# Investigation of the Utility of Optical Spectroscopy for Management of Chronic Respiratory Disease through Investigation of Sputum Mucin Glycosylation Patterns

**Brilliant, Charles D.**

# Investigation of the Utility of Optical Spectroscopy for Management of Chronic Respiratory Disease through Investigation of Sputum Mucin Glycosylation Patterns

Prifysgol Abertawe
Swansea University

**Charles David Brilliant BSc (Hons)**

**Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy**
**Swansea University**
**2018**

## Abstract

Chronic respiratory diseases such as chronic obstructive pulmonary disease (COPD), lung cancer and cystic fibrosis (CF) are leading causes of disease and death in the UK. These diseases are difficult to monitor and diagnose in an efficient and timely manner. In this work, Fourier-transform infrared spectroscopy (FTIR) was used to investigate sputum samples from lung cancer, COPD and CF patients for specific IR-spectral markers which could be utilised for the diagnosis and management of these conditions. It is proposed that spectral changes correspond to structural changes to sputum mucins, which could be biomarkers for COPD progression and exacerbation, for lung cancer diagnosis, and for CF patient monitoring.

In this study, sputum samples from COPD patients were obtained for FTIR analysis to generate a spectral library for use in creation of a generalised predictive model for COPD exacerbation. It was found that COPD exacerbations are a highly complex and heterogeneous condition, which made the generation of a generalised predictive model for exacerbation problematic. One model developed in this work demonstrated a capability to determine exacerbation from COPD baseline samples with 80% sensitivity and 48% specificity. Small correlations were found between peak positions and absorbance intensities around wavenumbers associated with mucin glycoprotein structural change, and physiological factors, such as smoking status or lung function.

FTIR spectroscopy was shown to have a very high power for distinguishing lung cancer sputum samples from non-cancer respiratory disease sputum samples, using a simple protocol with no sample pre-processing and linear regression modelling. A series of diagnostic algorithms were developed and were shown to have a greater than 90% sensitivity and specificity for detecting lung cancer from raw sputum.

FTIR was also successfully utilised in the monitoring of CF patient sputum samples for the presence of a novel inhaled therapeutic, OligoG, during and after treatment. It was found that FTIR spectroscopy can readily detect very low concentrations of OligoG in sputum with no sample processing or targeting of the therapeutic necessary.

In conclusion, FTIR was shown to be a powerful tool for analysis of raw sputum, capable of providing high-quality molecular structural information pertaining to the mucin glycoproteins, and proposed changes to these structures. Using this information, it is possible to distinguish lung cancer sputum from other respiratory disease sputum, and monitor the levels of a novel therapeutic in CF patient lungs. Associations were made between FTIR spectral features and physiological factors of COPD patients, however further work is needed to fully evaluate if COPD exacerbation can be predicted using this method.

## Declaration

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ……C.D.Brilliant…… (candidate)

Date ……07/05/19……

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ……C.D.Brilliant…… (candidate)

Date …… 07/05/19……

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access approved by the Swansea University.**

Signed …… C.D.Brilliant…… (candidate)

Date …… 07/05/19……

# Table of Contents

## Publications

1. Lydia C. Powell, Manon F. Pritchard, Elaine L. Ferguson, Kate A. Powell, Shree U. Patel, Phil D. Rye, Stavroula-Melina Sakellakou, Niklaas J. Buurma, **Charles D. Brilliant**, Jack M. Copping, Georgina E. Menzies, Paul D. Lewis, Katja E. Hill & David W. Thomas (2018) Targeted Disruption of the Extracellular Polymeric Network of *Pseudomonas aeruginosa* Biofilms by Alginate Oligosaccharides *NPJ Biofilms and Microbiomes* 4(13) https://doi.org/10.1038/s41522-018-0056-3

2. Ali Albalawi, **Charles D. Brilliant**, Alessandro Chiasera, Hrvoje Gebavi, Rolindes Balda, Maurizio Ferrari, Wilfried Blanc, Wedad Albalawi, Hazel Hung, Alexander Quandt, Anna Lukowiak, Stefano Taccheo (2018) Analytical modelling of Tm-doped tellurite glass including cross-relaxation process *Optical Materials* https://doi.org/10.1016/j.optmat.2018.06.037

## Acknowledgements

First I would like to thank my supervisor, Professor Paul Lewis, whose support and enduring patience across the last 4 years has been invaluable to the completion of this thesis. The freedom to explore my own ideas and "chase the rabbit" has been one of the most enjoyable aspects of this project.

I must also thank all past and present members of the Respiratory Diagnostics Group, and the Air Pollution Research Group in Swansea University. In particular, I would like to extend thanks to Drs Georgina Menzies and Aaran Lewis, whose guidance throughout the early-stages of this project was crucial.

Next, I thank all of the collaborators with whom I have worked throughout this thesis, who not only have provided excellent research ideas but also many, many, many sputum samples to "play" with. In particular, the team at Cwm Taf UHB worked tirelessly to help ensure SPEDIC was a success.

My family have been a constant source of encouragement and gave me the strong determination to finish. I especially must thank my wonderful Mum, whose expert proof-reading skills are second-to-none.

And finally, I must thank my loving partner Jenny, whose unrelenting love and unwavering support has been a guiding light all the way from the very start, to the very end of this project.

This work is dedicated to my Grampa and constant source of inspiration,
Reginald Arthur Gapper,
who sadly passed away during the completion of this thesis and is sorely missed.

## List of Figures

## List of Tables

## Abbreviations

4MGS – 4 Meter Gait Speed

AFM – Atomic Force Microscopy

AlO – Aluminium Oxide

ASL – Airway Surface

ATR – Attenuated Total Reflectance

ATR-FTIR – Attenuated Total Reflectance-Fourier-Transform Infrared Spectroscopy

ATR-VFIR – Attenuated Total Reflectance-Variable Filter Infrared Spectroscopy

AU – Absorbance Units

BMI – Body Mass Index

BSM – Bovine Submaxillary Mucin

Ca- Cancer

CF – Cystic Fibrosis

CFG – Consortium for Functional Glycomics

COPD – Chronic Obstructive Pulmonary Disease

CT – Computed Tomography

bCOPD – bronchoscopy Chronic Obstructive Pulmonary Disease

rCOPD – random Chronic Obstructive Pulmonary Disease

CYP – Cytochrome P450

DNA – Deoxy-ribonucleic Acid

DTGS – Deuterated Triglycine Sulphate

E-Cigarette – Electronic Cigarette

EGFR – Epidermal Growth Factor Receptor

EPHX – Enzyme Microsomal Epoxide Hydroxylase

ER- Oestrogen Receptor

ERK – Extracellular Signal-Regulated Kinase

$FEV_1$ – Forced Expiratory Volume in 1 second

FTIR – Fourier-transform Infrared Spectroscopy

FVC – Forced Vital Capacity

FWHM – Full width at Half Maximum

GalNAc – N-Acetyl Galactosamine

GlcNAc – N-Acetyl Glucosamine

GP – General Practitioner

GOLD - Global Initiative for Chronic Obstructive Lung Disease

HT-FTIR – High-throughput Fourier-transform Infrared Spectroscopy

IFN-γ – Type II Interferon gamma

IL- Interleukin

KBr – Potassium Bromide

LC – Lung Cancer

LDCT – Low-Dose Computed Tomography

Le – Lewis Antigen

LPS – Lipopolysaccharide

LTA – Lipoteichoic Acid

MCC – Mucociliary Clearance

MMP12 – Matrix Metalloproteinase-12

MPU – Metallised-Polyurethane/Foam Laminate

mRNA – Messenger Ribonucleic Acid

NC-Non-Cancer

NFAT - Nuclear Factor of Activated T-cell

NLST – National Lung Screening Trial

NK Cell – Natural Killer Cell

NSCLC – Non-Small Cell Lung Carcinoma

OligoG - OligoG

OPUS – Optics User Software

PAH – Polycyclic Aromatic Hydrocarbons

PC – Principle Component

PCA – Principle Components Analysis

PE - Polyethylene

PRISMA-7 – Program of Research to Integrate Services for the Maintenance of Autonomy

PU – Polyurethane

QOL – Quality-of-Life

Q-Value – Quotient Value

SCC – Squamous Cell Carcinoma

SCLC – Small Cell Lung Carcinoma

SED – Squared Euclidean Distance

SHP – Spectral Histopathology

SNP – Single Nucleotide Polymorphism

T-FTIR – Transmission-Fourier-transform Infrared Spectroscopy

TLR – Toll-like Receptor

TNFα – Tumour Necrosis Factor alpha

TR – Tandem Repeat

TUGT – Timed Up and Go Test

VFIR – Variable Filter Infrared Spectroscopy

WHO – World Health Organisation

ZnS – Zinc Sulphide

ZnSe – Zinc Selenide

# Chapter 1
# General Introduction: Chronic Diseases of the Respiratory System and Vibrational Spectroscopic Techniques for Biofluid Analysis

## 1.0 Chapter Aims

This chapter seeks to provide a comprehensive review of the current literature focussing on chronic diseases of the respiratory system, and vibrational spectroscopy. A brief overview of the human respiratory system will be given, before providing reviews of three major respiratory diseases which pose significant threats to public health; chronic obstructive pulmonary disease, lung cancer, and cystic fibrosis.

The second section of the literature review will focus on mucus and mucin biochemistry, providing a comprehensive examination of the role of mucus and mucins in the human respiratory system. It will also focus on how changes to mucus composition can contribute to chronic respiratory disease.

The final portion of the literature review will focus on vibrational spectroscopy, detailing the theoretical and mechanistic principles of various vibrational spectroscopic techniques, and how these techniques can be employed in patient biofluid analysis for disease diagnosis and/or management.

This chapter will also set-out the research questions and aims & objectives of this thesis.

## 1.1 The Respiratory System and Chronic Diseases of the Respiratory System

### 1.1.1 A Brief Overview of the Physiology of the Human Respiratory System

In order to fully comprehend how physiological changes are brought about in the development of respiratory disease and how these changes can affect the patient, it is important to have an understanding of the healthy human respiratory system. The physiology of the human respiratory system will be briefly described.

The human respiratory system is a highly-specialized set of organs and structures in which the primary function of gas exchange between the alveoli and the blood is carried out. To effectively and efficiently accomplish this function, the lung requires a very large internal surface area through which the gases will diffuse. The internal surface area of the lungs was estimated to range between 30-50m$^2$ (Hasleton, 1972), but a more recent study focussing on the deposition of bioaerosols within the lung used an estimated average internal surface area of adult lung to be 77.9m$^2$ (Guha, Hariharan, & Myers, 2014). The airways are made up of many different tissues including alveoli, where gas exchange occurs, and the bronchi and bronchioles,

through which the inhaled and exhaled air passes. The bronchi and bronchioles divide and subdivide many times, creating a vastly complex and intricate branching structure which allows the exchange of approximately 6 litres of gas per minute (Rackley & Stripp, 2012). The epithelial lining of the upper and large respiratory tracts is pseudostratified, with the lower and small being columnar and cuboidal (Crystal, Randell, Engelhardt, Voynow, & Sunday, 2008). The airway epithelium is highly specialised with many different types of cells including basal cells, goblet cells and ciliated cells performing multiple functions to ensure homeostasis in the lung. The composition of the airway epithelium includes a high number of mucus-producing cells; submucosal glandular cells and, lining the epithelium there are surface mucus/goblet cells (Vareille, Kieninger, Edwards, & Regamey, 2011). These cells express and secrete mucin proteins, which form a mucus gel lining the airways. This thin protective layer forms the airway surface liquid (ASL) layer, the main function of which is to maintain correct hydration of the airway and to form a physical and chemical barrier in which foreign particles and pathogens are trapped and subsequently removed through the action of the ciliated cells by mucociliary clearance (MCC) (Knowles & Boucher, 2002).

Toll-like receptors (TLRs) are fully expressed within the airway epithelium and have a direct role in regulating the immune response through immune cell recruitment and cytokine signalling (Parker & Prince, 2011). For example, lipopolysaccharide (LPS), a well-characterised molecule found in bacterial cell walls capable of inducing inflammation, is known to directly stimulate TLR-4 leading to cytokine and interferon production (Lu, Yeh, & Ohashi, 2008).

Immune cells including eosinophils, neutrophils, macrophages, dendritic cells, T-cells and natural killer (NK) cells are also present in the epithelial lining and are a vital part of the innate and adaptive immune response in the lungs (Figure 1-1). Together they form a highly important system, vital for the protection of the airways against bacterial and viral infection (Vareille *et al.*, 2011).

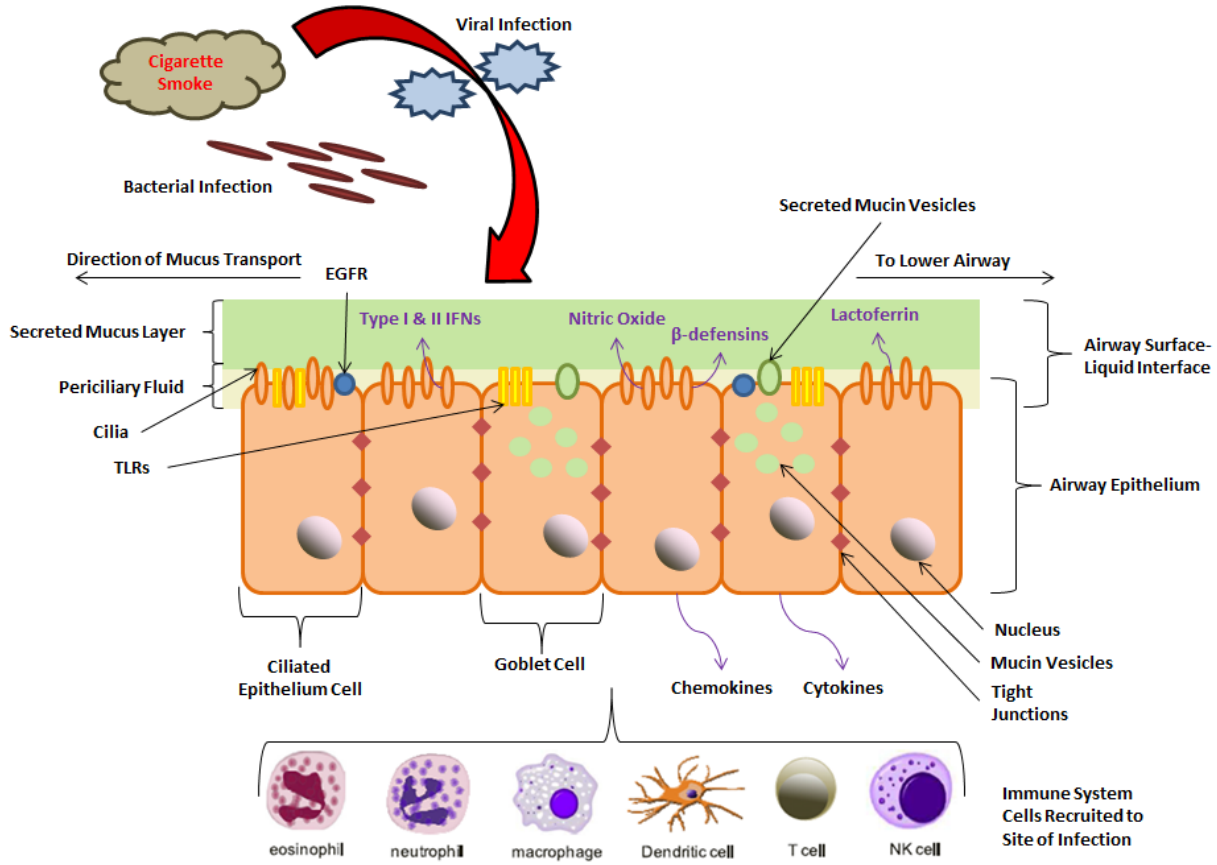


**Figure 1-1: The airway epithelium acts as a physical barrier against pathogens and foreign particles, such as cigarette smoke. Mucus is secreted from mucus-secreting goblet cells into the ASL and traps any pathogens or particles which enter the airway; ciliary beating action pushes the mucus out of the airway. Infection from bacteria or virus, or assault by foreign particles induces an initial innate immune response, with Type I and II interferons (IFNs), nitric oxide, β-defensins and lactoferrin being released into the ASL. TLR-binding is important for activation of signal-transduction pathways which lead to chemokine and cytokine release. Chemokines and cytokines are released by the epithelial cells, recruiting eosinophils, neutrophils, macrophages, dendritic cells, T-cell and NK-cells to the site of infection. (Figure and information adapted from Parker & Prince, 2011; Qiu *et al.*, 2015; Thorley *et al.*, 2011; Vareille *et al.*, 2011). ASL – Airway surface liquid; TLRs – Toll-like Receptors; NK-cells – Natural Killer cells; IFNs – Interferons; EGFR – epithelial growth factor receptor**

All of these highly-specialized cells work together to ensure the airways are kept in good health and free from infection. However when one or more parts of the respiratory system stop working correctly, it can contribute to the pathogenesis of a multitude of chronic respiratory disorders, including asthma, cystic fibrosis, bronchiectasis, emphysema, COPD and lung cancer. Diseases such as these represent a major burden on healthcare systems across the world; according to the World Health Organisation (WHO), chronic respiratory disease is responsible for 3.8 million deaths per year, equivalent to approximately 9% of all deaths worldwide (World Health Organization, 2018). Therefore, it is important that efforts are focussed on

reducing this mortality burden, through enhanced diagnostic technologies to diagnose diseases more quickly and efficiently and novel therapeutics to reduce symptom burden on the patient and improve their quality of life.

### 1.1.2 Chronic Obstructive Pulmonary Disease

#### 1.1.20 Introduction to Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease (COPD) is a serious, debilitating and progressive disease of the respiratory tract, characterized by irreversible chronic air-flow reduction which is associated with an abnormal inflammatory response of the lungs to toxins and pollutants. Airflow is limited by chronic obstruction of the airways through increased mucus production and reduced mucus clearance via MCC activity. The principle cause of COPD is smoking, but exposure to environmental and occupational pollutants and chemicals can have a significant impact on disease progression (Blanc *et al.*, 2009; Health and Safety Executive, 2014). COPD patients are also at an increased risk of developing lung cancer, and the two conditions are thought to be linked by inflammatory pathways (Houghton, Mouded, & Shapiro, 2008).

COPD is a term used to describe progressive and irreversible respiratory diseases involving chronic decline in lung function. This term mainly incorporates the diseases chronic bronchitis (CB) and emphysema. CB occurs due to hypersecretion of mucus and narrowing of the bronchi, caused by inappropriate inflammation responses, and emphysema is a permanent destructive enlargement of the airspaces within the lung without accompanying fibrosis of the lung tissue (Health and Safety Executive, 2014).

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) in their latest report (Global Initiative for Chronic Obstructive Lung Disease, 2018) define COPD as follows;

> *"Chronic Obstructive Pulmonary Disease (COPD) is a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases"*

There are approximately 3 million deaths due to COPD per year, and it is projected to rise to become the 3rd leading cause of death by 2030, due to a predicted increase in tobacco usage (Global Burden of Disease Collaborators, 2015; World Health Organization, 2008). Within England and Wales, it is estimated that there are approximately 900,000 COPD patients, although this number could be much higher due to under-diagnosis with the true number being between 1.5 million and 3.8 million (Health and Safety Executive, 2014). COPD is consistently responsible for between 25,000 and 30,000 deaths per annum for the last 25 years in the UK (Health and Safety Executive, 2014). A recent study estimated how the prevalence of COPD and the associated costs to healthcare providers will grow until 2030; the authors calculated that in England alone, numbers of COPD patients could grow from 950,000 in 2011, to 1.3 million in 2030: an increase of 38%. They also estimated costs to rise from £1.5 billion, to over £2.3 billion by 2030 (McLean *et al.*, 2016).

| Stage | $FEV_1/FVC$ and $FEV_1$ Scores | Characteristics |
|---|---|---|
| **0: At risk** | Normal | Chronic cough and sputum production, but lung function is normal. |
| [redacted] | | |
| | $FEV_1 < 30\%$ | failure.<br>Chronic cough and sputum production also present |

**Table 1-1: Classification stages of COPD, with increasing severity of symptoms and decrease in lung function between each stage; FEV: Forced expiratory volume in 1 second; FVC: Forced vital capacity; COPD: Chronic obstructive pulmonary disease. Information adapted from Rabe *et al.* (2007)**

A diagnosis of COPD is considered in any patient who displays the symptoms of cough, sputum production, dyspnoea, and/or a history of exposure to risk factors for the disease, such as smoking or heavy exposure to a variety of dusts and chemicals found in an occupational capacity. To confirm a diagnosis of COPD, lung function is measured using the 'Forced Expiratory Volume in one second' ($FEV_1$) and the 'Forced

Vital Capacity' (FVC). The diagnosis of partially irreversible air-flow limitation is confirmed by the presence of a post-bronchodilator $FEV_1$ of < 80% of the predicted value, in combination with a $FEV_1/FVC$ < 70% (Rabe *et al.*, 2007). Example spirometry traces from GOLD for normal and obstructed airways are shown Figure 1-2 (Global Initiative for Chronic Obstructive Lung Disease, 2018). As COPD progresses, it is classified into separate stages, based upon lung function and symptoms. Each sequential stage describes an increase in severity of the disease as illustrated in Table 1-1 (information derived from Rabe *et al.*, (2007)).

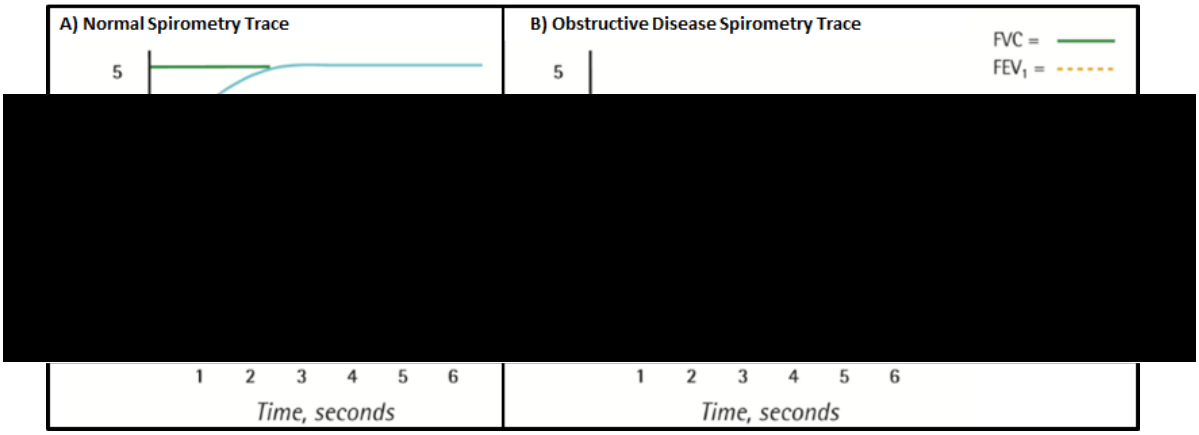


**Figure 1-2: Example spirometry traces for A) normal airways with no obstructive disease, B) obstructed airways. FEV1 is indicated by the orange-dashed line and FVC by the green line, drawn at the tangent of the plateau of the trace. FEV1 – Forced expiratory volume in 1 second; FVC – Forced vital capacity. Figure adapted from Global Initiative for Chronic Obstructive Lung Disease, (2018)**

COPD being a complex disease, with many different causes and stages, is associated with different comorbidities each influencing the disease outcome for the patient. A recent review aimed to categorise COPD phenotypes according to these various factors (Mirza & Benzo, 2017). The authors identified 7 distinct phenotypes of COPD – these are summarised in Table 1-2. All of these patients will have a confirmed diagnosis of COPD but will require different levels of care. Because of this, patients will utilise their local healthcare resources in diverse ways. By assigning a patient to a sub-category of COPD, the most appropriate care can be more readily and efficiently applied. It is important to recognise that diversity among patients exists and that it can have a significant impact of the quality and efficacy of care. However, it is also important to recognise that because of this diversity, these 7 phenotypes may not be complete, and that a single patient may be able to be classified into two or more phenotypes. It could be possible to further classify into subgroups of multiple phenotypes, for example Frequent-Exacerbator-&-Rapid-Decliner-Phenotype or

Comorbid-&-Physical-Frailty-Phenotype, each with their own specific treatment regime. Therefore, care must be taken when analysing patient groups based solely on this limited classification.

| Phenotype Classifier | Clinical Importance | Phenotypic Markers | Phenotype-Specific Management |
|---|---|---|---|
| **Asthma-COPD Overlap Phenotype** | Increased frequency and severity of exacerbations.<br>Faster $FEV_1$ decline.<br>Increased healthcare utilization.<br>Increased comorbidity burden and mortality. | Responds well to bronchodilators.<br>Sputum eosinophilia.<br>History of asthma.<br>High total IgE.<br>History of atopy. | Inhaled corticosteroids.<br>Bronchodilators.<br>Address atopy. |
| **Frequent Exacerbator Phenotype** | Faster decline in lung function.<br>Increased healthcare utilization.<br>Increased mortality.<br>Increased risk of depression. | 2 or more exacerbations per year. | Bronchodilators.<br>Inhaled corticosteroids.<br>Anti-inflammatory treatment. |
| **Upper Lobe-Predominant Emphysema Phenotype** | Potentially substantial symptomatic benefit with surgical lung volume reduction. | CT scans show predominant upper lobe emphysema. | Consider surgical lung volume reduction. |
| **Rapid Decliner Phenotype** | High mortality. | Rapid decline of lung function.<br>Usually younger.<br>Poor nutritional status.<br>No major comorbidities. | Early subspecialty and lung transplant evaluation. |
| **Comorbid phenotype** | Increased healthcare utilization.<br>Poorer QOL.<br>Increased mortality risk. | High comorbidity burden.<br>Ischaemic heart disease.<br>Congestive heart failure, diabetes, and high BMI are most common. | Aggressive management of comorbid disease alongside optimal respiratory therapy. |
| **Physical-Frailty Phenotype** | High symptom burden.<br>Frequent exacerbations.<br>Poor functional capacity.<br>Poor self-management of disease. | Screening:<br>4MGS (<0.8 m/s)<br>TUGT (>10 seconds)<br>PRISMA-7 questionnaire (score ≥ 3)<br>Diagnosis: Fried criteria – meets ≥ 3 of 5 criteria | Pulmonary rehabilitation.<br>Screening of COPD patients for frailty recommended. |
| **Emotional-Frailty Phenotype** | Increased healthcare utilization.<br>Poorer QOL.<br>Poor disease self-management. | High depression and anxiety scores.<br>Fear of breathlessness. | Health coaching.<br>Cognitive therapy.<br>Pharmacological management. |

**Table 1-2: COPD phenotypes and phenotype-specific identifiers and management techniques, as described in Mirza & Benzo (2017): 4MGS: 4 meter gait speed; PRISMA: Program of Research to Integrate Services for the Maintenance of Autonomy; TUGT: Timed Up and Go test; QOL: Quality of Life; COPD: Chronic Obstructive Pulmonary Disease; FEV1: Forced expiratory volume in 1 second**

### 1.1.21 Genetic and Environmental Risk Factors

COPD results from the interplay between genetic susceptibility and exposure to environmental stimuli. Smoking cigarettes has been shown to be the leading cause (Forey, Thornton, & Lee, 2011), but non-smokers may be at risk from environmental and occupational pollutants (Pauwels *et al.*, 2001). A person's early upbringing can

have a significant influence on their chance of developing COPD. Svanes and colleagues (2010) described factors including maternal, paternal and childhood asthma, maternal smoking, and childhood respiratory infections as "childhood disadvantage factors". These factors are associated with a substantial increase in COPD risk, permanently lowering lung function shown by a significantly associated reduced $FEV_1$ value which doesn't improve with age. The impact of one or more of these childhood disadvantage factors was shown to be as large as heavy smoking (Svanes *et al.*, 2010). The risk of developing COPD in non-smokers is increased if the patient is exposed to outdoor air pollution, second-hand smoke, occupational dust and fumes, and biomass smoke. The risk is also increased if the patient has previously been infected with tuberculosis (Eisner *et al.*, 2010). There is also a link between economic privation and development of COPD. Smokers who are economically deprived, or come from areas of high deprivation, are more likely to develop COPD than their non-smoking counterparts, but are also at a significantly increased risk of lung disease compared to smokers who are less deprived (Sherratt, Field, & Marcus, 2017). Recent evidence also suggests that electronic cigarette (e-cigarette) use could be a similar influencing factor on COPD development as conventional tobacco cigarette smoking. Alterations to the ratio of mucin glycoproteins in the lung, alongside significant increases in neutrophil granulocyte and neutrophil elastase levels have been observed in e-cigarette users, which are comparable to the changes observed in tobacco cigarette smokers (Reidel *et al.*, 2018). However, this study did not make a distinction between e-cigarette users who had never smoked tobacco and those who had previously smoked cigarettes. There is evidence that suggests long-term smoking cessation leads to the sputum proteome of healthy former smokers becoming more similar to that of those who have never smoked (Titz *et al.*, 2015). However, it is difficult to ascertain if the changes observed in e-cigarette users were due to current e-cigarette smoking status, or if changes caused by previous tobacco smoking were still evident in these participants as there was also no distinction of the e-cigarette users by the length of time they had ceased smoking (Reidel *et al.*, 2018). Despite this, the study still demonstrates that e-cigarettes may be harmful and could contribute to the progression of lung diseases, including COPD.

There is also evidence to suggest that COPD has genetic associations. Currently, the only well-established genetic cause of COPD is α-1 antitrypsin deficiency. An antiprotease, α-1 antitrypsin is a regulator of neutrophil elastase in the

lung (Haq *et al.*, 2016). Through decreased regulation of neutrophil elastase, a deficiency in α-1 antitrypsin can lead to damage to lung tissue (Kawabata, Hagio, & Matsuoka, 2002). However this cause of COPD is rare, being present in 1-2% of all COPD patients (Gooptu, Ekeowa, & Lomas, 2009). Although relatively rare, this does provide evidence for the concept that an imbalance in proteolytic enzymes is involved in the pathogenesis of emphysema and COPD (Hogg, 2004). Other possible genetic causes of COPD have been suggested through work performed in genome-wide association studies. Regions on chromosome 4 near HHIP and in FAM13A, and regions on chromosome 15 in CHRNA and IREB2 have been identified to be unambiguously associated with increasing COPD susceptibility (Pillai *et al.*, 2009).

Not all polymorphisms have a negative effect on COPD progression. Some protective genetic factors for individuals classified as being at high-risk of developing COPD have been identified. For example, a minor allele of a single-nucleotide polymorphism (SNP) in the gene for matrix metalloproteinase 12 (MMP12) (rs2276109 [–82A→G]), has been shown to have a significant protective effect on lung function in children with asthma and adult smokers as well as significantly reducing the incidence of COPD in adult smokers (Hunninghake *et al.*, 2009). MMP12 is produced by macrophages, the primary inflammatory cells that respond to smoke (Niewoehner, Kleinerman, & Rice, 1974). Increased MMP12 function in the airways has been shown to lead to degradation of elastin and is thought to be closely linked to COPD progression, although the exact mechanisms by which MMP12 can influence COPD progression are unclear (Hunninghake *et al.*, 2009).

Another example of a protective polymorphism is found in the enzyme microsomal epoxide hydroxylase (EPHX1). EPHX1 is an enzyme strongly expressed in the bronchial epithelium and is involved in the initial breakdown of exogenous toxins, such as polycyclic aromatic hydrocarbons (PAHs) produced by smoking. Two functional polymorphisms of EPHX1 exist; Tyr113His and His139Arg, or “fast-" and “slow-acting”, respectively. The Tyr113His polymorphism has also been shown to have a protective effect in high COPD-risk populations (Brøgger, Steen, Eiken, Gulsvik, & Bakke, 2006). However, high EPHX1 activity has also been associated with an elevated risk of tobacco-related lung cancer, although the authors did speculate that cigarette smoking could affect enzyme activity status (X. Li *et al.*, 2011).

It is important to note that the role of genetic factors in non-smokers with COPD has been less-widely studied, so care must be taken when interpreting findings from studies comparing genetic susceptibility of smokers to non-smokers. However, it has been implied that genetic factors are at least as important in COPD development among non-smokers as they are among smokers (Eisner *et al.*, 2010).

### 1.1.22 Pathophysiology of COPD

The pathological changes associated with COPD are observed in the airways, lung parenchyma, and pulmonary vasculature (Hogg, 2004). Characteristics of the pathological changes observed in COPD are as follows: chronic inflammation of the airways coupled with elevated numbers of inflammatory cells; remodelling and thickening of the airway walls due to repeated damage and repair; loss of elasticity by emphysematous destruction of parenchyma; chronic overproduction of mucus in the upper airway, and the formation of mucus plugs in lower airways (Hogg, 2004). All of these factors contribute to a progressive and partially-irreversible air-flow limitation, measured by a reduction in $FEV_1$ and $FEV_1/FVC$ values. In general, these changes are more markedly increased in more severe disease states (Global Initiative for Chronic Obstructive Lung Disease, 2018).

Cigarette smoke is known to be one of the greatest risk factors for COPD (Forey *et al.*, 2011); much of the advice for COPD treatment is based around smoking cessation (Pauwels *et al.*, 2001). The multiple pathways associated with mucus production and cell proliferation (Gensch *et al.*, 2004; Lemjabbar *et al.*, 2003) activated by cigarette smoke are well understood and it is clear how smoking can contribute to the thickening of airway walls and formation of mucus plugs, such as seen in Figure 1-3(B, and C).

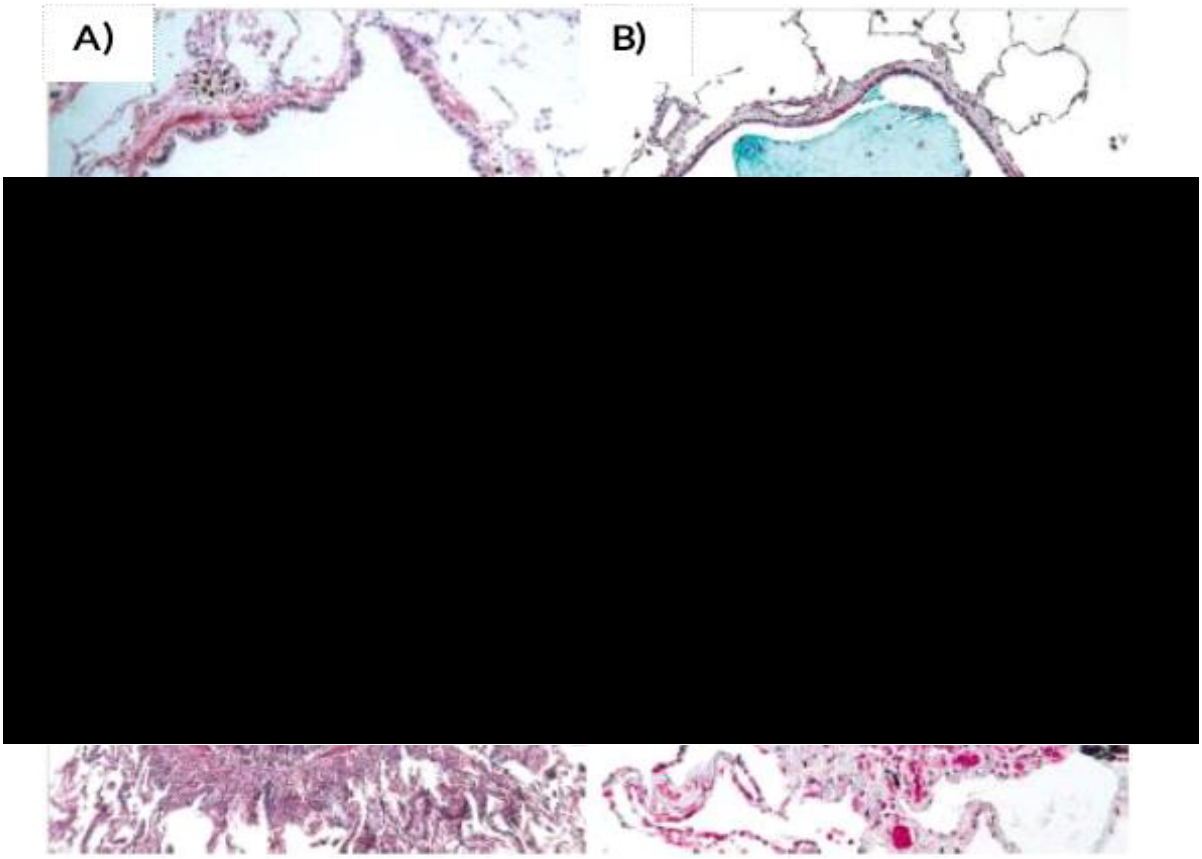


**Figure 1-3: Small airway obstruction stains, with increasing degrees of severity. (A) Normal small airway for comparison. No inflammation, no mucus overproduction. (B) Small airway with a mucus plug, but little or no evidence of hyperplasia. Mucus probably formed in the upper airway and aspirated into the small airway. This is indicative of mucus over-production and/or impaired MCC ability in the upper airway. (C) Acutely-inflamed airway with a markedly thickened wall. A large plug formed of mucus and cells, which was likely produced in the small airway, partially fills the lumen. (D) The airway is severely restricted and surrounded by connective tissue. Normal enlargement of the lumen and enlargement of the epithelial lining during lung inflation may not be possible (adapted from Hogg, 2004).**

Oxidative stress has been suggested to be an important mechanism by which COPD pathogenesis is amplified. Reactive oxygen species (ROS) have been identified as a cause of inflammation in lung tissue. The hydroxyl radical ($^{\bullet}OH^{-}$) and the superoxide anion ($O_2^{-}$) are highly unstable and are capable of inducing oxidative stress in a biological system. These ROS are continuously generated either endogenously by metabolic reactions, or exogenously from air pollution or cigarette smoke. Smoking induces MUC5AC production and secretion through activation of an epidermal growth factor receptor (EGFR)-dependant mechanism (Takeyama *et al.*, 2001) and also through an EGFR-independent mechanism triggered by ROS (Gensch *et al.*, 2004).

Smokers who show symptoms of COPD have been shown to have increased numbers of neutrophils, macrophages, $CD45^+$, $CD8^+$ cells and goblet cells compared to non-smokers. This is indicative of a sustained immune response and airway remodelling (Seatta *et al.*, 1998). The action through which cigarette smoke triggers cell proliferation is via the action of ROS-stimulated ADAM 17, (also known as tumour necrosis factor α-converting enzyme (TACE)) and amphiregulin (Lemjabbar *et al.*, 2003). Amphiregulin is a ligand for EGFR: binding of amphiregulin to EGFR activates the MAPK signal transduction pathways and triggers cellular proliferation. This increase in cellular proliferation in the airways would have a significant effect on the thickening of the airway wall and subsequent restriction of the airway itself. These same pathways have been implicated in increased mucin gene transcription and are thought to be crucial to mucus hypersecretion mechanisms (Thai, Loukoianov, Wachi, & Wu, 2008).

As seen in Figure 1-4, EGFR is phosphorylated in response to both smoke and lipoteichoic acid (LTA). Tobacco smoke generates ROS, through an NADPH oxidase dependant reaction, which in turn activates ADAM 17 (TACE) leading to amphiregulin cleavage. Conversely LTA stimulates platelet-activating factor receptor (PAFR), which consequently triggers ADAM 10, eventually leading to heparin-binding EGF-like growth factor (HBEGF) cleavage. HBEGF is also a ligand for EGFR and so HBEGF cleavage can lead to an increase in the transcription of proliferative and mucin production genes.

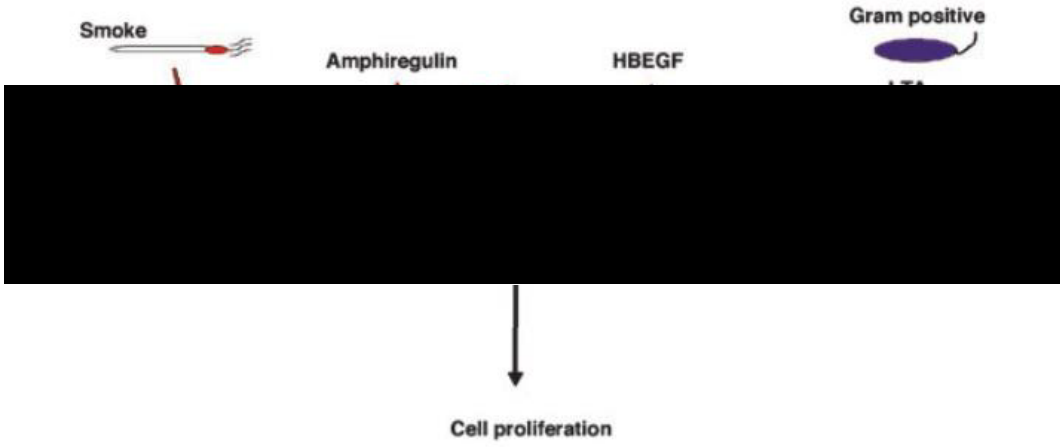


**Figure 1-4: A model for the activation of the EGFR signalling cascade by tobacco smoke and LTA from Gram-positive bacteria (adapted from Lemjabbar *et al.*, 2003)**

Another signalling pathway involved in COPD pathogenesis is the interleukin (IL)-17/IL-6 paracrine-autocrine loop. IL-17 is an important cytokine for autoimmunity, contributing towards the response of the immune system to bacterial and viral infection (Onishi & Gaffen, 2010), and is more highly expressed in the presence of cigarette smoke (Montalbano *et al.*, 2015). Evidence for the involvement of IL-17 in COPD pathogenesis is shown by the observation that IL-17 acts in tandem with IL-6 through a JAK2-dependent pathway to stimulate MUC5AC and MUC5B expression in the bronchial epithelial cells. It was also demonstrated that both IL-16 and IL-17 directly interact with the promoter region of MUC5B, stimulating expression (Y. Chen *et al.*, 2003). Increased IL-6 levels have been associated with increased severity of COPD, and IL-17 expression may be elevated by exposure to cigarette smoke (Chang *et al.*, 2014; Liang, Zhang, & Song, 2013).

IL-17 stimulates IL-6 expression and secretion in bronchial epithelial cells (Fossiez *et al.*, 1996), so it is easy to see how, in the presence of persistent inflammatory stimuli such as those present in the COPD lung and the observation that IL-17 is elevated in end-stage COPD (Roos *et al.*, 2015), a positive feedback loop for mucin gene expression can be formed, leading to increased MUC5AC and MUC5B secretion, contributing towards COPD pathogenesis.

### 1.1.23 COPD Exacerbation Mechanisms

Exacerbations of COPD are defined as an event in the natural course of the disease that are characterised by an acute worsening of the patient's baseline dyspnoea, cough, and/or sputum production beyond day-to-day variations, and are sufficient to warrant a change in management (Papi, Luppi, Franco, & Fabbri, 2006). The state of health of a COPD patient is heavily influenced by the frequency of acute exacerbations, with those being prone to more frequent exacerbations undergoing more hospital admissions and generally having a marked decrease in their quality-of-life (QOL) (Seemungal *et al.*, 1998). Exacerbations of COPD can be triggered by a variety of factors including bacterial and viral infection, and air pollutants (Papi *et al.*, 2006).

Exacerbation increases inflammation and oxidative stress in the lung, possibly due to the large numbers of neutrophils that are activated in response to the enhanced levels of proinflammatory cytokines such as IL-8 and tumour necrosis factor-α (TNFα) present in the COPD lung (Noguera *et al.*, 2001; Zheng *et al.*, 2017). Additionally, during acute exacerbation the levels of proinflammatory cytokines are

further increased, alongside an increased load of the transmembrane mucin MUC1 (Zheng *et al.*, 2017). MUC1 has functions associated with many signal-transduction pathways, including the MAP kinase pathway (Singh & Hollingsworth, 2006), which has previously been shown to contribute to increased expression of mucin genes (K. S. Song *et al.*, 2003). Levels of MUC1 expression has also been correlated with colonization by *Pseudomonas aeruginosa* and *Streptococcus pneumoniae* in the lung, with an increase in MUC1 expression observed after bacterial infection suggesting a potential protective effect of MUC1 (Dhar, Ng, Dunne, & Sutton, 2017; Kato *et al.*, 2017). MUC1-mediated increases in MUC5AC/B via MAP kinase pathways could therefore be one of the mechanisms through which mucus is hyper-secreted into the airways in colonized COPD lungs.

The lower airways of between 25-50% of COPD patients are colonized by bacteria, in particular *Haemophilus influenzae*, *S. pneumoniae*, and *Moraxella catarrhalis* (Papi *et al.*, 2006). The extent of colonization is closely correlated with the severity of disease and cigarette smoking (Monso *et al.*, 1999). Bacterial colonization of the airways directly influences airway inflammation, as it is known that bacterial exoproducts are mediators of the inflammatory response. In this way bacterial colonization could have a significant effect on the frequency and severity of COPD exacerbations. *H. influenzae* has been shown to induce more airway inflammation through increased neutrophil activation and as such, colonization of a COPD lung by *H. influenzae* could lead to a significantly increased risk of exacerbation compared to patients infected by other colonizing strains such as *S. pneumoniae* (Chin *et al.*, 2005).

Respiratory viruses have been found in approximately 50% of currently exacerbating COPD patients, compared to fewer than 20% of stable COPD patients. The most common viruses that have been reported to be in COPD sputa are rhinovirus, picornavirus, influenza A and respiratory syncytial virus (RSV). It is of note that these viruses were generally not detected in the nasal lavage of the patients (Rohde *et al.*, 2003; Seemungal, Harper-Owen, Bhowmik, Jeffries, & Wedzicha, 2000). Viral exacerbations are also associated with symptomatic colds and an increased recovery time compared to non-viral exacerbations (Seemungal, Donaldson, Bhowmik, Jeffries, & Wedzicha, 2000).

It has also been suggested that elevated levels of air pollution are key contributors towards COPD patient mortality, giving strong evidence that exacerbations can be caused by air pollution. Strong associations between elevated

levels of sulphur dioxide ($SO_2$), and nitrogen dioxide ($NO_2$) levels and COPD patient mortality were shown in a 5-year time-series analysis, with weaker associations shown between the levels of pollution and the general, non-COPD population (Garcia-Aymerich, Tobías, Antó, & Sunyer, 2000).

Interestingly, some results have shown that COPD exacerbation could be influenced by gender. It has been observed that more female than male COPD patients in the USA are dying, and where smoking is implicated in the development of COPD, women develop more severe COPD symptoms after fewer pack-years compared to men (Han *et al.*, 2007). Additionally, a recent multicentre study of women (n = 1,369) found that lung function declines more rapidly in post-menopausal women than in pre-menopausal women (Triebner *et al.*, 2017).

This observed increase in female deaths from COPD has been attributed to the increased levels of the oestrogens, a family of sex steroid hormones which are generally expressed at higher levels in women (Tam *et al.*, 2011). This explanation is probable as it is known that the oestrogens – including oestrone, oestradiol, and oestriol – and their associated receptors, oestrogen receptor (ER) α and ERβ, are expressed in the large bronchi and lung tissue respectively (Koehler, Helguero, Haldosén, Warner, & Gustafsson, 2005), therefore any alterations in oestrogen levels could affect the physiology of the lungs and large bronchi. Oestradiol is known to increase the expression of cytochrome P450 (CYP) enzymes, such as CYP1A1 (Mollerup, Ryberg, Hewer, Phillips, & Haugen, 1999), which are responsible for the metabolism of cigarette smoke into Phase I intermediate metabolites. These metabolites can cause damage to the lung through oxidative stress when the Phase II enzymes, which conjugate the metabolites to allow them to be excreted, become saturated. It is suggested that the oestrogen family of hormones contribute to oxidative stress in the lungs through this upregulation of Phase I enzymes (Tam *et al.*, 2011).

Evidence of physiological changes in lung tissue caused by alterations in oestrogen expression comes from the observation that oestradiol can cause significant increases in MUC5AC synthesis in a human bronchial epithelial cell line ($1HAE_o$), via action of nuclear factor of activated T-cell (NFAT) (Tam, Wadsworth, Dorscheid, Man, & Sin, 2014). MUC5B synthesis in human bronchial epithelial cells has also been shown to be upregulated by elevated oestradiol concentrations (Choi *et al.*, 2009). Therefore it is plausible that increased levels of oestrogen can lead to an increase in

mucin synthesis in the bronchioles and could be an important indicator of a patient's likelihood of exacerbation. Combining these increased levels of mucin synthesis with the previously stated observations of increased expression of oestradiol contributing to increased oxidative stress in female lungs through the CYP enzyme pathway, it is possible to suggest that the increased rate of COPD mortality in women, and the worsened symptoms in women who smoke may be due to the naturally higher levels of oestrogens.

Another mechanism which has been identified as a potential contributor towards exacerbation symptoms is an imbalance in secreted protease and antiprotease enzymes during time of exacerbation. A small study (n = 9) of COPD patients found that an increase in secreted antiproteases at the start of an exacerbation contributed to higher mucin concentration and stability and thus contribute towards mucus obstruction. Whilst a very small study, these results are still interesting as the authors reported that elevated levels of mucin glycoproteins MUC5AC and MUC5B persisted up to 5 weeks after the onset of initial symptoms (Chillappagari *et al.*, 2015).

Unfortunately it is hard to have an accurate picture of the true volume of COPD exacerbations, as most generally go unreported. One study found that around 50% of all exacerbations are not reported, as COPD patients tend to be accustomed to frequent symptom changes and thus do not feel the need report each occurrence (Seemungal *et al.*, 1998). This has a detrimental effect on the management and treatment of COPD, and therefore can have a significant impact on the patient's health.

### 1.1.24 Diagnosis of COPD and Exacerbation

A diagnosis of COPD is considered when a patient presents with the symptoms of cough, increased sputum production, dyspnoea, and/or a history of exposure to risk factors for the disease, as previously described. Currently COPD diagnosis is confirmed by spirometry, or a lung function test, which is the gold standard as it is the most reproducible, standardised and objective way of measuring lung function and air-flow limitation (Global Initiative for Chronic Obstructive Lung Disease, 2018). Spirometry involves measuring the patient's $FEV_1$ and FVC. A post-bronchodilator $FEV_1$ score of < 80% and a $FEV_1$/FVC score of < 70% which is not fully reversible is indicative of reduced lung function and therefore is used to confirm suspected COPD (Pauwels *et al.*, 2001). Although this method is readily available and

provides reproducible results, it is not a reliable indicator of disease progression, or exacerbation prediction (X. Chen, Xu, & Xiao, 2013).

By detecting biomarkers of disease progression and exacerbation, it is possible to identify which patients are most at risk, and therefore it could be possible to reduce patient morbidity. It has been observed that COPD exacerbations occur in clusters, with a period of elevated risk for a secondary exacerbation of approximately 8 weeks after the initial occurrence (Hurst *et al.*, 2009). There is a strong requirement for a low-cost, non-invasive and reliable method to detect COPD progression and exacerbation at an earlier stage in order to more effectively manage and treat the disease.

### 1.1.3 Lung Cancer

#### 1.1.31 Introduction to Lung Cancer

Lung cancer is a malignant tumour found within the tissues of the lung which is characterised by uncontrolled cell proliferation. Generally speaking, lung cancer can be categorised into two major types: small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC), although these can be further divided into subtypes; for example, common forms of NSCLC are adenocarcinoma and squamous-cell carcinoma (SCC).

Lung cancer is a major cause of cancer death in the UK, contributing to 23.6% and 13.6% of cancer deaths in males and females, respectively (Stewart & Wild, 2014). From 1995-2009, 552,143 patients in the UK were diagnosed with lung cancer, and during this time the 5-year survival rate improved from 7.3 to 9.6% (Allemani *et al.*, 2015). This however is still extremely low, and the main contributing factor to this poor prognosis is the inability of current diagnostic techniques to detect early-stage disease, when it is more easily treated. A review comparing 1-year lung cancer survival rates and stage-at-diagnosis found that on average 71.1% of patients diagnosed with stage I NSCLC survived at least 1-year after diagnosis, compared to an average of 15.5% of stage IV patients. The same review also found that 1-year survival of SCLC patients is also affected by stage of diagnosis, falling from 55.9% for stage I & II patients, to 14.4% for stage IV patients (Walters *et al.*, 2013).

#### 1.1.32 Lung Cancer Diagnosis

Currently, the most commonly used methods for detection of lung cancer include flexible bronchoscopy, CT-scan and X-ray (Latimer & Mott, 2015; Sutedja,

2003). However, these techniques are not effective for early detection of the disease, as evidenced by the extremely poor rate of diagnosis of early-stage disease (Allemani *et al.*, 2015; Morgan & Wilkes, 2017), and the high rate of late-stage diagnosis as seen in Figure 1-5 (Cancer Research UK, 2017).

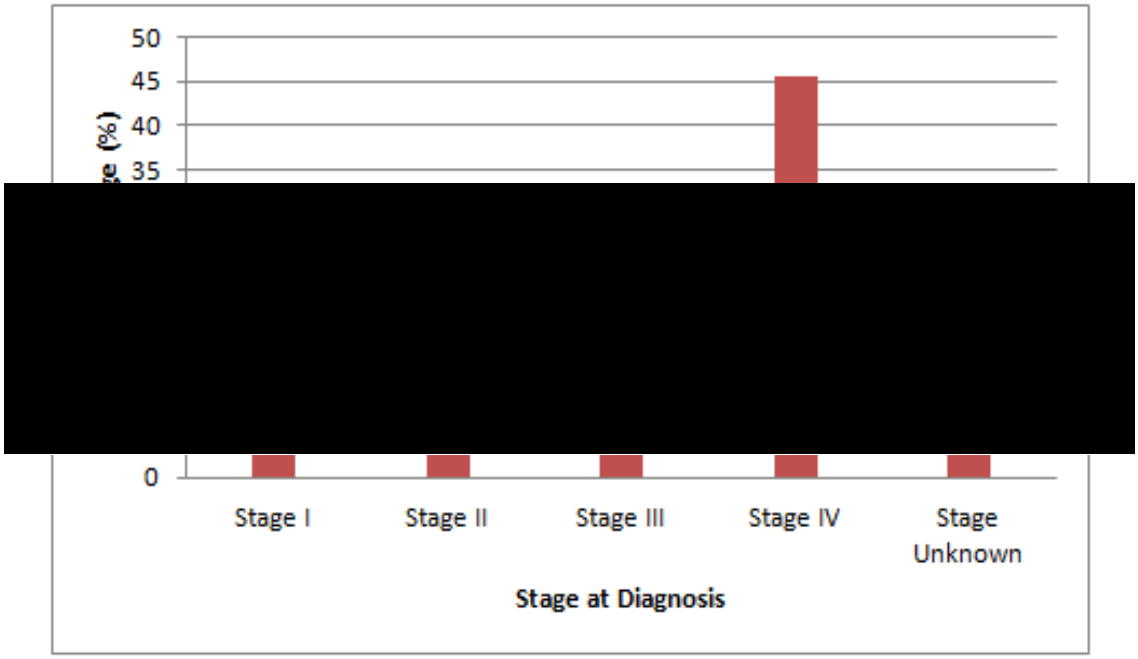


**Figure 1-5: Average percentage of lung cancer cases by T stage at diagnosis: diagnosis of late-stage disease is much more common than early-stage. Data obtained from Cancer Research, 2017**

Flexible bronchoscopy has been shown to have an overall sensitivity for lung cancer diagnosis of 88%, however this sensitivity drops markedly for peripheral lesions of < 2cm in diameter to 34% (Rivera, Mehta, & Wahidi, 2013). Two common forms of bronchoscopy, auto-fluorescence imaging (AFI) and white-light bronchoscopy (WLB), have sensitivity scores of 80%, and 53.3%, respectively, while the specificities were 83.3% for AFI, and 50% for WLB (Andolfi *et al.*, 2016). Patients who are suspected of lung cancer also undergo diagnostic X-rays and computed-tomography (CT) scans to identify regions of malignant tissue. Almost 1 in 4 (23%) of diagnostic X-rays have been shown to provide a false negative results (Stapley, Sharp, & Hamilton, 2006), whilst CT-scan has been shown to have 88.9% sensitivity and 92.6% specificity for diagnosis of lung cancer in a study comparing X-ray to CT-scanning (Toyoda, Nakayama, Kusunoki, Iso, & Suzuki, 2008). However, CT-scanning is limited by the potential for over-diagnosis and causing radiation-related harm to the patient (Hoffman & Sanchez, 2017; Mulshine & D'Amico, 2014). Evidently, there is

also a clear unmet need for a highly-sensitive and specific diagnostic tool, capable of diagnosing both centrally- and peripherally-located lesions.

### 1.1.33 Environmental and Genetic Risk factors for Lung Cancer Development

Smoking was first identified as a major contributing factor to the development of lung cancer in 1950 (Doll & Hill, 1999). In the years 2015/16 smoking was attributed as a contributing factor in 54% of all cancer deaths from cancers that can be caused by smoking (*Statistics on Smoking England: 2017*, 2017). The WHO identifies tobacco use as a major contributor to developing lung cancer, but there are also important environmental and occupational risks, as well a multitude of genetic factors (Stewart & Wild, 2014). Indeed, it is estimated that 10-25% of lung cancers worldwide occur in patients who are reported to have never smoked cigarettes and lung cancer in never-smokers is more common in women than in men (Couraud, Zalcman, Milleron, Morin, & Souquet, 2012). Existence of comorbid conditions has an influencing factor on lung cancer development; for example, COPD patients have a higher risk of developing lung cancer as their forced expiratory volume in one second ($FEV_1$) declines (Skillrud, Offord, & Miller, 1986; Tockman, Anthonisen, Wright, & Donithan, 1987). In fact, COPD and lung cancer are proposed to be related through a chronic inflammatory mechanism (Houghton *et al.*, 2008), so the finding that the risk of developing lung cancer increases as COPD progresses is not surprising. Exposure to airborne pollutants has also been linked to development of lung cancer. Environmental exposure to tobacco smoke, air pollution, and cooking- and heating-fuel fumes have all been linked to an elevated risk of lung cancer in never-smokers and the general population (Couraud *et al.*, 2012).

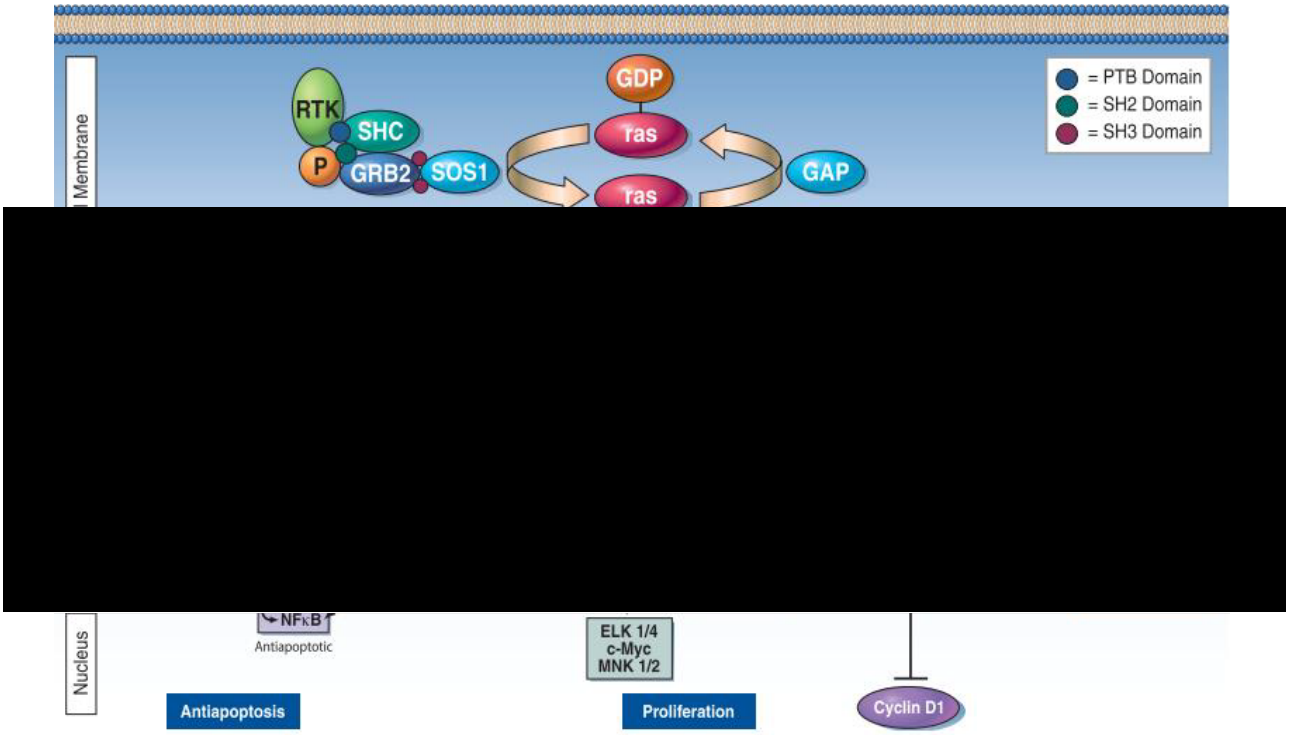


**Figure 1-6: *K-Ras* signalling transduction pathway: Activation of the *k-Ras* pathway leads to cell proliferation and antiapoptosis through the Raf-1/MEK/ERK and PI3K/Akt pathways, respectively. BAD = Bcl-2–antagonist of cell death; ELK = ephrin receptor EphB1 precursor; GDP = guanosine diphosphate; IKKB = inhibitor of NF-κB kinase β subunit; MNK = MAP kinase interacting S/T kinase; MST1 = mammalian sterile 20–like 1; NORE1 = novel Ras effector 1; PDK1 = phosphoinositide-dependent kinase–1; PKB = protein kinase B; PLA2 = phospholipase A2; ROS = reactive oxygen species. Figure adapted from Aviel-Ronen, Blackhall, Shepherd, & Tsao, (2006)**

In addition to environmental contributors to lung cancer risk, there are also a multitude of genetic factors which have been associated with an elevated risk for lung cancer. Activation/stimulation of oncogenes is a common genetic mechanism through which cancer can develop. One such oncogene is the GTPase *k-Ras*, which is involved in signal transduction pathways leading to cell proliferation (Figure 1-6). Overactive mutated forms of k-Ras have been found in 20-30% of all NSCLC (Aviel-Ronen *et al.*, 2006).

As well as overactive oncogenes, inactivation/suppression of tumour-suppression genes can contribute to an increased risk of lung cancer. One tumour-suppressor gene, of which mutated forms have been well associated with an increasing risk of cancer, is *p53*. Mutated, inactive forms of *p53* have been found in more than 50% of NSCLC cases

In addition to contributing to an increased overall risk for developing lung cancer, certain mutations have been shown to be associated with specific forms of

lung cancer developing. For example, EGFR mutations in lung adenocarcinoma have also been found to correlate with a greater number of metastatic lesions in the lung and brain, and a lower lymph node stage (Enomoto, Takada, Hagiwara, & Kojima, 2013). Also, a recent study identified upregulation of *COL5A1* as a key determinant of metastasis in lung adenocarcinoma. Observations of cell-growth-inhibition in *COL5A1* knockdown-human adenocarcinoma metastatic cells, and expression of *COL5A1* was observed to be higher in adenocarcinoma patients, with an associated increase in recurrence and decreased survivability, all contributed to this conclusion (Liu *et al.*, 2017).

### 1.1.34 Lung Cancer Screening for Early Detection

According to Cancer Research UK, the most common route to diagnosis of lung cancer is presentation of the patient to the emergency room, with more than 35% of patients being diagnosed this way (Figure 1-7). Also, a higher proportion of stage IV carcinomas are diagnosed in the emergency room than in any other route to diagnosis (Figure 1-8).

