buffy-coat and whole blood compared to that in plasma. These findings are in agreement with the difference spectra represented in Section 6.3.1.1 indicating that whole blood and buffy-coat spectra have higher distribution of protein compared to the plasma spectra.

Figure 6-42 shows, the relative peak intensity ratio of the bands assigned to carbohydrates (1005 and 1140 cm^{-1}) and proteins (amide I, II and III) in different combinations. In this case, such ratios indicate changes between carbohydrate and protein distribution in different blood components. All combinations of carbohydrate/protein peak intensity ratios are very similar and indicate a strong correlation with higher carbohydrate distribution in whole blood and plasma compared to that in buffy-coat. These findings are in agreement with the difference spectra represented indicating that plasma spectra have higher distribution of carbohydrates compared to the spectra for whole blood and buffy-coat.

The relative peak intensity ratio of the bands assigned to nucleic acids (967, 1079 and 1236 cm^{-1}) and proteins (amide I, II and III) in different combinations are demonstrated in Figure 6-43. The ratio of peak intensity corresponding to all nucleic acid and proteins, differ greatly between whole blood, plasma and buffy-coat. All combinations of nucleic acid/protein peak intensity ratios are very similar and indicate a strong

correlation with higher nucleic acid distribution in buffy-coat compared to that in whole blood and plasma.

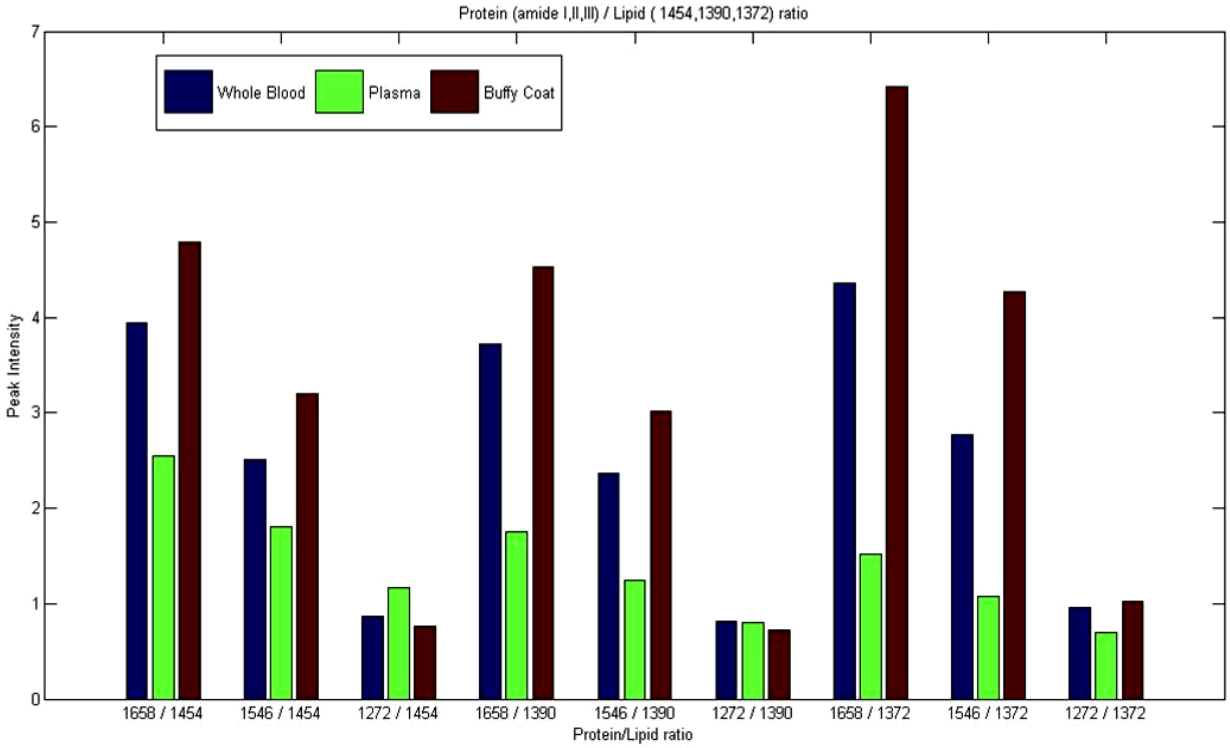


Figure 6-41: Correlation between protein (amide I, II and III) and lipid (1454, 1390 and 1372 cm^{-1}) peak intensity ratios

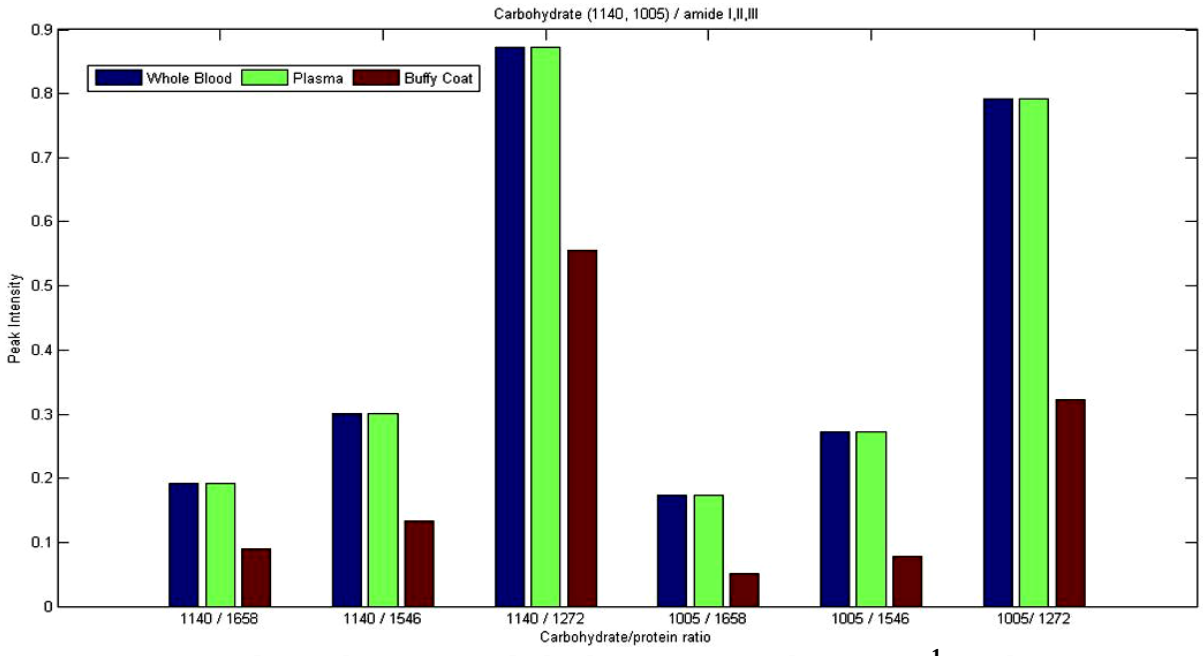


Figure 6-42: Correlation between carbohydrate (1005 and 1140 cm^{-1}) and protein (amide I, II and III) peak intensity ratios

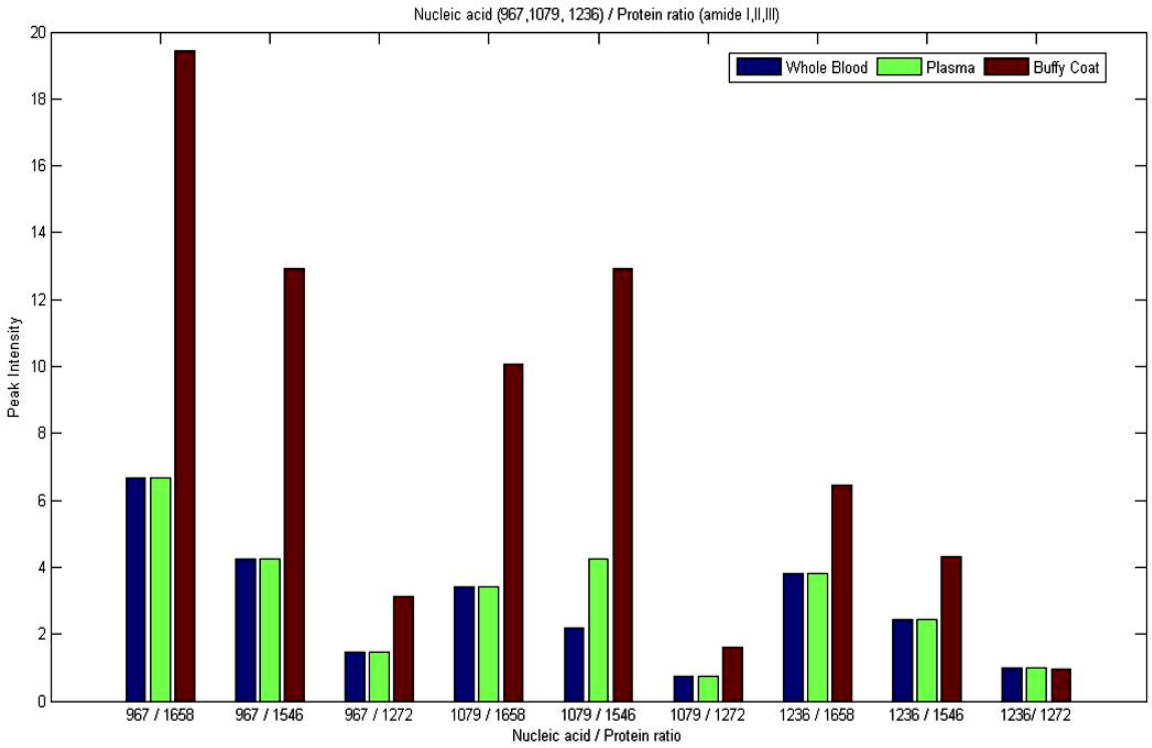


Figure 6-43: Correlation between nucleic acid (967, 1079 and 1236 cm^{-1}) and proteins (amide I, II and III) peak intensity ratios

6.3.2 Empirical analysis of chronic lymphocytic leukaemia (CLL) blood specimens

6.3.2.1 Empirical analysis: CLL Blood

12 blood samples were collected from patients with chronic lymphocytic leukaemia (CLL). Blood components such as plasma and buffy-coat were isolated using standard procedures as described in Section 5.2.3. White light images and spectral measurements were obtained for whole blood, plasma and buffycoat as described in Section 5.5 and 5.9.3.

A background spectrum was obtained from a blank area of the CaF_2 slide and ratioed against the sample spectrum each time. The optimised standard point measurement parameters described in Section 5.8 were used to measure FT-IR spectra from whole blood, plasma and buffycoat samples. Point spectra obtained in transmission mode obtained from the blood component at different regions were averaged using an in-

house Matlab programme. The total number of point spectra collected for whole blood, buffy-coat and plasma are 711, 623, and 338 respectively.

A plot of the mean FT-IR spectra of the 3 separate blood components from CLL blood normalised to the intensity at 1656 cm^{-1} (amide I) is shown in Figure 6-44 for comparison. Significant spectral differences are noted between the three blood components. These were identified using a standard empirical method of assigning key peak positions. Spectrum for individual blood component is sited in Appendix E. The peak positions of the major vibrational modes are labelled. The full list of peak assignments are summarised in Appendix D.

The representative mean spectra for CLL whole blood, plasma and buffy-coat are very similar. However, there is a degree of variations in peak positions and intensities as shown in Figure 6-44. The most distinctive peaks in the IR absorption spectra for whole blood, plasma and buffy coat are assigned to the absorption modes of proteins the most intense is the amide I band, centred near 1656 cm^{-1} for buffycoat and plasma spectra and 1658 cm^{-1} for whole blood spectra. The amide II peak are assigned at approximately 1546/1547 cm^{-1} for whole blood, plasma and buffy coat spectra The amide III band arise at 1233-1242 cm^{-1}. Lipid contributions in whole blood, plasma and buffycoat spectra are attributed at 1452-1467 cm^{-1} and 1307- 1399 cm^{-1} bands. A small peak at 1741 cm^{-1} observed in the buffy-coat spectra may correspond to C=O stretching vibrations from phospholipids in membranes. This peak is absent in the spectra for whole blood and plasma.

Symmetric and anti-symmetric stretching vibrations from the phosphate group are seen in the FT-IR spectra of whole blood and buffy-coat in the region between 950 and 1280 cm^{-1}. For buffy-coat spectra, nucleic acid signals contribute at 1084 cm^{-1}, 1056cm^{-1} and 972 cm^{-1}. For whole blood spectra, nucleic acid signals contribute at 1106 and 1170 cm^{-1}. The peak at 968/972 cm^{-1} is absent in the spectra for whole blood. Nucleic acid contributions are attributed at 1086cm^{-1}, 1055cm^{-1} and 968 cm^{-1}.

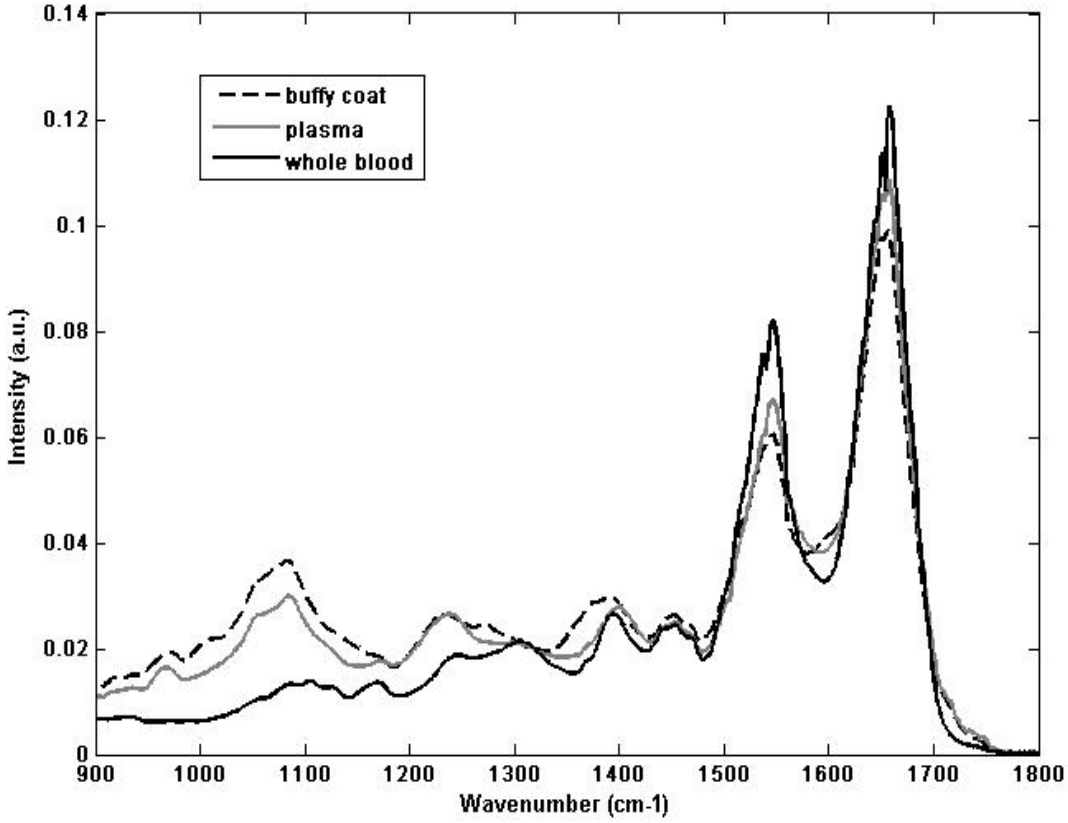


Figure 6-44: Represents the average IR spectra for whole blood, plasma and buffy-coat in the range 900-1800 cm^{-1} for CLL blood

6.3.2.2 Difference Spectra: CLL Blood

The difference spectrum was calculated using an in-house Matlab programme. Figure 6-45 shows the resulting spectra obtained from taking the difference between mean whole blood from plasma. Figure 6-46 shows the resulting spectra obtained from subtracting the mean whole blood spectra from mean buffy-coat spectra. The difference spectra obtained subtracting the mean plasma spectra from mean buffy-coat spectra is represented in Figure 6-47. Table 6-25 lists the biochemical constituents in high concentrations as positive peaks in a certain blood component and those with lower concentrations as negative peaks in other blood component.

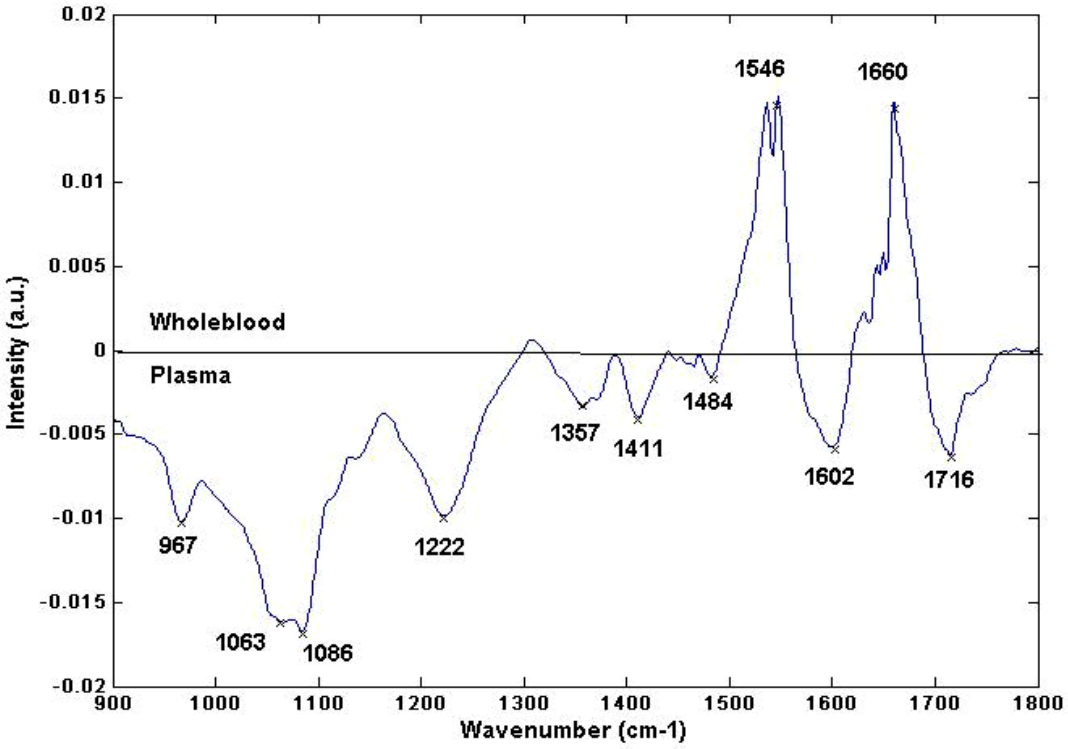


Figure 6-45: Difference spectra obtained by subtraction of whole blood spectrum from the spectrum obtained from plasma

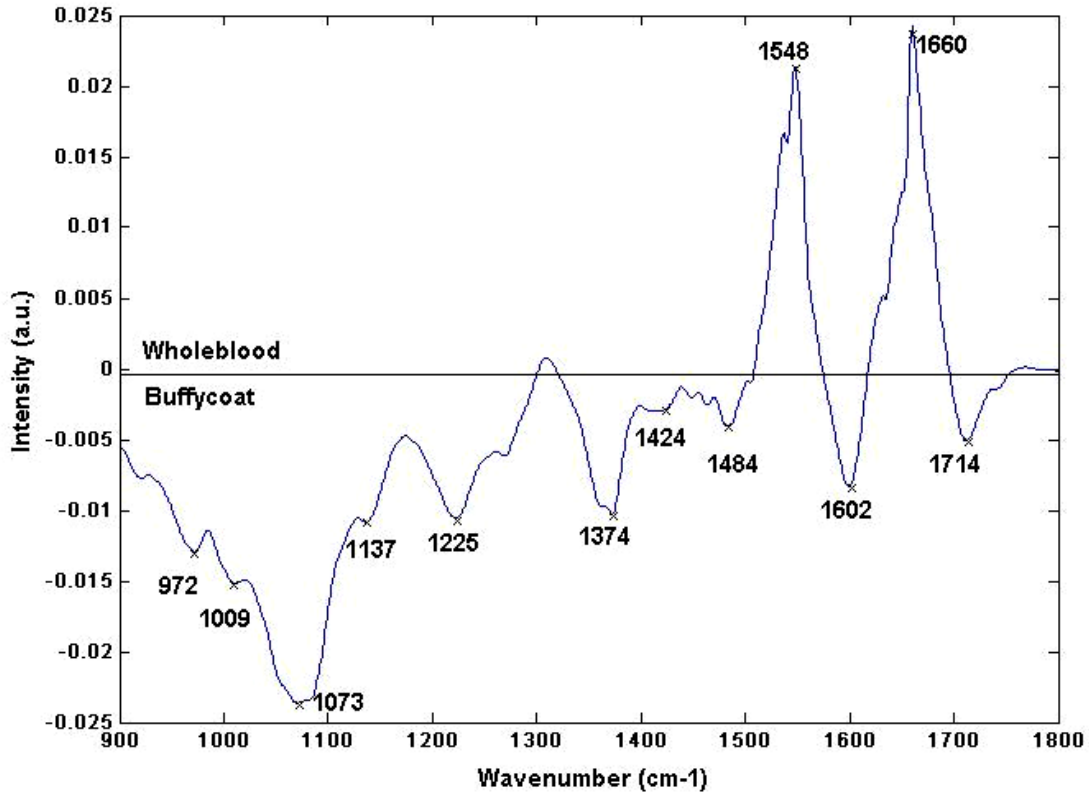


Figure 6-46: Difference spectra obtained by subtraction of whole blood spectrum from the spectrum obtained from buffycoat

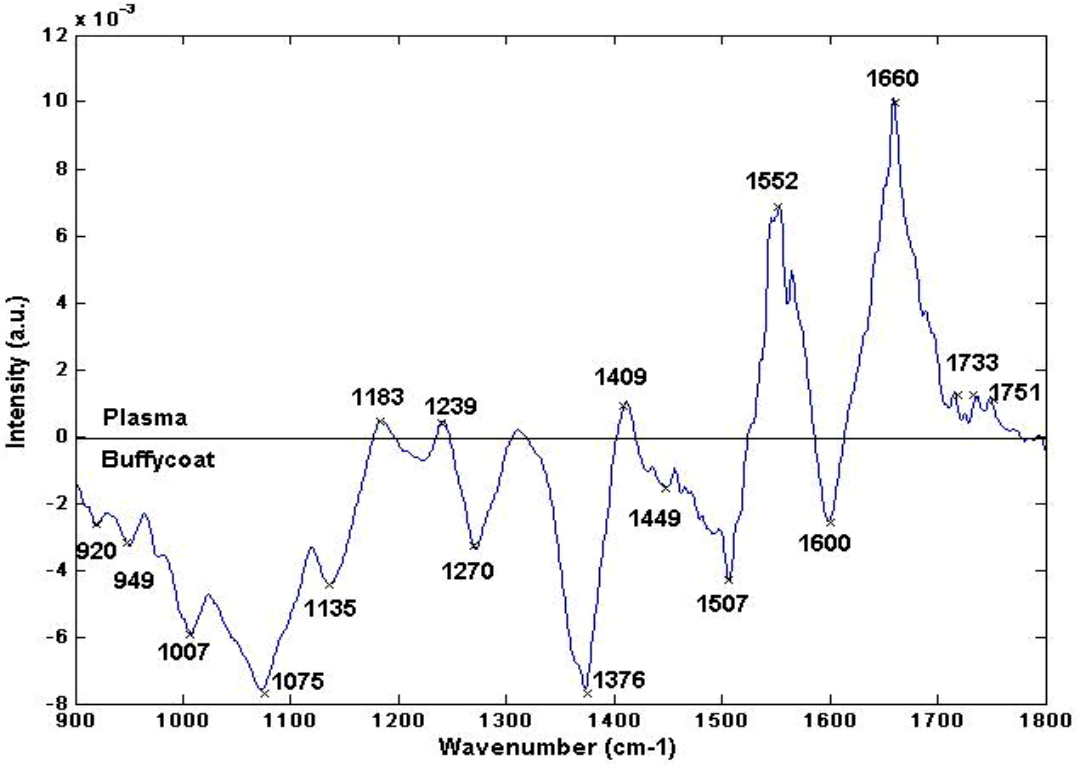


Figure 6-47: Difference spectra obtained by subtraction of plasma spectrum from the spectrum obtained from buffycoat

	Whole blood - Plasma		Whole blood - Buffycoat		Plasma - Buffycoat	
	Peak	**Assignment**	**Peak**	**Assignment**	**Peak**	**Assignment**
Positive Peaks	**1660** **1546**	Amide I Amide II	**1660** **1548**	Amide I Amide II	**1660** **1552** **1733-1751** **1409** **1239** **1183**	Amide I Amide II Lipids, esters, C=O str Lipids (C-H, CH_2) PO_2^- str Carbohydrates, C-O, C-OH str
Negative Peaks	**967** **1063-1086** **1222** **1357** **1411-1482** **1602** **1716**	Phosphorylated proteins, RNA P=O str of $>PO_2$ C-O, C-OH, C-C, C-O-C Amino acids, Lipid (COO, CH_3) Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) ???? C=O in nucleic acids	**972** **1009** **1073** **1137** **1225** **1374** **1424-1484** **1602** **1714**	Phosphorylated proteins, RNA Carbohydrates, C-O, C-OH str P=O str of $>PO_2$ Carbohydrates, C-O, C-OH str Carbohydrates, C-O, C-OH, C-C, C-O-C Lipids, CH_3 Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) ???? C=O in nucleic acids	**920** **949** **1007** **1135** **1270** **1376** **1449** **1507** **1600**	??? ??? Carbohydrates, C-O, C-OH str Carbohydrates, C-O, C-OH str Amide III Lipids, CH_3 Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) ??? ???

Table 6-25: Lists positive and negative peaks obtained from the difference spectra between whole blood, plasma and buffy-coat samples with peak assignment

6.3.2.3 Peak Intensity Ratio: CLL Blood

Figure 6-48 shows, the relative peak intensity ratio of the bands assigned to proteins (amide I, II and III) and lipids (1452 and 1393 cm^{-1}) in various combinations. It has been found that the peak intensity ratio corresponding to all proteins and lipids peaks; differ greatly between whole blood, plasma and buffy-coat spectra. Most combinations of protein/lipid peak intensity ratios are very similar and indicate a strong correlation with higher protein distribution in buffy-coat and whole blood compared to that in plasma. This is not the case in peak ratios at 1236/1452 cm^{-1} and 1236/1393 cm^{-1}. This could be due to the peak assigned at 1236 cm^{-1} has overlapping vibrational modes corresponding to lipids and amide III protein.

The relative peak intensity ratio of the bands assigned to nucleic acid (967, 1079 and 1236 cm^{-1}) and protein (amide I, II and III) in different combinations are demonstrated in Figure 6-49. The ratio of peak intensity corresponding to all nucleic acid and proteins, differ greatly between whole blood, plasma and buffy-coat. All combinations of nucleic acid/protein peak intensity ratios are very similar and indicate a strong correlation with higher nucleic acid distribution in buffy-coat compared to that in whole blood and plasma.

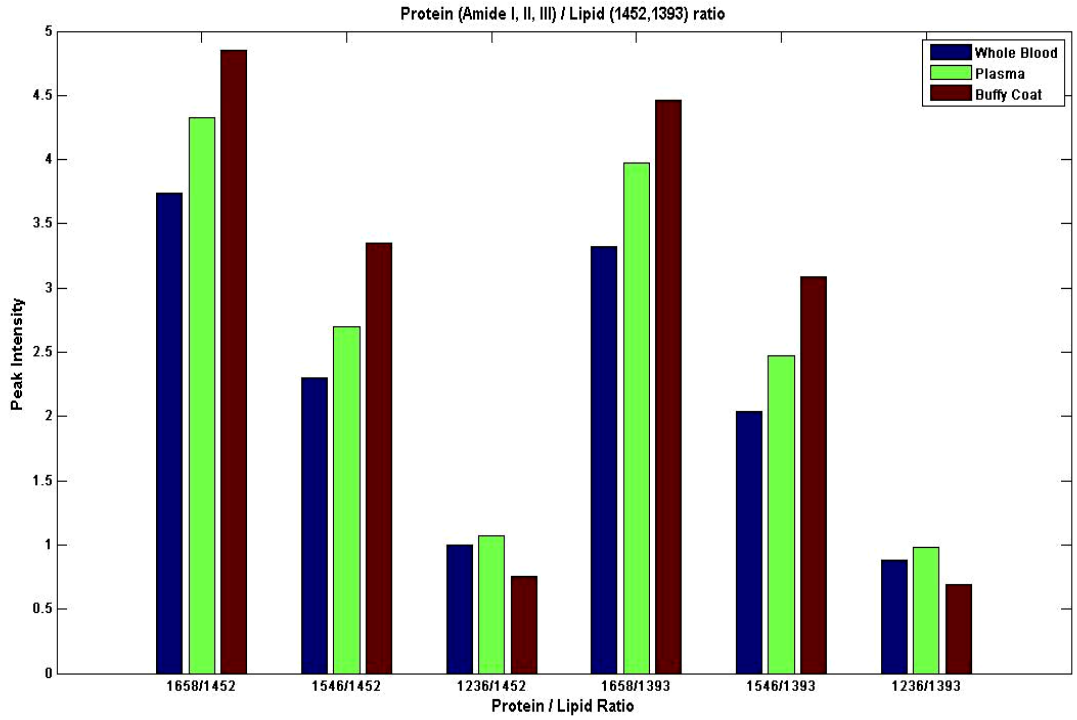


Figure 6-48: Correlation between protein (amide I, II and III) and lipid (1452 and 1393 cm^{-1}) peak intensity ratios

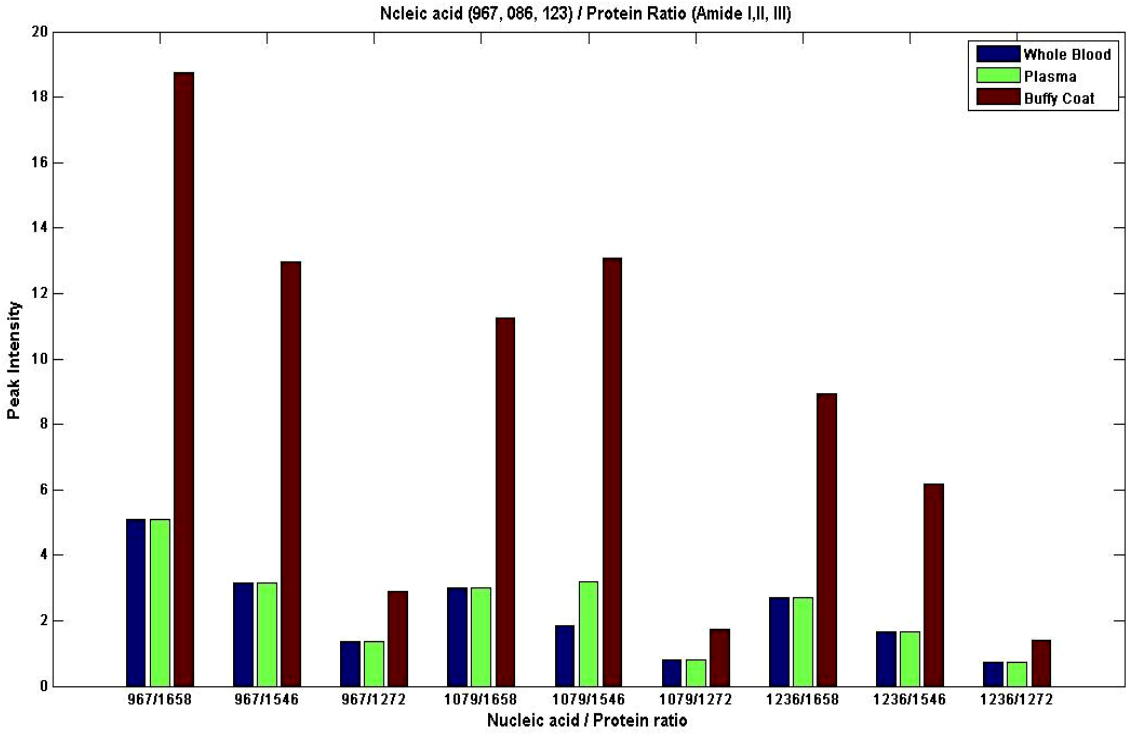


Figure 6-49: correlation between protein (amide I, II and III) and nucleic acid (967, 1079 and 1236 cm^{-1}) peak intensity ratios

6.3.3 Empirical analysis: Normal Blood vs CLL Blood Comparison

The comparison between the FT-IR spectra of normal whole blood and CLL whole blood is shown in Figure 6-50. Figure 6-51 represent the comparison between the FT-IR spectra of normal plasma and CLL plasma. The comparison between the FT-IR spectra of normal buffycoat and CLL buffycoat is shown in Figure 6-52. The full list of peak assignments are summarised in Appendix D.

The representative mean spectra for normal whole blood and CLL whole blood are very similar with very little variations in peak intensities as shown in Figure 6-50. A degree of variation throughout the fingerprint region (900-1800 cm^{-1}) for spectra belonging to normal and CLL plasma is demonstrated (Figure 6-51). Significant band shifts and changes in peak intensity are observed. Major band shifts are seen in peak at 1078 cm^{-1} to 1082 cm^{-1} and 1371 cm^{-1} to 1401 cm^{-1} in the CLL plasma spectra. The significant reduction in the intensity of various peaks corresponding to lipids, carbohydrates, amide III and amino acids indicate a general reduction in plasma metabolites, higher metabolic activity and lower levels of nutrition in leukaemia cells. These finding coincide with a study by Erukhimovitch 2005, which reported reduction in the intensity of peak assigned at 1056 cm^{-1} (carbohydrates), 1270 cm^{-1} (Amide III) and 1592 cm^{-1} (amino acids) in plasma samples from patients suffering from CLL compared to healthy subjects.