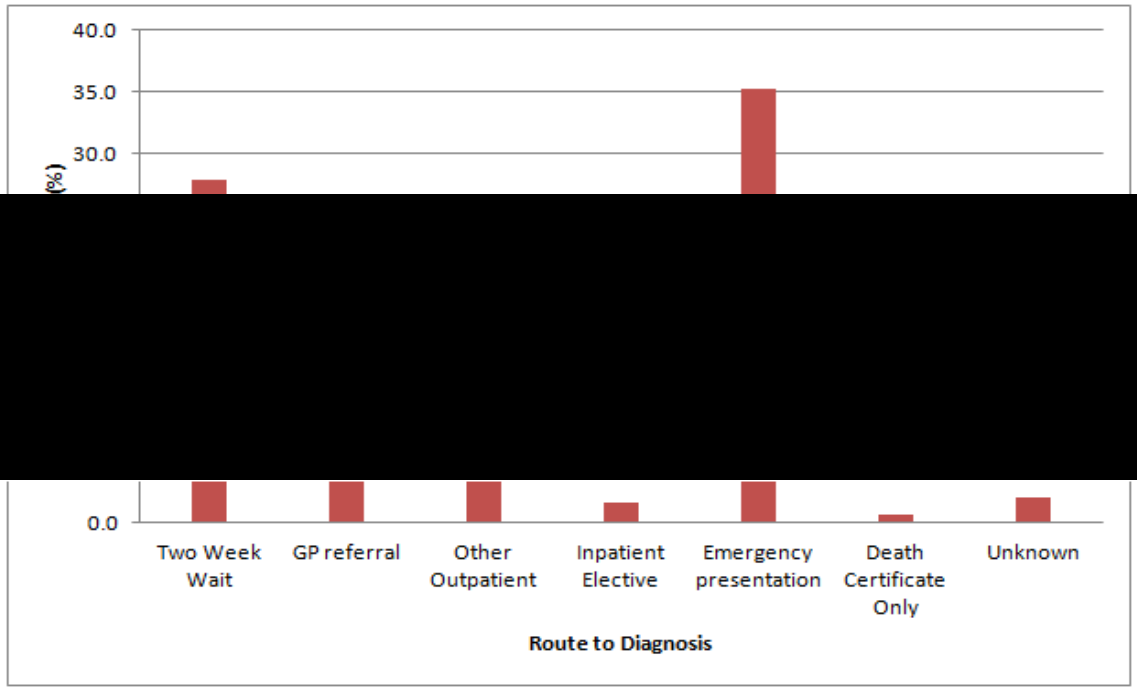


**Figure 1-7: Percentages of lung cancer diagnoses by each route to diagnosis: The data show how emergency presentation is the most common way patients receive a diagnosis of lung cancer, followed by the Two Week Wait, and general practitioner (GP) referral. Data obtained from Cancer Research UK, 2017**

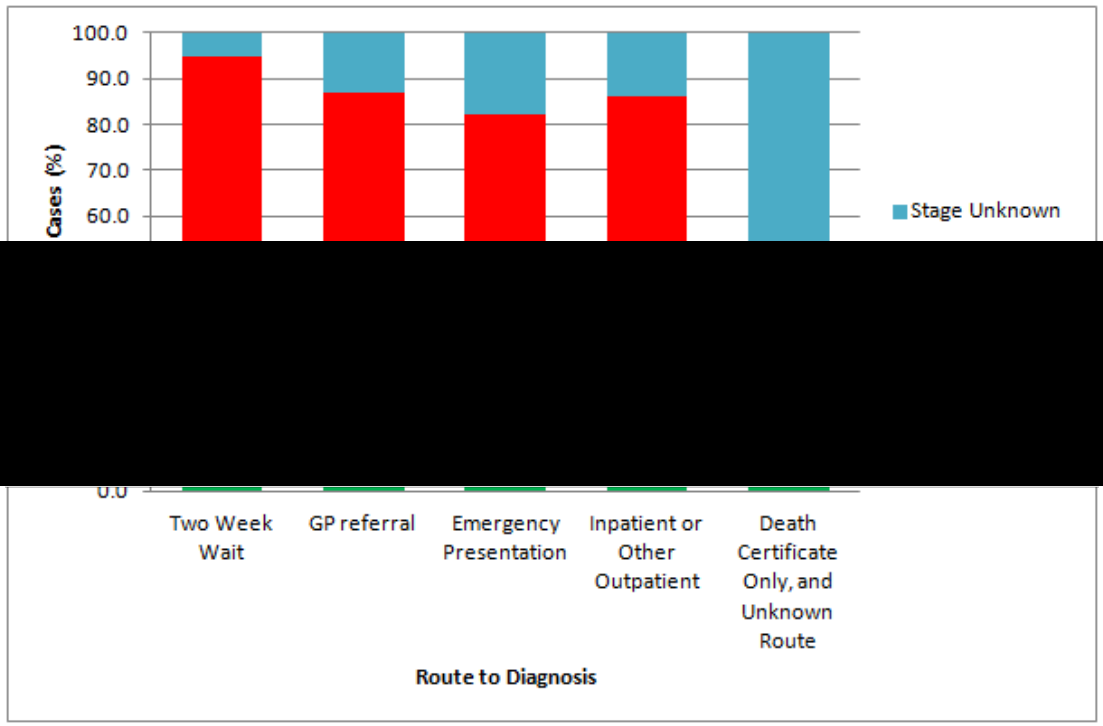


**Figure 1-8: Percentages of stages (stage I: green; stage II: yellow; stage III: orange; stage IV: red, unknown stage: blue) at diagnosis, split by route to diagnosis: a higher proportion of stage IV cases are diagnosed by emergency presentation than stage I-III cases. Additionally, whilst the proportion of stage IV cases are highest, there are more patients diagnosed at stages I-III in the two week wait and GP referral pathways than in the other routes to diagnosis. Data obtained from Cancer Research UK, 2017**

These data from Cancer Research UK suggest that patients who are in contact with their health service and/or their GP have a greater chance of being diagnosed at an earlier stage of disease, and therefore have a greater chance of survival (Figure 1-7, Figure 1-8 & Figure 1-9). This clearly demonstrates the need for an effective screening programme which can be implemented into the diagnostic pathway. Through regular screening of a high-risk population of patients, it would be possible to increase the likelihood of diagnosing lung cancer at earlier stages, and therefore increase the chances of survival.

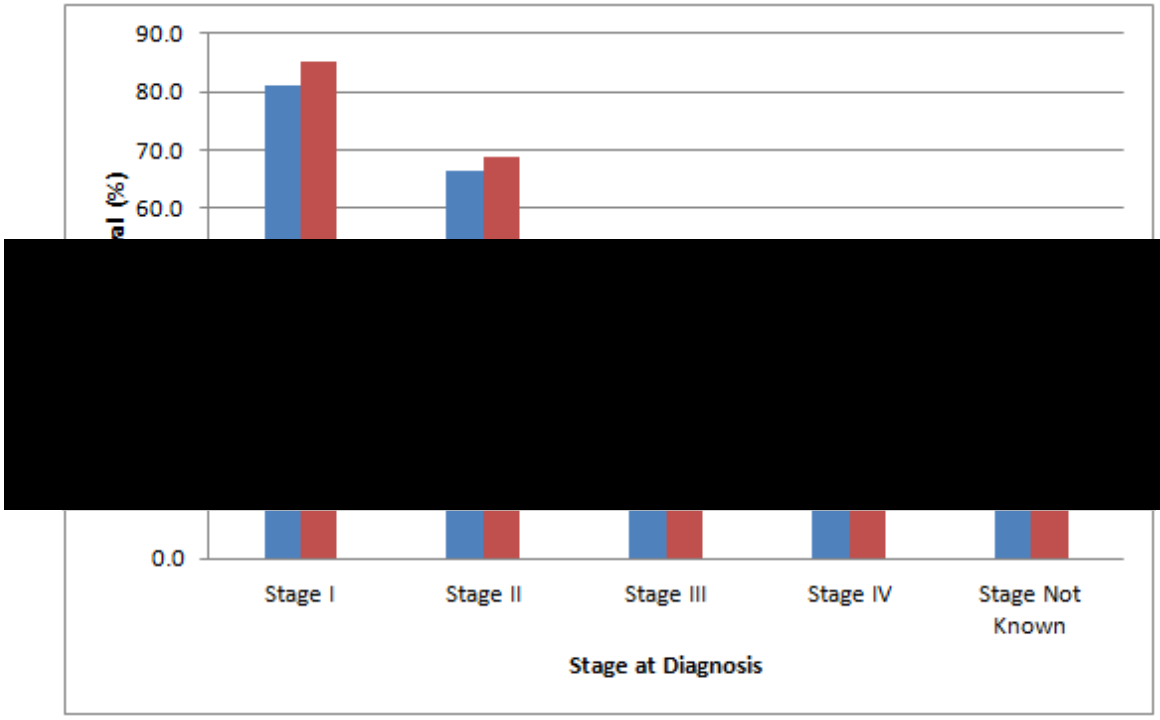


**Figure 1-9: Net one-year survival percentages for lung cancer patients, distinguished by stage at diagnosis. The data show that women (red) have a larger one-year survival percentage than men (blue) at all stages, but this survival percentage decreases sharply for both genders as the stage at diagnosis increases. Data obtained from Cancer Research UK, 2017**

Screening of high-risk respiratory disease patients, such as COPD patients, for early signs of lung cancer could improve early diagnosis and therefore could lead to an improvement in prognosis. As previously stated, there is evidence that diagnosis at later stages has a negative effect on the patient's overall prognosis, with poorer survival rates reported in those patients (Cancer Research UK, 2017; Walters *et al.*, 2013). However, any screening programme must be effective at detecting early-stage lung carcinoma, and must contribute to a reduction in mortality in the target population.

A meta-analysis of lung cancer screening programme trials has found that screening programmes where frequent chest X-rays was the modality of choice were associated with an 11% increase in patient mortality compared to less frequent screening (Manser *et al.*, 2013). Reviews by Mulshine & D'Amico (2014) and Hoffman & Sanchez (2017) of screening programmes for lung cancer detection identified evidence from the National Lung Screening Trial (NLST), that low-dose CT (LDCT) screening was effective at reducing lung cancer mortality in at-risk populations through an associated 20% decrease in death rate in patient groups that underwent yearly screening. The reviews also identified a range of issues with LDCT screening

which must be addressed in any potential screening technology. Among the issues identified were over-diagnosis, physical harm caused by the technique to the patient, and integration of the screening technology into primary care (Mulshine & D'Amico, 2014). Indeed, it was reported by the NLST that 26% of LDCT screening tests were positive; however 96.4% of all positive screenings were false positives. The guidelines for detection of abnormal nodules have now been changed to reflect this, although this has led to a sensitivity decrease of 8.6-15.2% (Hoffman & Sanchez, 2017). The same review also identified the cost of implementing LDCT screening within the USA, with a potential cost of $6.8 billion over a 5-year time frame. Despite this cost, it is suggested that LDCT screening could be cost-effective with an estimated $81,000 per quality-adjusted life-year, although this was variable depending on the subgroup of patient screened with the highest-risk patients showing most cost-effectiveness (Hoffman & Sanchez, 2017). A lower-cost technology would be more cost-effective for all patient groups screened and could therefore be more easily rolled-out to a wider population.

Sputum cytology is another commonly-used tool for lung cancer diagnosis (Rivera *et al.*, 2013; Thunnissen, 2003; Toyoda *et al.*, 2008). It has been suggested to have limited effectiveness as a stand-alone screening technology (Manser *et al.*, 2013). However, the potential for using sputum cytology for early-stage cancer detection can be enhanced through combination with DNA analysis and nuclear image analysis (Thunnissen, 2003), although this adds extra complexity, time and financial requirements.

LDCT screening programs have seen good success in reducing lung cancer mortality, but there are significant issues and costs that must be overcome if it is to become widespread and routine. There is a clear and strong need for a highly-sensitive and specific, low-cost, rapid and easily-implementable screening technology for lung cancer diagnosis.

### 1.1.4 Cystic Fibrosis

#### 1.1.41 Introduction to Cystic Fibrosis

Cystic fibrosis (CF) is an autosomal recessive disorder which primarily affects the airways, through airway-restriction by increased mucus load (Voynow, Gendler, & Rose, 2006). There are approximately 85,000 patients with CF worldwide and this number is projected to increase by 75% by 2025 (Burgel *et al.*, 2015; Kris De Boeck &

Amaral, 2016). This is mainly due to an increasing life-expectancy for CF patients, coupled with a decreasing mortality rate – see Figure 1-10 (Cystic Fibrosis Foundation, 2016). Despite this increase in life-expectancy, the leading cause of death for CF patients is still respiratory failure (Elborn, 2016).

A major characteristic of CF is an increased mucus load within the airways due to a reduced mucus clearance capacity (Elborn, 2016). The increased mucus load restricts the airflow, reducing lung capacity and has a significant impact on the QOL for the patient, with patients generally reporting a decrease in QOL over time (Uchmanowicz, Jankowska-Polanska, Rosinczuk, & Wleklik, 2015). Patients must undergo constant physical therapy and monitoring in order to appropriately manage their disease and improve their QOL (Button *et al.*, 2016; Feiten *et al.*, 2016).

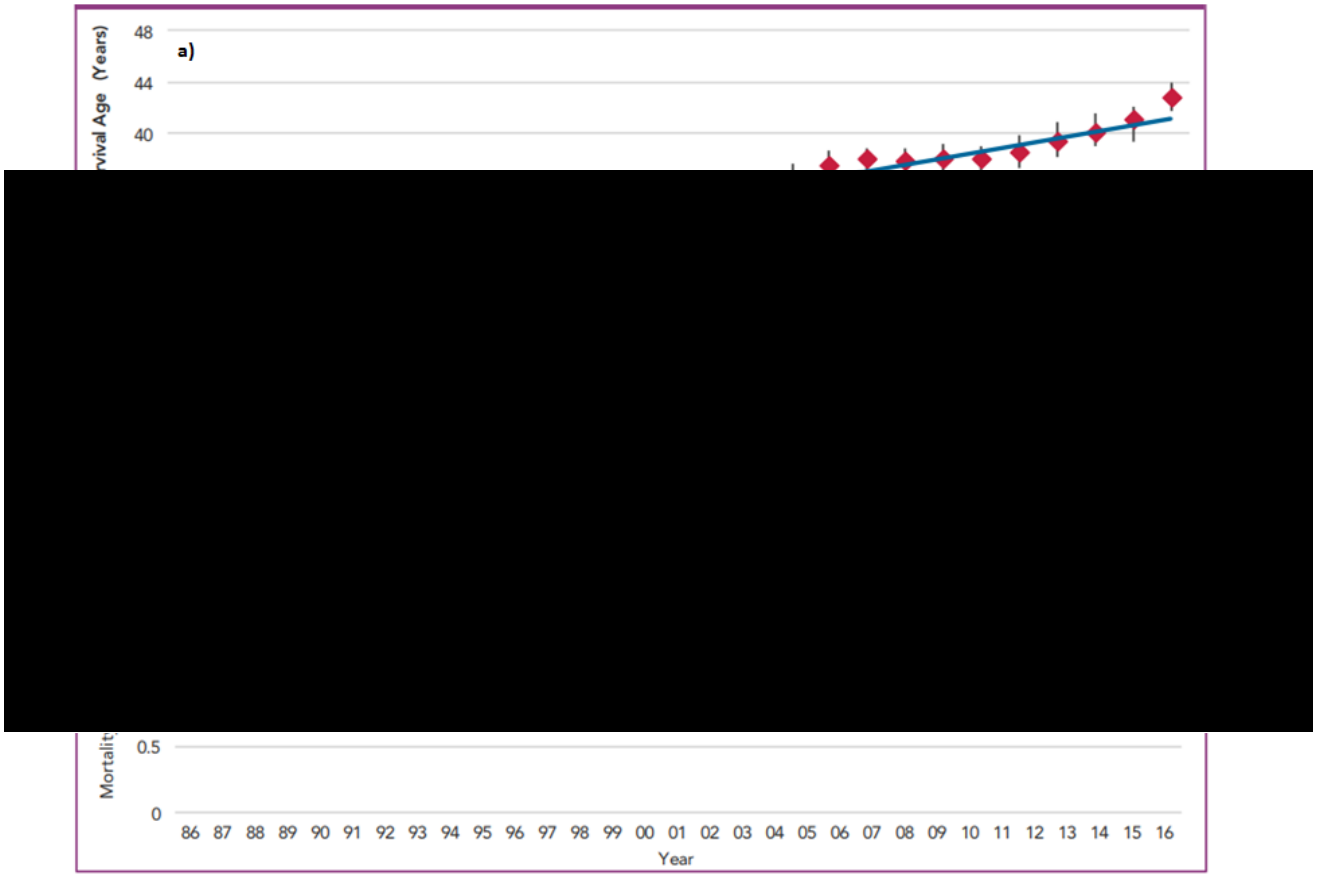


**Figure 1-10: a) Median predicted survival of CF patients from 1986 to 2016 in 5 year increments, and b) mortality rate per 100,000 people with CF, from 1986 to 2016. There is an overall trend of predicted survival increasing, coupled with a decreasing mortality rate (figures adapted from Cystic Fibrosis Foundation, 2016)**

CF is primarily caused by mutations or deficiency in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, with the most common mutation being the homozygous Phe508del CFTR mutation (K. De Boeck, Zolin,

Cuppens, Olesen, & Viviani, 2014a). The *CFTR* gene encodes for the CFTR transmembrane chloride and bicarbonate ion channel and mutations in this protein lead to dysregulation of epithelial fluid transport (Gustafsson *et al.*, 2012). Because of this, CF mucus contains less fluid and is therefore less hydrated than non-CF mucus. As a direct consequence of this, MCC activity is impaired in CF-patients (Lee & Foskett, 2014). Dysregulation of the CFTR channels can also lead to an elevation in salt levels in CF airway mucus, which has been shown to inhibit the activity of antimicrobial peptides (Zabner, Smith, Karp, Widdicombe, & Welsh, 1998). Due to the combination of reduced MCC activity and inhibition of antimicrobial peptides, CF patients are more susceptible to airway infection.

As a consequence of these inhibited activities, most CF patients are persistently colonized by biofilm-forming *Pseudomonas aeruginosa* and *Burkholderia* strains (Powell *et al.*, 2014; Wagner & Iglewski, 2008). A bacterial biofilm is formed when bacteria adhere to solid surfaces and secrete extracellular polysaccharides to form a matrix to which more bacteria cells can adhere. Biofilms also contribute antimicrobial-agent resistance, including antibiotics, further increasing the susceptibility of CF patients to persistent infection (Donlan, 2001).

Pulmonary infection by *Aspergillus, Candida, Lomentospora* and *Scedosporium* species is also common in CF patients (Garczewska, Jarzynka, Kuś, Skorupa, & Augustynowicz-Kopeć, 2016; Schwarz *et al.*, 2018)

### 1.1.42 Treatments for Cystic Fibrosis

Currently, treatments for CF are focussed on the reduction of the severity of symptoms, with most CF patients undergoing regular pulmonary and physical therapy (Cystic Fibrosis Foundation, 2015). Management of CF is complex, but recently developed targeted therapies for CF have seen success; one prominent new therapy, Ivacaftor, has been shown to improve lung function, and reduce the risk of pulmonary exacerbations through a potentiating effect on CFTR (Ramsey *et al.*, 2011). However, there is limited evidence that Ivacaftor has significant antimicrobial and antibiofilm properties, especially when targeted against *P. aeruginosa* colonies and biofilms (Payne *et al.*, 2017; Reznikov *et al.*, 2014). Therefore, therapies based on administration of Ivacaftor would require additional therapeutic input in order to reduce the bacterial load on the patient.

Tezacaftor and Lumacaftor are two novel therapeutic CFTR modulators which recently have been shown to function as combination drugs alongside Ivacaftor under the names of Symdeko and Orkambi, respectively. This work does not focus on these therapies so they will only be described in brief detail.

The clinical effectiveness of Tezacaftor-Ivacaftor combination therapy has been studied. One randomised controlled trial of 510 patients showed a 35% lower rate of pulmonary exacerbation in patients who had received Tezacaftor-Ivacaftor (Taylor-Cousar *et al.*, 2017). Another study found that the combination therapy of Tezacaftor-Ivacaftor is safe, with only moderate side effects reported (Davies *et al.*, 2018).

Lumacaftor-Ivacaftor has been shown to be effective in CF patients with the homozygous Phe508del CFTR mutation in a study of 1108 patients. It was found that approximately twice any patients who received the combination of Lumacaftor-Ivacaftor experienced an improvement in predicted $FEV_1$(%) of 5% or greater from baseline, than patients who received placebo alone. Similarly, patients who received Lumacaftor-Ivacaftor treatment also experienced a longer time to first pulmonary exacerbation after starting treatment than patients in the placebo cohort (Wainwright *et al.*, 2015). A recent review on the safety of Lumacaftor-Ivacaftor combination therapy found that, although the overall risk-benefit ratio was positive in favour of the therapy, care must be taken when prescribing to certain populations. It was found that Lumacaftor-Ivacaftor therapy may worsen liver function in CF patients with advanced liver disease, and it may also cause cataracts (Guevera & Mccolley, 2017).

Another novel inhaled therapeutic for CF which is currently under development is OligoG. OligoG is a low weight alginate oligosaccharide extracted and purified from the stem of brown seaweed (*Laminaria hyperborean*) and composed of a repeating chain of α-L-guluronate subunits. This thesis has a focus on the action of OligoG in the CF lung, and therefore OligoG will be described in further detail.

Studies have been carried out focussing on the clinical effectiveness of OligoG, which have shown that OligoG has promise as a treatment of CF. It has been shown to have antimicrobial, antibiofilm, calcium-chelating and mucolytic properties (Ermund *et al.*, 2017; Powell *et al.*, 2013; Roberts *et al.*, 2013; Vitko *et al.*, 2016), all of

which can contribute to a reduction of CF respiratory symptoms. Crucially, OligoG has been shown to have a potent antibiofilm affects against bacterial biofilms, which are commonly found in the CF lung (Powell *et al.*, 2013). *Burkholderia* and *P. Aeruginosa* colonisations of the CF lung are also significantly disrupted by OligoG treatment, through interactions between OligoG and the bacterial surface which inhibit bacterial motility (Powell *et al.*, 2014). OligoG has been shown to have an potentiating effect on antibiotics, increasing the effectiveness of azithromycin, aztreonanm ceftazidime, erythromycin, and oxytetracycline against Gram negative bacteria up to 512-fold (Khan *et al.*, 2012). OligoG has also been shown to be effective against fungal infections commonly found in the CF lung. Anti-fungal treatments nystatin, amphotericin B, fluconazole, miconazole, voriconazole, and terbinafine were all shown to have their activity enhanced by OligoG, with nystatin showing the most improvement with 16-fold reduction in minimum inhibitory concentration of nystatin required against *Aspergillus* and *Candida* strains (Tøndervik *et al.*, 2014).

Research is currently underway to fully elucidate the mechanisms utilised by OligoG within the airway mucus. Phase I, IIa and IIb studies have been carried out and have demonstrated that OligoG is safe for human use, with no intolerance up to 540 mg/day (https://ClinicalTrials.gov, NCT00970346, NCT01465529, NCT02157922 & NCT02453789). Although OligoG can be administered by inhalation daily without adverse effects, there is limited understanding of how OligoG is eliminated from the lung (Pritchard *et al.*, 2016). Other inhaled therapeutics are eliminated from the airways through a combination of macrophage uptake and alveolar clearance (Ibrahim & Garcia-Contreras, 2013). It is reasonable to hypothesise that OligoG could be eliminated from CF airways in a similar manner, but this is currently unconfirmed and requires further testing to confirm.

## 1.2 Sputum, Mucus and Mucins

### 1.2.1 Sputum

Sputum is the expectorated secretion of mucus from the upper airways, usually produced by patients with chronic inflammation of their lungs (Voynow & Rubin, 2009). It is a gel-like substance comprised mainly of mucus and is produced in the upper airways. It has a critical protective role in the lung, forming the first barrier to pathogens and foreign particles that enter the lung. The composition of sputum is known to change in chronic disease patients, and it is thought that alterations to the

biochemistry of a patient's sputum can have a direct effect on the progression of chronic lung disease (Kirkham *et al.*, 2008).

### 1.2.2 Mucus and Mucins

Mucus is a highly viscoelastic substance which coats all of the organs exposed to the external environment, including the respiratory, gastrointestinal, reproductive, and oculo-rhino-otolaryngeal tracts (Bansil & Turner, 2006). In the airways, the respiratory mucus gel forms the first line of defence against pathogens, airborne toxins and pollutants, which are trapped in the luminal mucus layer and constantly removed from the airway by MCC (K. C. Kim *et al.*, 1997).

The main macromolecular constituent of mucus is mucin; mucin is a general term for the large family of glycoproteins found in mucus secretions (Thornton, Rousseau, & McGuckin, 2008). Mucins are very large glycoproteins which may be secreted to form mucus gels, or mucins may be membrane-bound (Bansil & Turner, 2006). The other major components of mucus are water (~95% w/w), immunoglobulins, lipids, inorganic salts, cholesterol and other proteins (Bansil, Stanley, & LaMont, 1995). Mucins are responsible for the biological, physical and chemical properties of mucus and provide the structural matrix from which its protective properties can be derived (Georgiades, Pudney, Thornton, & Waigh, 2014). Therefore it is easy to see how abnormalities in mucus and mucin production can often lead to the development of serious pathological conditions and even death of patients. MUC5AC and MUC5B are the main constituents of airway mucus, and therefore are the most important mucins with regards to chronic respiratory diseases. Currently, 21 human mucin genes have been identified and endorsed by the HUGO gene nomenclature committee: *MUC1-MUC8, OVGP1, MUC12, MUC13, EMCN, MUC15-MUC17 and MUC19-MUC22* (Yates *et al.*, 2017), and the messenger ribonucleic acid (mRNA) expression values of these mucin genes in many different tissues have been characterised by the GeneAtlas® survey (Su *et al.*, 2004).

There is evidence to suggest that the cystic fibrosis transmembrane conductance regulator (CFTR) channel acts as a bicarbonate channel and is required for normal mucin secretion (Borowitz, 2015; Gustafsson *et al.*, 2012). When mucins are secreted into the lung they are initially in tightly-packed, high calcium-ion concentration granules as an N-terminal concatenated ring platform. These tightly-packed rings then expand as the calcium is removed by bicarbonate, increasing the

pH and forming the mucin mesh component of mucus, (Gustafsson *et al.*, 2012). Hydrogen bonds and $Ca^{2+}$ mediated cross-links are also required for proper formation of the mucus mesh (Meldrum *et al.*, 2018).

### 1.2.3 Protein Glycosylation and Glycan Synthesis

Mucins are highly glycosylated and the structure of the glycans is highly important in determining the biophysical properties of the mucin and mucus. Therefore it is key to have an understanding of the mechanisms of protein glycosylation and glycan synthesis. These will be briefly described.

Protein glycosylation is described as the attachment of polysaccharide chains, or glycans, to proteins and is the most observed and diverse post-translational modification of proteins, being found on cytosolic, cell-surface and secreted proteins (Shental-Bechor & Levy, 2008). A key feature of glycosylation is a high degree of heterogeneity in the sites of glycosylation and the structures of the glycans (Packer, Lawson, Jardine, Sanchez, & Gooley, 1998). Mammalian glycosylation occurs on the nascent protein chain as it travels through the endoplasmic reticulum lumen and Golgi complex. Glycosylation patterns on proteins have crucial roles in establishing protein thermostability and correct folding, as well as molecular recognition and immune responses (Shental-Bechor & Levy, 2008). There are two types of glycosylation; O- and N-glycosylation. Both types are distinct with respect to the amino acid residue at which the glycan is attached, and the required precursor molecules.

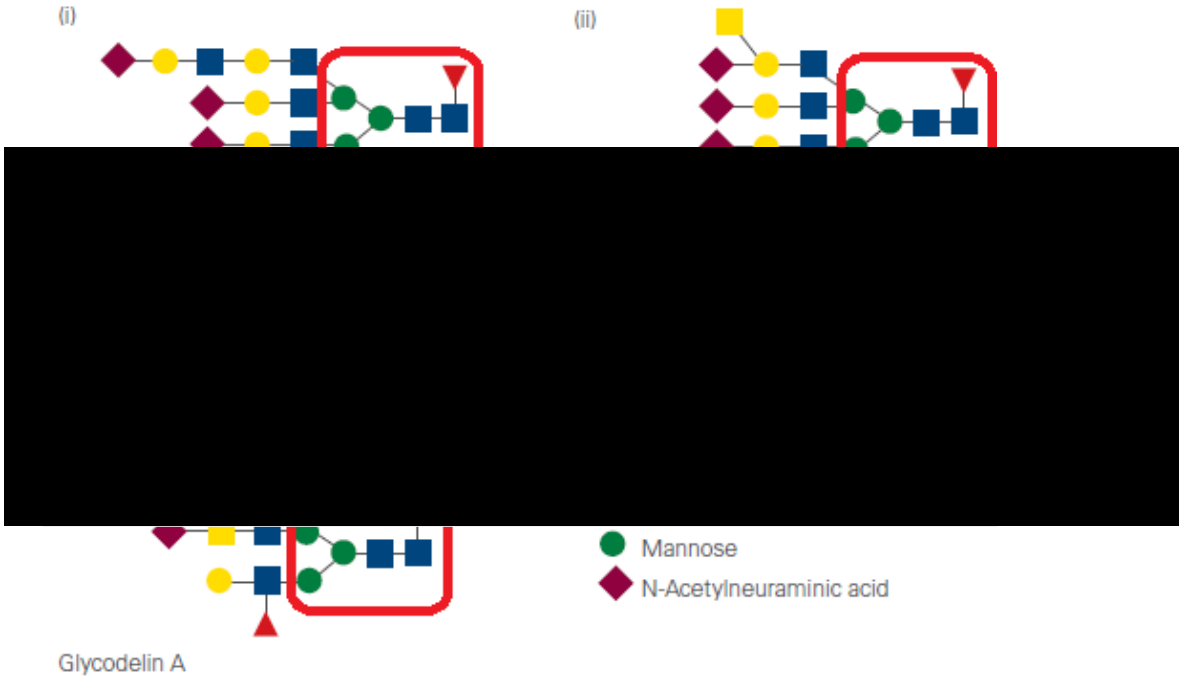


**Figure 1-11: Examples of N-glycans in Consortium for Functional Glycomics (CFG) format and the glycoproteins or cell types in which they are found. The conserved core structure, formed of two N-Acetylglucosamine (GlcNAc) and three mannose residues, is encircled by the red box to highlight its position (adapted from Easton, 2011).**

N-glycosylation requires a 14-mer precursor structure covalently bonded to an asparagine residue within the consensus sequence of Asn-X-Ser/Thr, where X is any amino acid except proline. Not every consensus sequence can be glycosylated, due to the 3-dimensional folding of the protein preventing access by oligosaccharyl transferase to the protein backbone. After binding, the precursor molecule is trimmed down to a conserved core structure, shown in Figure 1-11, containing two N-Acetylglucosamine (GlcNAc) residues and three mannose residues (Easton, 2011). Glycosyltransferases then carry out the step-wise addition of monosaccharide building blocks leading to the formation of complex and diverse structures with multiple branching chains. N-glycosylation does not occur on mucin glycoproteins and hereinafter will not be further explained.


**Figure 1-12: Examples of O-glycans and the glycoprotein or the cell types in which they are found. Note all O-glycans start with an N-Acetylgalactosamine residue bound to either a serine or threonine amino acid residue, but there is no single conserved core structure (adapted from Easton, 2011).**

O-glycosylation is the only type of glycosylation modification that occurs on mucin glycoproteins. It contrasts with N-glycosylation in that there is no conserved consensus sequence on which glycosylation occurs, and no single core glycan structure (Figure 1-12). Instead O-glycosylation occurs almost exclusively on serine or threonine amino acid residues (Easton, 2011). O-glycans tend to be much smaller than N-glycans, usually being comprised of around 3 to 6 monosaccharide units. They always are attached to the protein backbone via an N-Acetylgalactosamine residue but there is no conserved core structure; rather a family of eight core structures can be formed, although only four are commonly observed on mucin protein backbones.

The first residue is invariably N-Acetylgalactosamine (GalNAc), transferred to the protein backbone by GalNAc transferase. Specific glycosyltransferase enzymes then perform additions of monosaccharide units in a manner comparable to the assembly of N-glycans. The chain is initially elongated by addition of galactose (Gal) or GlcNAc residues to form one of the core O-glycan structures (Figure 1-13). These glycan cores are then modified by additional glycosyltransferase enzymes, until the chain is terminally glycosylated by addition of fucose, sialic acid (N-acetylneuramic acid), or a sulphate residue (Rose & Voynow, 2006). Interestingly, the GlcNAc-6-O-sulfotransferase from human bronchial mucosa is only active on terminal GlcNAc

residues (Degroote *et al.*, 1997), clearly imply that sulphate residues can only ever be located at the termini of the glycan structures.

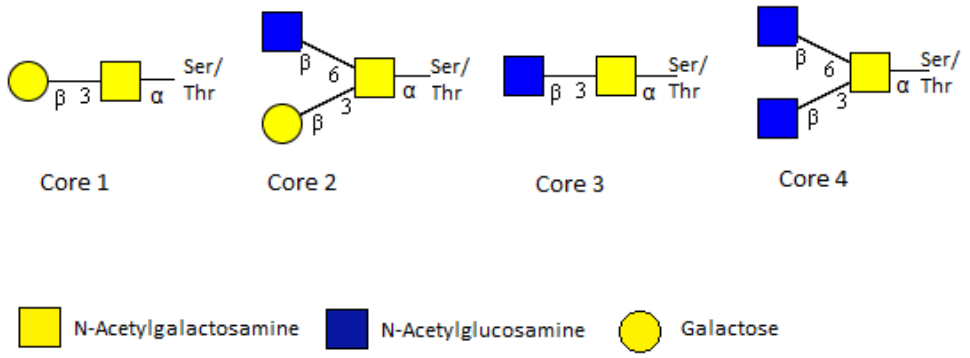


**Figure 1-13: Four of the core structures of O-glycans commonly found in glycoproteins. A GalNAc monosaccharide is covalently bonded via an α-bond to either a serine or threonine residue on the mucin protein backbone. The glycan chain is then extended with the addition of galactose and GlcNAc residues to form the four common O-GalNAc structures. Further additions of monosaccharide residues then extend these core structures, forming complex, diverse and possibly branched glycan chains.**

Through progressive addition of monosaccharides to the glycan chain, it is possible for the Lewis antigen blood-group determinants to be formed within the glycan structure. These structures are associated with inflammation and are characteristic of respiratory disease (Davril *et al.*, 1999; Degroote *et al.*, 2003; Kirkham *et al.*, 2008). The presence of Lewis antigens within the glycan structure confers a negative charge to the mucin and can influence the rheology of the mucus (A. T. Lewis, Jones, Lewis, Jones, & Lewis, 2013).

### 1.2.4 Mucin Glycoprotein Structure

Although the mucin family is large and diverse, mucin glycoproteins share many characteristics. In general, mucins are large, highly glycosylated, viscoelastic glycoproteins (Figure 1-14) synthesized in specialized goblet cells found in the surface epithelium, and mucous cells of the submucosal glands. They have a high molecular weight of between 0.5-2 MDa, are approximately 80% carbohydrate by weight, with a large number of O-glycans (Bansil & Turner, 2006). Mucins also possess an extensive number of tandem repeats (TR) in the backbone of the protein. It is these TR domains that distinguish mucins from other glycoproteins (Rose & Voynow, 2006). These TR regions have a very high proportion of serine and threonine residues, which are the sites of O-glycosylation (K. C. Kim *et al.*, 1997).

Despite the similarities throughout the mucin glycoprotein family, there are some important differences between each mucin. Mucins can be subdivided into two

main groups; secretory or membrane bound (Rose & Voynow, 2006). Within these groups there is a variation in the structure of the mucin based on size, and the presence or absence of defining features such as TR units, cysteine-rich motifs and von Willebrand factor (VWF) D4, C1 or C2 domains (Rose & Voynow, 2006). The secretory mucins are capable of forming oligomers with one another – the backbone of the mucus gel. Evidence suggests that both MUC5AC and MUC5B mucin monomers dimerize through C-terminal disulphide linkages. These mucin dimers then form oligomeric chains through disulphide linkages of the N-termini (Sheehan *et al.*, 2004; Wickström, Davies, Eriksen, Veerman, & Carlstedt, 1998). These oligomers then interact with one another through H-bonding and intermolecular forces to form the mucus gel.

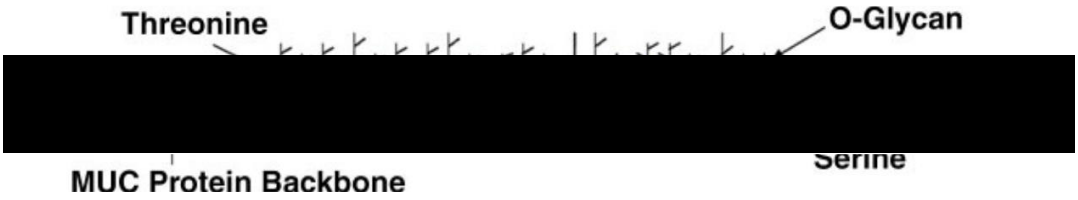


**Figure 1-14: A schematic drawing of a secretory mucin glycoprotein. The MUC protein backbone is represented by the blue, yellow and green sections. The central region (yellow) represents the tandem repeat region, which is the area where glycan attachment occurs on serine and threonine residues, indicated by white and black circles respectively (adapted from Rose & Voynow, 2006).**

### 1.2.5 Mucin-mesh structure

As stated above, the main components of mucus responsible for its visco-elastic gel-like properties are the high-molecular weight mucin glycoproteins (Bansil & Turner, 2006). Mucins polymerize to form long chains, linked by disulphide bonds at the C- and N-termini (Thornton *et al.*, 2008). The mucin chains have an overall negative charge and therefore bind readily with positive ions (Bansil & Turner, 2006). Inter-chain interactions between salivary MUC5B chains has been shown to be mediated by $Ca^{2+}$ ions (Raynal, Hardingham, Sheehan, & Thornton, 2003). More recently, high-concentrations of $Ca^{2+}$ have been identified as crucial for proper packaging and secretion of MUC5B into the airway, and failure to reabsorb calcium and uncouple the calcium cross-links in the mucus packages after secretion can impede mucin expansion and therefore has a major impact on the physical properties of the mucus gel (Ridley *et al.*, 2014).

Exposure of the mucus to oxidating agents, such as ROS, can have significant impacts on the physical properties of the mucus gel. Oxidation of airway mucus has

been found to increase the number of inter-chain disulphide cross-links in healthy airway mucus causing an increase in elasticity of the mucus. In the same study, mucus from CF patients was shown have abnormally high concentrations of oxidised disulphide cross-links which correlated with the level of ROS in the respective samples (Yuan *et al.*, 2015).

### 1.2.6 Mucin Gene Expression and Regulation

Multiple *in situ* hybridisation studies have determined that MUC5AC and MUC5B are differentially expressed throughout normal, healthy respiratory tissues. MUC5AC is strongly expressed in the goblet cells found in the surface epithelium, and is not expressed in the submucosal glands of the trachea and bronchi. Conversely, MUC5B is expressed at very low levels in the goblet cells and moderate to high levels of MUC5B expression are found in the submucosal glands (Audie *et al.*, 1993; Buisine *et al.*, 1999; Reid, Gould, & Harris, 1997). Neither MUC5AC nor MUC5B were found to be expressed in healthy adult bronchioles nor the alveoli (Buisine *et al.*, 1999), so presence of these mucins in respiratory disease patients' lower airways is indicative of abnormal mucus production and/or impaired MCC activity. Mucus hypersecretion is one of the main symptoms of COPD and is indicative of an increased rate of production of MUC5AC and MUC5B from the goblet cells and submucosal glands, which could partially result from goblet cell hyperplasia (Molfino & Jeffery, 2007; Saetta *et al.*, 2000). Indeed, MUC5AC expression is known to be elevated compared to MUC5B in stable COPD patients (Caramori *et al.*, 2009). It has been shown that goblet cell hyperplasia is accompanied by an increase in sialic acid residues in infected murine intestines (Y. S. Kim & Ho, 2010). This could be relevant to chronic respiratory disease in humans as MUC5AC and MUC5B have been shown to be well-conserved across species (Desseyn, Aubert, Porchet, & Laine, 2000).

The main gel-forming mucins are coded for by the *MUC2*, *MUC5AC*, *MUC5B* and *MUC6* genes. These genes are found in a cluster on 11p15.5 and are thought to have evolved from one common ancestor (Desseyn *et al.*, 2000). Regulation of mucin gene expression occurs via many distinct mechanisms, whereby one regulatory agent can cause the up- or down-regulation of multiple MUC genes, through the activation of multiple pathways. These regulators are usually inflammatory/immune response mediators and are a diverse group of molecules, although they can be broadly categorised into four groups; (I) inflammatory cytokines, (II) bacterial products, (III) growth factors, and (IV) environmental chemicals or pollutants (Thai *et al.*, 2008).

Specific mucin genes can be regulated at both the transcriptional and post-transcriptional level by specific regulators. For example, the agonist lipopolysaccharide (LPS) found on the cell wall of Gram-negative bacteria such as *Pseudomonas aeruginosa*, has been shown to transcriptionally upregulate both MUC2 and MUC5AC via distinct mechanisms. Such mechanisms include the upregulation of MUC2 via an Src-dependent NF-κB activation (J.-D. Li *et al.*, 1998), and MUC5AC upregulation by TNF-α converting enzyme (TACE) mediated TGFα release and EGFR activation (Shao, Ueki, & Nadel, 2003). Another way mucin gene expression is post-transcriptionally regulated is through TNF-α

In the lungs, mucin proteins are synthesised in two main types of secretory cells; submucosal glandular cells and the surface mucous/goblet cells that line the epithelium. The main mucins synthesised in the lung epithelium are MUC5AC and MUC5B. The expression and synthesis of these mucins is tightly regulated to ensure that over-production/hypersecretion of mucins does not occur. In non-COPD sputum, the ratio of MUC5AC to MUC5B is much higher than that of COPD sputum (Kirkham *et al.*, 2008), which suggests the regulation of expression of the airway mucins is altered during disease states.

*MUC5AC* is located on chromosome 11p15.5, in a cluster along with *MUC5B, MUC2* and *MUC6* and is primarily expressed in surface goblet cells in the lungs (Hovenberg, Davies, & Carlstedt, 1996). *MUC5AC* is strongly regulated by inflammatory cytokines, bacterial exoproducts, growth factors, proteases, pollutants and viruses. A common pathway utilised by many of these regulatory molecules is the MAP kinase pathway, with the ERK and p38 molecules (Figure 1-15).

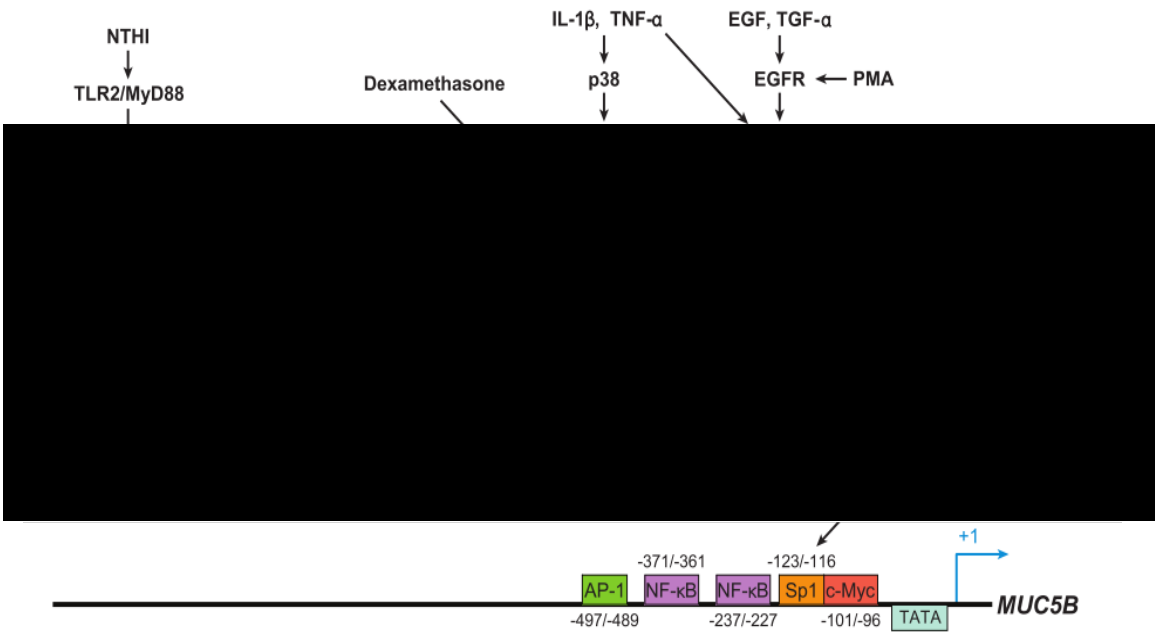


**Figure 1-15: A schematic of the 5' promoter regulatory regions of MUC5AC and MUC5B, showing some specific regulatory and signal transduction pathways involved in the expression of MUC5AC and MUC5B (adapted from Thai *et al.*, 2008).**

*MUC5B* is located next to *MUC5AC* in the 11p15.5 cluster and is almost exclusively expressed in the submucosal glands of the airways (Gosalia, Leir, & Harris, 2013). *MUC5B* expression has been shown to be stimulated in an autocrine/paracrine loop by the action of the cytokines IL-6 and IL-17 (Y. Chen *et al.*, 2003). IL-17 induces IL-6 secretion in human airway epithelial cells, through a Janus kinase (JAK)-dependant pathway, whereupon binding of IL-17 to IL-17R activates JAK2. This in turn leads to IL-6 production and secretion, which will bind with IL-6R. Binding of IL-6R leads to ERK signalling, which is known to implicated in upregulation of mucin genes.

### 1.2.7 Use of Sputum for Biomarker Discovery in Chronic Respiratory Disease

As detailed above, sputum production is elevated through multiple mechanisms during the course of chronic respiratory disease. As a consequence of this, sputum is a readily available, non-invasive sample source from these patients, which makes it an attractive prospect for biomarker discovery. Sputum has been shown to be a valuable source of biomarkers for respiratory disease. For example, a recent study has shown how asthma and COPD patients can been differentiated from COPD-asthma-overlap syndrome patients through significantly elevated levels of neutrophil gelatinase-associated lipocalin (NGAL) in COPD-asthma-overlap patient's sputum (Iwamoto *et al.*, 2014).

Sputum cytology has been used to accurately detect early-stage lung cancer in high-risk groups, however it requires a highly-trained and experienced operative, and strict technical control between laboratory environments to be most effective (Endo *et al.*, 2015). Its use is currently limited in lung cancer screening studies (Manser *et al.*, 2013), despite having a higher chance for detection of early-stage tumours than X-ray screening (Thunnissen, 2003).

Examination of sputum samples from CF patients revealed that severity of lung disease in CF patients has been correlated to levels of surface-bound neutrophil elastase on sputum-neutrophils, which has subsequently been suggested to be a novel biomarker for CF lung disease (Dittrich *et al.*, 2018).

### 1.2.8 Alterations to Respiratory Mucin Glycan Structure & Composition in Chronic Respiratory Disease

Key to the pathogenesis of COPD is an alteration in the properties of the mucus layer that lines the airways. Aside from a much greater amount of mucus being produced (i.e. mucus hypersecretion), the composition of the mucus is altered. The ratio of MUC5B : MUC5AC is changed, so much greater amounts of MUC5B are produced in COPD patients (Kirkham *et al.*, 2008; Kirkham, Sheehan, Knight, Richardson, & Thornton, 2002). Because of the highly-diverse nature of the monosaccharide building blocks of the glycans, the properties of the glycans can vary greatly. Glycans composed with more N-acetylneuraminic acid or sulphated residues have an overall greater negative charge and are more acidic (A. T. Lewis *et al.*, 2013). Therefore it is possible for overall acidity of the mucus to be affected by the levels of N-acetylneuraminic acid and sulphated residues. The degree of fucosylation of the mucins can also affect the rheology of the mucus, as an increase of fucosylation has been shown to increase hydrophobicity of the mucins (A. T. Lewis *et al.*, 2013).

There can be different glycoforms of specific mucins based on charge differences – specifically MUC5B has two glycoforms, a low-charge form and a high-charge form. The high-charge form is composed with more sialic acid, as well as more sulphated residues forming sulphate esters. A change in the balance of the charged glycoforms of mucins can have serious implications for the patient's health, as demonstrated by the finding of a highly-viscous mucous plug composed of mainly the low-charge glycoform of MUC5B in the lungs of a patient in *status asthmaticus* (Sheehan, Howard, Richardson, Longwill, & Thornton, 1999). Similarly, a recent study

focusing on children with acute asthma also identified increased MUC5AC and low-charge glycoform of MUC5B present in airway mucus (Welsh *et al.*, 2017).

Differences in the amounts of low- or high-charge glycoforms of mucins can have significant effects on the overall pH of the mucus due to the different levels of sialic acid and sulphated residues. Both increased sialylation and sulphation displayed by the high-charge glycoforms are associated with increased acidity. A study performed by Davril and colleagues demonstrated significant differences in the levels of sialic acid and sulphate in airway mucin glycans between infected and non-infected cystic fibrosis (CF) and chronic bronchitis (CB) patients. Average sulphation of bronchial mucins of CF patients was found to be higher than that of CB patients. Sulphation levels are also significantly increased in infected CF and CB patients compared to non-infected CF and CB patients. The sialic acid content of the same bronchial mucins was observed to be significantly higher in infected patients than in non-infected patients (Davril *et al.*, 1999).

Elevated levels of oestradiol have been shown to be linked to increased fucosylation of MUC5AC, as oestradiol causes an increase in fucosyltransferase mRNA expression in human airway epithelial cell lines (Tam *et al.*, 2014). This is interesting as if it is taken together with the observation that increased fucosylation causes the hydrophobicity of the mucins to increase (A. T. Lewis *et al.*, 2013), it is possible to suggest that increasing oestradiol levels can be associated with mucin hydrophobicity and mucus viscosity, therefore having the potential to be a reliable indicator of exacerbation risk.

## 1.3 Vibrational Spectroscopy Methods

### 1.3.1 Theoretical and Mechanistic Principles of Current Techniques

#### 1.3.1.0 Fourier Transform Infrared Spectroscopy

Fourier transform infrared (FTIR) spectroscopy is a highly-sensitive analytical method which is capable of rapidly analysing structural changes in molecules. Briefly, FTIR measures chemical bond vibrations by measuring infrared (IR) absorbance by a sample - or transmission through a sample - and then produces an infrared spectrum based on the absorptive properties of that sample which is measured in absorbance units (AU).

The theoretical basis of FTIR function is absorption of IR light, by molecular bonds such as N-H, O-H and C=O. IR light is directed at the sample which is partially absorbed in a molecule/bond-specific manner. Molecular bonds have well-defined energy levels and exist at a ground state in terms of vibrational energy. These bonds can absorb IR light, become excited and transition between energy levels. Each type of bond will absorb IR light at specific wavenumbers, which correspond to the frequency of vibration of the bond, and a spectrum of absorbance can be generated based on the total composition of molecular bonds in the sample.

By using IR light of many wavelengths to analyse the sample, it is possible to construct a distinctive IR molecular fingerprint based on the principle that different chemical bonds absorb different wavelengths. FTIR produces both narrow and broad spectral bands, which are easily resolved, molecule-specific and directly provide information on the biochemical composition of the sample. Samples generally require very little preparation before analysis by FTIR; for example, the sample does not need to be mixed into a matrix. Indeed, biofluids taken from patients can be analysed almost immediately by pipetting directly onto the device and taking a spectrum. One disadvantage of FTIR is that samples must be as free from water as possible, because water is a strong absorber of IR light and will mask all other bonds that may be present in the sample. Therefore the sample to be analysed must be completely dry before any successful analysis can be carried out as the –OH bonds in water have a very strong absorption band at 1640cm$^{-1}$ and will also obscure any peaks generated from the sample at lower wavenumbers (Oberg & Fink, 1998). When planning FTIR experiments on biofluids, drying time of the samples must be taken into account, which will be an influencing factor on which FTIR technology to use, sample preparation methods, and the number of samples which can analysed.

Mechanistically, the core of an FTIR spectrometer is an interferometer, which is comprised of a beam-splitter and two mirrors, one fixed and one movable, for the generation of a variable optical-path difference between two beams. The initial results appear as an interferogram, which is subsequently Fourier-transformed to produce an absorbance spectrum. The source emits an IR beam which is divided by the beam-splitter into two beams of equal energy. One beam is refracted towards the fixed mirror and the other beam is directed towards the movable mirror, which reflects it towards the beam-splitter. The two beams recombine and interfere with one-another

in either a constructive or destructive way, depending on the optical path difference. The recombined beam is then passed through the sample before reaching the detector. The detector measures the light intensity relative to the position of the movable mirror and creates an interferogram. This interferogram is then Fourier transformed into a readable spectrum by the computer (Barth, 2007). This spectrum contains a mix of sharp and broad, resolvable bands which are unique to the molecular content of the sample.

There are multiple configurations that FTIR spectrometers can be made from, with the two most common being attenuated total reflectance (ATR)-FTIR and transmission-FTIR (T-FTIR). ATR-FTIR and T-FTIR have some important similarities, which makes the two techniques comparable. For example, both techniques have been shown to produce almost identical information on a protein's structure (Goormaghtigh, Gasper, Bénard, Goldsztein, & Raussens, 2009), but there are important differences that must be taken into account.

#### 1.3.1.1 Transmission-FTIR

Transmission-FTIR (T-FTIR) is a common form of FTIR technology which is utilised in a broad manner of applications. The sample to be analysed is placed onto an IR-transmissible substrate, such as calcium fluoride ($CaF_2$), silicon, or low-E glass slides. Samples are applied in a liquid state and are allowed to dry to form a thin film. T-FTIR is not typically suited to analysis of solid or powder samples without sample preparation, and liquid-sample analysis must be carried out using a liquid-transmission cell. Tissue samples can be readily examined by T-FTIR, however they must first be fixed to the slide using a suitable fixing agent, which must then be removed prior to spectrum acquisition to remove its signal from the fixing agent and reduce scattering (Pilling & Gardner, 2016). T-FTIR has been used for high-definition tissue imaging in pathology, with spatial resolutions of 5-6µm achievable, allowing for extremely highly detailed IR-images of the tissue section (Pilling & Gardner, 2016). It has been shown that IR spectral histopathology (SHP) can diagnose lung cancer with accuracy comparable to multi-panel immuno-histochemistry (Bird *et al.*, 2012).

In T-FTIR, the IR beam passes directly through the sample and the IR-transmissible sampling substrate. Generating a good quality spectrum is dependent on the IR beam being able to penetrate and transmit through the sample and substrate with minimal losses due to scattering of the beam. Beam scattering can be

reduced by ensuring the sample dries evenly over the sampling area, minimising variation in the topography of the sample and reducing the "coffee-ringing" effect, whereby high concentrations of the sample are drawn to the edges of the drying sample spot. This effect can be reduced through the use of a strict experimental design (Baker *et al.*, 2015), however this adds complexity, time and cost to the procedure.

#### 1.3.1.2 Attenuated Total Reflection-FTIR

Attenuated total reflection (ATR)-FTIR is another common configuration of FTIR technology. It is capable of rapidly generating structural information of the chosen sample without the need for preparation of thin-films or the use of liquid cells, unlike in T-FTIR. In ATR-FTIR, the sample to be analysed is placed directly on top of an IR-transparent crystal. The IR-beam is refracted into and internally reflected through the crystal, creating an evanescent wave of IR radiation which penetrates into the sample applied to the crystal. IR radiation is absorbed by the sample in a molecular bond specific manner and a spectrum of absorbance is generated (Figure 1-16).

ATR crystals are usually diamond or germanium (Ge) due to these materials' high hardness and resistance to scratches, but can also be zinc selenide (ZnSe), zinc sulphide (ZnS), silicon, or potassium bromide (KBr), depending on the intended application. Each material has specific advantages and disadvantages relating to their properties, which in turn has implications regarding the suitability of the crystal for use in certain settings and environments, and for analysis of certain samples. For example, whilst diamond and Ge have a very high hardness and are suitable for use with most samples, they come with a very high cost which can limit the potential applications an FTIR device equipped with such crystals can be used in. Other crystals, including ZnS and ZnSe, however, have a much lower associated cost and so are more appropriate for use in settings where low-cost-of-use/low-purchase-cost are more important; for example, as a point-of-care device. Other crystals, such as KBr, may not be suited for use on aqueous samples, including biofluids, as KBr is soluble in water.

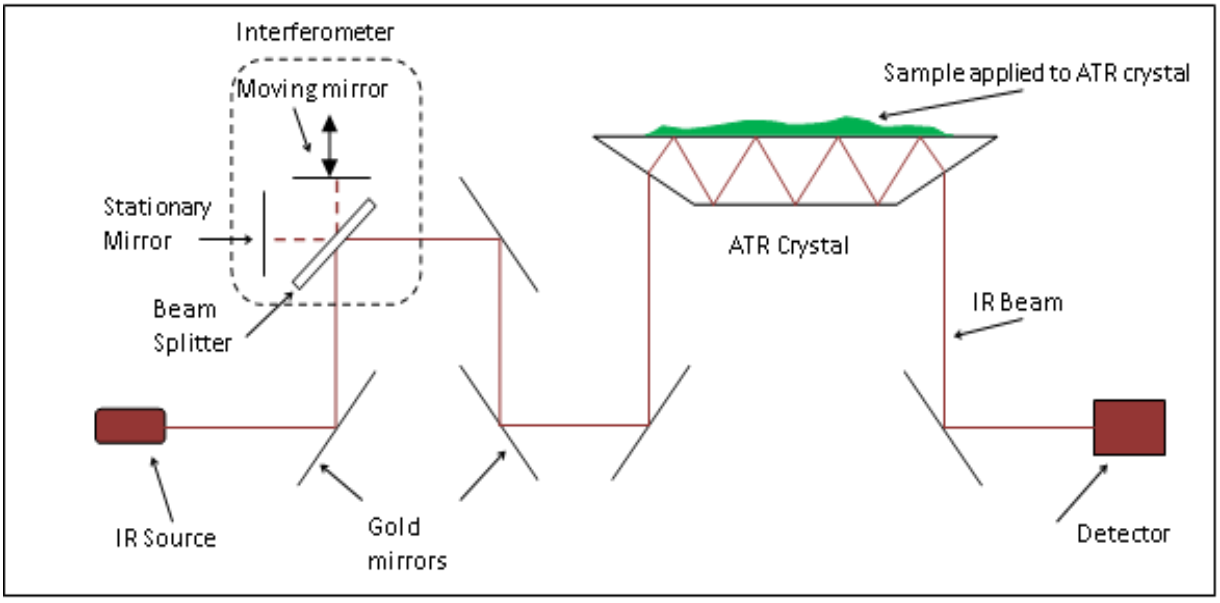


**Figure 1-16: A simplified schematic drawing of an ATR-FTIR spectrometer, showing the IR source, interferometer (dotted perimeter), ATR crystal and detector. The detected signal (interferogram) is interpreted by a connected computer and Fourier-transformed to generate the spectrum.**

Due to the rapid nature of ATR-FTIR, experiments designed around ATR-FTIR are able to monitor real-time structural alterations caused by environmental change through modification of environmental conditions (e.g. temperature, pressure, pH). ATR-FTIR overcomes many of the disadvantages associated with T-FTIR methods. For example, T-FTIR requires the use of expensive and fragile slides, usually made of $CaF_2$ for sample preparation, whereas ATR-FTIR does not. However, ATR-FTIR still requires that the sample be completely free of water, to prevent masking of peaks.

ATR-FTIR requires very little sample preparation prior to spectral acquisition, with spectra being able to be generated immediately after a sample is placed onto the sampling module. ATR-FTIR is ideal for rapid analysis of both fluid and solid/powder samples, although solid/powder samples must be held in place to ensure close contact between the sample and ATR crystal. Another advantage of ATR-FTIR is the speed-of-spectrum acquisition; it is possible to use ATR-FTIR to quickly and easily perform quality analysis throughout processes by monitoring molecular bond composition of a sample as it passes through an assay, for example cleavage of glycan chains from a protein backbone by β-elimination.

There are limitations to ATR-FTIR, for example, due to the very low penetration depth of the IR beam into the sample, it must be in extremely close

contact with the ATR crystal. This is achieved by directly spotting the samples onto the ATR crystal and allowing it to dry, forming a thin-film directly on top of the crystal. When analysing solid samples with ATR-FTIR, this contact is created by applying external pressure using a clamp to press the sample into the ATR-crystal. Another limitation of ATR-FTIR is the long spectral acquisition time of multiple samples. Individual samples can be rapidly analysed, however ATR-FTIR of many samples can be time-consuming, especially if the samples have to be dehydrated prior to spectral acquisition, as no automated high-throughput ATR system is available (Baker *et al.*, 2015).

#### 1.3.1.3. Attenuated Total Reflectance Variable Filter Infrared Spectroscopy

ATR-Variable filter infrared spectroscopy (ATR-VFIR) is a variation of ATR-FTIR spectroscopy. The spectrometer is constructed with zero moving parts and utilises a linear variable filter (LVF) detector. An example of an ATR-VFIR spectrometer is the IR Sphinx, manufactured by Comline Spectrolytic (Wackersdorf, Germany). This spectrometer is lightweight and hand-held, and therefore is truly portable. It is also relatively low-cost compared to other full size ATR units such as the Bruker Alpha, and it is these qualities that have created an interest in utilising ATR-VFIR as a potential tool for use in the clinic for monitoring of disease

ATR-VFIR instruments make use of a linear variable filter detector, consisting of a variable-wavelength filter, which allows specific wavelengths to be transmitted through at specific regions of the filter, mounted on top of a piezoelectric crystal wedge. The detector present in the IR Sphinx is the Pyreos Thin Film Pyroelectric Linear 128 Element Line Sensor Array, with a lead-zirconate-titanate (PZT) filter capable of allowing wavelengths from 1818-909cm$^{-1}$ to be transmitted (Pyreos, part no: PY0738).

#### 1.3.1.4 Final Summary Comparison of T-FTIR, ATR-FTIR & ATR-VFIR

The various FTIR technologies described so far each have their own unique features, advantages and disadvantages. Hereinafter, these will be directly compared to highlight key differences between each technology.

All of the technologies described are forms of vibrational spectroscopy, measuring the absorbance of infrared radiation through excitation of the vibrational energy of molecular bonds present in the sample of interest. FTIR spectrometers require an interferometer to spatially deconstruct the IR beam, whereas the VFIR

spectrometer has no need of an interferometer, and therefore has no moving parts and is less complex than conventional FTIR spectrometers. This reduced complexity allows VFIR spectrometers to be low cost, and small enough to be hand-held and portable, qualities which not applicable to conventional FTIR spectrometers. However, they are currently not capable of the same level of spectral resolution as FTIR spectrometers, and therefore applications in which they could be used may be limited.

As previously stated, FTIR spectroscopy can be performed in several configurations, with the two of the most common being T-FTIR and ATR-FTIR. The main difference between these configurations is the sampling mode. T-FTIR passes the IR beam directly through the sample which is held on an IR transmissible substrate, whereas in ATR-FTIR the beam is refracted and internally reflected through an IR transmissible crystal, such as diamond or KBr, on to which the sample is placed. In T-FTIR the IR beam is transmitted completely through the sample and substrate, but in ATR-FTIR IR radiation only penetrates a few microns into the sample, at the point of internal reflection in the crystal.

### 1.3.2 Use of Vibrational Spectroscopic Methods in Disease Management and Diagnosis

Due to its inherent ease of use, high reproducibility and non-invasiveness, FTIR has previously been applied with success to a range of biofluids and tissue samples. The technique is capable of analysing microlitre volumes of sample, with minimal sample preparation required.

A great deal of work has been carried out to establish FTIR as a suitable technology for diagnosis of various cancers and other diseases. An increasing number of publications show that FTIR is a viable tool for analysing biofluid samples (Baker, Trevisan, Bassan, Bhargava, & Butler, 2014), and diagnosis and monitoring of diseases such as brain cancer (Hands *et al.*, 2016), cystinuria (Oliver *et al.*, 2016), breast cancer (Backhaus *et al.*, 2010) and oropharyngeal cancers (Menzies *et al.*, 2014).

Whiteman *et al.* performed a small preliminary study into the use of FTIR spectroscopy as a diagnostic tool for detection of biomarkers of exacerbation in COPD patients' sputum. It was found that the infra-red spectra of nebulised sputum collected from COPD patients and healthy volunteers were significantly different in the amide II and glycogen-rich regions, but not the amide A region despite COPD

patients having more peaks over a wider range in this region (Whiteman, Yang, Jones, & Spiteri, 2008). This study demonstrated the proof of concept, that it is possible to use FTIR to characterise differences in COPD patients' sputum from healthy sputum. The authors attributed the observed spectral differences to changes in mucin protein secondary structure and glycosylation. In fact, FTIR is capable of determining protein secondary structure by determination of the exact positions of the absorbance bands within amide I. α-helices, and random coil structures are associated with major absorption bands centred around 1656, and 1650cm$^{-1}$, respectively whilst β-sheets produce a range of bands from 1624 to 1642cm$^{-1}$, an absorbance maximum seen around 1636cm$^{-1}$ (Goormaghtigh *et al.*, 2009; Kong & Yu, 2007; S. P. Lewis, Lewis, & Lewis, 2013).

FTIR has previously been shown to be a valuable tool in the diagnosis of lung cancer; work within our group demonstrated the efficacy with which FTIR can be utilised to analyse sputum samples from patients and determine the presence of lung cancer with high sensitivity and specificity (Paul D. Lewis *et al.*, 2010). It was also reported that FTIR spectra suggested a diagnosis of cancer in 48% of patients where no tumour was visible during bronchoscopy. This is highly important as it shows how FTIR could be used as more accurate diagnostic tool, rather than just bronchoscopy and pathology alone.

Lewis *et al.* have identified a large number of discrete sugar- and sulphate-associated infrared absorption peaks that can be used to predict the presence of the major mucin-linked carbohydrates and mucin modifications observed in the IR spectrum of sputum (A. T. Lewis *et al.*, 2013). These specific absorption peaks are able to distinguish differences in glycosylation patterns between COPD and non-COPD sputa. This has implications for the rapid distinction of COPD from other respiratory diseases, for example asthma or bronchiectasis.

## 1.4 Aims and Objectives

This thesis is focussed on evaluating the utility of FTIR for regular use on sputum samples from chronic respiratory disease patients. Sputum samples to be analysed by FTIR will be obtained from COPD patients for the prediction of exacerbation, from lung cancer patients for the diagnosis of lung cancer, and from CF patients for the investigation of the interactions of the novel therapeutic OligoG with sputum.

1. **Evaluate FTIR to determine changes to IR spectral patterns from baseline to exacerbation**
2. **Evaluate the sensitivity and specificity of the hand-held FTIR device for detection of COPD in sputum and develop a standardised protocol for sputum analysis by ATR-VFIR using a sampling strip**
3. **Further examine FTIR as a potential tool for early lung cancer diagnosis and lung cancer staging through analysis of raw sputum**
4. **Investigate the utility of FTIR for monitoring of concentration in sputum of an inhaled therapeutic, and interaction with sputum mucins, in spontaneous cystic fibrosis patient sputum**

**Objective 1: Evaluate FTIR to determine changes to IR spectral patterns from baseline to exacerbation.**

Based on previous knowledge of predicted changes to sialylated glycans in mucins during exacerbation, the student will use FTIR to generate an infrared spectral library for samples from the same patients at baseline, and at the start of their next exacerbation. These samples have already been collected by respiratory nurses, as part of the existing MEDLUNG study. This work will provide preliminary data leading to a large study for full evaluation of FTIR as a method for long-term monitoring of COPD patients and predicting exacerbation. Replicate spectra will be generated using a Bruker Alpha and high-throughput Bruker Vertex with an HTS-XT attachment. Using ratios of absorbencies at 1650cm$^{-1}$ against absorbencies of wavenumbers specific to Lewis antigen structures, we have developed a regression model that provides a single threshold score where patient scores below are predicted as baseline and scores equal/above are predicted as exacerbation. For sample size calculation we use an ROC curve to determine accuracy and set an area under the curve at 0.8 with a null hypothesis of 0.5. For Type I and II error rates at 0.01 we require a minimum sample size of 72; i.e. 36 samples at baseline followed by 36 samples at exacerbation to test model accuracy.

**Objective 2: Evaluate the sensitivity and specificity of the IR Sphinx for detection of COPD in sputum and develop a standardised protocol for sputum analysis by ATR-VFIR using a sampling strip.**

Using glycan-associated wavenumbers, evaluate the miniaturised FTIR device to detect COPD on a retrospective sputum sample set of adequate statistical power for COPD baseline cases. This data will allow the development of generalized linear models to provide the likelihood of COPD disease future trial evaluation. Using backward selection, models will be evaluated in terms of goodness of fit (inc. $R^2$ and likelihood ratio), significance of model coefficients and validated by resampling via bootstrapping and test data sets. Models will be selected according to sensitivity and specificity in classifying COPD.

**Objective 3: Further examine FTIR as a potential tool for early lung cancer diagnosis and lung cancer staging through analysis of raw sputum.**

Through utilisation of high-throughput FTIR technology, retrospectively collected sputum samples obtained through the MEDLUNG study from patients attending bronchoscopy clinics under suspicion of lung cancer will be analysed. Samples were collected with patient consent prior to bronchoscopy by instructing the patient to produce sputum by cough. Patients who receive a final clinical diagnosis of "lung cancer" will be compared to those who receive a final diagnosis of "COPD" or "non-COPD respiratory disease". Through comparison of the patient's clinical data (cancer status, morphology, stage) with replicate FTIR spectra of their raw sputum, generalized linear models will be developed to predict the likelihood of lung cancer in these patients. Regression models will be evaluated based on sensitivity and specificity scores for this known sample set. FTIR will be investigated for its utility to distinguish different lung cancer subtypes (adenocarcinoma, squamous cell, SCLC) and TNM stages. Through statistical modelling using multivariate techniques, variation in sputum spectra between cancer subtypes and TNM stages will be quantified for regression and classification model construction.

**Objective 4: Investigate the utility of FTIR for monitoring of concentration in sputum of an inhaled therapeutic, and interaction with sputum mucins, in spontaneous cystic fibrosis patient sputum.**

FTIR spectroscopic techniques will be applied to sputum samples collected from CF patients who were taking part in a clinical trial of a novel therapeutic drug, in order to evaluate interaction between the drug and patient's sputum. This will be achieved through statistical evaluation and comparison of IR spectra of the drug and

of CF-patient sputum. Alterations to the sputum spectrum caused by presence of the drug will be quantified using similarity and distance modelling.

# Chapter 2
# Materials & Methods

## 2.1 Sputum and Patient Data Collection

### 2.1.1 MEDLUNG – diagnostic study

Informed consent to provide a sputum sample was obtained as part of the Medlung observational study (loco-regional ethical committee approval 05/WMW01/75, study trial UKCRN ID 4682). Spontaneous sputum was collected from patients during visits to bronchoscopy clinics throughout the duration of the study at the following hospital sites: Great Western Hospital, Swindon; St. George's Hospital, London; Queen Elizabeth Hospital, Birmingham; New Cross Hospital, Wolverhampton. King's Mill Hospital, Nottingham; Milton Keynes Hospital, Milton Keynes; Chelsea and Westminster Hospital, London; Prince Philip Hospital, Llanelli; Bronglais Hospital, Aberystwyth. Patient details were collected regarding: patient medical history, drug history, age, biological sex, smoking status & pack-years. The patient's final clinical diagnosis, including staging, histological data and final source of diagnosis, where appropriate, was also recorded. All details were stored on a secure server at www.onedrive.com and accessed electronically. After providing full, informed consent, patients were instructed to provide a single sample of sputum by coughing-up from their airways. All sputum samples were immediately stored at -80°C in the hospital until transit to Swansea University on dry ice. Samples were stored in Swansea University at -80°C until FTIR analysis.

MEDLUNG recruited 1270 patients in total, over the period from June 2006 to December 2015. Of these 1270 patients, samples from 252 patients with confirmed lung cancer were available in Swansea University for the commencement of this work in 2014. An additional 132 COPD patients had also been recruited, of which 89 whose disease was classified as currently baseline, and 43 who were classified as currently exacerbating.

### 2.1.2 SPEDIC – longitudinal study

SPEDIC was a year-long study performed throughout the duration of this thesis, in collaboration with Cwm Taf University Health Board, and was designed to track the progress of COPD patients' disease (loco-regional ethical committee approval 15/LO/1703, study trial UKCRN ID 32064). The aim of this study was to evaluate the feasibility of using FTIR analysis to investigate COPD patient sputum samples in order to reliably predict an exacerbation prior to the exacerbation occurring. The study recruited patients who attended the respiratory clinic in Prince

Charles Hospital (Cwm Taf, Merthyr Tydfil) after referral from their GP due to worsening COPD symptoms, or after presenting in the emergency room with an exacerbation of COPD. In total, 55 COPD patients who attended these clinics throughout the study course consented to take part in the study.

Patients recruited into the study gave informed consent to give up to 5 spontaneous sputum samples per week, which would then be analysed by FTIR spectroscopy at Swansea University. All samples were stored at -20°C in the patients' homes until collection and subsequent long-term storage at -80°C. There were two collections per patient per 7-day period, with a maximum of 4 days between sputum production and collection for storage at -80°C. The effects of short-term storage of sputum at -20°C were unclear, however there was no feasible alternative given the frequency of sample production throughout the study. A review of the literature found that short-term storage of mucus-based samples at -20°C does not cause significant degradation of the mucin glycoprotein structure as evidenced by no significant differences in rheological properties of mucus stored at -20°C or -80°C (Gastaldi, Jardim, & King, 2000). Further studies have also shown that the other components of mucus are not significantly affected by short-term cold storage at the comparatively increased temperatures of 2-8°C. One such study focused on cervical mucus and compared the levels of various cytokines in samples which had been refrigerated for 8 hours to those which were snap-frozen on dry ice. They found the yield of cytokines was not significantly different between storage modes, with the exception of TNFα, type II interferon gamma (IFN-γ), and IL-1β which all showed an increased yield in the refrigerated samples compared to the snap-frozen samples (Panicker, Meadows, Lee, Nisenbaum, & Unger, 2007). Another study found that short-term refrigeration of sputum does not cause significant differences in the levels of detectable cytokines nor *Mycobacterium tuberculosis* cultures (Kolwijck *et al.*, 2013). Taken together, these studies all suggest that short-term storage of sputum at -20°C before long-term storage at -80°C does not cause alterations to the molecular structure of the mucin glycoproteins nor to the other constituent components of mucus.

Patients filled in a validated patient health questionnaire providing information about various aspects of their current health status, including cough, sputum production, and difficulty performing light, moderate and strenuous physical

activity with patients rating each variable on a scale of 1-6, with higher values associated with worse symptoms. Current exacerbation status was confirmed by the health care assistant assigned to each patient. The patient health status data was stored on secure server at www.zoho.com and assessed electronically.

SPEDIC also recruited 70 non-COPD chronic respiratory disease patients. A key question facing clinicians is the distinction between COPD and other respiratory diseases, such as asthma or bronchiectasis. These patients' sputum samples were analysed by FTIR and compared to COPD patient sputum spectra in order to identify specific markers to distinguish COPD from non-COPD respiratory disease.

### 2.1.3 Novel Therapeutics in Cystic Fibrosis Trial

CF patients were recruited into randomised, double-blind, crossover phase IIb studies (NCT02157922 & NCT02453789). For this work, samples from 17 CF patients were used. Patients were administered the drug treatment (OligoG 0.2% w/v) or a placebo treatment (lactose 0.2% w/v) via inhalation daily for 4 weeks before crossing over and receiving the other treatment – see appendix 1.4 for study phase allocations for each patient. Sputum samples were provided by the patients before any treatment as a screening sample, and at regular 14 day intervals during both treatment phases. Samples were frozen at -80°C, transported to Swansea University on dry ice and then subsequently stored at -80°C until FTIR analysis was performed. The aims were to ascertain if FTIR analysis could be employed to determine OligoG presence in sputum after treatment, and to provide evidence for any possible interaction between OligoG and sputum mucins.

### 2.1.3 Spirometry

Lung function tests were performed in accordance with the joint American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines (Miller *et al.*, 2005) by trained National Health Service (NHS) research staff. Spirometry was used to measure the patient's FEV and FVC values to provide a quantitative measure of lung function to be compared to FTIR spectra of sputum, as it has been shown to have a high diagnostic accuracy for COPD (Schneider *et al.*, 2009). However, care was taken to not over-emphasise the importance of spirometry results in the analysis of this work as it is known to have limited accuracy. One study found that spirometry was performed to an acceptable standard in only 60% of 153 patients, with inability to

meet “end-of-test criteria” being the most common cause of failure, followed by poor repeatability (Hegewald, Gallo, & Wilson, 2016).

## 2.2 FTIR Spectroscopy

### 2.2.0 Sample preparation

All raw sputum samples in Swansea University were stored at -80°C when not in use. Samples were allowed to slowly thaw and reach room temperature prior to spectral analysis. No other sample preparation was carried for FTIR analysis.

### 2.2.1 ATR-FTIR

ATR-FTIR spectroscopy was carried out using a Bruker Alpha, with the ATR diamond single reflection sampling module attachment (Bruker Optics), equipped with a deuterated triglycine sulphate (DTGS) detector. Prior to each measurement, the diamond sampling window was cleaned in 70% ethanol, and a background spectrum was acquired. 2µl of each sample was directly pipetted onto the sampling window and allowed to dry in air at room temperature. Once dry, sample spectra were generated, with each spectrum consisting of an average of 24 scans per sample at $4cm^{-1}$ resolution, over the spectral range of 4000-400$cm^{-1}$. After the spectrum was generated, the sample was rehydrated using $dH_2O$ to facilitate cleaning and then the ATR crystal was cleaned with 70% ethanol and blue roll. The instrument was controlled by Optics User Software (OPUS) version 7.5 (Bruker).

### 2.2.2 High-throughput Transmission FTIR

High-throughput-transmission FTIR (HT-FTIR) was carried using the Bruker Vertex 70 with the HTS-XT module attachment, equipped with a DTGS detector. A silicon 96-well plate (Bruker Optics) was used as the FTIR substrate, which was cleaned in 70% ethanol prior to sampling. One position on the plate was always left clear to enable background measurements to be taken. 2µl of sample was spotted onto the sampling plate and allowed to dry in air at room temperature. Spectra were acquired in transmission mode within the spectral range of 4000 to 400$cm^{-1}$, at a resolution of $4cm^{-1}$ and were the result of 32 scans.

### 2.2.3 FTIR Microspectroscopy

FTIR microspectroscopy was performed using a Bruker Vertex 70 with a Hyperion 2000 IR microscope attachment, in reflectance mode. The spectrometer was equipped with a KBr beamsplitter and a liquid nitrogen-cooled mercury cadmium telluride

(MCT) detector. The knife-edge aperture was set to 50 × 50μm. A single background spectrum of a gold standard plate was taken prior to spectrum acquisition. Sputum was pipetted (2μl) onto silicon plate and allowed to dry in air. 900 spectra were acquired over an area of approximately 1000x1000μm within the sputum spot.

#### 2.2.4 OPUS controlling software

The Bruker Spectrometers were controlled, and the parameters of each spectrum acquisition were set, by OPUS version 7.5 (Bruker Optics) as described above.

#### 2.2.5 IR Sphinx

##### 2.2.5.1 IR Sphinx - Crystal

The IR-Sphinx (Spectrolytic Gmbh) was used for ATR-VFIR measurements. The ATR crystal was cleaned with 70% ethanol and blue roll before all measurements. A background scan was taken and then 20μl of sputum was pipetted directly onto the ATR crystal, and allowed to air-dry. A small fan was used to expedite the drying. Once the sputum was dry, 3 consecutive spectra were taken 2 minutes apart at 200 scans each and modulation frequency of 8Hz. An average of these spectra was taken if they displayed a similarity score greater than 0.98. Similarity assessment was carried out using the in-built "SIX" algorithm in the Sphinx Suite Software.

##### 2.2.5.2 IR Sphinx – Sample Strip

20μl of sputum was pipetted directly onto surface of the sample strip, and allowed to air-dry. A small fan was used to expedite the drying. Once the sputum was dry, a background spectrum was first taken, and then the sample strip was applied to the crystal and using a custom-made clamp, was held in place throughout the scan. 3 consecutive spectra were taken 2 minutes apart at 200 scans each and as above, similarity scores calculated prior to averaging.

##### 2.2.5.3 SphinxSuite controlling software

The IR Sphinx Spectrometers were controlled, and the parameters of each spectrum acquisition were set, by SphinxSuite version 1.4.7 (Spectrolytic).

## 2.3 Spectral Processing

### 2.3.1 FTIR Spectra

Each FTIR spectrum was processed using the in-built algorithms in OPUS for baseline correction using the rubberband method, min-max or vector normalisation, second-derivative calculation (window size = 9, derivative degree =2) and peak picking.

### 2.3.2 ATR-VFIR Spectra

Acquired spectra were processed using the in-built algorithms in SphinxSuite. Spectra were converted to absorbance and a Savitzky-Golay filter (Window size = 13; Polynomial order = 5; Derivation degree = 0) was applied to smooth the data. An extended multiplicative scattering correction (EMSC) was applied to all spectra.

## 2.4 Statistical Methods

Statistical programming was carried out using the statistical programming environment, R (R Core Team, 2016). Data reduction techniques and statistical modelling were carried out using the in-built and custom-made functions. Normality and/or skewedness of data was tested for using data visualisation by histograms, with Q-Q plots and with the Shapiro-Wilk test for normality. Significance was assessed by either the parametric T-Tests or the non-parametric Wilcoxon Rank Sum Test, depending on the normality of the data. One way ANOVA was also performed to assess variance between and within patient groups.

### 2.4.1 Principle Component Analysis

Principle components analysis (PCA) is a data reduction and visualisation technique which is used to determine patterns of variation within a dataset. Principle components (PCs) are calculated based on the summation of vectors drawn from the sum of individual variable components within the dataset. PCs can then be plotted as vectors and visualised in geometric space (Figure 2-1).

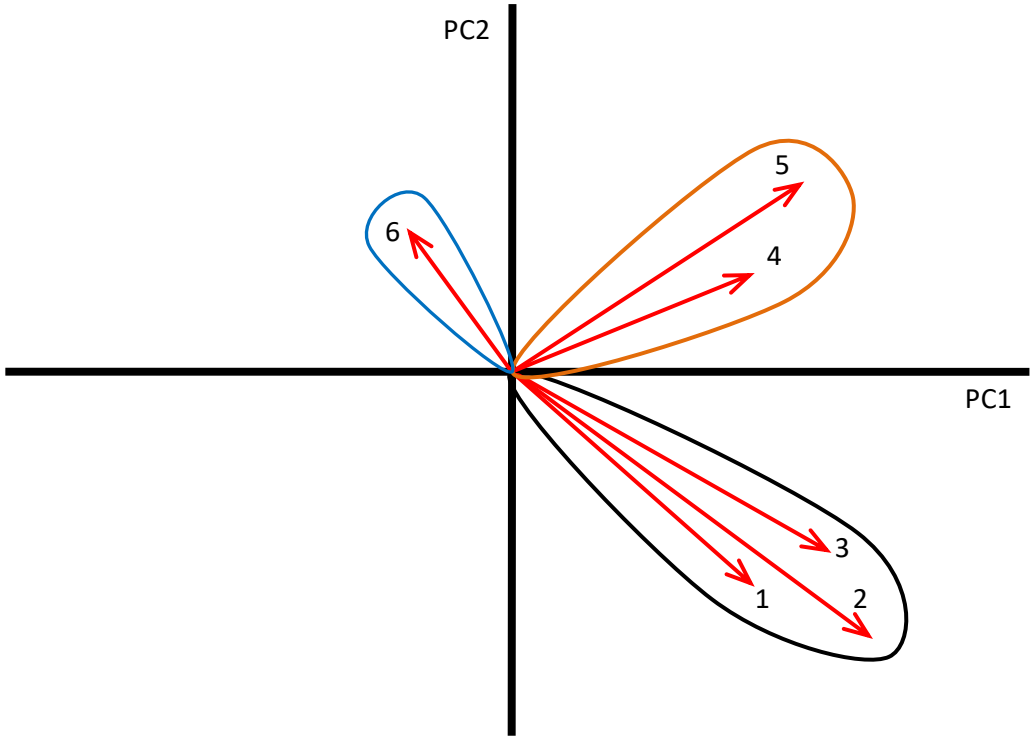


**Figure 2-1: Graphical representation of PCA, showing vectors 1-6 originating from the origin of PC1 and PC2. Vectors can be grouped according to the variance of the variables (black, orange and blue perimeters)**

PCA was used to reduce the spectral data into a set of variables (wavenumbers) and transform them into a Cartesian coordinate system that could be used to separate and display individual spectra in a scatterplot. The R "prcomp" (R Core Team, 2016) package was used to perform PCA on spectral data.

## 2.5 Incubation of Sputum with OligoG

Sputum samples from CF patients were incubated with 2% OligoG at a 1:10 (OligoG:sputum) concentration, for a final concentration of 0.2% OligoG in sputum, to mimic concentrations of OligoG used in previous studies (Hengzhuang *et al.*, 2016; Pritchard *et al.*, 2016). Incubations were performed at 37°C for 30 minutes prior to FTIR analysis.

## 2.6 OligoG, DNA and Calcium Ion Interaction

OligoG was incubated with 0mM (control), 1mM and 5mM $Ca^{2+}$ (Cole-Parmer) ions, and 0mM (control), 1mM and 10mM whole human DNA (Promega) for 30 minutes at 37°C. Table 2- below shows all final dilutions of OligoG, DNA, and $Ca^{2+}$ ions made up prior to incubation and subsequent high-throughput T-FTIR analysis, as described earlier in Section 2.2.2.

| Variable Modified | Final Concentrations | | |
|---|---|---|---|
| **DNA** | 2%OligoG + 0mM DNA + 0mM $Ca^{2+}$ (Control) | 2%OligoG + 1mM DNA + 0mM $Ca^{2+}$ | 2%OligoG + 10mM DNA + 0mM $Ca^{2+}$ |
| **$Ca^{2+}$** | | 2%OligoG + 0mM DNA + 1mM $Ca^{2+}$ | 2%OligoG + 0mM DNA + 5mM $Ca^{2+}$ |
| **$Ca^{2+}$ & DNA** | | 2%OligoG + 1mM DNA + 1mM $Ca^{2+}$ | 2%OligoG + 10mM DNA + 5mM $Ca^{2+}$ |

**Table 2-1: Concentrations of DNA and $Ca^{2+}$ ions incubated with 2% OligoG prior to FTIR analysis**

# Chapter 3
# Investigations of Sputum Mucins from Chronic Obstructive Pulmonary Disease Patients to Predict Exacerbation by FTIR

## 3.1 Introduction

### 3.1.1 COPD, Mucus and Mucin

COPD is a progressive and life-limiting condition of the respiratory system, characterised by moderate to severe airflow limitation, excess mucus and sputum production, and cough. The patient's airflow is limited by chronic airway obstruction by increased mucin production leading to an increased mucus load. Mucus clearance via MCC activity is also reduced. COPD is primarily caused by smoking, however environmental and occupational exposure to chemicals and other pollutants are also significant contributors towards COPD progression (Health and Safety Executive, 2014).

COPD patients frequently undergo episodic worsening of their disease, termed exacerbations. Exacerbations of COPD are defined as an event in the natural course of the disease that are characterised by an acute, non-fully reversible worsening of the patient's baseline dyspnoea, cough, and/or sputum production beyond day-to-day variations, and are sufficient to warrant a change in management (Papi *et al.*, 2006).

The state of health of a COPD patient is heavily influenced by the frequency of acute exacerbations, with those being prone to more frequent exacerbations undergoing more hospital admissions, and generally having a marked decrease in their QOL (Seemungal *et al.*, 1998). Currently there is no way to predict exacerbation occurrence, and exacerbations are usually self-reported. This is hypothesised to contribute to the phenomenon that approximately 50% of COPD exacerbations are unreported to the patient's clinician and therefore it can be assumed that approximately 50% are also untreated (Seemungal, Hurst, & Wedzicha, 2009). This represents a significant issue for management of COPD by patients and by clinicians, as due to the progressive nature of COPD, after each exacerbation the patient's symptoms worsen and their QOL decreases. This is highlighted by the finding that patients who had a higher exacerbation reporting rate had an improved quality of life compared to patients who did not report and subsequently did not receive treatment (Wilkinson, Donaldson, Hurst, Seemungal, & Wedzicha, 2004).

COPD phenotyping into seven distinct phenotypes has been shown to be effective for defining patients, leading towards clinically meaningful outcomes (Mirza & Benzo, 2017). The identified phenotypes are described in detail in Table 1-2, but

briefly are 1) asthma-COPD overlap, 2) frequent exacerbator, 3) upper lobe-predominant, 4) rapid decliner, 5) comorbid COPD, 6) physical frailty, and 7) emotional frailty. These phenotypes each have distinct causes and each has practical implications for the care of the patient. For example, comorbid COPD patients who have high BMI, congestive heart failure, and/or ischemic heart disease show higher levels of dyspnoea, lower QOL, increased health care utilization and increased mortality risk compared to COPD patients with similar airflow limitation but no comorbidity.

### 3.1.2 Fourier-Transform Infrared Spectroscopy

FTIR spectroscopy is a vibrational spectroscopy technique capable of rapidly identifying molecular structures in samples of interest through measurements of IR absorbance by molecular bonds. Every molecular bond has vibrational energy, which is increased through absorption of certain IR wavelengths, specific to that molecular bond. By targeting a sample with IR light composed of many wavelengths a spectrum of absorbance can be generated, corresponding to the molecular composition of the sample.

FTIR spectroscopy has been successfully applied in the analysis of biofluids from patients of many different diseases. It has been shown to be effective for detection of DNA and mannose-6-phosphase, a known breast and prostate cancer marker, in saliva from smokers (Rodrigues *et al.*, 2017). FTIR spectroscopy has also been shown to be capable of classifying and staging many different cancers including breast (Backhaus *et al.*, 2010), lung (Bird *et al.*, 2012; Paul D. Lewis *et al.*, 2010), oropharyngeal (Menzies *et al.*, 2014), and brain cancers (Hands *et al.*, 2016). This shows that FTIR is a powerful tool for determining subtle molecular changes in complex biofluids, which are regularly comprised of a mixture of proteins, sugars, lipids, cellular debris, and DNA.

FTIR has also been shown to be capable of determining protein secondary structure, through determination of α-helices, β-sheet, and random coil structures present in bovine submaxillary mucin (BSM), a mucin protein which is homologous to MUC5AC and MUC5B in terms of primary and secondary structures (S. P. Lewis *et al.*, 2013).

Whiteman *et al.* performed a small preliminary study into the use of FTIR spectroscopy as a diagnostic tool for detection of biomarkers of exacerbation in

COPD patients' sputum. It was found that the infra-red spectra of sputum collected from COPD patients and healthy volunteers were significantly different in the amide II and glycogen-rich regions, but not the amide A region despite COPD patients having more peaks over a wider range in this region (Whiteman *et al.*, 2008).

The main constituents of sputum are the high molecular weight mucin glycoproteins MUC5AC and MUC5B, which are heavily glycosylated with O-glycan structures covalently bonded to the polypeptide chain (Rose & Voynow, 2006; Wickström *et al.*, 1998). Expression of MUC5AC is known to be increased in COPD patients with stable disease, compared to smokers without respiratory disease and healthy non-smokers (Caramori *et al.*, 2009). COPD exacerbations are known to be related to inflammatory processes within the lung (Wedzicha & Seemungal, 2007). Additionally, levels of pro-inflammatory mediators, such as neutrophil elastase, hyaluronic acid metabolites, LPS, IL-8 and ROS have all been demonstrated to be linked to increased mucin expression (Bautista *et al.*, 2009; Bozinovski, Anthony, & Vlahos, 2014; Papakonstantinou *et al.*, 2015; J.-S. Song, Cho, Yoon, Moon, & Park, 2005). Expression of certain inflammatory mediators has also been shown to increase the expression of specific glycosyltransferases, which can alter the glycosylation patterns on the sputum mucins. For example, TNFα exposure can increase sialyltransferase activity in the bronchial mucosa (Delmotte *et al.*, 2002). Respiratory infections have been linked to elevated levels of sialylation and sulphation of mucin glycan structures in non-cancer respiratory disease patients (Davril *et al.*, 1999; Lo-Guidice *et al.*, 1997; van Halbeek *et al.*, 1994), again showing how response to inflammatory mediators and stimulants can change the glycan structure in diseased airway mucins. Evidence has also been demonstrated that alterations to the levels of specific mucin glycans in mucus can alter the viscosity of the mucus. Mediation of fucosylation and sialylation of glycan structures in MUC5AC by carbocysteine has been shown to reduce the viscosity of airway mucus (Ishibashi, Takayama, Inouye, & Taniguchi, 2010).

FTIR spectroscopy has previously identified wavenumbers uniquely associated with absorption peaks corresponding to specific individual glycan moieties, including fucose, galactose, GlcNAc, GalNAc, and sialic acid. It was also used to identify unique FTIR absorption peaks associated with Lewis X antigens, and sialylation and sulphation of these glycan structures (A. T. Lewis *et al.*, 2013).

Work previously carried out by the Swansea University Respiratory Diagnostics Group, as part of Aaran Lewis' PhD thesis (Figure 3-1), utilised a small cohort of COPD patient sputum samples (~40 total) and suggested that FTIR could be utilised to detect COPD exacerbation. The work presented in this chapter seeks to expand on this work, by increasing the number of patients tested and examining the patient groups more closely to establish if differences between COPD phenotypes can be identified through FTIR sputum analysis.

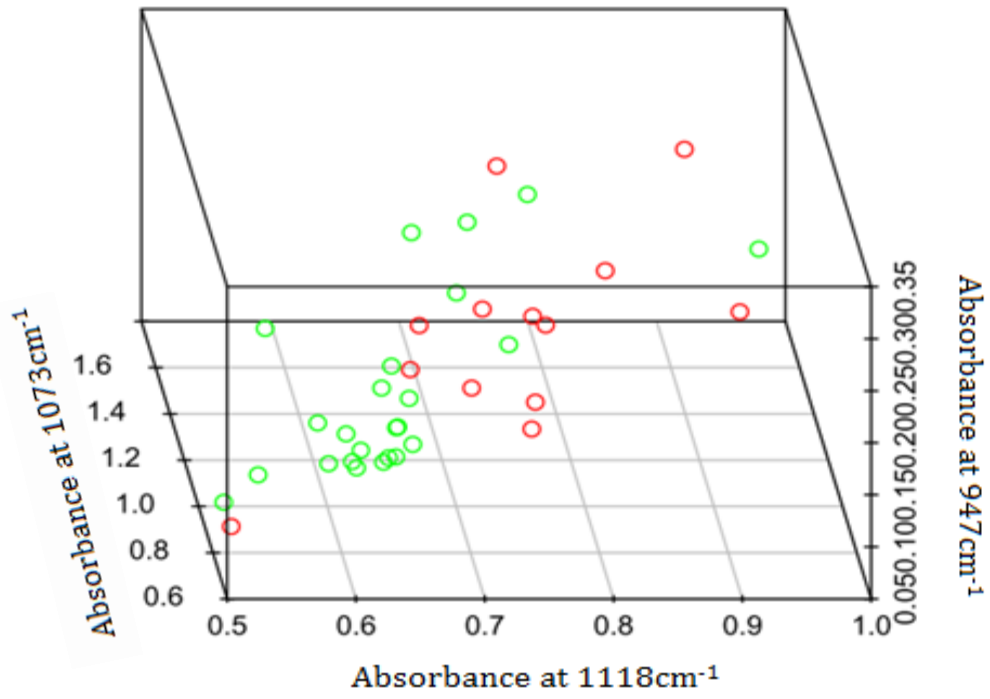


**Figure 3-1: Three-dimensional scatterplot from Aaran Lewis' PhD Thesis, showing separation of COPD baseline (green) from COPD exacerbation (red) patients using absorbance at 3 prominent wavenumbers; 1703, 1118 and 947cm$^{-1}$.**

### 3.1.3 Chapter Aims

The aims of this chapter are to investigate glycosylation patterns in sputum mucins using FTIR, in order to distinguish exacerbation of COPD from baseline, and to address the practical issues of performing FTIR analysis on viscous biofluids. This chapter evaluates FTIR as a tool for exacerbation detection, through monitoring of sputum mucin glycosylation patterns by applying FTIR analysis to raw sputum samples obtained from COPD patients. Through careful interpretation of IR spectra from sputum samples, differences observed in sample spectra derived from patients at baseline or exacerbation can be related to changes in the molecular structure of the

sputum. Models that distinguish exacerbation sample spectra from baseline COPD sample spectra will be constructed.

To achieve these goals, FTIR analysis was applied to raw sputum samples obtained from patients with a diagnosis of COPD who were either confirmed by a clinician to be currently undergoing an exacerbation, or to be at baseline. Spectral data was then interrogated by statistical and multivariate analysis to identify regions of the spectrum able to differentiate between the disease groups.

A longitudinal study was also carried out, analysing COPD patient sputum samples throughout the course of one year. This was done to show the potential of FTIR spectroscopy to detect changes to the glycosylation patterns of sputum mucins in the days preceding an exacerbation. FTIR spectra of sputum samples collected during the study will be analysed for specific patterns which can be used in the creation of generalised predictive models for exacerbation.

## 3.2 Methods

### 3.2.1 Patient Recruitment and Sample Classification

As described in Chapter 2, section 2.1.1, COPD patient sputum samples (n=132 were obtained from patients recruited through the MEDLUNG study. Patient samples were classified as "baseline" (n=89) or "exacerbation" (n=43) depending on the clinical diagnosis established at recruitment. Background clinical information of each patient was provided by the recruiting centre. Information provided included age, biological sex, smoking status (current, former, or never) and pack-years, occupation, lung function, medical and drug history, current infection status, and exposure to asbestos.

COPD patients were also recruited through the SPEDIC study, which was longitudinal study with a focus on using FTIR to identify spectral markers for exacerbation prediction. Patients recruited to the study provided 5 sputum samples per week whilst their day-to-day health and current exacerbation status was recorded. FTIR analysis was carried out on these samples as described in Chapter 2 Section 2.2 and the spectra compared to the patient's health data.

SPEDIC sputum samples were classified as either being "baseline", "exacerbation", "pre-exacerbation" or "post-exacerbation" (high-risk) according to the daily health data collected during the study. Sputum from exacerbation days were

identified through the reporting from patients and clinicians during the study. Samples belonging to the high-risk category were identified as being collected less than 8 weeks after the last reported exacerbation event, as patients within this period have an elevated risk of a secondary exacerbation (Hurst *et al.*, 2009). Pre-exacerbation samples were identified as those being collected 7 or fewer days prior the first reported day of an exacerbation; 7 days was chosen as the cut-off for the pre-exacerbation phase after consulting with respiratory clinicians identified 7 days as a suitable time period for an "early warning system" of COPD exacerbation.

SPEDIC also recruited 70 non-COPD chronic respiratory disease patients. A key question facing clinicians is the distinction between COPD and other respiratory diseases, such as asthma or bronchiectasis. These patients' sputum samples were analysed by FTIR and compared to COPD patient sputum spectra in order to identify specific markers to distinguish COPD from non-COPD respiratory disease.

#### 3.2.1.1 Phenotyping Patients

COPD patients were grouped into phenotypes using the patient's clinical information, collected at time of recruitment. COPD phenotypes were classified using the following parameters, where information was available: 1) smoker/non-smoker, 2) COPD stage by FEV1 value only, as the FVC/FEV1 ratio was not available for any patients, and 3) presence of comorbidities which may influence the progression of COPD.

### 3.2.2 FTIR Spectroscopy

FTIR spectroscopy was carried out in accordance to the protocols outlined in Chapter 2, section 2.2.1 & 2.2.2. Spectral processing was performed as described in Chapter 2, section 2.3.

### 3.2.3 Spectral Analysis and Model Building

IR peak identification and characterisation was performed using the peak picking algorithm in OPUS (Bruker Optics) set to a 10% threshold. Peak widths were measured using the same tool and are displayed as full width at half maximum (FWHM).

FTIR spectra of COPD patient sputum from each classification group were analysed for significant differences using a Wilcox rank-sum test to identify IR regions which could be used to distinguish baseline from exacerbation samples.

Multivariate statistical modelling using PCA was applied to visualise the data points and how baseline and exacerbation samples cluster in a two-dimensional scatterplot. Regression models were constructed based on these PCA scatterplots for sensitivity and specificity calculations.

### 3.2.4 Calculation of IR Spectral Ratios

Ratios of absorbencies at specific biologically significant wavenumbers were calculated to generate a single value, or quotient, to represent the spectrum. This quotient ("Q-value") calculation method was used as a data reduction technique to enable correlation of a whole spectrum to current patient health data.

## 3.3 Results

First, COPD patient sputum samples were collected from patients attending bronchoscopy clinics and emergency departments in hospitals from across the UK. These patients were part of the MEDLUNG study, and were classified at recruitment as being either "baseline COPD" or "exacerbation of COPD". These samples were analysed by FTIR spectroscopy for sensitive and specific markers which could be used to distinguish baseline COPD from exacerbations of COPD.

### 3.3.1 FTIR spectrum generation and COPD sputum characterisation

FTIR spectra were generated in triplicate for COPD baseline (n=89) and COPD exacerbation (n=43) sputum samples. These spectra were processed by vector-normalisation and baseline correction to allow comparisons between sputum samples from different patients, reducing impact from variations introduced during sampling and spectral acquisition.

Average raw absorbance spectra of the COPD baseline and exacerbation patient cohorts show a high degree of similarity, although differences in raw absorption intensity can be observed. The average COPD exacerbation spectrum, shown in Figure 3-2, display a general increase of absorption at the major glycogen-associated peaks (between 1200 and 900cm$^{-1}$), compared to the average baseline spectrum. The difference spectrum (Figure 3-3) identifies regions of the average spectra where the differences between exacerbation and baseline are greatest, with large peaks identifiable at well characterised glycan-associated wavenumbers; 1586, 1458, 1419, 1106, 1030, 990 and 923cm$^{-1}$ (Kačuráková & Mathlouthi, 1996; Khajehpour, Dashnau, & Vanderkooi, 2006; A. T. Lewis *et al.*, 2013; Wiercigroch *et al.*, 2017). This suggests an increase in overall glycan content in COPD exacerbation sputum. This

increase in sugar may be due to an overall increase in glycosylation of the mucin glycoproteins.


**Figure 3-2: Min-max normalised absorbance ATR-FTIR average spectra of the COPD baseline cohort (blue) and COPD exacerbation cohort (red), from 1800-900cm$^{-1}$**

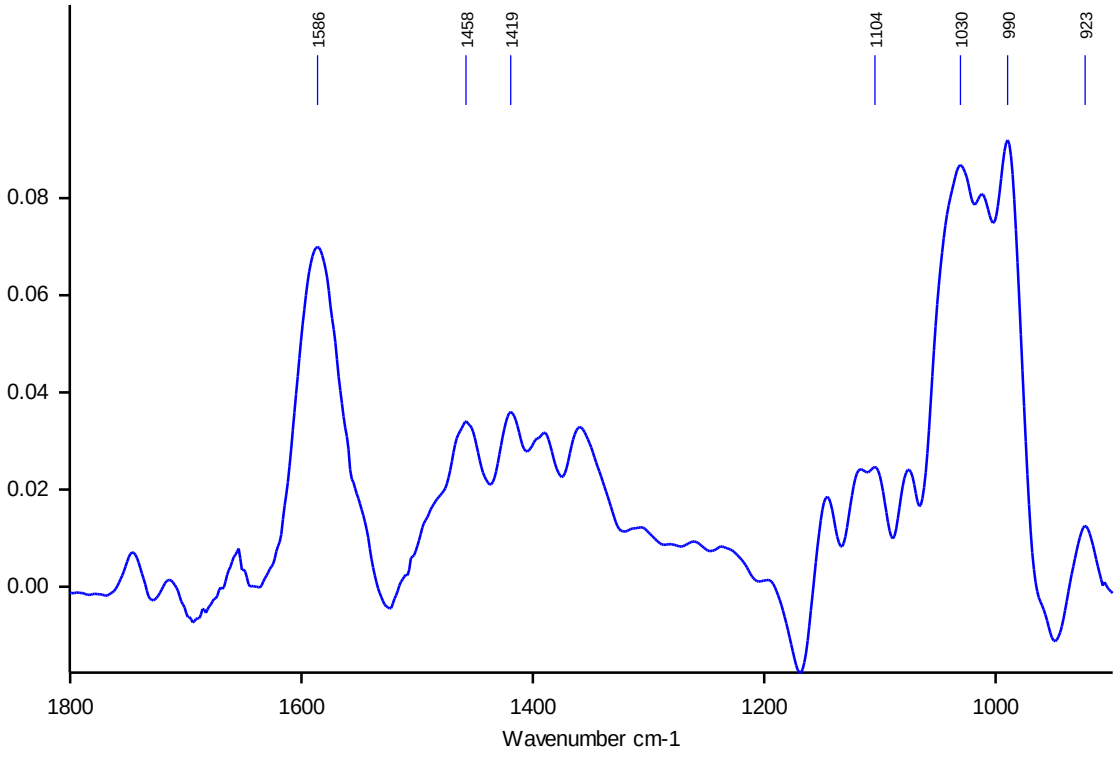


**Figure 3-3: Average difference spectrum of min-max normalised baseline and exacerbation sputum absorbance spectra with peak positions indicating regions of the**

**spectrum where average exacerbation absorbance is greater than the average baseline absorbance.**

Using wavenumbers which have been associated with specific mucin glycan structures (Kačuráková & Mathlouthi, 1996; Khajehpour *et al.*, 2006; A. T. Lewis *et al.*, 2013; Wiercigroch *et al.*, 2017), and mucin protein secondary structure (S. P. Lewis *et al.*, 2013), it is possible to characterise the COPD sputum FTIR spectrum and annotate with proposed structures. Shown in Figure 3-4 is an average FTIR spectrum of all baseline COPD patient sputum samples, annotated with proposed molecular structures for specific absorption bands.

The amide I peak is centred at 1636cm$^{-1}$, suggesting that the major constituent of the protein secondary structure is β-sheets (S. P. Lewis *et al.*, 2013), followed by the amide II peak found at 1544cm$^{-1}$ corresponding to N-H bonds, typically found within the protein chain. Absorbance bands around 1586cm$^{-1}$ are associated with amino-glycans, such as GlcNAc, GalNAc and sialic acid (Khajehpour *et al.*, 2006). Sulphated glycan residues produce absorbance bands at 1240cm$^{-1}$ (S=O), 1153cm$^{-1}$ (S-O), 1116cm$^{-1}$ (S=O), and 990cm$^{-1}$ (S-O) (A. T. Lewis *et al.*, 2013); indicated by the orange lines on Figure 3-4. Sialylation of glycan structures produces absorbance bands at 1161cm$^{-1}$, corresponding to C-N bonds in the acetyl functional group, and at 1130cm$^{-1}$, corresponding to C-O bonds located in the hexose ring (Khajehpour *et al.*, 2006); indicated by the red lines in Figure 3-4. Absorbance bands around 1456 and 1418cm$^{-1}$ are associated with C-$H_2$ and H-C-OH bonds, respectively (Kačuráková & Mathlouthi, 1996; Wiercigroch *et al.*, 2017). It is also possible to identify particular glycan linkages indicating increased fucosylation of the glycan structures, such as the Fuc-(α1-3)-GlcNAc linkage which produces an absorption band at 1020cm$^{-1}$; indicated by the purple line in Figure 3-4. If a glycan structure contains a second fucose residue in a Fuc-(α1-2)-Gal configuration, peaks shifts towards 1164cm$^{-1}$ can be observed (A. T. Lewis *et al.*, 2013).

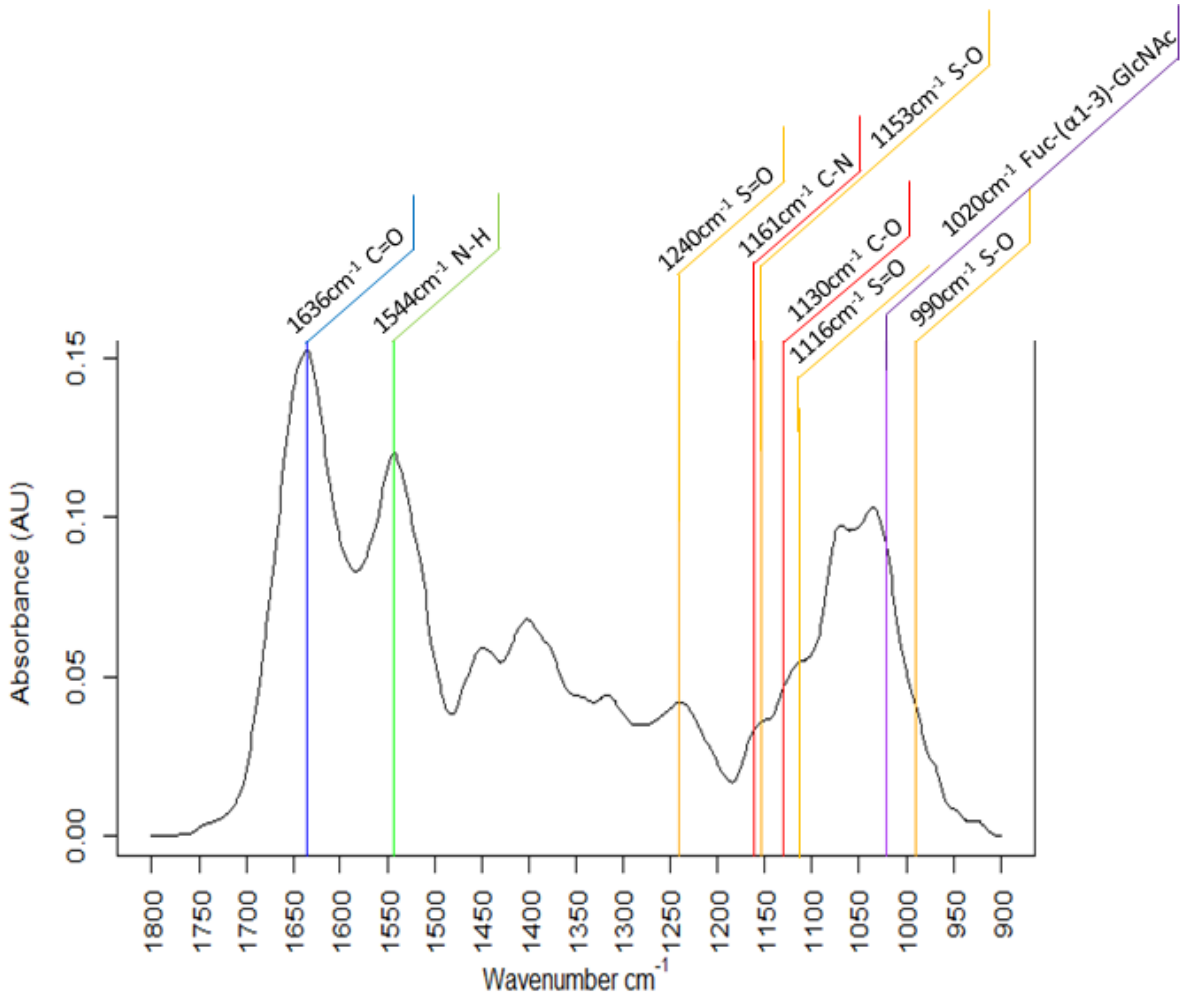


**Figure 3-4: Average vector-normalised, baseline-corrected FTIR spectrum of all COPD patient sputum samples. Wavenumbers corresponding to absorbance bands associated with sialylated (red), sulphated (orange), and fucosylated (purple) glycan residues are shown. The peak positions at 1636cm$^{-1}$ and 1544cm$^{-1}$ corresponding to amide I (blue) and amide II (green), respectively, are also indicated.**

Absorption intensity and peak positions were examined around these absorbance bands, for statistically significant differences between COPD baseline and exacerbation sputa.

### 3.3.2 Distinction of Baseline from Exacerbation

#### 3.3.2.1 Absorbance spectra

The absorbance spectra were examined for differences between absorbance intensities at wavenumbers associated with molecular structures hypothesised to be associated with COPD exacerbation. However, significance testing by the Wilcoxon rank sum identified no wavenumbers which showed a statistically significant ($p < 0.05$) difference in absorbance when comparing COPD baseline to exacerbation sputum spectra.

First, the absorbance spectra were examined in the same manner as in Aaran Lewis's previous work, in order to replicate the findings on the larger dataset. Unfortunately it was not possible to replicate the previous findings using the same methods (Figures 3-5 and 3-6). The methodologies used to generate the data in both cases were identical, thus reducing the possibility of operator error being the cause of the inability to reproduce the findings. It is more likely that the irreproducibility is due to the previous work being carried out on a comparatively small dataset (n = 24 baseline, and n = 13 exacerbation samples), which may have caused any observed trends or differences between patient groups to be due to statistical chance rather than statistically significant differences in sputum structure.

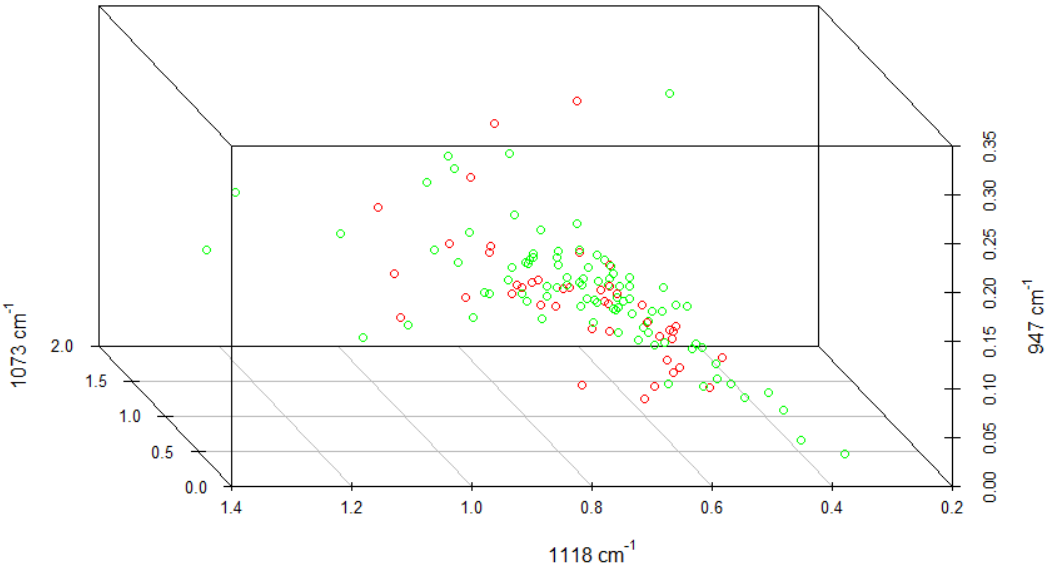


**Figure 3-5: Three-dimensional scatterplot of baseline-corrected, min-max normalised absorbance values at 1073, 1118 and 947cm$^{-1}$.**

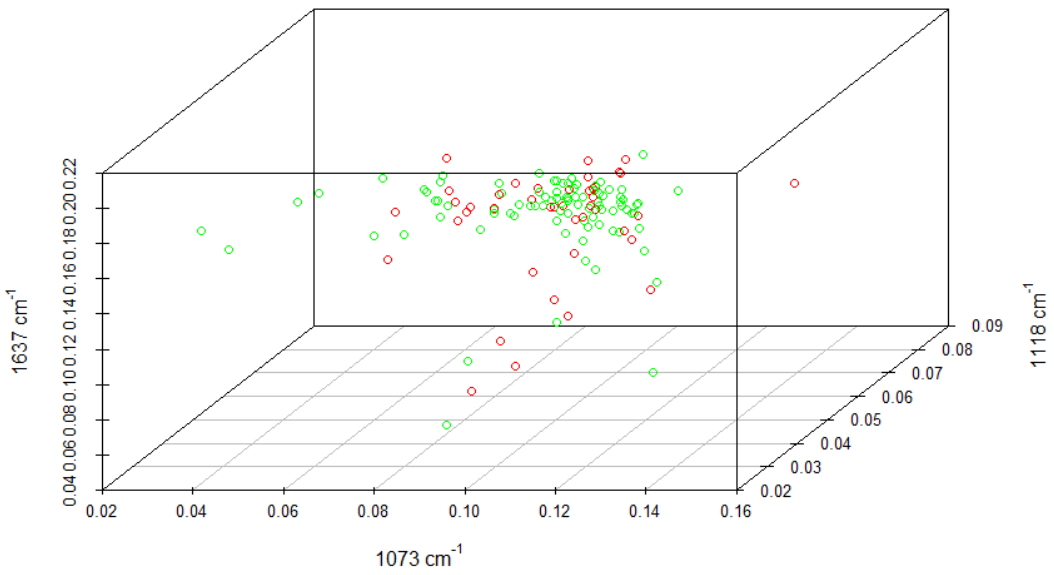


**Figure 3-6: Three-dimensional scatterplot of baseline-corrected, min-max normalised absorbance values at 1073, 1118 and 1637cm$^{-1}$.**

PCA was carried out on the dataset in an attempt to account for the inter-sample variation. To ensure that the PCA would be focussed in spectral regions with the most chance of discriminating baseline from exacerbation, the peaks identified in the difference spectrum in Figure 3-3 were used in the PCA algorithm.

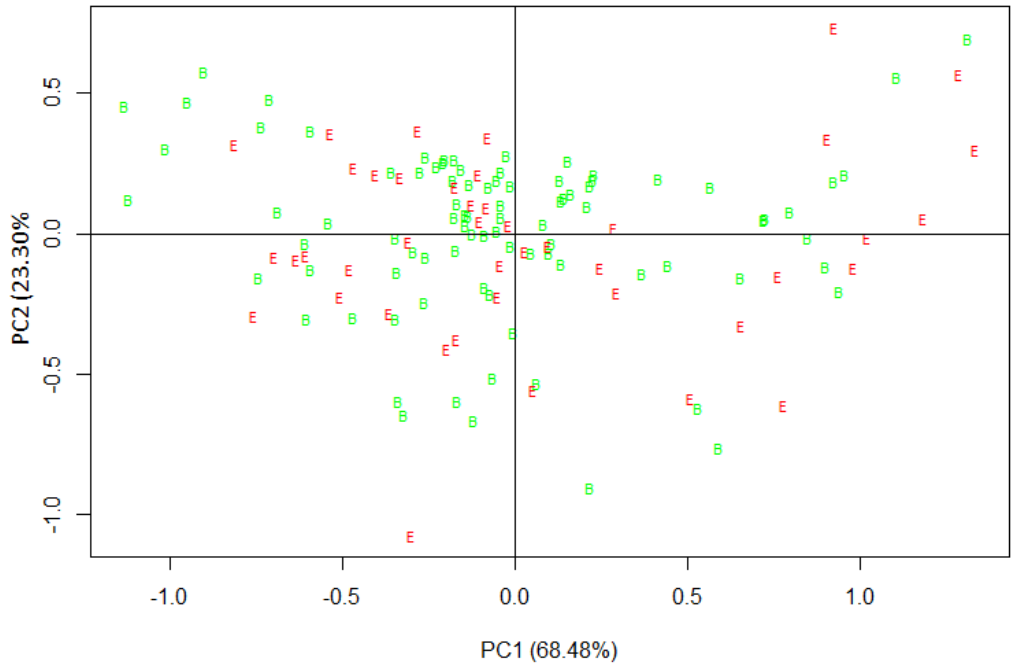


**Figure 3-7: Two-dimensional PCA scatterplot showing the variance explained by PC1 and PC2. Baseline (green) and exacerbation (red) spectra cluster together with no separation observed.**

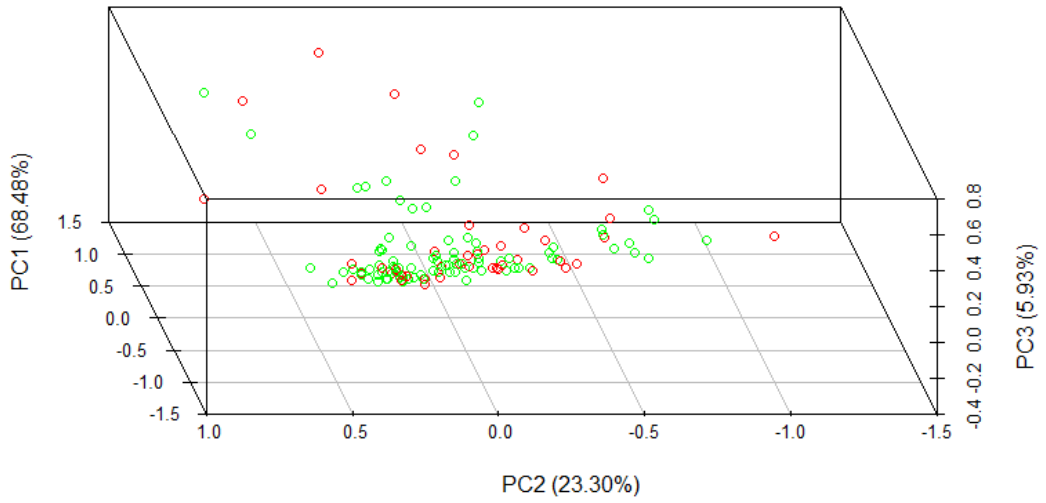


**Figure 3-8: Three-dimensional scatterplot showing how baseline (green) and exacerbation (red) spectra cluster together based on PC1, PC2 and PC3.**

PCA analysis on the absorbance spectra showed poor separation of baseline and exacerbation patient samples (Figure 3-7 & Figure 3-8). This suggests that a higher resolution method for determining differences between absorbance bands, such as second-derivative spectrum analysis, may be required to distinguish the patient groups from one another.

### 3.3.2.2 Second Derivative Spectra

Second derivative spectra were calculated to increase the resolution of peak detection across the spectrum, by calculating a rate of change within a defined window of the absorbance spectrum (Figure 3-9). A second-derivative peak is identified when the rate of change within the absorbance spectrum changes direction, and can be used to more detect absorbance distributions which comprise large, broad peaks such as those observed in the glycogen-rich region.

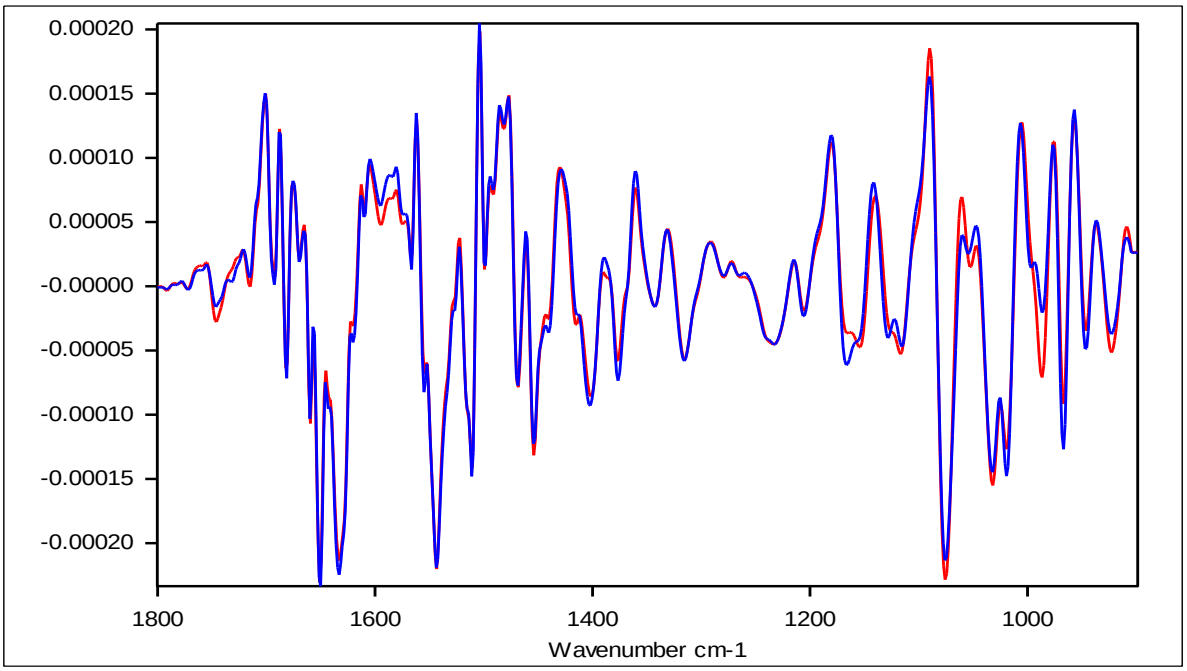


**Figure 3-9: Average second-derivative spectra of COPD baseline (blue) and exacerbation (red) sputum samples, from 1800-900cm$^{-1}$. Spectra show a high degree of similarity, but small variations can be seen in the glycogen-rich region (1200-900cm$^{-1}$)**

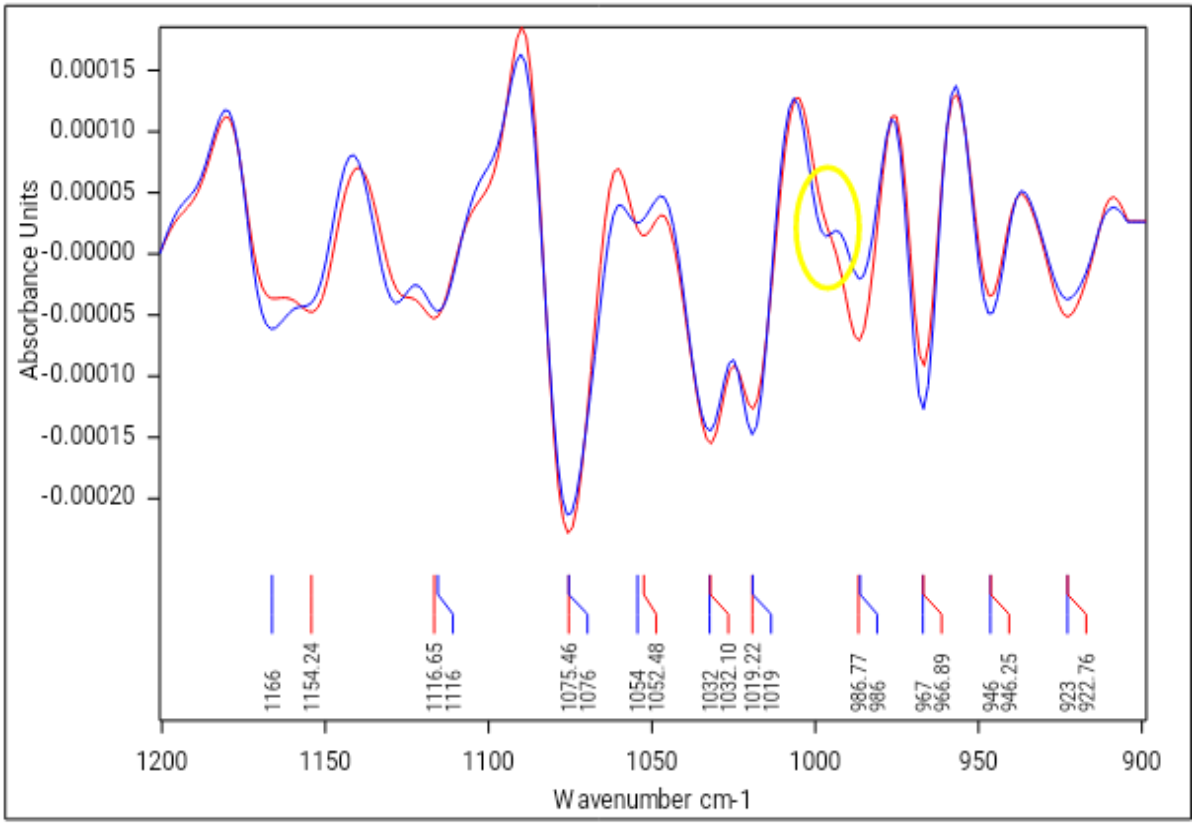


**Figure 3-10: Average second-derivative spectra of COPD baseline (blue) and exacerbation (red) sputum samples, focussed on 1200-900cm$^{-1}$ and major peak positions highlighted for clarity. An apparent peak shift can be observed from around 1166cm$^{-1}$ in the baseline spectrum to around 1154cm$^{-1}$ in the exacerbation spectrum. Another probable candidate for peak shift is from around 1054cm$^{-1}$ to 1052cm$^{-1}$ from baseline to exacerbation. An increase in second-derivative absorption can be seen at around 986cm$^{-1}$ in the exacerbation spectrum compared to the baseline spectrum, along with a broadening of the peak to obscure the secondary peak at around 998cm$^{-1}$ (yellow circle).**

The second-derivative spectra show an extremely high degree of similarity across the spectral region 1800-900cm$^{-1}$, although some subtle differences in peak position and absorbance can be observed (Figure 3-10). To highlight further possible regions of interest, a spectral subtraction was performed, subtracting the exacerbation spectrum from the baseline spectrum (Figure 3-11).