Changes in peak intensities are also seen between normal buffycoat and CLL buffycoat spectra as represented in Figure 6-52. The most significant changes are seen in the 900-1200 cm^{-1} region, corresponding to the vibrational modes of nucleic acids. This indicates that the concentration of nucleic acids increase in leukaemia cells due to DNA replication. The increase in the nucleic acid in CLL buffycoat coincides with study by Boydston-White 1999, which suggest that concentration of nucleic acids increase in leukaemia cells during the replication of DNA. Changes in intensities are also observed in Amide I and II peaks.

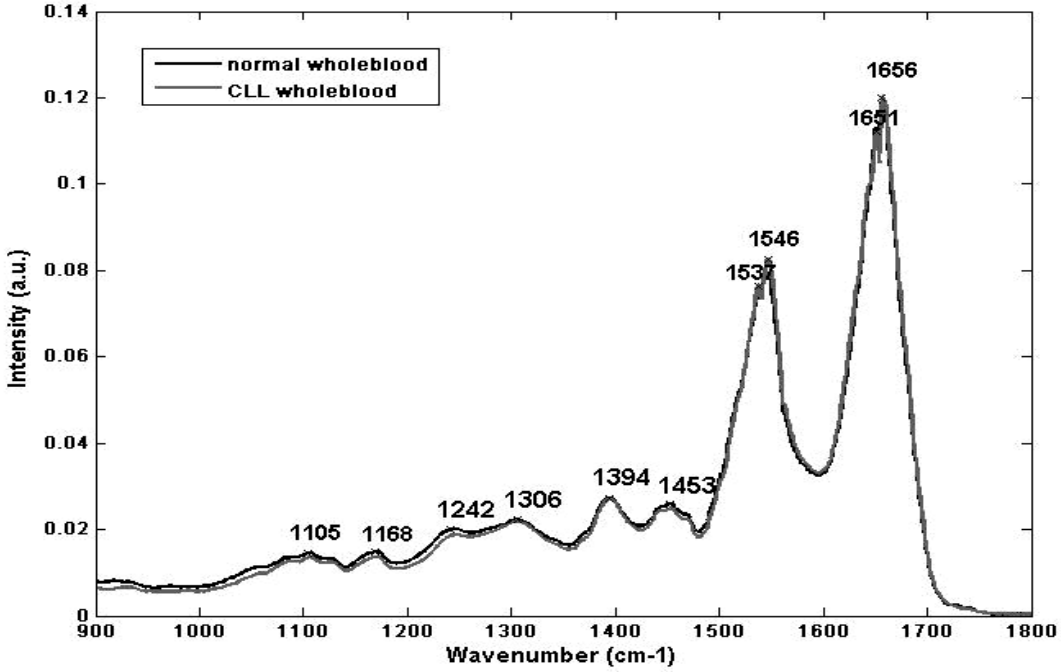


Figure 6-50: Represents the average IR spectra for normal whole blood and CLL whole blood in the range 900-1800 cm^{-1}

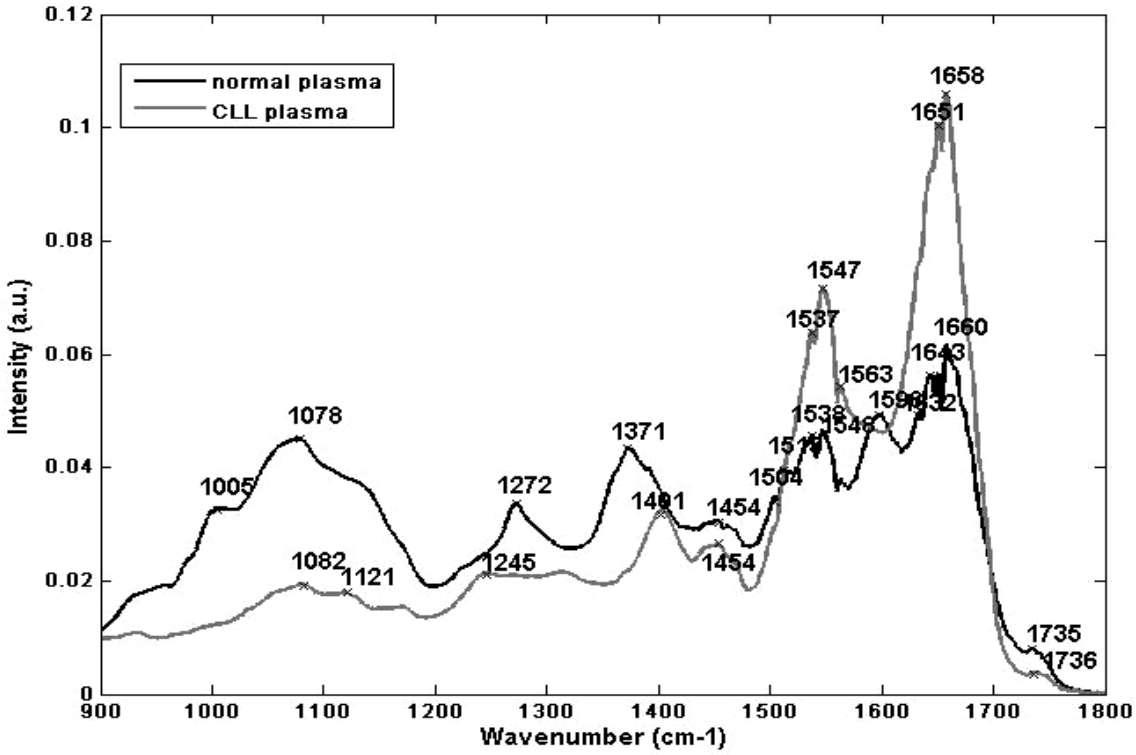


Figure 6-51: Represents the average IR spectra for normal plasma and CLL plasma in the range 900-1800 cm^{-1}

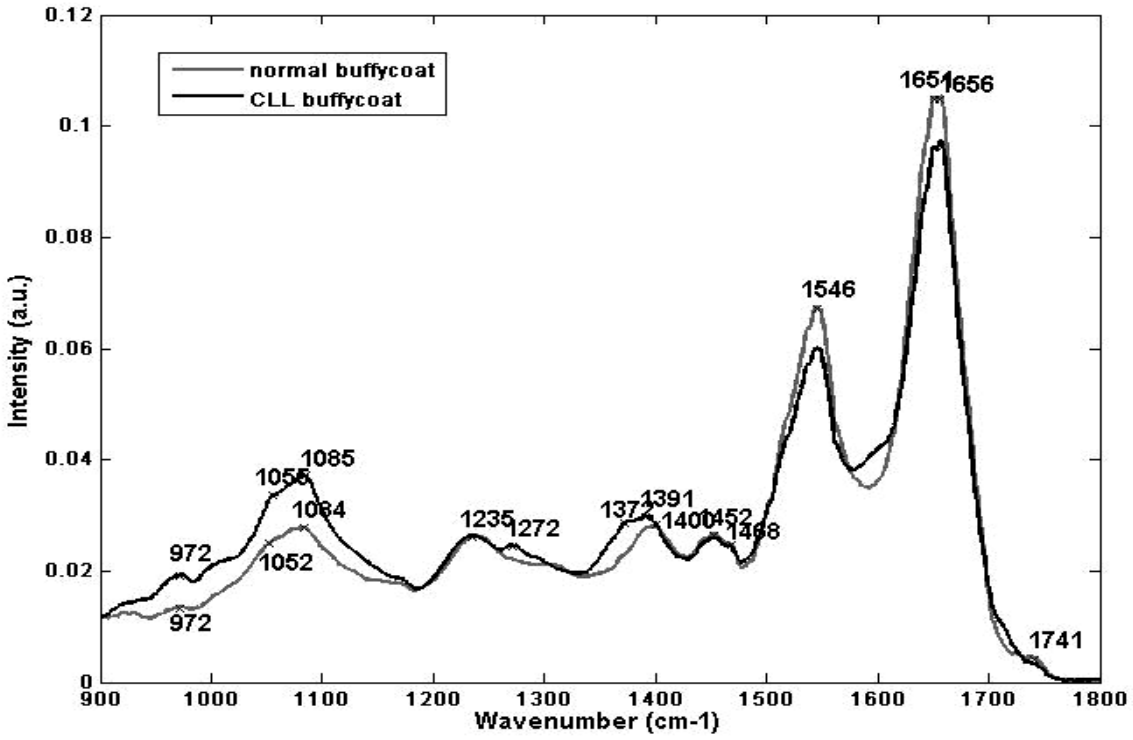


Figure 6-52: Represents the average IR spectra for normal buffycoat and CLL buffycoat in the range 900-1800 cm^{-1}

6.3.3.1 Difference Spectra: Normal vs CLL Blood

The difference spectrum was calculated using an in-house Matlab programme. Figure 6-53 shows the resulting spectrum obtained from subtracting the mean normal whole blood spectrum from mean CLL whole blood spectrum. Figure 6-54 shows the resulting spectra obtained from taking the difference between mean normal plasma spectrum and mean CLL plasma spectrum. The difference spectra obtained subtracting the mean normal buffycoat spectra from mean CLL buffy-coat spectra is represented in Figure 6-55. Table 6-26 lists the biochemical constituents in high concentrations as positive peaks in a certain blood component and those with lower concentrations as negative peaks in other blood component.

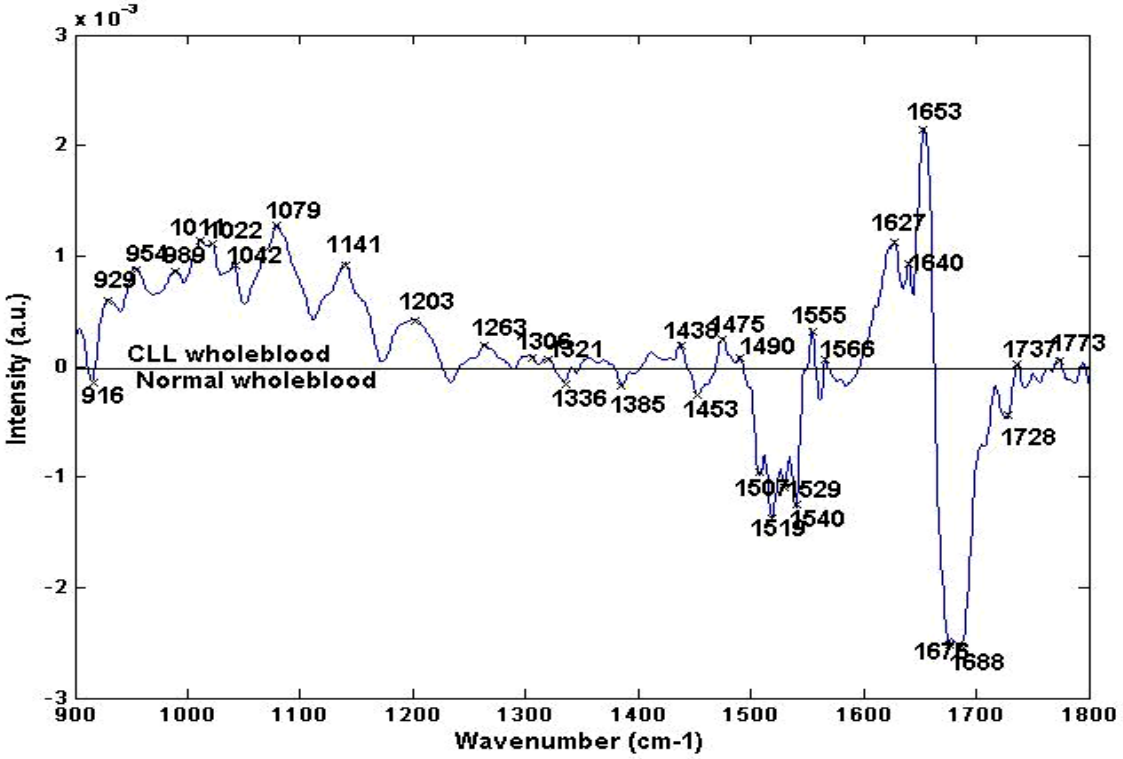


Figure 6-53: Difference spectra obtained by subtraction of normal whole blood spectrum from the spectrum obtained from CLL whole blood

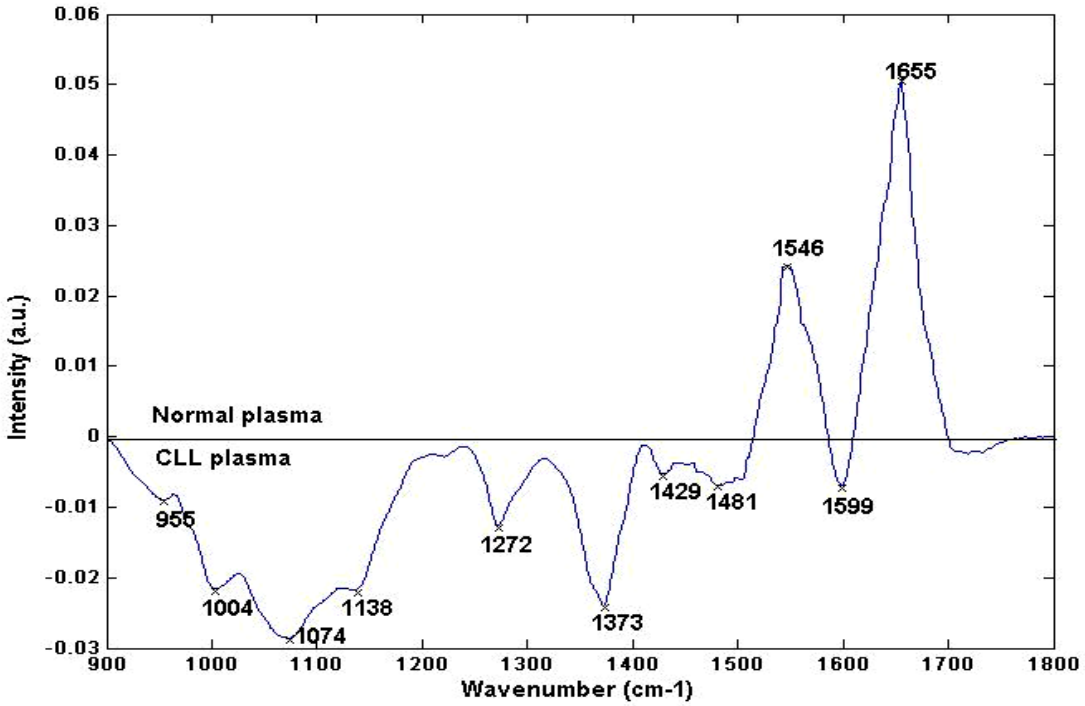


Figure 6-54: Difference spectra obtained by subtraction of normal plasma spectrum from the spectrum obtained from CLL plasma

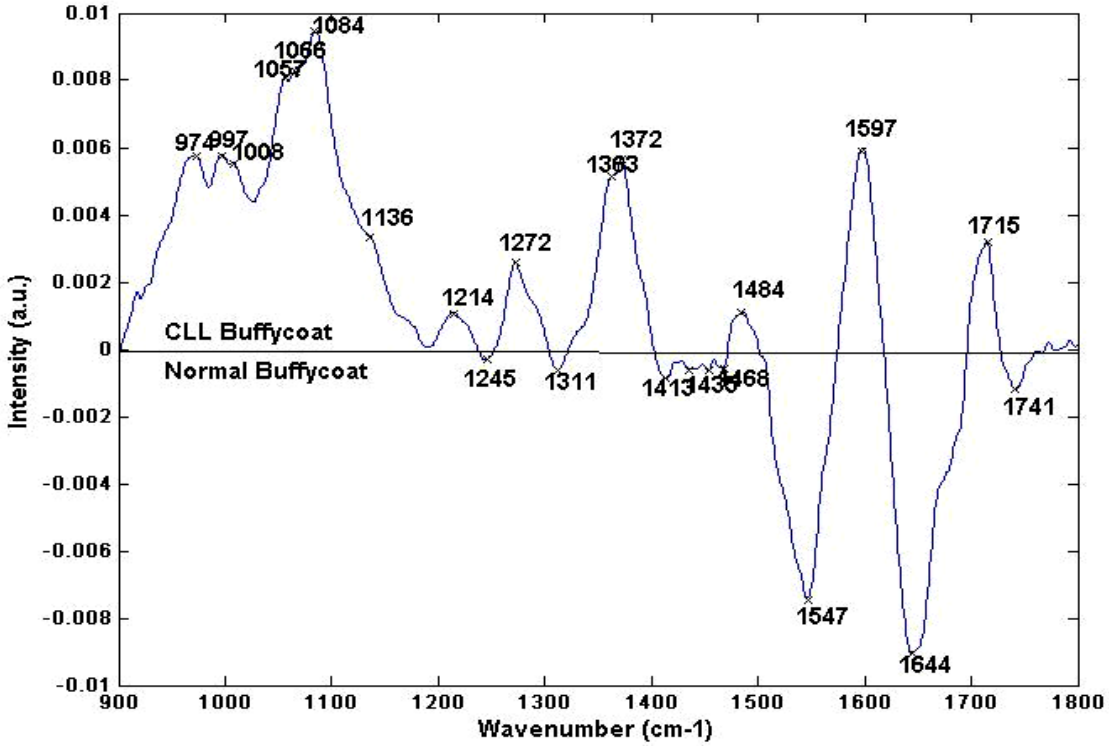


Figure 6-55: Difference spectra obtained by subtraction of normal buffycoat spectrum from the spectrum obtained from CLL buffycoat

	Normal–CLL whole blood		Normal–CLL Plasma		Normal–CLL buffycoat	
	Peak	**Assignment**	**Peak**	**Assignment**	**Peak**	**Assignment**
Positive Peaks	**929** **954** **989** **1011** **1022** **1042** **1079** **1141** **1203** **1263** **1306** **1321** **1438-1490** **1555-1565** **1627** **1640** **1653** **1737-1773**	??? ??? Phosphorylated proteins, RNA Carbohydrates, C-O, C-OH str ??? Carbohydrates, C-OH P=O str of $>PO_2$ Carbohydrates, C-O, C-OH str ??? Carbohydrates, C-O, C-OH, C-C, C-O-C ??? Lipids (CH_3) Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) Amide II ??? ??? Amide I Lipids, esters, C=O str	**1655** **1546**	Amide I Amide II	**974-997** **1008** **1057** **1066-1084** **1136** **1214** **1272** **1383-1372** **1484** **1597** **1715**	Phosphorylated proteins, RNA Carbohydrates, C-O, C-OH str Carbohydrates, C-OH P=O str of $>PO_2$ Carbohydrates, C-O, C-OH str ??? Carbohydrates, C-O, C-OH, C-C, C-O-C Amino acids (C-O of COO-) Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) Amide II C=O in nucleic acids
Negative Peaks	**916** **1336** **1385** **1453** **1507** **1515** **1529** **1540** **1675** **1688** **1728**	??? Lipids (CH_3) Amino acids (C-O of COO-) Lipid (CH def of $>CH_2$) ??? ??? ??? ??? Amide I ??? Lipids, esters, C=O str	**955** **1004** **1074** **1138** **1272** **1373** **1429-1481** **1599**	Phosphorylated proteins, RNA Carbohydrates, C-O, C-OH str P=O str of $>PO_2$ Carbohydrates, C-O, C-OH str Carbohydrates, C-O, C-OH, C-C, C-O-C Lipids, CH_3 Lipids (C-H, CH_2 wagging vib, CH_3 sym bending vib and C-O str in carbohydrates) ???	**1245** **1311** **1413** **1438** **1547** **1644** **1741**	Carbohydrates, C-O, C-OH, C-C, C-O-C ??? Amino acid C-O of sym COO- Carbohydrates, C-O, C-OH str Carbohydrates, C-O, C-OH str Amide II Amide I Lipids, esters, C=O str

Table 6-26: Lists positive and negative peaks obtained from the difference spectra between normal and CLL whole blood spectrum; normal and CLL plasma spectrum and normal and CLL buffycoat spectrum with peak assignment

6.3.3.2 Peak Intensity Ratio: Normal vs CLL Blood

Figure 6-56 shows, the relative peak intensity ratio of the bands assigned to proteins (amide I, II and III) and lipids (1454 and 1390 cm^{-1}) in various combinations for the three blood components of normal and CLL blood. It has been found that the peak intensity ratios corresponding to all proteins and lipids peaks; differ greatly between; normal and CLL whole blood, normal and CLL plasma and normal and CLL buffy-coat spectra. In general, a slightly higher protein/lipid distribution is seen in normal whole blood spectra compared to that of CLL whole blood spectra. The protein/lipid distribution in normal plasma spectra is significantly higher than in spectra of CLL plasma. The reduction in the protein/lipid peak intensity ratio in CLL plasma coincides with study by Erukhimovitch 2005 as previously discussed in Section 5.3.3. A slightly higher protein/lipid distribution is seen in CLL buffycoat spectra compared to that of normal buffycoat spectra. This is not evident in the protein/lipid peak ratios at 1272/1454, 1658/1390 and 1272/1390 cm^{-1}.

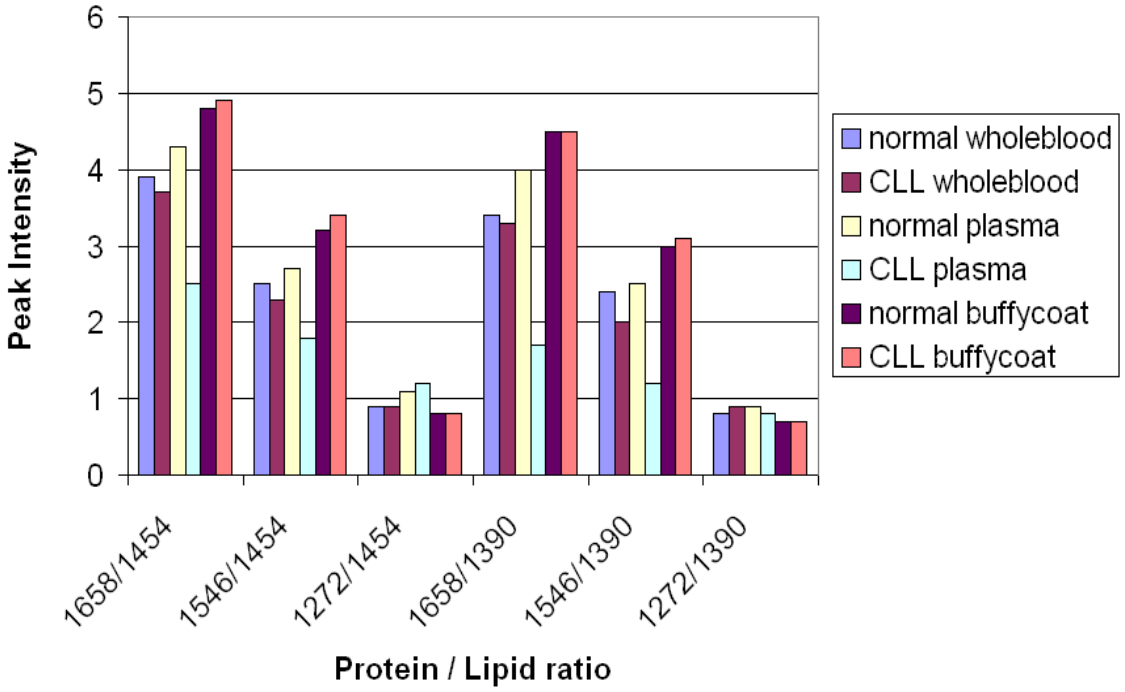


Figure 6-56: correlation between protein (amide I, II and III) and lipid (1454 and 1390 cm^{-1}) peak intensity ratios in the three blood components (whole blood, plasma and buffycoat) of normal and CLL blood

Figure 6-57 shows, the relative peak intensity ratio of the bands assigned to nucleic acids (967, 1079 and 1236 cm^{-1}) and proteins (amide I, II and III) in various combinations for the three blood components of normal and CLL blood. In general, a higher nucleic acid/protein distribution is seen in CLL whole blood spectra compared to that of normal whole blood spectra. The nucleic acid/protein distribution in normal plasma spectra is higher than in spectra of CLL plasma. There is a significant increase in the nucleic acid/protein distribution in CLL buffycoat compared to that of normal buffycoat spectra. The increase in the nucleic acid/protein peak intensity ratio in CLL buffycoat coincides with study by Boydston-White 1999 as previously discussed in Section 5.3.3.The observed results suggest that concentration of nucleic acids increase in leukaemia cells during the replication of DNA.

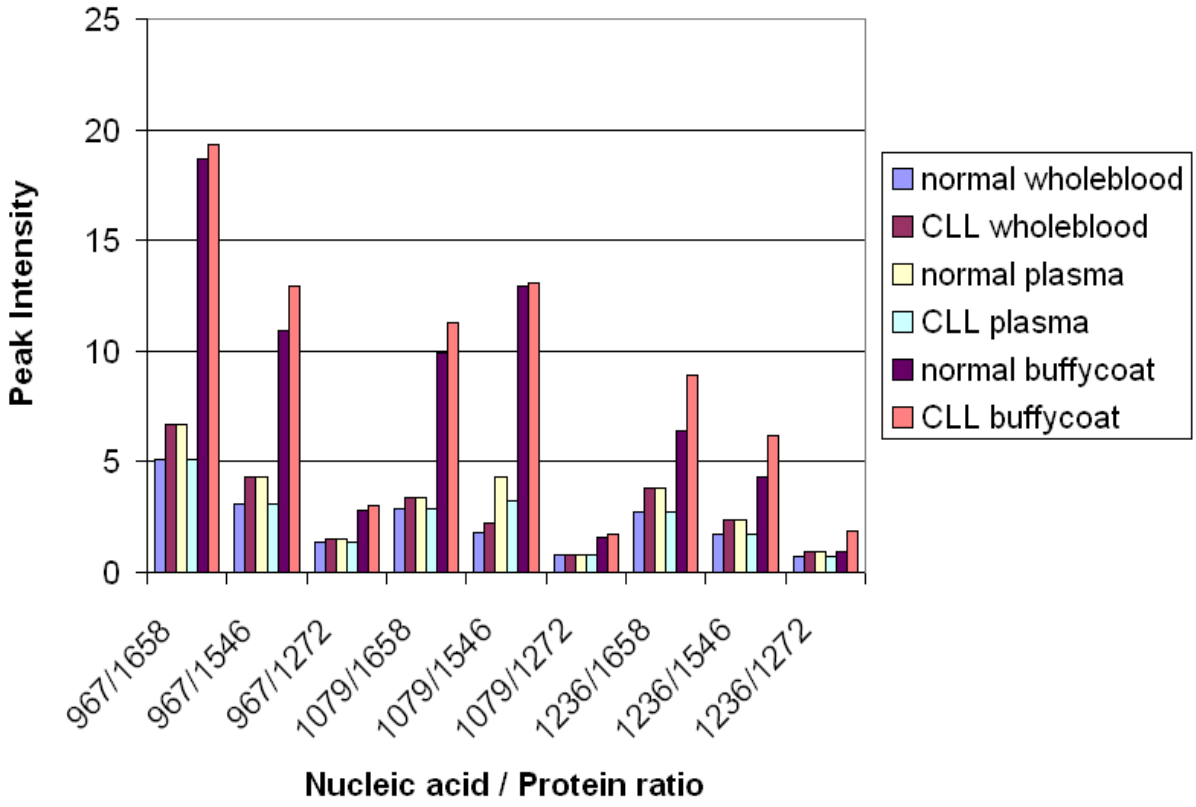


Figure 6-57: correlation between nucleic acid (967, 1079 and 1236 cm^{-1}) and protein (amide I, II and III) peak intensity ratios in the three blood components (whole blood, plasma and buffycoat) of normal and CLL blood

6.3.4 Multivariate Analysis

All point spectra were combined into clinically significant groups for the construction of a multivariate linear discriminant model. Initially, PCA was used to generate data for linear discriminant analysis as described in Section 4.2.2.1. LDA is explained in Section 4.2.2.2. All multivariate analysis is performed using in-house Matlab programmes. Samples from the normal and CLL pathological groups were rearranged into the three clinically significant subsets:

1) Dataset 1 comprised of only whole blood point spectra from the normal and CLL groups.
2) Dataset 2 comprised of only plasma point spectra from the normal and CLL groups.
3) Dataset 3 comprised of only buffycoat point spectra from the normal and CLL groups.

Subsequently classification models were constructed using these three clinical subsets. To interpret biochemical variation, ANOVA was performed on all datasets to demonstrate principal components with maximal variance. Table 1 situated in Appendix F reports principal components with highest F values after ANOVA of the first 25 PC for datasets 1, 2 and 3. Table 1 also reflect the corresponding % variance. The most significant PC are plotted against each other in Figures 1-3 located in Appendix F to observe the degree of natural separation between blood groups. Fcrit for datasets 1, 2 and 3 are 6.69, 6.72 and 6.61 respectively.

All 25 of the calculated PCs were used as input into an individual LDA model to maximise the variance between the normal and CLL blood groups and minimise the variance between two blood groups. This required one linear discriminant function for optimal separation of the blood groups. Two-dimensional scatter plot of the linear discriminant scores from subsets 1, 2 and 3 are displayed in Figures 6-58 – 6-60. The efficiency of the PCA-fed LDA classification model was evaluated using leave-one-sample-out cross validation. The classification results, reflecting the number of spectra correctly classified into either normal blood or CLL blood group in datasets 1, 2 and 3 as well as their sensitivity and specificity are presented in Tables 6-27 – 6-29.

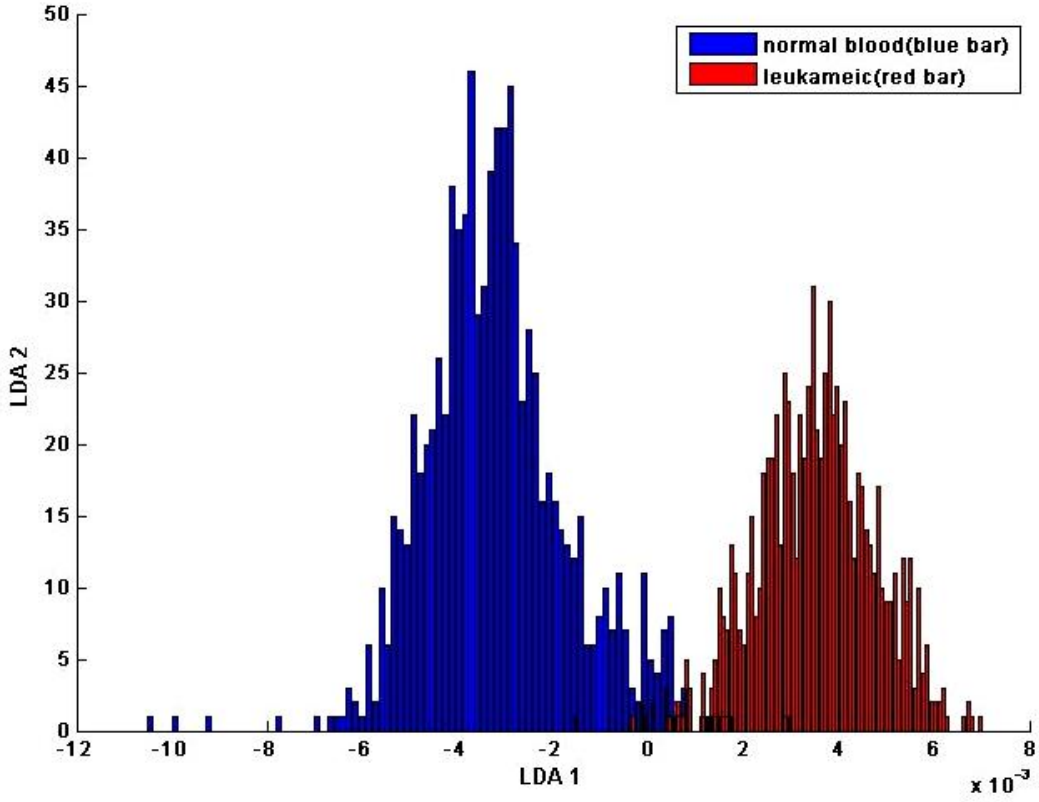


Figure 6-58: Histogram of LDA results for dataset 1

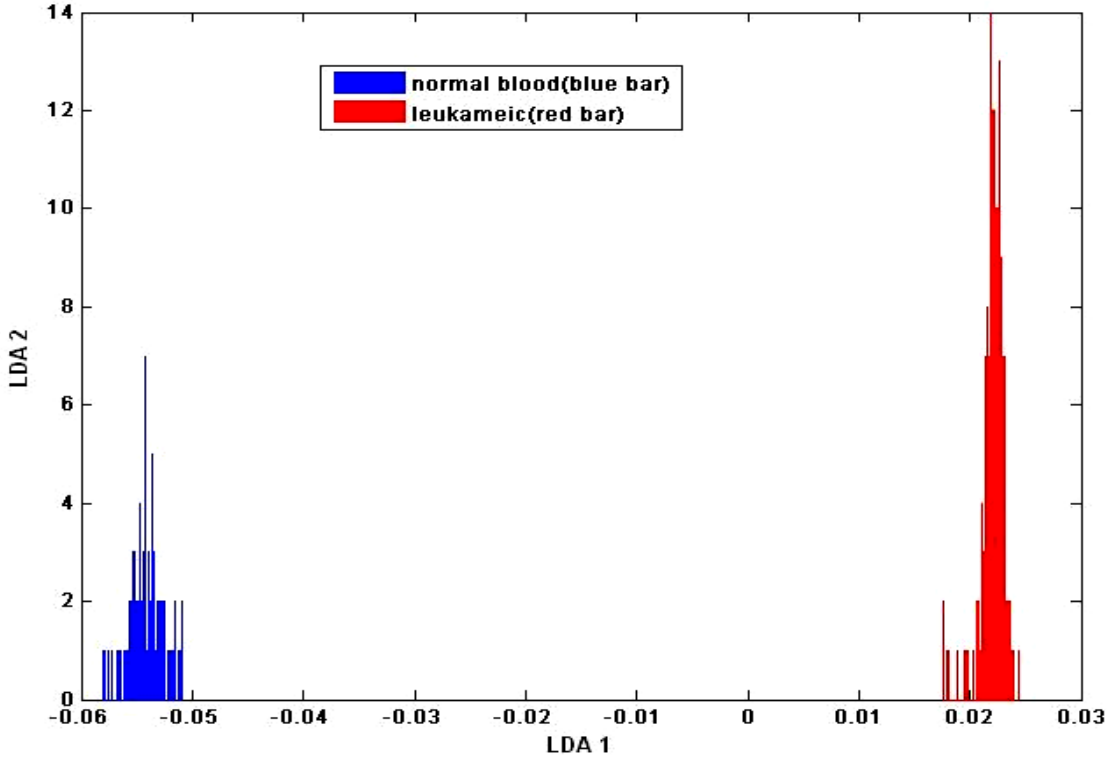


Figure 6-59: Histogram of LDA results for dataset 2

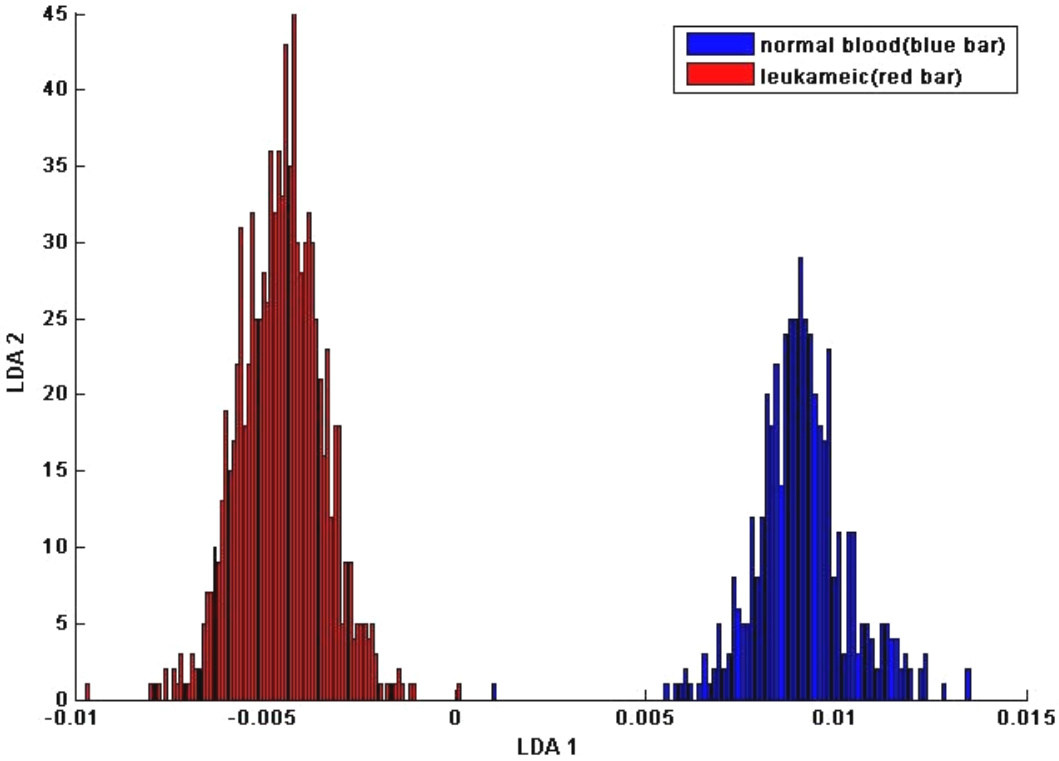


Figure 6-60: Histogram of LDA results for dataset 3

Blood Group	Blood Group Classification		
	Normal whole blood	CLL whole blood	Total No.
Normal whole blood	**545**	366	**911**
CLL whole blood	170	**655**	**825**
Number correct	**545**	**655**	
			1736
			Average
Sensitivity (%)	**59.8**	**79.3**	**69.6**
Specificity (%)	**79.4**	**59.8**	**69.6**

Table 6-27: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 1. 70% spectra are correctly classified

Blood Group	Blood Group Classification		
	Normal plasma	**CLL plasma**	**Total No.**
Normal plasma	**32**	64	**96**
CLL plasma	0	**235**	**235**
Number correct	**32**	**235**	
			331
			Average
Sensitivity (%)	**33.3**	**100**	**66.7**
Specificity (%)	**100**	**33.3**	**66.7**

Table 6-28: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 2. 80.6% spectra are correctly classified

Blood Group	Blood Group Classification		
	Normal buffycoat	**CLL buffycoat**	**Total No.**
Normal buffycoat	**473**	7	**480**
CLL buffycoat	1	**958**	**959**
Number correct	**473**	**958**	
			1438
			Average
Sensitivity (%)	**98.5**	**99.9**	**99.2**
Specificity (%)	**99.9**	**98.5**	**99.2**

Table 6-29: Summary of leave-one-sample-out cross validation results and sensitivity and specificity values for dataset 3. 99.5% spectra are correctly classified

6.4 PLS Analysis

In this study prior to spectroscopic measurements, Full Blood Count (FBC) was taken at the Haematology Laboratory for each buffycoat sample as a standard procedure for measuring concentration of cellular analytes. Concentrations of six cellular analytes were measured as a result of full blood count as shown in Table 6-30. These include neutrophils, lymphocytes, monocytes, eosinophils and basophils. After spectroscopic measurements and the pre-processing of spectra, partial least squares (PLS) regression was performed on buffycoat spectra from normal and CLL blood samples. It was used to construct a linear predictive model for each cellular analytes amount based on the spectrum obtained from normal and CLL buffycoat samples.

	Cellular Analyte [10^9 /L]				
Sample Number	Neutrophil	Lymphocyte	Monocyte	Eosinophil	Basophil
Normal Buffy Coat					
1	3.04	2.87	0.41	0.18	0.07
2	3.83	2.37	0.38	0.52	0.02
3	3.99	1.71	0.83	0.08	0.01
4	2.46	1.96	0.4	0.2	0.01
5	7.02	2.36	0.55	0.2	0.02
6	5.71	3.14	0.51	0.15	0.02
7	3.65	2.58	0.39	0.07	0.02
8	4.24	2.14	0.31	0.07	0.02
9	2.14	1.82	0.41	0.13	0.02
10	2.49	2.07	0.49	0.21	0.03
11	6.55	2.53	0.56	0.16	0.05
12	3.13	2.91	0.41	0.1	0.05
CLL Buffycoat					
1	7.31	0.39	0.88	0.23	0.01
2	0.68	3.36	0.51	0.08	0.01
3	3.2	1.72	0.86	0.5	0.1
4	2.22	0.56	0.07	0.06	0.01
5	2.05	35.22	0.59	0.03	0.04
6	4.46	20.83	0.44	0	0.02
7	0	0	0	0.19	0.03
8	0	0	0	0.47	0.22
9	0	0	0	0.35	0.24
10	3.17	3.41	0.7	0.09	0.01
11	4.04	21.2	1.19	0.29	0.05
12	0.09	272.6	5.99	0.13	2.4

Table 6:30: Concentrations of 6 cellular analytes for normal buffycoat samples and CLL buffycoat samples determined by standard full blood count assessment

The spectral data and analyte concentrations were mean-centred prior to calibration of the training set. Example of a root mean square error as a result of PLS analysis of infrared spectra is illustrated in Table 6-31 for all five cellular analytes. Figure 6-61 and Table 6-31 shows variance captured by 20 latent variables (LV). The optimum number of LVs required to capture variance is determined through calculating the PRESS values. The first 10 LV's capture most of the variance. The X-Block (predictor variable) captured higher % variance compared to that captured by the Y-Block (predicted variable).

	X Block		Y Block	
LV	This LV	Total	This LV	Total
1	44.19	44.19	9.89	9.89
2	14.62	58.81	13.44	23.34
3	11.15	69.96	14.65	37.99
4	11.55	81.52	4.46	42.45
5	6.55	88.06	6.93	49.37
6	2.94	91.01	10.18	59.55
7	2.00	93.01	5.05	64.6
8	2.56	95.57	2.83	67.43
9	1.15	96.72	4.76	72.2
10	0.36	97.08	3.14	75.33
11	0.33	97.42	3.24	78.58
12	0.30	97.72	2.59	81.17
13	0.27	97.99	2.39	83.57
14	0.16	98.15	2.06	85.62
15	0.33	98.48	0.57	86.19
16	0.09	98.57	1.57	87.77
17	0.15	98.72	0.73	88.49
18	0.07	98.79	1.24	89.73
19	0.06	98.85	0.97	90.7
20	0.11	98.96	0.44	91.14

Table 6-31: Percent Variance Captured by PLS Model by 20 LVs

After the training had been performed on the calibration set and the LV's had been determined, the spectra of the validation set were subjected to cross validation. Data was divided into contiguous blocks during the cross-validation i.e. normal (the calibration/training dataset) and leukaemia (the validation dataset) so that data is serially correlated. The calibration (training) dataset comprised of 437 spectra (10 normal buffycoat samples) and the validation dataset consisted of 627 spectra (12 CLL blood samples). The cellular analyte concentrations of the 12 CLL samples investigated by FT-IR spectroscopy were predicted using PLS analysis. The difference between the predicted and the actual concentration is illustrated for all 5 cellular analytes in Figure 6-62. Each symbol on the plot shows the samples in the test set only. The cumulative error for each analyte is represented as a) 0.6949, b) 0.6600, c) 0.8534, d) 0.8424 and e) 0.9158. The average sum squared error is 1.411.

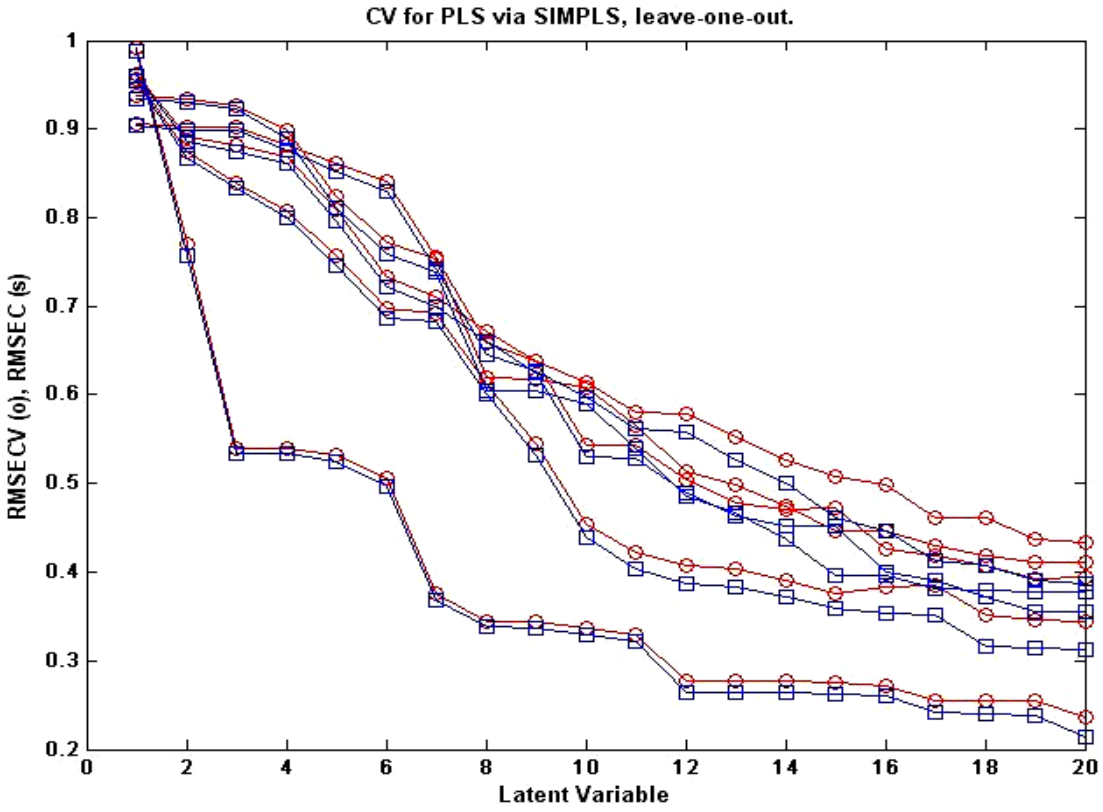


Figure 6-61: Shows trends in RMSECV and RMSEC with increasing number of latent variables for neutrophil, lymphocyte, monocyte, eosinophil and basophil qualitative model. The first 10 LVs capture most of the variance and chosen as optimal

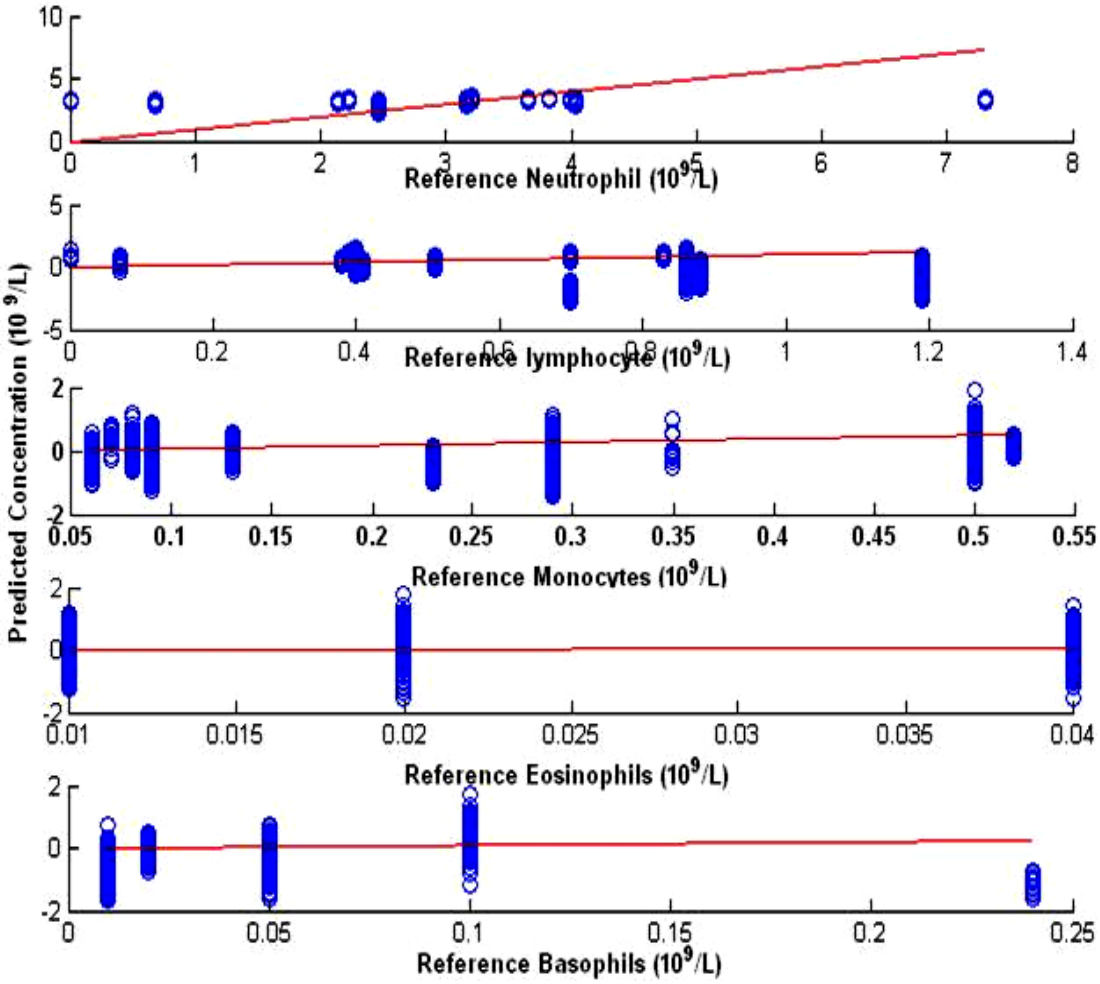


Figure 6-62: Scatterplots comparing blood a) neutrophil, b) lymphocyte, c) monocyte, d) eosinophil and e) basophil concentrations predicted on the basis of the infrared spectra of dry buffycoat films to actual concentration. Each circle corresponds to the validation test set and the solid line is the line of identity

The preliminary PLS analysis used in this study used concentrations obtained from full blood count (FBC) measurements. FBC values obtained from both normal and CLL buffycoat samples show very little variations in concentration values. Hence, PLS in this case is not accurate and has not been able to predict the true concentration values. Many studies have made use of PLS for quantitative clinical applications in predicting concentration of proteins, glucose triglycerides, urea, and cholesterol in serum (Shaw 2000; Petibios 1999) and plasma (Deleris 2003). The quantification of cellular analytes such as neutrophil, lymphocyte, monocyte, eosinophil and basophil in dried buffycoat film is a new concept investigated in this thesis, which has not been addressed in the literature to date.

The idea behind this initial sub-study was to determine whether concentrations of different cellular analytes including neutrophil, lymphocyte, monocyte, eosinophil and basophil can be predicted in CLL samples. This may reveal important characteristic features including:

- origin: whether the leukaemia or lymphoma arise from a lymphoid progenitor or myeloid progenitor

- Type of cell that is involved in differentiation and maturation: whether the cells involed are lymphocytes (T cells and B cells) or myeloids (basophils, megakaryocytes, neutrophils, monocytes, and esinophils).

These characteristic features may further help clinicians to accurately screen cellular profile of blood samples and classify the type of leukaemia or lymphoma depending on cell lineages; determine the stage cell differentiation and maturation and further understand disease progression. It can be concluded that the quantitative analysis of cellular analytes in dried buffycoat films based on FT-IR spectroscopy and PLS analysis cannot as yet compete with present day laboratory diagnostics. Currently, the PLS model constructed is not complete due to time restrictions and is imperfect. Much work is required in understanding the concept of PLS. This is briefly outlined in the future work section.

6.5 Summary

Significant spectral differences were noted between the three blood components of normal and CLL blood including whole blood, buffycoat and plasma. Further spectral variations between the different blood components were emphasised using difference spectra. Peak intensity ratios were found to be in agreement with the difference spectra indicating higher protein distribution in buffy-coat and whole blood spectra compared to that in plasma in both normal and CLL blood. Higher nucleic acid distribution is seen in buffy-coat spectra compared to that in whole blood and plasma. The spectra of plasma mainly comprised of lipids, carbohydrates, amino acids and lactate contributions.

Multivariate analysis reveals that PCA-fed-LDA classification model constructed using spectra from buffycoat samples shows better performance compared to that using whole blood or plasma with higher sensitivity and specificity values. The prediction of the five most frequently occurring cellular components of blood was attempted with PLS regression, which was applied to buffycoat spectra. FBC values obtained from both normal and CLL buffycoat samples show very little variations in concentration values. Hence, PLS in this case is not accurate and has not been able to predict the true concentration values.

Chapter 7: Conclusions

The scope of this thesis was to investigate the application of FT-IR spectroscopy as a potential diagnostic tool for the detection of leukaemia and lymphoma. This mainly consisted of accurate spectral measurements the target cell line and blood samples; extracting the relevant biochemical information buried in the spectrum and application of multivariate statistical methods to convert experimental data to useful diagnostic information. A combination of spectral amalysis and multivariate statistical analysis was used for both quantification of biochemical constituents and classification of cell lines derived from human T-cell lymphoma, B- cell lymphoid and myeloid leukaemia as well as blood samples from healthy subjects and patients with CLL. The spectral data were subjected to both unsupervised and supervised methods of multivariate analysis; in particular principal component analysis fed linear discriminate analysis (PCA-fed LDA). Leave one out cross validation (LOOCV) was also used to examine the quality of spectroscopic classification. Cellular analyte concentrations were predicted by means of multivariate methods such as partial least-squares (PLS) regression.

7.1 Leukaemia and lymphoma cell line classification model performance

The FT-IR spectra acquired from T-cell lymphoma, B- cell lymphoid, and myeloid leukaemia cell lines reveal that peak shapes and positions are repeatable although the peak intensities tend to vary due to differences in the bio-chemical constituents. Most significant changes in peak intensities occur in the fingerprint region between the 900-1800 cm^{-1} corresponding to the proteins (Amide I, II and III), nucleic acids, amino acids, lipids and carbohydrates. Overall, the data showed that the relative intensities of DNA and amide III modes increase in malignant cells, indicating increased cellular nuclear content. Although these findings are consistent with those described in various publications (Boydston-White 1999; Boydston-White 2005 and Beekes 2007), these can not solely be used as a diagnostic algorithm.

Significant differences have been demonstrated between the spectra from air-dried and frozen cells in terms of peak intensities and area under the curve. Findings suggest that freezing cell line samples preserves cellular composition and reduces the rate of cellular

damage. This is consistent with other studies conducted by Wolkers 2007 and Mourant 2003. Visual inspections reveal that the IR absorption spectra of cell lines are complex, with hundreds of data points due to the sheer number of contributing bio-molecules and therefore contain many overlapping bands. In addition, the spectral homogeneity of spectra from cells belonging to similar lineage makes the visual interpretation of individual spectra nearly impossible. It is therefore essential that univariate methods are combined with both supervised and unsupervised methods of multivariate analyses for a meaningful interpretation.

The initial imaging experiments of the B-cell lymphoid, myeloid leukaemia and T-cell lymphoma cell lines have demonstrated problems associated with poor region selection and contributions from CaF_2 possibly resulting in poor quality spectral datasets. A preliminary unfiltered five cell line group classification model using 50 frozen and air-dried samples correctly classified 54% of 18556 spectra. The three cell line group model also using 50 frozen and air-dried samples correctly classified 77.1% of 19903 spectra. This demonstrates that higher numbers of spectra are classified correctly in the three cell line group models compared to the five cell line group models using LOOCV. The highest sensitivity was achieved for three cell line group datasets ranging between 47-75% compared to that achieved by five cell line model 44 -71%. The highest specificity was also achieved for three cell line group datasets ranging between 72-88% compared to that achieved by five cell line model 74-85%.

Furthermore, the use of the frozen spectra improves the performance of classification model considerably. Finding show that 73.3% of 9920 spectra are correctly classified in the frozen three cell line group datasets, whereas in the air-dried three cell line group only 41.5% of 9983 spectra are correctly classified. The frozen three cell line model achieved sensitivity and specificity of 47.1% and 72.2% respectively. In contract the dried three cell line model achieved sensitivity and specificity of 69.3% and 86.8% respectively. These findings confirmed the feasibility of using FT-IR spectroscopy to measure and discriminate between different cell lines groups, covering the range of haematological malignancies. The use of FT-IR spectral classification models for the discrimination of leukaemia and lymphoma have not been demonstrated to date.

However, similar methods employing PCA-fed-LDA and LOOCV have been used previously in investigation of synovial fluids from joints affected by rheumatoid arthritis and osteoarthritis (Shaw 1995) and classification of gastric inflammation and malignancy on endoscopic biopsies (Li 2005).

7.2 PCA-fed-LDA Classification Model Optimisation

Various modalities were investigated in order to optimise the PCA-fed-LDA classification model. These include assessing the

- effects of increasing PCs input for LDA
- effects of Savitsky-Golay filters
- application of standard deviation to filter outliers from the spectral dataset
- application of an absorption filter tool on spectral dataset

Optimisation of the PCA-fed-LDA classification model is important for the accurate classification of leukaemia and lymphoma derived cell lines. Findings reveal that methods of optimisation have significant improvements on the performance of the PCA-fed LDA classification model. Increasing the numbers of PCs input into the PCA-fed-LDA classification model from 5 PCs to 25 PCs is shown to increase the prediction achieved. In addition, using the only the frozen spectral dataset significantly improved the performance of the PCA-fed LDA classification model. Furthermore, the application of 1^{st} and 2^{nd} derivative reduces the percentage of spectra correctly classified in the frozen spectral dataset. 0^{th} derivative has a significant impact on the performance of the PCA-fed LDA classification model with higher percentages of the spectra correctly classified.

Applying a range of standard deviation filtering (+/- 5 to +/- 1.5) to the frozen spectral dataset has shown to have a considerable impact. Standard deviation of +/- 1.5 improved the performance of classification model resulting with highest sensitivity and specificity values. Although this an effective method of identifying poor quality spectra (outliers), a large number of spectra are eliminated from the dataset. The use of absorbance spectral filter tool to filter out weak/ low absorbance or saturated spectra is a

better approach of identifying poor quality spectra. The spectra retained after the application of the absorbance spectral filter tool is of better quality and more stable. This may indicates that performance of the PCA-fed LDA classification model is improved when the absorbance filter tool is applied compared to the use of standard deviation. However, as with the application of standard deviation filter, a huge number of spectra are eliminated from the dataset. Furthermore, significant improvements on the performance of the PCA-fed LDA classification model was achieved using three cell line group model compared to that of five cell line group model.

Overall, cross-validation testing of LDA models developed with statistically significant principal components has demonstrated that 25 PC's are sufficient to produce a good model. In addition, 0^{th} derivative and the absorbance filter tool on frozen spectral dataset are the standard parameters for constructing an optimised PCA-fed-LDA classification model as 90% of 3183 spectra were correctly classified in the frozen, filtered five cell line group model. In the frozen filtered three cell line group model 88% of 2972 spectra were correctly classified

7.3 Normal and CLL blood classification model performance

To avoid loss of huge amounts of spectra from original datasets as a result of standard deviation and absorbance spectral filter tool, targeted point spectra measurements were collected from 10 normal and 12 CLL blood samples. Several regions of whole blood, plasma and buffycoat preparations were probed for their IR signature to evaluate spectral variations. The point spectra measurement overcomes contributions from CaF_2 and acquires a more pure spectrum of better quality. Hence, this leads to improved performance of classification model. The limitation with point spectra measurements is that it can be time consuming and hence produce a smaller spectral dataset.

Significant spectral differences were noted between the three blood components of normal and CLL blood including whole blood, buffycoat and plasma. Further spectral variations between the different blood components were emphasised using difference spectra. Peak intensity ratios were found to be in agreement with the difference spectra indicating higher protein distribution in buffy-coat and whole blood spectra compared to

that in plasma in both normal and CLL blood. Higher nucleic acid distribution is seen in buffy-coat spectra compared to that in whole blood and plasma. The spectra of plasma mainly comprised of lipids, carbohydrates, amino acids and lactate contributions.

Variations are also observed in peak intensity ratios between normal and CLL whole blood spectra; normal and CLL buffy coat spectra and normal and CLL plasma spectra. In general higher protein/lipid distribution is seen in normal whole blood, plasma and buffycoat spectra compared to that of CLL whole blood, plasma and buffycoat spectra. Furthermore, higher nucleic acid/protein distribution is seen in CLL whole blood and buffycoat spectra compared to that of normal whole blood and buffycoat spectra. The increase in the nucleic acid/protein peak intensity ratio in CLL buffycoat coincides with study by Boydston-White 1999, Schultz 1996; Schultz 1997 and Bendetti 1984. The observed results suggest that concentration of nucleic acids increase in leukaemia cells during the replication of DNA.

Multivariate analysis reveals that PCA-fed-LDA classification model constructed using spectra from buffycoat samples shows better performance compared to that using whole blood or plasma with higher sensitivity and specificity values. It has been demonstrated that higher number of spectra are classified correctly in the classification model constructed from buffycoat spectra. Using LOOCV, the whole blood model correctly classified 70% of 1736 spectra, measured on 22 samples (sensitivity 69.6% and specificity 66.7%). The plasma model correctly classified 80.6% of 331 spectra (sensitivity and specificity of 66.7%). The buffy coat model correctly classified 99.5% of 1438 spectra (sensitivity 99.2% and specificity 99.2%).

The prediction of the five most frequently occurring cellular components of blood was attempted with PLS regression, which was applied to buffycoat spectra in the mid-infrared fingerprint region from 1800–900 cm^{-1}. The calibration (training) set for PLS regression was composed of cellular component of neutrophil, lymphocyte, monocyte, eosinophil and basophil from normal buffycoat samples at different concentrations. The cellular analyte concentration of the 12 CLL samples (validation dataset) was predicted using PLS analysis. FBC values obtained from both normal and CLL buffycoat samples

show very little variations in concentration values. Hence, PLS in this case is not accurate and has not been able to predict the true concentration values.

7.4 Clinical relevance

The lack of studies in the literature comparing spectral differences between multiple subtypes of leukaemia and lymphomas has been highlighted. This thesis has served to improve the results within the literature. This study has demonstrated the ability of FT-IR spectroscopic classification models, to differentiate up to five cell line groups covering a range of haematological malignancies as well normal and CLL blood samples. Furthermore, combination of FT-IR spectroscopy with multivariate feature extraction and classification methods has the potential to become an ideal complementary tool to conventional assays for early detection of leukaemia and lymphoma. Additionally, as only small blood samples are obtained during routine check-up, this will be ideal for FT-IR spectroscopic measurements as only small sample volume is required.

FT-IR spectral imaging demonstrated potential for differentiating different cell line types and blood samples. The pseudo-colour maps of the principal component scores calculated from the FT-IR spectral dataset have been shown to provide information on the distribution of biochemical constituents. Further work in this area encompassing biochemical studies would hopefully elicit more information and add to the understanding of complex biochemical constituents at cellular level. Infrared imaging provides a potential window into the disease processes, revealing compositional and structural information. It also provides the added benefit that instrumentation does not require reagents to promote a visible change in colour. It also ensures that all IR compositional data is stored in a single data set, permitting the rapid extraction of information of components of interest. These initial results further indicate that this approach provides the added benefit that instrumentation does not require staining reagents to promote a visual change in colour and ensures that all compositional data is stored in a single dataset. With further improvements and analytical work, an objective discrimination method can be developed.

Chapter 8: Future Work

The potential of FT-IR and chemometrics as a rapid and accurate diagnostic tool is clearly demonstrated by the results presented in this thesis. However, more work would be needed to see this through to clinical application and this study leads to many further research recommendations, which fall into three distinct areas:

- Improvements in study design
- Improvements in chemometric techniques
- Exploitation of other optical diagnostic techniques

8.1 Improvements in study design

In regards to the cell line work, cell preparation and sampling methods need to be standardised. Further controlled experimental work need to be carried out to account for the extent of air-drying and freezing of cell lines samples and the affect this may have on the cell viability. Future experimental designs must incorporate obtaining and comparing spectra from samples air-dried and freezed ranging at 1hour to 24 hours. These observations can be further extended to days and weeks to provide a better understanding of the effects of air-drying and freezing on cell line samples over a period of time and provide a measure of degree of cellular decomposition.

To futher limit other compounding factors that may effect the biochemical composition of the spectral fingerprint, future experimental designs must also consider using cell lines at a specific phase of the cell cycle (G1, G2 and S phase). Results from the cell line work presented in this thesis do not take into account the possible varations that may occur within cells in different phases of the cell cycles. Sample preparation methods including cell cultering and mitotic selection (or "mitotic shake-off") have been previously used to isolate cell at different stages of the cell division cycle (Boydston-White 2006). By isolating cells at different stages of the cell cycle, changes in the FT-IR spectra of individual cells can be correlated with their biochemical age.

The cross validation testing of the FT-IR classification models indicated that sufficient population variance was not accounted for within the spectral datasets. This was particularly the case for the spectral dataset constructed from blood sample measurements. Therefore larger spectral models are required to account for the variation in spectra expected to be encountered in clinical practice. Ideally, the use of an unknown test set would provide a more robust independent analysis of sensitivity and specificity of the technique.

Further evaluation of the clinical relevance of FT-IR spectroscopy is required. It will be interesting to see whether the FT-IR spectroscopy is able to classify other rare forms of leukaemia such as acute myeloid leukaemia (AML), acute lymphoblastic leukaemia (ALL) and chronic lymphocytic leukaemia (CML). There is scope to broaden the applications of the study design to include fine needle aspiration biopsy (FNAB) sample from lymph nodes and bone marrow to identify spectral change associated with stages of cell differentiation and maturation as well as cell lineages involved. This may predict early changes before the disease progresses therefore further rigorous testing of models constructed from testing of heterogeneous material such as FNAB sample is also required. Although this was planned in the original ethics application, due to time constraints, further work was not possible.

Other avenues to exploit involve monitoring effectiveness of drugs during chemotherapy using FT-IR spectroscopy. Key objective would be to 1) assess the potential of FT-IR spectroscopy for the follow-up of leukaemia treatment 2) assess spectral and biochemical changes that occur in white blood cells of a CLL patient during chemotherapy and monitor and evaluate WBC populations before and after as well as during the various stages of chemotherapy. These investigations would be useful to indentify possible biomarkers which may indicate successful remission for the patient.

This study has demonstrated the potential of FT-IR spectroscopy as a complementary analytical tool in haematology. Although, the ability to influence early detection is an appealing prospect, there is still much work to be undertaken before it can be clinically implemented. Large multicentre clinical trials are required to determine the sensitivity

and specificity, cost effectiveness and effect on clinical outcome of FT-IR spectroscopy for the detection of haematological malignancies. In addition, this will also overcome discrepancies associated with system standardisation and transferability. This will indeed require setting up multicentre sites on a very large scale and expanding collaboration with surgeons, haematologists and scientists.

8.2 Improvements in chemometric techniques

Improvements in the spectral analysis techniques are also required. Selection of significant spectral regions may enable faster measurement and processing and may improve prediction by removing unrelated data. Incorporating the entire spectral region (4000 – 700cm^{-1}) must be considered to account for variations in lipid and fatty acid profile of normal and cancer cells. As previously mentioned much work is required to optimise the PLS model with improvements in the calibration dataset and testing the quality of predictions using an independent test set. Furthermore, it will be useful to investigate alternative and advanced analytical methods such as:

- Cluster analysis as a complimentary imaging technique
- Artificial neural network (ANN) as classification model
- Biochemical fitting using non-negative least squares to estimate relative concentrations of biochemical components in normal and CLL cells as well as different subtypes of leukaemia cells.

Much improvement is required in the optimisation of the PLS model. Various factors need to be considered to improve our understanding. These include:

- Improve calibration dataset with more variation in concentration. This may be achievable by preparing artificial mixtures, which involves isolating different cellular analytes from the original buffycoat sample maybe with an aid of a flow cytometer coupled to a sorter or a traditional centrifugal method (where different cell types separate due to variation in density). Then spiking serum/plasma samples with known concentration of cellular analytes. Each serum/plasma sample shall have a known concentration of individual cellular analyte (neutrophil, lymphocyte, monocyte, eosinophil or basophil) ranging from cell

count of 10^20/L to 10^5/L. This may achieve a more reliable and accurate calibration dataset.

- Testing the quality of predictions using an independent test set, this involves using a series of samples that have been left out of the original calculation.