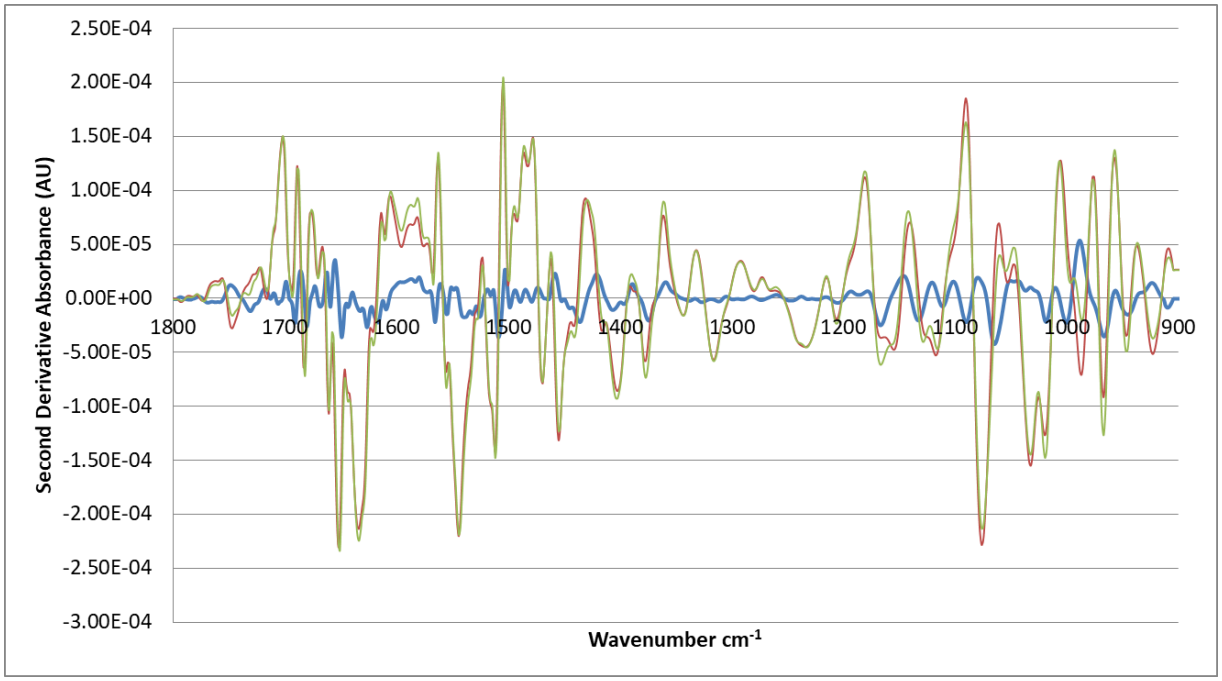


**Figure 3-11: Average second-derivative spectra from 1800-900cm$^{-1}$ of COPD baseline (green) and exacerbation (red) sputum samples, with the difference spectrum (blue) shown to highlight probable spectral regions which may differentiate COPD baseline from exacerbation.**

Because the differences between the average spectra are small, the subtractive spectrum displays small peaks which are difficult to differentiate from surrounding noise, especially in the amide I and amide II regions. However, examining the glycogen-rich region reveals some broader subtractive peaks which may be indicative of differences between the populations (Figure 3-12). These regions are centred around 1066cm$^{-1}$ (purple box), 1020cm$^{-1}$ (yellow box), and 988cm$^{-1}$ (black box).

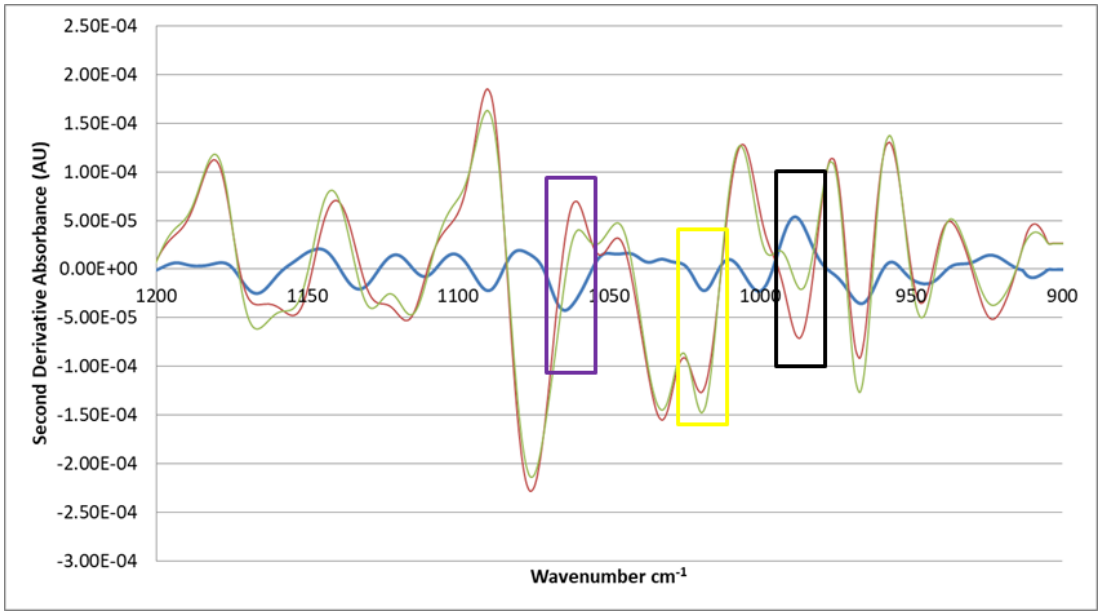


**Figure 3-12: Average second-derivative spectra from 1200-900cm$^{-1}$ of COPD baseline (green) and exacerbation (red) sputum samples, with the difference spectrum (blue) shown to highlight probable spectral regions which may differentiate COPD baseline from exacerbation. Regions of interest (purple, yellow and black boxes) have been highlighted as possible regions of difference between COPD baseline and exacerbation spectra.**

The average second-derivative spectra of the baseline and exacerbation sample spectra are highly similar, with an extremely high Spearman's Rho for correlation of 0.9853, indicating a very low level of variation between the sample groups. Significance testing using the non-parametric Wilcoxon Rank Sum of the second-derivative absorbance values indicated a panel of 18 non-contiguous wavenumbers which showed a significant difference ($p < 0.05$) in second-derivative absorbance between COPD baseline and exacerbation samples. The distributions of these wavenumbers are shown in the boxplots below (Figure 3-13 to Figure 3-30).

All of the wavenumbers examined above show a statistically significant difference at the 95% confidence level between mean second-derivative absorbencies for each patient cohort; however there was a large overlap between each group for each wavenumber. Multiple hypothesis testing was carried out, which indicated a new α-level of $0.05/636 = 7.862 * 10^{-5}$. None of the calculated p-values were below this new α-level, so it was deemed that second-derivative absorbance values of key wavenumbers could not show a statistically significant difference between baseline and exacerbating COPD patients.

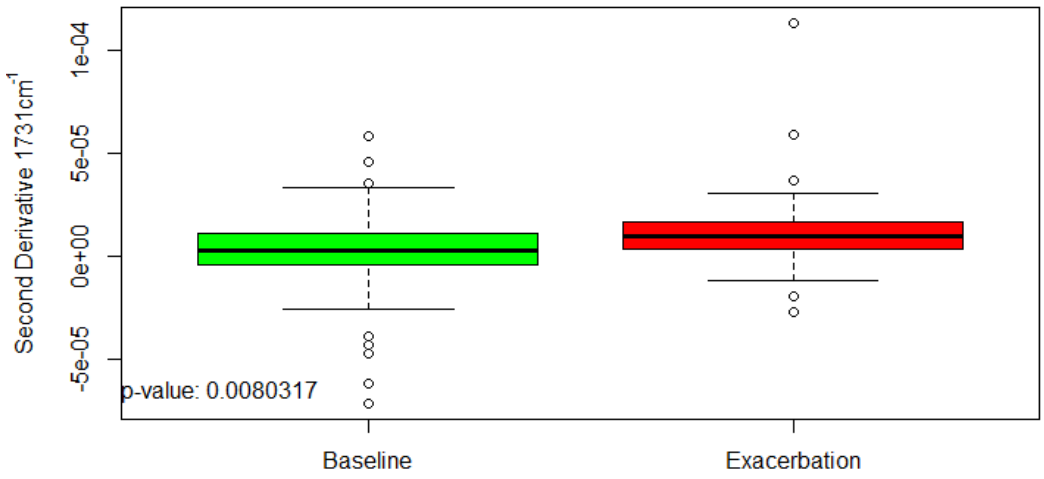


**Figure 3-13: Distribution of absorbance values at 1731cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

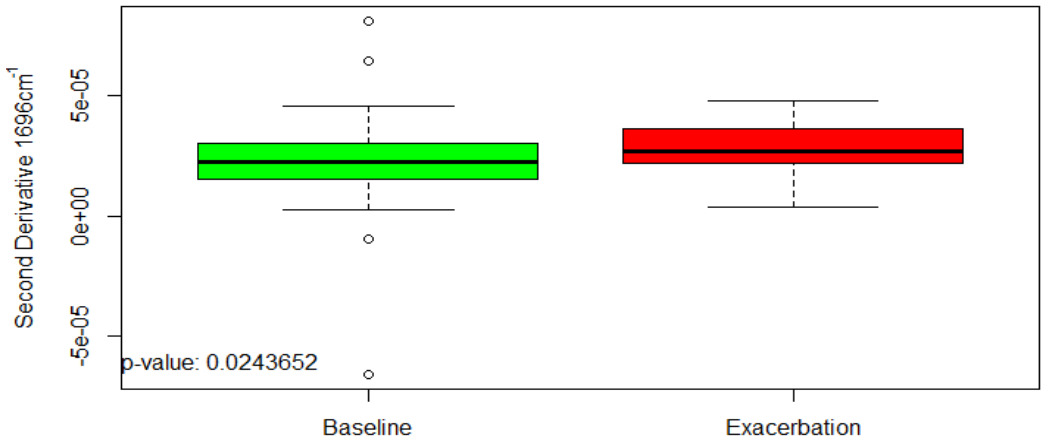


**Figure 3-14: Distribution of absorbance values at 1696cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

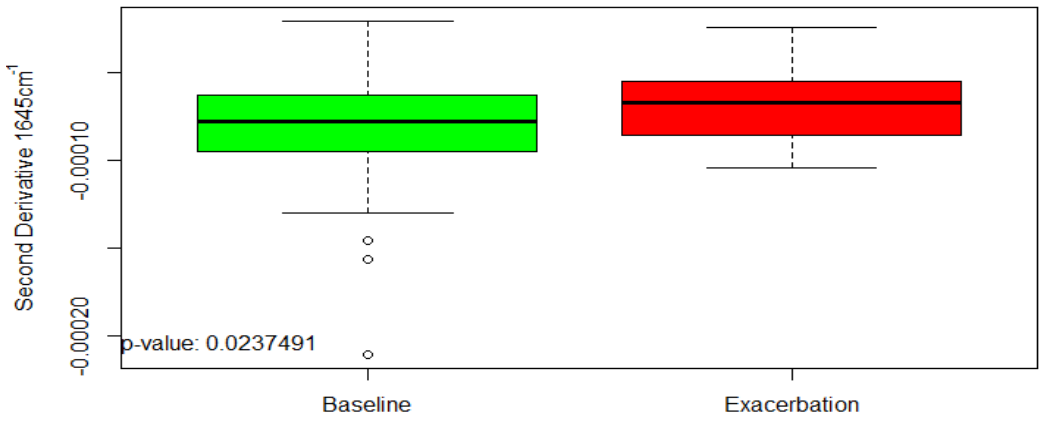


**Figure 3-15: Distribution of absorbance values at 1645cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

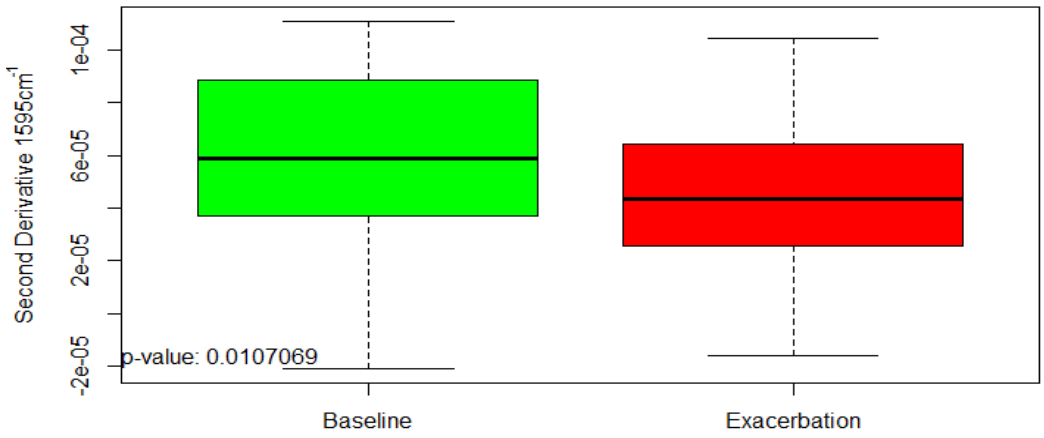


**Figure 3-16: Distribution of absorbance values at 1595cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

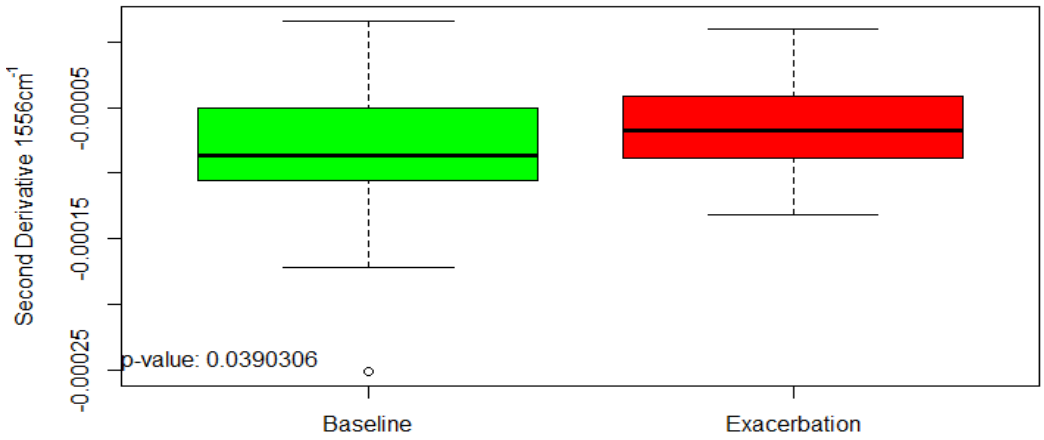


**Figure 3-17: Distribution of absorbance values at 1556cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

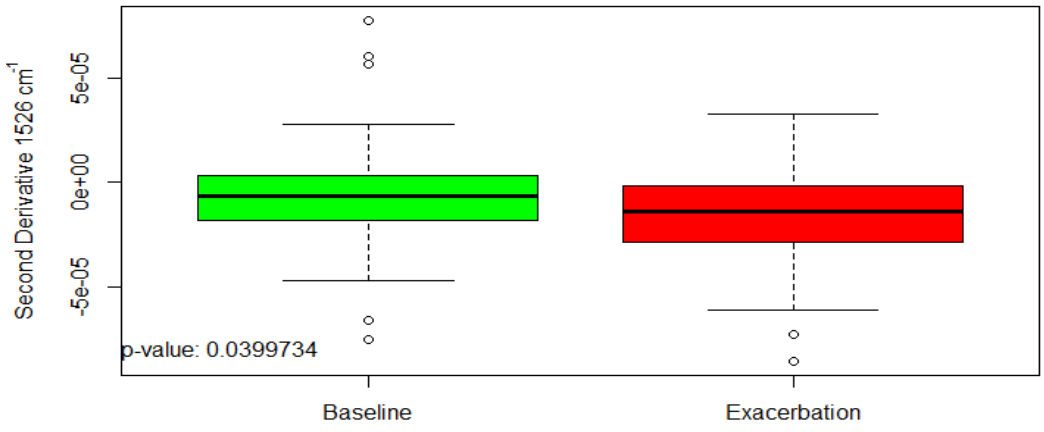


**Figure 3-18: Distribution of absorbance values at 1526cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

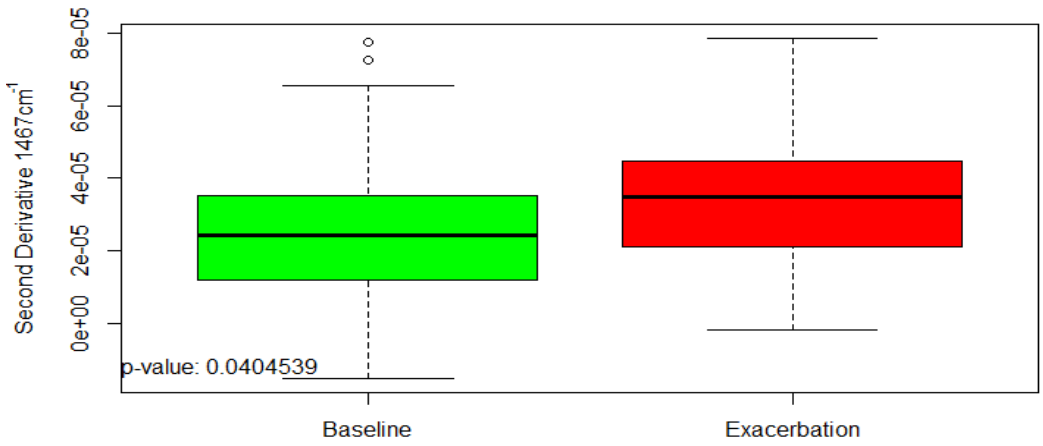


**Figure 3-19: Distribution of absorbance values at 1467cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

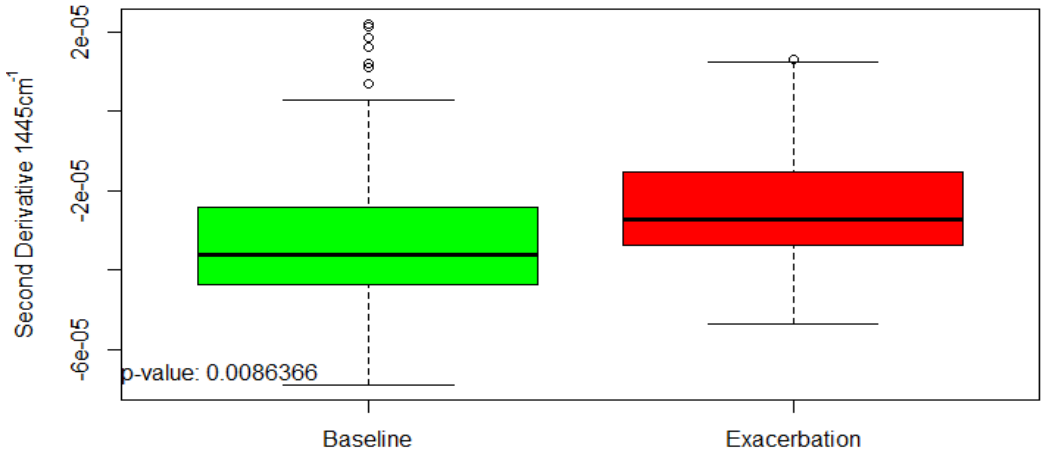


**Figure 3-20: Distribution of absorbance values at 1445cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

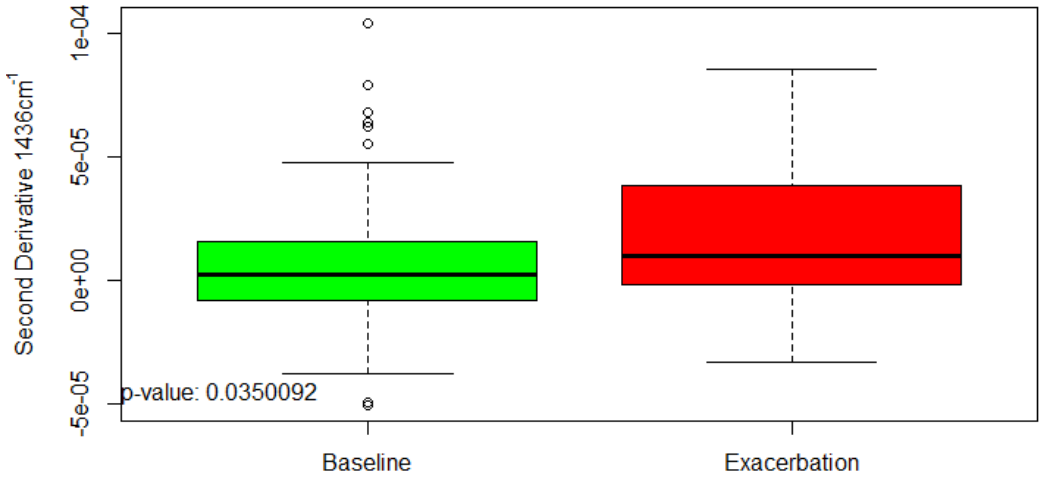


**Figure 3-21: Distribution of absorbance values at 1436cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

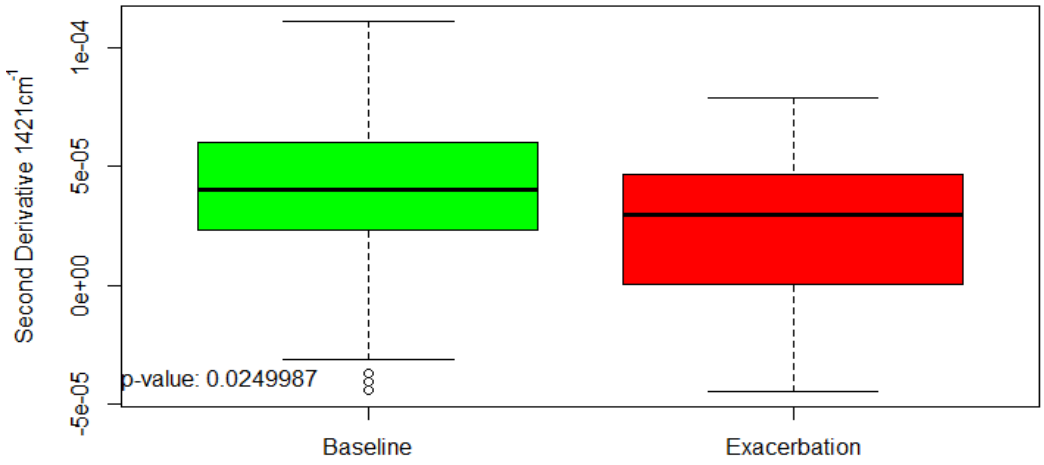


**Figure 3-22: Distribution of absorbance values at 1421cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

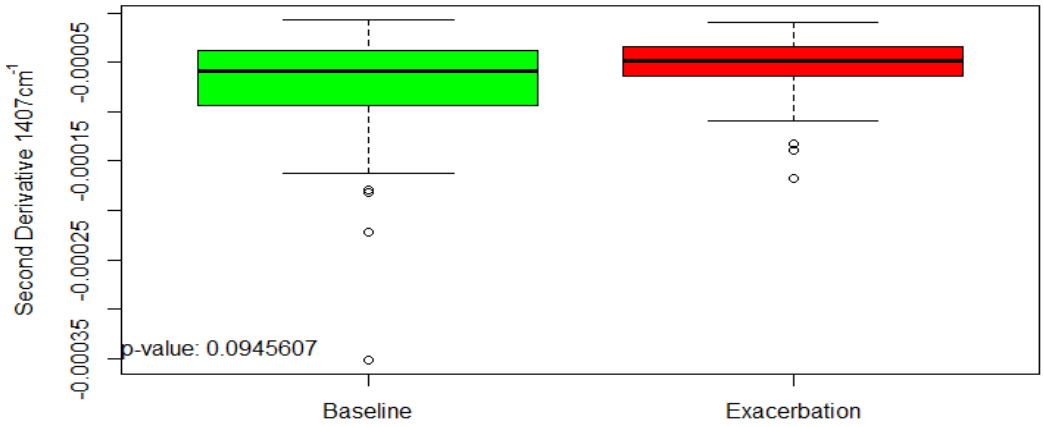


**Figure 3-23: Distribution of absorbance values at 1556cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

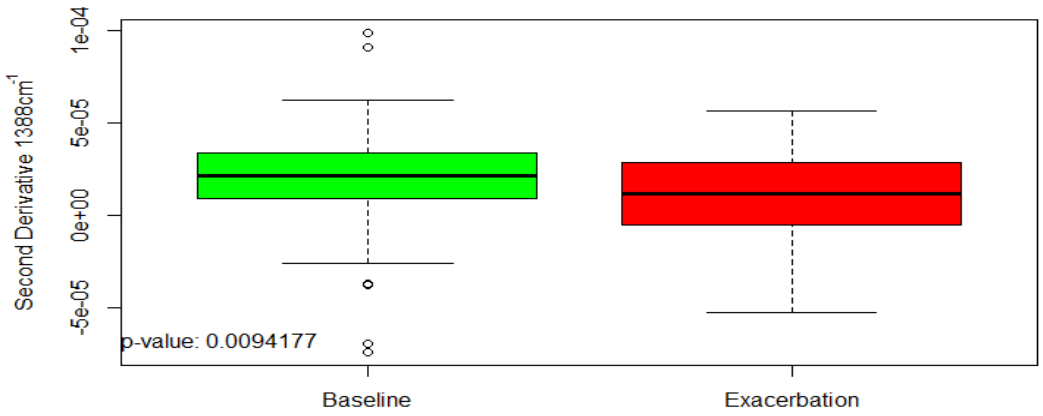


**Figure 3-24: Distribution of absorbance values at 1556cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

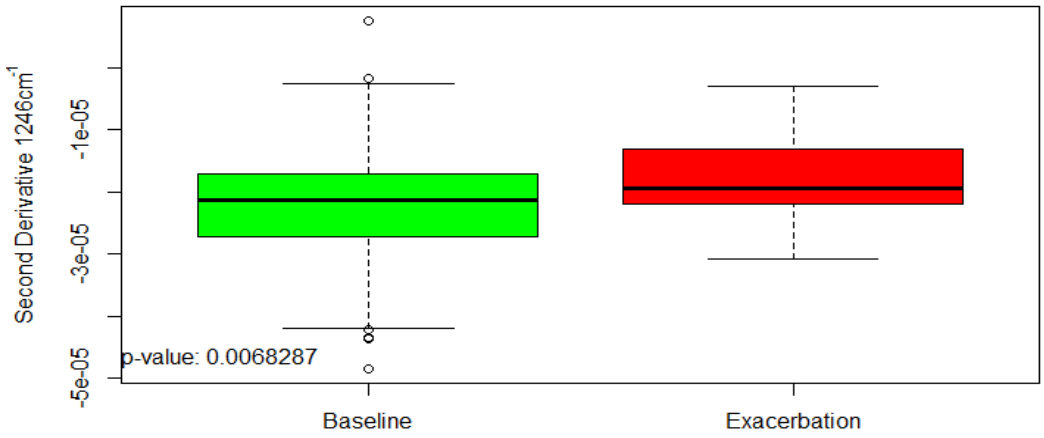


**Figure 3-25: Distribution of absorbance values at 1246cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

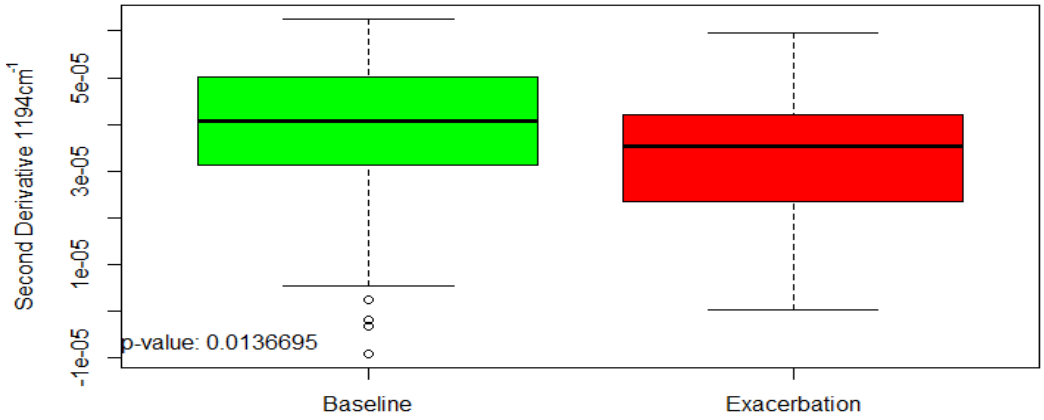


**Figure 3-26: Distribution of absorbance values at 1194cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

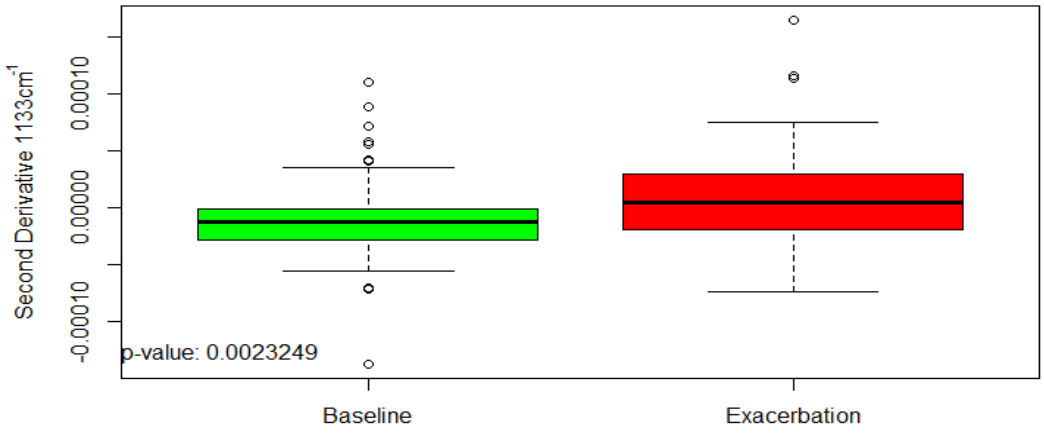


**Figure 3-27: Distribution of absorbance values at 1133cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

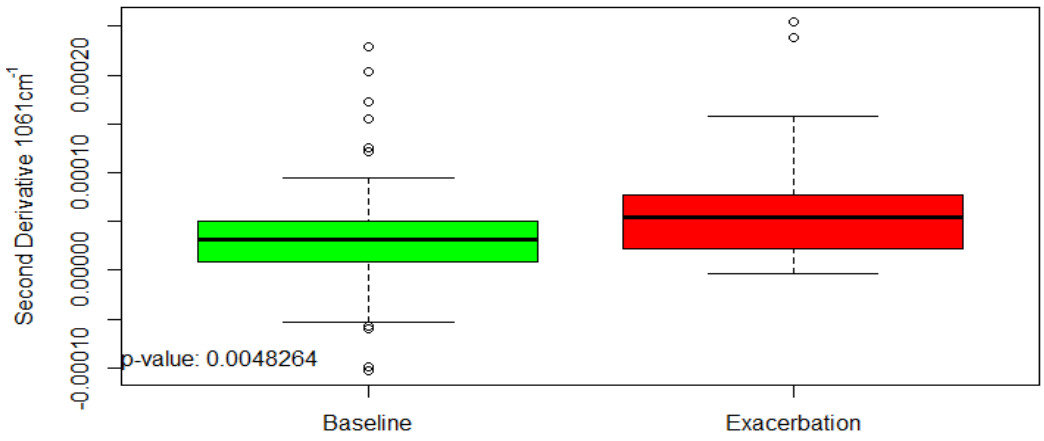


**Figure 3-28: Distribution of absorbance values at 1061cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

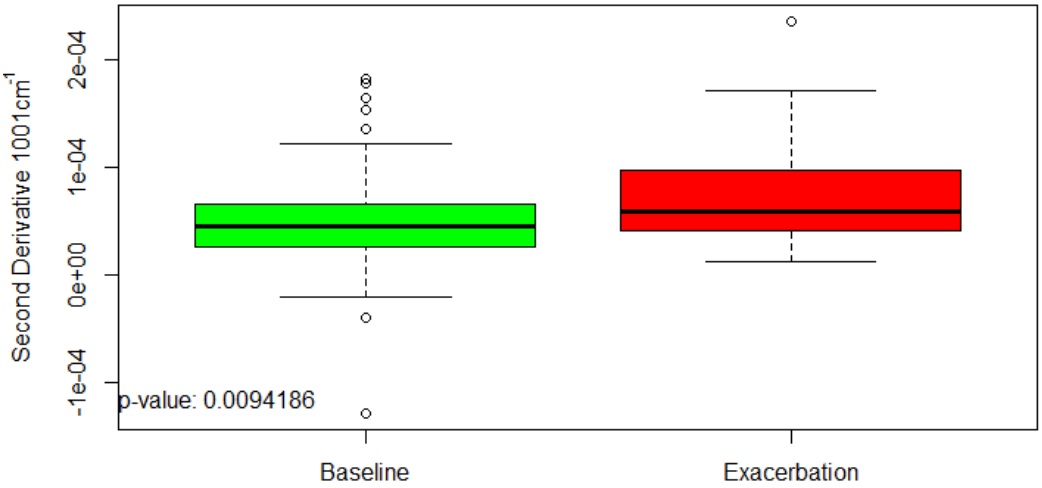


**Figure 3-29: Distribution of absorbance values at 1001cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

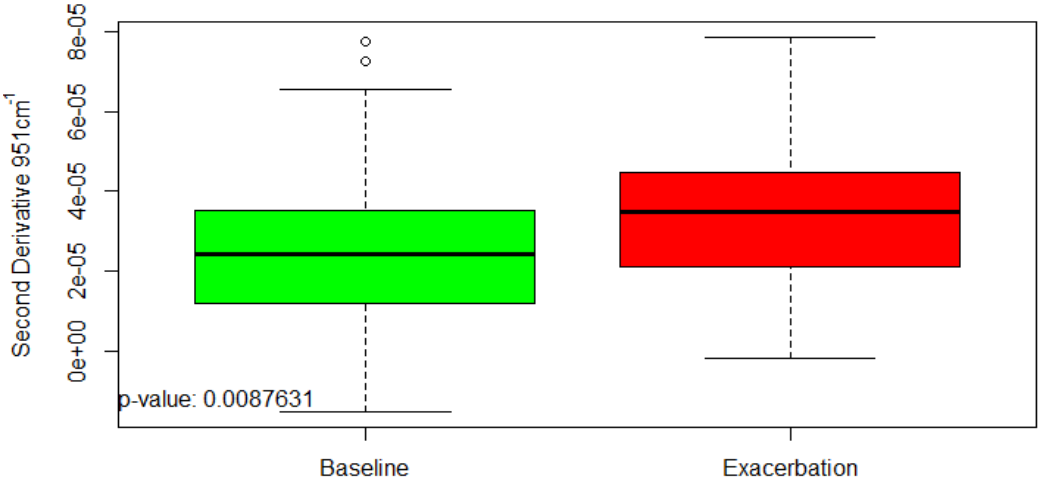


**Figure 3-30: Distribution of absorbance values at 951cm$^{-1}$ in COPD baseline (green) and exacerbation (red) sputum second-derivative spectra**

PCA analysis was carried out on all wavenumbers that were deemed to be significant at the 95% confidence level, in order to identify wavenumbers which were the main sources of variation within the patient cohorts.

PCA was not able to differentiate COPD baseline from exacerbation with a high degree of sensitivity or specificity. This was not surprising, given that the distributions of absorbencies at these wavenumbers showed a large degree of overlap. The lack of clustering of patient spectra did not allow classification of groups of variance, so it was not possible to identify and group the patients by sources of variance. Both patient cohorts were found in all four PCA quadrants, with no clear clustering of groups (Figure 3-31).

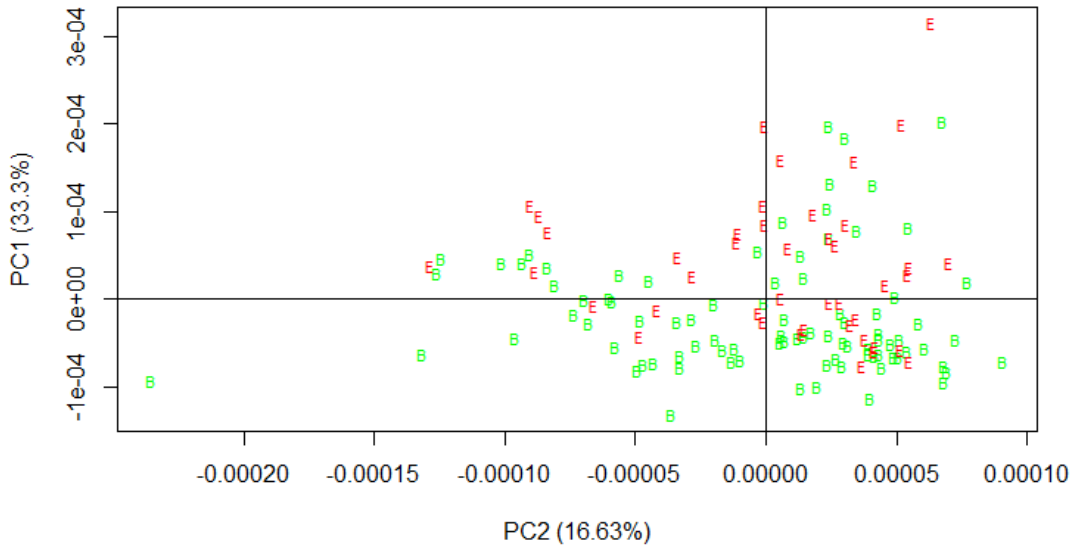


**Figure 3-31: PCA scatterplot of PC1 vs PC2 based on the 18 wavenumbers which were indicated to be statistically significant ($p < 0.05$) for determining COPD baseline (green) from exacerbation (red). The principle components which account for the most variation between samples are shown.**

### 3.3.3 Distribution of Peak Positions in Second Derivative Spectra

Second derivative spectra were calculated from the corresponding absorbance spectra for each sample and peak picking analysis was carried out within the 1800-900cm$^{-1}$ fingerprint region. Peak detection analysis was performed on the second-derivative spectra rather than the absorbance spectra because it is possible to discriminate small peaks and shoulders that would otherwise be obscured by larger absorbance peaks. In this way, loss or gain of secondary absorbance bands that could be indicative of exacerbation may be identified.

The detected peaks in all patient spectra were grouped together by spectral region and COPD status. The distributions of peak positions in wavenumbers were examined using histograms (Figure 3-32 to Figure 3-53) and boxplots (Figure 3-54, Figure 3-55, and Figure 3-56), and significance was assessed using the non-parametric Wilcoxon rank sum test, with statistical significance being found at the 95% confidence level.

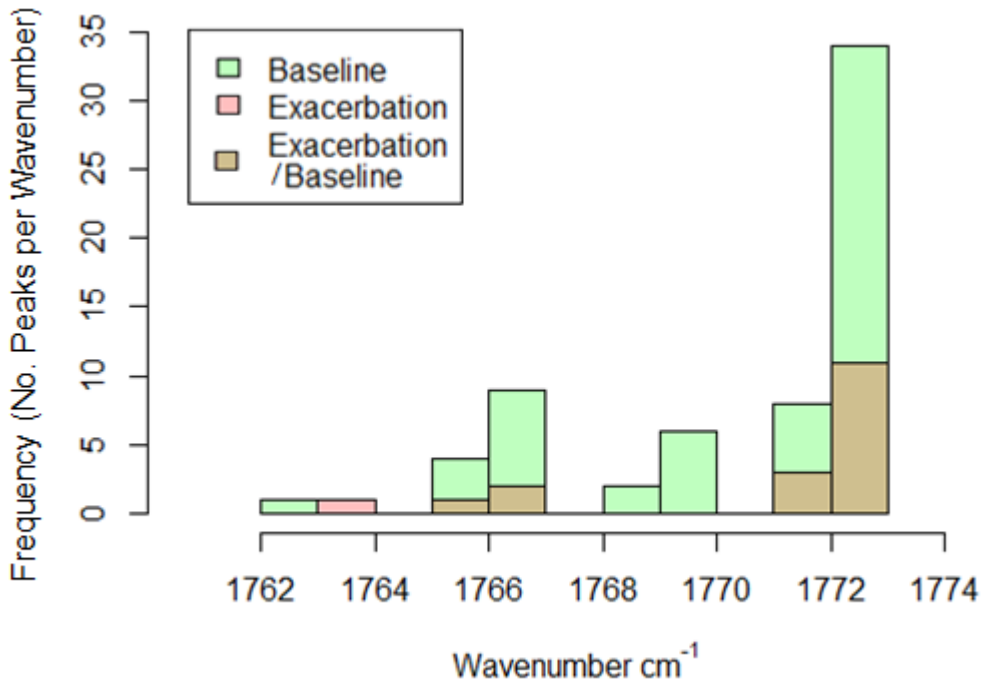


**Figure 3-32: Histograms showing distribution of wavenumber position of detected peaks around 1770cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

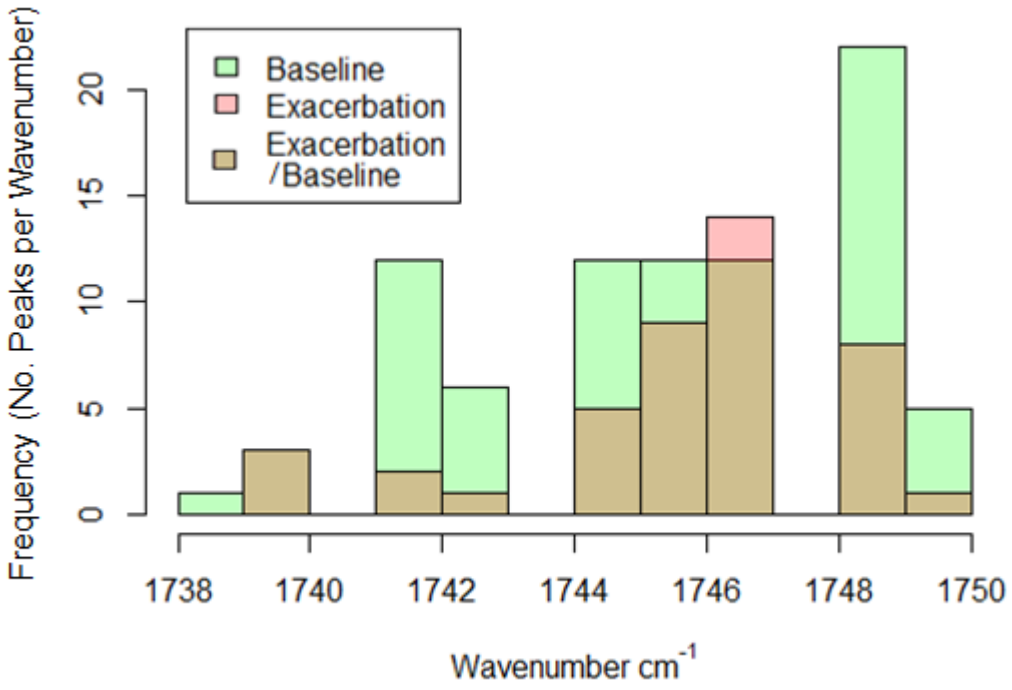


**Figure 3-33: Histograms showing distribution of wavenumber position of detected peaks around 1745cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

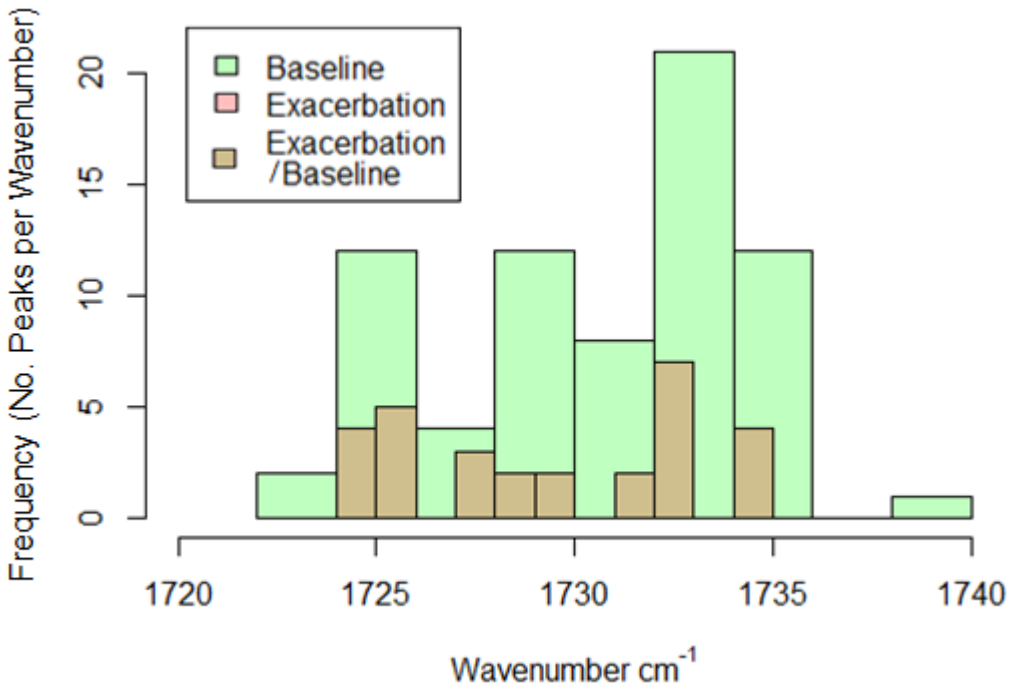


**Figure 3-34: Histograms showing distribution of wavenumber position detected peaks around 1730cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

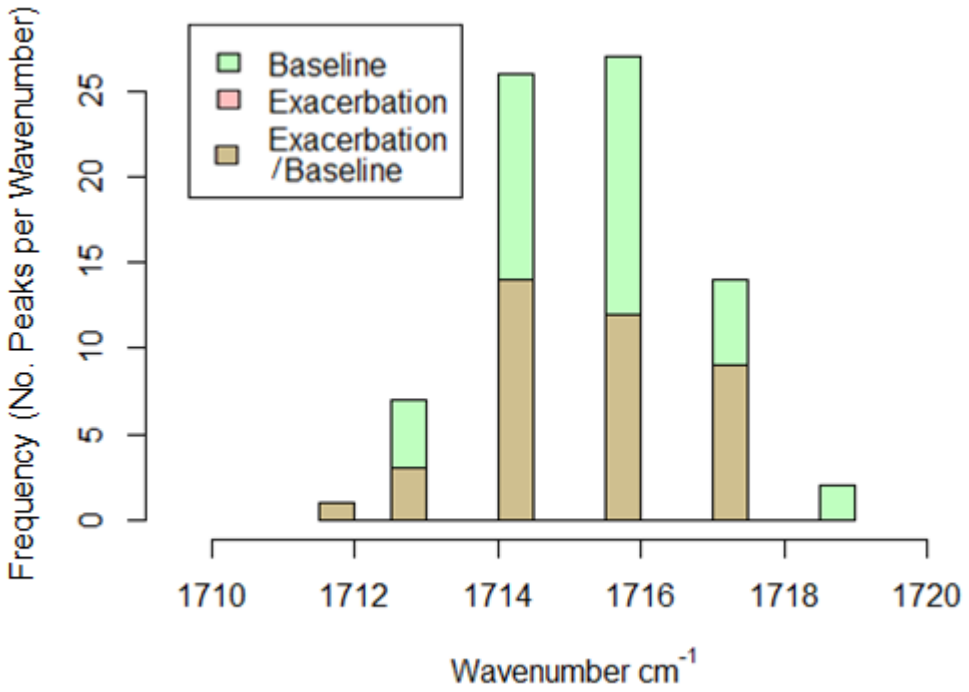


**Figure 3-35: Histograms showing distribution of wavenumber position of detected peaks around 1714cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

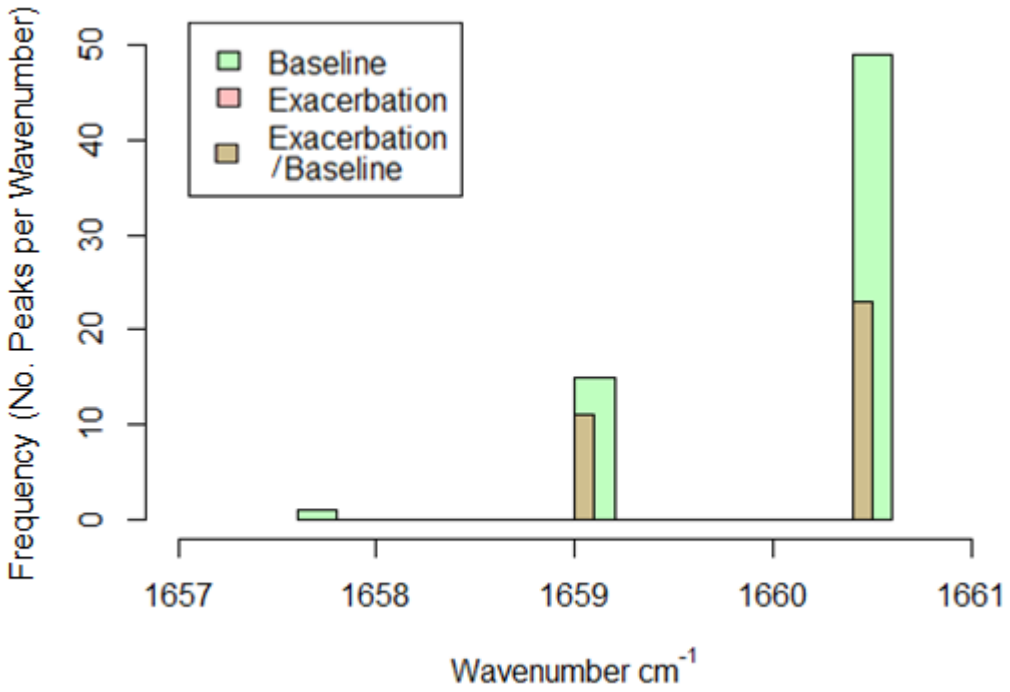


**Figure 3-36: Histograms showing distribution of wavenumber position of detected peaks around 1659cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

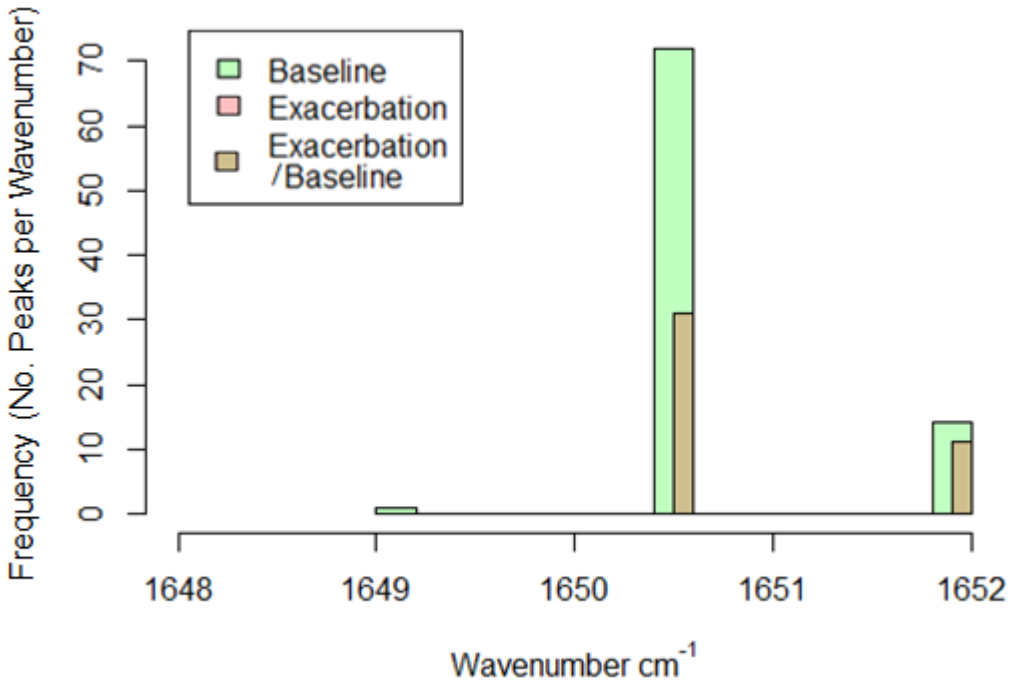


**Figure 3-37: Histograms showing distribution of wavenumber position of detected peaks around 1650cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

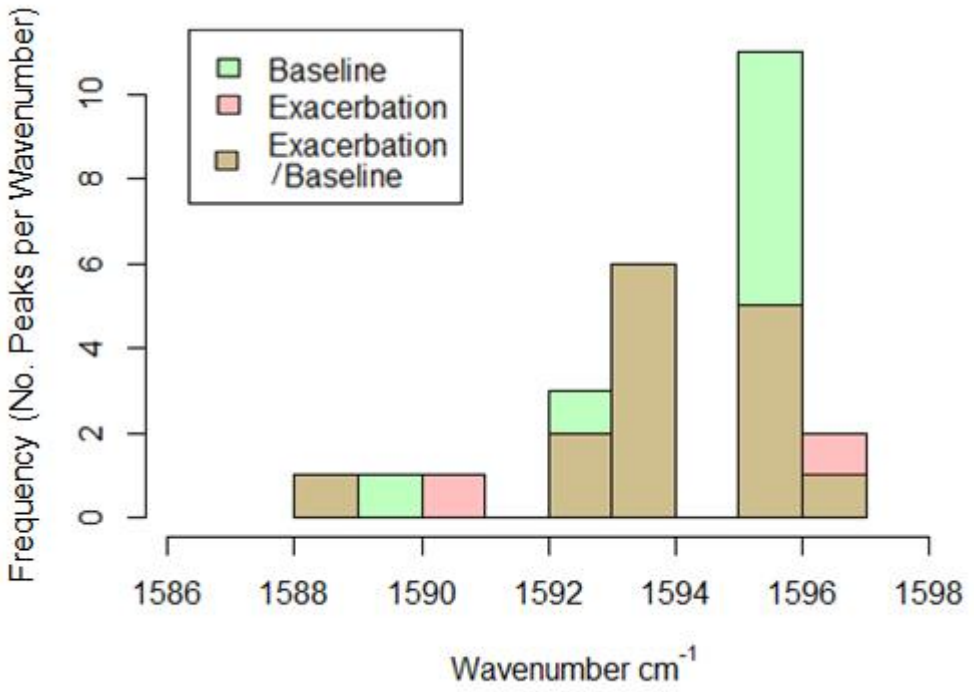


**Figure 3-38: Histograms showing distribution of wavenumber position of detected peaks around 1594cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

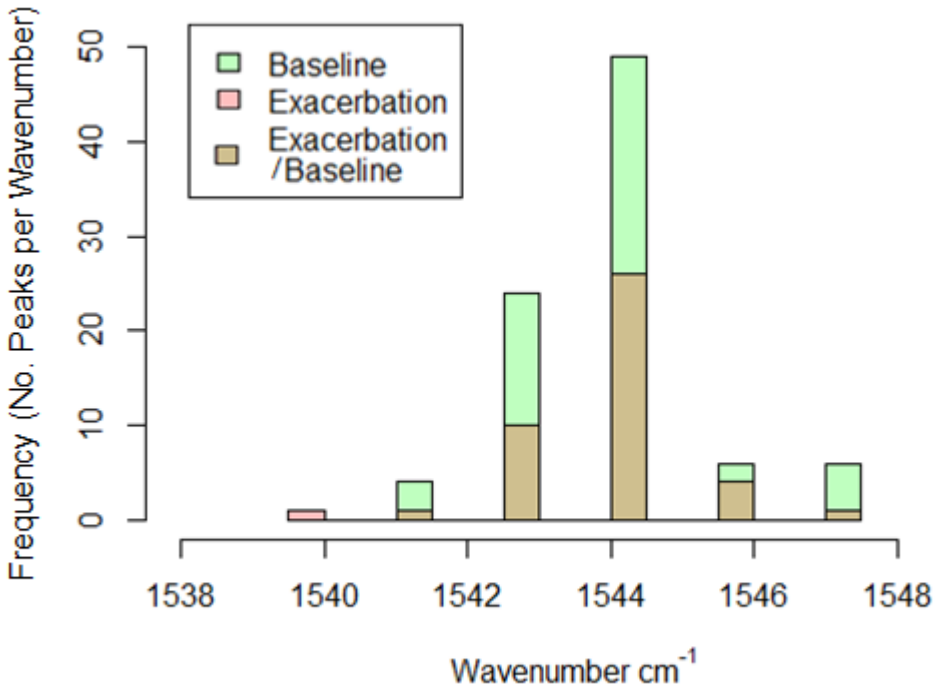


**Figure 3-39: Histograms showing distribution of wavenumber position of detected peaks around 1544cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

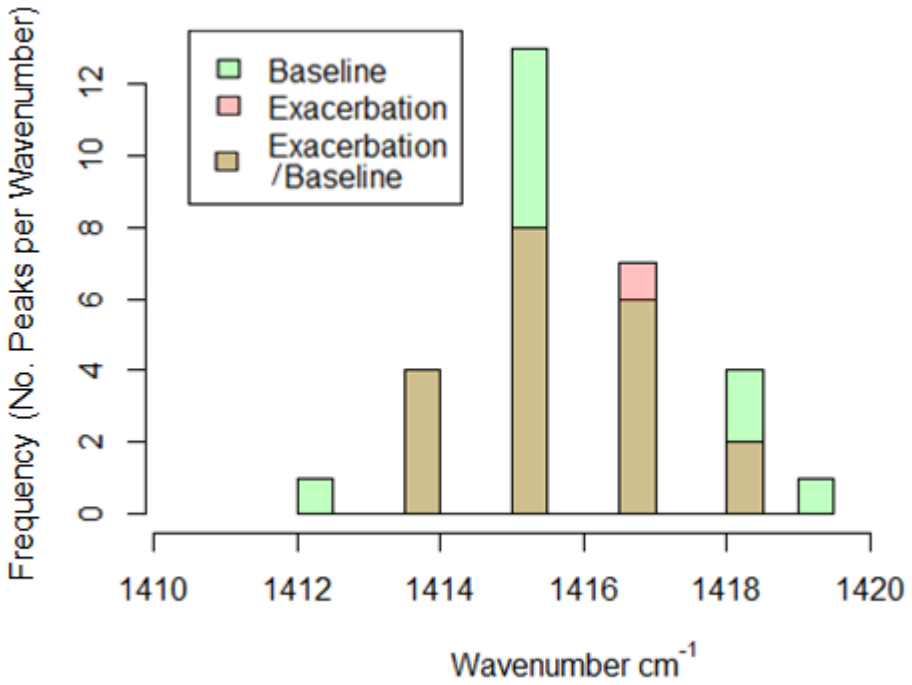


**Figure 3-40: Histograms showing distribution of wavenumber position of detected peaks around 1415cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

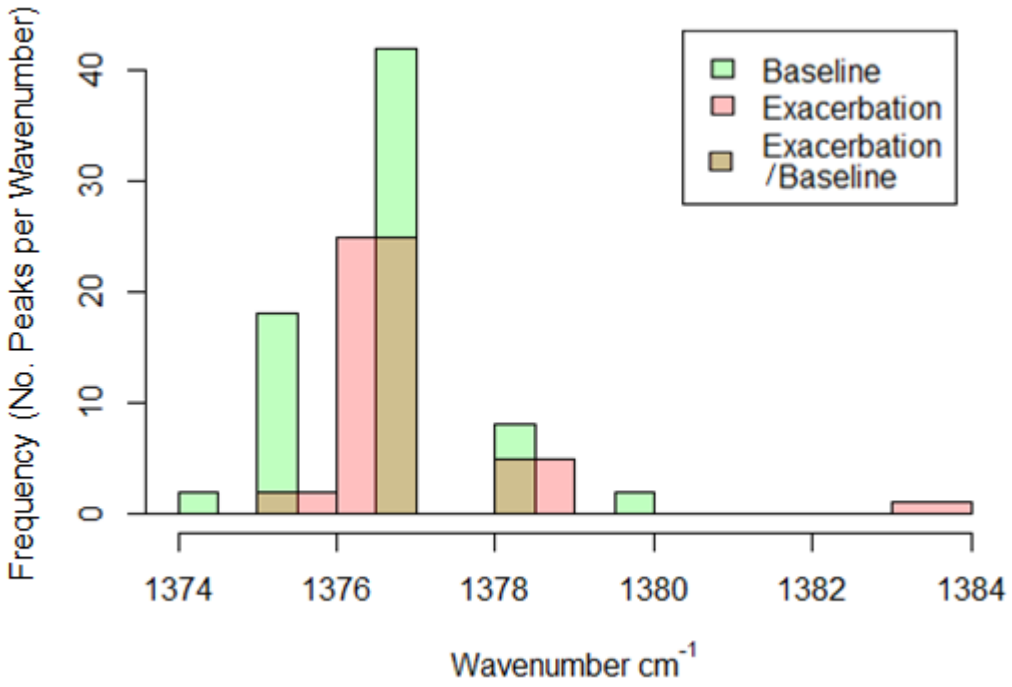


**Figure 3-41: Histograms showing distribution of wavenumber position of detected peaks around 1376cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

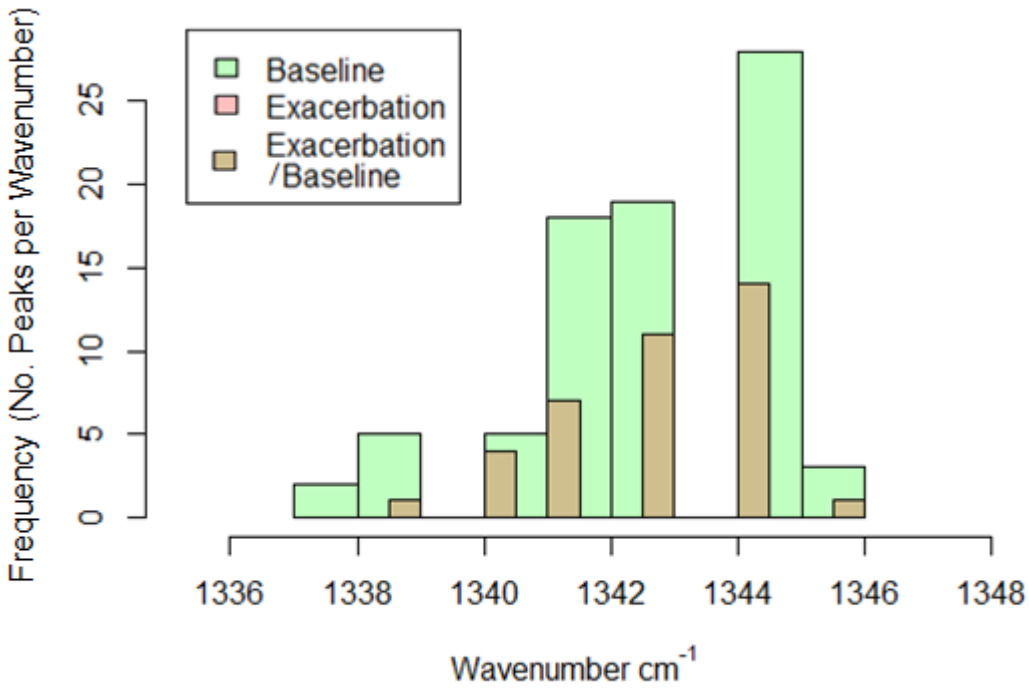


**Figure 3-42: Histograms showing distribution of wavenumber position of detected peaks around 1342cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

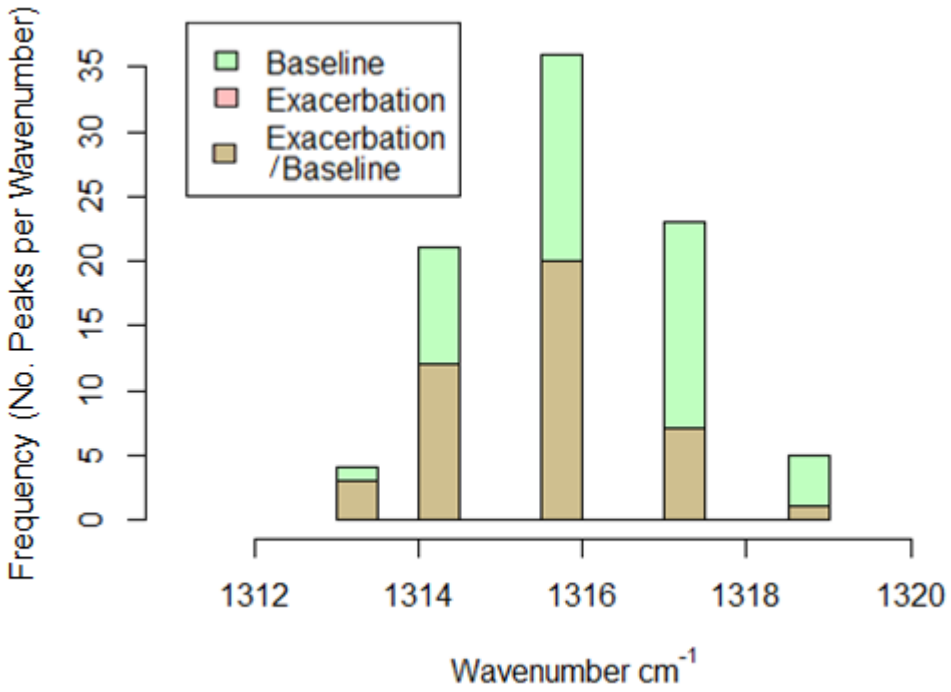


**Figure 3-43: Histograms showing distribution of wavenumber position of detected peaks around 1316cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