8.3 Exploitation of other optical diagnostic techniques

Exploitation of Raman spectroscopy as a complementary technique is necessary for further biochemical analysis at single cell level. Raman spectrometry benefits from a higher spatial resolution (1-2 μm) compared to FT-IR spectrometry (10-20 μm). This will allow biochemical analysis of single cells at sub-cellular level providing qualitative and quantitative information on the molecular composition of various regions of a cell (e.g. cytoplasm, nucleus and cell membranes). Use of near-infrared radiation (NIR) Raman spectroscopy also overcomes the limitation of spectral interference of water and provides a higher spatial resolution which allows detection of biochemical constituents present within a single cell and at a sub-cellular level. Furthermore, due the differences in the selection rules, the analysis of Raman spectra can provide complementary biochemical information reflecting Raman inelastic scattering of non-polar bonds (e.g. C-C,C=C, and aromatic rings) that give a stronger Raman signals which are otherwise weak in the infrared spectra.

ATR-FTIR spectroscopy is a technique enabling the measurement of aqueous solutions. The ATR accessory eliminates sample preparation required for FT-IR analysis by transmission techniques and thereby speeds the measurement. In this way, ATR FT-IR spectroscopic studies of liquid cell line suspensions and blood samples could be feasible.

References

1. Alexander, F.E. (1995). The Search for Causes of the Leukaemias. *European Journal of Cancer*, 31A (6), 863-867.

2. Alfano, R.R., Tata, D.B., Cordero, J., Tomashefsky, P., Longo, F.W. and Alfano, M.A. (1984). Laser induced fluorescence spectroscopy from native cancerous and normal tissue. *IEEE J Quant Elect*, 20**,** 1507–1511.

3. Andrus, G.L.P. and Strickland, D.R. (1998). Cancer Grading by Fourier Transform Infrared Spectroscopy. *Biospectroscopy*, 4, 37-46.

4. Andrus, P.G. (2006). Cancer Monitoring by FT-IR Spectroscopy. *Technology in Cancer Research and Treatment*, 5(2),157-167.

5. Basso, G., Case, C. and Dell'Orto, M.C. (2007). Diagnosis and genetic subtypes of leukaemia combining gene expression and flow cytometery. *Blood. Cells. Mol. Dis*., 39(2), 164-8.

6. Beekes, M., Lasch, P., and Naumann, D. (2007). Analytical applications of Fourier transform- infrared (FT-IR) spectroscopy in microbiology and prion research. *Veterinary Microbiology*. 123 (4), 305-319.

7. Benard, C.W., Bloomfield, C., Bonadonna, G., Byron, W.B. Glatstein, E., Hoppe, R., Rosenberg, S.A., Rudders, R., and Simon, R. (1985). Classification of non-Hodgkin's lymphomas. Reproducibility of major classification system. *Cancer*, 55, 91-95.

8. Benedetti, E., Bramanyi, E. and Papinechi, F. (1997). Determination of the relative amount of nucleic acids and proteins in leukemic and normal lymphocytes by means of Fourier transform infrared microspectroscopy. *Applied spectroscopy*, 51 (6), 792-797.

9. Benedetti, E., Papineschi, F., Vergammi, P., Consolini, R. and Spremolla, G. (1984). Analytical infrared spectral differences between human normal and leukaemic cells (CLL). *Leukaemia Research*, 8, 483-489.

10. Berger, A.J., Koo, T.W., Itzkan, I., Horowitz, G., and Feld, M.S. (1999). Multicomponent Blood Analysis by Near-Infrared Raman Spectroscopy. *Applied Optics*, 38 (13), 2916-292.

11. Bigio, I.J., Bown, S.G. and Briggs, G (2000). Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results. *J. Biomed. Opt*., 5, 221-228.

12. Binet, J. L., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F., Thomas, M., Tchernia, G.,

Jacquillat, C., Boivin, P., Lesty, C., Duault, M. T., Monconduit, M., Belabbes, S., Gremy, F. (1981). Classification of leukaemia. *Leukaemia*, 48 (1), 198-206.

13. Bonnet, D. and Dick, J.E. (1997), Human acute myeloid leukaemia is organized as a hierarchy that originates from a primitive haematopoietic cell. *Nat Med*, 3, 730–737.

14. Bonnier, F., Rubin, S., Venteo, L., Krishna, C.M., Pluot, M., Baehrel, B, Manfait, M and Sockalingum, G. (2006). In-vitro analysis of normal and aneurismal human ascending aortic tissues using FT-IR microspectroscopy. *Biochemica et Biophysica Acta*, 1758, 968-973.

15. Bossart, R., Keller H., Kellerhals, H. and Oelichmann, J. (2003). Principal components analysis as a tool for identity control using near-infrared spectroscopy. Journal of *Molecular Structure*, 661-662, 319-323.

16. Boydston-White, S., Gopen, T., Houser, S., Bargonetti, J. and Diem, M. (1999). Infrared spectroscopy of human tissue. Infrared spectroscopic studies of myeloid leukaemia (ML-1) cells at different phases of the cell cycle. *Biospectroscopy*, 5, 219-227.

17. Boydston-White, S., Romeo, M., Chernenko, T., Regina, A., Miljkovic, M. and Diem, M. (2006). Cell-cycle-dependent variations in FT-IR micro-spectra of single proliferating HeLa cells: Principal component and artificial neural network analysis. *Biochemica et Biophysica Acta*, 1758, 908-914.

18. Boydston-White. S., Chernenko. T., Regina, A., Miljkovic, M., Matthaeus, C., and Diem, M. (2005). Microspectroscopy of single proliferating HeLa cells. Vibrational Spectroscopy, Vol. 38 169-177.

19. Braziel, R.M., Shipp, M.A., Feldman, A.L., Espina, V., Winters, M., Jaffe, E.S., Petricoin, E.F. and Liotta, L.A (2003). Molecular Diagnostics. Haematology, 279-293.

20. Burattini, E., Malvezzi-Campeggi, F., Chilosi, M., Conti, C., Ferraris, P., Monti, F., Sabbatini, S., Tosi, G., Zamo, A. (2007). FPA micro spectral imaging of non-Hodgkin lymphomas. *Journal of Molecular Structure*, 834-836,170-175.

21. Byler, D.M. and Susi, H. (1986). Examination of the Secondary Structure of Proteins by Deconvolved FT-IR Spectra. *Biopolymers*, 25, 469-487.

22. Cancerstats Leukaemia and Lymphomas: UK 2005.

23. Cantor, A.B. and Orkin, S.H. (2001), Haematopoietic development: a balancing act. *Curr Opin Genet Dev*, 11, 513–519.

24. Carmona, P., Monzon, M., Monleon, E., Badiola, J.J. and Monreal, J. (2005). In vivo detection of scapie cases from blood by infrared spectroscopy. *Journal of General Virology*, 86, 3425-3431.

25. Chan, J.W., Taylor, D.S., Lane, S., Zwerdling, T., Tuscano, J. and Huser, T. (2008), Non-destructive Identification of Individual Leukemia Cells by Laser Tweezers Raman Spectroscopy. *Analytical Chemistry,* 80, 2180-2187.

26. Chan, J.W., Taylor, D.S., Zwerdling, T., Lane, S.M., Ihara, K. and Huser, T. (2006). Micro-Raman Spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophysical Journal*, 90, 648-656.

27. Chang, S.K., Pavlova, I., Marin, N., Follen, M., Richards-Kortum, R. (2005). Fluorescence spectroscopy as a diagnostic tool for detecting cervical pre-cancer. *Gynecology Oncology*, 99 (3), S61-S63.

28. Chapman, D., Jackson, M. and Haris, P.I. (1989). Investigation of Membrane Protein Structure using Fourier Transform Infrared Spectroscopy. *Biochemical Society Transactions*, 17, 617-619.

29. Chicken, D. W., Johnson, K. S., Falzon, M.R., Lee, A.C., Briggs, G., Pickard, D., Bigio, I.J., Bown, S.G. and Keshtgar, M.R.S. (2004). Optical biopsy utilising elastic scattering spectroscopy for intra-operative diagnosis of sentinel lymph node metastases. *Journal of Clinical Oncology, 2004 ASCO Annual Meeting Proceedings (Post-Meeting Edition)*, 22, 841 (abstract).

30. Chiriboga, L., Xie, P., Vigorita, V., Zarou, D., Zakim, D. and Diem, M. (1998). Infrared spectroscopy of human tissue. II. A comparative study of spectra of biopsies of cervical squamous epithelium and of exfoliated cervical cells. *Biospectroscopy*, 4 (1), 55-59.

31. Cohenford, M.A. and Rigas, B (1998). Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implications for tumour biology. *Proc. Natl. Acad. Sci*., 95, 15327-32.
32. Cohenford, M., Godwin, T., Cahn, F., Bhandare, P., Caputo, T. and Rigas, B. (1997). Infrared spectroscopy of normal and abnormal cervical smears: Evaluation by principal component analysis. *Gynecology Oncology*, 66, 59-65.

33. Colley, C.S., Kazarian, S.G., Weinberg, P.D., Lever, M.J. (2004). Spectroscopic imaging of arteries and atherosclerotic plaques. *Biopolymers*, 74(4), 328-35.

34. Crow, P., Molckovsky, A., Stone, N., Uff, J., Wilson, B., Wongkeesong, L.M. (2005). Assessment of fiberoptic near-infrared Raman spectroscopy for diagnosis of bladder and prostate cancer. *Adult Urology*, 65(6), 1126 - 1130.

35. Crow, P., Stone, N., Kendall, C., Uff, J., Farmer, J.A.M., Barr, H. and Wright, M.P.J. (2003). The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma in vitro. *British Journal of Cancer*, 89, 106-108.

36. Crow, P., Barrass, B., Kendall, C., Hart-Prieto, M., Wright, M., Persad, R . and Stone, N (2005). The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *British Journal of Cancer*, 92, 2166–2170.

37. Debatin, K.M., Stahnke, K., and Fulda, S. (2003). Apoptosis in haematological disorders. *Seminars in Cancer Biology*, 13, 149-158.

38. Deleris, G. and Petiois, C. (2003). Application of FT-IR spectrometry to plasma contents analysis and monitoring. *Vibrational spectroscopy*, 32,129-136.

39. Dhar, A., Johnson, K.S., Novelli, M.R., Bown, S.G., Bigio, I.J., Lovat, L.B. and Bloom, S.L. (2006). Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique. *Gastrointestinal Endoscopy*, 63(2), 257-261.

40. Diem, M., Chiriboga, L. and Yee, H. (2000). Infrared spectroscopy of human cells and tissue.VIII. Strategies for analysis of infrared tissue mapping data and application to liver tissue. *Biopolymers (Biospectroscopy)*, 57, 282-290.

41. Diem, M., Chiriboga, L., Lasch, P. and Pacifico, A. (2002). IR spectra and IR maps of individual and cancerous cells. *Biopolymers (Biospectroscopy)*, 67, 349-353.

42. Diem, M., Griffiths, P.R. and Chalmers, J.M. (2008). Vibrational Spectroscopy for Medical Diagnosis. John Wiley and Sons Ltd, England.

43. Domen, J. (2001). The role of apoptosis in regulating haematopoietic stem cell numbers. *Apoptosis*, 6, 239–252.

44. Dong, H.Y., Harris, N.L., Preffer, F.I. and Pitman, M.B. (2001). Fine-needle aspiration biopsy in the diagnosis and classification of primary and recurrent lymphoma: a retrospective analysis of the utility of cytomorphology and flow cytometry. Modern Pathology, 14, 472–81.

45. Dou, X., Yamaguchi, Y., Yamamoto, H., Uenoyama, H., and Ozaki, Y., (1996). Biological Applications of Anti-Stokes Raman Spectroscopy: Quantitative Analysis of Glucose in Plasma and Serum by a Highly Sensitive Multichannel Raman Spectrometer, *Applied Spectroscopy*, 50 (10), 1301-1306.

46. Downing, J.R. and Shannon, K.M. (2002). Acute Leukaemia: A paediatric perspective. *Cancer Cell*, 2, 437-445.

47. Dukor, R.K., Liebman, M.N. and Johnson, B.L. (1998). A new, non-destructive method for analysis of clinical samples with FT-IR microspectroscopy: Breast cancer tissue as an example. Cellular and molecular biology, 44(1), p.211-7.

48. Eckel, R., Huo, H., Guan, H.W., Hu, X., Che, X. and Huang, W.D. (2001). Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue. Vibrational Spectroscopy, 27(2), 165-173.

49. Enejder, A.M.K., Koo, T.W., Oh, J., Hunter, M., Sasic, S., Feld, M.S. and Horowitz, G.L. (2002). Blood analysis by Raman spectroscopy, *Optics Letters*, 27 (22), 2004-2006.

50. Enver, T. and Greaves, M. (1998), Loops, lineage, and leukaemia. *Cell*, 94, 9–12.

51. Erukhimovitch, V., Talyshinsky, M., Souprun, Y. and Huleihel, M. (2002). Spectroscopic characterization of human and mouse primary cells, cell lines and malignant cells. *Photochemistry and Photobiology*, 74 (4), 446-451.

52. Erukhimovitch, V., Talyskinsky, M., Souprun, Y., Huleihel, M. (2006). FT-IR spectroscopy examination of leukaemia patient plasma. *Vibrational Spectroscopy*, 40, 40-46.

53. Fabian, H., Lasch, P., Boese, M. and Haensch, W. (2003). Infrared Microspectroscopic imaging of benign breast tumour tissue sections. *Journal of Molecular Structure*, 661-662, 411-417.

54. Fialkow, P.J., Singer, J.W., Adamson, J.W., Vaidya, K., Dow, L.W., Ochs, J. and Moohr, J.W. (1981), Acute nonlymphocytic leukaemia: heterogeneity of stem cell origin. *Blood*, 57,1068–1073.

55. Fisher, A.G. (2002), Cellular identity and lineage choice. *Nat Rev Immunol* , 2, p. 1–6.

56. Gaigneaux, A., Ruysschaert, J.M. and Goormaghtigh, E. (2002). Infrared spectroscopy as a tool for discrimination between sensitive and multiresistant K562 cells. *Eur. J. Biochem*, 269,1968-1973.

57. Gao, T., Feng, J. and Ci, Y. (1999). Human breast carcinomal tissues display distinctive FT-IR spectra: implication for the histological characterization of carcinomas. *Analytical Cellular Pathololog*y, 18, 87–93.

58. Gasparri, F. and Muzio, M. (2003). Monitoring of apoptosis of HL60 cells by Fourier-transform infrared spectroscopy. *Biochemical Journal*, 369, 239-248.

59. Gendrin, C., Roggo, Y. and Collet, C. (2008). Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review. *Journal of Pharmaceutical and Biomedical Analysis*, 48, 533- 553.

60. German, M.J., Hammiche, A., Ragavan, N., Tobin, M.J., Cooper, L.J., Matanhelia, S.S., Hindley, A.C., Nicholson, C.M., Fullwood, N.J., Pollock, H.M. and Martin F.L. (2006). Infrared spectroscopy with multivariant analysis potentially facilitates the segregation of different types of prostate cell. *Biophysical Journal*, 90, 783-3795.

61. Grimbergen, M.C.M., Van Swol, C.F.P., Draga, R.O.P., Van Diest, P., Verdaasdonk, R.M., Stone, N., Bosch, J.H.L.R. (2009). Bladder cancer diagnosis during cystoscopy using Raman spectroscopy. Proceedings of the SPIE, 7161, 716114-716114-6.

62. Groves, F.D, Linet, M.S. and Devesa, S.S (1995). Patterns of Occurrence of the Leukaemias. *European Journal of Cancer,* 31A (6), 941-949.

63. Haka, A.S., Volynskaya, Z., Gardecki, J.A., Nazemi, J., Lyons, J., Hicks, D., Fitzmaurice, M., Dasari, R. R., Crowe, J. P. and Michael S. Feld (2006). In vivo Margin Assessment during Partial Mastectomy Breast Surgery Using Raman Spectroscopy. *Cancer Research*, 66(6), 3317-3322.

64. Hans, C.P., Weisenburger, D.D., Vose J.M., Lynette, M.H., Lynch, J.C., Aoun, P., Greiner, T.C., Chan, W.C., Bociek, R.G., Bierman, P.J. and Armitage, J.O. (2003). A significant diffuse component predicts for inferior survival in grade 3 follicular lymphoma, but cytologic subtypes do not predict survival. *Blood*, 101, 2363–7.

65. Harvey, T.J., Faria, E.C., Henderson, A. and Gazi, E. (2008). Elastic scattering spectroscopy in the diagnosis of melanocytic lesions: comparison with clinical and histopathological diagnosis. *J. Biomed. Opt*., 13 (6).

66. Haynes, C.L., Ranjit, C., Xiaoyu Y.Z. and Van Duyne R.P. (2005). Surface-enhanced Raman sensors: early history and the development of sensors for quantitative biowarfare agent and glucose detection. *Journal of Raman Spectroscopy*. 36, 471–484.

67. Hehn, S.T., Grogan, T.M. and Miller, T.P. (2004). Utility of fine-needle aspiration as a diagnostic technique in lymphoma. *Journal of Clinical Oncology*, 22, 3046–52.

68. Hoffbrand, A.V. and Hamblin, T.J. (2007). Is CLL an appropriate label for all patients how meet the diagnostic criteria of chronic lymphocytic CLL. *Leukaemia Research*, 31, 273-275.

69. Hoffbrand, A.V., Moss, P.A.H and Pettit, J.E. (2006). Essential Haematology, Fifth Edition. Blackwell Publishing.

70. Hosafci, G., Klein, O., Oremek, G and Mantele, W. (2007). Clinical chemistry without reagents? An infrared spectroscopic technequies for determination of

clinically relevant constituents of body fluids. *Anal. Bioanal. Chem*, 387, 1815-1822.

71. Hui, H.K., and Po, H.H. (2005). Surface-enhanced Raman scattering (SERS) spectroscopy technique for lactic acid in serum measurement. Plasmonics: Metallic Nanostructures and Their Optical Properties III. *Proceedings of the SPIE*, 5927, 395-402.

72. Isaacson, P.G. (2000). The current status of lymphoma classification. *British Journal of Haematology*, 109, 258-266.

73. Jaffe, E.S., Harris, N.L., Stein, H., Verdiman, J.W., editors, 2001. World Health Organisation Classification of Tumors. Pathology and Genetics of Tumors of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press.

74. Jamin, N., Miller, L., Moncuit, J., Fridman, W.H., Dumas, P. and Teilluad, J.L. (2003). Chemical heterogeneity in cell death: combined synchrotron IR and fluorescence microscopy studies of single and apoptotic and necrotic cells. *Biopolymers*, 72, 366-373.

75. Jess, P.R.T., Smith, D.D.W, Mazilu, M., Dholakia, K., Riches, A.C., Herrington, S. (2007). Early detection of cervical neoplasia by Raman spectroscopy. International Journal of Cancer, 121 (12), 2723-2728.

76. Kanter, E., Majumder, S., Kanter, G., Woeste, E., Mahadevan-Jansen, A. (2009). Effect of hormonal variation on Raman spectra for cervical disease detection. *American Journal of Obstetrics and Gynecology*, 200 (5), 512.e1-512.e5.

77. Katz, R. (1997). Pitfalls in the diagnosis of fine-needle aspiration of lymph nodes. *Monographs in Pathology*, 39, 118–33.

78. Kendall, C., Stone, N., Shephard, N., Gebos, K., Warren, B., Bennett, R. and Barr, H. (2003). Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *Journal of Pathology*, 200, 602-609.

79. Khanmohammadi, M., Nasiri, R., Ghasemi, K., Samani, S. and Garmarudi, A.M. (2007). Diagnosis of basal cell carcinoma by infrared spectroscopy of whole blood samples applying soft independent modelling class analogy. *Journal of Cancer Research Clinical Oncology*, 133, 1001-1010.

80. Kneipp, K., Haka, A.S., Kneipp, H., Badizadegan, K., Yoshizawa, N., Boone, C., Shafer-Peltier, K.E., Motz, J.T., Dasari, R.R., and Feld, M.S. (2002). Surface-enhanced Raman Spectroscopy in single living cells using gold nanoparticles. *Applied spectroscopy*, 56, 15-154.

81. Kocjan, G. (1997). The role of FNAC in diagnosis of lymph node enlargements. *Cytopathology*, 8, 2–3.

82. Krafft, C., Sobottka, A.B. Geiger, K.D. Schackert, G. and Salzer, R. (2006). *Analytical and Bioanalytical Chemistry*, 387, 1669-1677.

83. Krafft, C., Codrich, D., Pelizzo, G. and Sergo, V. (2008). Raman and FT-IR imaging of lung tissue: Methodology for control samples. *Vibrational Spectroscopy*, 46, 141-149.

84. Krafft, C., Knetschke, T., Funk, R.H.W. and Salzer, R. (2005). Identification of organelles and vesicles in sigle cells by Raman microsctroscopic mapping. *Vibrational Spectroscopy*, 38, 85-93.

85. Krafft, C., Knetschke, T., Siegner, A., Funk, R.H.W. and Salzer, R. (2003). Mapping of single cells by near infrared Raman microspectroscopy. *Vibrational Spectrsocopy*, 32, 756-83.

86. Krishna, C.M., Sockalingum, D. Ganesh. K.G., Rubin, S., Vasudevan, B.K. and Manfait, Michel. (2005). Micro-Raman spectroscopy of mixed cancer cell populations. *Vibrational Spectroscopy*, 38, 95-100.

87. Krishnan, R.S. and Shankar, R.K. (1981). Raman Effect: History of the discovery. *Journal of Raman Spectroscopy*, 10 (1), 1-8.

88. Kuimova, M.K., Chan, K.L., Kazarian, S.G. (2009) Chemical imaging of live cancer cells in the natural aqueous environment, *Applied Spectroscopy*, 63(2), 164-71.

89. Kwong, Y.L. and Chan, T.K. (1988). Seminar in Haematology: Current concepts in Haematology: (1) Normal Haemopoiesis. *Journal of the Hong Kong Medical Association*, 40 (3), 171-173.

90. Landgen, O., Porwit MacDonald, A., Tani, E., Czader, M., Grimfors, G., Skoog, L., Ost, A., Wedelin, C., Axdorph, U., Svedmyr, E. and Bjorkholm, M. (2004). A prospective comparison of fine-needle aspiration cytology and histopathology in the diagnosis and classification of lymphomas. *Hematololology Journal*, 5(1), 69–76.

91. Lasch, P. and Naumann, D. (2006). Spatial resolution in infrared microspectroscopic imaging of tissues. *Biochimica et Biophysica Acta*, 1758**,** 814-829.

92. Lasch, P., Boese, M., Pacifico, A. and Diem, M. (2002). FT-IR spectroscopic investigations of single cells on the subcellular level. *Vibrational Spectroscopy*, 28, 147-157.

93. Lasch, P., Chiriboga, L., Yee, H. and Diem, M (2002). Infrared spectroscopy of human cells and tissue: detection of disease. *Technology in Cancer Research and Treatment*, 1 (1), 1-7.

94. Lasch, P., Hensch, W., Naumann, D. and Diem, M (2004). Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim. Biophys. Acta*, 1688, 176-186.

95. Lau, D.P., Huang, Z., Lui, H., Anderson, D.W., Berean, K., Morrison, M.D., Shen, L., and Zeng, H., (2005). Raman Spectroscopy for Optical Diagnosis in the Larynx: Preliminary Findings. *Lasers in Surgery and Medicine* 37, 192–200.

96. Levi, F., Lucchini, F. and Negri, E. (2000) Trends in cancer incidence and mortality from leukaemia in subsequent age groups. *Leukaemia*, 14, 1980-1985.

97. Li, Q.B., Sun, X.J., Xu, Y.Z., Yang, L.M., Zhang, Y.F., Weng, S.F., Shi, J.S. and Wu, J.G. (2005). Diagnosis of gastric inflammation and malignancy on endoscopic biopsies based on Fourier Transform Infrared Spectroscopy. *Clinical Chemistry*, 51, 346-350.

98. Lioe, T.F.H., Allen, D.C. and Spence, R.A. (1999). The role of fine needle aspiration cytology (FNAC) in the investigation of superficial lymphadenopathy: uses and limitations of the technique. *Cytopathology*, 10, 291–7.

99. Lovat, L.B., Johnson, K., Mackenzie, G.D., Clark, B.R., Novelli, M.R., Davies, S., O'Donovan, M., Selvasekar, C., Thorpe, S.M., Pickard, D., Fitzgerald, R., Fearn, T., Bigio, I. and Bown, S.G. (2006). Elastic scattering spectroscopy accurately detects high grade dysplasia and cancer in Barrett's oesophagus. *Gut*, 55, 1078-1083.

100. Low-Ying, S., Shaw, R.A., Leroux, M. and Mantsch, H.H. (2002). Quantitation of glucose and urea in whole blood by mid-infrared spectroscopy of dry films. *Vibrational spectroscopy*, 28, 111-116.

101. Lui, K.Z., Ahmed, M.K., Dembinski, T.C., Mantsch, H.H. (1997). Prediction of fetal lung maturity from near-infrared spectra of amniotic fluid. *International Journal of Gynecology & Obstetrics*, 57,161-8.

102. Lui, K.Z, Schultz, C.P, Johnston, J.B, Mantsch, H.H (1997). Comparison of Infrared spectra of CLL cells with their ex-vivo sensitivity (MTT assay) to chlorambucil and cladribine. *Leukaemia Research*, 21, 1125-1133.

103. Lui, K.Z., Li, J., Kelsey, S.M. and Mantsch, H.H. (2001). Quantitative determination of apoptosis on leukaemia cells by infrared spectroscopy. *Apoptosis*, 6, 267-276.

104. Lui, K.Z., Schultz, C.P., Salamon, E.A., Man, A. and Mantsch, H.H. (2003). Infrared spectroscopic diagnosis of thyroid tumours. *Journal of Molecular structure*, 661-662, 397-404.

105. Lui, K.Z., Shaw, R.A., Man, A., Dembinski, T.C. and Mantsch, H.H. (2002). Reagent free, simultaneous determination of serum cholesterol in HDL and LDL by infrared spectroscopy. *Clinical Chemistry*, 48 (3), 499-506.

106. Lui, K.Z., Shi, M.H., and Mantsch, H.H. (2005). Molecular and chemical charcterisation of blood cells by infrared spectroscopy: A new optical tool in haematology. *Blood cells, Molecules and Disease*, 35, 404-412.

107. Lyng F.M, Faolain E.O, Conroy J, Meade A.D, Knief P, Duffy B, Hunter M.B, Byrne J.M. Kelehan P and Byrne H.J (2007). Vibrational spectroscopy for cervical cancer pathology from biochemical analysis to diagnostic tool. *Experimental and Molecular Pathology*, 82, 121-129.

108. Mahadevan-Jansen, A. and Richards-Kortum, R. (1996). Raman spectroscopy for the detection of cancer and precancers. *Journal of Biomedical Optics*, 1(1), 31-70.

109. Malini, R., Venkatakrishna, K., Pai, K.M., Rao, L., Kartha, V.B. and Krishna, C.M. (2006). Discrimination of normal, inflammatory, premalignant and malignant oral tissue: A Raman spectroscopic study. *Biopolymers*, 81, 179-193.

110. Manoharan, R., Baraga, J., Rava, P., Dasari, R., Fitzmaurice, M., Feld, M. (1993). Biochemical analysis and mapping of atherosclerotic human artery using FT-IR Microspectroscopy. *Atherosclerosis*, 103(2), 181-93.

111. Matthaus, C., Miljkovic, M., Romeo, M., Boydston-White, S. and Diem, M. (2006). Raman and infrared mico-spectral imaging of mitotic cells. *Applied Spectroscopy*, 60 (1), 1-8.

112. Maziak, D.E., Do, M.T., Shamji, F.M., Sundaresan, S.R., Perkins, D.G., Wong, P.T.T. (2007). Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: An exploratory study. *Cancer Detection and Prevention*, 31(3), 244-253.

113. McCarthy, K.P. (1997). Molecular diagnosis of lymphomas and associated diseases. *Cancer and Metastasis Reviews*, 16, 109-127.

114. McClain, B.L., Clark, S.M., Gabriel, R.L. and Ben-Amotz, D (2000). Educational applications of IR and Raman Spectroscopy: A comparison of experiment and theory. *Journal of Chemical Education*, 77 (5), 654-660.

115. Mordechai, J., Ramesh, J., Huleihel, M., Cohen, Z.K.O., Talyshinsky, M., Cahana, A., Salman, A., Sahu, R.K., Guterman, H. and Mordechai, S. (2004). Studies on acute human infections using FT-IR microspectroscopy and cluster analysis. *Biopolymers*, 73, 494-502.

116. Mourad, W.A., Tulbah, A., Shoukri, M., Al-Davel, F., Akhtar, M., Ali, M.A., Hainau, B. and Martin, J. (2003). Primary diagnosis and REAL/WHO

classification of non-Hodgkin's lymphoma by fine-needle aspiration: cytomorphologic and immunophenotypic approach. *Diagnostic Cytopathology*, 28, 191–5.

117. Mourad, W.A., Al-Nazer, M. and Tulbah, A. (2003). Cytomorphologic differentiation of Hodgkin's lymphoma and Ki-1+ anaplastic large cell lymphoma in fine needle aspirates. *Acta Cytol*, 47, 744–8.

118. Mourant, J.R., Gibson, R.R., Johnson, T.M., Carpenter, S., Short, K.W., Yamada, Y.R and Freyer, J.P (2003). Methods for measuring the infrared spectra of biological cells. *Physics in Medicine and Biology*, 48, 243-257.

119. Mourant, J.R., Yamada, Y.R. Carpenter, S., Dominique, L.R. and Freyer, J.P. (2003). FT-IR Spectroscopy Demonstrates Biochemical Differences in Mammalian Cell Cultures at Different Growth Stages. *Biophysical Journal*, 85 (3), 1938-1947.

120. Ng, L.M. and Simons, R. (1999). Infrared spectroscopy. *Analytical Chemistry*, 71, p343R–350.

121. Notingher, I., Bisson, I., Polak, J.M. and Hench, L.L. (2004). In situ spectroscopic study of nucleic acids in differentiating embryonic stem cells. *Vibrational Spectroscopy*, 35, 199-203.

122. Notingher, I., Jell, G., Notingher, P.L., Bission, I., Tsigkou, O., Polak, J.M., Steven, M.M. and Hench, L.L. (2005). Multivariate analysis of Raman spectra for in vitro non-invasive studies of living cells. *Journal of Molecular Structures*, 744-747, 179-185.

123. Olsen, R.J., Chang, C.C., Henrick, J.L., Zu, Y. and Ehsan, A. (2008). Acute leukaemia Immunohistochemistry: A systematic diagnostic approach. *Arch. Pathol. Lab. Med*.,132, 462-475.

124. Omberg, K., Osborn, J., Zhang, S., Freyer, J., Mourant, J., Schoonover, J. (2002). Spectroscopic study of human lung epithelial cells (A549) in culture: living cells versus dead cells. *Applied Spectroscopy*, 56, 813–819.

125. Paluszkiewicz, C. and Kwiatek, M. (2001) Analysis of human cancer prostate tissues using FT-IR microspectroscopy and SRIXE techniques. *Journal of Molecular Structure*, 565-566, 329-334.

126. Petibois, C., Rigalleau, V., Melin, A.M., Perromat, A., Cazorla, G., Gin, H. and Deleris, G. (1999) Determination of glucose in dried serum samples by Fourier-Transform Infrared Spectroscopy. *Clinical Chemistry*, 45 (9), 1530-1535.

127. Petiois C, Gionnet K, Goncalves M, Perromat A, Moenner M and Deleris G (2006). Analytical performances of FT-IR spectroscopy and imaging for

concentration measurements within biological fluids and cells. *The Analyst*, 131, 640-647.

128. Pevnser, A. and Diem, M. (2003). IR spectroscopic studies of major cellular components III. Hydration of protein, nucleic acid and Phospholipid Films. *Biopolymers (Biospectroscopy)*, 72, 282-289.

129. Pichardo-Molino, J.L., Frausto-Reyes, C., Barbosa-Garcia, O., Huerta-Franco, R., Gonzalez-Trujillo, J.L., Ramirez-Alvarado, C.A., Gutierrez-Juarez, G. and Gutierrez, M. (2007). Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers Med Sci*, 22, 229-236.

130. Polli, E. and Semenza, G (1955). Deoxyribonucleic acids from normal and leukaemic human leukocyte. *Inst. Lomb. Sci. Lett (Rend Sci).*, 89, 160-195.

131. Pui, C.H., Boyett, J.M., Relling, M.V., Harrison, P.L., Rivera, G.K., Behm, F.G., Sandlund, J.T., Ribeiro, R.C., Rubnitz, J.E., Gajjar, A., Evans, W.E. (1999). Sex differences in prognosis for children with acute lymphoblastic leukaemia. *Journal of Clinical Oncology*, 17, 18-824.

132. Puppels, G.J., Garritsen, H.S.P., Segers-Nolten, G.M.J., De Mul, F.F.M. and Greve, J. (1991). Raman microscopic approach to the study of human granulocytes. *Biophysical Journal*, 60, 1046-1056.

133. Rai, K.R., Sawitsky, A., Cronkite, E.P., Chanana, A.D., Levy, R.N., and Pasternack, B.S.(1975). Clinical staging of chronic lymophocytic leukaemia. *Blood*, 46, 219-234.

134. Ramesh, A., Salman, A., Hammody, Z., Cohen, B., Gopas, J., Grossman, N. and Mordechai, S. (2001). FT-IR Mivroscopic studies on normal and H-ras oncogene transfected cultured mouse fibroblasts. *Euro. Biophy. J.* 30, 250-5.

135. Ramesh, J., Kapelushnik, J., Mordehai, J., Mose, A., Huleihel, M., Erukhimovitch, V., Levi, C and Mordechai, S. (2002). Novel methodology for the follow-up of acute lymphoblastic leukaemia using FT-IR microspectroscopy. *Journal of Biochemical and Biophysical Methods*, 51 (3), 251-261.

136. Redaelli, A., Laskin, B.L., Stephens, J.M., Botteman, M.F., Pashos, C.L. (2004). The clinical epidemiological burden of chronic lymphocytic leukaemia. *Eur. J. Cancer. Care (England)*, 13, 279-287.