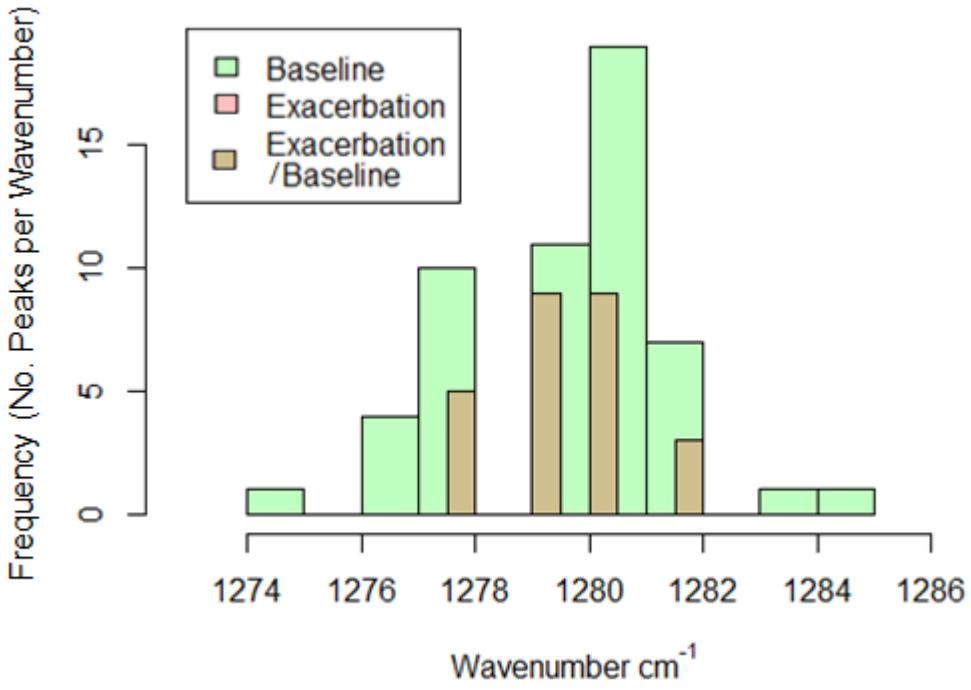


**Figure 3-44: Histograms showing distribution of wavenumber position of detected peaks around 1280cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

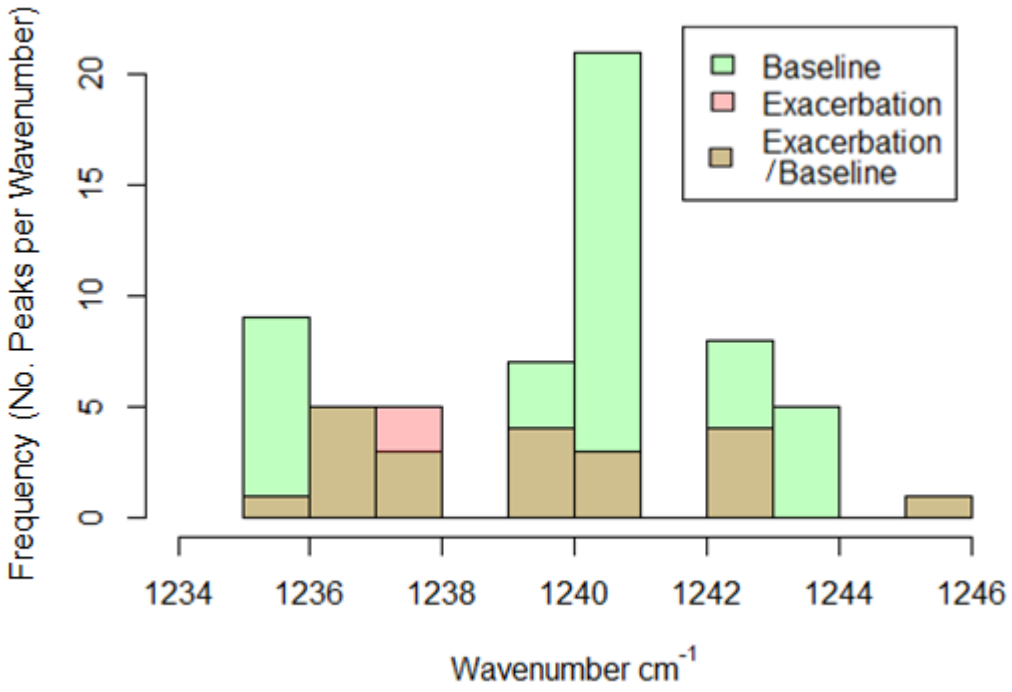


**Figure 3-45: Histograms showing distribution of wavenumber position of detected peaks around 1240cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

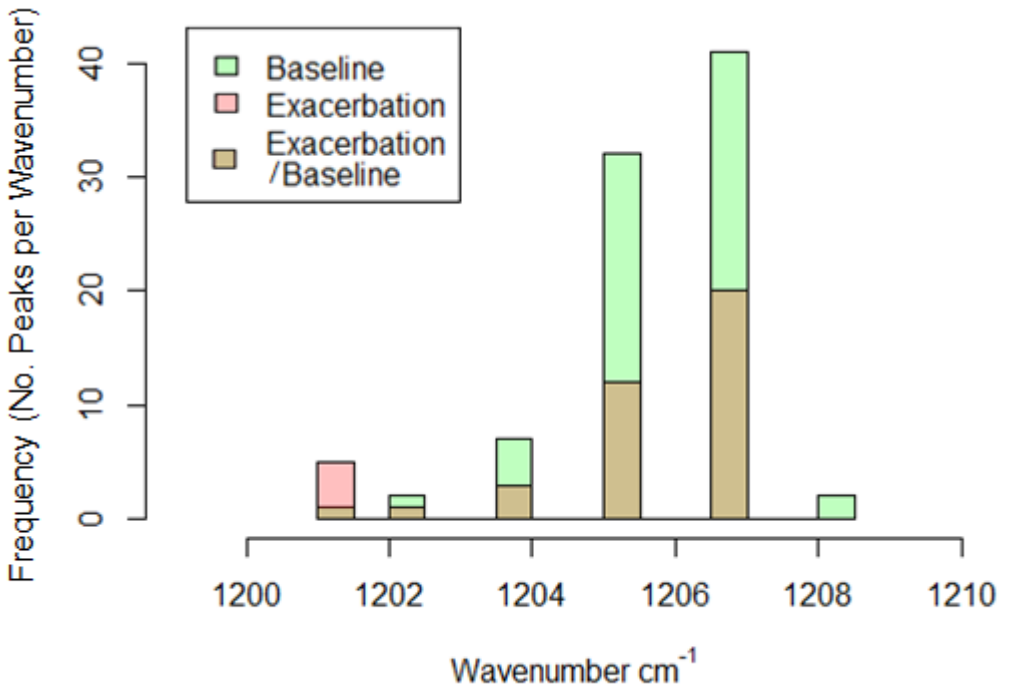


**Figure 3-46: Histograms showing distribution of wavenumber position of detected peaks around 1205cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

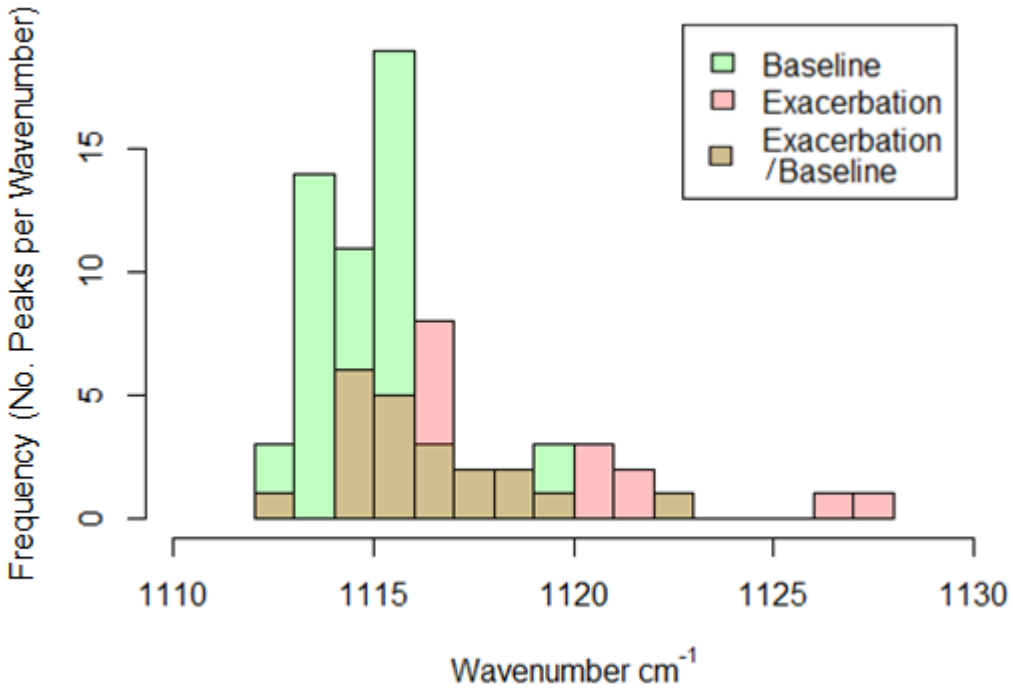


**Figure 3-47: Histograms showing distribution of wavenumber position of detected peaks around 1115cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

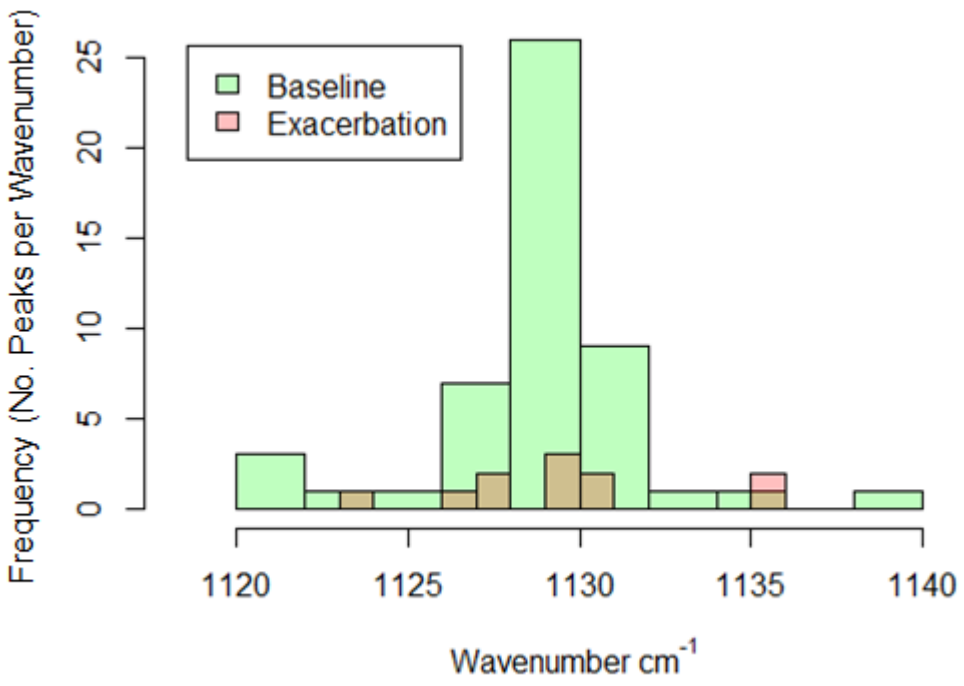


**Figure 3-48: Histograms showing distribution of wavenumber position of detected peaks around 1128cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

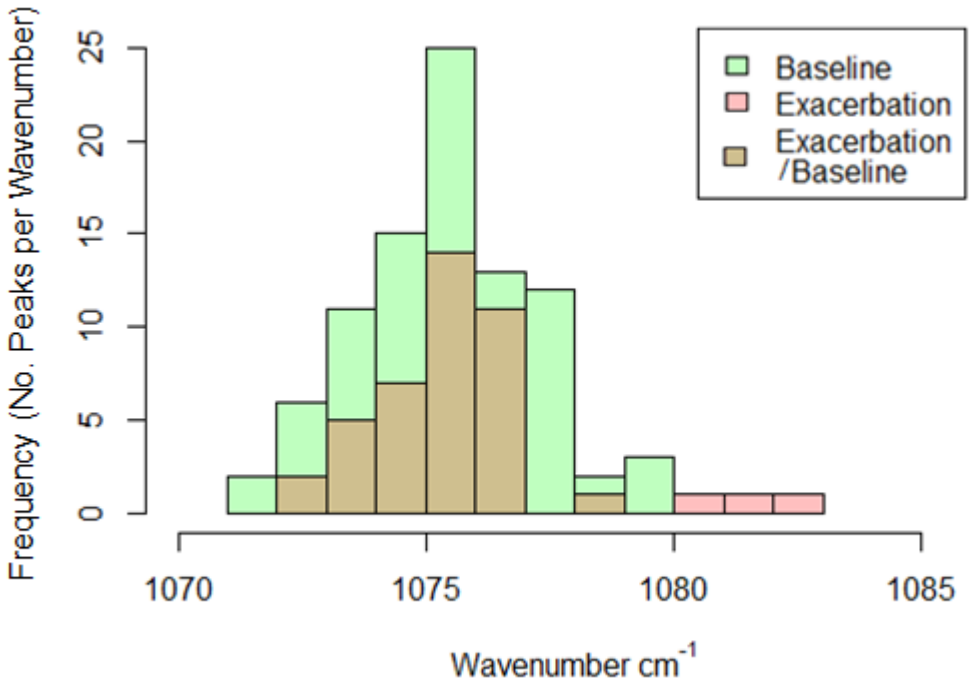


**Figure 3-49: Histograms showing distribution of wavenumber position of detected peaks around 1076cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

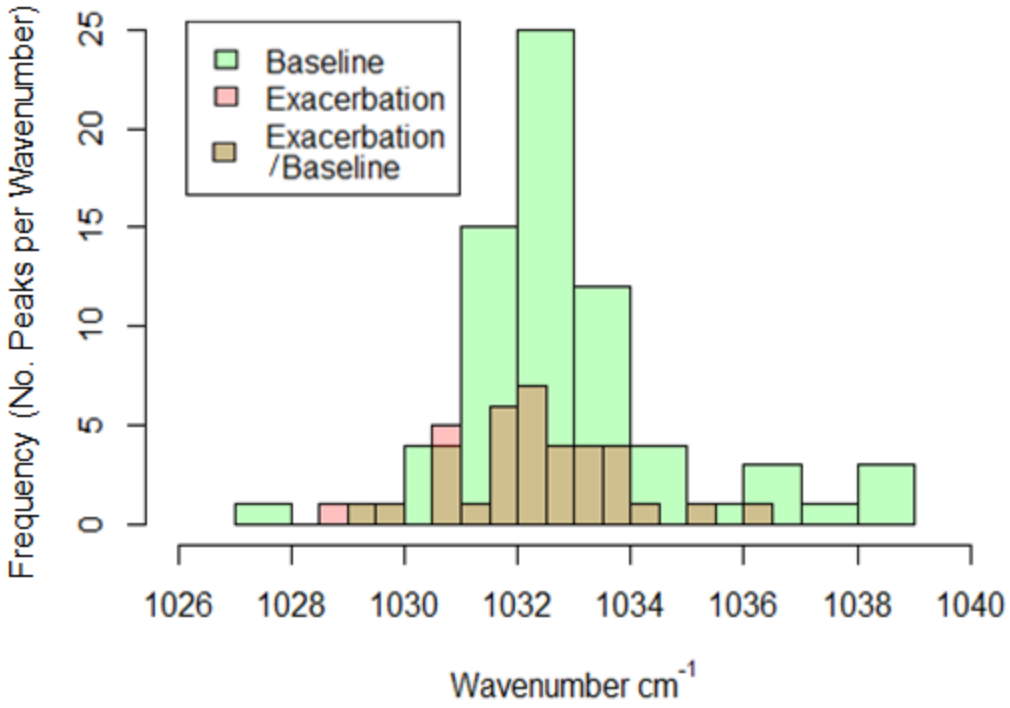


**Figure 3-50: Histograms showing distribution of wavenumber position of detected peaks around 1033cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

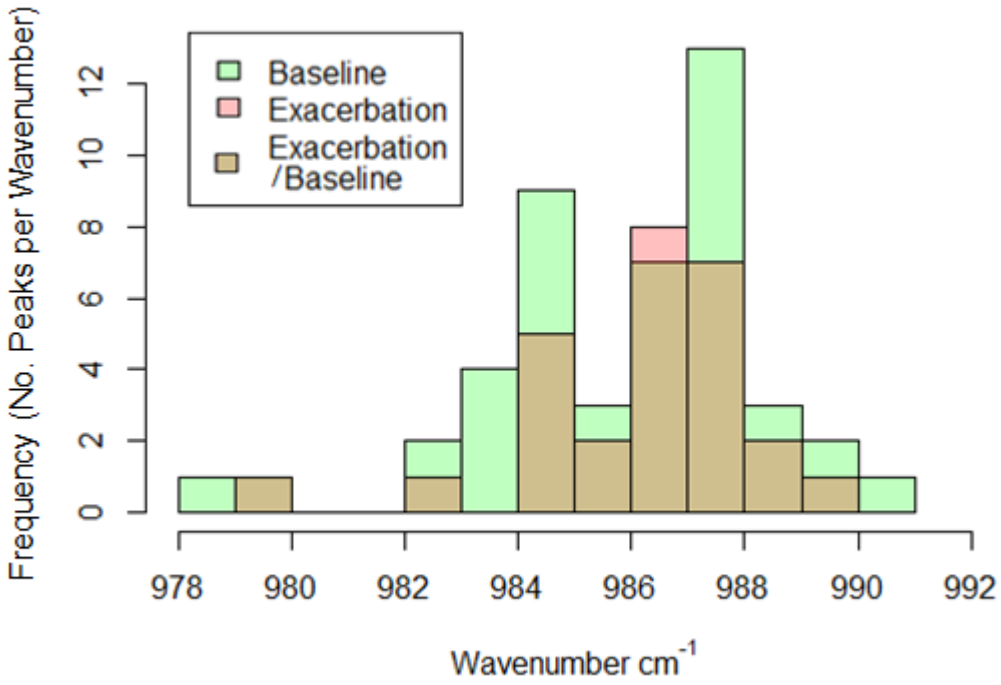


**Figure 3-51: Histograms showing distribution of wavenumber position of detected peaks around 986cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

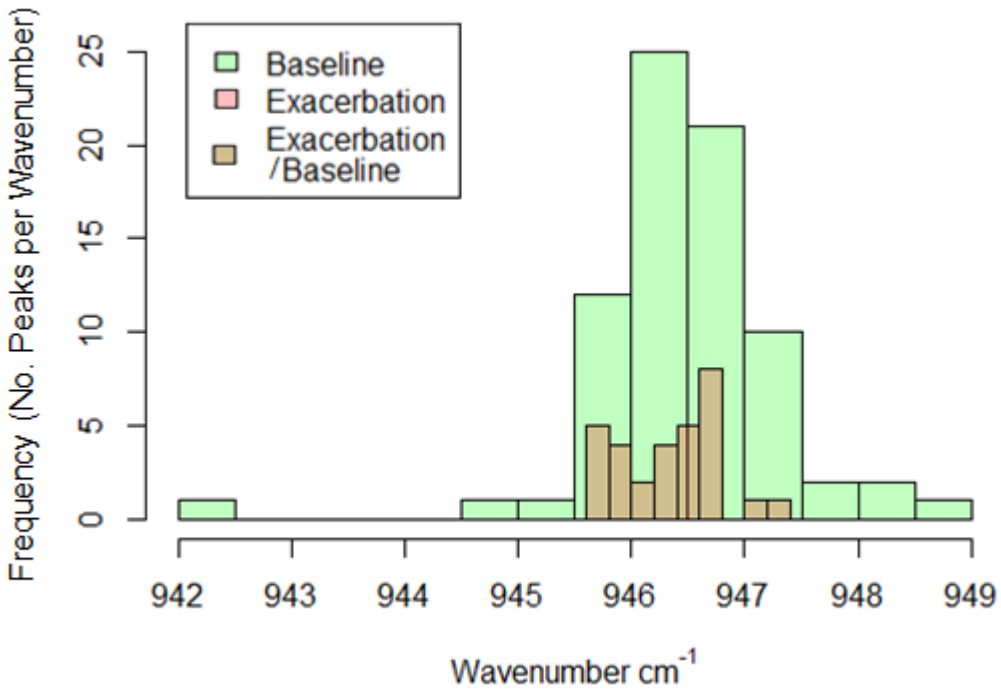


**Figure 3-52: Histograms showing distribution of wavenumber position of detected peaks around 946cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

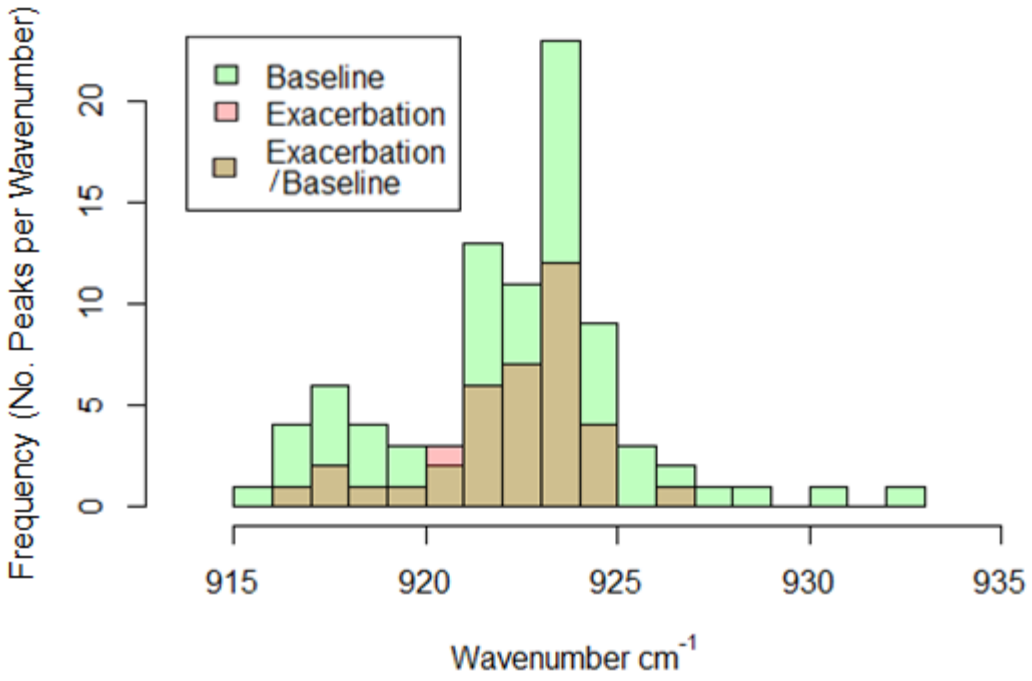


**Figure 3-53: Histograms showing distribution of wavenumber position of detected peaks around 922cm$^{-1}$ in the second-derivative spectra of COPD baseline (green) and exacerbation (red) patients: overlapping colours show both exacerbation and baseline patients with peaks at the same wavenumber**

The histograms show that second-derivative peak distribution is very similar in both COPD baseline and COPD exacerbation sputum. Statistical significance testing showed that the distribution of peak positions was not significantly different between COPD baseline and exacerbation patients in all but two detectable second-derivative peaks; 1115 and 1376cm$^{-1}$.

A peak in the second-derivative spectra centred around 1129cm$^{-1}$ was identified in 52.81%, and 30.23% of baseline spectra and exacerbation sputum spectra, respectively. This difference was shown to be statistically significant at the 95% confidence level ($p=0.014$), suggesting that peak loss around 1129cm$^{-1}$ may be indicative of exacerbation, however the low number of baseline samples the peak is detected in suggests this peak is not common in COPD sputum second-derivative spectra. Therefore using peak loss around 1129cm$^{-1}$ as an exacerbation marker could lead to a large number of type I errors in a detection protocol.

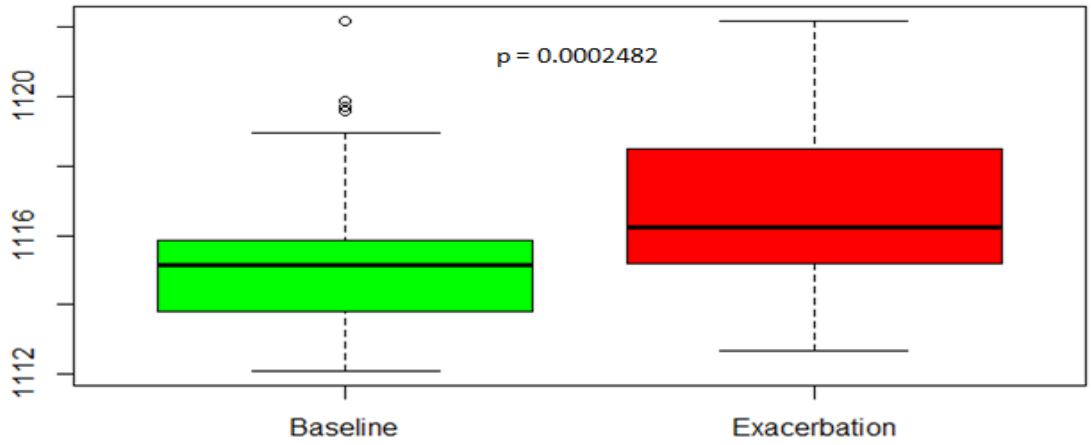


**Figure 3-54: Distribution of wavenumber position of peaks detected around 1116cm$^{-1}$ in the second-derivative spectra of COPD baseline and exacerbation patients.**

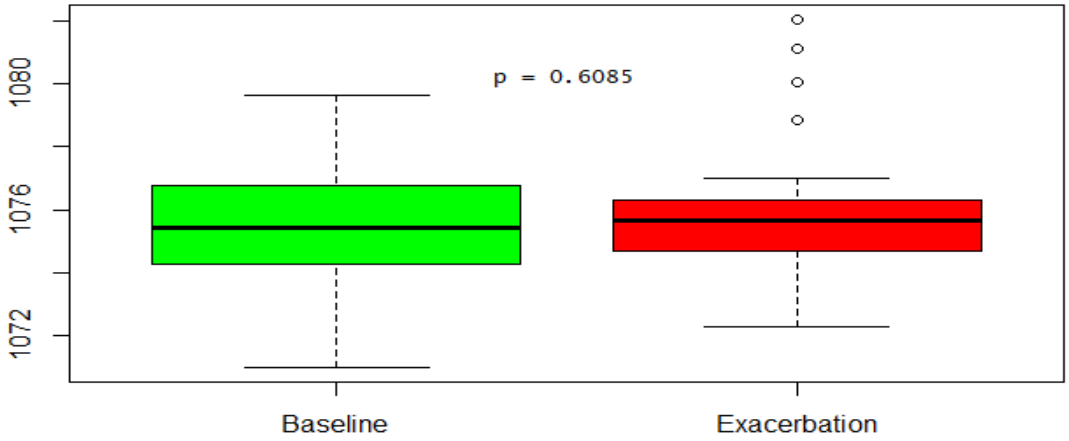


**Figure 3-55: Distribution of wavenumber position of peaks detected around 1076cm$^{-1}$ in the second-derivative spectra of COPD baseline and exacerbation patients**

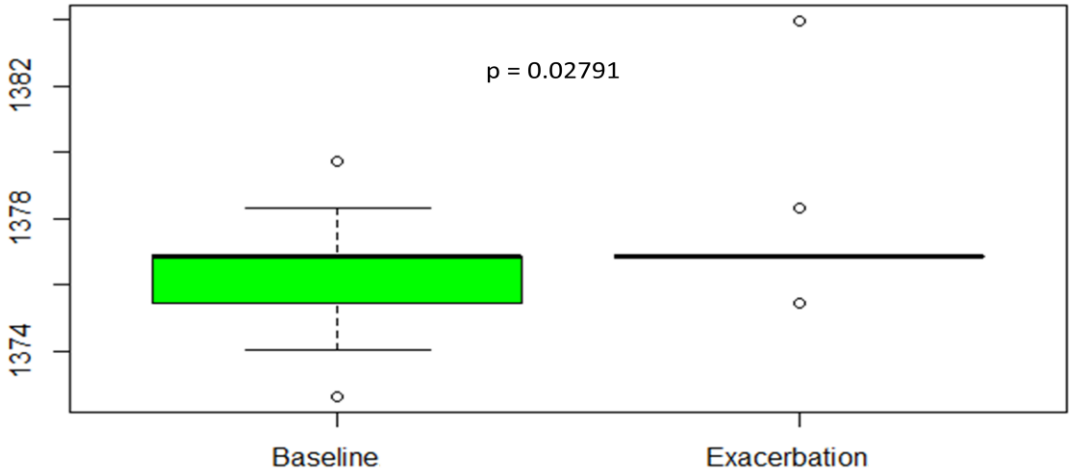


**Figure 3-56: Distribution of wavenumber position of peaks detected around 1076cm$^{-1}$ in the second-derivative spectra of COPD baseline and exacerbation patients**

After examination of the distributions of detected peak positions in second-derivative spectra, it was found that there were two regions in the fingerprint region around which peaks were detected at statistically significantly ($p < 0.05$) different wavenumbers for COPD baseline and exacerbation sputum. These regions were centred at approximately 1376 and 1116$cm^{-1}$.

A statistically significant ($p < 0.05$) peak shift was observed in second-derivative spectra of COPD exacerbation sputum, from 1115.508$cm^{-1}$ in baseline sputum to 1116.984$cm^{-1}$ in exacerbation sputum. This peak was detectable in 68.54% of baseline sputum spectra and 72.09% of exacerbation spectra, and this difference was not found to be statistically significant ($p=0.086$), indicating that the peak is a common peak readily detectable in the majority of COPD patient sputum samples.

The distribution of peaks around 1376$cm^{-1}$ was different however, with both baseline and exacerbation sputum spectra sharing the same median peak position at 1376.88$cm^{-1}$. However the exacerbation sputum spectra showed very little variation from the median, whereas the baseline spectra displayed much more variation. This peak was detectable in 82.02%, and 76.74% of baseline and exacerbation second-derivative spectra, respectively. Therefore, using presence of a peak at 1376.88$cm^{-1}$ would be weak evidence for indication of exacerbation, as too many patients at baseline would display this peak. An alternative to this approach would be to use presence of peaks at other wavenumbers within this region as evidence for not indicating an exacerbation sample.

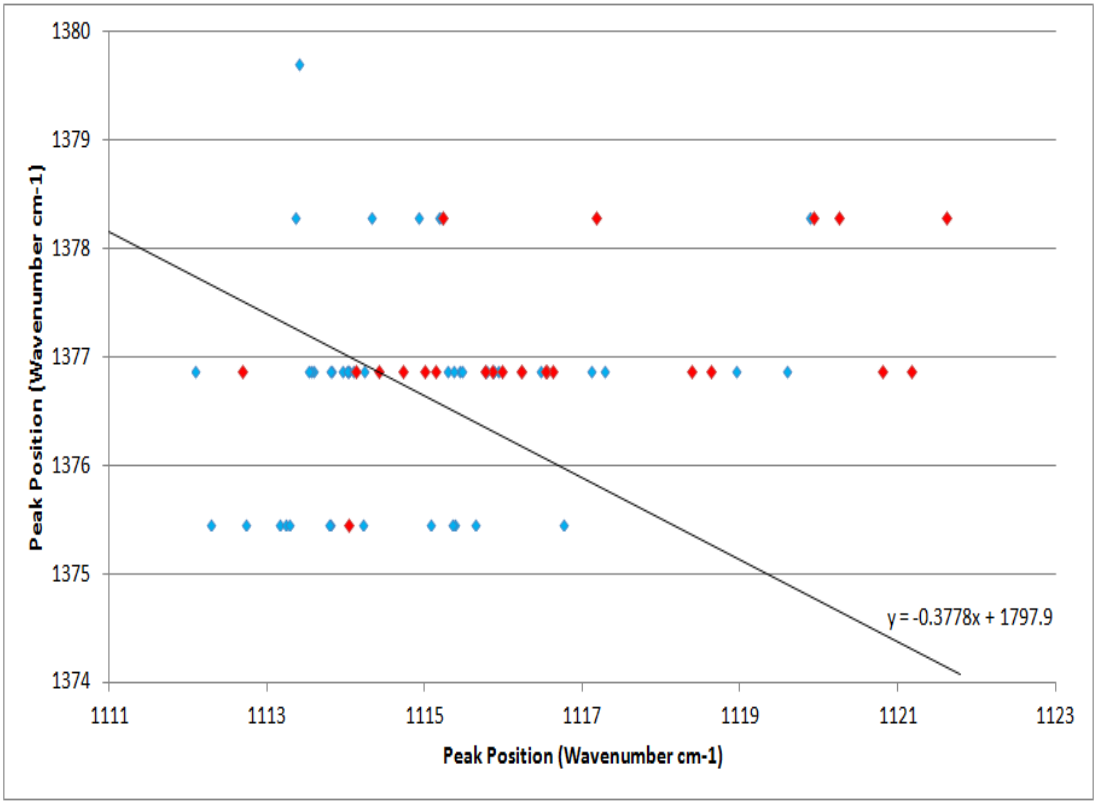


**Figure 3-57: Two dimensional linear regression model comparing positions of detected peak at approximately 1376 and 1116cm$^{-1}$ in COPD baseline (blue) and exacerbation (red) sputum second-derivative spectra**

A linear regression model comparing the detected peak positions at around 1116cm$^{-1}$ and 1376cm$^{-1}$ was generated (Figure 3-57). Using the model, COPD exacerbation spectra can be differentiated from baseline with sensitivity and specificity of 80.76% and 48.00%, respectively. Both peaks were observed in 58.01% of all patients, with no statistically significant difference observed between patient groups for the detection of peaks.

### 3.3.4 Distinction of COPD patients based on Physiological Measures

One possible factor which may contribute to the difficulty to distinguish baseline COPD from exacerbation is the fact that COPD exacerbation is a heterogeneous condition, with a multifactorial aetiology. It may be possible to distinguish exacerbation from stable disease with greater power if some of the following confounding variable could be controlled for.

At the time of recruitment to MEDLUNG, patients were asked to provide information regarding their state of health and medical background. Alongside their current COPD status (baseline or exacerbation) information pertaining to the patients' biological sex, age, medical history and comorbidities, drug history, current

smoking status and pack-years, current infection status, occupation, likely exposure to asbestos, and lung function by FEV1 (%) were all collected. Patients were grouped by common features within each category and these groups were analysed to assess the level of variation within the sputum FTIR spectra that each variable may be responsible for. Correlation between non-categorical data, such as pack-years or FEV1, and absorbance intensity and peak position was evaluated.

#### 3.3.4.1 Infection

Infection of the airways is a common cause of exacerbation in COPD patients, with increased bacterial colonisation being evident in an estimated 69% of exacerbation cases (Wedzicha & Seemungal, 2007) and viral infection being the primary cause of an estimated 22-23%. Examination of the patient clinical information shows that the majority of exacerbating patients (25 infections out of 35 total exacerbation samples) had a bacterial infection in their airways, compared to only 7 confirmed baseline COPD patients. Due to this low number of infected baseline patients, all patients were grouped by infection status regardless of current COPD status. This was deemed appropriate for two reasons: 1) the current interest was to observe differences specific to infection status, and 2) the majority of exacerbation samples collected were reported to be infective exacerbations, and that exacerbations of COPD have been heavily associated with increased bacterial and viral loads in the patient's airways (Huang & Boushey, 2015; Seemungal, Harper-Owen, *et al.*, 2000; Wedzicha & Seemungal, 2007).

FTIR analysis to distinguish infected vs non-infected COPD patient sputum was carried out. The average spectra of the infected and non-infected cohort were plotted to visualise any differences between groups. The infected sputum spectra were found to show a slight decrease in absorption at the protein associated amide I (~1630cm$^{-1}$) and amide II (~1530cm$^{-1}$) regions, alongside a concurrent slight increase in absorption at 1032 and 990cm$^{-1}$ (Figure 3-58), which are associated with sialylated and sulphated Lewis antigen species in mucin glycans, respectively (A. T. Lewis *et al.*, 2013). Sulphated and sialylated glycans are associated with infection and exacerbation in CF and CB patients (Davril *et al.*, 1999; Lhermitte *et al.*, 1991). This suggests that relative to the overall protein content of sputum, the levels of sulphated and sialylated glycans are increased during times of pulmonary infection in COPD patients.

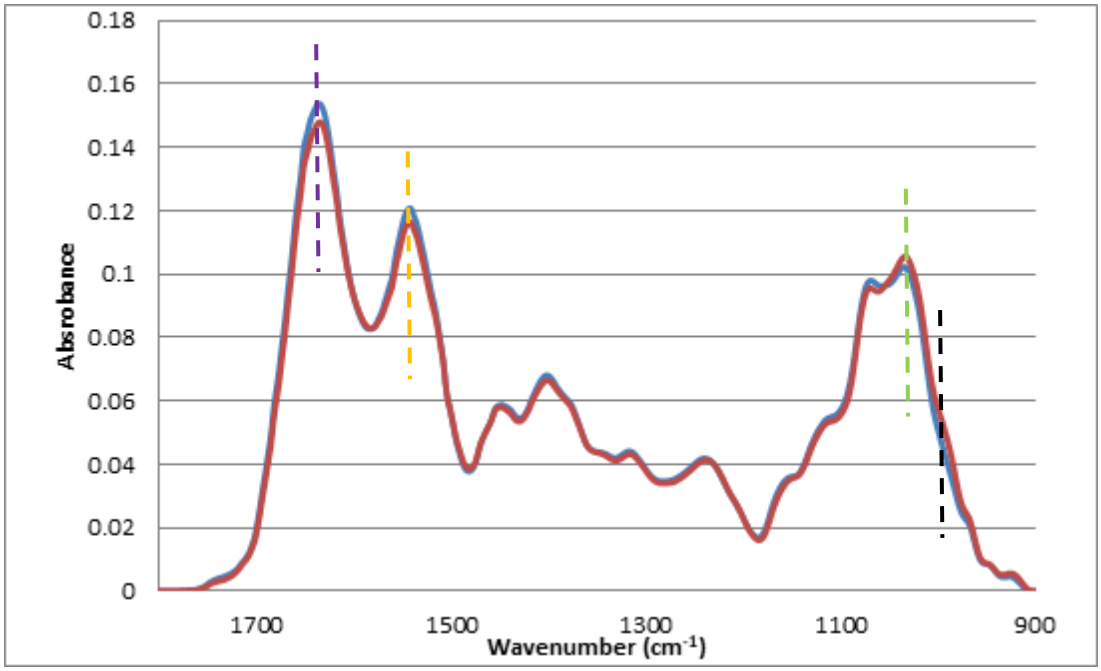


**Figure 3-58: Vector-normalised, baseline-corrected absorbance spectra of infected sputum (red) and non-infected sputum (blue). A slight decrease in absorption can be observed in the amide I (purple dashed) and amide II (orange dashed) regions, along with a concurrent slight increase in absorption at 1032 (green dashed) and 990cm$^{-1}$ (black dashed) can be observed.**

### 3.3.4.2 Biological Sex

It has been shown that COPD is becoming more common in women, and female COPD patients develop more severe symptoms after fewer pack-years than male patients (Han *et al.*, 2007). The biological rationale for this observation has been discussed in chapter 1 (section 1.1.23), but briefly, it is thought that the mechanism behind women's increased susceptibility to COPD is based on oestrogen levels, which are higher in women. MUC5AC and MUC5B synthesis in human bronchial epithelial cells can be upregulated by oestradiol (Choi *et al.*, 2009; Tam *et al.*, 2014), therefore the increased levels of oestrogen in women could contribute to a more rapid worsening of COPD symptoms through more rapid mucin synthesis and secretion.

The COPD patients were grouped by biological sex as determined on the patient recruitment form and statistically significant differences between the two cohorts was sought for each detected peak position and absorbance intensity. No peak position changes could be identified, but small changes in absorption within the glycogen-rich region and at approximately 1580cm$^{-1}$ could be identified. However these were not deemed to be statistically significant at the 95% confidence level.

### 3.3.4.3 Smoking Status

Smoking is known to be the major contributing factor to COPD progression, with smoking activating many of the same signal transduction pathways implicated in COPD disease progression and exacerbation (Lemjabbar *et al.*, 2003; Seatta *et al.*, 1998; Thai *et al.*, 2008).

Patients were grouped by their current smoking status (current-, ex-, or never-smoker) and by pack-years. The vast majority of patients were reported as being ex-smokers (n = 82, 62.12%), with 34 (25.76%) reporting as current-smokers and 7 (5.30%) reporting as never-smokers. As the number of never-smokers was very low, statistical significance was only sought between current- and ex-smokers.

Correlation between the number of pack-years and FTIR spectral features was calculated by Spearman's Rho. A small but significant positive correlation (rho = 0.3, p = 0.0041) was found between pack-years and the wavenumber position of the second-derivative peaks found around 1115cm$^{-1}$ (Figure 3-59).

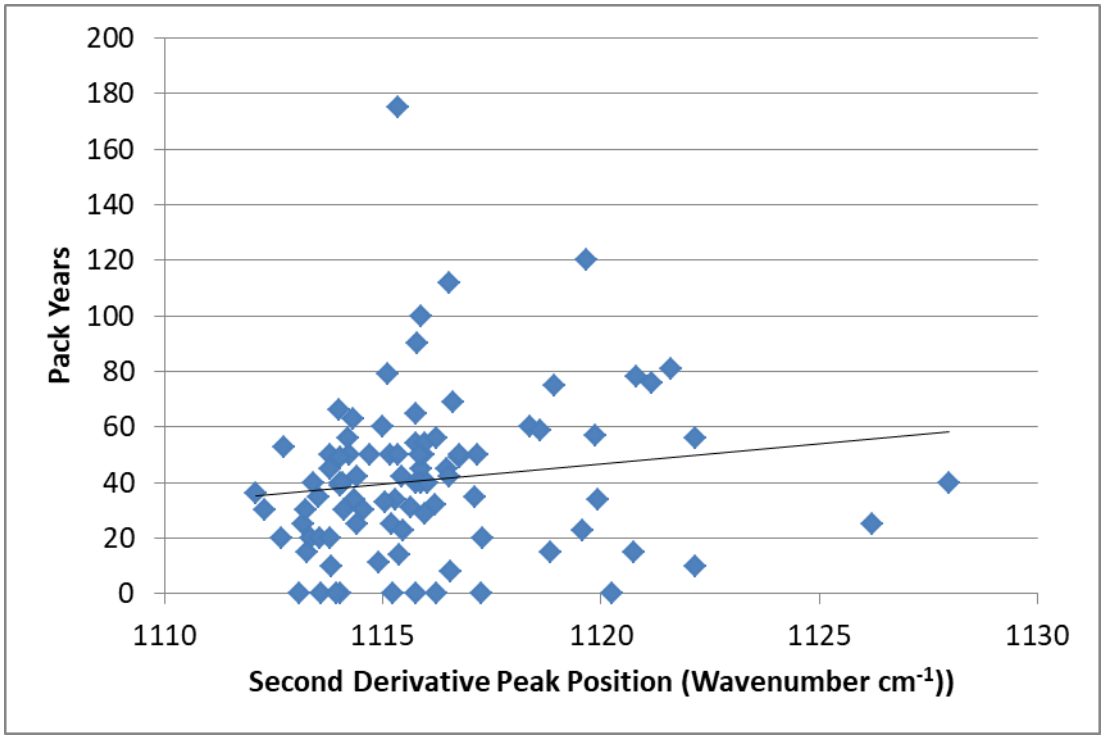


**Figure 3-59: A shift in peak position from around 1115cm$^{-1}$ towards higher wavenumbers and pack-years in all COPD patients was found to have a small positive correlation (rho = 0.3, p = 0.0041)**

A another small but significant (rho = -0.233, p=0.026) negative correlation was found between pack-years and the wavenumber position of the second-derivative

peaks found around 1019cm$^{-1}$ (Figure 3-60), although examination of the scatterplot suggests that the significance may be influenced by the presence of outliers.

Indeed, once the smoking pack-years outliers have been removed, the correlation scores between pack-years and second-derivative peak position drops from 0.3000 to 0.2014, and from -0.2330 to -0.1480 for peak positions around 1115 and 1019cm$^{-1}$, respectively. Neither of these new correlation coefficient scores were deemed to be statistically significant at the 95% confidence level.

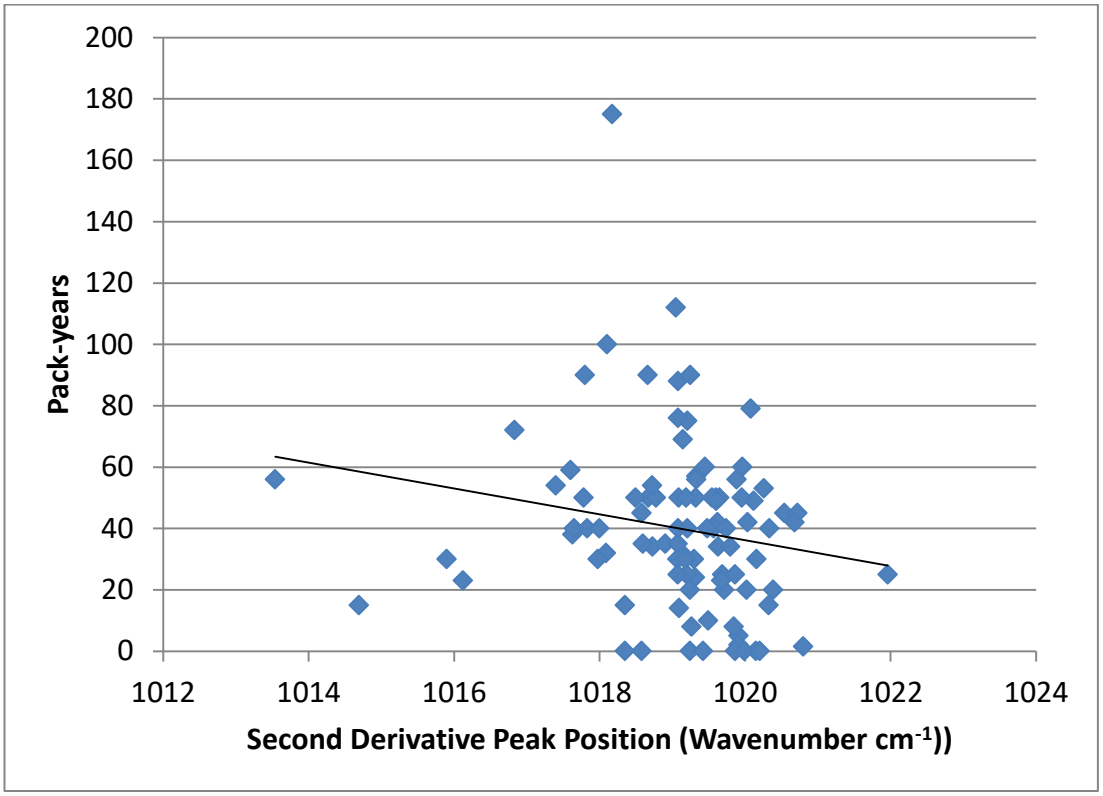


**Figure 3-60: A shift in peak position from around 1020cm$^{-1}$ towards lower wavenumbers and pack-years in all COPD patients was found to have a small negative correlation (rho = -0.233, p = 0.0026)**

### 3.3.4.4 Occupation

Certain occupations have been linked to an increased risk of COPD. Occupations where this risk is highest are those where the worker is exposed to dusts, fine particulate matter, volatile organic chemicals and polycyclic aromatic hydrocarbons (Health and Safety Executive, 2014). Such occupations include, but are not limited to, those in construction, mining, textiles and farming.

Patients were grouped according to their occupational risks for COPD, into high-occupational-risk and low-occupational-risk. All patients who were identified as

having worked in a high-risk environment, such as construction, mining, or farming, were classified as high-occupational-risk (n = 76). Further subclassification was not possible as there was no information regarding the/any specific risks each patient faced at their place-of-work; for example the load of the exposure during work, or length of time the patient was exposed to the pollutant. Low-risk occupations, such as clerical or office staff were placed into the low-risk-occupation cohort (n = 56).

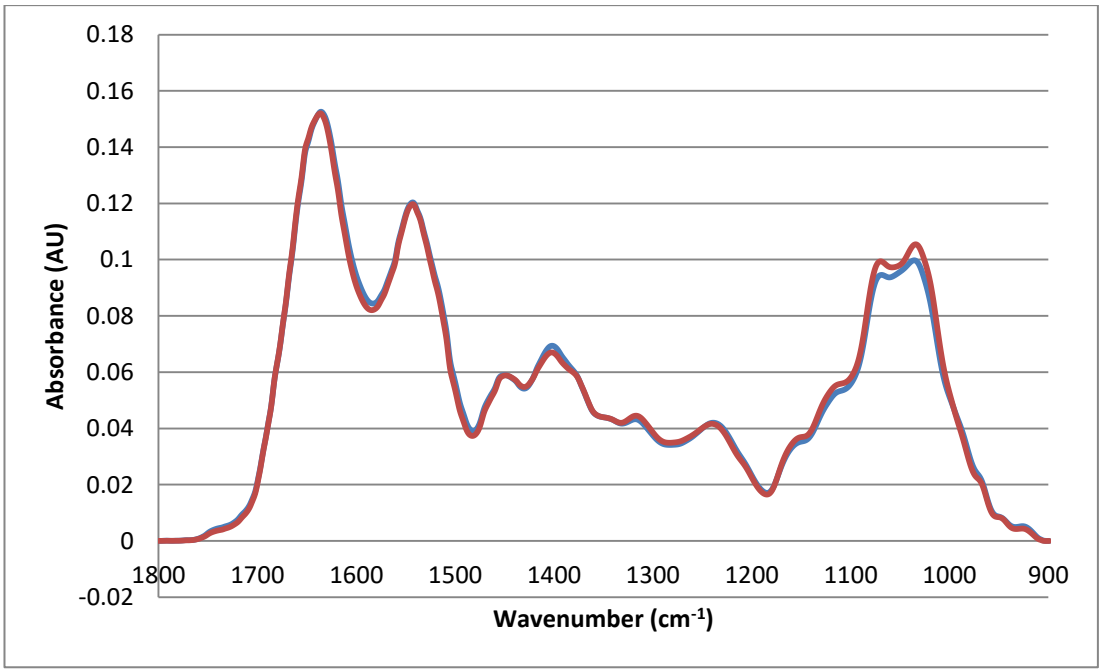


**Figure 3-61: Vector-normalised, baseline-corrected average spectra of the higher-risk job cohort (red) and the lower-risk job cohort (blue). A significant increase in absorbance can be seen across the glycogen-rich region**

Statistically significant differences in absorbance from 1132 to 1103cm$^{-1}$ and from 1082 to 1076cm$^{-1}$ were detected (Figure 3-61). These wavenumbers are important due to their association with modifications to glycan moieties in sputum mucins. 1080cm$^{-1}$ is a major glycosylation peak, associated with C-O stretching moieties within the glycan pentose and hexose rings, and 1116cm$^{-1}$ has been associated with sulphated and sialylated Lewis antigen structures, which are indicative of inflammation (Delmotte *et al.*, 2002; A. T. Lewis *et al.*, 2013). This suggests that an occupation which places the patient at higher risk of COPD may contribute towards increased modification of the mucin glycan structures by Lewis antigens.

### 3.3.4.5 Asbestos Exposure

Asbestos is known to be a major cause of lung disease, eventually causing asbestosis in exposed individuals. It is thought that exposure to fine particles, like

asbestos, could cause exaggerated lung tissue damage in individuals with impaired MCC activity, like COPD patients (Seibold *et al.*, 2011).

Patients were classified into cohorts of "non-exposure" (n = 82), or "exposure" (n = 47) according to self-reported exposure to asbestos at any time in their life. Three patients provided "not-known" for this question and were excluded from the analysis. Asbestos exposure was not associated with any statistically significant change in peak position or absorbencies at any wavenumbers. This is despite the average spectra showing a relative decrease in absorption within the amide I region, coupled with a slight increase in the glycosylation rich region, similar to the spectral effects seen in sputum from infected COPD patients (Figure 3-62).

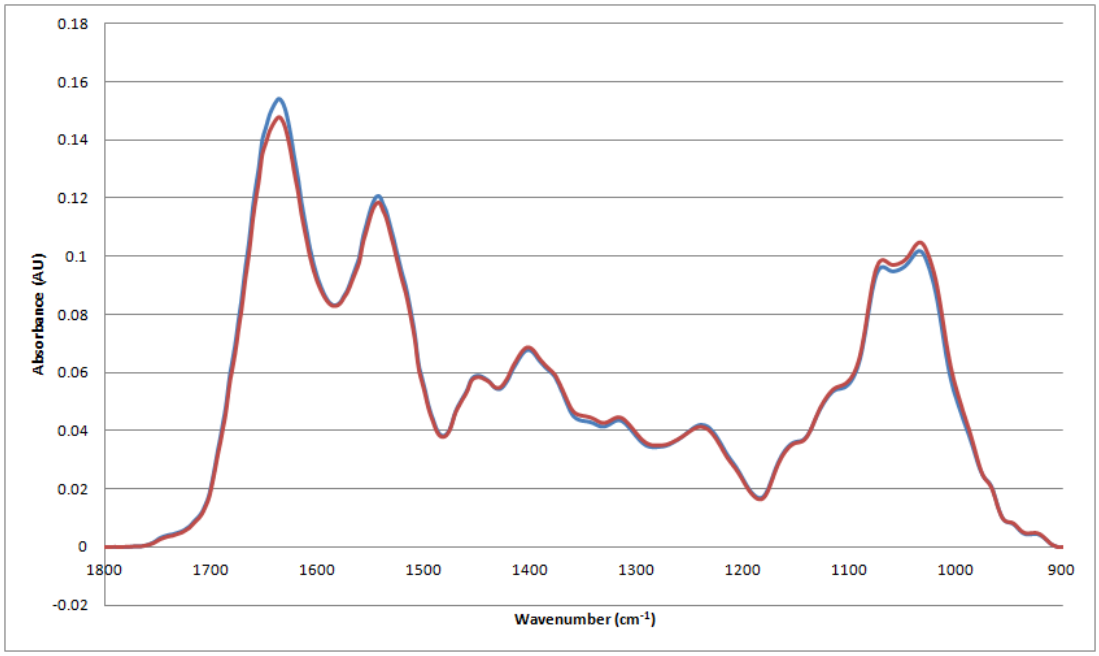


**Figure 3-62: Vector-normalised, baseline-corrected average spectra of sputum from COPD patients who have been exposed to asbestos (red) and those who have not been exposed (blue). Small differences in absorption can be seen within the glycosylation-rich region, along with a larger, but not statistically significant ($p > 0.05$), decrease in absorption at approximately 1630cm$^{-1}$, the protein-associated amide I region.**

### 3.3.4.6 Lung Function

Lung function was assessed by spirometry, and quantified using FEV1 (%), which is the gold standard for assessing lung function in COPD patients (Pauwels *et al.*, 2001). Unfortunately COPD stage information was not available due to the patients only providing a single FEV1 score at time of recruitment. The patients' FVC scores, which were required for the FVC/FEV1 ratio, a crucial component for estimating COPD stage alongside FEV1 (%) – see chapter 1 for a more detailed

description on COPD stages – were not collected by the recruitment teams. Historical FVC/FEV1 ratios, even where available, were not deemed to be appropriate for use in this work as COPD is a progressive disease and it was likely that the patients' lung function could have declined since their most recent spirometry session.

FEV1 (%) scores were correlated against second-derivative peak positions, which indicated two regions of wavenumbers where the second-derivative peaks positions demonstrated statistically significant correlation with the FEV1 (%) value. The median peak positions indicated were 1279cm$^{-1}$ (rho = 0.262, p = 0.045) (Figure 3-63), and 922cm$^{-1}$ (rho = 0.24, p = 0.0149) (Figure 3-64).

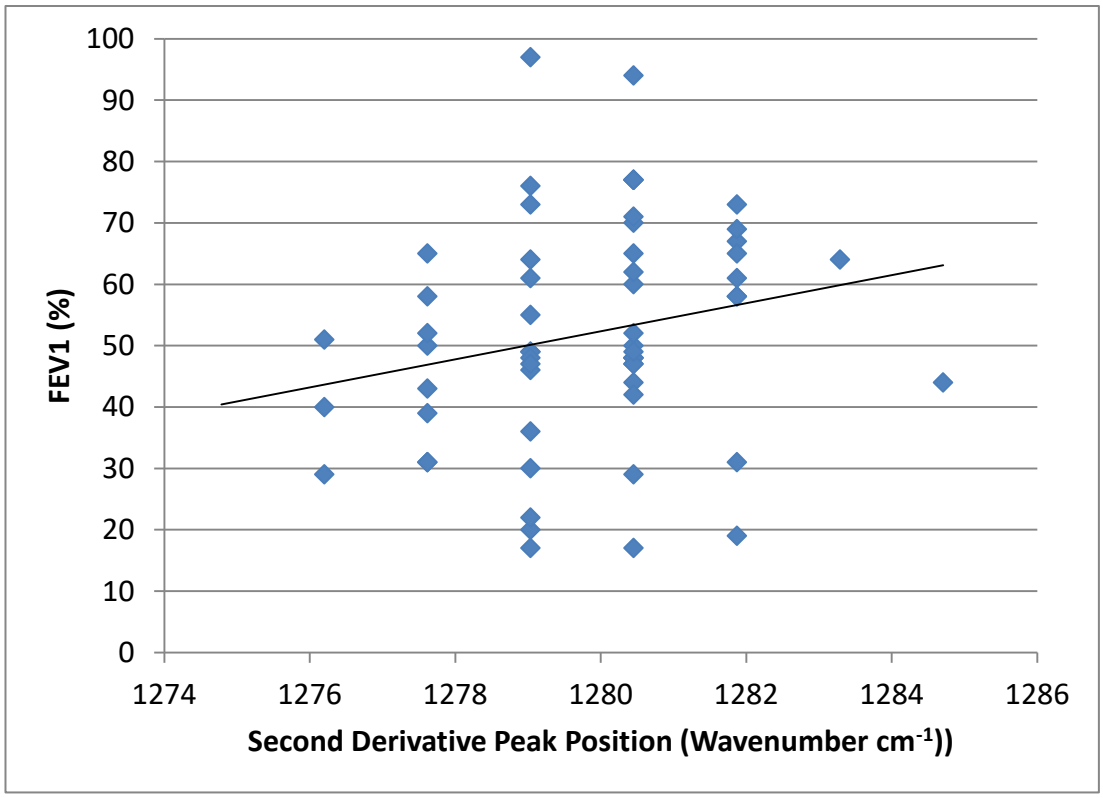


**Figure 3-63: Correlation plot showing a small but significant correlation between FEV1 (%) and peak shift from approximately 1276 to 1284cm$^{-1}$ as FEV1 (%) increases (rho = 0.262, p = 0.045)**

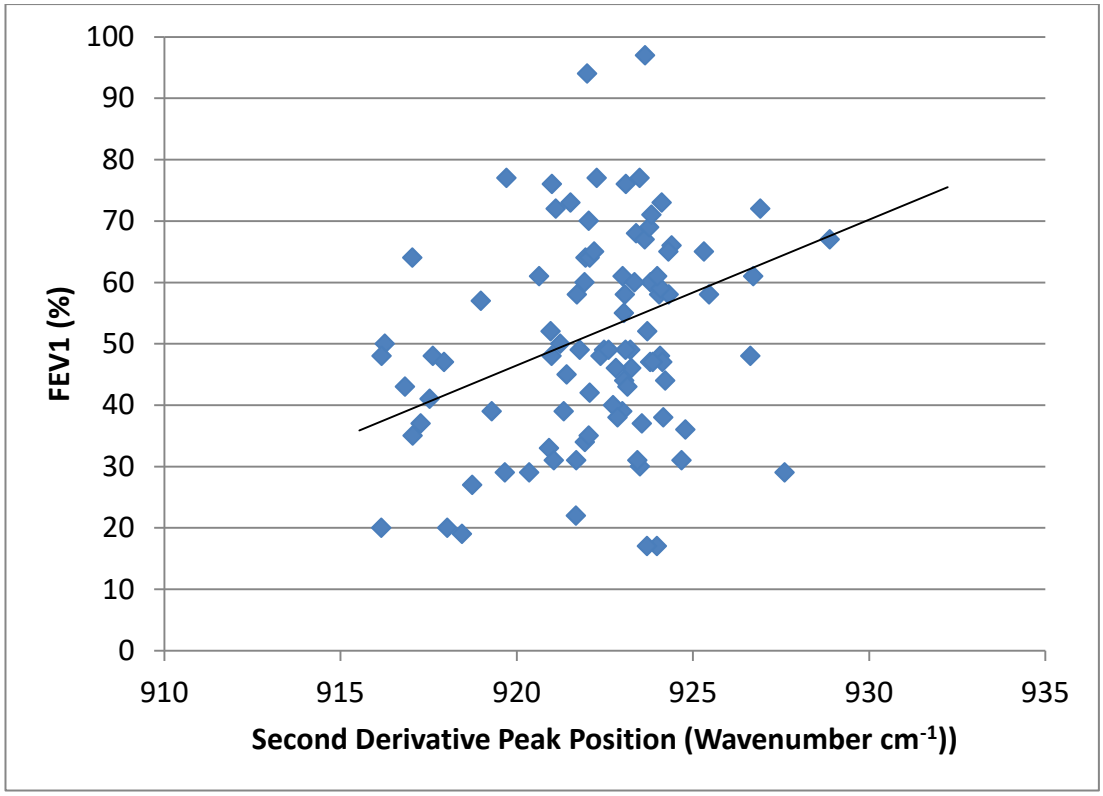


**Figure 3-64: Correlation plot showing a small but significant correlation between FEV1 (%) and peak shift from approximately 916 to 928cm$^{-1}$ as FEV1 (%) increases (rho = 0.24, p = 0.0149)**

### 3.3.5 SPEDIC Longitudinal Study

The data shown so far from the MEDLUNG study have not conclusively shown how FTIR spectroscopic analysis could differentiate COPD exacerbation from baseline. Moreover, merely differentiating exacerbation from baseline is not necessarily clinically useful, as each exacerbation causes a non-fully reversible worsening of symptoms and leads to COPD progressing (Seemungal, Donaldson, *et al.*, 2000). It is therefore more expedient to predict when an exacerbation is likely to occur, prior to onset of symptoms and hospitalisation as patients who are deemed to be high-risk for exacerbation can receive targeted preventative therapies (National Institute for Health and Care Excellence, 2018). Indeed, the COPD patients recruited during the MEDLUNG study were recruited as a "snap-shot" of their disease, with little information provided about the patient's exacerbation status, only whether or not they were currently exacerbating. One such example of missing information is the time since the most recent and time to the next exacerbation episode; exacerbations are known to occur in temporal clusters, with an initial exacerbation increasing the likelihood of a subsequent exacerbation within 8 weeks (Hurst *et al.*, 2009).

SPEDIC was a longitudinal study performed during the completion of this thesis, which recruited 55 COPD patients for monitoring over the period of 12 months which aimed to address some of the issues faced by MEDLUNG. Patients provided 5 sputum samples per week over the course of the study which were then analysed by FTIR spectroscopy. Patients were regularly assessed by healthcare assistants and data regarding the patient's state of health and exacerbation status were gathered.

The main aim of the SPEDIC study was to monitor the patient's disease as it progressed over time, and identify IR spectral features which could be used to predict exacerbation prior to reporting of symptoms by the patient or healthcare assistants. Identification of such IR spectral features would be invaluable for assessing a patient's disease state and could inform therapy choice. Through close monitoring of the patients by clinicians throughout the study, it was very clear when a patient was truly at their baseline and when they were exacerbating and at what stage they were during the exacerbation. In this way, SPEDIC has more clarity pertaining to the patient's current state of COPD than the MEDLUNG study described earlier in this chapter. With samples being available from each patient on most days, it will be possible to closely track their disease and identify biomarkers of exacerbation before the symptoms may become apparent to either the patient or the clinician and therefore, before the exacerbation can be diagnosed and treated. Identification of these markers will allow the classification of another subgroup of COPD patients – the pre-exacerbation/high-risk group.

#### 3.3.5.1 Inflammatory Quotient Time-Series Plots

Inflammatory markers are known to be elevated at exacerbation, and COPD disease progression is well linked to inflammatory markers in sputum (Parr, White, Bayley, Guest, & Stockley, 2006; Seemungal *et al.*, 2001). Additionally, airway inflammation has been concluded to be a driver for sputum heterogeneity (Whiteman *et al.*, 2008). Some markers of inflammation in the lung such as sialic acid, or sulphated Lewis antigens have been previously characterised by FTIR analysis and wavenumbers positions of peaks associated with these structures have been identified (Kačuráková & Mathlouthi, 1996; Khajehpour *et al.*, 2006; A. T. Lewis *et al.*, 2013; S. P. Lewis *et al.*, 2013; Wiercigroch *et al.*, 2017). Shown in appendix 3 are time series plots showing a quotient value (Q-value) calculated from a series of inflammation-associated wavenumbers of each spectrum for every patient. These Q-values are classified by the point in the baseline-exacerbation cycle the patient was

currently in, as defined by the reported health status data collected during the study. The categories are true baseline (C), currently exacerbating (E), pre-exacerbation being within 7-days prior to the first reported exacerbation day (P), and post-exacerbation high-risk period being within 8 weeks from the last reported exacerbation day and not within 7 days prior to the next reported exacerbation (H). It was hypothesised that due to the inflammatory nature of COPD exacerbations, IR absorption at inflammation-associated wavenumbers would be increased, relative to absorption at wavenumbers associated with protein and glycan structure. The Q-values, were calculated using the following equation:

$$Q = \left( \frac{\left( \frac{\sum \left( \log \sqrt{Abs(1670^2, 1589^2, 1377^2, 1168^2, 1103^2, 1054^2, 1035^2, 1020^2, 997^2, 986^2, 967^2)} \right)}{11} \right)}{\left( \frac{\sum \left( \log \sqrt{Abs(1655^2, 1543^2, 1078^2)} \right)}{3} \right)} \right)$$

**Equation 1: Calculation of inflammatory Q-value, where Abs(X) is equal to the absorbance value at the specified wavenumber in $cm^{-1}$. Each absorbance value is squared and square-rooted to give the absolute values. The log of each value is then summed and an average is taken. The inflammation-associated wavenumbers are normalised by division of the overall mucin contribution.**

The inflammatory Q-values were plotted in chronological order, and each sample classified according to the categories described. Analysis of the plots was carried out, looking for patterns among the Q-values that were prescriptive of an exacerbation occurrence within 7 days. A five-point moving average trend line was used to visualise overall trends occurring throughout a week, as the patients provided a maximum of 5 samples per week. Absolute rate-of-change of the inflammatory Q-values from sample-to-sample was calculated to quantify the degree of change between each sampling day. Where consecutively collected samples were collected with a multiple day gap, differences between sample spectra were averaged out by dividing by the number of days between each sample.

#### 3.3.5.2 Predicting Exacerbation from Inflammatory Q-Values

The inflammatory Q-values were plotted against the patient's sample classifications to identify any recurring patterns of the Q-values that may occur in the days prior to the first reported day of exacerbation. The plots are shown in appendix 3. Each point on the graphs represents a single sputum sample from a single day. For

example, each exacerbation episode is by a sequential series of "E" markers, with each marker representing a single sputum sample received during the exacerbation episode.

As the inflammatory Q-value is composed of absorbencies derived from wavenumbers which have been associated with an increased absorbance due to the presence of molecular structures associated with inflammation, it was hypothesised that an increase in an inflammatory Q-value may precede an exacerbation event. Additionally, as COPD is a chronic condition associated with sustained inflammatory responses, each patient was hypothesised as having a "normal" or "acceptable" level of airway inflammation which would not constitute an exacerbation. Therefore the sequences were interrogated for an increase in Q-value, which surpassed the average level of inflammatory Q-value from all previous non-exacerbation samples for that patient, and was sustained over at least two days prior to the first reported day of exacerbation.

The number of days prior to the first reported day of exacerbation where a positive prediction was made were plotted to visualise the distribution. The mode was found to be 10 days, with a higher proportion of positive predictions being made within 15 days of the first reported day (Figure 3-65).

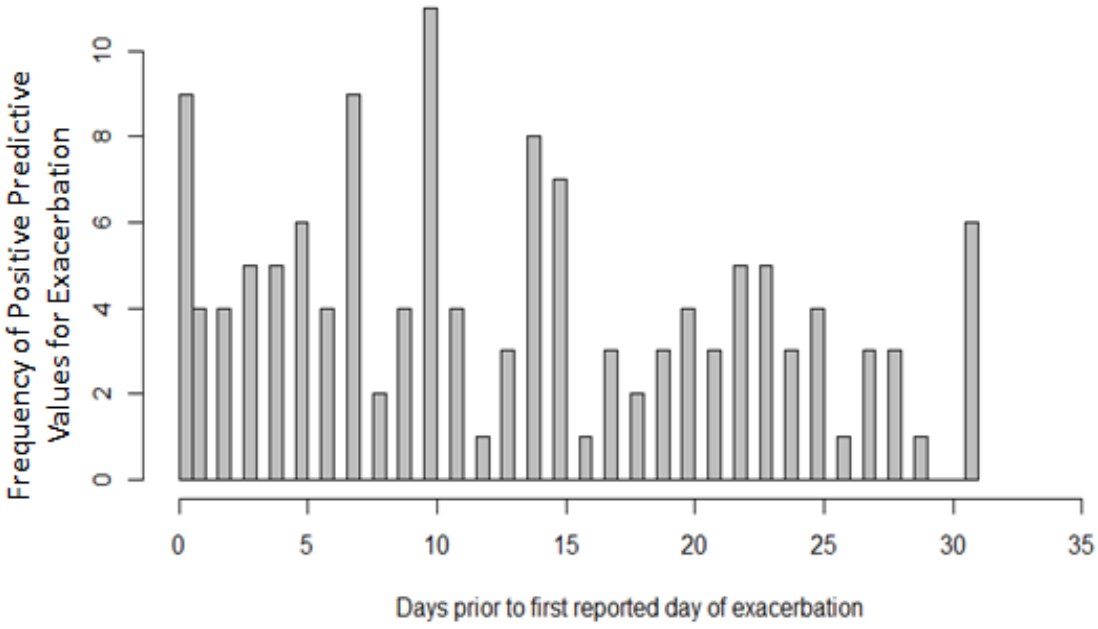


**Figure 3-65: Histogram showing the distribution of positive predictive values in the days leading up to the first reported day of exacerbation. A greater distribution of**

**positive predictive values can be seen within 15 days of the first reported day of exacerbation onset.**

Throughout the course of the study, 56 individual exacerbation episodes were recorded by the patients and a positive prediction was made within 15 days of the first reported day of exacerbation 36 times (60.71%). However, a correct positive prediction was made a total of 578 times in 4204 individual samples. Discounting multiple predictions for the same exacerbation episode, an exacerbation episode prediction was made 141 times. This suggests 1) a high false positive rate for predicting exacerbation, or 2) the possibility of the patients not reporting all of their exacerbations, possibly due to being accustomed to short-term changes in their symptoms which may mask an exacerbation, as has been reported previously (Seemungal *et al.*, 2009).

## 3.4 Discussion

### 3.4.1 MEDLUNG

The major aim of this chapter was to further the work previously undertaken within the Respiratory Diagnostics Group at Swansea University, by extending the number of COPD patients analysed. An initial FTIR analysis was carried out on 89 COPD baseline and 43 COPD exacerbation sputum samples. The first aim was to replicate and expand the work carried out by Aaran Lewis on the small subset of COPD sputum samples, by increasing the numbers of patients analysed. However, it was not possible to replicate the results from the previous work, with multivariate analysis and three-dimensional scatterplots being unable to distinguish baseline COPD sputa from exacerbations. This is most likely due to the small patient numbers used in the original work compared to the larger dataset used in this current work. In order to fully ascertain if COPD baseline could be distinguished from exacerbation using FTIR spectroscopy of raw sputum, a more in-depth analysis of the spectra was required. Deconvolution of the spectra was carried out by second-derivative analysis to increase the resolution of secondary peaks and absorbance distributions which were obscured by the major amide I, amide II and glycogen-rich bands.

Second derivative peak analysis highlighted two peak positions which were found to be statistically significantly different between COPD baseline and exacerbation sputum. A statistically significant peak shift from approximately 1115cm$^{-1}$ in baseline samples to approximately 1116cm$^{-1}$ was detected, possibly indicating a

change in the sulphation of the glycan residues. The second statistically significant (two-fold) peak shift was from approximately 1376cm$^{-1}$ in exacerbation samples to wavenumbers within the range of 1372 to 1380cm$^{-1}$ in baseline samples. The majority of exacerbation samples displayed a peak at 1376cm$^{-1}$ but this was less common in baseline patients, who displayed a greater distribution of peak positions. A peak being present at 1376cm$^{-1}$ is not itself enough to be indicative of exacerbation, however presence of a peak at the wavenumbers from approximately 1372 to 1380cm$^{-1}$ with exception of 1376cm$^{-1}$ may be indicative of the patient not currently having an exacerbation episode. No other statistically significant peak shifts were observed across the spectrum.

PCA analysis also struggled to differentiate baseline COPD from exacerbations, with little clustering of groups and no separation of groups from one another.

One possible reason for the difficulty in differentiating between COPD baseline and exacerbation is the innate variability of the disease. It is well known that there are multiple phenotypes of COPD, each requiring its own distinct course of treatment and management (Mirza & Benzo, 2017). Therefore, an in-depth analysis of the COPD patient cohorts was performed by examining the measured physiological characteristics of each patient to ascertain if they could be a source of inter-patient variation within each cohort, and could therefore be controlled.

Patients were categorised according to their personal and clinical information and statistical associations between the spectra and physiological data were sought. High-risk occupations, increased smoking status and infection were all correlated with an increase in absorption at the sulphate-associated wavenumber 1116cm$^{-1}$, or peak shift around 1116cm$^{-1}$. This supports the hypothesis that a patient's personal and physiological factors can influence their sputum's molecular characteristics, which are detectable in the FTIR spectra. These factors should be controlled for when making determinations about COPD patients and exacerbation detection.

### 3.4.2 SPEDIC

The SPEDIC study was a year-long longitudinal study of 55 COPD patients during which the patients provided five samples of sputum per week, and their current exacerbation status was recorded. FTIR analysis was carried out on the sputum samples in order to compare the molecular structural information to the patient's current state of health, and to generate predictive models for exacerbation.

The previous MEDLUNG study was limited by the lack of prior and follow-up information provided for each patient, so it was unclear if any of the baseline patients had experienced an exacerbation within the previous eight weeks, or if they had an exacerbation within a short period of time of recruitment. This meant that it was unclear if any pre- or post-exacerbation biological alterations to the mucins were present which may have increased the variation within the baseline cohort.

The FTIR spectrum of each sputum sample was converted into an inflammatory Q-value, which was based on the contributions of a panel of inflammation-associated wavenumbers, normalised against overall mucin content. Patterns in this inflammatory Q-value over time were sought by interrogating each patient's time series with a predictive algorithm based on the hypothesis that inflammation-associated structures may become more prevalent in sputum in the days preceding exacerbation, and a concordant increase in inflammatory Q-value may be observed. The predictive algorithm was found to successfully predict approximately 60% of all individually reported exacerbation episodes; however the number of false positive predictions was very high. This could indicate that the measure used may not accurately reflect the true nature of inflammation within the COPD lung. A follow-up study during which physiological measurements of inflammation are taken over time and the correlated to FTIR spectra of sputum would provide insight into how inflammation effects the sputum molecular structure in 'real-time'.

An unexpected outcome of the SPEDIC study was an apparent overall reduction in exacerbations for all patients. Out of 55 patients, only 26 patients recorded at least one exacerbation episode during the whole study period. This was confirmed as each patient was visited by a designated healthcare assistant at least twice per 7 day period, and each reported exacerbation was confirmed by either the healthcare assistant or, in the case of patients who were hospitalised due to exacerbation, the patient's medical notes upon admission to hospital.

Additionally, fewer hospitalisation days due to COPD exacerbation were recorded during the study (280) for all patients compared to the previous year (652). This was also surprising, but through consultations with the clinical team who met with the patients, it became clear that many of the patients were previously unaware of how to manage their condition in a proactive, rather than reactive manner. This suggests that the patients became more educated regarding their disease. The clinical

team reported that the patients also showed a change in perception of their disease, as they started to actively engage with their disease and take steps to avoid progression towards exacerbation. Whilst this is clearly a confounding factor for the SPEDIC study and the work presented within this thesis, it is inarguably a highly positive outcome for the patients involved.

This has led to the hypothesis that through increased engagement of COPD patients with healthcare professionals, and ensuring that they are considering their disease more carefully, a reduction in COPD exacerbation admissions to hospitals could be achieved, along with an associated reduction in COPD-associated costs to the healthcare system. As this hypothesis is based on a single study of a low number of patients over a single year, it is important to consider that it is possible the reduction in exacerbation days is due to chance alone; however the finding presented in this work is promising and warrants further investigation with larger patient cohort in a multi-centre study over a longer timeframe.

## 3.5 Conclusions

FTIR is a rapid, low-cost and non-invasive technology, capable of reliably providing detailed molecular structural information about a sample of interest. The work presented in this chapter adds to the already large volume of work demonstrating the utility of FTIR spectroscopy for biofluid analysis. Reproducible and reliable spectra were acquired for 132 COPD patients using a simple protocol involving no sample preparation, and detailed structural information was obtained for each of these samples. However, it has also shown that FTIR spectroscopy may not be suitable for COPD exacerbation detection. One two-dimensional linear model based on second-derivative peak positions demonstrated a capability to determine exacerbation from baseline samples with 80% sensitivity and 48% specificity in sputum samples which contained detectable peaks within the regions of interest.

Additionally a longitudinal study of COPD patients which aimed to use FTIR to monitor COPD patient sputum samples was not able to generate a generalised predictive model for COPD exacerbation prediction. One of the main reasons for this is hypothesised to be that the limited number of patients enrolled in the study was not enough to overcome the inter-patient variability which is innate to COPD. A new, larger cohort study, during which the patient's physiological factors are regularly measured, would provide valuable insight towards how the inflammatory processes in

the respiratory system change over time, and how these changes cause alterations to the biophysical properties of sputum.

Engagement of COPD patients by healthcare assistants appears to reduce the healthcare service requirements of these patients. This hypothesis is based on the observation of reduced hospitalisation days due to exacerbation during the SPEDIC study compared to the year before, but is based only on small number of patients. Further work is required to fully establish the relationship between COPD patient engagement by healthcare professionals and subsequent utilisation of healthcare service resources by the patients.

# Chapter 4

# Evaluation and Development of Hand-held Attenuated Total Reflectance Variable Filter Infrared Spectroscopy for Potential Use by Clinicians and Patients to Manage COPD

## 4.1 Introduction

### 4.1.1 Introduction to ATR-VFIR

Attenuated total reflectance variable filter infrared spectroscopy (ATR-VFIR) is a relatively new technology within the confines of IR spectroscopy. It is based around the use of a linear variable filter (LVF) to split the IR light into its constituent wavenumbers. The filter is laid on top of a photodetector array, which converts the IR radiation into an electrical signal to be interpreted by a computer – a schematic drawing in Figure 4-1 shows a basic layout of a VFIR spectrometer with LVF detector. These filters are of a lower cost and are significantly smaller in size than the interferometers found in conventional FTIR spectrometers. They also have another advantage in that they contain no moving parts, as the beam can be collimated over the whole LVF through the use of stationary optics (Emadi, Wu, De Graaf, & Wolffenbuttel, 2011).

Miniaturised spectrometers have lower-quality optics than their benchtop counterparts (Wolffenbuttel, 2005), which can lead to lower-quality spectra being generated. However, it has been shown that miniaturised spectrometers are capable of highly-specific discrimination between molecular species, giving comparable results to benchtop spectrometers (Bentini *et al.*, 2006; Unger, Pfeifer, & Siesler, 2016).

The development of optical spectrometers with LVF array detectors has led to an increase in miniaturised, hand-held, and portable spectrometers. One commercially, available miniaturised IR spectrometer is the IR Sphinx (Spectrolytic GmbH), originally designed for oil quality analysis. This chapter will explore and evaluate the IR Sphinx with respect to its capabilities for analysis of more complex biofluids and its suitability for the clinical environment.

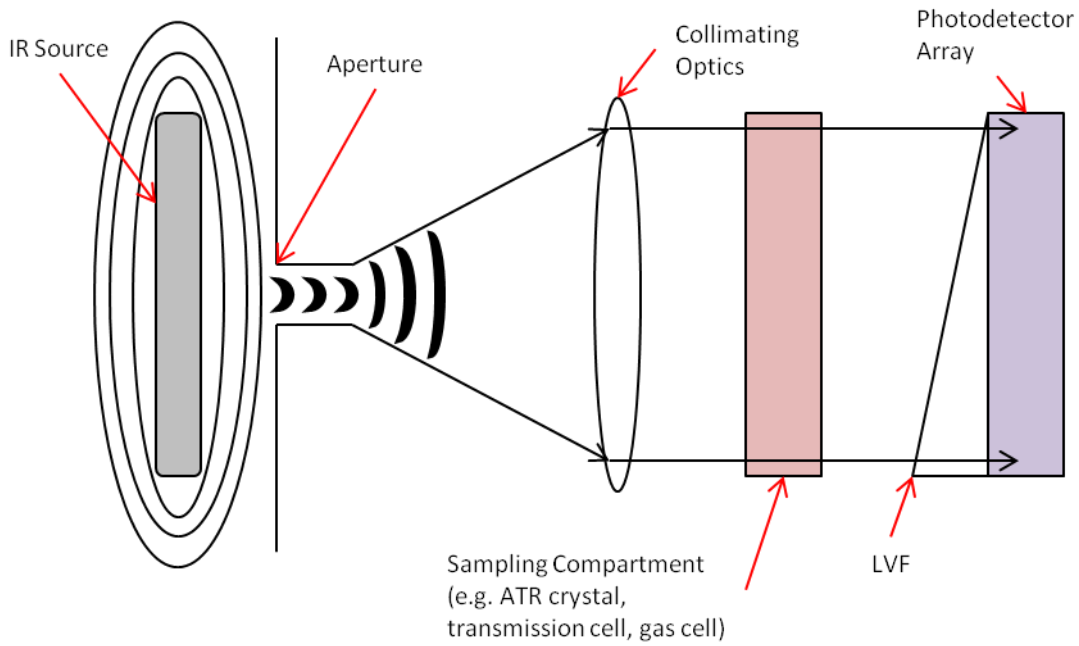


**Figure 4-1: Schematic drawing of a VFIR spectrometer, showing an IR source emitting IR radiation through a slit and collimating optics, which focus the beam onto the LVF filter and detector array.**

### 4.1.2 Development of a Sputum Sampling Strip for ATR-VFIR Analysis of Sputum

#### 4.1.2.1 Rationale and Key Criteria

A key component of the IR Sphinx ATR-VFIR system is the use of a large zinc sulphide (ZnS) ATR-crystal. This crystal has a lower-cost than diamond, but still retains a high hardness and so is resistant to scratching during sampling and cleaning. However, when sputum is applied to a surface and allowed to dry, it becomes extremely hard and is difficult to remove without soaking and scrubbing. Any potential device that requires a labour-intensive cleaning protocol after each use will not be suitable for clinical use. Additionally, whilst the ZnS crystal is resistant to scratching, with heavy use in the clinical environment, with potentially many uses throughout the day, there is a risk that the device could become damaged if sputum is being directly applied. For these reasons, a low-cost, low-IR-absorbing, disposable sampling-strip is required to be used to apply the sputum, which would then be applied to the spectrometer by a clamping mechanism.

For the sampling-strip to be suitable for IR analysis, it must conform to certain criteria:

1) The IR substrate must not interfere with the sputum spectrum through addition of spectral artefacts due to absorption, or scattering by the substrate.
2) The substrate must not react with sputum, forming new molecular interactions which would introduce additional peaks into the spectrum.
3) The sputum must adhere strongly to the substrate and not be removed during the spectrum acquisition process.
4) All of the components used in the manufacture of the sampling-strip must be of a low enough cost to ensure that using the sampling-strip as a disposable item is feasible.
5) The strip must be able to be clamped in place against the ATR crystal without damage occurring to the strip or device, whilst ensuring that there is consistent contact between sample and crystal.

#### 4.1.2.2 Construction of the Sample Strip

As previously stated, there are certain criteria to which the sampling strips must conform, and therefore the construction of the strips must also meet them. Shown in Figure 4-2 is a basic schematic describing how the individual components of the strip are arranged. The hard plastic backing acts as a solid base to which the other components are attached. Additionally the backing is used for handling the strip. The low-IR-absorbing substrate is the surface onto which sputum is spread and allowed to dry. This substrate is laid over a compressible tape, which is present to cushion the substrate between the plastic backing and ATR crystal, and to evenly distribute the clamping force across the whole strip.

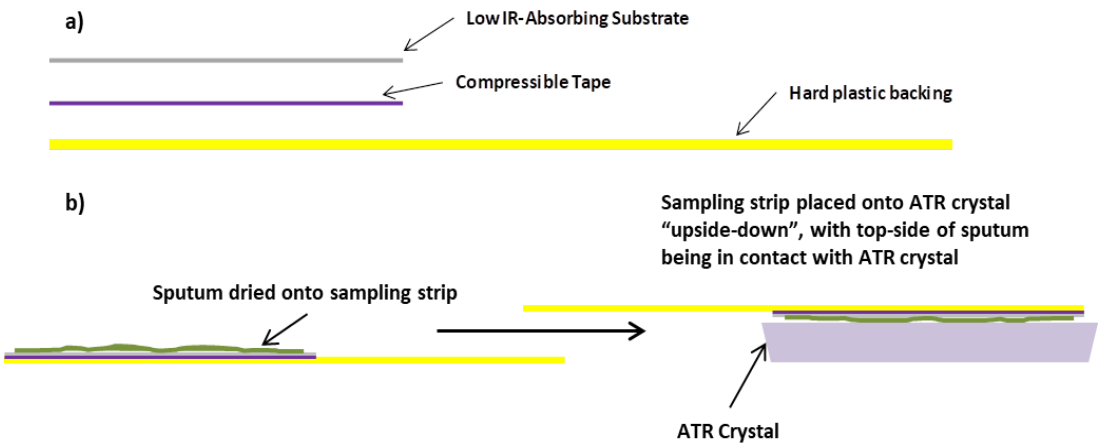

**Figure 4-2: (a) Schematic of sampling strip construction, and (b) sampling strip use, showing sputum dried onto the substrate surface and the strip being placed "upside-down" on the ATR-crystal. A clamping force maintains close contact between the sample and ATR crystal during measurement**

#### 4.1.2.3 Selection of Low-IR-Absorbing Substrate

Low-IR-absorbing materials are a crucial component of all FTIR substrates, and care must be taken to ensure that the chosen substrate allows maximum signal, with minimal substrate interference. Previously, ATR-FTIR has been performed using highly-IR-reflective low-E or gold-coated slides, but the cost of these slides is prohibitive for use within the clinical environment where substrates must be single-use (Cui, Butler, Martin-hirsch, & Martin, 2015). Aluminium oxide (AlO) foil has been shown to have a low absorption of IR in the mid-IR range, and is substantially lower-cost than conventional ATR substrates (Cui *et al.*, 2015; Vahur, Teearu, Peets, Joosu, & Leito, 2016). For these reasons, AlO was selected as a potential candidate substrate for the sampling strip and was subjected to further testing.

#### 4.1.2.4 Selection of Compressible Tape

ATR-FTIR spectroscopy requires excellent contact between the sample and ATR-crystal, therefore one of the chief components of the sampling strip would have to be a suitable material that would facilitate this contact. Additionally, the ATR crystal of the IR Sphinx is very large, (17mm*25mm = 425mm$^2$), thus requiring a large amount of the sample to be evenly applied to generate a spectrum with maximum signal and minimum noise. Applying the samples with a disposable sampling strip requires that the sample strip creates and maintains this excellent contact between the sample and ATR crystal. The need for a suitable material that would evenly distribute applied downward pressure through the sample onto the ATR crystal was therefore identified.

### 4.1.3 Chapter Aims

The primary aim of this chapter is to evaluate a miniaturised ATR-VFIR spectrometer, the IR-Sphinx (Spectrolytic GmbH). This spectrometer will be evaluated with respect to its utility as a clinical tool in the diagnosis of COPD exacerbation through sputum analysis, but this is not the major aim of this work. Comparisons of spectral quality and COPD exacerbation discrimination will be made between the IR Sphinx and the laboratory FTIR spectrometers used in Chapter 3. The IR Sphinx will also be assessed for its suitability for the clinical environment based on criteria as defined by consultations with clinicians.

As part of the clinical evaluation of the IR Sphinx, a hand-held sampling strip on which to apply sputum prior to IR measurements will also be developed. Different low-IR-absorbing substrates and backing materials will be tested, and comparisons will be made between spectra generated using the sample strip and those generated through direct application of the sputum to the ATR crystal. Spectral quality will be assessed through signal-to-noise ratio, and similarity index calculations and spectra will be examined for introduction of spectral artefacts caused by the sampling strip.

## 4.2 Methods

### 4.2.1 Spectral Generation

IR spectra were generated as described in Chapter 2, section 2.2.

### 4.2.2 Spectral Processing

All spectra were processed as described in Chapter 2, section 2.3.

### 4.2.3 Sample Strip Materials

Metallised-polyurethane foam was obtained from Coveris (Louth, UK). AlO film was obtained from Swansea University Welsh Centre for Printing and Coating (Swansea, UK). Compressible foams were obtained from TESA (Milton Keynes, UK) and 3M (Bracknell, UK).

### 4.2.4 Density Testing of Compressible Foam

The density of compressible foams used in the construction of the sample strip was calculated by measuring the dimensions of a known sample of foam using callipers, and measuring the weight using digital scales with a margin of error of 0.05g. Measurements were taken in triplicate and averaged.

## 4.3 Results

### 4.3.1 IR Sphinx Evaluation

A hand-held, portable and low-cost spectrometer, such as the IR Sphinx, capable of distinguishing molecular structural differences in complex biofluids has great potential to be useful within the clinical environment for rapid identification of disease biomarkers. The work outlined in Chapter 3 did not definitively show how high resolution FTIR spectroscopy could distinguish sputum samples collected from COPD patients with baseline disease or at time of exacerbation.

The IR Sphinx was evaluated for its ability to generate reliable spectra of sputum samples obtained from currently exacerbating COPD patients, and COPD patients currently at baseline. Sputum from COPD patients was used for this work because of the sample volumes required by the IR Sphinx were greater than which was previously required, especially when taking into account sample replicates, and COPD patient sputum was in much greater supply. Determination of baseline from exacerbation was not the primary aim of this work, rather the aim was to assess the reliability of the IR Sphinx, and develop protocols to improve the suitability of the device for a clinical application.

The spectra generated by the IR Sphinx were also compared to those generated by the Bruker Alpha, a benchtop, portable ATR-FTIR spectrometer to provide a benchmark.

#### 4.3.1.1 IR Sphinx Resolution Characterisation

First, IR spectra generated by the IR Sphinx were assessed for quality, by comparing spectral resolution to that of spectra generated by the Bruker Alpha, through comparison of isopropyl alcohol (IPA) spectra generated using the IR Sphinx under normal operating conditions, and the Bruker Alpha at decreasing resolutions from 2 – 256cm$^{-1}$ (Figure 4-3). A direct comparison of an IPA spectrum generated using the IR Sphinx and Bruker Alpha at 32cm$^{-1}$ shows a great deal of similarity, although the three peaks observable in the Bruker Alpha spectrum from 1500-1300cm$^{-1}$ are not individually resolvable in the IR Sphinx spectrum.

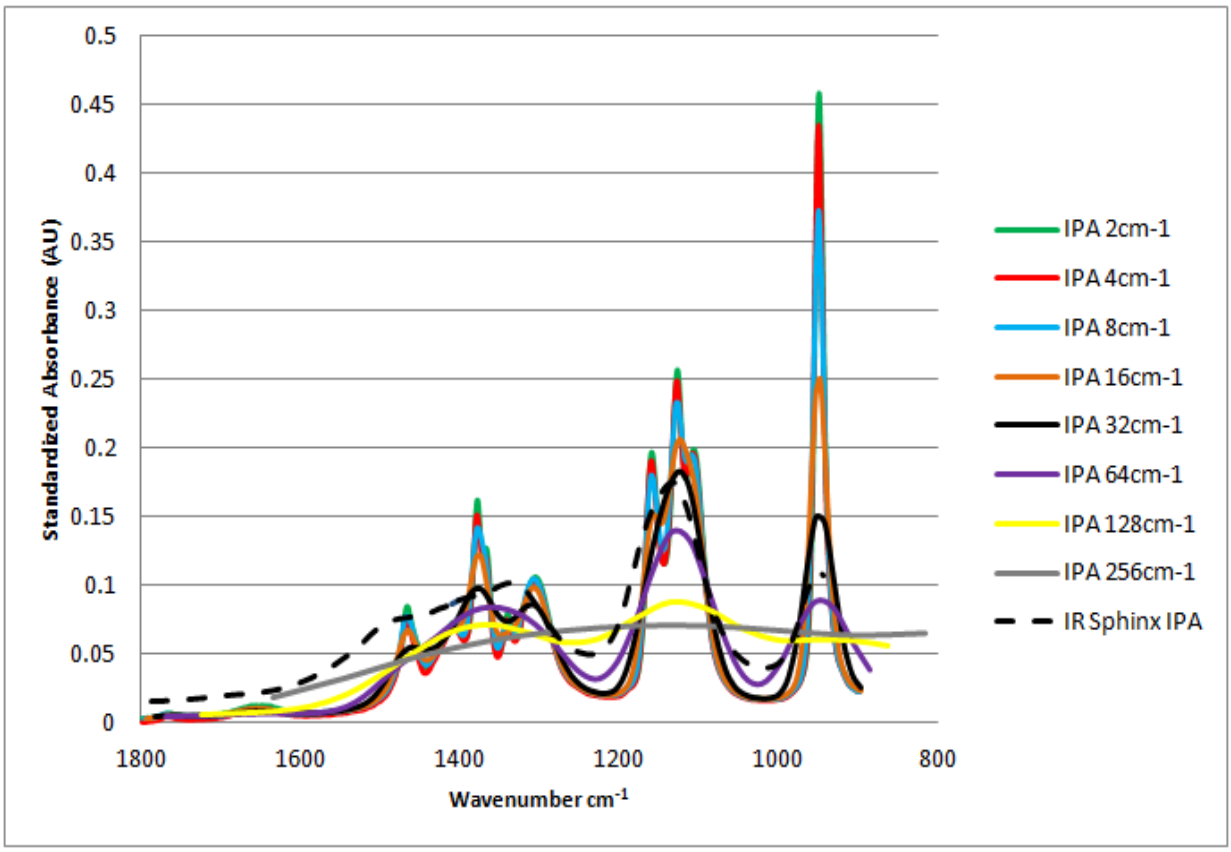


**Figure 4-3: A series of ATR-FTIR spectra of IPA acquired at a range of spectral resolutions decreasing from 2cm$^{-1}$ to 256cm$^{-1}$ using the Bruker Alpha (coloured, solid lines spectra), and an IPA spectrum acquired using the IR Sphinx (black dashed)**

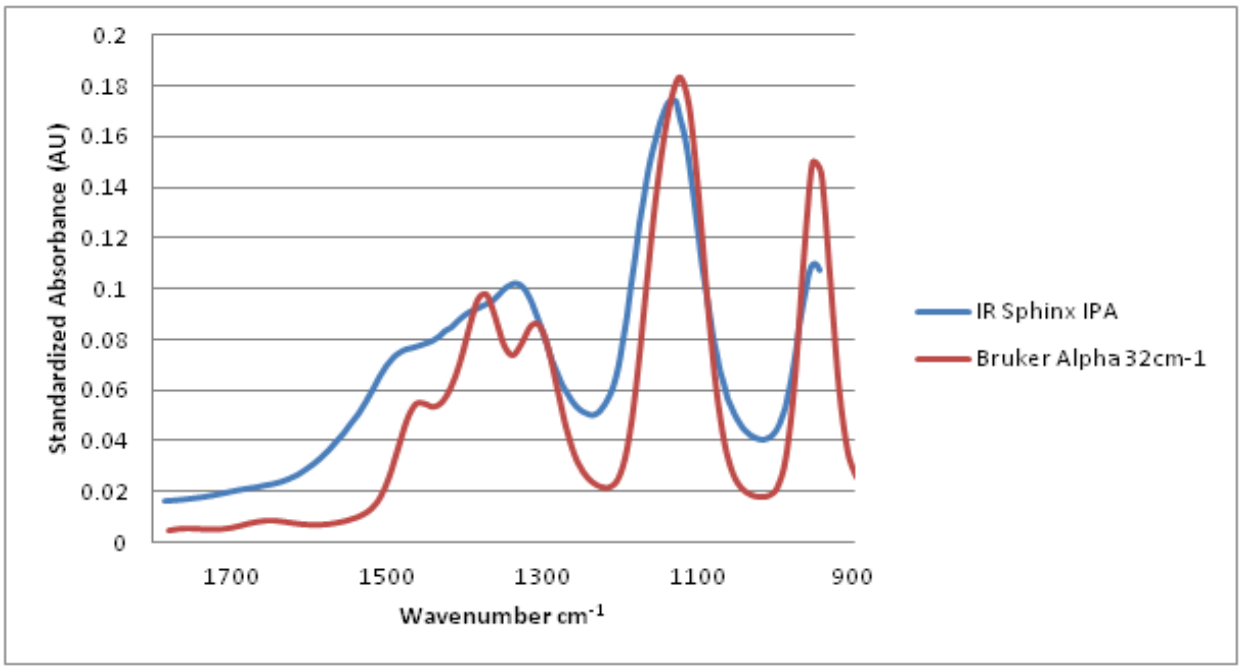


**Figure 4-4: A direct comparison of the Bruker Alpha IPA spectrum at 32cm$^{-1}$ resolution and the IR Sphinx IPA spectrum.**

Spectral resolution of the IR Sphinx was found to be lower than that of the Bruker Alpha, with little to no distinction of the peaks within the amide I and amide II regions, and no resolution of individual peaks through the glycogen-rich region. The overall resolution of the IR Sphinx was estimated to be approximately 30cm$^{-1}$ due to the similarity of the spectra shown in Figure 4-4.

#### 4.3.1.2 IR Sphinx COPD Sputum Spectra and Exacerbation Detection

COPD patient sputum ATR-VFIR spectra (n=129) were acquired using the IR Sphinx in triplicate. Figure 4-5(a) shows the average COPD baseline spectrum, with broad absorbance bands clearly visible at amide I and II, and at the glycogen-rich region. However it is difficult to resolve individual peak positions; for example, distinguishing the exact position of the amide II peak from the amide I peak is problematic due to the lack of trough between the two peaks which is characteristic of the region.

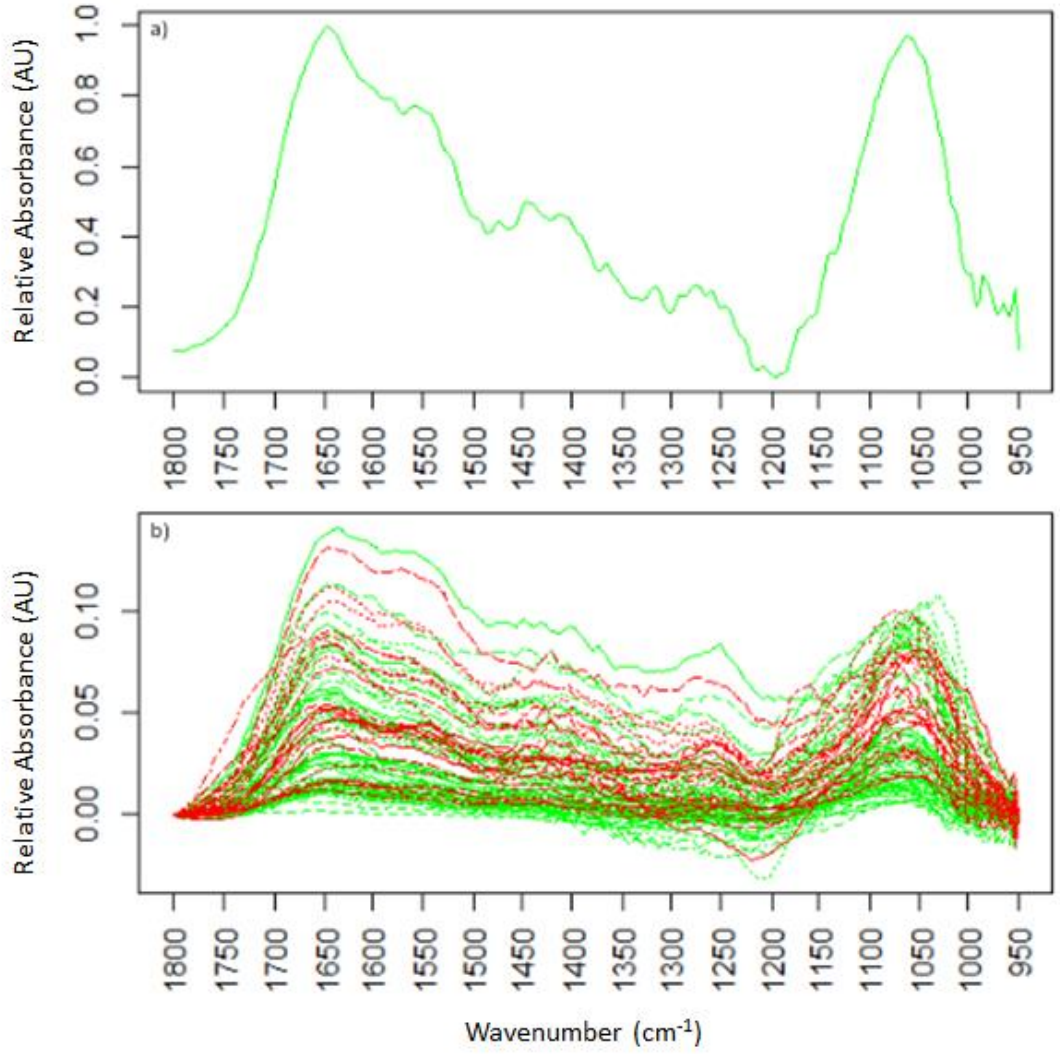

**Figure 4-5: Infrared spectra generated by the IR Sphinx of a) min-max normalized, baseline-corrected average COPD baseline spectrum, and b) all baseline-corrected, mean-normalized COPD patient sputum spectra (red – exacerbator, green – baseline).**

PCA was performed on absorbencies at wavenumbers shown to be statistically significant ($p < 0.05$) after multiple-hypothesis testing for determining baseline from exacerbation samples. The six most significant wavenumbers selected for PCA were 992, 1022, 1127, 1289, 1337 and 1700cm$^{-1}$ (Figure 4-6).

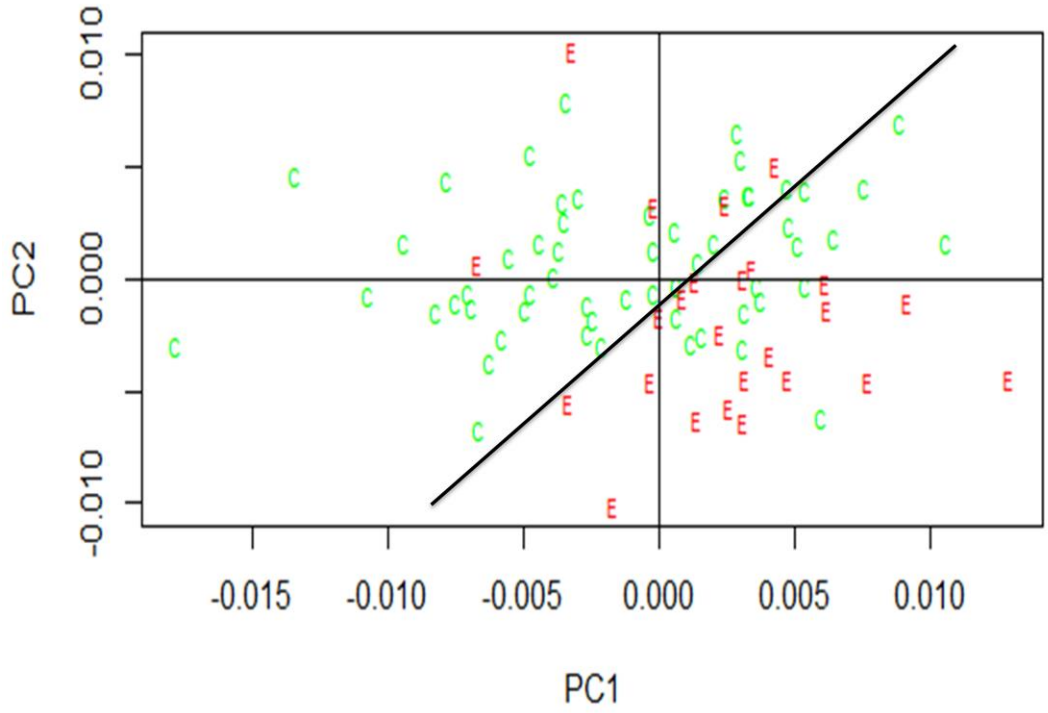


**Figure 4-6: PCA scatterplot showing a clustering of exacerbation samples in PC1+ PC2-. The regression model shown distinguishes COPD baseline from COPD exacerbation samples with a sensitivity and specificity of 54.05% and 88.09%, respectively. Sensitivity can be improved whilst retaining >80% specificity by determining baseline from exacerbation based on the sample falling within $PC1^{+}PC2^{-}$, giving sensitivity and specificity scores of 60% and 81.48%, respectively.**

### 4.3.2 Development of the Sample Strip IR Substrate

As part of the clinical evaluation and development of the hand-held IR Sphinx spectrometer, a need for a disposable sampling strip was identified. There were certain criteria that the sample strips needed to adhere to in order to be suitable for use with the IR Sphinx. The substrate onto which the sample was deposited must not interfere with the spectrum and must have low-IR absorption properties, sputum must adhere strongly to the substrate, the whole strip must be of low enough cost to

be used as a disposable item, and the strip must be capable of being clamped onto the ATR crystal without damage occurring to the strip or crystal.

Aluminium oxide (AlO) film was selected as a potential substrate for the sampling strip due its well-documented low-IR absorption properties. Through collaboration with WCPC at Swansea University, a thin (<100μm) polyurethane (PU) film metallised with AlO was obtained.

### 4.3.3 Testing the AlO Substrate: Bruker Alpha

-

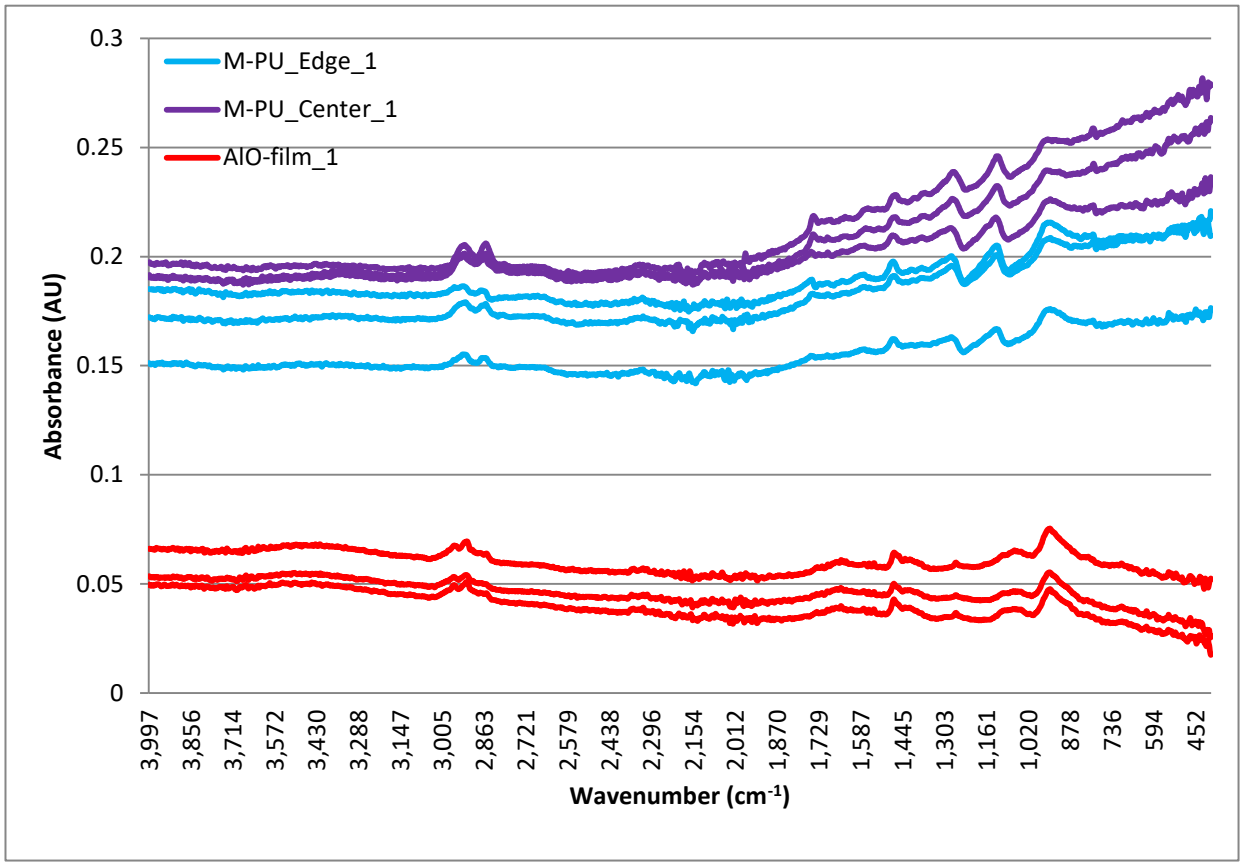


**Figure 4-7: ATR-FTIR spectra from 4000-450cm$^{-1}$ of the AlO substrate, on different backing materials; (red) AlO deposited onto a PU film, placed onto adhesive PE foam and then secured on a hard plastic backing (AlO film), and (blue & purple) AlO deposited directly onto PU foam by metallisation (M-PU) and secured onto a hard plastic base, directly measured by ATR-FTIR at the centre (purple), or the edge (blue) of the M-PU strip.**

Replicate spectra of the M-PU and AlO-film substrates were generated using the Bruker Alpha (Figure 4-7). The M-PU spectra show a higher level of absorption across the whole mid-IR spectrum, compared to the AlO-film spectra. Also, the M-PU spectra show increasing absorption from approximately 1800cm$^{-1}$ and down towards the N-IR region. The fingerprint region (1800-900cm$^{-1}$) contains the most spectral

information, so an increasing baseline of absorption due to the sampling substrate is undesirable.

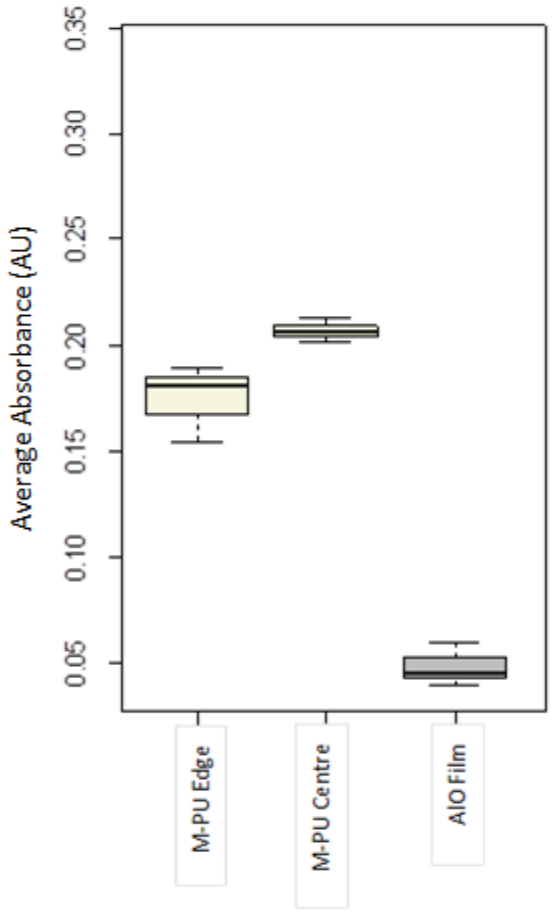


**Figure 4-8: Distribution of average absorbencies of whole ATR-FTIR spectra (n=6) of M-PU and AlO film. The M-PU spectra were generated from the centre and the edges of the metallised surface of the M-PU. This was to provide insight into the uniformity of the metallised surface across the wide surface area of the sampling strip. The AlO film shows a significantly decreased average absorbance level compared to the M-PU. Additionally, there is a significant difference between the sampling modes of the M-PU, suggesting that the metalized coating is not consistent across the strip.**

It was found that the AlO coating on the M-PU is easily disrupted. Figure 4-9 shows pictures taken of the M-PU before (Figure 4-9 A) and after (Figure 4-9 B & C) ATR-FTIR measurements using the Bruker Alpha. This is further evidence that the M-PU substrate is inappropriate for use as a sampling strip for routine ATR-VFIR analysis. Therefore, AlO deposited onto PU film (AlO film) was the substrate of choice for the sampling strip.

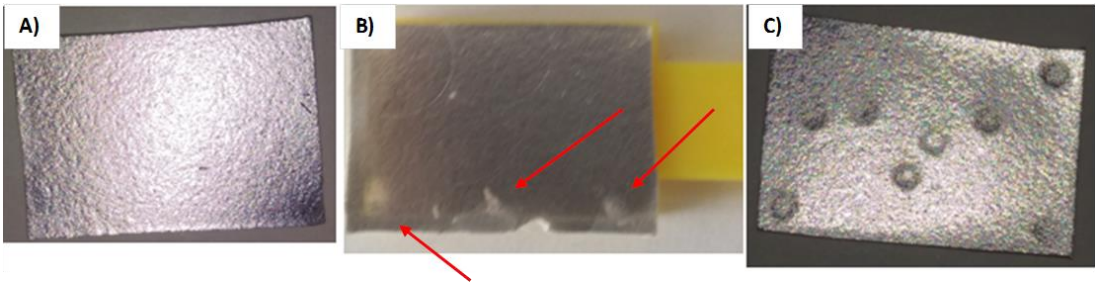


**Figure 4-9: Photographs of the M-PU strips before ATR-FTIR measurements (A) and after ATR-FTIR measurements (B & C). B shows the M-PU with plastic base and C shows it without. The plastic base appears to protect the M-PU from most severe wear, although there is evidence of AlO disbonding near the edges (red arrows).**

### 4.3.4 Testing the AlO Substrate: IR Sphinx

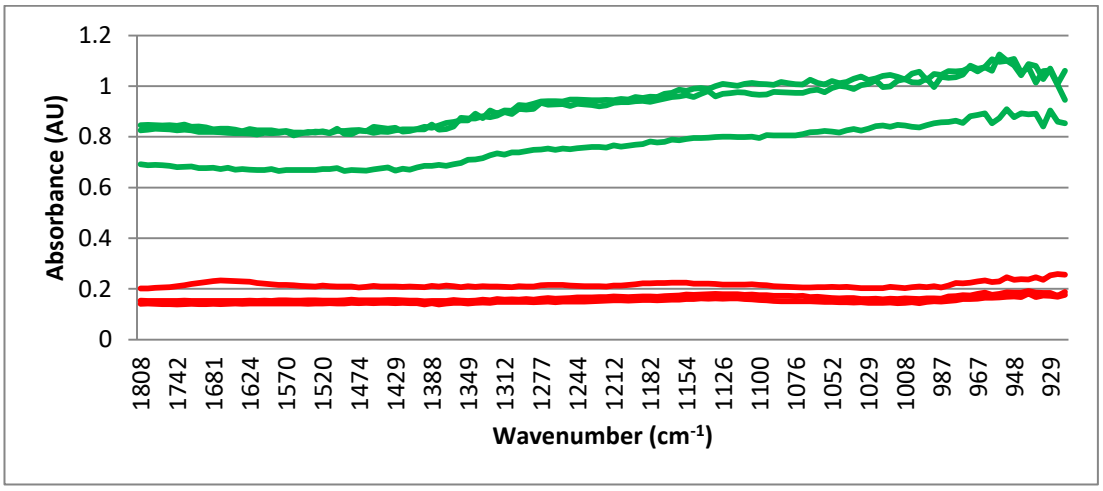


**Figure 4-10: Replicate ATR-VFIR (IR Sphinx) spectra of the plain AlO substrate placed onto M-PU (green) or the AlO film on foam (red). The plain M-PU shows a markedly higher level (approximately 4x) of absorbance compared to the AlO film, and a drifting baseline towards the lower wavenumber region.**

Figure 4-10 and Figure 4-11 show how the different deposition methods of AlO influence the spectra generated using the IR Sphinx. As with the Bruker Alpha, the average absorbance across the spectrum was significantly ($p < 0.05$) increased in M-PU spectra, compared to AlO film spectra. This shows how the metallisation process of the PU foam does not produce a layer of AlO of sufficient thickness to ensure that the IR beam does not penetrate through the AlO layer and become absorbed by the underlying substrate layer. This is further evidence that AlO film is a more suitable substrate for ATR-VFIR analysis.

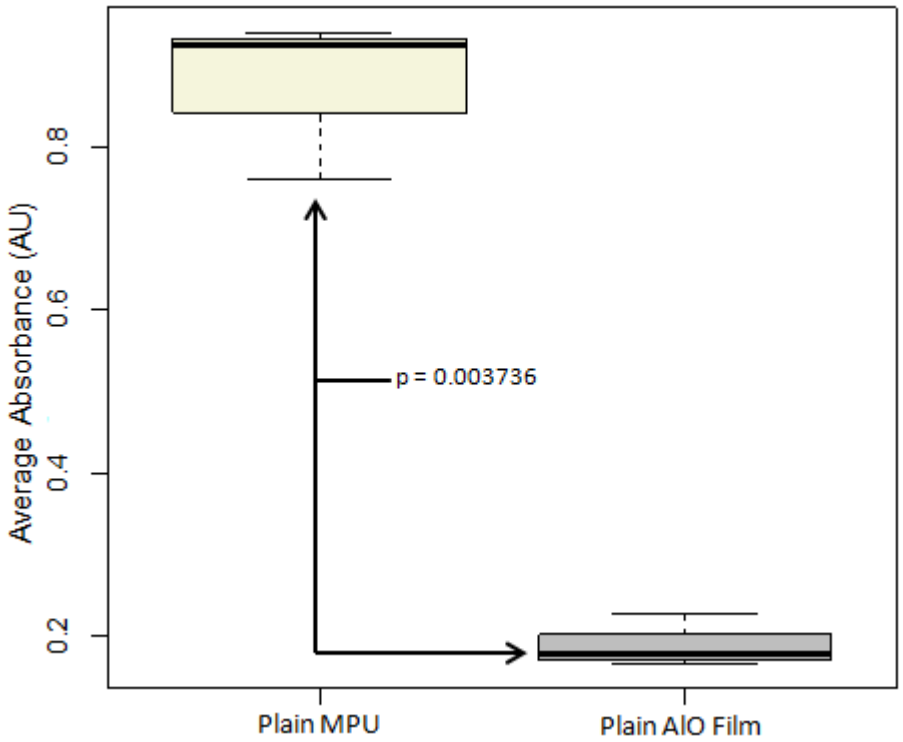


**Figure 4-11: Boxplots of average absorbencies across the whole IR spectrum from 1800:950cm$^{-1}$ of plain MPU and AlO substrates generated with the IR Sphinx (n=6). Significance testing using the Wilcoxon-rank sum test shows the AlO film substrate causes significantly less absorbance across the whole spectrum ($p < 0.05$).**

### 4.3.5 Backing material testing

For the sampling strip to produce spectra of optimal quality there must be extremely close contact between the sampling surface and ATR crystal. The ATR crystal of the IR Sphinx is large, with a total area of 425mm$^2$, compared to the 4mm$^2$ of the diamond ATR crystal found in conventional benchtop FTIR spectrometers, such as the Bruker Alpha. This large surface area increases the potential for variation in how samples could be applied to the ATR crystal, both with and without the use of an external sampling strip. The sampling strip was designed with compressible foam underneath the IR-substrate (Figure 4-2(a)), to spread the applied load from the clamping force during IR-spectrum acquisition and ensure equally close contact between the sample and ATR crystal surfaces. Multiple compressible foams with differing hardnesses, thicknesses and densities were tested to ensure that the optimal material was used for the best possible quality spectrum acquisition (Table 4-1).

| Name | Supplier | Material | Thickness (µm) | Density (kg/m³) | Hardness |
|---|---|---|---|---|---|
| ***E1715h*** | 3M | PE Foam | 380 | 480 | Medium-Firm |
| ***E1515h*** | 3M | PE Foam | 380 | 465 | Medium |
| ***E1120h*** | 3M | PE Foam | 559 | 384 | Soft |
| ***Softprint 52122*** | TESA | PE Foam | 500 | 330* | Soft |
| ***xSOFT 52223*** | TESA | PE Foam | 500 | 270* | Extra Soft |

**Table 4-1: List of the foams analysed and a description of properties - *values not provided by manufacturer, calculated in house; PE = polyethylene.**

Porcine gastric mucin (PGM) was used as a control sample for testing sampling strips with different compressible foams, to ensure that any differences seen are due to the properties of the foam. PGM was applied directly to the AlO-film and allowed to dry at room temperature for one hour. Replicate (n=12) ATR-VFIR spectra of the PGM dried on the sampling strips with each backing material were acquired and assessed for overall spectral quality using the following parameters; signal-to-noise ratios (SNR), variation of overall absorbance intensity and peak positions variation.

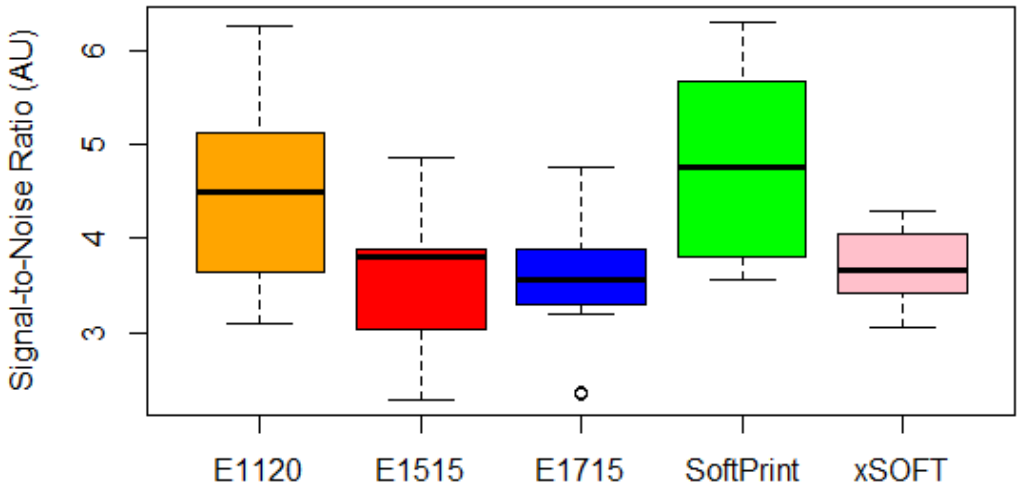


**Figure 4-12: Distribution of SNRs from all spectra acquired using each backing material.**

ATR-VFIR spectra of PGM were first evaluated in terms of overall SNR, which was calculated using the in-built signal-to-noise algorithm in OPUS, with parabolic fitting enabled. The 3M-E1120 tape showed the greatest variation in range of SNR, whereas the TESA xSOFT showed the least variable SNR values. The TESA SoftPrint showed the greatest overall SNR, although the range was large and overlapped strongly with the other foam tapes.

Spectra were also evaluated for variation in peak positions within the glycogen-rich region of the spectrum, and average absorbance across the spectrum and specified regions of the spectrum.

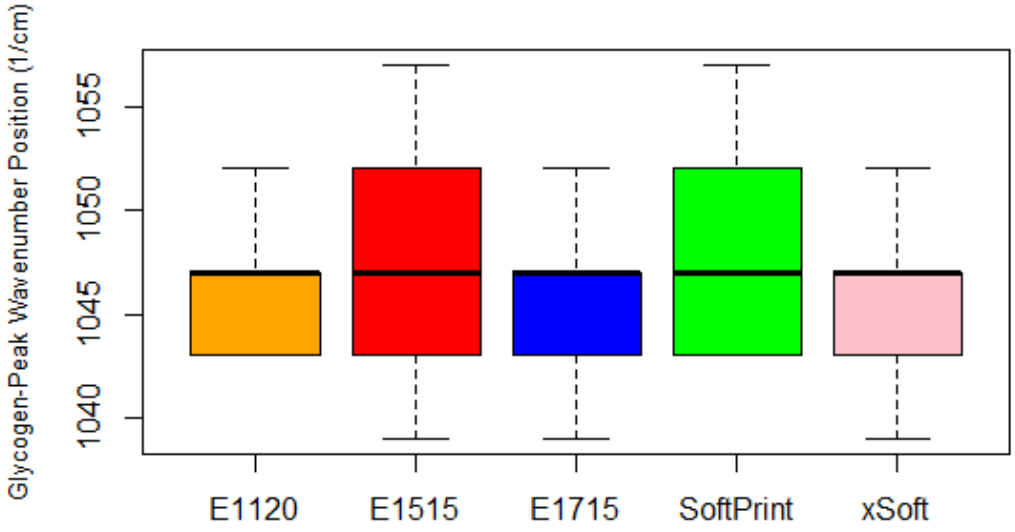


**Figure 4-13: Distribution of the wavenumber position ($cm^{-1}$) of the major glycogen peak found in PGM using each backing material.**

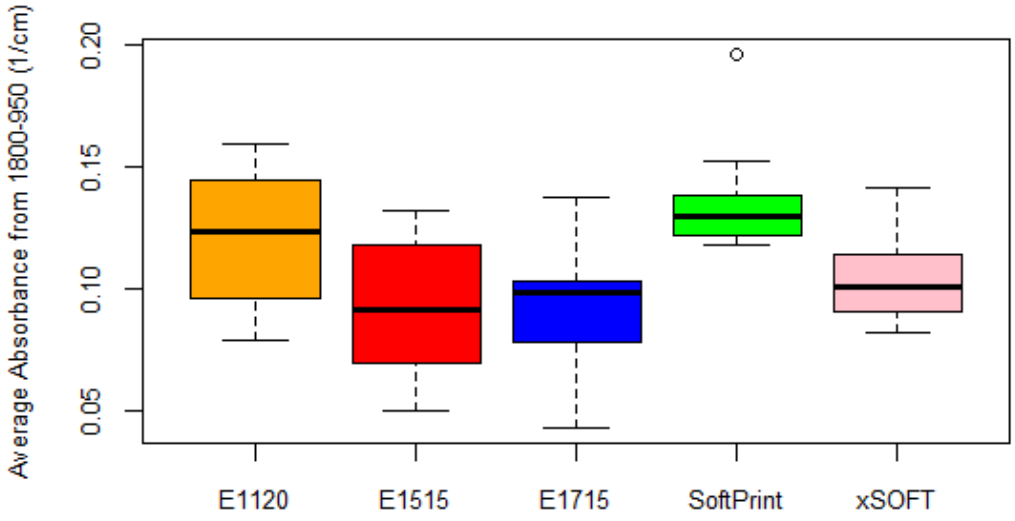


**Figure 4-14: Distribution of average absorbencies across the full spectrum - 1800-900$cm^{-1}$. Variance is lowest in the Softprint and xSOFT tapes**

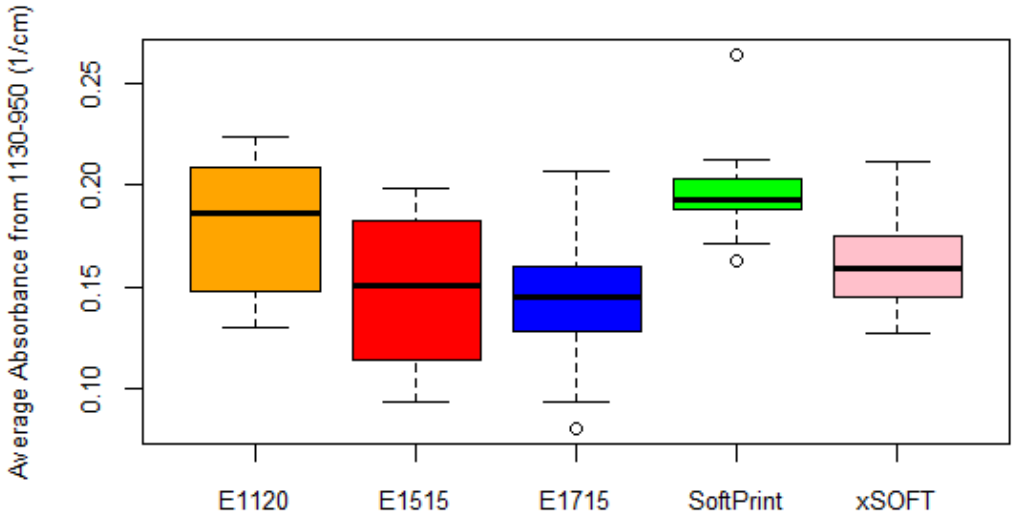


**Figure 4-15: Distribution of average absorbencies within the major glycogen-associated peak (1130-950cm$^{-1}$).**

| | Mean Absorbance 1130-950cm$^{-1}$ | Variance | Mean Major Glycogen Peak Position | Variance | Mean SNR | Variance |
|---|---|---|---|---|---|---|
| **E1120** | 0.179011 | 0.0012 | 1046.167 | **11.0606** | 4.53075 | 1.00674 |
| **E1515** | 0.148525 | 0.0013 | 1047.167 | 38.6970 | 3.520917 | 0.58614 |
| **E1715** | 0.14348 | 0.0014 | 1045.727 | 16.0182 | 3.514818 | 0.5005 |
| **SoftPrint** | 0.197294 | **0.0006** | 1048.167 | 27.0606 | 4.814333 | 1.10665 |
| **xSOFT** | 0.161803 | **0.0006** | 1045.833 | 14.6970 | 3.708167 | **0.1453** |
| **ANOVA Between Groups (p)** | 0.00063489 | | 0.671685744 | | 0.000220598 | |

**Table 4-2: Mean absorbance from 1130-950cm$^{-1}$, mean glycogen peak position, and mean SNR across the whole spectrum, with variances for each group shown. *Lowest variance values for each parameter are bolded for clarity*. The Softprint and xSOFT tapes both show the lowest variance in terms of absorbance intensity, and the xSOFT shows the least variable SNR. The 3M-E1120 tape showed the least variance for detection of the major glycogen peak. Single factor ANOVA shows that the variation between the groups of backing materials is statistically significantly different at the 95% confidence level when assessing the mean absorbance and SNR, but not the peak position.**

Spectra generated using the xSOFT tape showed the lowest variation in terms of average absorbance, and SNR (Table 4-2). The observed variation in peak position was lowest for the E1120 tape, although the peak positions in spectra acquired with xSOFT tape were not shown to be statistically significantly different between the E1120 and xSOFT tapes at p=0.822163 (Table 4-3). This suggested that usage of the xSOFT tape could produce more reliable spectra than the other tapes in terms of absorption intensity and SNR, whilst not significantly increasing variation in the wavenumber positions of the major peaks.

| | | Two-sided T-Test Results | | | |
|---|---|---|---|---|---|
| | | **E1120** | **E1515** | **E1715** | **SoftPrint** |
| **Glycogen Peak Position** | **E1515** | 0.62971 | | | |
| | **E1715** | 0.77869 | 0.51389 | | |
| | **SoftPrint** | 0.27601 | 0.67352 | 0.21967 | |
| | **xSOFT** | 0.82216 | 0.53514 | 0.94898 | 0.14152 |
| **Average Absorbance 1130-950cm$^{-1}$** | **E1515** | **0.04692*** | | | |
| | **E1715** | **0.02744*** | 0.74646 | | |
| | **SoftPrint** | 0.15107 | **0.00113*** | **0.00083*** | |
| | **xSOFT** | 0.16995 | 0.30590 | 0.18216 | **0.00367*** |
| **Signal-to-Noise Ratio** | **E1515** | **0.01157*** | | | |
| | **E1715** | **0.01056*** | 0.98435 | | |
| | **SoftPrint** | 0.50626 | **0.00256*** | **0.00234*** | |
| | **xSOFT** | **0.01875*** | 0.45912 | 0.43302 | **0.00851*** |

**Table 4-3: Results of statistical significance testing by T-test for differences in detected peak positions, average absorbance from 1130-950cm$^{-1}$, and SNR across the spectrum due to the use of the backing materials. Significance was sought at $p<0.05$, and all significant differences found due to the backing material are indicated with an asterisk.**

### 4.3.6 Comparison of ATR-FTIR Spectra with and without the Sampling Strip

The sample-strip was first tested using the benchtop Bruker Alpha spectrometer, to perform direct comparisons between sputum spectra acquired using the standard ATR-FTIR protocol, as used in Chapter 3, and sputum spectra acquired with the sampling strip. The sampling strip spectra were examined for introduction of spectral artefacts, shifts in baseline, shifts in peak position, relative to ATR-FTIR spectra acquired without the strip. One COPD patient's sputum sample (TR06) which had previously been observed to produce easily reproducible spectra, and which was in plentiful supply in order to carry out many repeats was chosen to assess the reproducibility of IR-spectra generated using the sampling strip.