137. Rehman, S., Movasaghi, Z., Tucker, A.T, Joel, S.P., Darr, J.A., Ruban, A.V. and Rehman, I.U. (2007). Raman spectroscopic analysis of breast cancer tissues: identifying differences between normal, invasive ductal carcinoma and ductal carcinoma in situ of the breast tissue. *Journal of Raman Spectroscopy*, 38, 1345–1351.

138. Richard, S.J., and Jack, A.S. (2003). The development of integrated haematopathology laboratories: a new approach to the diagnosis of leukaemia and lymphoma. *Clin. Lab. Haem.*, 25, 337-342.

139. Rigas, B., LaGuardia, K., Qiao, L., Bhandare, P. S., Caputo, T. and Cohenford, M. A. (2000). Infrared spectroscopic study of cervical smears in patients with HIV: implications for cervical carcinogenesis. *Journal of Laboratory and Clinical Medicine,* 35, 26–31.

140. Rigas, B., Morgello, S., Goldman, I. and Wong, P. (1990). Human colorectal cancer displays abnormal Fourier-transform infrared spectra. *Proc. Natl. Acad. Sci.* 87, 8140-8144.

141. Rohleder, D., Kocherscheidt, G., Gerber, K., Kiefer, W., Kohler, W., Mocks, J., and Petrich, W. (2005). Comparision of mid-infrared and Raman spectroscopy in the quantitative analysis of serum. *Journal of Biomedical Optics*, 10, 3, 031108-1-10.

142. Rohleder, D., Kiefer, W. and Petrich, W. (2004). Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy. *Analyst***,** 129, 906 – 911.

143. Romeo, M., Mohlenhoff, B and Diem, M (2006). Infrared microspectroscopy of human cells: causes for the spectral variance of oral mucosa (buccal cells). *Vibrational Spectroscopy*, 42, 9-14.

144. Romeo, M., Mohlenhoff, B., Jennings, M. and Diem, M (2006). Infrared micro-spectroscopic studies of epithelial cells. *Biochimica et Biophysica Acta-Biomembranes*, 1758 (7), 915-922.

145. Romeo, M.J. and Diem, M (2005). Infrared spectral imaging of lymph nodes: strategies for anlysis and artefact reduction, *Vibrational Spectroscopy*, 38, 115-119.

146. Sahu, R.K., Argov, S., Salman, A., Huleihel, M., Grossman, N., Hammody, Z., Kapelushnik, J. and Mordechai, S. (2004). Characteristics absorbance of nucleic acids in the mid-IR region as possible common biomarkers for diagnosis of malignancy. *Technology in Cancer Research and Treatment*, 3 (6), 629-638.

147. Sahu, R.K., Argov, S., Salman, A., Zelig, U., Huleihel, M., Grossman, N., Hammody, Z., Kapelushnik, J and Mordechai, S. (2005). Can Fourier transform infrared spectroscopy at higher wavenumber (mid –IR) shed light on biomarkers for carcinogensis in tissue? *Journal of Biomedical Optics*, 10(5), 054017-1 - 054017-10.

148. Sahu, R.K., Zelig, U., Huleihel, M., Brosh, N., Talyshinsky, M., Ben-Harosh, M., Mordechai, S. and Kapelushnik, J. (2006). Continuous monitoring of WBC

(biochemistry) in an adult leukaemia patient using advanced FT-IR-spectroscopy. *Leukaemia Research*, 30 (6), 687-693.

149. Salman, A., Sahu, R.K., Bernshtain, E., Zelig, U., Goldstein, J., Walfisch, S., Argov, S and Mordechai, S (2004). Probing cell proliferation in the human colon using vibrational spectroscopy: a novel use of FT-IR micro spectroscopy. Vibrational spectroscopy, 34, 301-308.

150. Sarver, R.W. and Krueger, W.C. (1991). Protein Secondary Structure from Fourier Transform Infrared Spectroscopy: A Data Base Analysis. *Analytical Biochemical*, 194, 89-100.

151. Savitzky and Marcel J.E. Golay (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36, 1627–1639.

152. Scarisbrick, J.J., Pickard, D., Lee, A. and Yu, Raymond (2004). Elastic scattering spectroscopy in the diagnosis of melanocytic lesions: comparison with clinical and histopathological diagnosis. *Journal of the American Academy of Dermatology*, 50 (3), 6.

153. Schultz, C.P., Lui, K.Z., Johnston, J.B. and Mantsch , H.H. (1997). Prognosis of Chronic Lymphocytic Leukaemia from Infrared Spectra of Lymphocytes. *Journal of Molecular Structure*, 408/409, 253-256.

154. Schultz, C.P., Lui, K.Z., Johnston, J.B. and Mantsch, H.H. (1996). Study of Chronic Lymphocytic Leukaemia Cells by FT-IR Spectroscopy and Cluster Analysis. *Leukaemia Research*, 20 (8), 649-655.

155. Sharwani, W. Jerjes, V. Salih, B. Swinson, I. Bigio, M. El-Maaytah, C. Hopper (2006). Assessment of oral premalignancy using elastic scattering spectroscopy. *Oral Oncology*, 42 (4), 343-349.

156. Shaw R.A, Manfield J.R., Rempel S.P., Low-Ying S, Kupriyanov V.V, Mantsch H.H (2000). Analysis of biobediacl spectra and images: from data to diagnosis. *Journal of Molecular Structure (Theochem)*, 500, 129-138.

157. Shaw, R.A., Kotowich, S., Eysel, H.H., Jackson, M., Thomson, G.T. and Mantsch, H.H. (1995). Arthritis diagnosis based upon the near-infrared spectrum of synovial fluid. *Rheumatology International*, 15, 159-165.

158. Shaw, R.A., Kotowich, S., Leroux, M. and Mantsch, H.H. (1998). Multi-analyte serum analysis using mid-infrared spectroscopy. *Annals of Clinical Biochemistry*, 35, 624-632.

159. Shaw, R.A., Low Ying, S., Leroux, M. and Mantsch, H.H. (2000). Toward reagent-free clinical analysis: quantitation of urine urea, creatinine and total protein from the mid-infrared spectra of dried urine films. *Clinical Chemistry*, 46, 1493-1495

160. Shen, Y.C., Davies, A.G., Linfield, E.H., Elsey, T.S., Taday, P.F. and Arnone, D.D. (2003). The use of Fourier-transform infrared spectroscopy for the quantitative determination of glucose concentration in whole blood. *Phys. Med. Biol*., 48, 2023-2032.

161. Shetty, G., Kendall, C., Shepherd, N., Stone, N. and Barr, H. (2006) Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. *British Journal of Cancer*, 94, 1460-1464.

162. Siebert, F. and Hildebrandt, P (2008). Vibrational Spectroscopy in Life Science: Tutorials in Biophysics. WILEY-VCH, Weinheim.

163. Sindhuphak, R., Issaravanich, S., Udomprasertgul, V., Srisookho, P., Warakamin, S., Sindhuphak, Sirin., Boonbundarlchai, R. and Dusitsin, N. (2003). A new approach for the detection of cervical cancer in Thai women. *Gynecologic Oncology*, 90(1), 10-14.

164. Snook, R.D., Harvey, T.J., Faria, E,C.,and Gardner, P. (2009). Raman tweezers and their application to the study of singly trapped eukaryotic cells. *Integr. Biol.*, 1, 43 – 52.

165. Song, L.M. and Wilson, B.C. (2005). Optical detection of high-grade dysplasia in Barrett's Esophagus. *Techniques in Gastrointestinal Endoscopy*, 1096-2883, 78-88.

166. Stone, N., Stavroulaki, P., Kendall, C., Birchall, M. and Barr, H. (2000). Raman spectroscopy for early detection of laryngeal malignancy: Preliminary results. *The Laryngoscope*, 110, 1756-1763.

167. Susi, H. and Byler M.D. (1988). Fourier Deconvolution of the Amide I Raman of Proteins as Related to Conformation. *Applied Spectroscopy*, 42(5), 819-825.

168. Swinson, B., Jerjes, W., El-Maaytaah, M., Norris, P. and Hopper, C. (2006). Optical techniques in diagnosis of head and neck malignancy. *Oral Oncology*, 42 (3), 221-228.

169. Szczepanski, T., Vincent, H.J., and Jacques, J.M (2003). Classification system for acute and chronic leukaemia. *Best Practice & Research Clinical Haematology*, 16 (4), 561-582.

170. Tefferi, A. (2008). The history of myeloproloferative disorders: before and after Dameshek. *Leukaemia*, 22, 3-13.

171. Tenen, D.G. (2003), Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer*, 3, 89–101.

172. Tsiftsoglou, A.S., Pappas, I.S. and Vizirianakis, I.S. (2003). Mechanisms involved in the induced differentiation of leukaemic cells. *Pharmacology & Therapeutics*, 100, 257-290.

173. Vo-Dinh, T., Allain, L.R. and Stokes, D.L. (2002). Cancer gene detection using surface-enhanced Raman scattering (SERS). *Journal of Raman Spectroscopy*, 33, 511-516.

174. Wallace, M. B., Perelman, Backman, V., Crawford, J. M., Fitzmaurice, M., Seiler, M., Badizadegan,K., Shields, S. J., Itzkan, I., Dasari, R. R., van Damm J. and Feld, M. S. (2000). Endoscopic detection of dysplasia in patients with Barrett's esophagus using light-scattering spectroscopy. *Gastroenterology*, 119(3), 677-682.

175. Wang, H.P., Wang, H. C. and Huang, Y. J. (1997). Microscopic FT-IR studies of lung cancer cells in pleural fluid. *Science of the Total Environment*, 204, 283–287.

176. Wang, J.S., Shi, J.S., Xu, Y.Z., Duan, X.Y., Zhang, L., Wang, J., Yang, L.M., Weng, S.F., Wu, J.G. (2003). FT-IR spectroscopic analysis of normal and cancerous tissues of esophagus. *World Journal of Gastroenterology*, 9(9), 1897-9.

177. Wang, T.D., Triadafilopoulos, G., Crawford, J.M., Dixon, L.R., Bhandari, T., Sahbaie, P., Friedland, S., Soetikno, R., and Contag, C. H. (2007). Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. *Proc. Natl. Acad. Sci.*, 104(40), 15864–15869.

178. Wolkers, W.F., Balasubramanian, S.K., Ongstad, E.L., Zec, H.C. and Bischof, J.C. (2007). Effects of freezing on membranes and proteins in LNCaP prostate tumor cells. *Biochim Biophys Acta.*,1768(3), 728–736.

179. Wong, P., Wong, R., Caputo, T., Godwin, H., Rigas, B. (1991). Infrared spectroscopy of exfoliated human cervical cells: Evidence of extensive structural modifications at the molecular level. *Experimental and Molecular Pathology*, 55, 269-284.

180. Wood, B. R., Chiriboga, L., Yee, H., Quinn, M. A., McNaughton, D. and Diem, M. (2004). Fourier transform infrared (FT-IR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium. *Gynecologic Oncology*, 93, 59-68.

181. Wood, B., Quinn, M., Burden, F. and McNaughton, D. (1996). An investigation into FT-IR spectroscopy as a bio-diagnostic tool for cervical cancer. *Biospectroscopy*, 2, 001-011.

182. Wood, B.R. and McNaughton, D. (2002). Raman excitation wave;ength investigation of single red blood cells in vivo. *Journal of Raman Spectroscopy*, 33, 517-523.

183. Wood, B.R., Quinn, M.A., Tait, B., Ashdown, M., Hislop, T., Romeo, M. and McNaughton, D (1998). *Biospectroscopy*, 4, 75-91.

184. Xie, Y., Davies, S.M. and Xiang, Y. (2003). Trends in leukaemia: Incidence and survival in the United States (1973-1998). *Cancer*, 97, 2229-35.

185. Xun-Ling, Y., Rui-Xin, D., Lei, Z., Xue-Jun, Z., Zong-Wang, Z. (2005). Raman spectra of single cell from gastrointestinal cancer patients. World Journal of Gastroenterology, 11(21), p. 3290-3292.

186. Yano, K., Ohoshima, S., Gotou, Y., Kumaido, K., Moriguchi, T. and Katayama, H.(2000). Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool? *Analytical Biochemistry*, 15 (2), 218-225.

187. Young, N.A. and Al-Saleem, T. (1999). Diagnosis of lymphoma by fine-needle aspiration cytology using the revised European–American classification of lymphoid neoplasms. *Cancer*, 87, 325–45.

Websites

www.thermo.com

www.naturalypure.com/HerbalLymphaticSystemMedicine.htm

www-sdc.med.nagasaki-u.ac.jp/n50/disaster/Leukemia-E.html

www.droid.cuhk.edu.hk/web/specials/lymph_nodes/lymph_nodes.htm

www.med-ed.virginia.edu/courses/path/innes/wcd/lymphoma.cfm

www.biomedcentral.com

Appendix

Appendix A

1. Lymphatic System

The lymphatic system - the target of lymphomas - is a subsystem of the circulatory system. The components of the lymphatic system are divided into two groups: the primary organs and the secondary organs. The bone marrow and the thymus gland are primary organs. They regulate the production blood-forming (haematopoietic) elements and differentiation of lymphocytes that make up the immune system of the body. The secondary organs include the lymphatic (a network of very fine vessels that connect the lymph nodes scattered throughout the body), lymph nodes, aggregated lymphoid tissue and spleen. These secondary organs are involved in lymphatic functions including; assisting in circulating body fluids by transporting excess fluids away from interstitial spaces in body tissue and return it to the bloodstream to prevent the backflow of the lymph fluid and transporting infection-fighting cells called lymphocytes to combat bacteria and foreign entities and removes cell debris by phagocytosis forming part of the body's immune system.

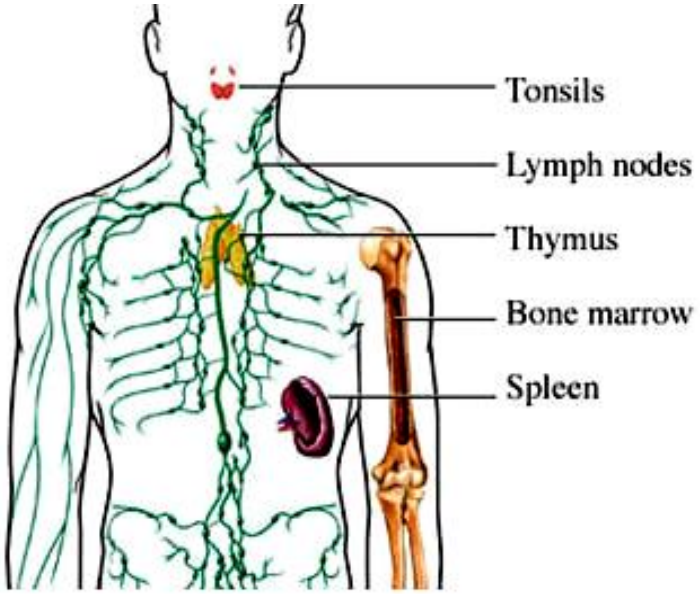


Figure 1: The lymphatic system.
Source: www.naturalypure.com/HerbalLymphaticSystemMedicine.htm

2. Bone Marrow

Bone marrow (or "medulla ossea") is the soft tissue comprising the cavities of large bones and is the most active haematopoietic organ in the body. It is a network of connective tissue fibres, adipocytes, blood vessels, and blood-producing (haematopoietic) stem cells (figure 21). As needed, the HSCs differentiate to become a particular type of cell- a RBC (erythrocytes), platelet (thrombocytes) or WBC (lymphocytes), and only after maturation these cells are released from the myeloid tissue of the bone marrow into the blood stream. The immature T-lymphocytes move to the thymus for final development, but the B- lymphocytes remains in the bone marrow during maturation. After maturation, both B- and T- lymphocytes circulate in the lymph and accumulate in secondary lymphoid organs, where they await recognition of antigens.

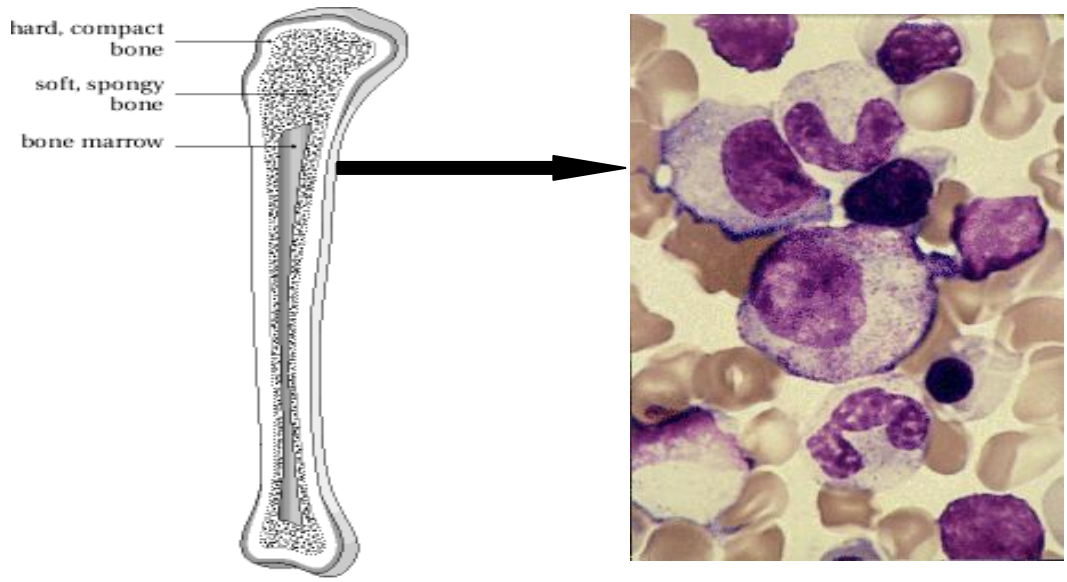


Figure 2: Structure and characteristic appearance of bone marrow tissue.
Source: www-sdc.med.nagasaki-u.ac.jp/n50/disaster/Leukemia-E.html

B-lymphocytes play a large role in the humoral immune response. They react against antigens (invading bacteria or viruses and toxins) that produce an immune response in the body by making antibodies. Antibodies are antigen specific, and the immune system is able to remember each antigen it fights. Once a B-cell makes antibodies

against a certain antigen, e.g., bacteria, it keeps a memory of that antigen. If the antigen appears again, the B-cell can produce a large number of antibodies very rapidly. In this way, a second infection with that bacterium is often prevented.

T cells play a central role in cell-mediated immunity. There are four different types of T cells with different immunological functions:

- Cytotoxic T cells destroy virally infected cells and tumor cells, and are also implicated in transplant rejection.
- Helper T cells divide rapidly and secrete cytokines that regulate or "help" the immune response. They also enhance the production of antibodies by B cells.
- Inflammatory T cells that recrit macrophages and neutrophils to the site of infection or other tissue damage.
- Natural Killer T cells posses the ability to recognise and luse virus infected cells and certain tumour cells as a natural function without requiring antibodies or other mediation. They are effective early in the course of viral infection, and may limit the spread of infection during this early stage.

3. Lymph nodes

Lymph nodes (small bean-shaped organ, 2.5 cm in diameter) are specialized collections of lymph tissue which are supported within by a meshwork of connective tissue called reticulin fibres. Lymph nodes occur along the entire length of the lymphatic system and tend to increase in size as they become closer to the thoracic duct. They appear to be strategically distributed in chains or clusters in the neck, armpits, pelvis, abdomen and groin. Lymph nodes are enclosed by a capsule of connective tissue and comprised of several lymph nodules, which are populated by dense aggregates of lymphocytes (T-cells and B-cells) and macrophages that collect and destroy bacteria and viruses as well as many cells in the precancerous stages of development. The nodules are separated by spaces called "lymph sinuses." The lymph node filters lymph which is brought to the cortex of the node by the afferent lymphatics**,** and there are several per node. The lymph is then filtered for antigens and particulate matter in the cortex, and an immune response

is generated, if necessary. The filtered lymph leaves the node through one or two efferent vessels near an indentation called the "hilum." Blood vessels also enter and exit the node at the hilum.

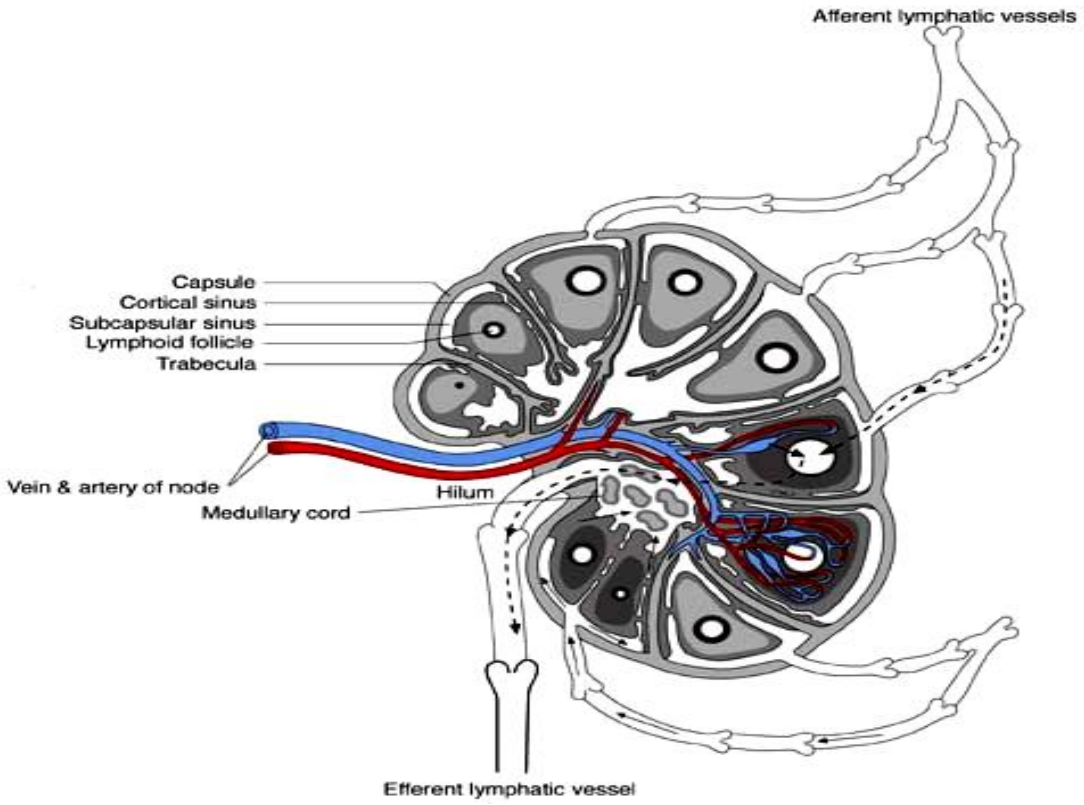
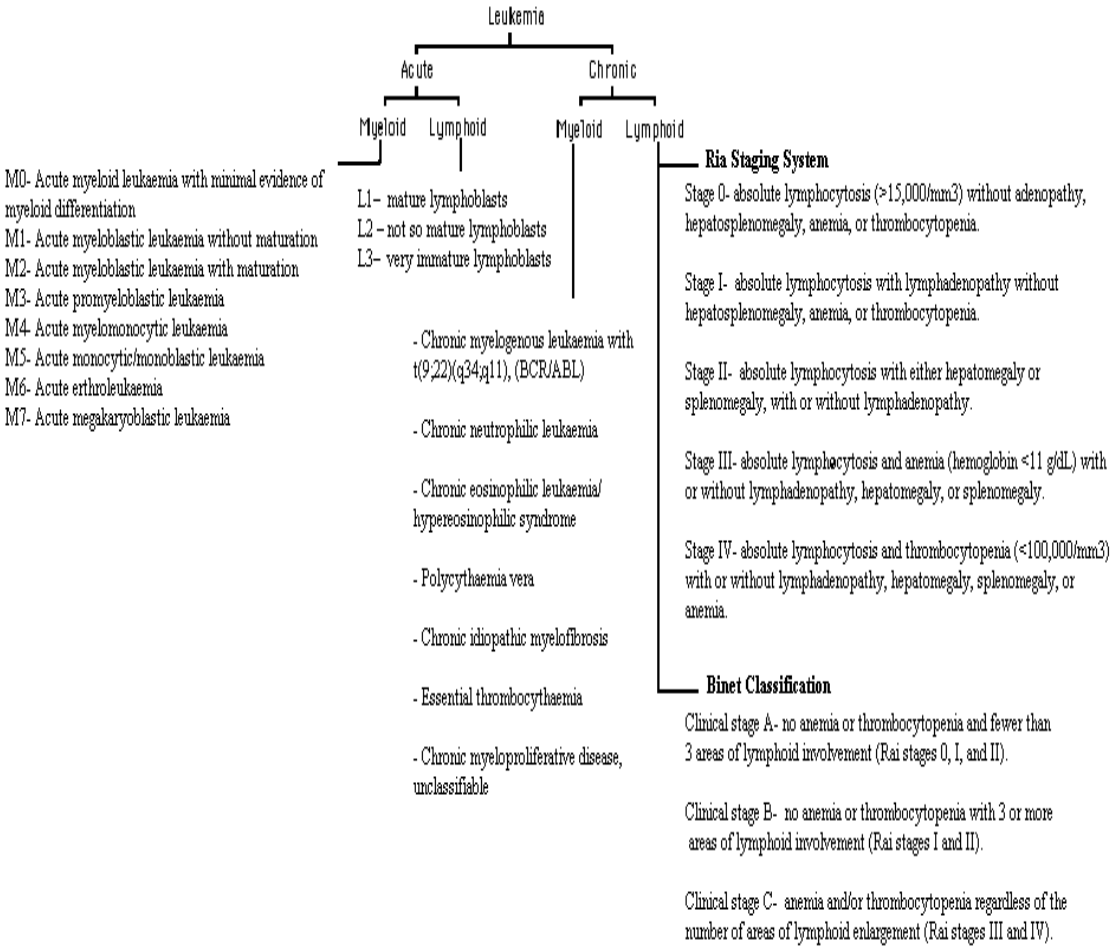


Figure 3: Illustration of a lymph node.
Source: www.droid.cuhk.edu.hk/web/specials/lymph_nodes/lymph_nodes.htm

Appendix B


Leukemia
Acute
Chronic
Myeloid
Lymphoid
Myeloid
Lymphoid
M0- Acute myeloid leukaemia with minimal evidence of myeloid differentiation
M1- Acute myeloblastic leukaemia without maturation
M2- Acute myeloblastic leukaemia with maturation
M3- Acute promyeloblastic leukaemia
M4- Acute myelomonocytic leukaemia
M5- Acute monocytic/monoblastic leukaemia
M6- Acute erthroleukaemia
M7- Acute megakaryoblastic leukaemia
L1– mature lymphoblasts
L2 – not so mature lymphoblasts
L3– very immature lymphoblasts
- Chronic myelogenous leukaemia with t(9;22)(q34;q11), (BCR/ABL)
- Chronic neutrophilic leukaemia
- Chronic eosinophilic leukaemia/ hypereosinophilic syndrome
- Polycythaemia vera
- Chronic idiopathic myelofibrosis
- Essential thrombocythaemia
- Chronic myeloproliferative disease, unclassifiable
Ria Staging System
Stage 0- absolute lymphocytosis (>15,000/mm3) without adenopathy, hepatosplenomegaly, anemia, or thrombocytopenia.
Stage I- absolute lymphocytosis with lymphadenopathy without hepatosplenomegaly, anemia, or thrombocytopenia.
Stage II- absolute lymphocytosis with either hepatomegaly or splenomegaly, with or without lymphadenopathy.
Stage III- absolute lymphocytosis and anemia (hemoglobin <11 g/dL) with or without lymphadenopathy, hepatomegaly, or splenomegaly.
Stage IV- absolute lymphocytosis and thrombocytopenia (<100,000/mm3) with or without lymphadenopathy, hepatomegaly, splenomegaly, or anemia.
Binet Classification
Clinical stage A- no anemia or thrombocytopenia and fewer than 3 areas of lymphoid involvement (Rai stages 0, I, and II).
Clinical stage B- no anemia or thrombocytopenia with 3 or more areas of lymphoid involvement (Rai stages I and II).
Clinical stage C- anemia and/or thrombocytopenia regardless of the number of areas of lymphoid enlargement (Rai stages III and IV).

Appendix C

Clinical staging of both HL and NHL is used for determining the extent of the disease and is done once the diagnosis is confirmed. Clinical stages depend on the number of lymph nodes affected, location and distribution of the affected lymph nodes and how far it has spread throughout the body (figure 8). This gives rise to four clinical stages according to the Ann Arbour Staging classification system:

- **Stage I:** tumour hasn't spread. Apparent involvement of a single lymph node region or a single organ, such as bone.
- **Stage II:** More than one tumour, all found in lymph nodes on the same side of the diaphragm. Stage II can also mean that a tumour has spread to another organ, but that it is close to the original lymph node tumour.
- **Stage III:** More than one tumour with the tumours found on different sides (above and below) of the diaphragm. There may be tumours in the spleen or more than one tumour in nearby organs.
- **Stage IV:** Many tumours spread throughout an organ such as the liver or stomach, as well as in the lymph nodes. Stage IV can also refer to a single tumour in another organ, plus tumours in distant lymph nodes.

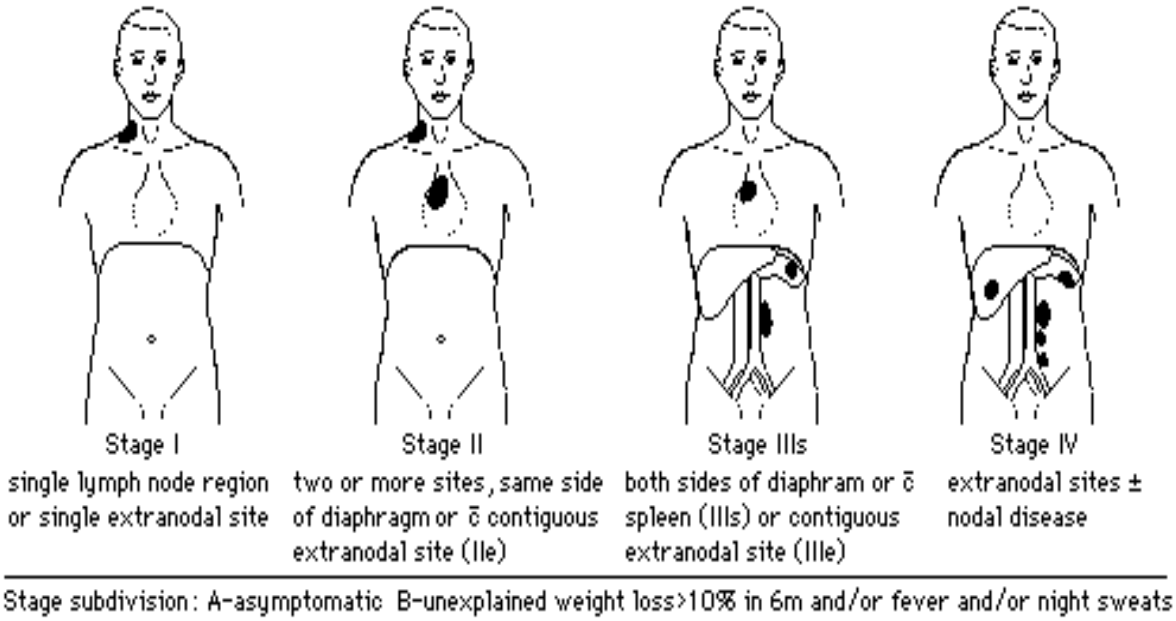


Figure 1: Clinical staging of Hodgkin's and non-Hodgkin lymphoma.
Source: www.med-ed.virginia.edu/courses/path/innes/wcd/lymphoma.cfm

Appendix D

Table 1: Infrared Peak Assignment

v: stretching vibrations, δ: bending (scissoring) vibrations, s: symmetric, as: asymmetric

Molecular Motion	**Wavenumber (cm^{-1})**	**References**
v(C=O): Amide I,	1652-1656	Schultz C.P et al., 1996
	1643	Erukhimovitch. V et al., 2005
	1643	Erukhimovitch. V et al., 2002
	1600-1720	Deleris and Petibois 2003
	1589-1713	Deleris and Petibois 2005
	1620-1690	Lasch. P et al., 2002
	1640-1660	Benedetti et al., 1997
	1640	Ahmad et al., 2000
	1655	Nevilliappan et al., 2002
	1645	Boydston-White et al., 2005
	1650	Boydston-White et al., 1999
	1658	Chiriboga et al., 1998
	1655	Chiriboga et al., 1998
	1657	Sahu et al., 2005
	1650	Diem et al., 2004
	1635	Colley et al., 2004
	1650	Diem et al., 2000
	1640	Gasparri et al., 2003
	1652	Lui et al., 1997
	1645	Moss et al., 2005
	1610-1690	Jackson et al., 1995
	1645	Krafft et al., 2005
	1650	Beleites et al., 2005
	1655	Liu Kan-Zhi et al., 2002
	1655	Benedetti et al., 1998
	1620-1690	Andrus and Strickland 1998
	1620-1690	Mourant et al., 2003
α-helices	1645-1662	Byler and Susi 1988
ß-sheet	1613-1637	Byler and Susi 1988
turns	1662-1682	Byler and Susi 1988
coils	1637-1645	Byler and Susi 1988
δ(N-H): Amide II, α-helix, proteins	1540	J.Sule-Suso et al., 2005
	1544	Erukhimovitch. V et al., 2005
	1544	Erukhimovitch. V et al., 2002
	1480-1600	Deleris and Petibois 2003
	1474-1589	Deleris and Petibois 2005
	1530-1570	Lasch. P et al., 2002
	1530-1545	Benedetti et al., 1997
	1545	Ahmad et al., 2000

	1545	Nevilliappan et al., 2002
	1540	Boydston-White et al., 2005
	1530	Boydston-White et al., 1999
	1541	Chiriboga et al., 1998
	1545	Sahu et al., 2005
	1530	Colley et al., 2004
	1550	Diem et al., 2000
	1538	Gasparri et al., 2003
	1542	Lui et al., 1997
	1550	Jackson et al., 1995
	1543	Krafft et al., 2005
	1530	Beleites et al., 2005
	1546	Liu Kan-Zhi et al., 2002
	1549	Benedetti et al., 1998
	1530-1570	Andrus and Strickland 1998
Amide III (C-N stretching and N-H bending)	1300	J.Sule-Suso et al., 2005
	1254	Beleites et al., 2005
	1270	Erukhimovitch. V et al., 2005
	1240-1340	Lasch. P et al., 2002
	1200-1330	Benedetti et al., 1997
	1240-1310	Ahmad et al., 2000
	1245	Boydston-White et al., 1999
	1307/1242	Chiriboga et al., 1998
	1304	Krafft et al., 2005
	1230	Benedetti et al., 1998
	1240-1340	Andrus and Strickland 1998
	1240-1340	Gasparri et al., 2003
CH_3 deformation	1450	J.Sule-Suso et al., 2005
CH_3 asymmetric deformation	1455	Schultz C.P et al., 1996
	1455	Lasch. P et al., 2002
	1455	Ahmad et al., 2000
δCH_2 lipid acyl bending	1468	Schultz C.P et al., 1996
	1467	Liu Kan-Zhi et al., 2002
PO_2^- symmetric stretching	1080	J.Sule-Suso et al., 2005
	1080	Nevilliappan et al., 2002
	1087	Schultz C.P et al., 1996
	1082	Erukhimovitch. V et al., 2002
	1095	Chiriboga et al., 1998
	1080	Gasparri et al., 2003
	1000-1150	Ramesh et al., 2002
	1083	Lasch. P et al., 2002
	1084-1088	Ahmad et al., 2000
	1087	Lui et al., 1997
	1000-1100	Ramesh et al., 2002
	1084	Jackson et al., 1995
	1088	Liu Kan-Zhi et al., 2002
PO_2^- asymmetric stretching	1240	J.Sule-Suso et al., 2005

	1240	Schultz C.P et al., 1996
	1237	Lasch. P et al., 2002
	1220-1250	Ahmad et al., 2000
	1240	Nevilliappan et al., 2002
	1243 (1200-1400)	Erukhimovitch. V et al., 2002
	1170-1313	Ramesh et al., 2002
	1230	Chiriboga et al., 1998
	1237	Colley et al., 2004
	1240	Gasparri et al., 2003
	1240	Lui et al., 1997
	1200-1245	Ramesh et al., 2002
	1226	Jackson et al., 1995
	1242	Liu Kan-Zhi et al., 2002
DNA	966	Schultz C.P et al., 1996
	968	Lasch. P et al., 2002
	970	Benedetti et al., 1997
	970	Nevilliappan et al., 2002
	965	Boydston-White et al., 1999
	965	Chiriboga et al., 1998
	966	Sahu et al., 2005
	965	Gasparri et al., 2003
	966	Lui et al., 1997
	950	Mourant et al., 2003
	965	Ramesh et al., 2002
	965	Ramesh et al., 2003
Nucleic acids (DNA:RNA)	1020	Andrus and Strickland 1998
	1000-1200	Ramesh et al., 2002
	1060-1095	Boydston-White et al., 1999
	1000-1150	Chiriboga et al., 1998
	1080	Sahu et al., 2005
	1080/1235	Diem et al., 2004
	1000/1080	Colley et al., 2004
	1000/1140	Gasparri et al., 2003
	1070/1100	Mourant et al., 2005
	1080/1234	Diem et al., 2002
	1020/1121	Ramesh et al., 2003
	1080/1240	Beleites et al., 2005
RNA $\nu(C_2=O)$	1960-1705	Benedetti et al., 1997
$\nu\delta$ ring of Uracyl	996	Benedetti et al., 1997
ν(C-O) ribose	1015	Benedetti et al., 1997
	1121	Andrus and Strickland 1998
	1085	Boydston-White et al., 1999
ν(C=O)lipids, cholesterol esters, triglyceride	2850 and 2920	Schultz C.P et al., 1996
	1732-1739	Deleris and Petibois 2003
	1713-1739	Deleris and Petibois 2005

	1737 1700-1750 2800-3000 2850 2800-2950 1730 1742 1740	Lasch. P et al., 2002 Erukhimovitch. V et al., 2002 Andrus and Strickland 01998 Mourant et al., 2005 Mourant et al., 2003 Beleites et al., 2005 Mourant et al., 2003 Lasch et al., 2002
v(C-C/C-O): saccharides, glucose, lactate, glycerol C = O (ester) stretch	1397 900-1300 900-1300 1061-1138 1150 1170 900-1500 1099	Schultz C.P et al., 1996 Erukhimovitch. V et al., 2005 Deleris and Petibois 2003 Deleris and Petibois 2005 Lasch. P et al., 2002 Colley et al., 2004 Diem et al., 2000 Mourant et al., 2003
v(COO^-): amino acids	1360-1430 1360-1430 1370-1420 1397 1395 1400	Erukhimovitch. V et al., 2005 Deleris and Petibois 2003 Deleris and Petibois 2005 Lasch. P et al., 2002 Chiriboga et al., 1998 Jackson et al., 1995
δas (CH_3), δas (CH_2),: fatty acids, phospholipids and triglyceride δs (CH_2): δs (CH_3), δs (CH_2)	1430-1480 1430-1480 1468 1400-1455 1448 1452/1467 1467 1468 1446 1378 1352 1352	Erukhimovitch. V et al., 2005 Deleris and Petibois 2005 Lasch. P et al., 2002 Nevilliappan et al., 2002 Colley et al., 2004 Lui et al., 1997 Jackson et al., 1995 Beleites et al., 2005 Liu Kan-Zhi et al., 2002 Liu Kan-Zhi et al., 2002 Lasch et al., 2002 Erukhimovitch et al., 2002
δNH_2 (amino acids)	1592 1560-1630 1455 1400 1517 (tyrosine) 1582 (arginine)	Erukhimovitch. V et al., 2005 Deleris and Petibois 2003 Chiriboga et al., 1998 Mourant et al., 2003 Jackson et al., 1995 Jackson et al., 1995
Carbohydrate Carbohydrate: C-OH stretching	1056 1050 1155	Erukhimovitch. V et al., 2005 Lasch. P et al., 2002 Nevilliappan et al., 2002
v(CH): unsaturated fatty acids, cholesterol esters	3000-3020 3000-3020	Deleris and Petibois 2003 Deleris and Petibois 2005

vs (CH) saturated fatty acids	3081 2934	Colley et al., 2004 Colley et al., 2004
vas(CH_3): (methyl) cholesterol esters, triglyceride	2950-2990 2950-2990	Deleris and Petibois 2003 Deleris and Petibois 2005
vas(CH_2): long chain fatty acids, phospholipids	2880-2950 2880-2950 2870-2960 2945-2980 2800-3000 2926 2850	Deleris and Petibois 2003 Deleris and Petibois 2005 Nevilliappan et al., 2002 Gasparri et al., 2003 Ramesh et al., 2002 Kan-Zhi Liu., 2002 Lasch. P et al., 2002
vs(CH_3) cholesterol esters, triglyceride, glycerol	2860-2880 2860-2880 2800-3000 2800-3000	Deleris and Petibois 2003 Deleris and Petibois 2005 Sahu et al., 2005 Ramesh et al., 2002
vs(CH_2): Fatty acids, phospholipids and triglyceride vas(CH_2): vs(CH_2):	2819-2996 2920 2800-3000 2800-3000 2830-2870 2800-3000 2851/2924 2852 2852, 2874	Deleris and Petibois 2003 Lasch. P et al., 2002 Ramesh et al., 2002 Sahu et al., 2005 Deleris and Petibois 2003 Sahu et al., 2005 Colley et al., 2004 Kan-Zhi Liu., 2002 Lasch et al., 2002
-CO-O-C asymmetric stretch	1173	Lasch. P et al., 2002
-CO-O-C symmetric stretch	1063	Lasch. P et al., 2002
Albumin	1488-1600	Deleris and Petibois 2003
Glucose	997-1062	Deleris and Petibois 2003
Fibrinogen	1363-1428	Deleris and Petibois 2003
Immunoglobulin- G_2	1622-1652	Deleris and Petibois 2003
Lactate	1115-1134	Deleris and Petibois 2003
Glycerol	837-882	Deleris and Petibois 2003
Triglyceride	1091-1122 2800-3200	Deleris and Petibois 2003 Kan-Zhi Liu., 2002
Cholesterol	1033-1074 1700-1800 2800-3000	Deleris and Petibois 2003 Kan-Zhi Liu., 2002 Kan-Zhi Liu., 2002
Cholesterol esters	1147-1186	Deleris and Petibois 2003
Immunoglobulin- G_1	1316-1419	Deleris and Petibois 2003
α_{1-} Antitrypsin	691-712	Deleris and Petibois 2003
α_{2-} Macroglobulin	983-1116	Deleris and Petibois 2003
Transferin	1292-1338	Deleris and Petibois 2003
Apolipoprotein- A_1	1427-1484	Deleris and Petibois 2003

Urea	1632-1652	Deleris and Petibois 2003
Apolipoprotein- B	2835-2862	Deleris and Petibois 2003
Immunoglobulin- M	1360-1428	Deleris and Petibois 2003
Apolipoprotein- C_3	1660-1676	Deleris and Petibois 2003
Immunoglobulin- A	1372-1418	Deleris and Petibois 2003
Immunoglobulin- G_4	1505-1538	Deleris and Petibois 2003
Immunoglobulin- G_3	1628-1652	Deleris and Petibois 2003
Immunoglobulin- D	1464-1479	Deleris and Petibois 2003
Haptoglobulin	1378-1428	Deleris and Petibois 2003
α_{1-} Acid glycoprotein	1279-1337	Deleris and Petibois 2003
Fatty acyl moieties	2819-2996	Deleris and Petibois 2003
Amino acids	1369-1430	Deleris and Petibois 2003
Glycogen C-O stretching or C-OH bending	1025-1047	Nevilliappan et al., 2002
C-OH groups of serine, threonine and tyrosine (cell proteins)	1170	Nevilliappan et al., 2002
Collagen	1031-1082 1204,1240 and 1283	Andrus and Strickland 1998 Jackson et al., 1995
Glycogen	1025 900-1185 1150 950-1200 1300 1280 900-1200	Andrus and Strickland 1998 Sahu et al., 2005 Diem et al., 2000 Diem et al., 2002 Mourant et al., 2003 Chiriboga et al., 1998 Chiriboga et al., 1998

Appendix E

Normal blood components spectra

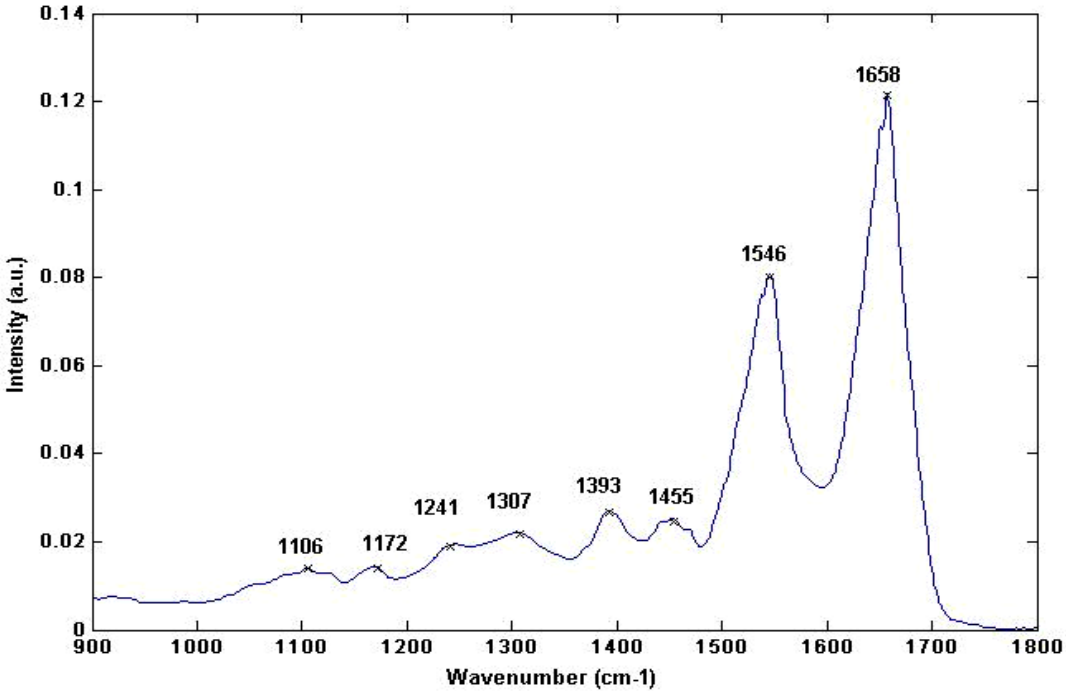


Figure 1: Represents the average IR spectra for whole blood (normal) in the range 900-1800 cm^{-1}

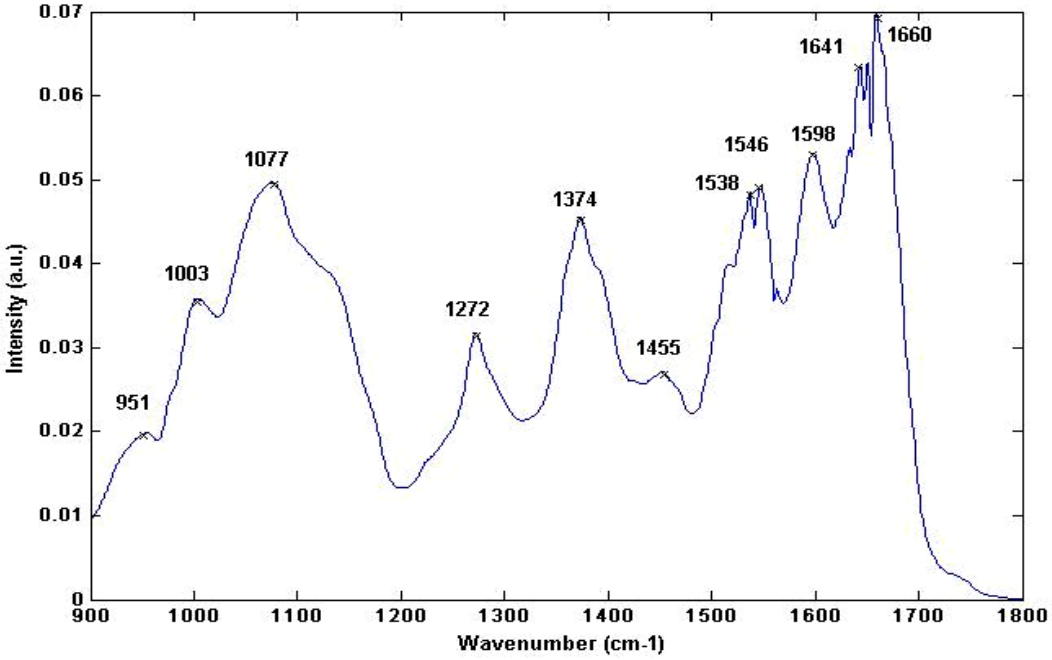


Figure 2: Represents the average IR spectra for plasma (normal) in the range 900-1800 cm^{-1}

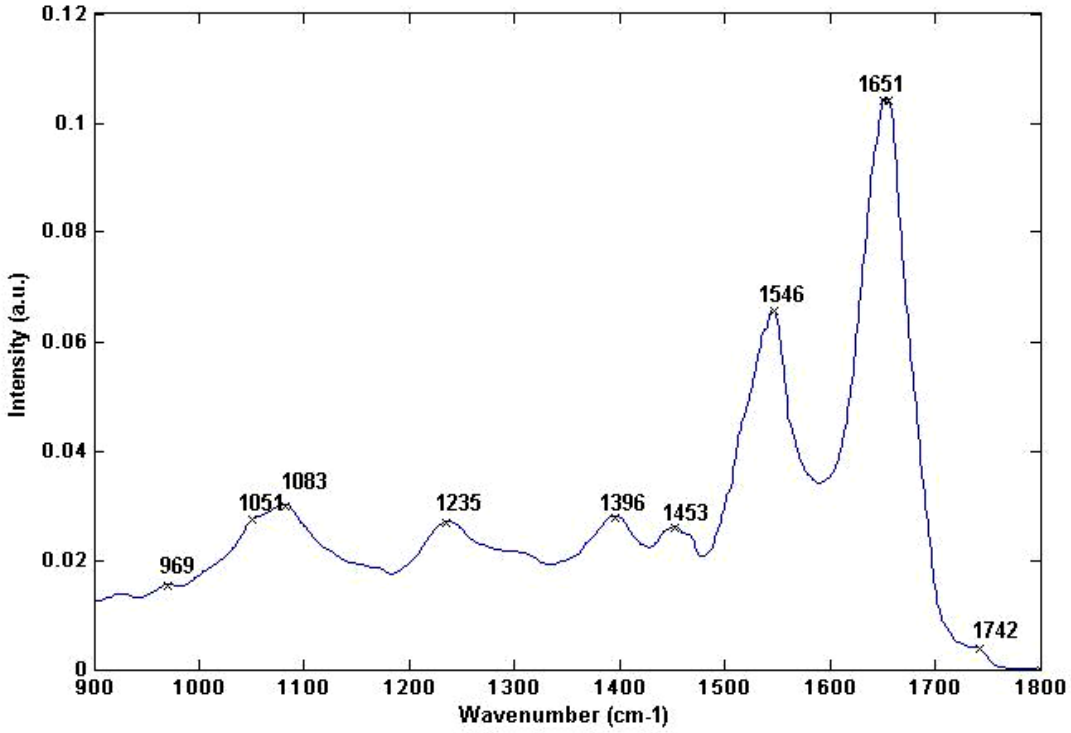


Figure 3: Represents the average IR spectra for buffycoat (normal) in the range 900-1800 cm^{-1}

CLL Blood Component Spectra

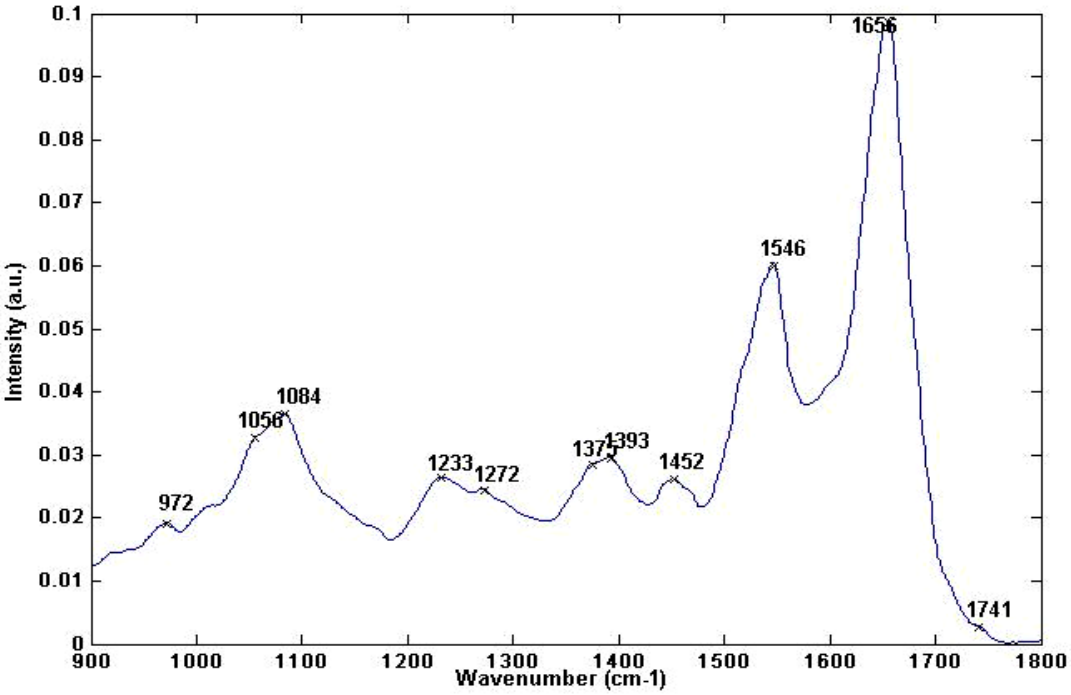


Figure 4: Represents the average IR spectra for buffycoat (CLL) in the range 900-1800 cm^{-1}

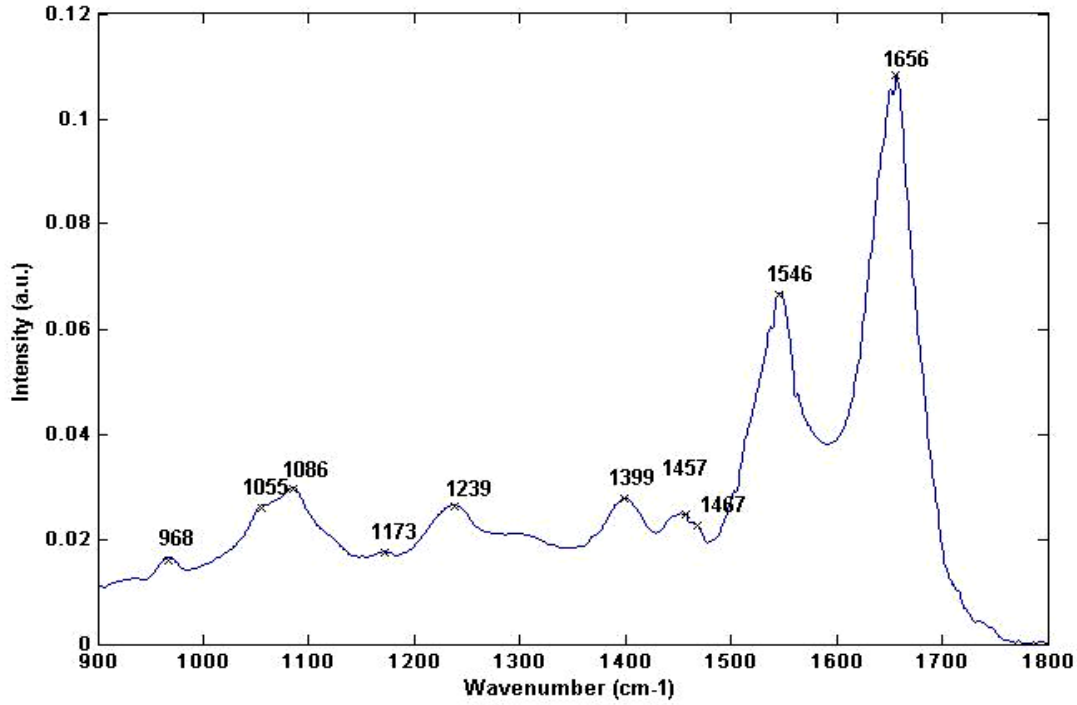


Figure 5: Represents the average IR spectra for plasma (CLL) in the range 900-1800 cm^{-1}

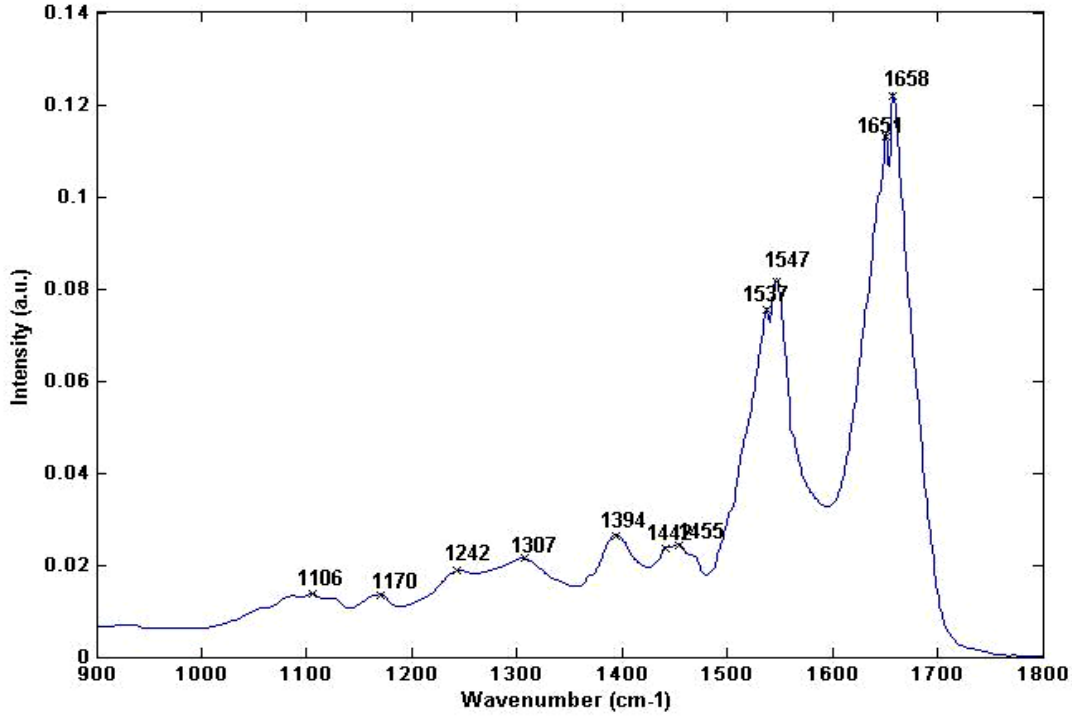


Figure 6: Represents the average IR spectra for whole blood (CLL) in the range 900-1800 cm^{-1}

Appendix F

Principal Components	F-Values			% Variance		
	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3
1	9.9298	11485.1	382.4847	55.2016	85.5357	59.6138
2	1.8567	0.596	143.185	74.5043	95.2741	81.6748
3	3.0947	4.546	177.9044	88.7037	98.6421	88.6043
4	46.7434	0.9492	84.8351	93.3296	99.3189	93.0126
5	1217.807	0.1339	6.8547	95.3255	99.5061	95.3197
6	36.1405	0.0006	134.7953	96.5466	99.6603	96.4974
7	9.2294	2.2741	22.8498	97.4737	99.79	97.4897
8	187.5219	0.0352	29.9669	98.0436	99.8683	98.1046
9	21.3296	0.1529	62.683	98.5315	99.91	98.5419
10	5.2501	0.0178	41.3933	98.8596	99.9298	98.7856
11	34.547	0.0406	12.6872	99.0678	99.9441	99.0038
12	142.087	0.0057	69.7272	99.2148	99.957	99.2032
13	63.6541	0.0028	0.0014	99.3255	99.965	99.329
14	49.8919	0.0715	14.1777	99.4199	99.9714	99.429
15	3.0017	0.0024	0.0604	99.4956	99.9764	99.5101
16	14.9621	0.0004	0.8801	99.5598	99.98	99.5748
17	16.3289	0.0056	1.2848	99.6099	99.9834	99.6233
18	0.1409	0.0044	0.0594	99.6512	99.9855	99.6625
19	0.0364	0.0002	18.6598	99.6865	99.987	99.6955
20	0.1748	0.0474	0.1224	99.7211	99.9883	99.724
21	0.0003	0.0074	4.5083	99.7507	99.9894	99.7503
22	10.9174	0.0018	3.0323	99.7765	99.9905	99.7724
23	3.5482	0.0154	0.954	99.799	99.9913	99.7907
24	7.8314	0.0057	0.1705	99.8173	99.992	99.8056
25	3.7248	0.0127	0.9301	99.8317	99.9926	99.8175

Table 1: ANOVA F-value and percentage of variance corresponding to the first 25 PCs for datasets 1, 2 and 3. The highest F- values highlighted in red represent the most significant PCs, which are then plotted against one another in a two-domensional scatter plot to visualise the natural clustering of cell line populations.

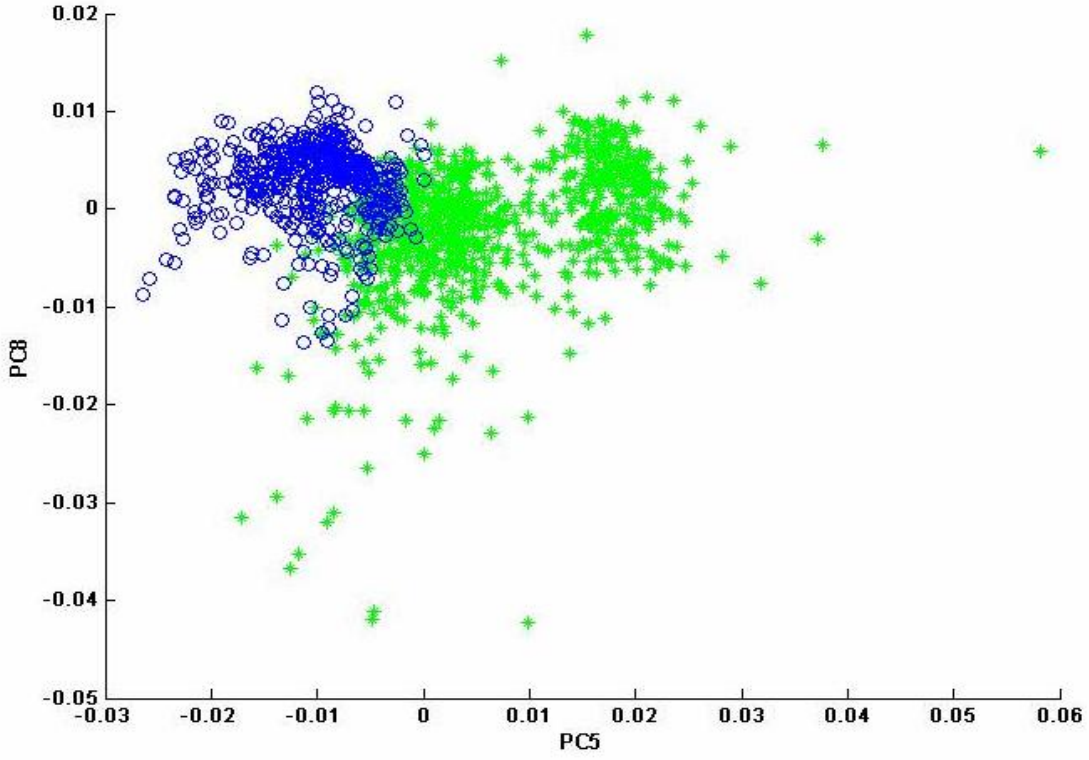


Figure 1: PC scores scatter plot (PC5 vs. PC8) represents clustering of normal and CLL whole blood (dataset 1). Each symbol represents one spectrum. Normal whole blood= green stars and CLL whole blood = blue circles

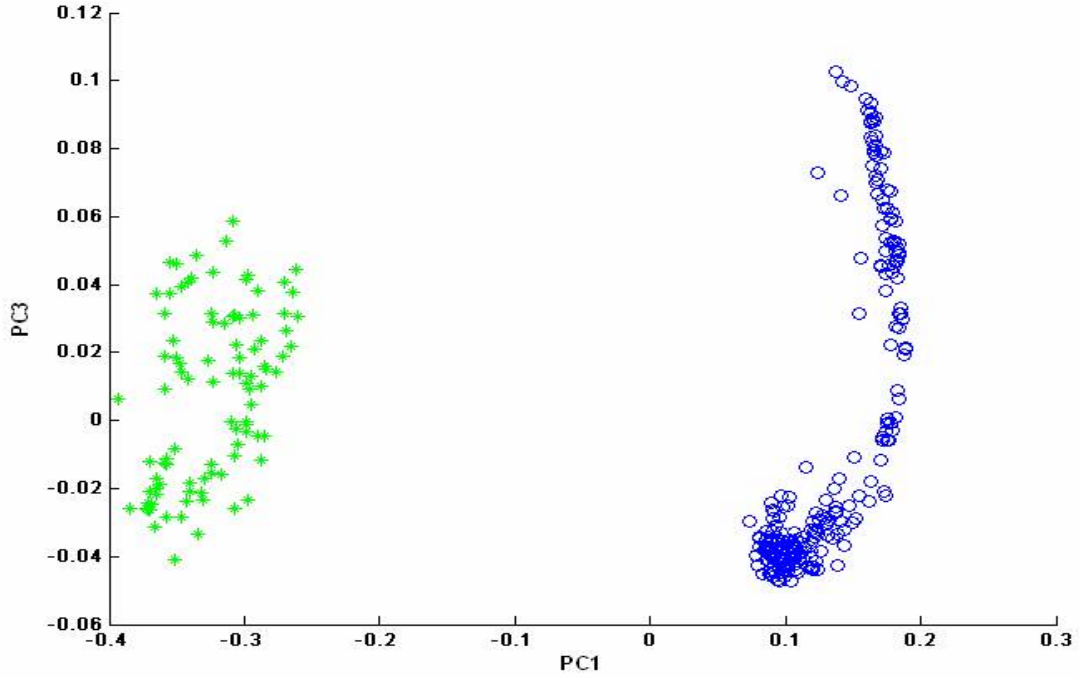


Figure 2: PC scores scatter plot (PC1 vs. PC3) represents clustering of normal and CLL whole blood.(dataset 2). Normal plasma= green stars and CLL plasma = blue circle

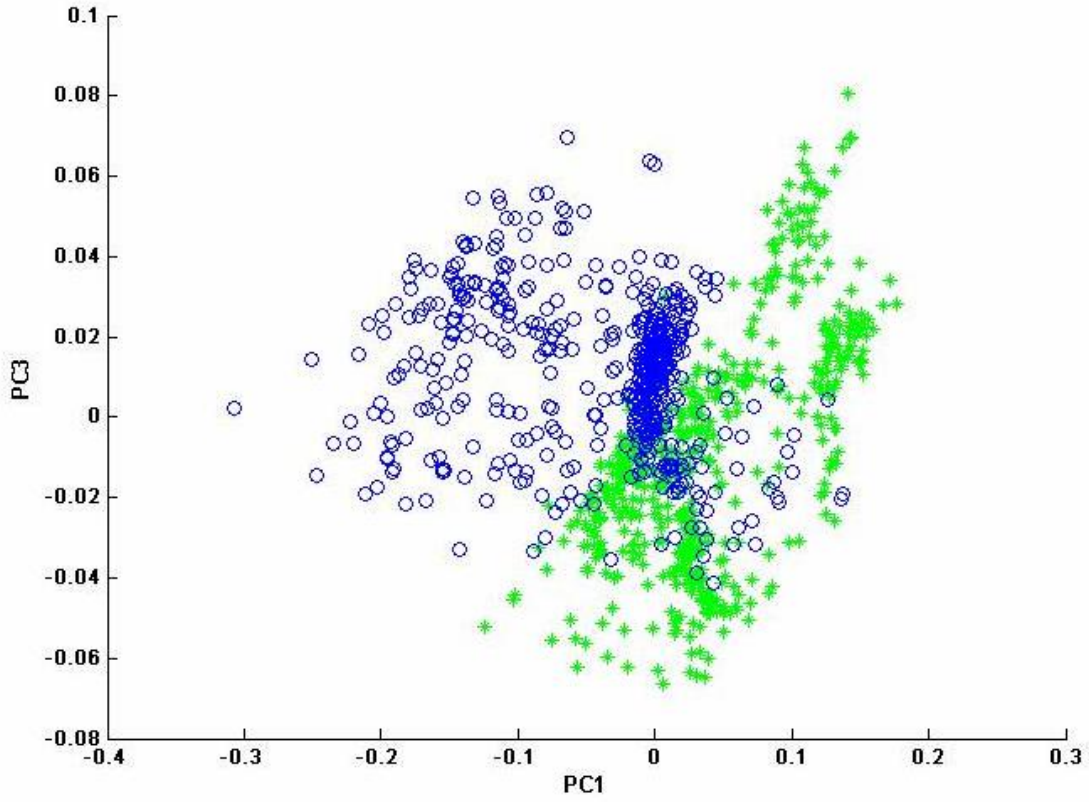


Figure 3: PC scores scatter plot (PC1 vs. PC3) represents clustering of normal and CLL whole blood.(dataset 2). Normal buffycoat= green stars and CLL buffycoat = blue circles

Appendix G

List of Publications and Posters

Publications

Babrah, J., McCarthy, K., Lush, R.J., Rye, A.D., Bessant, C. and Stone, N. (2009). Fourier transform infrared spectroscopic studies of T-cell lymphoma, B-cell lymphoid and myeloid leukaemia cell lines. Analyst, 134, 763–768.