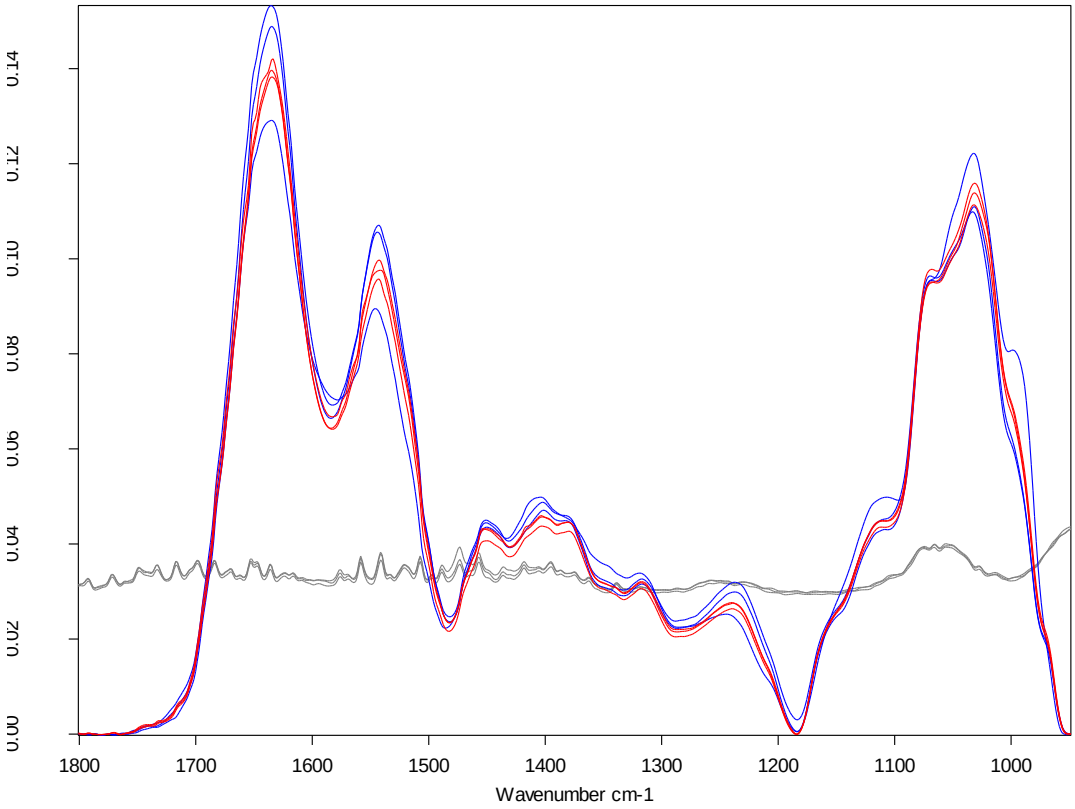


**Figure 4-16: Vector-normalized, baseline-corrected ATR-FTIR spectra of COPD patient sample TR06, dried onto a sampling strip (red) and directly onto the ATR crystal (blue) in triplicate. Raw spectra of plain sampling strip (grey) in triplicate are also shown for comparison of AlO absorbance features to absorption profiles of typical sputum spectra.**

A visual examination of the spectra in Figure 4-16 shows that the vector-normalised, baseline-corrected replicate ATR-FTIR spectra of sample TR06 do not show a large degree of variation between sampling modes. There are no visible spectral artefacts, or major peak position shifts introduced by the sampling strip. A statistical analysis of the TR06 patient sputum spectra followed to confirm the significance of any changes that the sampling strip may introduce.

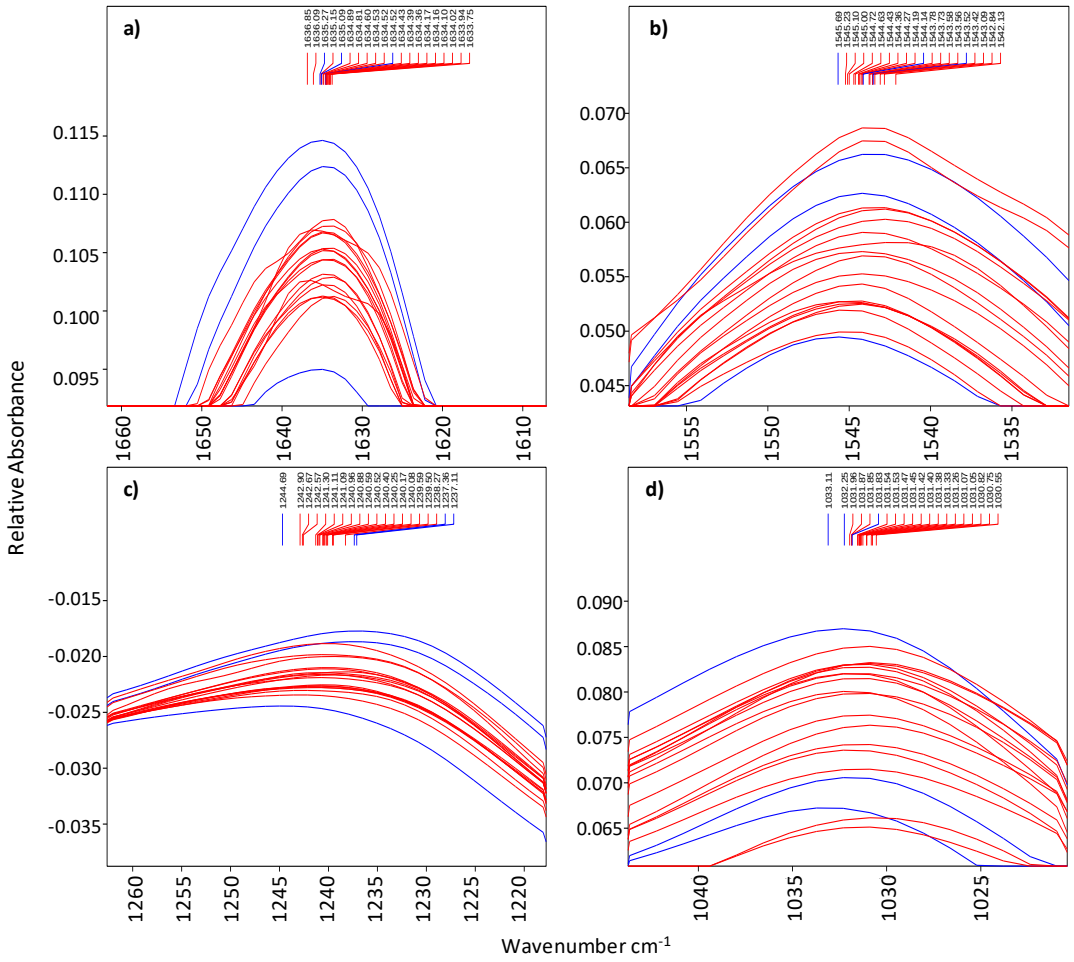


**Figure 4-17: Baseline-corrected, vector-normalized ATR-FTIR spectra of COPD patient sputum sample TR06 dried onto the sampling strip (red, 16 replicates), and onto the ATR crystal (blue, triplicate). Exact peak positions of major peaks found within (a) 1660 – 1610cm$^{-1}$, (b) 1555 – 1535cm$^{-1}$, (c) 1260 – 1220cm$^{-1}$, and (d) 1040 – 1025cm$^{-1}$ regions are highlighted above the spectra and shown in figure 4-4.**

The location and number of major IR peaks in ATR-FTIR spectra within amide I, amide II, or glycogen-rich regions (Figure 4-16 and Figure 4-17) does not significantly change because of the use of the sampling strip. Subtle variation in peak positions can be seen, however the range of the peak positions is small, and could be due to variations of the molecular structure of the sputum sample. The heterogeneous nature of sputum has been explored in this thesis in Chapter 3; therefore it is reasonable that the absence of observations of major peak position alterations in spectra generated using the sampling strip could support the hypothesis that the subtle peak position variation could be due to intra-sample variation. Variation between and within sampling modes was assessed by ANOVA (Table 4-4). Variation within each sampling mode was found to be statistically significantly

greater ($p < 0.05$) than the variation introduced by the sampling strip at all peak positions except for the glycogen-rich peak position.

| Spectral Region | Variation Within Sampling Modes | Variation Between Sampling Modes | P-value |
|---|---|---|---|
| **Amide I** | 10.2036 | 0.2457 | 0.5308 |
| **Amide II** | 13.6585 | 0.4233 | 0.4778 |
| **Sulphate** | 57.9309 | 3.1373 | 0.3508 |
| **Glycogen-rich** | 3.2993 | 2.8442 | 0.0014 |

**Table 4-4: ANOVA results, no significant variation between sampling modes for the amide I, amide II and sulphate regions, but statistically significant difference in variation between the sampling modes was detected within the glycogen-rich region.**

| | SS | WO | SS | WO | SS | WO | SS | WO |
|---|---|---|---|---|---|---|---|---|
| | Amide I | | Amide II | | Sulphate | | Glycogen-rich | |
| Mean Peak Position | 1634.65 | 1634.96 | 1544.04 | 1544.45 | 1240.83 | 1239.72 | 1031.37 | 1032.40 |
| P-value | 0.3443 | | 0.5981 | | 0.6987 | | 0.09715 | |
| 95% CI | -1.0551589<br>0.4314089 | | -2.864628<br>2.045878 | | -9.37816<br>11.60691 | | -2.5478322<br>0.4257489 | |

**Table 4-5: Mean peak positions of ATR-FTIR spectra generated using the sampling strip (SS) and without (WO) the sampling strip. P-values with 95% confidence intervals (CI) show degree of significance of differences in peak position due to the sampling strip.**

Normality testing with a Shapiro-Wilk test for normality suggested that the peak position data were normally distributed, so significance testing was performed using Welch's two-sided T-test, with an $\alpha$-level of 0.05. A statistically significant difference was found at the glycogen-rich region peak position ($p < 0.05$) due to use of the sampling strip (Table 4-5).

### 4.3.7 Comparison of COPD Sputum ATR-FTIR Spectra Generated with and without the Sampling Strip

ATR-FTIR spectra of COPD patient sputum samples were generated using the Bruker Alpha ATR spectrometer with and without the use of the sampling strip. This was to assess if using the sampling strip introduced a source of error or variation into the acquired spectra. Replicate (n=3) ATR-FTIR spectra of 118 COPD patient sputum samples were processed and averaged according to Chapter 2, Section 2.2.3. The average spectra for each patient were acquired for each sampling mode were then analysed for significant differences. The average spectra are shown in appendix 4.

| | | | | | |
|---|---|---|---|---|---|
| **Average Peak Position** | **1638.128** | **1543.403** | **1239.567** | **1071.261** | **1035.405** |

| | | | | | |
|---|---|---|---|---|---|
| **p-two-sided** | 1.30E-09 | 0.03545 | 0.0001507 | 6.82E-05 | 1.49E-09 |
| **p-one-sided (greater)** | 6.52E-10 | 0.9823 | 0.9999 | 3.41E-05 | 7.47E-10 |
| **p-one-sided (less)** | 1 | 0.0177 | 7.54E-05 | 1 | 1 |
| **Mean difference ($cm^{-1}$)** | 3.3887 | -0.6479 | -1.4940 | 1.4628 | 2.412 |
| **99 % CI** | 2.0428 | -1.4452 | -0.4964 | 0.5401 | 1.486 |
| | 4.7347 | 0.1494 | -2.4914 | 2.3854 | 3.339 |

**Table 4-6: Average major peak positions detected in absorbance spectra with significance values from paired T-tests for differences in peak positions in sputum spectra generated using the two sampling modes.**

Statistical significance testing using paired T-tests showed that there are statistically significant ($p < 0.05$) differences in detected peak positions at all major peaks when comparing spectra generated using the sampling strip or by depositing directly onto the ATR crystal (Table 4-6). The distribution of peak positions across all samples was also found to be statistically significantly different between the sampling modes (Figure 4-18 to Figure 4-22). The sampling strip appears to contribute to a greater degree of variance in detected absorbance peak position across the infrared spectrum. This greater variance could either be due to the AlO foil interacting with the IR beam, or due to variations in the sputum sample itself. However, the IR-absorbance of aluminium foil is very low and the spectrum is almost featureless, with the exception of a small peak at $974cm^{-1}$ (Figure 4-7). It is unlikely, however, that this peak should influence the position of the detected peaks studied in this work.

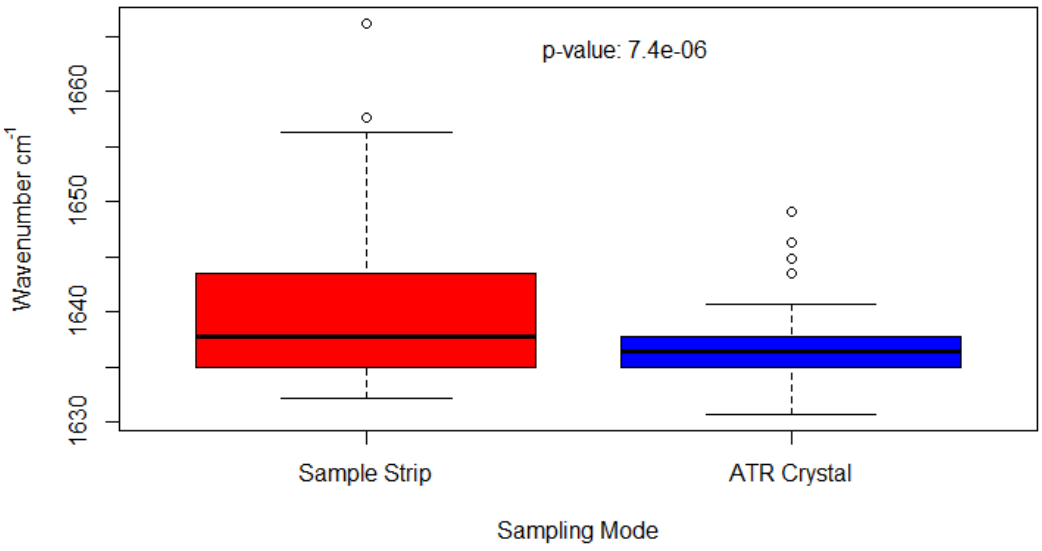


**Figure 4-18: Comparison of the distributions of all detected absorbance spectrum peak positions ($cm^{-1}$) within the amide I region from the use of the sampling strip (red) and the ATR crystal (blue)**

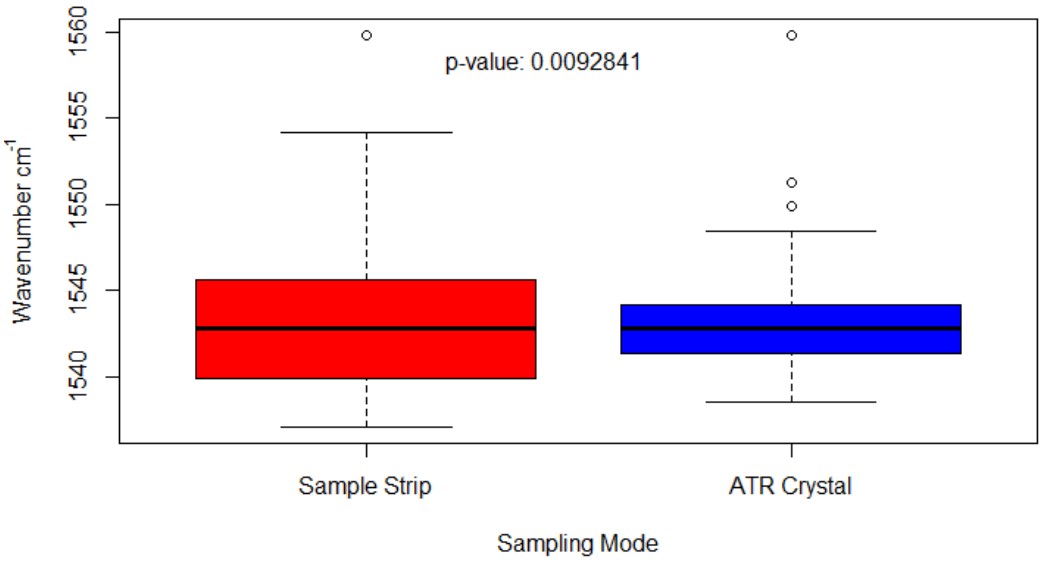


**Figure 4-19: Comparison of the distributions of all detected absorbance spectrum peak positions ($cm^{-1}$) within the amide II region from the use of the sampling strip (red) and the ATR crystal (blue)**

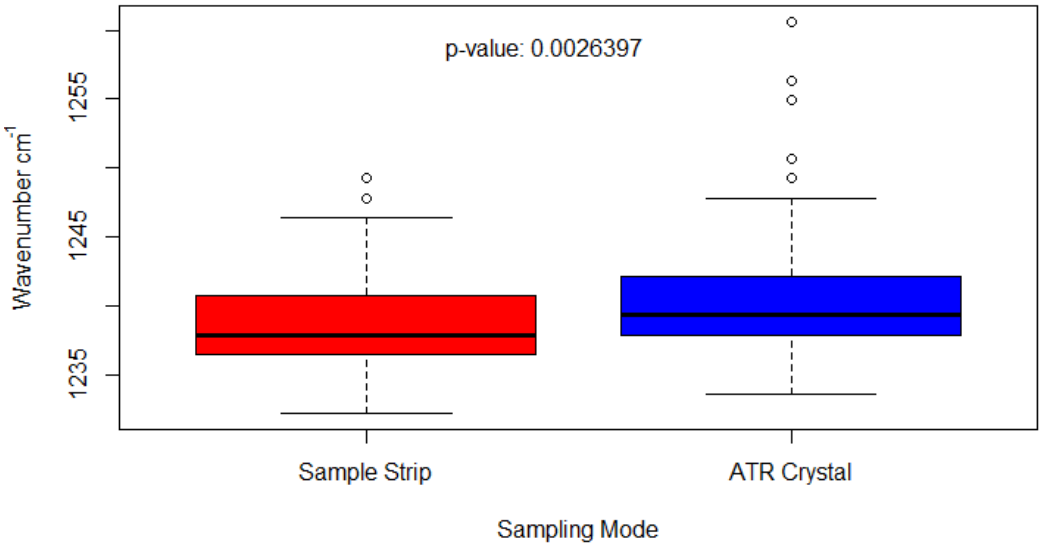


**Figure 4-20: Comparison of the distributions of all detected absorbance spectrum peak positions ($cm^{-1}$) within the sulphate (S=O stretching) region from the use of the sampling strip (red) and the ATR crystal (blue)**

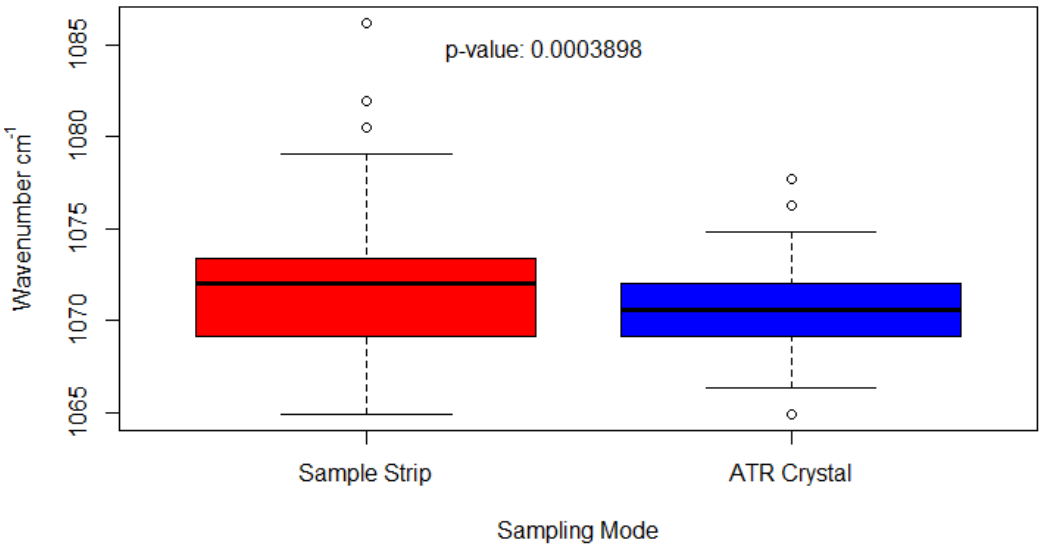


**Figure 4-21: Comparison of the distributions of all detected absorbance spectrum peak positions ($cm^{-1}$) within the glycogen-rich region from 1065 – 1085$cm^{-1}$ from the use of the sampling strip (red) and the ATR crystal (blue)**

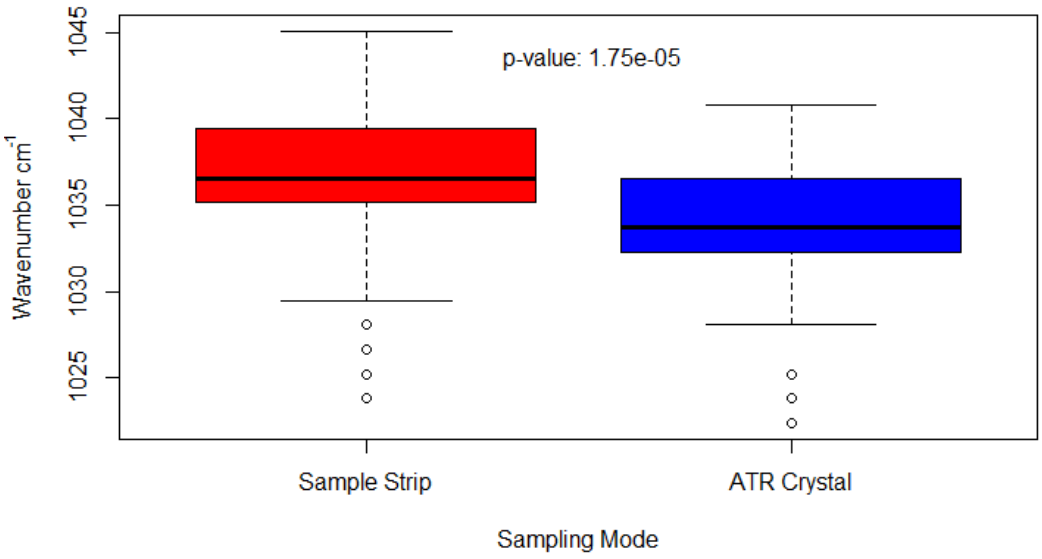


**Figure 4-22: Comparison of the distributions of all detected absorbance spectrum peak positions ($cm^{-1}$) within the glycogen-rich region from 1025 – 1045$cm^{-1}$ from the use of the sampling strip (red) and the ATR crystal (blue)**

### 4.3.8 Further Development of ATR Sampling Strip

A recent, exciting development in the field of ATR-FTIR has seen the introduction of a novel design of ATR crystal, using etched silicon wafers. Silicon wafers have been used as substrates for T-FTIR for many years, but recently IRUBIS GmbH have developed silicon wafers for ATR spectroscopy. A simplified schematic of a silicon ATR wafer is shown in Figure 4-23. They claim that the silicon ATR wafer produces spectra of comparable quality to that produced using a conventional ZnSe

or diamond crystal. These wafers are approximately €40 each, and therefore represent a significant reduction in cost for an ATR spectrometer compared to conventional diamond, ZnSe, or KBr crystals. This reduction in cost per unit ATR crystal raises the possibility of using these or similar silicon ATR crystals as a disposable sampling substrate. For this reason, silicon ATR wafers were acquired from IRUBIS and evaluated for their inter- and intra-wafer reproducibility of spectra, and the overall quality of spectra produced by these wafers.

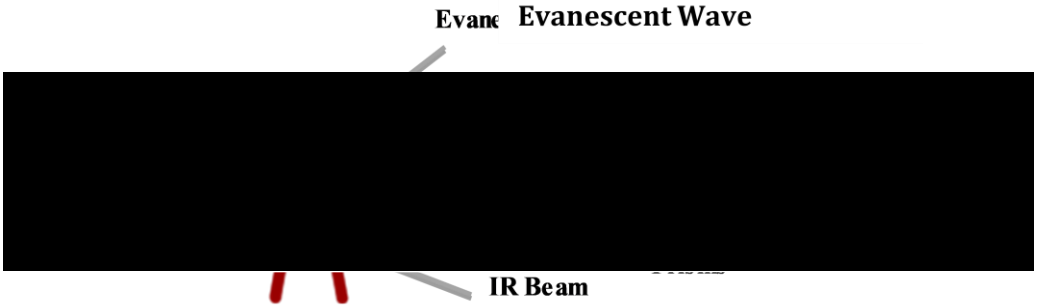


**Figure 4-23: Schematic of silicon ATR wafer, showing etched prisms and beam path refracting through the prism. Figure adapted from IRUBIS GmbH, 2018 (https://irubis.com/products/atr-crystals/single-reflection-atr-crystal/)**

Fifteen silicon ATR wafers were obtained from IRUBIS for direct comparison to ATR-FTIR spectra produced by the benchtop Bruker Alpha instrument utilised earlier in this thesis. It was necessary to use an ATR-adapter in the sample compartment of the Bruker Vertex, onto which the ATR wafers were placed for analysis. One COPD patient's sputum (TR01) was chosen for this test to ensure that variation due to sampling was limited. Sputum was pipetted (2µl) onto each wafer and allowed to dry before FTIR analysis. Replicate spectra (n=7) were generated for each wafer using a new background spectrum each time, to reduce any potential influence from changing environmental conditions. The spectra generated are shown in Figure 4-24 with low inter- and intra-wafer variability seen. The spectra can be seen to group together by wafer, although this could be due to variation in sampling when applying the sputum. This is confirmed in Figure 4-25, where the pattern of wafer groups is distinct from that shown in Figure 4-24. The light-green spectra in Figure 4-25 are seen to group away from the other spectra, which suggests that an error in sampling occurred. Such an error could be that the sputum was applied too thickly or too was not applied completely on the active area of the ATR wafer. Sputum which is applied too thickly takes more time to fully dry, but also it is possible that water contained in the sample could become trapped at the interface

between the sample and wafer. Presence of water in a sample is known to adversely affect spectrum acquisition (Baker *et al.*, 2014; S. P. Lewis *et al.*, 2013; Oberg & Fink, 1998), therefore it is reasonable to hypothesise that retained water could be a source of sampling error. Another potential source of error is the possibility of the ATR wafer being incorrectly loaded into the spectrometer prior to spectrum acquisition. However this is unlikely as all three replicate spectra show the same error. Furthermore, if the error was due to the loading of the wafer into the spectrometer it would be reasonable to assume that such an error could be present in the spectra for the other wafers, however examination of the these spectra shows no such error. Because of this, care should be taken to ensure that sputum is not applied to the wafers too thickly, and all samples should be fully dry before spectrum acquisition.

To further test the concordance between silicon ATR-wafers, the experiment was repeated using a different condition for the background spectrum. A background spectrum of every wafer was collected, and an average of each of these was calculated. For comparison, spectra were generated using the specific background spectrum for each wafer, and the average background spectrum (Figure 4-26). A mean spectrum for each experimental condition was calculated (Figure 4-27), showing a Pearson's correlation coefficient of 0.999998 ($p < 2.2*10^{-16}$) between the two conditions of background spectrum. This implies that variation introduced by the wafers is extremely low. Correlation between the wafers was also assessed by comparing the average spectrum for each wafer to every other spectrum generated using the silicon ATR wafers. The median correlation coefficient of each wafer was shown to be rho > 0.99, suggesting a high degree of similarity between spectra of the same sample generated using different ATR wafers.

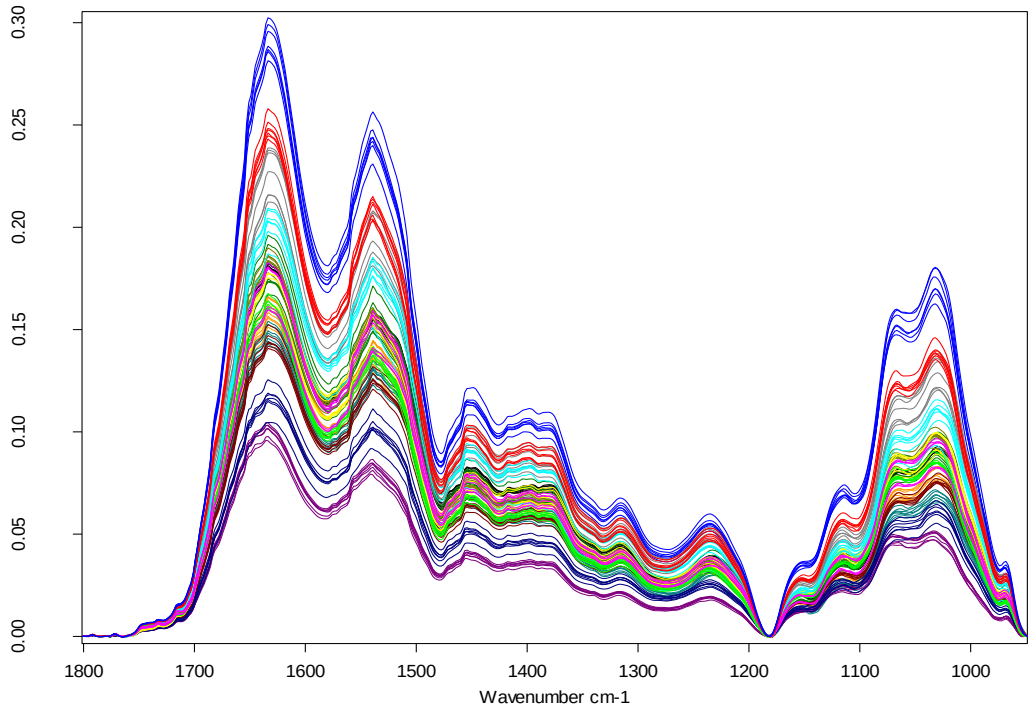


**Figure 4-24: First experiment of the silicon ATR wafers. Baseline-corrected ATR-FTIR spectra (n=105) of COPD sputum (TR01) generated using the silicon ATR wafer. Each spectrum is coloured according to the wafer used, with 7 replicate spectra per wafer.**

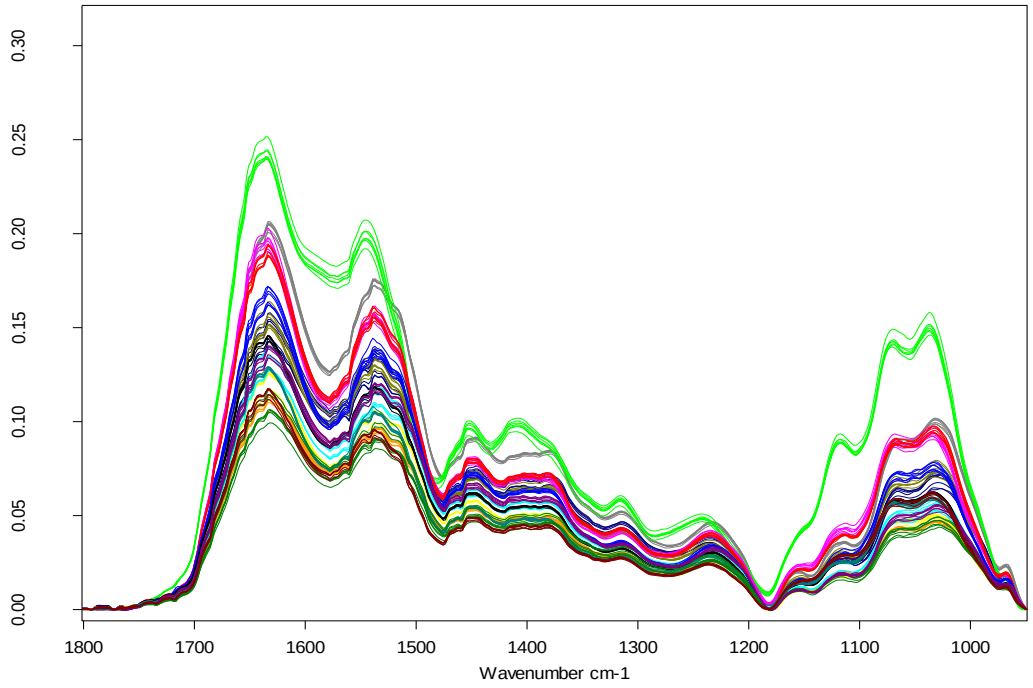


**Figure 4-25: Second experiment of the silicon ATR-wafers. Colour coding of wafers is the same as in Figure 4-24.**

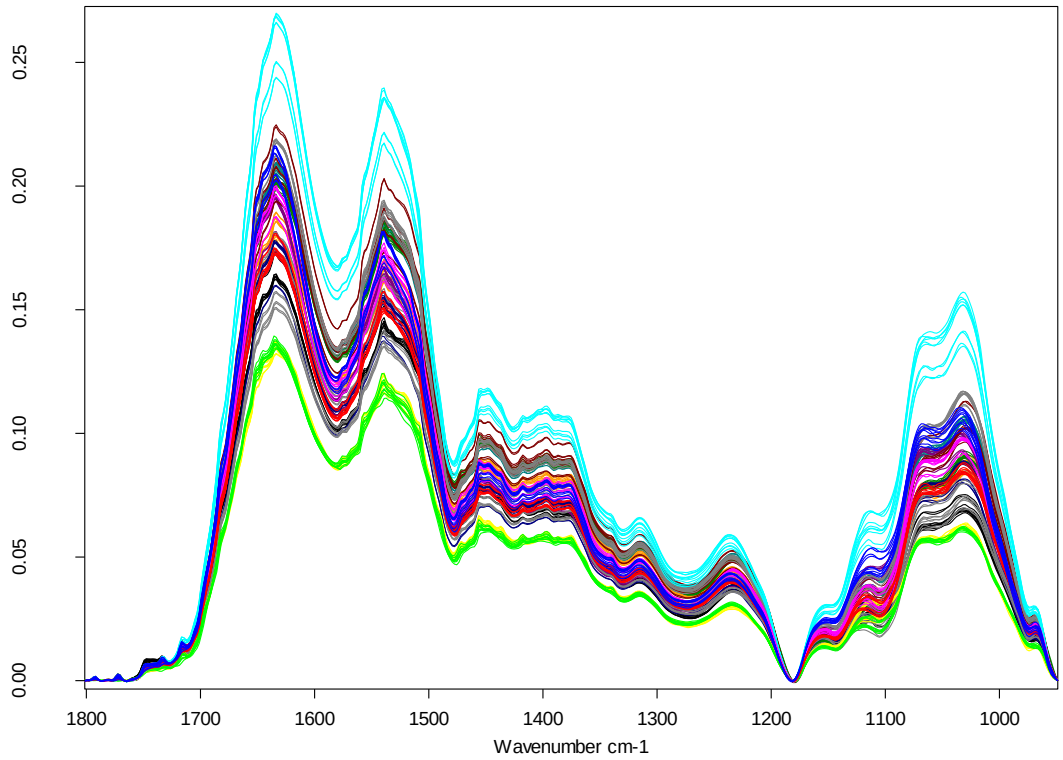


**Figure 4-26: Baseline-corrected ATR-FTIR spectra (n=150) generated using background spectra specific to the wafer (n=75, 5 per wafer), and an average background spectrum consisting of background spectra from all wafers (n=75, 5 per wafer). Colour coding of spectra as in Figure 4-24.**

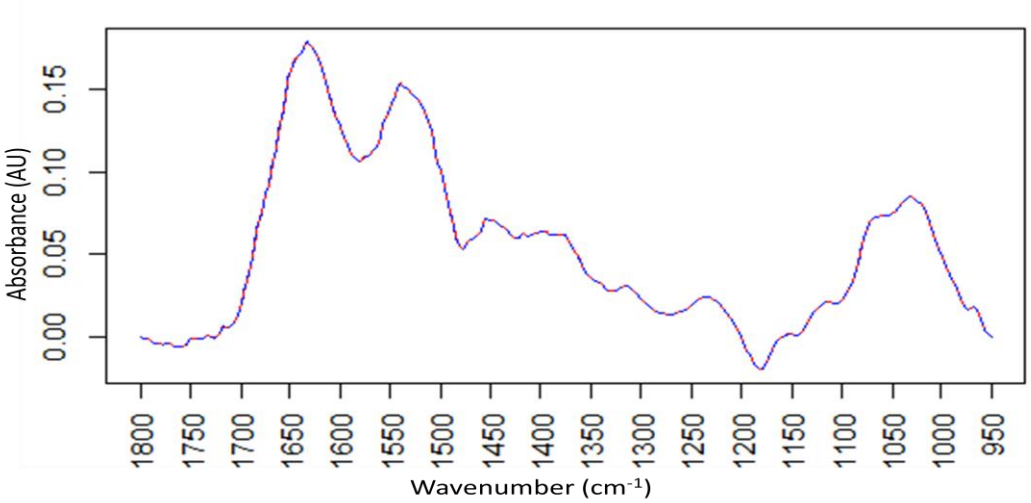


**Figure 4-27: Baseline-corrected mean spectra of TR01 generated using the silicon ATR wafers, with wafer-specific background spectrum (red), and the average background spectrum (blue). Pearson's correlation coefficient of 0.999998 ($p < 2.2*10^{-16}$) showing very high degree of similarity between spectra generated using the different background conditions.**

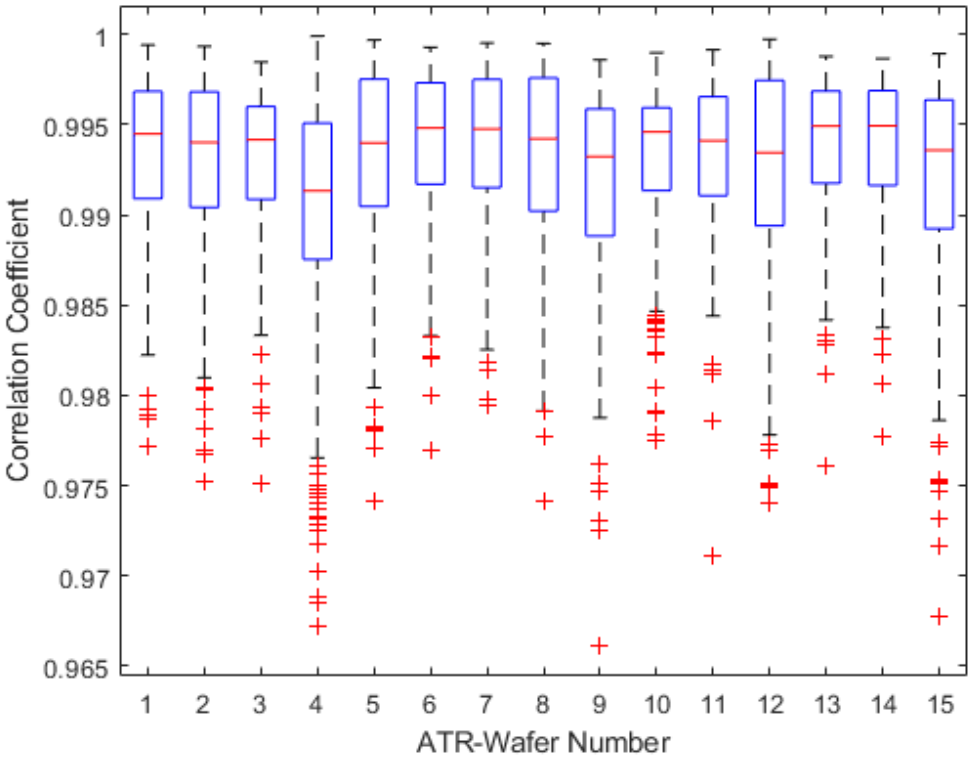


**Figure 4-28: Distribution of Spearman's Rank correlation coefficient between each wafer's average spectrum and all other spectra generated using the silicon ATR wafers. The median correlation coefficient of each wafer to all other spectra is >0.99, implying excellent correlation and a high degree of similarity between spectra generated using each wafer.**

Next, the spectra generated by the ATR wafers in the Bruker Vertex were compared to spectra previously generated by the Bruker Alpha ATR-FTIR spectrometer. Porcine gastric mucin (PGM) was used as a standard sample to compare across devices. Repeat spectra (n=15 per device) were generated in the same manner as described earlier. Analysis of spectra focused on identifying differences between raw spectra generated on either device, which may constitute the aberrant addition of spectral artefacts. For this reason only raw, unprocessed spectra were examined. Shown in Figure 4-29 are the average raw spectra of PGM for both devices.

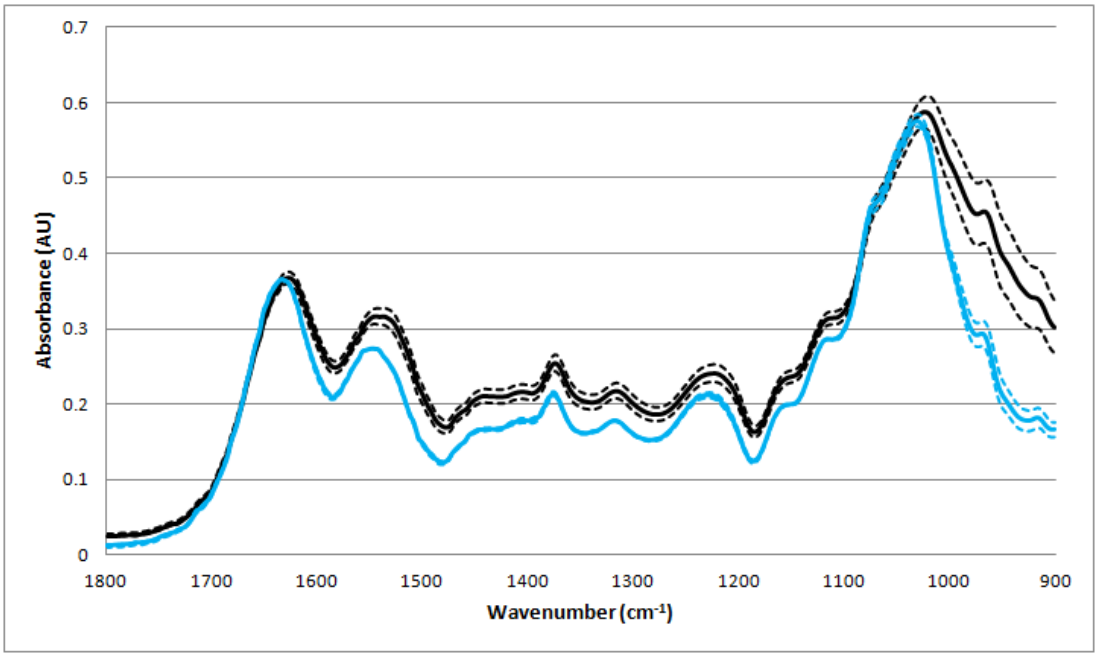


**Figure 4-29: Average raw spectra of PGM generated with ATR wafers (black) and the Bruker Alpha (blue), from 1800 – 900 $cm^{-1}$. The dashed lines represent 95% confidence levels from the mean spectrum.**

Initial investigation of the raw spectra shows an overall increase in spectral drift towards the N-IR region in the PGM spectra generated with the ATR wafers. The wafers could be the cause of this drift; however it could also be due to the experimental design. It was necessary to use two spectrometers for this experiment, which have different architecture and detectors. It is possible that the use of different spectrometers is the reason for this difference in drift. No peak gain or loss was observed within the fingerprint region, and all peaks show the same overall shape, with no apparent distortion introduced by the wafers. Subtle variations can be observed in peak position at the amide I and glycogen-rich regions, but statistical analysis highlighted zero peaks positions which showed a statistically significant ($p < 0.05$) shift. Similarity between the ATR-wafer spectra and Bruker Alpha ATR spectra was quantified using Spearman's Rank, with the average correlation coefficient between sampling modes calculated as 0.935617, suggesting excellent similarity between raw spectra.

The work presented in this chapter overall suggests that the silicon ATR wafers do not introduce a source of variation when performing ATR-FTIR analysis of sputum, and therefore are a suitable alternative substrate for ATR-FTIR analysis. They also produce spectra of excellent quality, which are highly comparable to spectra of the same sample generated using a diamond ATR FTIR crystal.

## 4.4 Discussion

### 4.4.1 The IR Sphinx for Generation of COPD Sputum Spectra

The IR-Sphinx is a hand-held, miniaturised IR spectrometer, based on ATR-VFIR technology, with a MEMS detector. It contains zero moving parts, with the IR beam being split into constituent wavenumbers by the LVF filter which forms part of the MEMS detector. It also utilises a large ZnS ATR crystal. For these reasons, the IR-Sphinx is a lower-cost alternative to the laboratory systems such as the Bruker Alpha. The IR-Sphinx was evaluated in this thesis to ascertain its capability for generating reliable spectra from complex biofluids. To this end, COPD patient sputum from baseline and exacerbating patients were used. It was found that the IR Sphinx is capable of generating reproducible sputum spectra, which were, in general, comparable to those generated by a higher-resolution device, although much spectral information was lost.

The spectra produced by the IR Sphinx are of a lower resolution than that produced by the Bruker Alpha. The Bruker Alpha is capable of producing spectra at a resolution of $4cm^{-1}$; however comparative testing using IPA spectra provided an estimate of resolution of approximately $30cm^{-1}$ for the IR Sphinx. This decrease in resolution is evident when directly comparing spectra for the same sample. Key spectral features, such as the amide I, and the major glycogen-rich region peak and a peak at around $1240cm^{-1}$ are all identifiable. However, the trough between amide I and II is almost completely lost, leading to difficulties in resolving the major amide II peak, and there is little resolution of secondary absorbance bands across the spectrum.

There is a loss of resolution of the minor peaks present across the spectrum, for example, the secondary peak seen at approximately $1125cm^{-1}$, and the minor peak identified at $967cm^{-1}$ in the Bruker Alpha spectra are not clearly resolvable in the IR Sphinx spectra. Additionally, the major peak at amide II is lost, forming only a small shoulder in the amide I peak. This drop in the ability to resolve individual peaks and shoulders limits the ability of the IR Sphinx to provide spectra from which accurate ratios of glycosylation patterns can be derived.

Shown in Figure 4-6, is a PCA scatterplot based on absorbencies at six wavenumbers shown to be significantly ($p < 0.05$) different between the two patient groups - 992, 1022, 1127, 1289, 1337 and $1700cm^{-1}$. This PCA was examined in terms of

its sensitivity and specificity for distinguishing COPD baseline and exacerbation sputum samples, with the best sensitivity of 60% and specificity of 81.48% if the sample fell within $PC1^{+}PC2^{-}$. A linear model provided increased specificity, but decreased sensitivity of 88.09% and 54.05% respectively. However, one must be careful when interpreting these results. The IR Sphinx is a lower resolution device, compared to the Bruker Alpha and Bruker Vertex instruments used in Chapter 3. The work presented in the previous chapter was not able to demonstrate that high resolution FTIR spectroscopy could discriminate COPD exacerbation from baseline in a reliable method. This heavily calls into question the accuracy of the results generated by the IR Sphinx for this chapter – why is the lower resolution IR Sphinx capable of achieving sensitivity and specificity scores for COPD exacerbation detection which the high-resolution benchtop instruments cannot? One possible explanation for this could relate to the respective sizes of the sampling crystals on each instrument. The Bruker Alpha has a small single bounce diamond ATR crystal, with a surface area of $4mm^2$, whereas the IR Sphinx has a much larger active area for sampling, of $425mm^2$. The architecture of the LVF detector present in the IR Sphinx requires that any sample is spread evenly across the ATR crystal to ensure that all wavelengths of IR light passing through the sample are measured by the detector. Therefore, much larger sample volumes are required, so a greater averaging effect of any subtle spectral differences could be present. It could be possible that this reduction in spectrum variability leads to an overall reduction in variability within the patient cohorts, which in turn allows a multivariate data reduction technique such as PCA to more easily group each patient cohort. However, care should be exercised when interpreting spectral results from low resolution instruments such as the IR Sphinx.

### 4.4.2 Design of the Sampling Strip

The IR Sphinx was originally designed as a robust, portable IR spectrometer, mainly for use as a rapid analytical tool in challenging environments, such as for oil quality analysis. For this reason it has been developed with no moving parts, and is encased in a rugged aluminium shell with a rubber bumper strip to protect it from impact. It also has many grooves running along its sides, for the user to grip the device more easily. The ATR crystal is recessed into the body of the aluminium shell. These design choices for the IR Sphinx are ideal for its originally intended purpose; however they make the device unsuitable for the clinical environment, for use by

either clinicians or patients. It must be completely redesigned prior to implementation into patient care and disease management.

Part of the redesign process of the IR Sphinx involved the designing of a novel, low-cost, disposable sampling strip, to which sputum samples will be applied for analysis. The sampling strip was constructed with a low IR-absorbing substrate, placed on top of the adhesive compressible foam to ensure that the pressure from clamping the strip in place was applied evenly across the sampling surface. The compressible foam was then laid on top of a rigid plastic backing, for handling and to maintain the shape of the sampling strip (Figure 4-2).

A literature search suggested AlO would be a suitable low-cost, low IR-absorbing substrate for FTIR analysis. Subsequent IR-absorption testing using the benchtop Bruker Alpha FTIR spectrometer showed that AlO deposited onto a PE film (AlO-film) absorbed less IR radiation across the mid-IR region of 4000-400cm$^{-1}$ than AlO directly deposited onto PU foam (M-PU) (Figure 4-7). Additionally it was found that the degree of IR-absorption by M-PU varied depending on the location within the sampling strip from which the spectrum is taken (Figure 4-8), which suggested the metallisation process of the PU foam was not uniform. Visual inspection of the M-PU after FTIR analysis revealed that the metallisation of the foam by AlO was not durable, as it was easily removed and damaged during spectrum acquisition (Figure 4-9). All of these factors contributed to the final decision that AlO-film was the substrate of choice for a low-cost and reliable sampling strip.

After testing the AlO-film for raw absorbance, sputum was dried onto the film and FTIR analysis on these sputum samples was carried out. Initial testing was performed using only one COPD patient sputum sample (TR06), which had previously been shown to produce spectra with a high degree of reproducibility, and which was in high enough supply for all desired replicate spectra. COPD sample TR06 was dried on the sampling strip prior to ATR-FTIR analysis (n=16), and these spectra were compared to TR06 sample spectra previously generated for Chapter 3 (n=3). It was found that the two sampling modes produced highly comparable IR-spectra, with an ANOVA test showing very little variation found between and within sampling modes for each peak position (Table 4-4) and no significant differences being shown in peak position between sampling modes (Table 4-5).

The sampling strip was required to contribute to maintaining close contact between the sputum sample and the ATR-crystal, with the assistance of a bespoke, in-house made clamp. To achieve this, the sampling strip was designed with compressible foam underneath the AlO-film substrate, in order to spread the applied clamping force across the whole sample and ATR-crystal. It was hypothesised that the properties of the foam could influence the quality of the produced spectra, by having a direct influence on the ability of the sampling strip to press the sputum sample against the ATR crystal. Multiple compressible foams from two manufacturers were evaluated to ensure the optimal material was used; from 3M the E1715h, E1515h, and E1120h tapes, and from TESA the Softprint 52122, and xSOFT 52223 tapes were evaluated. These foams were selected based on their varying hardnesses, densities and thicknesses. The performance of the tapes were evaluated based on the quality of the spectra produced by the IR Sphinx, using PGM dried onto sample strips constructed with the separate tapes. After comparisons of SNR, and variation within absorbance levels and peak positions in spectra produced when using each foam, it was found that the soft, thicker tapes (TESA xSOFT, TESA SoftPrint, and 3M-E1120h) performed best, with the lower variance in the spectral parameters observed. This suggests that softer, more pliable foams could be more suitable to applying and distributing the clamping force required to firmly and evenly hold the sampling strip in place.

### 4.4.3 Testing the Sampling Strip with COPD Sputum

Further testing was then carried out on all available COPD patient sputum samples (n = 118). It was found that AlO-film produced IR-spectra comparable to spectra generated using the standard operating procedure of drying sputum onto the ATR crystal using the Bruker Alpha. The mean correlation coefficient of the baseline-corrected, vector-normalised average-spectra of each sample generated using both sampling modes was found to be rho = 0.9787. This high degree of correlation is indicative of good similarity between sputum dried onto the ATR crystal and the sampling strip.

Statistical significance testing showed the peak positions of spectra generated using the sampling were significantly different from those generated by applying the sample directly to the ATR-crystal of the Bruker Alpha (Table 4-6). This suggests that use of the sampling strip does lead to significant changes in the features of any IR spectrum generated. The mean differences between the peak positions for each

sampling mode are shown to be relatively small, with the largest mean difference being observed at the peak detected at approximately 1638cm$^{-1}$. No significant correlation was identified between the shifts detected at amide I and shifts at any other major peak, suggesting that the causes of shifts at each peak are independent of one another. One suggested cause of peak shift is the differing effect of water interference in the spectrum when changing the sampling mode.

Figure 4-2(b) shows how the sample is inverted onto the ATR crystal after drying, with the topside of the dried sputum being in direct contact with the ATR crystal. This is in contrast to conventional ATR-FTIR in which the sample is dried directly on the ATR crystal so that the underside of the sample is in contact with the crystal. Water is known to produce strong peaks within the amide I region (1660-1620cm$^{-1}$) and can influence the absolute peak position within this region (Baker *et al.*, 2014; Barth, 2007). Additionally, it is also known that biofluids do not dry evenly on a surface: the edges of a sample drying faster than the centre leading to a coffee-ringing effect with higher levels of sample-molecules concentrating at the edges of the sample. Also the degree of coffee-ringing is directly influenced by the initial concentration of the sample molecules, with a stronger effect seen in more dilute biofluid samples (Baker *et al.*, 2015). The amide I peaks detected in the sampling strip cohort could be split into two distinct groups; centred at approximately 1636cm$^{-1}$, and centred at approximately 1650cm$^{-1}$). Infrared spectroscopy of proteins in an aqueous medium requires a short path length, and high concentration (Barth, 2007). It could be that by re-orientating the sample 'upside-down' for the measurement on an ATR crystal, the path length to higher concentrations of proteins may be shortened. Due to the coffee ringing effect, water is pulled by osmotic pressure towards the edges of the drying sample, and away from the centre.

No novel major peaks were identified or lost in any COPD sputum spectra generated with the sampling strip when compared with those generated using the standard procedure. These findings overall suggest that a sampling strip comprised of AlO and a compressible tape is a reliable method for the generation of sputum spectra.

### 4.4.4 Silicon ATR Wafers

A novel form of ATR-FTIR using silicon ATR wafers was tested in this chapter. The silicon wafers are a low-cost alternative to conventional ATR crystals, in which

diamond, ZnS/Se, or KBr are regularly used. These materials are high-cost, brittle, or soluble in water, making them unsuitable for use in the clinical environment without significant protocols in place to reduce the risk of damage to the device. The reproducibility of spectra generated using the silicon ATR wafers was assessed, and was shown to be extremely good (Figure 4-24 to Figure 4-27). Spectra were shown to have very little variation in absorbance intensity, or peak position, and any variation observed was deemed to be due to variations within the sample as variations specific to a particular wafer were not consistent.

### 4.4.5 Conclusions

The IR Sphinx is a robust, portable IR spectrometer, which has been shown to be capable of generating IR spectra of the complex biofluid, sputum. Analysis of COPD patient sputum using the IR Sphinx appears to suggest that differentiation of COPD baseline and exacerbation samples, using a simple analytical protocol, combined with a multivariate statistical approach is possible. However, one must be careful, as the work presented in chapter 3 showed that differentiating COPD baseline from COPD exacerbation sputum samples using high-resolution FTIR spectroscopy is problematic. One possible explanation for this disparity could be due to the sample volumes required to generate a spectrum, with much larger volumes required by the IR Sphinx, leading to an overall averaging effect of any differences between patient samples. Care is required when interpreting these spectra; subtle variation of absorbance bands and calculation of spectral ratios of absorbance bands which are close together is not possible, and this could limit the IR Sphinx's usefulness.

The lower-cost of the IR Sphinx, combined with its portability and ease-of-use, compared to other benchtop FTIR spectrometers, raises the possibility that miniaturised IR spectrometers could be integrated into patient primary care for rapid analysis of biofluids and disease diagnosis. However, further work needs to be carried out to develop miniaturised IR spectrometers to be more suited to the clinical environment; for example, the components within the IR Sphinx, whilst lower-cost compared to conventional FTIR, are costly to replace if broken and the devices are not user-friendly for a patient or clinician to use quickly on the ward. Proper operation of the device requires a separate computer plugged in to control the software and collate the data. For the device to be made truly suitable for the clinical environment, the IR Sphinx and controlling software must be incorporated into a single device, and be made of components that are economical to replace.

Disposable sampling strips have been developed which partially address some of the issues faced by the IR Sphinx. They are extremely low-cost and allow the user to quickly scan many samples without lengthy drying times after sample application to the ATR crystal. Applying the patient's sputum to the sampling strip allows the user to analyse samples without touching the ATR crystal, and therefore minimises the chances of breakages occurring during operation or cleaning. The sampling strips are held in place on the window using a clamp which can be set to not apply a force in excess of the breaking-point of any component of the IR Sphinx, further reducing the likelihood of breakages. High resolution spectroscopy of sputum dried onto the sampling strips did not identify any new spectral artefacts caused by the strips, suggesting that they are suitable for use as an FTIR substrate. However, statistically significant peak shifts were observed, at all major peaks across the spectrum, this was not shown to be consistent in all patient samples, suggesting that spectrum variation may be related to sputum sample variation.

Use of novel silicon wafers as ATR crystals has been shown to be a highly effective potential substrate for FTIR analysis of sputum, providing reproducible spectra of comparable quality to those generated using a conventional diamond ATR crystal. Further work into development of disposable sampling strips for clinical FTIR analysis should focus on optimising the use of silicon wafer ATR crystal. At €40 per crystal they are not yet of low-enough cost to be considered truly disposable. However optimisation through a reduction in size of the crystals, and up-scaled production could lead to much lower costs and the potential to use these crystals as a disposable item in the clinical environment.

# Chapter 5
# Evaluation of FTIR for Rapid Diagnosis of Lung Cancer and Distinction of Lung Cancer Subtypes

## 5.1 Introduction

### 5.1.1 Lung Cancer

Worldwide, lung cancer represents a burden on healthcare systems and is a major cause of mortality. Alongside bowel cancer, lung cancer is the second most common cause of cancer in adult males (13%) with prostate cancer being the most common at 26%. Lung cancer is also the second most common in adult females (12%), after breast cancer which accounts for 31% of all cancer cases in women (Cancer Research UK, 2018). In the UK, lung cancer is the most common cause of cancer death, currently accounting for approximately 21% of all cancer deaths (Cancer Research UK, 2018). Lung cancer patients also have a very poor 5-year survival rate of <10%, which is primarily due to a majority of patients being diagnosed only after the disease has progressed too far to be easily treated (Allemani *et al.*, 2015; McPhail, Johnson, Greenberg, Peake, & Rous, 2015; Morgan & Wilkes, 2017). The WHO identifies tobacco use as a major contributor to developing lung cancer, but there are also important environmental and occupational risks, as well a multitude of genetic factors. COPD patients have a higher risk of developing lung cancer as their forced expiratory volume in one second ($FEV_1$) declines (Skillrud *et al.*, 1986; Tockman *et al.*, 1987). Chronic inflammation is the proposed mechanism through which lung cancer and COPD are thought to be related (Houghton *et al.*, 2008). Diagnosis of lung cancer is influenced by the patients' background with a late-stage diagnosis more likely in the presence of comorbidities and disability (Ram, Young, Wook, Kook, & Hyock, 2017), with COPD as a comorbid condition being strongly associated with stage-independent poor survival (Gao *et al.*, 2016). Indeed, some COPD patients, especially those who have frequent exacerbations, can be accustomed to frequent changes in their condition (Seemungal *et al.*, 1998), and this may contribute further to a late-stage diagnosis of cancer as persistent changes to symptoms could be attributed to an exacerbation. There is an unmet need for a diagnostic tool to more readily identify lung cancer at an early-stage and to distinguish it from COPD.

The recommended procedure for the diagnosis of lung cancer is flexible bronchoscopy and has been shown to have an overall sensitivity for lung cancer diagnosis of 88%. However this sensitivity drops markedly for peripheral lesions of <2cm in diameter to 34% (Rivera *et al.*, 2013). There are two common forms of bronchoscopy, auto-fluorescence imaging (AFI) and white light bronchoscopy (WLB),

which have sensitivity scores of 80%, and 53.3%, respectively, while the specificities were 83.3% for AFI, and 50% for WLB (Andolfi *et al.*, 2016). Patients who are suspected of lung cancer also undergo diagnostic X-rays and computed-tomography (CT) scans to identify regions of malignant tissue. Almost 1 in 4 (23%) of diagnostic X-rays have been shown to provide false negative results (Stapley *et al.*, 2006). CT-scanning has greater sensitivity and specificity scores of 88.9% & 92.6%, respectively (Toyoda *et al.*, 2008), however it is costly and requires highly-trained operators to achieve this level of accuracy. Evidently, there is also a clear unmet need for a rapid, inexpensive, highly-sensitive and specific diagnostic tool, capable of diagnosing both centrally- and peripherally-located lesions.