Babrah, J., McCarthy, K., Lush, R.J., Rye, A.D., Bessant, C. and Stone, N. (2007). FT-infrared spectroscopic studies of lymphoma, lymphoid, and myeloid leukemia cell lines. Proc. SPIE, 6628, 66280R.

Posters

Vibrational spectroscopy for the diagnosis of leukaemia, SPEC 2006, Heidelberg, Germany

Cancer diagnostics using infrared spectroscopy, Cranfield Health Conference 2006.

Fourier Transform-Infrared Spectroscopic Studies of Lymphoma, Lymphoid and Myeloid Leukaemia Cell Lines, British Society of Haematology, 2007

Prizes

Poster prize for the poster titled “Cancer diagnostics using infrared spectroscopy”, Cranfield Health Conference 2006

FT-Infrared Spectroscopic Studies of Lymphoma, Lymphoid and Myeloid Leukaemia Cell Lines

Jaspreet Babrah [1,4], Keith P. McCarthy [3], Richard Lush [2], Adam D. Rye [2], Conrad Bessant [4] Nicholas Stone [1,4].

[1] Biophotonics Research Group, Gloucestershire Royal Hospital, Gloucester, UK
Tel: 01452 395708, Email: j.babrah@medical-research-centre.com
[2] Department of Haematology, Gloucestershire Royal Hospital, Gloucester, UK
[3] Department of Histopathology, Gloucestershire Royal Hospital, Gloucester, UK
[4] Cranfield Health, Cranfield University, Silsoe, UK.

ABSTRACT

This paper presents a novel method to characterise spectral differences that distinguish leukaemia and lymphoma cell lines. This is based on objective spectral measurements of major cellular biochemical constituents and multivariate spectral processing. Fourier transform infrared (FT-IR) maps of the lymphoma, lymphoid and myeloid leukaemia cell samples were obtained using a Perkin-Elmer Spotlight 300 FT-IR imaging spectrometer. Multivariate statistical techniques incorporating principal component analysis (PCA) and linear discriminant analysis (LDA) were used to construct a mathematical model. This model was validated for reproducibility. Multivariate statistical analysis of FTIR spectra collected for each cell sample permit a combination of unsupervised and supervised methods of distinguishing cell line types. This resulted in the clustering of cell line populations, indicating distinct bio-molecular differences. Major spectral differences were observed in the 4000 to 800 cm^{-1} spectral region. Bands in the averaged spectra for the cell line were assigned to the major biochemical constituents including; proteins, fatty acids, carbohydrates and nucleic acids. The combination of FT-IR spectroscopy and multivariate statistical analysis provides an important insight into the fundamental spectral differences between the cell lines, which differ according to the cellular biochemical composition. These spectral differences can serve as potential biomarkers for the differentiation of leukaemia and lymphoma cells. Consequently these differences could be used as the basis for developing a spectral method for the detection and identification of haematological malignancies.

Keywords: lymphoid leukaemia cell lines, myeloid leukaemia cell lines, lymphoma cell line, haematological malignancies, Fourier transform infrared (FT-IR) spectroscopy, multivariate statistical analysis, principal component analysis (PCA), linear discriminant analysis (LDA).

1. INTRODUCTION

Fourier transform infrared (FT-IR) spectroscopy is a vibrational technique that monitors changes in biochemical composition of cells at a molecular level. Haematological malignancies manifest at single cell level by subtle biochemical changes, which are often not obvious in histopathological studies. The non-invasive FT-IR spectroscopic method has been proposed to overcome limitations of current "gold standards" providing a diagnostic tool for haematological malignancies without the need for staining. Automated, repetitive analyses can therefore be carried out at a low cost with the added advantage of utilizing small sample volumes. In addition, FT-IR spectroscopy can distinguish differences in the characteristics of diverse molecules by probing chemical bond vibrations. These molecular and sub-molecular profiles can be used to define and differentiate cell types. The appeal of these factors has forced the exploitation of FT-IR spectrometers that combine superior spectral sensitivity with powerful multivariate statistical analysis. This permits essentially all of the relevant biochemical information buried in the spectrum to be usefully extracted.

FT-IR spectroscopy uses the mid-infrared (IR) part of the electromagnetic radiation (4000-400 cm^{-1}) to obtain an IR spectrum. It is based on the principle that when a sample is interrogated with an IR beam, the functional groups within the sample absorb the IR radiation. This induces vibrational motions (bending, stretching, etc) of covalently bonded moieties. The IR absorptions, when plotted against the wavelength of the light, produce a unique "biochemical fingerprint" of the molecular species present in the sample[1]. If this molecular fingerprint is modified by a disease process, then FT-IR spectroscopy can be used to detect and monitor the disease processes. Particular absorptions are assigned to various functional groups in an attempt to extract biochemical information. In complex biological materials such as cells, the major IR absorption bands arise from N-H, C=O, C-H and P=O bonds; found in proteins (present in cytoplasm), lipids (originate in cell membranes) and nucleic acids (found in nucleus)[2]. The sum or the specific arrangement of the bands in the IR spectra can provide qualitative and semi-quantitative information. This can be used to distinguish between normal and cancerous cells or tissues.

To date FT-IR spectroscopy has proven to be a powerful analytical tool to probe the structure and confirmation of lipids, proteins and nucleic acids[3] and also of complex biological materials such as tissues, body fluids and cell cultures[4,5]. The ability of FT-IR spectroscopy to detect changes in the chemical composition and morphology of cells led to its use in distinguishing differences among cell populations; for example, between normal and cancerous cells[6,7] as well as characterisation of cancerous state using proliferative cell lines at various stages of the cell cycle[8]. Consequently, FT-IR spectroscopic study has become an active area of research in the field of haematology, particularly in detecting and monitoring characteristic changes in molecular compositions and structures that accompany the transformation from a normal to a cancerous state. There have been various studies of FT-IR spectroscopy into haematological malignancies, including studies of differentiation between normal leucocytes and leukaemia cells based on qualitative and quantitative differences at the DNA, protein and lipid levels[9, 10]; follow-up chemotherapy assessments to predict drug-sensitive or drug-resistant[11-13], determination of apoptosis in leukaemia cells[14, 15] and differentiation of clinically aggressive (high-grade) and clinically indolent (low-grade) lymphomas based on the nucleic acid absorbance ratio[16].

Here we describe a novel method based on a combination of mid-FTIR spectroscopy of dry films with advanced multivariate statistical analysis. There are several reasons to favour this approach; only microliters of cell culture are required for measurement, the cultured cells are much more homogeneous with high sensitivity to certain key analytes. This report presents preliminary results obtained from the analysis of lymphoma, myeloid and lymphoid leukaemia cell lines based on spectral variations due to biochemical differences. The spectral data were subjected to both unsupervised and supervised methods of multivariate statistics, in particular principal component analysis fed linear discriminate analysis (PCA-fed LDA) to elucidate the spectral variations between the lymphoma, myeloid and lymphoid leukaemia cell lines. The enduring aim of this study is to demonstrate that multivariate statistical methods can differentiate between cell types, which may aid in the development of a diagnostic algorithm to analyse blood and bone marrow aspirate specimens.

2. MATERIALS AND METHODS

2.1 CELL CULTURE

Lymphoma (*Karpas*), lymphoid (*REH* and *ACV*) and myeloid (*HL60* and *Meg01*) leukaemia cell lines were obtained from the Section of Haemato-Oncology, at the Institute of Cancer Research, Sutton, UK. *REH* and *ACV* are derived from human lymphoid leukaemia, *HL60* and *Meg01* are derived from human myeloid leukaemia and Karpas 299 is derived from Hodgkin's lymphomas. The cells were grown as a suspension in 80-90% RPMI 1640 medium containing 10-20% fetal bovine serum (FBS). Cells were maintained at 37ºC in a humidified atmosphere with 5% CO_2.

2.2 SAMPLE PREPARATION

Prior to spectroscopy, cells were washed and diluted (1:5) in cold, sterile saline (0.9% NaCl) solution. The samples were pelleted by centrifugation at 3000rpm for 3min to concentrate the cells. The pellet was re-suspended in 0.9% saline solution and the washing cycle was repeated twice more. Typically 50 µL of the final diluted cell sample was transferred onto two CaF_2 slides by Cytospin™ (1500 rpm for 5 min), a method formerly developed for cytology. The aim of this technique is to produce a monolayer of cells. The slides were then allowed to air dry at room

temperature for 30 min as thin circular disk films (diameter of 2-3 mm). The cell samples were examined by FT-IR spectroscopy and a transmission spectrum was acquired for the resulting films. This approach finesses the difficulties associated with strong water absorptions by simply eliminating water from the specimen. In addition to eliminating the spectral inference of water, this approach can provide better spectral resolution by the virtue of eliminating the water/solute interactions.

2.3 SPECTRAL MEASUREMENT

The infrared spectral data presented in this study were collected using a Perkin-Elmer Spectrum One/Spotlight FT-IR Imaging 300 System based at Gloucestershire Royal Hospital. This instrument is incorporated with a computerised microscope XYZ stage, visual image capture software and a highly sensitive, liquid nitrogen-cooled 16 x 16 element HgCdTe (MCT) focal plane array detector with a step scanning interferometer. Visual images of lymphoid, myeloid and lymphoma cell lines were collected via a CCD camera under white light LED illumination in transmission mode. The desired regions of the visual image map were selected and measured using a resolution of 8 cm^{-1} with a step size of 6.25 μm and 128 scans per pixel (spp) to increase the signal-to-noise ratio. Absorption spectra were acquired in the wavenumber range of 4000 to 720 cm^{-1}. Measurements of each cell type were repeated three times with subsequent sets (1-3) prepared at different times. Background scans were obtained from a region of no sample and ratioed against the sample spectrum each time.

2.4 DATA ANALYSIS/PRE-PROCESSING

Prior to multivariate statistical analysis, the absorption spectra were converted using PS convert (Perkin Elmer) to a more compact data format. MATLAB (version 6.1, © The MathWorks Inc.1994-2006) was used to pre-process the data. This aims to enhance the spectral features that incorporate information regarding biochemical composition and effectively to suppress or eliminate unnecessary artefacts. The dataset was cropped in spectral dimension (typically between 1800 – 800 cm^{-1}). The truncated spectra were then smoothed using a 9-point Savitsky-Golay algorithm and normalized to minimise intensity variations between the spectral data set. All the spectra were then mean centred which calculates the average spectrum of the data set and subtracts that average from each spectrum. Peak positions were determined by using Spectrum/spotlight software (version 5.0.01). For each cell type, the mean spectrum was taken as an average of three different regions at various sites of the sample. Two different multivariate statistical analyses were employed in this study using MATLAB. This allows virtually all of the data present within the IR spectrum to be utilised. As the variation in spectral morphology seen between the different cell lines involves subtle differences across the range of wavenumbers, multivariate statistical analysis allows better separation of the cell line types.

Firstly, an unsupervised PCA was applied to the spectral dataset in order to compress data, without losing relevant information. It is a well-established method ideally suited to distinguish and visualise small, reoccurring spectral variations in large data sets. It was used to describe the major trends of variation in the data and to establish whether or not spectra of cells group into classes due to cell type. Therefore, it can be used as a first step for classification of the 5 cell line groups. In this study, it was employed to generate pseudocolour maps of the significant principal component scores for each cell sample. ANOVA was performed to determine which PCs best differentiate between the cell line groups. The differences in the spectra across each cell sample were used for the clustering of cell line populations. Spectra from different cell line types were assigned different colours. The colour coded scatter plots enabled the natural separation of the data to be visualized.

The second methodology was a supervised LDA, where the number of groups (cell line class identity) is dictated rather than discovered during the analysis. The scores for a specified number of principal components are used as the input data for LDA. For the purpose of this study, LDA was used to, without altering the data contained within the spectra emphasize the clustering of different cell line types following PCA. LDA for the different cell line types involves finding a linear discriminate function (LDF) that maximises the separation in discriminate space between different cell line types, whilst minimising the separation within each group. The combination of these two statistical methods referred to as PCA fed LDA was used to create algorithms for the different cell line types. The constructed algorithm was tested for its ability to correctly predict the cell line type. The results are presented as the proportion of spectra correctly predicted or classified into a group.

3. RESULTS AND DISCUSSION

In the present study, FT-IR spectroscopy of pure lymphoma (*Karpas*), lymphoid (*REH* and *ACV*) and myeloid (*HL60* and *Meg01*) cell lines was carried out. We demonstrated reproducibility of the spectroscopic and multivariate statistical methods to distinguish between cell line types. Figure 1 shows white light image and pseudocolour maps, of total intensity at each point, in transmission mode of REH cell line. The pseudocolour maps shows absorbencies of varying intensities, where blue indicates low and red indicates high absorbance values. These were produced for all five cell lines by using a motorized XYZ stage to collect spectra, in an XY co-ordination, from a selected area. This way each pixel region on the map represents the spectrum for that point. The variation of spectral differences across a sample was established by calculating principal components. Each PC is composed of PC loads, which are the spectral components and PC scores, which relates to the abundance of that spectral component in the sample. Figure 2 show the PC scores and loads of the first 10 principal components for *REH*. As the number of principal components increase more subtle variations can be seen. The score map from the second principal component was used to identify the major differences between spectra across the map. Peak assignment of the loads demonstrates further variation in the distribution of lipids, proteins and nucleic acids. Averaged spectra from the three regions of interest were plotted for comparison and identification of spectral differences as shown in figure 3. The spectrum in figure 4 is the mean representative spectrum for the *REH* cell line highlighting the prominent absorptions. Pseudocolour maps as well as their representative mean spectra were obtained for each cell type, but have not been presented in this paper due to the sheer number of data sets. Figure 5 represents the mean spectra from three sets of measurements obtained for each cell line (*REH, ACV, HL60, Meg01* and *Karpas*) as representative spectra for further spectral analysis. Figure 5 also shows the peak intensity variations between the mean IR spectra of the five cell lines groups.

The IR spectra of cell lines are complex, with hundreds of data points due to the sheer number of contributing bio-molecules and therefore contain many overlapping bands. In addition, the spectral homogeneity of spectra from cells belonging to similar lineage makes the visual inspection of individual spectra nearly impossible, and requires the use of methods of multivariate analyses for a meaningful interpretation. The subsequent results from these evaluation techniques are discussed in sections 3.2 and 3.3.

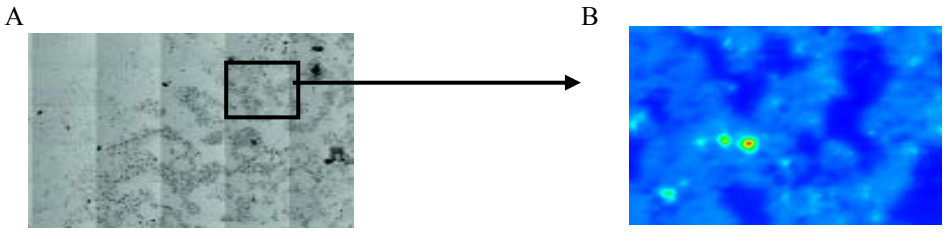


Figure 1: a) White light image of *REH* cell line and b) pseudocolour map of *REH* cell line in transmission mode. Black box highlights the region of interest.

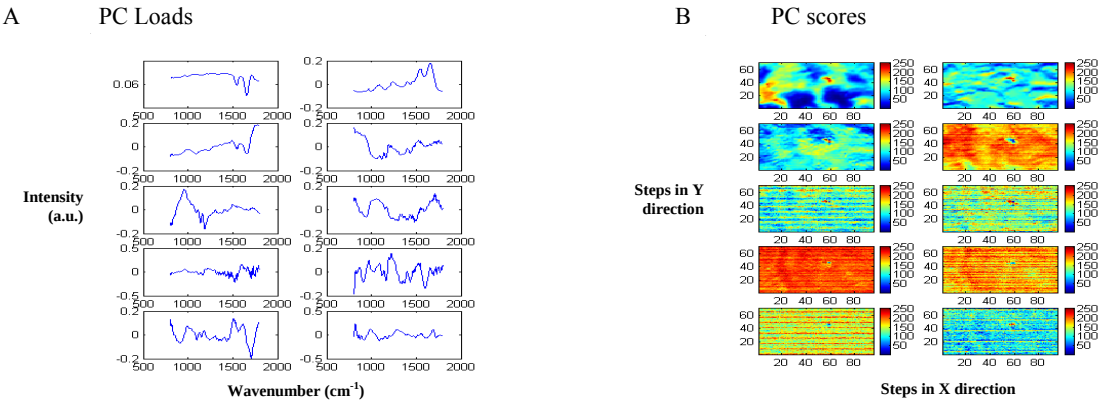


Figure 2: First 10 PC loads and corresponding pseudocolour score maps for the *REH* cell line.

A B

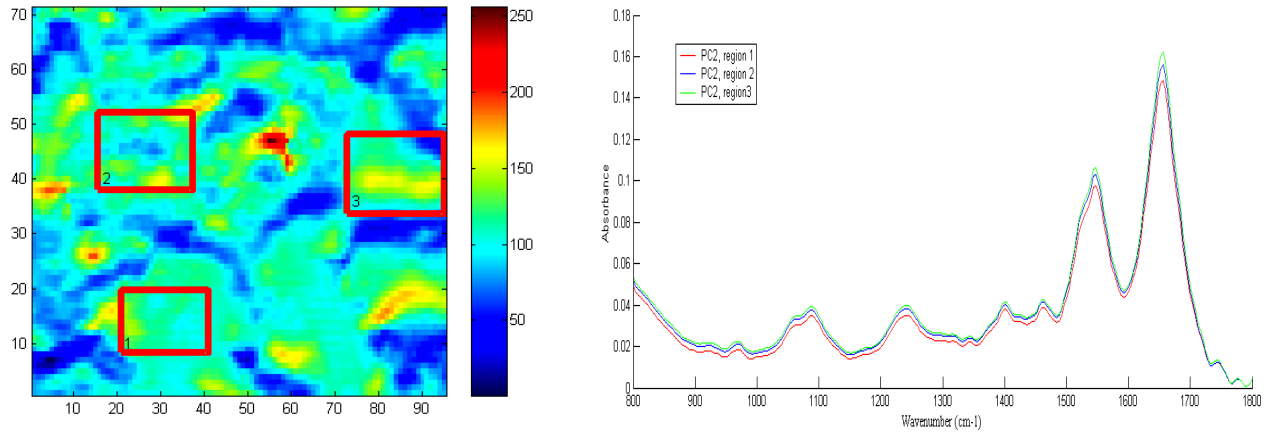


Figure 3: a) PC 2 score map with three selected regions of interest marked, b) mean IR spectra from three selected areas, region 1 (red line), region 2 (blue line) and region 3 (green line).

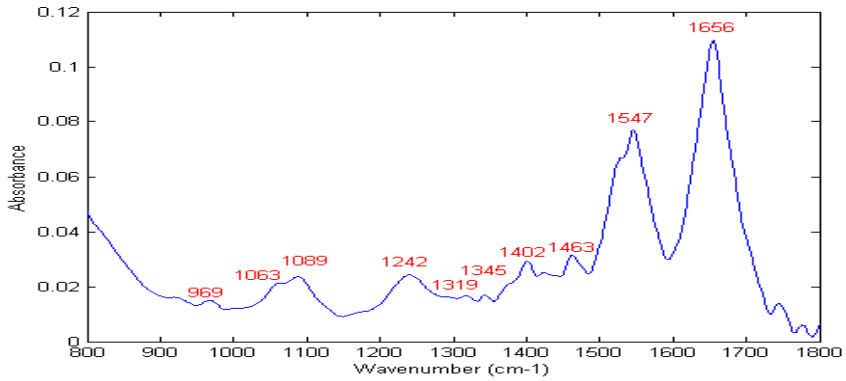


Figure 4: Representative mean spectra for REH cell line indicating major bio-molecular constituents.

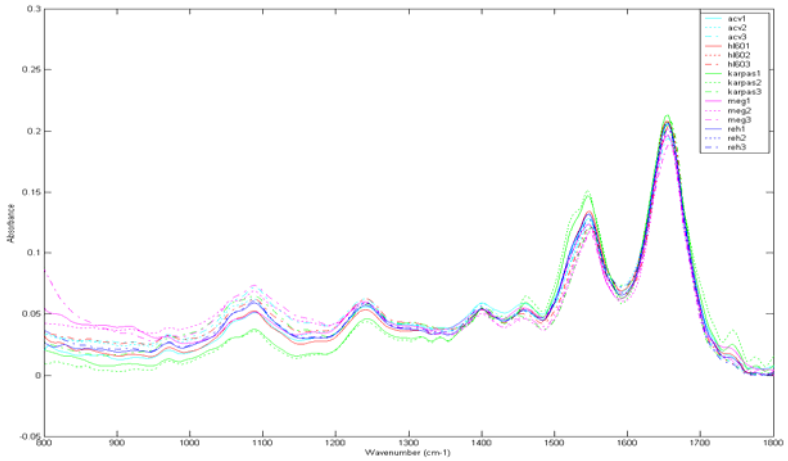


Figure 5: Mean normalised spectra of lymphoid (*REH* and *ACV*), myeloid (*HL60* and *Meg01)* and lymphoma (*Karpas*) cell lines from 3 sets of measurements. Set 1 is indicated by solid line (——), set 2 is indicated by dotted line (·····) and set 3 is indicated by dash-dot line (– · –).

3.1: PEAK INTENSITY ANALYSIS

The IR absorption spectra acquired in the region 800-1800 cm^{-1} using the methods described in section 2.3 were consistent with those described in various publications[8-10]. The IR spectra of cell lines show differences in the intensities of the peaks caused by differences in the bio-chemical constituents. This indicates considerable variation between the different cell line groups with subtle variation within the samples (figure 4 and 5). The most intensity variations are seen in the region of intensity 800-1300 cm^{-1} indicating major variations in the lipid and nucleic acid content of the cell lines. Figure 5 shows higher nucleic acid content in myeloid cell lines (e.g. *Meg01*) and lower amide I and amide II protein levels in lymphoma cell line (*karpas*). In comparison, lymphoma cell line has higher proteins levels and lower nucleic acid content.

Despite the minor spectral differences, the IR spectra of all cell lines are dominated by bands assigned to the absorption modes of proteins: the most intense is the amide I band, centred near 1656 cm^{-1}, which corresponds to the C=O stretching coupled to the N-H bending and the C-H stretching modes of peptidic bonds. It has been reported previously that the shape of the amide I band is influenced by the overall secondary structure of cellular proteins. In particular, the peaks assigned to α-helices, β-sheets, turns and coils have absorption maxima between 1645-1662 cm^{-1}, 1613-1637 cm^{-1}, 1662-1682 cm^{-1} and 1637-1645 cm^{-1} respectively. The amide II band at 1547 cm^{-1} arises from the vibrational modes that involve the C-N-H bending and C-N stretching of peptidic bonds. The amide III at 1242 cm^{-1} is contributed by proteins arising form C-N stretching and N-H bending. The peaks at 1400-1460 cm^{-1} are attributed to amino acids. The complex band pattern between 900-1300 cm^{-1} region represents the superimposed contributions of nucleic acids (DNA, RNA), carbohydrates and phosphates. In particular, the band centred at 1089 cm^{-1} is assigned to C-O bending modes of saccharides (glucose and fatty acids) and the peak at 969 cm^{-1} form the C-C stretching mode of (deoxy) ribose that is present in the DNA and RNA.

3.2: PRINCIPAL COMPONENT ANALYSIS

Multivariate analysis by PCA was initially used alone to classify the cell types for visual comparison. PCA allows the degree to which the algorithm groups spectra from the same cell line type, and separates the spectra from different cell line types to be represented visually using scatter plots. Scores of principal components (PC2 vs. PC4) were used for the plots with the position of each of the spectra marked on the plot. The separation of lymphoma and lymphoid and myeloid leukaemia cell lines produced is shown as two-dimensional plot in figure 6.

As a result of strong similarities in their spectral signatures, the lymphoid leukaemia cell types (*REH* and *ACV*) formed one cluster with overlapping spectra and the myeloid leukaemia cell types (*HL60* and *Meg01*) produced the second cluster again with overlapping spectra. The third cluster belongs to the lymphoma (*Karpas*) cell line. PC scatter plots reveal good homogeneity of the spectra belonging to a particular lineage. This indicates prominent spectral variation between the different cell lines with very little intra-variability within the samples.

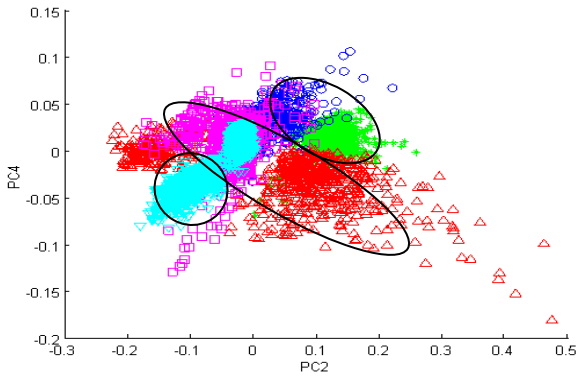


Figure 6: a) scores scatter plot (PC2 vs. PC4) represents clustering of lymphoid, myeloid and lymphoma cells lines. Each symbol represents one spectrum. Each symbol represents one spectrum. Lymphoid leukaemia cell lines (*ACV*= red upward triangle and *REH*= magenta squares), myeloid leukaemia cell lines (*HL60*= blue squares and *Meg01*= green stars) and lymphoma cell line (*Karpas*= cyan downward triangle).

3.3: LINEAR DISCRIMINANT ANALYSIS

The five cell line group classification model generated using PC fed LDA demonstrated good separation of the spectra in linear discriminant space. LDA was performed to maximise the distance between the groups, whilst minimising the distance between members within each group. This is illustrated as the two-dimensional plot of the scores of the LD1 against the scores of LD4 in figure 7. The training performance of the model was 95%. It also generated good sensitivity ranging from 94-99% and a specificity of 98-100% across the 5 cell line group classification model.

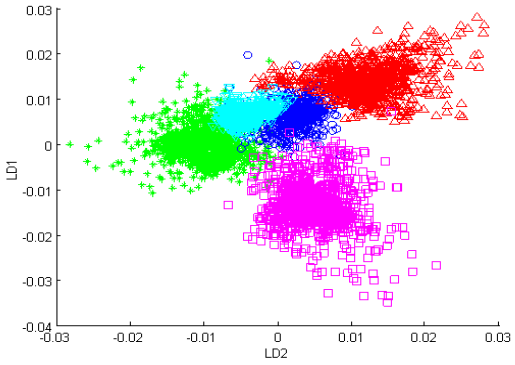


Figure 7: Scatter plot of the 5 cell line groups as generated by PCA fed LDA. Each symbol represents one spectrum. Each symbol represents one spectrum. Lymphoid leukaemia cell lines (*ACV*= red upward triangle and *REH*= magenta squares), myeloid leukaemia cell lines (*HL60*= blue squares and *Meg01*= green stars) and lymphoma cell line (*Karpas*= cyan downward triangle).

		FT-IR Analysis					
		ACV	**Hl60**	**Karpas**	**Meg01**	**REH**	**Total**
Class	**ACV**	1365	0	0	3	62	**1430**
	Hl60	59	1322	1	12	14	**1408**
	Karpas	0	0	1293	6	10	**1309**
	Meg01	0	0	56	2000	41	**2097**
	REH	118	0	0	1	1876	**1995**
	Total	**1542**	**1322**	**1350**	**2022**	**2003**	**8239**

Table 1: FT-IR five cell line group prediction model.

Class	**ACV**	**Hl60**	**Karpas**	**Meg01**	**REH**
FT-IR Sensitivity (%)	96	94	99	96	94
FT-IR Specificity (%)	98	100	99	99.7	98

Table 2: Sensitivities and specificities achieved by the five group classification model as predicted by PCA fed LDA.

4. CONCLUSION

This study shows that small spectral differences exist between the infrared spectra of lymphoma, lymphoid and myeloid leukaemia cells. Irrespective of subtle spectral differences between cell types, it is possible to classify cells of different lineages via its spectral signature and unsupervised and supervised methods of multivariate statistics in particular PCA fed LDA. Spectral differences between the cell lines differ according to the cellular biochemistry, which can serve as potential biomarkers lymphoma, lymphoid and myeloid leukaemia. These initial results further indicate that this approach could be useful for early diagnosis of haematological malignancies. It also provides the added benefit that instrumentation does not require reagents to promote a visible change in colour. It also ensures that all IR compositional data is stored in a single data set, permitting the rapid extraction of information of components of interest. Further work will be carried out using additional cell lines to assess and compare sensitivity, specificity and reproducibility of FT-IR spectroscopy as a complementary diagnostic tool in the field of clinical haematology. With some improvement in the technique and the analytical approach, a very objective discrimination method can be developed. This could be promising for investigations based on fine needle aspirations and blood specimens.

REFERENCES

1. Ellis D.I. and Goodacre R. "Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy". The Analyst. **Vol. 131**, 875-885, 2006.
2. Mourant J.R., Gibson R.R., Johnson T.M., Carpenter S., Short K.W., Yamada Y.R and Freyer J.P. "Methods of measuring the infrared spectra of biological cells". Physics in Medicine and Biology. **Vol. 48**, 243-257, 2003.
3. Mantsch H.H. and McElhaney R.N. "Application of infrared spectroscopy to biology and medicine". Journal of Molecular Structures. **Vol. 217**, 347-362, 1990.
4. Erukhimovitch V., Talyshinsky M., Souprun Y., and Huleihel M. "Spectroscopic characterization of human and mouse primary cells, cell lines and malignant cells". Photochemistry and Photobiology. **Vol. 74 (4)**, 446-451, 2002.
5. Delris G. and Petibois C. "Application of FT-IR spectroscopy to plasma contents analysis and monitoring". Vibrational Spectroscopy. **Vol. 32**, 129-136, 2003.
6. Cohenford M.A. and Rigas B. "Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: implication of tumour biology". Proc. Natl. Acad. Sci. USA. **Vol. 95**, 15327-15332, 1998.
7. Diem M., Boydston-White S. and Chiriboga L. "Infrared spectroscopy of cells and tissues: shinning light onto a novel subject". Applied Spectroscopy. **Vol. 53**, 148-161, 1999.
8. Boydston-White S., Gopen T., Houser S., Bargonetti J. and Diem M. "Infrared spectroscopy of human tissue. Infrared spectroscopic studies of myeloid leukaemia (ML-1) cells at different phases of the cell cycle". Biospectroscopy. **Vol. 5**, 219-227, 1999.
9. Benedetti E., Papineschi F., Vergammi P., Consolini R. and Spremolla G. "Analytical infrared spectral differences between human normal and leukaemic cells (CLL)". Leukaemia Research. **Vol. 8**, 483, 1984.
10. Schultz C.P., Lui K.Z., Johnston J.B., and Mantsch H.H. "Study of Chronic Lymphocytic Leukaemia Cells by FT-IR Spectroscopy and Cluster Analysis". Leukaemia Research. **Vol. 20 (8)**, 649-655, 1996.
11. Lui K.Z, Schultz C.P, Johnston J.B., Mantsch H.H. "Comparison of Infrared spectra of CLL cells with their ex-vivo sensitivity (MTT assay) to chlorambucil and cladribine". Leukaemia Research. **Vol. 21**, 1125-1133, 1997.
12. Schultz C.P., Lui K.Z., Johnston J.B. and Mantsch H.H. "Prognosis of Chronic Lymphocytic Leukaemia from Infrared Spectra of Lymphocytes". Journal of Molecular Structures. **Vol. 408/409**, 253-256, 1997.
13. Ramesh J., Kapelushnik J., Mordehai J., Mose A., Huleihel M., Erukhimovitch V., Levi C and Mordechai S. "Novel methodology for the follow-up of acute lymphoblastic leukaemia using FTIR microspectroscopy". Journal of Biochemical and Biophysical Methods. Vol. 51 (3), 251-261, (2002).
14. Lui K.Z, Li. J, Kelsey S.M and Mantsch H.H. "Quantitative determination of apoptosis on leukaemia cells by infrared spectroscopy". Apoptosis. **Vol. 6**, 267-276, 2001.
15. Gasparri. F and Muzio. M. "Monitoring of apoptosis of HL60 cells by Fourier-transform infrared spectroscopy". Biochemical Journal. **Vol. 369**, 239-248, 2003.
16. Andrus G.L. and Strickland R.D. "Cancer Grading by Fourier Transform Infrared Spectroscopy". Biospectroscopy. **Vol. 4**, 37-46, 1998.

ACKNOWLEDGEMENT

The financial support for this research project by Leukaemia and Intensive Chemotherapy Charity (LINC) is gratefully acknowledged. Further support has been received from Biophotonics and Haematology teams at Gloucestershire NHS Trust. Author wishes to thank the hard work from Professor Mel Greaves and Sue Colman, of Institute of Cancer Research, Sutton for culturing the cell lines and Carol Baxter of Cytology Department, Gloucestershire Royal Hospital for the preparation of the slides.

Fourier transform infrared spectroscopic studies of T-cell lymphoma, B-cell lymphoid and myeloid leukaemia cell lines

Jaspreet Babrah,*[a] Keith McCarthy,[c] Richard John Lush,[b] Adam David Rye,[b] Conrad Bessant[d] and Nicholas Stone*[a]

Received 9th May 2008, Accepted 8th December 2008
First published as an Advance Article on the web 19th December 2008
DOI: 10.1039/b807967f

This paper presents Fourier transform infrared (FT-IR) spectroscopy to characterise spectral differences that distinguish cells derived from human T-cell lymphoma, B-cell lymphoid, and myeloid leukaemia cell lines. This methodology is based on spectral measurements of major cellular biochemical constituents and multivariate spectral processing. Major spectral differences were observed in the 1800–900 cm^{-1} 'fingerprint' spectral region. Bands in the averaged spectra for each cell line were assigned to major biochemical constituents including: proteins, lipids, carbohydrates and nucleic acids. Multivariate statistical analysis of the spectra was carried out to develop a classification model to discriminate the five cell types. The results show that FT-IR spectroscopy displays high sensitivity and specificity when discriminating between T-cell lymphoma, B-cell lymphoid, and myeloid leukaemia cells based on intrinsic biomolecular signatures. FT-IR spectroscopy in combination with multivariate statistical analysis provides an important insight into T-cell lymphoma, B-cell lymphoid, and myeloid leukaemia cell line identification. In conclusion, this paper demonstrates a potential for this technique to be used in developing a clinical tool for the detection and identification of haematological malignancies.

1. Introduction

Leukaemia and lymphoma are haematological malignancies that result from the clonal expansion and accumulation of lymphoid or myeloid progenitors in the bone marrow.[1] Normally, 25% of the white blood count is made up of lymphocytes including both T- and B-cells, which play a major role in humoral and cellular immunity respectively.[2] At the time of diagnosis, the normal marrow cell population has usually been replaced by the leukaemia clone accompanied by haematogenous dissemination to a variety of extramedullary sites. Current immunological, cytopathological and molecular biology procedures identify variations in cellular morphology or granularity, staining patterns or use antibodies to detect abnormal leucocytes and cell-specific marker proteins.[3–5] Although these methods are well established, they are invasive, requiring extensive sample preparation and staining techniques as well as highly trained personnel.

The ultimate aim of this research is to provide practical, effective methods that will compete successfully on all levels with the present gold standards in diagnosis, analysis, and imaging of haematological malignancies. This will potentially provide new or complementary information to clinicians that is otherwise not available. Fourier transform infrared (FT-IR) spectroscopy, a non-invasive method, has been proposed to overcome some of the limitations of current 'gold standard' methods and therefore potentially provides a diagnostic tool for haematological malignancies without the need for staining. Automated, repetitive analyses can therefore be carried out at a low cost with the added advantage of utilizing a small volume of sample. FT-IR spectroscopy is a vibrational spectroscopic technique that monitors subtle changes in the biochemical composition of cells at a molecular level which are often not obvious in cytopathological or molecular studies. It uses the mid-infrared (mid-IR) part of the electromagnetic radiation frequencies (4000–400 cm^{-1}) to obtain an IR spectrum, which reflects a unique set of molecular vibrations. These infrared-active vibrations can then be correlated directly to the biochemical species and the resultant IR spectrum can then be described as a 'fingerprint' of the molecular species. If this molecular fingerprint is modified by a disease process then IR spectroscopy can be used to detect and monitor disease processes. FT-IR spectroscopy holds great promise as an analytical tool to probe chemical bond vibrations as well as compositional and structural changes at the molecular and sub-molecular levels with minimal sample preparation.

To date, FT-IR spectroscopy has proven to be a powerful analytical tool to probe the structure and confirmation of lipids, proteins and nucleic acids[6] and also of complex biological materials such as body fluids and cell cultures.[7,8] The ability of FT-IR spectroscopy to detect changes in the chemical composition and morphology of cells led to its use in distinguishing differences among cell populations; for example, between normal

[a]*Biophotonics Research Group, Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucester, UK GL1 3NN. E-mail: j.babrah@medical-research-centre.com; n.stone@medical-research-centre.com; Fax: +44 (0)8454 225485; Tel: +44 (0)8454 225486*
[b]*Department of Haematology, Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucester, UK GL1 3NN*
[c]*Department of Histopathology, Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucester, UK GL1 3NN*
[d]*Cranfield Health, Cranfield University, CranfieldBedfordshire, UK MK43 0AL*

and cancerous cells[9–11] as well as biochemical and morphological characterisation of myeloid leukaemia (ML-1), cervical cancer (HeLa) and murine embryonic stem (ES) cells at various differential phases of their life cycle and stages of cell proliferation.[12–15] A recent study pioneered by the Krafft group demonstrated that FT-IR microscopic imaging can identify the differentiation state of individual human mesenchymal stem cells.[16] A classification model based on multivariate statistical analysis was used to distinguish cell types by their IR spectroscopic fingerprint. Extensive FT-IR spectroscopic studies have also been performed on identification of various colonies of micro-organisms[17] as well as differentiation of bacterial species; for example, *Candida albicans*,[18] *Listeria* species[19] and *Escherichia coli* strains.[20] FT-IR spectroscopy has further been employed to investigate multiple aspects of leukocyte physiology, for example, to characterise differentiation and apoptotic processes,[21,22] to predict drug sensitivity and resistance in leukaemic patients undergoing chemotherapy and to monitor the response of leucocytes to chemotherapy.[23–26]

This paper investigates whether FT-IR spectroscopy can be applied as an alternative analytical technique in the classification of haematological malignancies using cell lines derived from leukaemia and lymphomas. This method combines superior spectroscopic sensitivity with powerful multivariate statistical analysis to extract the relevant biochemical information buried in the spectrum. These statistical techniques aid in the identification of characteristic features such as peak intensity changes and shifts, within the IR spectrum, and the construction of a classification model for distinguishing leukaemia and lymphoma cell lines. This may aid in the development of a diagnostic algorithm to ultimately analyse blood and bone marrow aspirate specimens in future spectroscopic studies.

2. Materials and methods

2.1 Cell culture

The following cells were obtained from the Section of Haemato-Oncology, at the Institute of Cancer Research, Sutton: *Karpas 299* (Human T-cell lymphoma cell line) cells, *REH* (Human acute lymphoid cell line) cells, RCH-*ACV* (Human acute lymphoid cell line) cells, *HL60* (Human acute myeloid leukaemia cell line) cells and *Meg01* (Human chronic myeloid cell line) cells. A total of 5 sets of the same cell types were supplied as separate batches. The cells were grown as a suspension in 80–90% RPMI 1640 medium containing 10–20% fetal bovine serum (FBS). Cells were maintained at 37 °C in a humidified atmosphere with 5% CO_2.

2.2 Sample preparation

Prior to spectroscopic measurements the cells were washed and diluted (1 : 5) in cold, sterile saline (0.9% NaCl) solution. The samples were pelleted by centrifugation at 3000 rpm for 3 min to concentrate the cells. The pellet was re-suspended in 0.9% phosphate buffer saline solution and the washing cycle was repeated twice more. Typically 200 μL of the final diluted cell samples was transferred onto CaF_2 slides by Shandon Cytospin® Cytocentrifuge (1500 rpm for 5 min), a method formerly developed for cytology as described below. The slides were then allowed to air-dry at room temperature under airflow for 30 min as thin circular disk films (diameter of 5 mm) and frozen at −80 °C. This approach overcomes the difficulties associated with strong water absorptions by simply eliminating water from the specimen. The slides were then examined by FT-IR spectroscopy and a number of transmission spectra were acquired for the resulting films.

2.3 Spectral measurements

The infrared spectral data presented in this study were collected using a Perkin-Elmer Spectrum One/Spotlight FT-IR Imaging 300 System based at Biophotonic Research Group, Gloucestershire Royal Hospital. This instrument is incorporated with a computerised microscope *x*-*y*-*z* stage, visual image capture software and a highly sensitive, liquid nitrogen-cooled 1 × 16 element HgCdTe (MCT) array detector with a step scanning interferometer. Visual images of lymphoid, myeloid and lymphoma cell lines were collected *via* a CCD camera under white light LED illumination in transmission mode. From the visual image, area of high cellular content was selected for spectral imaging. The average size of the area probed was approximately 200–400 μm for each cell line sample. Spectral images were measured by scanning the computer controlled microscope stage in an *x*-*y* direction in increments corresponding to the 6.25 μm step size, using a spectral resolution of 8 cm^{-1} and 128 scans per pixel (spp) to increase the signal-to-noise ratio. Transmittance spectra were acquired in the wavenumber range of 4000–720 cm^{-1}. Background scans were obtained from a region of clean calcium fluoride (*i.e.* area with no sample) to allow contributions to the spectrum from the instrument to be removed when ratioed against the tissue spectra during data processing. Measurements of each cell type were repeated five times with subsequent batch sets (1–5) prepared at different times to test for reproducibility.

2.4 Data analysis

The raw FSM (binary) file was imported into Matlab© using a combination of in-house functions and a Perkin-Elmer© data plugin script which converted and truncated (1800–900 cm^{-1}) the raw binary data into an ASCII format which can then be processed by Matlab©. This loading process also converted the data from transmittance into absorbance values. All subsequent multivariate data analysis was performed in Matlab© to perform data reduction and discriminant analysis as well as to carry out leave-one-out cross-validation using in-house designed scripts.

An in-house absorbance spectral filter tool was applied to the loaded hyperspectral data cube to filter out weak or low absorbance. All spectra outside the absorbance range of 0.2 (upper absorbance limit) and 0.15 (lower absorbance limits) were removed. This absorbance range was chosen to optimise the stability of the classification model. The absorbance filter tool in combination with other pre-processing techniques such as normalisation was employed to enhance the spectral features that incorporate information regarding biochemical composition and effectively suppresses or eliminates unnecessary artefacts, noise, calcium fluoride, water and CO_2 signals as well as variations in sample thickness. This absorbance spectral filter tool was applied to all raw FSM for all 5 cell samples from 5 subsequent batches.

The absorbance-filtered spectra for all 5 cell samples from each of the 5 batches were combined to create a new filtered dataset with most weak absorbance and saturated spectra removed. The dataset was then mean-centred prior multivariate classification to ensure that each spectrum could be statistically compared to the mean of the dataset. This was achieved by calculating the mean of the dataset and subtracting that mean from each spectrum.

The mean-centred data were inputted into a Principal Component Analysis (PCA)-fed Linear Discriminant Analysis (LDA) classification algorithm.[27,28] PCA was performed on all spectra to extract and visualise persistent features of spectra from individual cell types and to compare different cell types against each other. LDA was performed to assess the performance of the model in distinguishing the different cell line types. This method uses the scores from PCA as inputs to develop a pathology classification model. The first 25 principal component scores (describing the greatest variance in the dataset) were used as the input data for LDA. The constructed classification method was tested for its ability to correctly predict the cell line type. In this study we used leave-one-out cross-validation (LOOCV) to test the accuracy of the predictions of the PCA-fed LDA classification model. Classification accuracy was assessed whereby the class of spectra within a batch is predicted by holding out all of the spectra measured from a particular batch (the test set) and building an LDA model with the remaining data (the training set). This method is repeated so that each batch in the dataset is left out in turn and used once as the test data. Sensitivities and specificities were calculated as a measure of model accuracy and performance to evaluate the efficacy of classification systems. In our analysis, sensitivity is defined as the ratio of the number of true positives to the sum of true positives and false negatives and the specificity is defined as the ratio of the number of true negatives to the sum of true negatives and false positives – see eqn (1) and (2).

$$\text{Sensitivity} = \frac{\text{No. of true positives}}{\text{No. of true positives} + \text{No. of false negatives}} \times 100\% \quad (1)$$

$$\text{Specificity} = \frac{\text{No. of true negatives}}{\text{No. of true negatives} + \text{No. of false positives}} \times 100\% \quad (2)$$

3. Results

In the present study, FT-IR spectroscopic measurements of pure lymphoma (*Karpas 299*), lymphoid (*REH* and *ACV*) and myeloid (*HL60* and *Meg01*) cell lines were carried out. We demonstrated the reproducibility of the spectroscopic and multivariate statistical methods to distinguish between cell line types. Fig. 1a and 1b show a white light image and the corresponding average absorbance image for the *Meg01* cell line. The average absorbance image shows absorbencies of varying intensities, where blue indicates low and red indicates high absorbance values. These were produced for all five cell lines by using a motorized *x*-*y*-*z* stage to collect spectra, in an *x*-*y* co-ordination, from a selected area. This way each pixel region on the map represents the spectrum for that point. Fig. 1c represents the average absorbance image after the absorbance spectral filter tool was applied to filter out weak and saturated spectra. The spectral

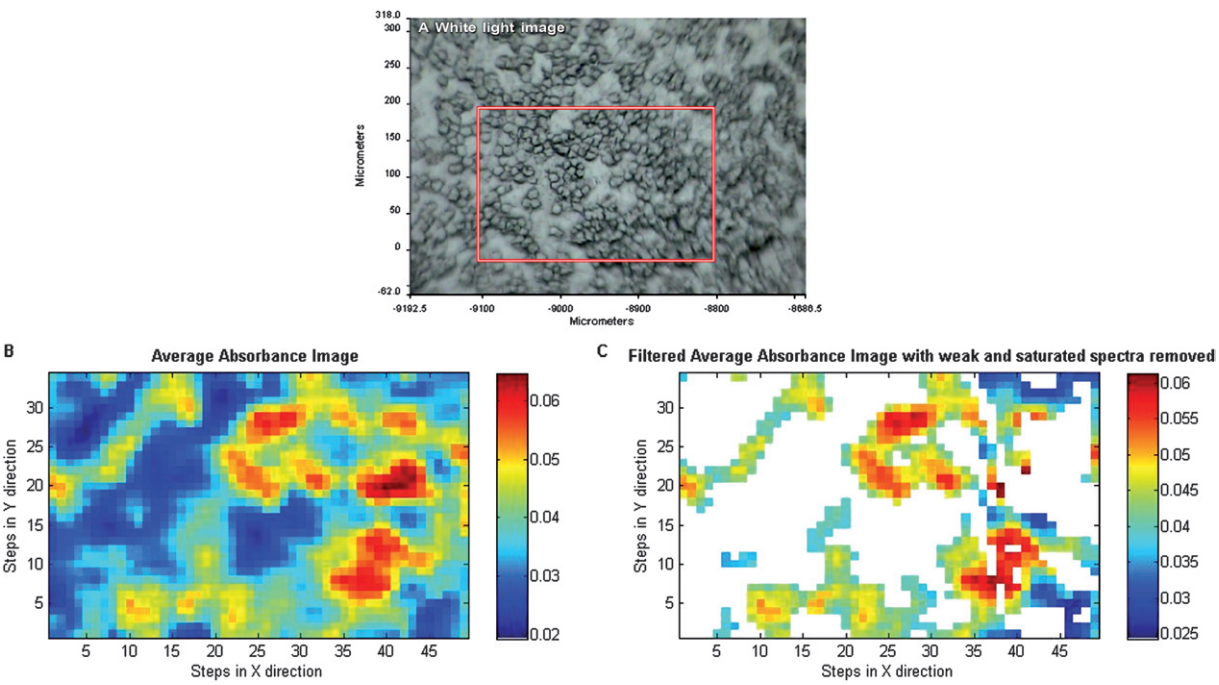


Fig. 1 (A) White light image of *Meg01* cell sample. The red box (200 × 300 μm) highlights the region of interest. (B) Corresponding average absorbance image of *Meg01*: the blue areas indicate low and red areas indicate high absorbance values. (C) Filtered average absorbance image: the white areas represent weak or saturated spectra that have been filtered out.

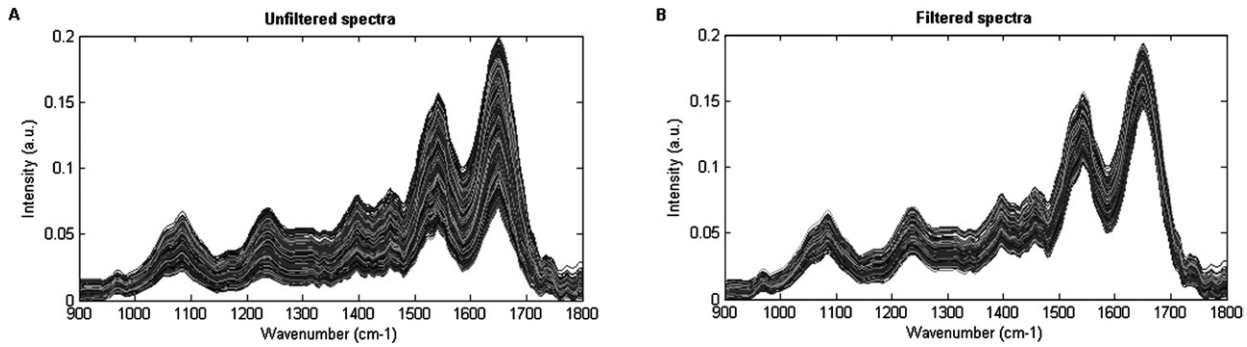


Fig. 2 (A) Unfiltered normalised spectral dataset and (B) remaining good quality normalised filtered spectra.

dataset shown in Fig. 2a and 2b represents the original unfiltered dataset and the remaining good quality filtered spectra. Average absorbance images as well as their representative mean spectra were obtained for each cell type, but have not been presented in this paper due to the sheer number of datasets.

3.1 Peak intensity analysis

In order to identify significant spectral signatures of the various cell types, a standard univariate method of assigning key peak positions was employed. This monitors vibrations due to pre-assigned specific functional groups and explores changes in these functional groups as potential biomarkers. The representative mean spectra from five sets of measurements for each cell line (*REH*, *ACV*, *HL60*, *Meg01* and *Karpas 299*) are depicted in Fig. 3. The mean spectra of the five cell types do have a degree of similarity. However, there are also many spectral features differentiating the different cell types due to slight variations in peak positions and intensities. The intensities of the IR spectra provide quantitative information while the absorption positions (wavenumber (cm^{-1})) reveal qualitative characteristics about the nature of the chemical bonds and their structure.

Peak shapes and positions are repeatable although the peak intensities tend to vary due to differences in the biochemical constituents. Most significant changes in peak intensities occur in the fingerprint region between 900 and 1500 cm^{-1} corresponding to the nucleic acids, amino acids, amide III, lipids and carbohydrates. There are also variations in peaks corresponding to amino acids (1401 cm^{-1}), lipids (1452 cm^{-1}) as well as the amide I (1655 cm^{-1}) and amide II (1549 cm^{-1}) peaks. Despite the minor spectral differences, all five cell types were found to display prominent characteristic spectral features. The most distinctive peaks in the IR absorption spectra are assigned to the absorption modes of proteins: the most intense is the amide I band, centred near 1655 cm^{-1}, which corresponds to the C=O stretching coupled to the N–H bending and the C–H stretching modes of peptide bonds.[19,20] It has been reported previously that the shape of the amide I band is influenced by the overall secondary structure of cellular proteins.[21] In particular, the peaks assigned to α-helices, β-sheets, β-turns and random coils have absorption maxima in the ranges 1645–1662 cm^{-1}, 1613–1637 cm^{-1}, 1662–1682 cm^{-1} and 1637–1645 cm^{-1} respectively. The amide II band at 1549 cm^{-1} arises from the vibrational modes that involve the C–N–H bending and C–N stretching of peptidic bonds.[22] The

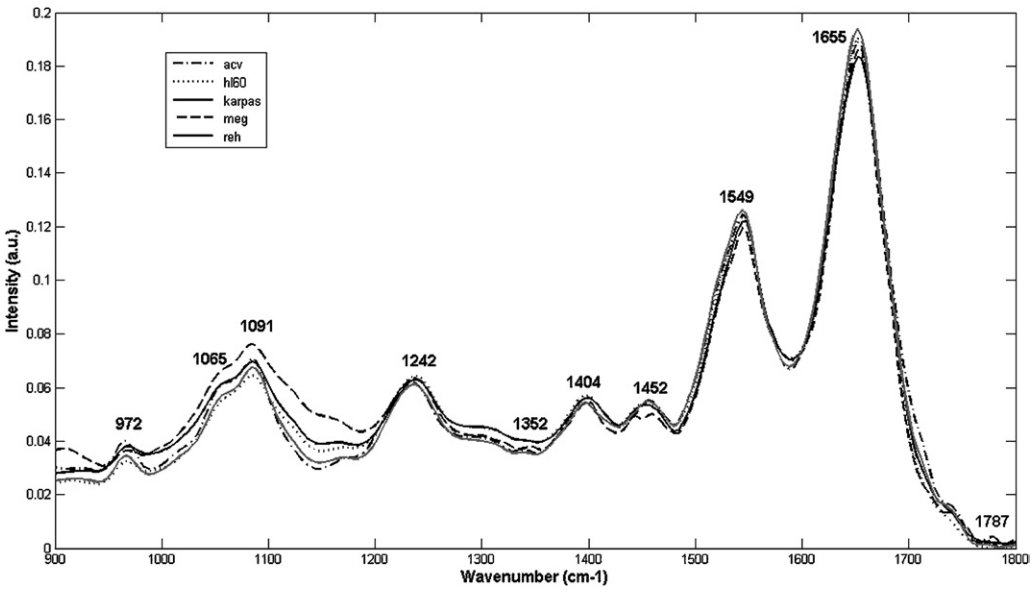


Fig. 3 Representative average IR spectra from five sets of measurements obtained for each cell line (*REH*, RCH-*ACV*, *HL60*, *Meg01* and *Karpas 299*) in the range 900–1800 cm^{-1}.

amide III at 1242 cm^{-1} is contributed by proteins arising from C–N stretching and N–H bending.[22,23] The peaks at 1400–1460 cm^{-1} are attributed to CH_2 and CH_3 deformation vibrations primarily due to lipid contribution. Nucleic acid signals contribute to the infrared spectra at 1091 cm^{-1} (sym. PO_2^-), 1065 cm^{-1} (PO_2^- and ribose groups) and 972 cm^{-1} (phosphodiester group present in the DNA and RNA).[20] Lipids have long hydrocarbon chains and therefore their have peaks at 1352 cm^{-1} (CH_3 vibration) and 1740 cm^{-1} (C=O bending) although the peak at 1787 cm^{-1} emerges somewhat as a shoulder (Fig. 3). Phosphate bands of phospholipids are also found between 1080 and 1230 cm^{-1}. The position and amplitude of these absorption bands (Fig. 3) are governed by the nature and concentration of the constituents, including the molecular conformation of the constituents and interactions among them. Overall, we show that the relative intensities of DNA and amide III modes increase in malignant cells, indicating increased cellular nuclear content. Although these findings are consistent with those described in various publications,[13,16,17] these cannot solely be used as a diagnostic algorithm.

Visual inspections reveal that the IR absorption spectra of cell lines are complex, with numerous data points due to the sheer number of contributing bio-molecules and therefore contain many overlapping bands. In addition, the spectral homogeneity of spectra from cells belonging to similar lineage makes the visual interpretation of individual spectra nearly impossible. It is therefore essential that these vibrational spectroscopic measurements are combined with univariate methods (peak assignment) as well as multivariate statistical analyses for a meaningful interpretation.

3.2 Multivariate analysis

The ability of FT-IR spectroscopy to distinguish between the cell types was assessed using multivariate analysis based on spectral differences within the dataset. PCA was used in this study as a technique that reduces the number of parameters needed to represent the variance in the spectral dataset. The principal components were then used to generate a linear discriminant model. PCA is a chemometric technique that can resolve a complete spectral dataset into a few key spectral components and can thus identify and isolate important trends within the dataset.[27] LDA applies a linear discriminant function that maximises the variance in the data between groups and minimises the variance between members of the same group.[28] All 5 cell types were discriminated as shown in Fig. 4. The classification model was tested using leave-one-out cross-validation (LOOCV). Classification results in Table 1 show that most of the spectra within the 5 cell line group model have been classified correctly with only a small number of spectra misclassified. The reason for this misclassification may be associated with small inconsistencies within the spectra belonging to leukaemia and lymphoma cell line populations. Misclassifications are also due to the relatedness of the cell type. Based on the cross-validation results, sensitivity and specificity values were calculated (Table 1). The model obtained sensitivity ranging between 74 and 100% and specificity ranging between 94 and 100%. Similar methods have been used previously in investigation of synovial fluids from joints affected by rheumatoid arthritis and osteoarthritis[29] and classification of gastric inflammation and malignancy on endoscopic biopsies.[30] Ideally, the use of an unknown test set would provide a more robust independent analysis of sensitivity and specificity of the technique and this is planned for future work.

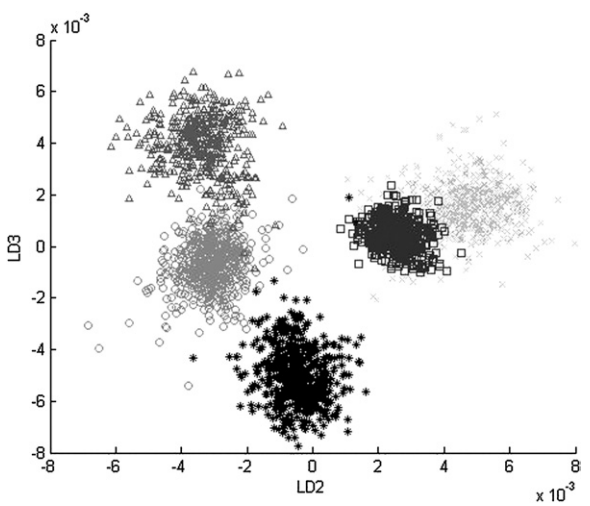


Fig. 4 Scatter plot (LD2 *vs.* LD3) of the 5 cell line groups as generated by PCA-fed LDA. Each symbol represents one spectrum. Lymphoid leukaemia cell lines (*ACV* = crosses and *REH* = stars), myeloid leukaemia cell lines (*HL60* = circles and *Meg01* = squares) and lymphoma cell line (*Karpas 299* = upward triangles).

Table 1 Summary of cross-validation results and sensitivity and specificity values

	Diagnosis (cell type)				
FT-IR prediction	*ACV*	*HL60*	*Karpas 299*	*Meg01*	*REH*
ACV	545	0	23	0	36
HL60	0	658	1	0	7
Karpas 299	0	74	543	0	8
Meg01	0	0	0	646	0
REH	3	83	81	0	475
Number correct	**545**	**658**	**543**	**646**	**475**
Sensitivity (%)	90.2	98.8	86.9	100	73.9
Specificity (%)	99.9	93.8	95.9	100	97.9

4. Conclusions

This study shows that, on first examination, the IR spectra are visually similar but detailed spectral analysis reveals noticeable variations that could give an insight into the biochemical differences that exist between the IR spectra of T-cell lymphoma, B-cell lymphoid and myeloid leukaemia cell lines. Irrespective of subtle spectral differences between cell types, it is possible to classify cells of different lineages *via* its spectral signature. Although the subtle spectral differences between leukaemia and lymphoma cell lines show potential for discrimination, the obtained spectra are complex and their exact molecular origins are only partially understood. We believe that further biochemical studies need to be applied to assess the contribution of complex biochemical constituents at the cellular level.

The present study is the first to demonstrate that methods of multivariate statistics, in particular PCA-fed LDA, have the

ability to group mid-FT-IR spectra by leukaemia and lymphoma cell type. Cross-validation reveals that a combination of FT-IR spectroscopy with multivariate feature extraction and classification methods has the potential to become an ideal complementary tool to conventional assays for early detection of leukaemia and lymphoma. These initial results further indicate that this approach provides the added benefit that instrumentation does not require contrast reagents to promote a visual change in colour and ensures that all compositional data are stored in a single dataset. With some improvements in the technique and the analytical approach, an objective discrimination method can be developed which combines pattern recognition and classification models. This could be promising for future investigations based on fine needle aspirations and blood specimens for diagnosing haematological malignancies.

Acknowledgements

The financial support for this research project by Leukaemia and Intensive Chemotherapy Charity (LINC) is gratefully acknowledged. The authors of this study would like to thank Professor Mel Greaves and Sue Colman based at the Institute of Cancer Research, Sutton for culturing the cell lines. We would also like to thank the hard work from Karol Baxter and Rajvinder Dhillion at the Cytology Department, Gloucestershire Royal Hospital for the preparation of the slides. Further support has been received from Haematology team and Dr Nicholas Stone, who holds a NIHR (Career Scientist) Senior Research Fellowship and leads the Biophotonics Research Group at Gloucestershire Royal Hospital. Martin Isabelle also deserves a special gratitude for all his contributions in developing in-house analysis programmes.

References

1 J. L. Binet, A. Auquier, G. Dighiero, C. Chastang, H. Piguet, J. Goasguen, G. Vaugier, G. Potron, P. Colona, F. Oberling, M. Thomas, G. Tchernia, C. Ja quillat, P. Boivin, C. Lesty, M. T. Duault, M. Monconduit, S. Belabbes and F. Germy, *Leukaemia*, 1981, **48**, 198–206.
2 A. B. Cantor and S. H. Orkin, *Current Opinion in Genetics and Development*, 2001, **11**, 513–519.
3 R. M. Braziel, M. A. Shipp, A. L. Feldman, V. Espina, M. Winters, E. S. Jaffe, E. F. Petricoin and L. A. Liotta, *Haematology*, 2003, 279–293.
4 G. L. Andrus and D. R. Strickland, *Biospectroscopy*, 1998, **4**, 37–46.
5 K. P. McCarthy, *Cancer and Metastasis Reviews*, 1997, **16**, 109–127.
6 H. H. Mantsch and R. N. McElhaney, *Journal of Molecular Structure*, 1990, **217**, 347–362.
7 V. Erukhimovitch, M. Talyshinsky, Y. Souprun and M. Huleihel, *Photochemistry and Photobiology*, 2002, **74**, 446–451.
8 G. Delris and C. Petibois, *Vibrational Spectroscopy*, 2003, **32**, 129–136.
9 M. A. Cohenford and B. Rigas, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 15327–15332.
10 M. Diem, S. Boydston-White and L. Chiriboga, *Applied Spectroscopy*, 1999, **53**, 148–161.
11 C. P. Schultz, K. Z. Lui, J. B. Johnston and H. H. Mantsch, *Leukaemia Research*, 1996, **20**, 649–655.
12 M. Romeo, B. Mohlenhoff, M. Jennings and M. Diem, *Biochimica et Biophysica Acta*, 2006, **1758**, 915–922.
13 S. Boydston-White, T. Gopen, S. Houser, J. Bargonetti and M. Diem, *Biospectroscopy*, 1999, **5**, 219–227.
14 S. Boydston-White, T. Chernenko, A. Regina, M. Miljkovic, C. Matthaeus and M. Diem, *Vibrational Spectroscopy*, 2005, **38**, 169–177.
15 S. Boydston-White, M. Romeo, T. Chernenko, A. Regina, M. Miljkovic and M. Diem, *Biochimica et Biophysica Acta*, 2006, **1758**, 908–914.
16 C. Krafft, R. Salzer, S. Seitz, C. Ern and M. Schieker, *Analyst*, 2007, **132**, 647–653.
17 M. Beekes, P. Lasch and D. Naumann, *Veterinary Microbiology*, 2007, **123**, 305–319.
18 C. Sandt, G. D. Sockalingham, D. Aubert, H. Lepan, C. Lepouse, M. Jaussaud, A. Leon, J. M. Pinon, M. Manfait and D. Toubas, *Journal of Clinical Microbiology*, 2003, **41**, 954–959.
19 A. Cecilia, J. Dietrich, M. Wenning and S. Scherer, *Analytical and Bioanalytical Chemistry*, 2008, **390**, 1639–1635.
20 D. Ami, L. Bonecchi, S. Cali, G. Orsini, G. Tonon and S. M. Doglia, *Biochimica et Biophysica Acta.*, 2003, **1624**, 6–10.
21 K. Z. Liu, L. Jia, S. M. Kelsey, A. C. Newland and H. H. Mantsch, *Apoptosis*, 2001, **6**, 267–276.
22 F. Gasparri and M. Muzio, *Biochemical Journal*, 2003, **369**, 239–248.
23 K. Z. Liu, C. P. Schultz, J. B. Johnston and H. H. Mantsch, *Leukaemia Research*, 1997, **21**, 1125–1133.
24 C. P. Schultz, K. Z. Lui, J. B. Johnston and H. H. Mantsch, *Journal of Molecular Structures*, 1997, **408/409**, 253–256.
25 J. Ramesh, J. Kapelushnik, J. Mordehai, A. Mose, M. Huleihel, V. Erukhimovitch, C. Levi and S. Mordechai, *Journal of Biochemical and Biophysical Methods*, 2002, **51**, 251–261.
26 C. P. Schultz, K. Z. Lui, J. B. Johnston and H. H. Mantsch, *Leukaemia Research*, 1996, **20**, 649–655.
27 P. Lasch and D. Naumann, *Cell. Mol. Biol.*, 1998, **44**, 189–202.
28 C. Krafft, A. B. Sobottka, K. D. Geiger, G. Schackert and R. Salzer, *Analytical and Bioanalytical Chemistry*, 2006, **387**, 1669–1677.
29 R. A. Shaw, S. Kotowich, H. H. Eysel, M. Jackson, G. T. D. Thomson and H. H. Mantsch, *Rheumatology International*, 1995, **15**, 159–165.
30 Q. B. Li, X. J. Sun, Y. Z. Xu, L. M. Yang, Y. F. Zhang, S. F. Weng, J. S. Shi and J. G. Wu, *Clinical Chemistry*, 2005, **51**, 346–350.