### 5.1.2 Fourier-Transform Infrared Spectroscopy

FTIR is a vibrational spectroscopy technique which detects molecular bond conformations through measuring the absorbance of IR light at specific frequencies. A specific type of molecular bond, for example C-O, will absorb IR light at a frequency that correlates with the vibrational frequency of the bond. The exact vibrational frequency of the bond is influenced by ionic, electrostatic and dipole-dipole interactions, as well as the stretching and bending moments of the bond. Changes to these properties can have a significant impact on the absorbance of IR-light and can therefore influence the overall spectrum.

A great deal of work has been carried out to establish FTIR as a suitable technology for diagnosis of various cancers and other diseases. An increasing number of publications show that FTIR is a viable tool for analysing biofluid samples (Baker *et al.*, 2014), and diagnosis and monitoring of diseases such as brain cancer (Hands *et al.*, 2016), cystinuria (Oliver *et al.*, 2016), breast cancer (Backhaus *et al.*, 2010) (Backhaus *et al.*, 2010) and oropharyngeal cancers (Menzies *et al.*, 2014). Previous work in our group has already shown FTIR to be capable at distinguishing lung cancer from COPD with a high sensitivity and specificity by analysing cell pellets isolated from spontaneous sputum samples (Paul D. Lewis *et al.*, 2010). Additionally, a US-based group has shown that spectral histopathology (SHP) can diagnose lung cancer with accuracy comparable to multi-panel immuno-histochemistry (Bird *et al.*, 2012). Both of these studies provide excellent evidence for the proof-of-concept of using FTIR to diagnose and detect lung cancer at an early-stage. However there are limitations to both of these studies, chiefly the low n-number and complex protocols involved. This chapter seeks to improve and extend beyond this proof-of-concept by increasing the

number of cases analysed, but also by becoming more clinically relevant with a simplified protocol.

### 5.1.3 Chapter Aims

The main aim of this chapter is to evaluate the utility of FTIR as a tool for lung cancer diagnosis through comparison of lung cancer and COPD patient sputum spectra. FTIR analysis performed on lung cancer and COPD patient raw sputum samples will be the basis for the development of predictive diagnostic models for lung cancer against the patients' confirmed clinical diagnosis. Models will be assessed in terms of their sensitivity and specificity scores for accurate lung cancer detection.

## 5.2 Methods

### 5.2.1 Patient Recruitment

The Medlung observational study (loco-regional ethical committee approval: 05/WMW01/75) recruited patients who attended bronchoscopy clinics across the UK under suspicion of lung cancer and were subsequently given a final clinical diagnosis of either "lung cancer" or "non-cancer". The non-cancer group could be split into those who had a diagnosis of COPD, and those who have other respiratory diseases including pneumonia, asthma, and/or tuberculosis. Patients were referred by their GPs to the bronchoscopy clinics in the recruiting hospitals after showing signs of lung cancer. Patients gave informed consent before providing a sample of spontaneous sputum. The final clinical diagnosis after bronchoscopy, biopsy and any other relevant scans was recorded along with the histological data for each cancer. Confirmed cancer cases and confirmed COPD cases make up the "Cancer", and "Bronchoscopy COPD control" (bCOPD) cohorts respectively.

The SPEDIC diagnostic study (loco-regional ethical committee approval: 15/LO/1703) recruited patients who presented at the respiratory clinic in Prince Charles Hospital (Cwm Taf, Merthyr Tydfil) due to showing symptoms of chronic respiratory disease, but not COPD. These patients had never attended the clinic before, had no previous diagnosis of COPD or COPD exacerbation and make up the "Non-Cancer-Non-COPD control" (NC) cohort. Additionally, 55 COPD patients were also recruited to a year-long longitudinal trial during which they provided 5 sputum samples per week. For the work presented in this chapter, 40 sputum samples from patients at baseline COPD were randomly selected from this pool of samples. These COPD patients are used as a random control, as the samples are taken from any point

in time when the patient is not presenting at hospital, and are treated as the "Random COPD control" (rCOPD) cohort.

In total, raw sputum samples were collected from 214 lung cancer patients with a confirmed metastatic status, 108 bCOPD patients, 40 randomly selected non-exacerbation sputum samples from the rCOPD cohort and 46 NC patients (Table 5-) – see appendix 1 for detailed patient information. Each sputum sample was stored at -80°C until required for FTIR spectrum generation.

| | Patients | Undistinguished NSCLC | Adenocarcinoma | Squamous | SCLC |
|---|---|---|---|---|---|
| Confirmed Cancer Diagnosis | 252 | 53 | 89 | 46 | 64 |
| Total Confirmed With Full Histology | 214 | 42 | 83 | 41 | 48 |
| Metastatic | 106 | 29 | 39 | 18 | 23 |
| Non-Metastatic | 108 | 13 | 44 | 23 | 25 |
| bCOPD cases | 108 total | 74 Confirmed COPD Baseline | | 34 Confirmed COPD Exacerbation | |
| rCOPD cases | 40 total | 40 Confirmed COPD Baseline > 6 weeks after most recent exacerbation | | | |
| NC cases | 46 total | Mixed cohort of diagnoses of non-cancer, non-COPD respiratory disease | | | |

**Table 5-1: Lung cancer and COPD patient cohort data; cancer histological subtypes and COPD exacerbation status recruited to study are also shown**

### 5.2.2 Spectrum Acquisition and Sample Processing

Transmission-FTIR (t-FTIR) was carried on raw sputum samples using a Bruker Vertex 70 with high throughput attachment (HTS-XT), a KBr beamsplitter, and a DGTS detector. Ninety-six well silicon plates (Bruker) were cleaned in 70% ethanol, rinsed with $dH_2O$ three times then air dried. Triplicate raw sputum samples were pipetted directly onto the plates and allowed to dry in atmospheric conditions. Once dry, spectra were generated at 32 scans per spectrum, with a fresh background spectrum taken between each sample spectrum. Each 96-well plate was scanned in triplicate, giving a total of 9 replicate spectra per sample.

### 5.2.3 Spectrum Processing

All spectra underwent a quality analysis before processing. A spectrum was discarded if the signal-to-noise ratio (SNR) was too low and/or if the baseline of the spectrum was sloped or oscillatory which is indicative of excessive scattering (Baker *et al.*, 2015). All sample replicates were averaged before vector-normalisation and baseline-correction using the OPUS 7.5 (Bruker) in-built baseline-correction, vector-normalisation algorithms. Second derivative spectra were generated using the Savitzky-Golay method with 9 smoothing points. Peak picking analysis was carried out using the in-built peak peaking algorithm in OPUS, set to a 10% threshold.

### 5.2.4 Statistical Analysis

Statistical tests were carried out using the programming environment R (R Core Team, 2016). Testing for distribution normality was carried out using the Shapiro-Wilk test, with an α-level set to 0.05, suggesting that the null hypothesis of the data being drawn from a normally-distributed population could be rejected, alongside Q-Q plots and histograms to visualise the distribution. As normality testing suggested non-normally distributed data, statistical significance was calculated using the non-parametric Mann-Whitney U Test, at the 95% confidence level. Multivariate analyses were carried out to establish how groups cluster using principle components analysis (PCA).

### 5.2.5 Model Building

Two dimensional scatterplots were produced comparing absorbencies at specific wavenumbers. Sensitivity and specificity calculations were carried out using the linear equations of the separator line. The wavenumbers of initial interest were those previously identified to be discriminatory for lung cancer, 966, 1024, 1051, 1411, 1577, and 1654cm$^{-1}$ (P. D. Lewis *et al.*, 2010), as well as other prominent peaks identified during peak picking analysis.

## 5.3 Results

High-throughput transmission FTIR spectroscopy was performed as described on sputum samples obtained from lung cancer patients recruited through the MEDLUNG study. In total, 214 lung cancer sputum samples were analysed by FTIR – absorbance spectra shown in appendix 2. These spectra were compared to FTIR spectra from 148 COPD sputum samples and 46 non-cancer-non-COPD (NC) respiratory disease sputum samples.

### 5.3.1 Normality and Significance Testing

Normality testing was carried out to ascertain how the absorbencies at each wavenumber were distributed across the cancer and control cohorts. A Shapiro-Wilk (SW) test for normality on five wavenumbers which correspond with positions of major peaks and troughs was initially carried out, and the results are summarised in Table 5-2. The results suggested that the null hypothesis that the data were drawn from a normally distributed population can be rejected, therefore indicating the data are non-normally distributed.

| Wavenumber ($cm^{-1}$) | Cancer | Non-Cancer |
|---|---|---|
| **1740** | < 2.2e-16 | 6.68E-15 |
| **1653** | < 2.2e-16 | 2.06E-14 |
| **1589** | 0.00102 | 0.006305 |
| **1410** | 3.11E-08 | 8.18E-06 |
| **1076** | 6.38E-07 | 0.0723 * |

**Table 5-2: Results from Shapiro-Wilk test for normality of distribution of absorbencies at 1740, 1653, 1589, 1410 and 1076$cm^{-1}$ in cancer and non-cancer control cohorts. P<0.05 suggests that the null hypothesis of normally distributed data can be rejected and the data are non-normally distributed. *P >0.05, the null hypothesis cannot be rejected, the data are normally distributed.**

Q-Q plots and histograms were drawn to visualise distribution (Figure 5-1). The distribution of absorbencies at 1653$cm^{-1}$ was shown to be heavily tailed, and the distribution at 1076$cm^{-1}$ can be said to be closer to normality, but still demonstrated a light skewing in both cancer and control cohorts. Therefore, combining these results with the Shapiro-Wilk normality results, the data can be said to be non-normally distributed. Thus, non-parametric statistical testing was carried out to determine significance of differences between absorbencies at wavenumbers of interest.

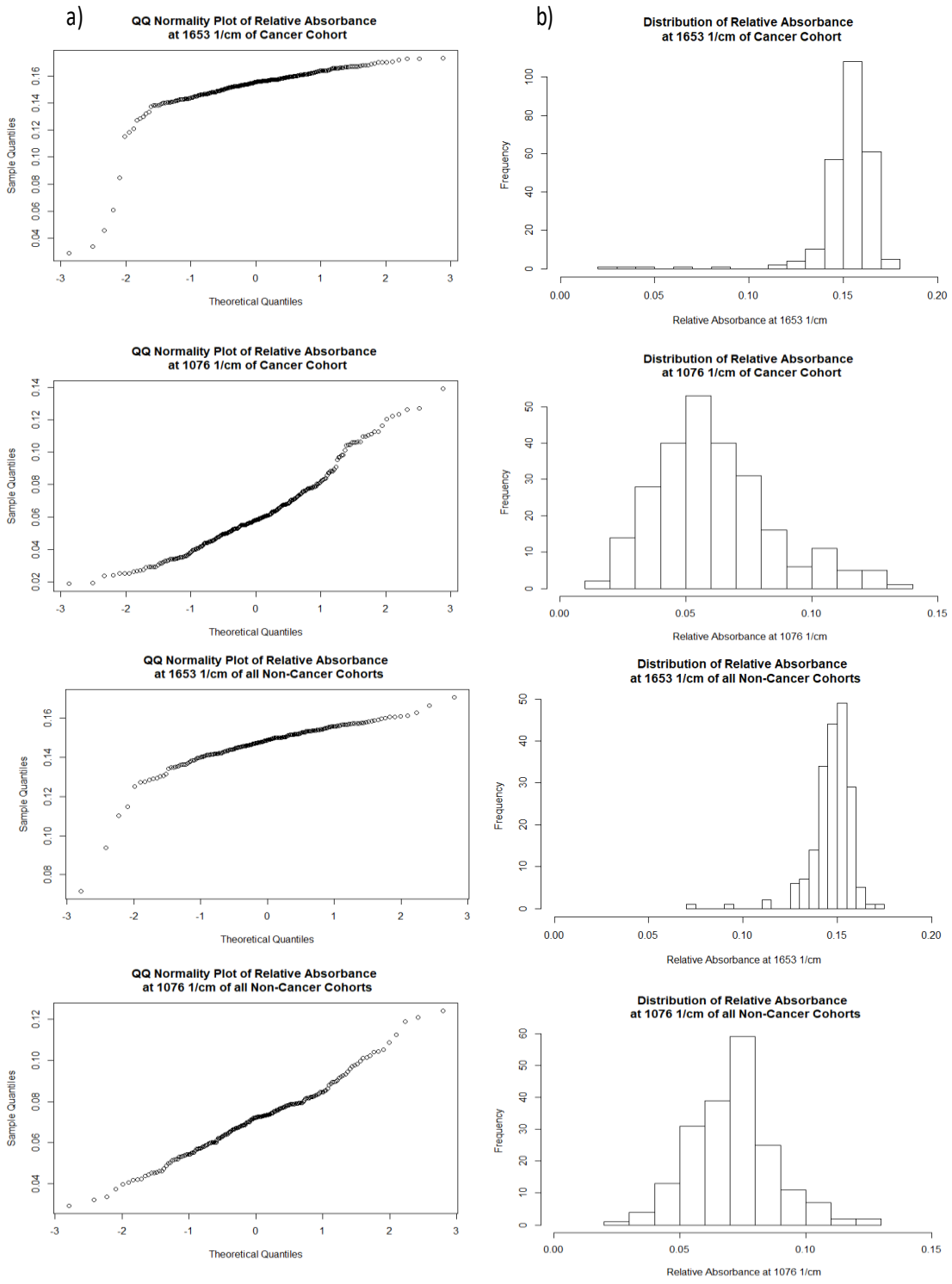


**Figure 5-1: A) QQ Normality plots and B) frequency histograms of distribution of absorbencies at 1653cm$^{-1}$ and 1076cm$^{-1}$. Plots of distribution at 1653cm$^{-1}$ show a heavy-tailed distribution. Distribution at 1076cm$^{--1}$ is closer to normality but the cancer cohort shows a light-tailed distribution and the control cohorts show a light skewing.**

As all wavenumbers tested were shown to be drawn from non-normally distributed data. The non-parametric Mann-Whitney U test was carried out to assess the statistical significance of any differences between the mean absorbencies.

### 5.3.2 Absorbance Spectra Diagnostic Models

Shown in Figure 5-2 are vector-normalised, baseline-corrected average absorbance spectra of the cancer, rCOPD, bCOPD and NC cohorts (red, blue, purple and green respectively). Whilst the average spectra appear to be highly similar, with little variation in peak position or relative absorbance, subtle differences can be identified.

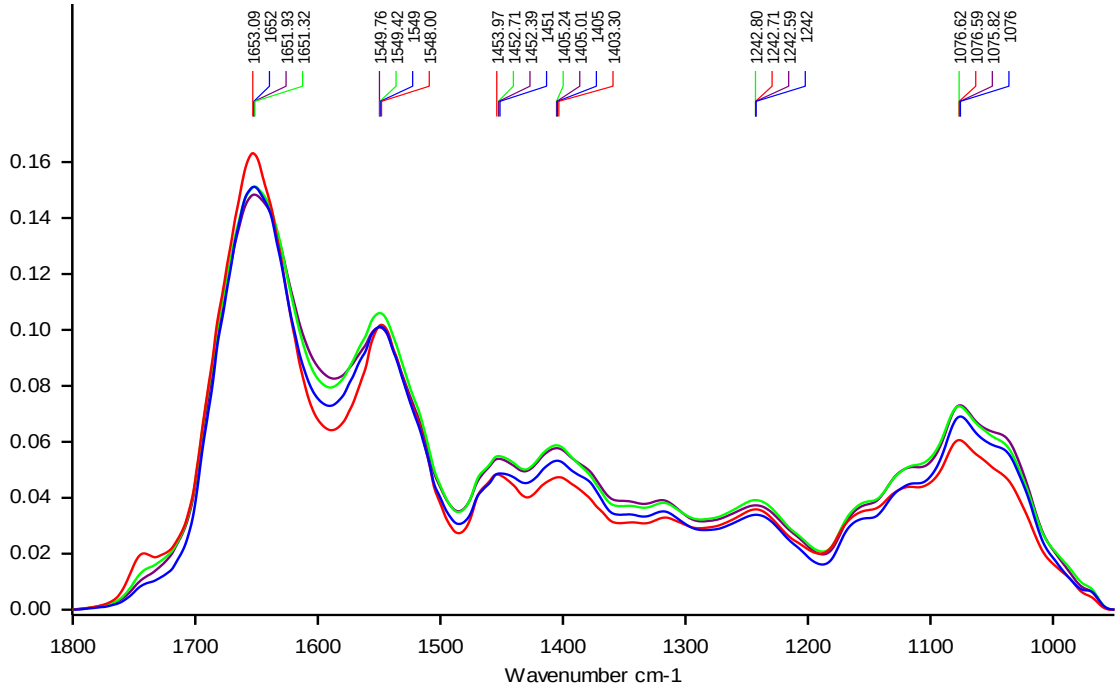


**Figure 5-2: Average vector-normalised, baseline-corrected absorbance spectra from 1800-950cm$^{-1}$ of cancer (red), rCOPD (blue), bCOPD (purple) and NC (green) cohorts, with major peak positions shown. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

For example, the relative intensity of multiple peaks and troughs can be seen to be different between cancer and non-cancer average spectra. For example, the relative absorbance at amide I (~1653cm$^{-1}$) is higher in the cancer average spectrum, and the major glycogen peak (~1076cm$^{-1}$) is lower compared to the non-cancer average spectra. The proposed vibrational mode of 1653cm$^{-1}$ is C=O stretching from a protein source. The proposed vibrational mode of the glycogen peak at ~1076cm$^{-1}$ is C-O stretching, from the alcohol groups found within individual monosaccharide moieties throughout the sputum. This suggests that overall glycosylation compared to protein content in lung cancer sputum could be reduced, compared to COPD and NC patient sputum. Additionally, the trough between amide I and amide II of the cancer spectrum appears to be lower than all of non-cancer spectra, whilst the amide

I peak shows a greater relative intensity than the non-cancer amide I peaks. This may suggest a reduction in the levels of amino-sugars such as sialic acid, *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine (GlcNAc) relative to the levels of protein present in lung cancer sputum, as amino-sugars have been shown to produce peaks at 1550, 1564 and 1628cm$^{-1}$ (S. P. Lewis *et al.*, 2013).

Figure 5-3 demonstrates how the different patient cohorts cluster and separate from each other based on the absorbance at 1076 and 1653cm$^{-1}$, as well as showing the overall trend within each patient cohort. There appears to be a trend for cancer sputum spectrum to show an increase in relative absorption at 1653cm$^{-1}$ alongside a decrease in relative absorption at 1076cm$^{-1}$. In fact, the absorption at these wavenumbers was found to be statistically significantly different when comparing the cancer cohort against all of the control cohorts simultaneously (Table 5-3). The patient groups cluster together well, with very little separation evident, especially within the NC and COPD control cohorts. Linear regression lines were drawn at three locations and the equations calculated. Using these equations the sensitivity and specificity of using only absorbance at two wavenumbers of normalized-baseline-corrected spectra was calculated.

| **Wavenumber (cm$^{-1}$)** | **Mean Cancer Spectrum Absorbance** | **Mean Non-Cancer Spectrum Absorbance** | **p-Value** |
|---|---|---|---|
| **1740** | 0.020078 | 0.011667 | 7.185e-15 |
| **1653** | 0.152049 | 0.146985 | 1.02*10-12 |
| **1589** | 0.065135 | 0.078781 | < 2.2e-16 |
| **1411** | 0.048932 | 0.055263 | 1.958e-09 |
| **1076** | 0.061379 | 0.071092 | 2.45*10-9 |

**Table 5-3: Results of significance testing, comparing the normalized absorbencies at each wavenumber between the cancer and non-cancer cohorts with a Mann-Whitney U test; all wavenumbers tested were shown to be highly-significantly different between the patient groups.**

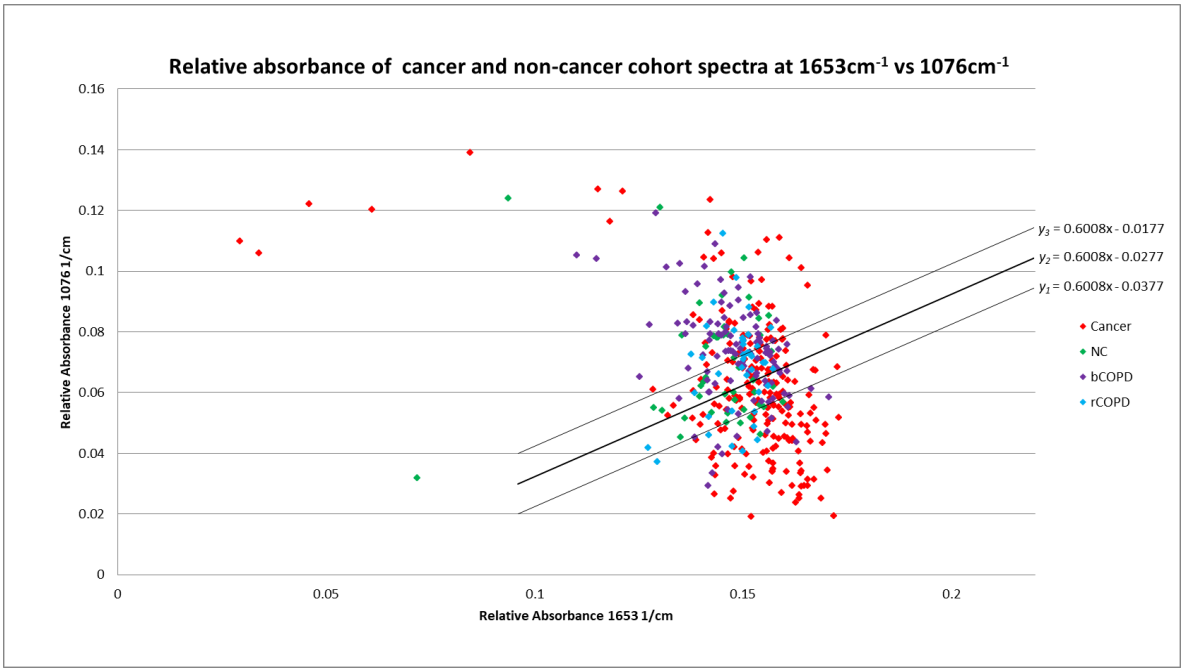


**Figure 5-3: Vector-normalised baseline-corrected absorbencies at 1076cm$^{-1}$ and 1653cm$^{-1}$ of all cancer (red), NC (green), bCOPD (purple) and rCOPD (blue) sample spectra. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

The calculated sensitivity and specificity scores for a model based on normalized absorbencies at 1653cm$^{-1}$ and 1076cm$^{-1}$ are shown below in Table 5-4. The cancer and non-cancer patient cohorts exhibit a large overlap, so the intercept of the linear separator was modified to optimise sensitivity and specificity scores. The most accurate equation was determined to be $y_2$, with sensitivity and specificity of 61.11% and 71.65% respectively. The other models demonstrated stronger specificity but poor sensitivity ($y_1$) or vice versa ($y_3$).

| ***1076 vs 1653 cm$^{-1}$*** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***$y_1$ = 0.6008x - 0.0377*** | | | ***$y_2$ = 0.6008x - 0.0277s*** | | | ***$y_3$ = 0.6008x - 0.0177*** | |
| **True Positives** | 109 | | **True Positives** | 154 | | **True Positives** | 188 |
| **False Positives** | 25 | | **False Positives** | 55 | | **False Positives** | 104 |
| **True Negatives** | 169 | | **True Negatives** | 139 | | **True Negatives** | 90 |
| **False Negatives** | 143 | | **False Negatives** | 98 | | **False Negatives** | 64 |
| **Sensitivity** (%) | *43.26* | | **Sensitivity** | *61.11* | | **Sensitivity** | *74.60* |
| **Specificity** (%) | *87.11* | | **Specificity** | *71.65* | | **Specificity** | *46.39* |

**Table 5-4: Sensitivity and specificity scores for determining lung cancer from non-cancer control groups based on the equations of the three lines shown in Figure 5-3.**

As seen in Table 5-3, other absorbance values at many wavenumbers were shown to be statistically significantly different between the cancer and non-cancer patient cohorts. Combinations of these wavenumbers were assessed for their sensitivity and specificity. The best performing linear regression model based on absorbance spectra was found to be 1589 vs 1740cm$^{-1}$, with sensitivity and specificity scores for lung cancer detection of 77.78% and 81.9%, respectively.

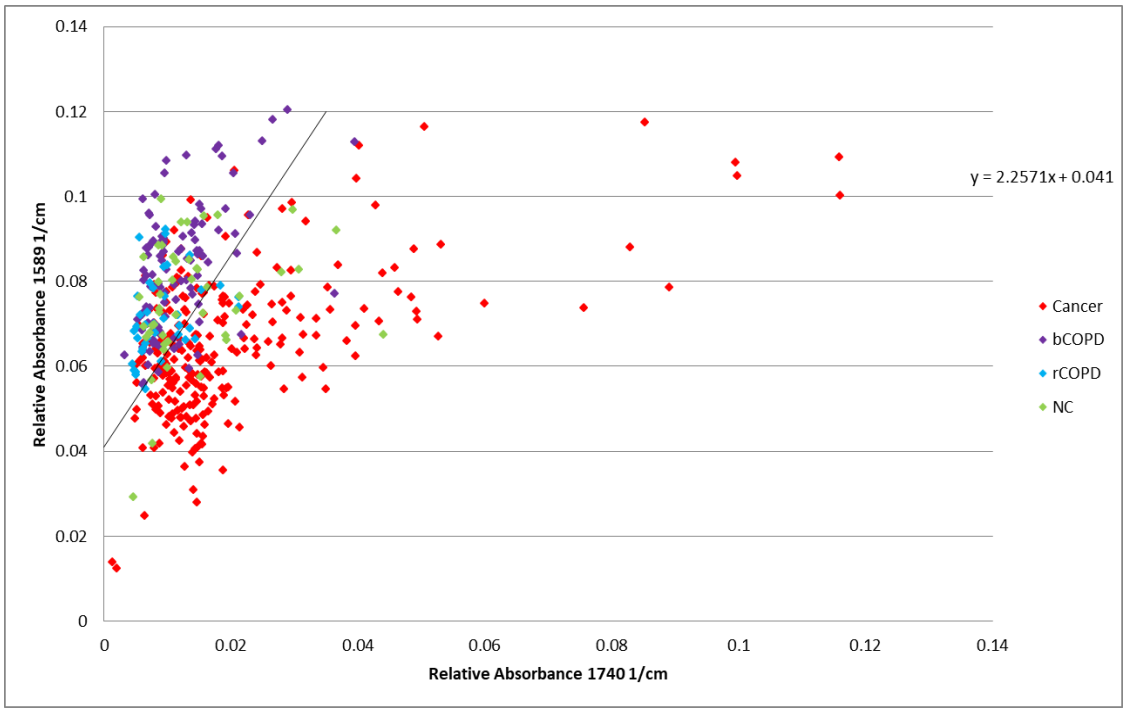


**Figure 5-4: Two-dimensional, linear separation model of normalised absorbencies at 1589 and 1740cm$^{-1}$, showing absorbencies from lung cancer (red), bCOPD (purple), rCOPD (blue) and NC (green) sputum sample spectra. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

The mean absorbencies at five wavenumbers hypothesised to have good power for lung cancer detection were found to be statistically significantly different at $\alpha$=0.05 (Table 5-3). However, the regression models shown Figure 5-3 and Figure 5-4 display poor separation of cancer and non-cancer patient cohorts, confirmed by overall sensitivity and specificity scores of <80%.

This implies that further methods such as deconvolution of the spectra through second-derivative calculation, or multivariate statistics are necessary to further elucidate any differences between the patient groups.

### 5.3.3 Multivariate Statistics on Vector-Normalised, Baseline-Corrected Absorbance Spectra

PCA was carried out using the same wavenumbers previously identified to be able to distinguish lung cancer from healthy controls (P. D. Lewis *et al.*, 2010). All wavenumbers were found to be statistically significantly different between the patient cohorts at $\alpha$=0.05 after multiple hypothesis testing.

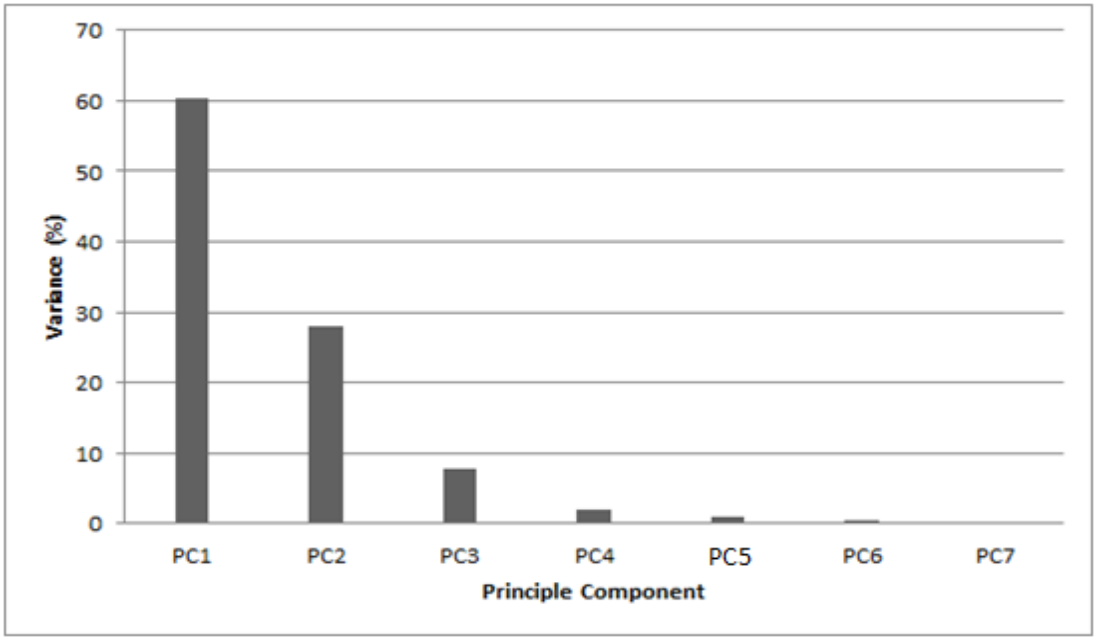


**Figure 5-5: Scree plot of percentage of variances accounted for by each PC, with approximately 60% of the variance being accounted for by PC1, 27% by PC2 and 8% by PC3. PC4, 5, 6, and 7 accounts for less than 5% of the variance each and so are not included in the analysis. PC – Principle component**

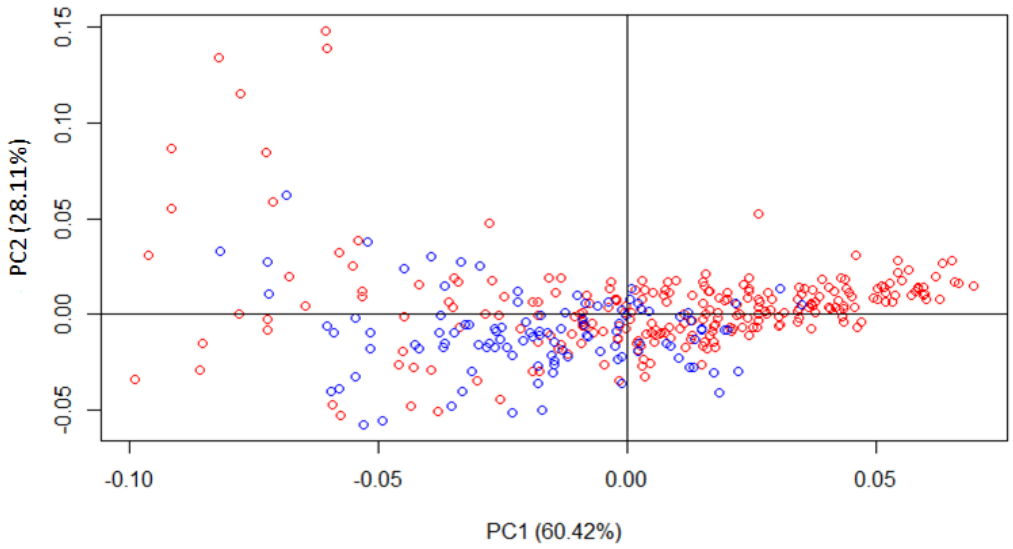


**Figure 5-6: PCA scatterplot of PC1 vs PC2 showing clustering of cancer (red) from non-cancer (blue) sputum sample spectra based on the six wavenumbers previously specified**

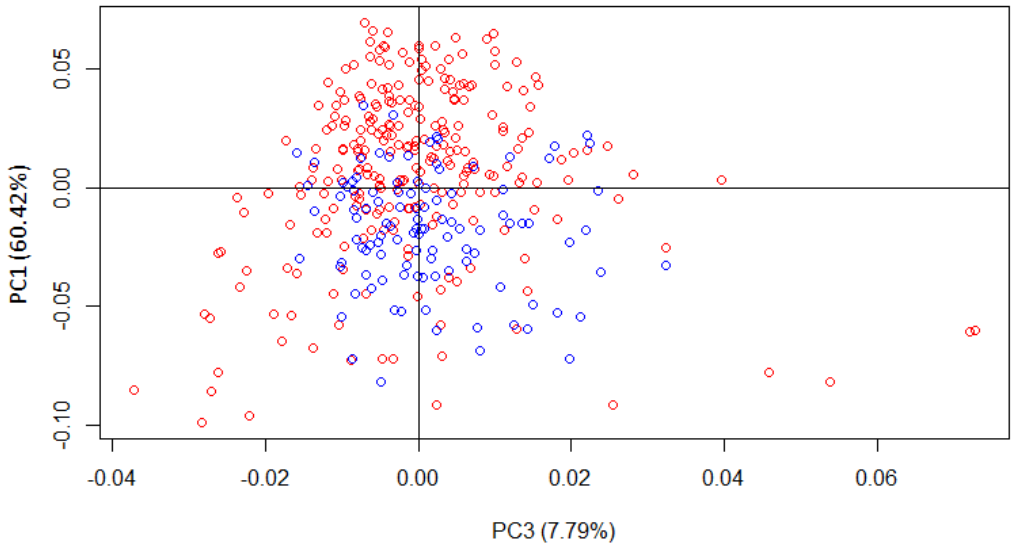


**Figure 5-7: PCA scatterplot of PC1 vs PC2 showing how the PCs separate cancer (red) from non-cancer (blue) sputum sample spectra based on the 6 wavenumbers previously specified**

PCA analysis on the absorbance spectra focussed on the wavenumbers described above was not able to increase the discrimination of cancer sputum from non-cancer respiratory disease sputum compared to the linear absorbance models. Two-dimensional PCA scatterplots containing PCs explaining more than 5% of the variances were plotted (Figure 5-6 and Figure 5-7). PC1 and PC2 (Figure 5-6) show a horseshoeing effect which is a statistical artefact and suggests any separation of patient groups in this manner is not genuine (P. D. Lewis & Menzies, 2015). Visualising PC1 and PC3 together (Figure 5-7) show how the cancer spectra tend to group towards PC1+ whilst the non-cancer spectra tend towards PC1-. PC3 does not display good separation of patient clusters, but the horseshoeing artefact seen in PC1/PC2 is not observed.

### 5.3.2 Second Derivative Absorbance Models

Second derivative spectra were calculated from the vector-normalised, baseline-corrected average spectra. A second-derivative spectrum is used to increase the sensitivity for peak finding by calculating the rate-of-change across a small window of the spectrum. In this way, the resolution of small peaks and shoulders within larger peaks is increased.

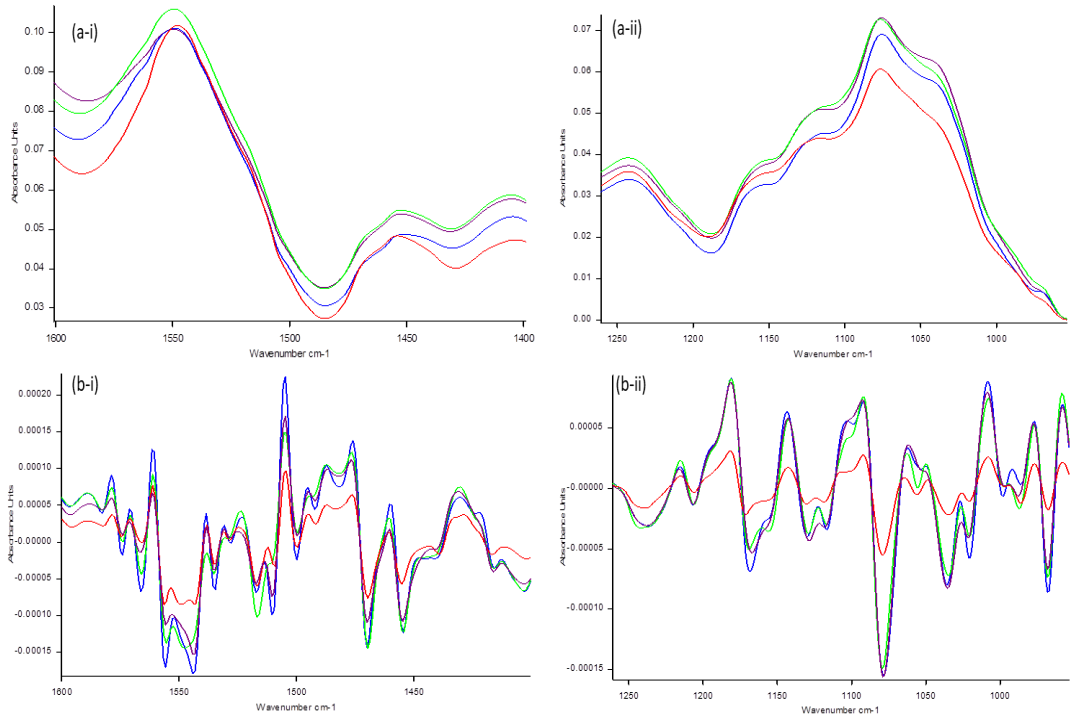


**Figure 5-8: (a) Absorbance spectra and (b) corresponding second-derivative spectra of (i) 1600-1400cm$^{-1}$ showing the amide II region, and (ii) 1250-900cm$^{-1}$ showing the glycogen-rich region of cancer (red), bCOPD (purple), rCOPD (blue) and NC (green) patient average spectra. Peak maxima in the absorbance spectra correspond with peak minima in the second-derivative spectra. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

The average second-derivative spectra were closely examined to identify regions of the spectra that could be used to distinguish cancer from non-cancer. Wavenumbers previously identified to have good discriminatory potential for lung cancer sputum were selected for examination (P. D. Lewis *et al.*, 2010), alongside novel wavenumbers being selected by examination of the average second-derivative spectra.

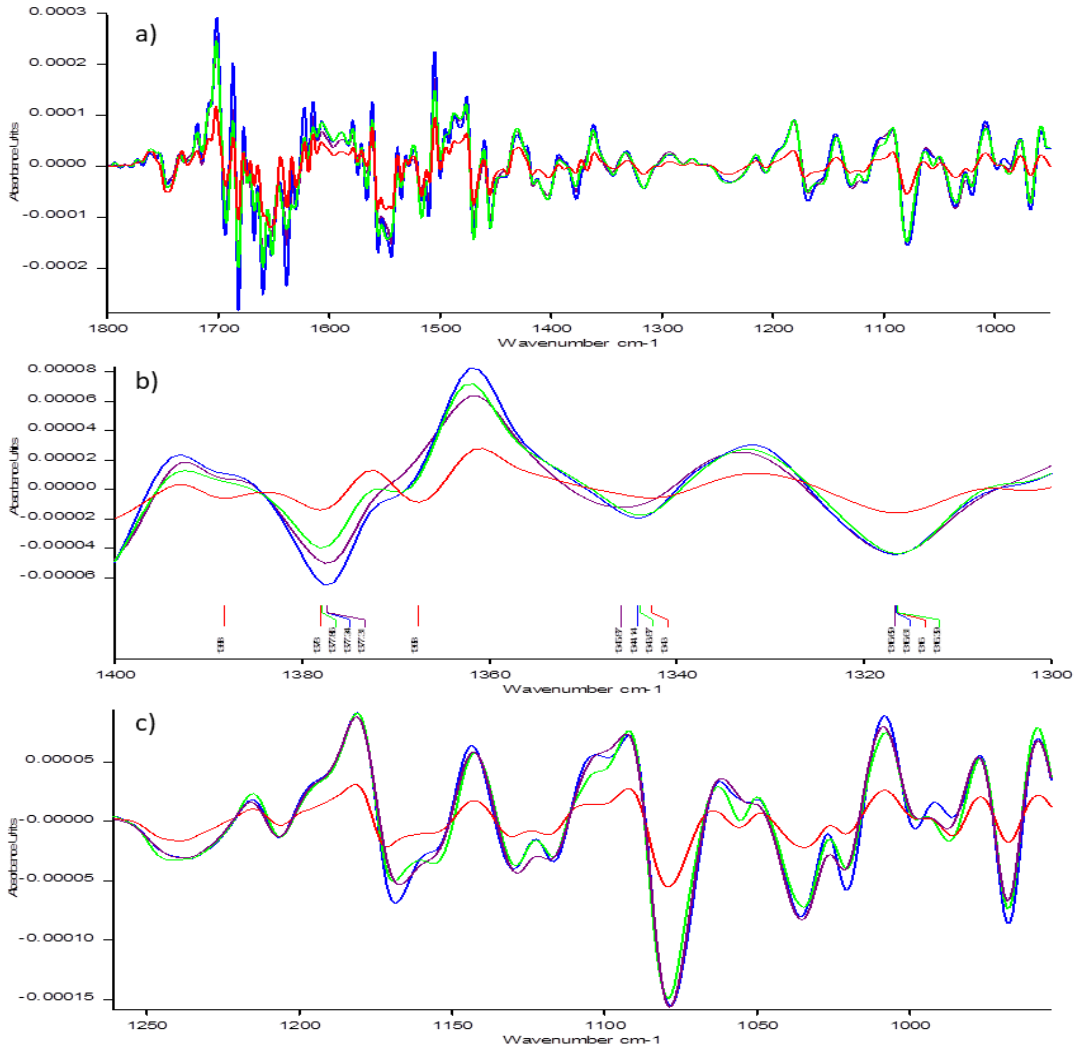


**Figure 5-9: Second derivative spectra from (a) 1800-950cm$^{-1}$ (b) 1400-1300cm$^{-1}$, and (c) 1250-950cm$^{-1}$ of cancer, rCOPD, bCOPD and non-cancer-non-COPD cohorts. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

Linear regression models were generated to examine how the second-derivative absorbencies at specific wavenumbers could separate the cohorts within two dimensions. The second-derivative absorbance at 967.598cm$^{-1}$ was initially chosen as a standard for plotting against other wavenumbers. This was because it was readily identifiable in all spectra and was calculated to have the lowest standard deviation and variance compared to all other major peaks detected in the cancer (Table 5-5), rCOPD (Table 5-6) and NC cohorts (Table 5-7). It has been previously shown to be associated with presence of nucleic acids and phosphorylation of proteins (Malins *et al.*, 2005; Maziak *et al.*, 2007) as well as the presence of the Lewis X (Le$^{x}$) antigen within the mucin glycans (A. T. Lewis *et al.*, 2013).

| **Mean Peak Position** | *1170.88* | *1129.58* | *1079.14* | *1056.84* | *1034.53* | *1020.45* | *998.01* | *985.79* | *967.59* |
|---|---|---|---|---|---|---|---|---|---|
| **Number of Samples Peak Detected (/60)** | 47 | 31 | 58 | 29 | 45 | 32 | 30 | 31 | 55 |
| **Mode** | 1171 | 1130 | 1079 | 1057 | 1036 | 1021 | 999 | 984 | 968 |
| **Std Dev** | 1.686 | 1.298 | 1.341 | 1.68 | 2.58 | 2.050 | 2.39 | 3.19 | 1.13 |
| **Min peak position** | 1167.70 | 1127.35 | 1071.48 | 1053.87 | 1027.70 | 1013.60 | 992.87 | 976.24 | 964.27 |
| **Max peak position** | 1173.53 | 1131.80 | 1081.16 | 1059.91 | 1040.11 | 1022.39 | 1002.05 | 991.36 | 970.06 |
| **Variance** | 2.783 | 1.631 | 1.767 | 2.73 | 6.51 | 4.07 | 5.55 | 9.86 | 1.25 |

**Table 5-5: Average major peak positions within the glycogen-rich region from a subset of 60 randomly selected lung cancer patient second-derivative spectra. Standard deviation and variance for each peak has been calculated and the lowest values are highlighted.**

| **Mean Peak Position** | *1168.41* | *1129.91* | *1078.65* | *1053.93* | *1035.57* | *1020.70* | *998.40* | *985.19* | *967.73* |
|---|---|---|---|---|---|---|---|---|---|
| **Number of Samples Peak Detected (/40)** | 39 | 36 | 40 | 11 | 38 | 36 | 34 | 23 | 40 |
| **Mode** | 1168 | 1130 | 1079 | 1055 | 1035 | 1021 | 999 | 984 | 968 |
| **Std Dev** | 2.15 | 1.46 | 1.91 | 2.99 | 1.53 | 0.64 | 0.69 | 1.34 | 0.41 |
| **Min peak position** | 1160.95 | 1124.97 | 1073.87 | 1049.58 | 1031.345 | 1019.55 | 996.64 | 983.64 | 966.76 |
| **Max peak position** | 1172.13 | 1132.31 | 1084.94 | 1059.82 | 1038.77 | 1022.35 | 999.66 | 989.52 | 969.08 |
| **Variance** | 4.52 | 2.07 | 3.54 | 8.179 | 2.271 | 0.401 | 0.456 | 1.727 | 0.16 |

**Table 5-6: Average major peak positions within the glycogen-rich region from all rCOPD patient second-derivative spectra. Standard deviation and variance for each peak has been calculated and the lowest values are highlighted.**

| **Mean Peak Position** | *1169.36* | *1129.39* | *1079.21* | *1055.87* | *1034.96* | *1020.42* | *998.83* | *986.14* | *967.91* |
|---|---|---|---|---|---|---|---|---|---|
| **Number of Samples Peak Detected (/46)** | 38 | 37 | 46 | 16 | 40 | 41 | 25 | 35 | 44 |
| **Mode** | 1169 | 1131 | 1078 | 1055 | 1033 | 1021 | 999 | 985 | 968 |
| **Std Dev** | 1.53 | 2.08 | 2.28 | 1.73 | 2.33 | 1.50 | 1.04 | 2.14 | 0.95 |
| **Min peak position** | 1166.18 | 1123.32 | 1073.24 | 1054.18 | 1027.53 | 1014.12 | 994.32 | 982.18 | 964.82 |
| **Max peak position** | 1172.17 | 1132.53 | 1085.07 | 1060.74 | 1039.13 | 1022.69 | 999.96 | 992.35 | 969.54 |
| **Variance** | 2.35 | 4.32 | 5.20 | 2.98 | 5.41 | 2.26 | 1.08 | 4.56 | 0.90 |

**Table 5-7: Average major peak positions within the glycogen-rich region from all NC patient second-derivative spectra. Standard deviation and variance for each peak has been calculated and the lowest values are highlighted.**

A series of linear regression models were built and tested for sensitivity and specificity for determining lung cancer from non-cancer respiratory disease, and are shown below:

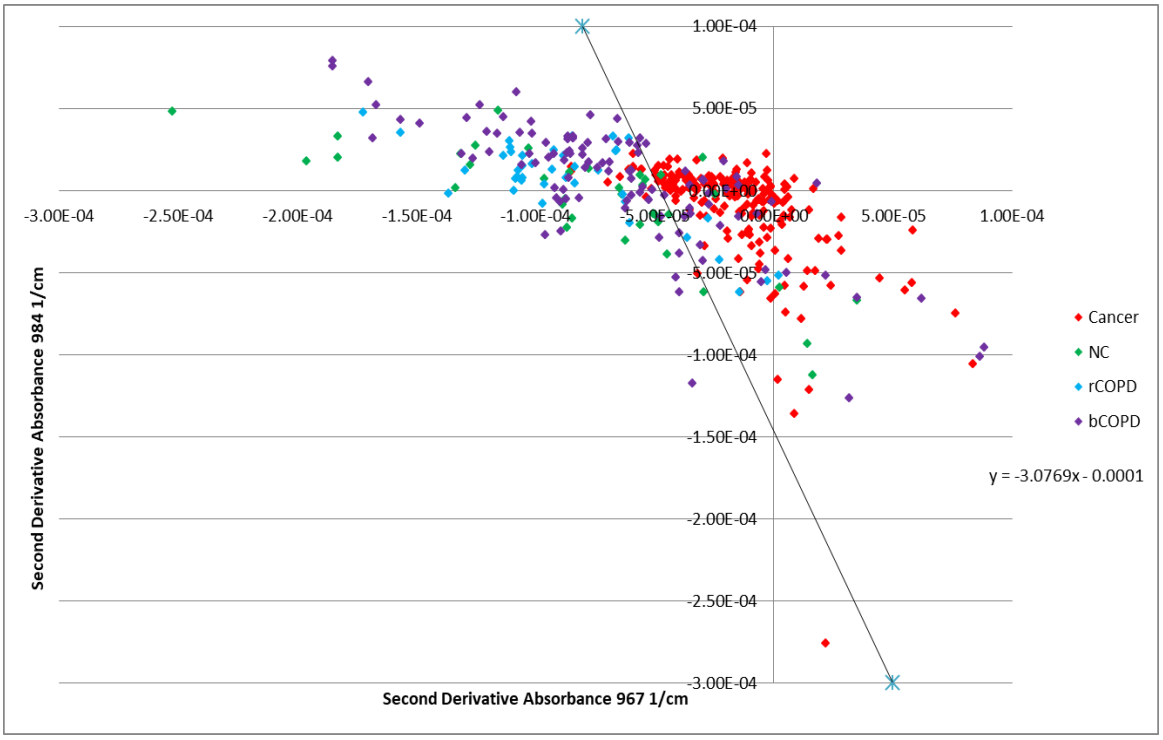


**Figure 5-10: Two dimensional scatterplot of second-derivative absorbencies at 967cm$^{-1}$ and 1051cm$^{-1}$, showing good separation of cancer (red) and non-cancer (blue, purple, green) sputum sample spectra, primarily along the 967cm$^{-1}$ axis. A linear regression line for cancer prediction is shown at 93.94% sensitivity and 75.26% specificity. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

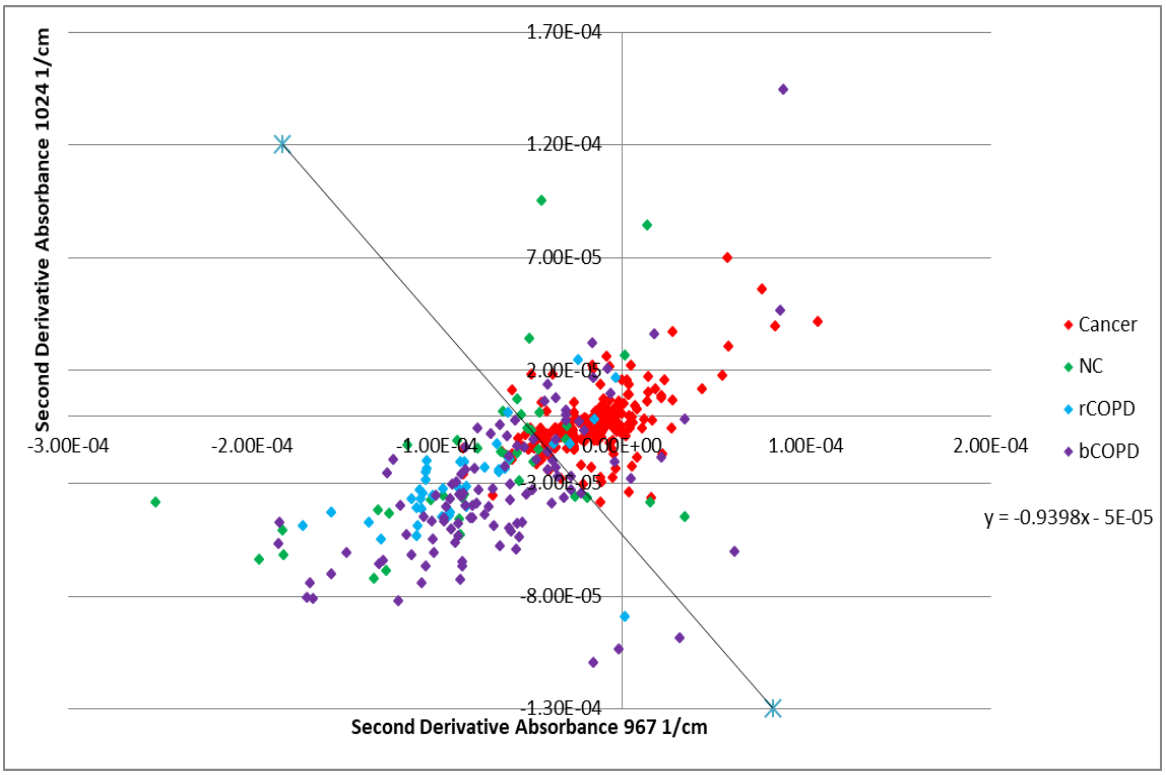


**Figure 5-11: Two dimensional scatterplot of second-derivative absorbencies at 967cm$^{-1}$ and 1024cm$^{-1}$, showing a linear relationship between the wavenumbers and separation of cancer (red) and non-cancer (blue, purple, green) sputum sample spectra across**

**both wavenumbers. A linear regression line for cancer prediction is shown at 85.58% sensitivity and 81.44% specificity. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

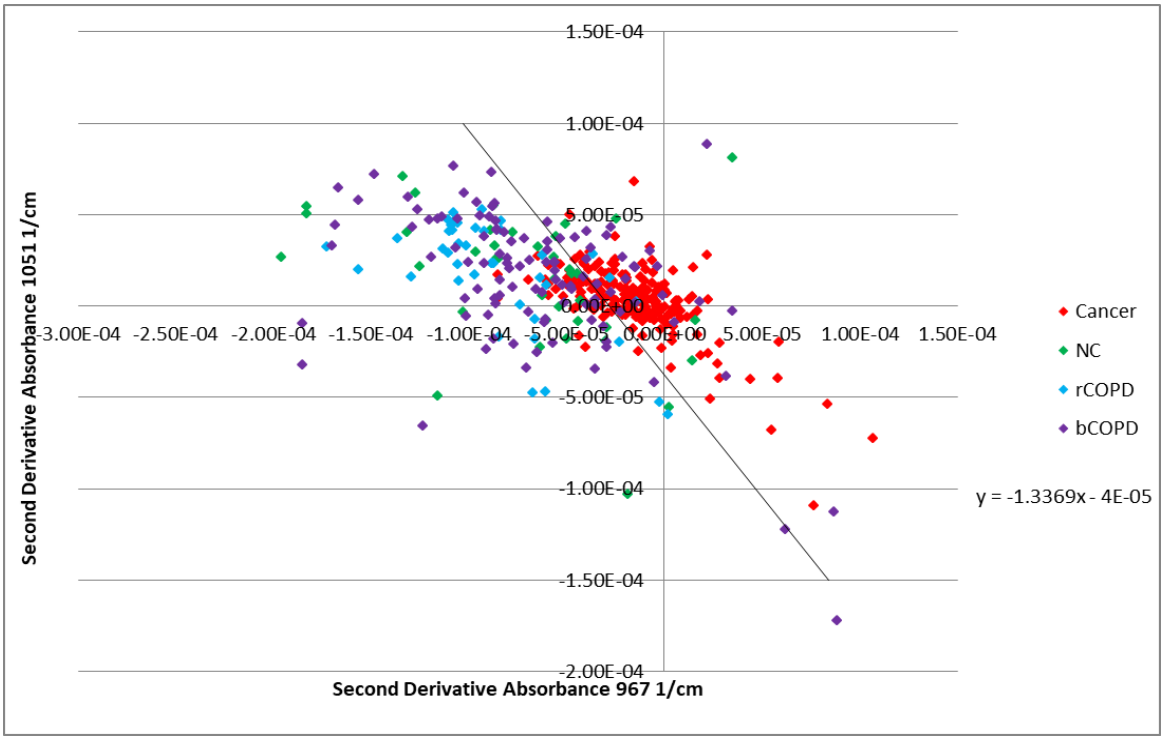


**Figure 5-12: Two dimensional scatterplot of second-derivative absorbencies at 967cm$^{-1}$ and 1051cm$^{-1}$, showing separation of cancer (red) and non-cancer (blue, purple, green) sputum sample spectra, primarily along the 967cm$^{-1}$ axis. A linear regression line for cancer prediction is shown at 81.31% sensitivity and 81.44% specificity.**

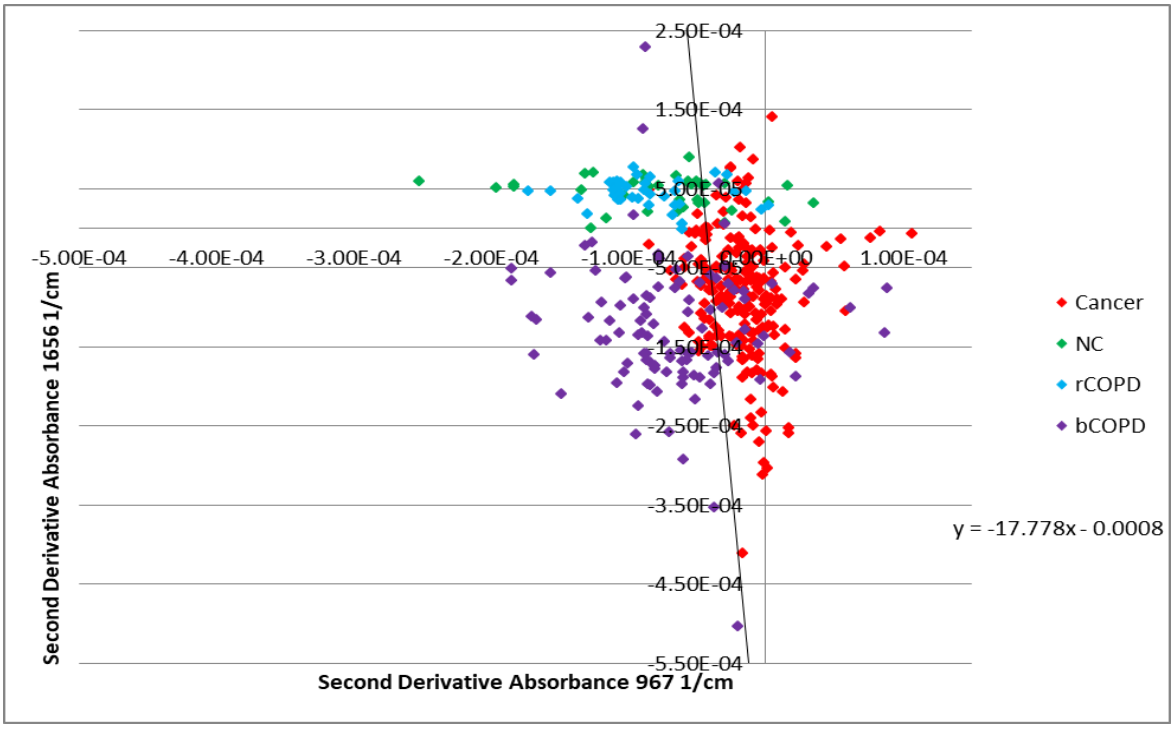


**Figure 5-13: Two dimensional scatterplot of second-derivative absorbencies at 967cm$^{-1}$ and 1656cm$^{-1}$, showing good separation of cancer (red) and non-cancer (blue, purple,**

**green) sputum sample spectra. A linear regression line for cancer prediction is shown at 83.18% sensitivity and 77.32% specificity. A distinct separation of the bCOPD (purple) and other non-cancer groups is also observed along the 1656cm$^{-1}$ axis. Reasons for this are explored in more detail subsequently. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

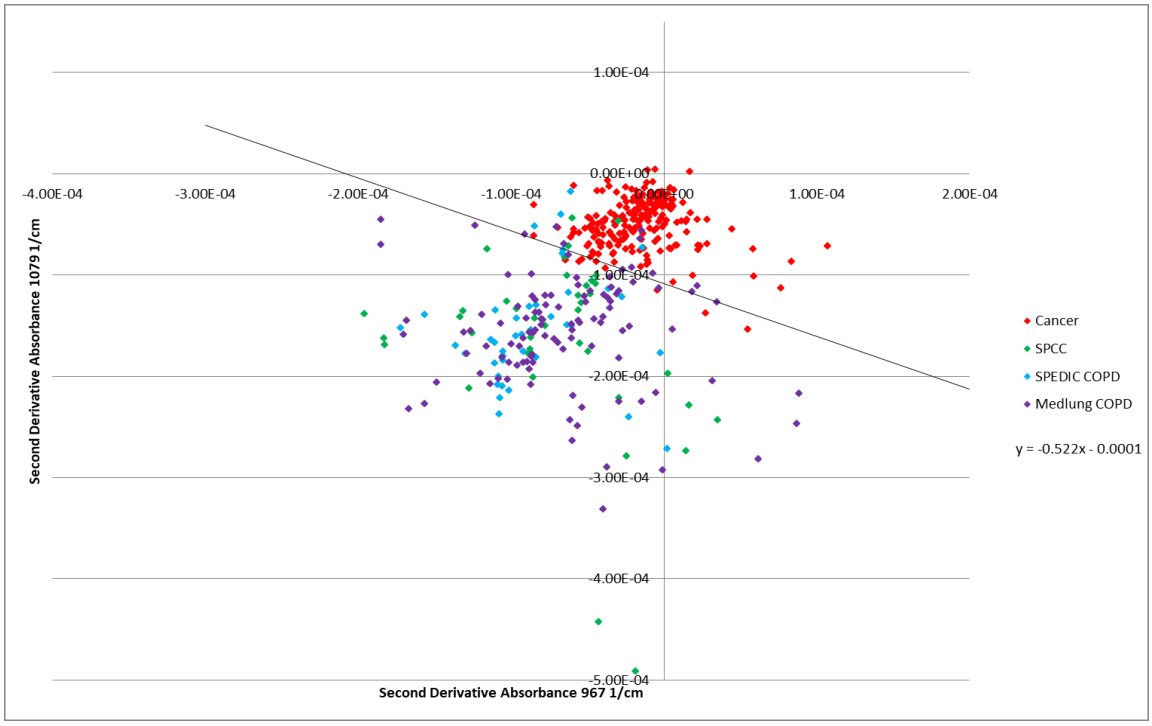


**Figure 5-14: Two dimensional scatterplot of second-derivative absorbencies at 967cm$^{-1}$ and 1079cm$^{-1}$, showing excellent separation of cancer (red) and non-cancer (blue, purple, green) sputum sample spectra. A linear regression line for cancer prediction is shown at 92.99% sensitivity and 94.33% specificity. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

| | 984 vs 967cm$^{-1}$ | 1024 vs 967cm$^{-1}$ | 1051 vs 967cm$^{-1}$ | 1079 vs 967cm$^{-1}$ | 1411 vs 967cm$^{-1}$ | 1577 vs 967cm$^{-1}$ | 1656 vs 967cm$^{-1}$ |
|---|---|---|---|---|---|---|---|
| **True Positives** | 178 | 181 | 174 | 199 | 178 | 192 | 178 |
| **False Positives** | 39 | 36 | 36 | 11 | 22 | 39 | 44 |
| **True Negatives** | 155 | 158 | 158 | 183 | 172 | 155 | 150 |
| **False Negatives** | 36 | 33 | 40 | 15 | 36 | 22 | 36 |
| **Sensitivity (%)** | 93.93 | 84.58 | 81.31 | 92.99 | 83.18 | 89.72 | 83.18 |
| **Specificity (%)** | 75.26 | 81.44 | 81.44 | 94.33 | 88.66 | 79.90 | 77.32 |

**Table 5-8: Sensitivity and specificity calculations for each linear model shown above. All models show good sensitivity and specificity, with the weakest being 1656 vs 967cm$^{-1}$ at 77.32% specific and the strongest 1079 vs 967cm$^{-1}$ at 94.33 specific.**

The linear models shown above provided an excellent proof-of-concept that lung cancer sputum could be distinguished from non-cancer based on second-

derivative absorbencies at defined wavenumbers. The wavenumbers examined thus far were those previously identified by the Respiratory Diagnostics Group at Swansea University (P. D. Lewis *et al.*, 2010), and not all of the regression models displayed >80% sensitivity and specificity. Examination of the second-derivative spectra in Figure 5-9 suggested that there could be novel wavenumbers that may also distinguish cancer from non-cancer. Therefore, further analysis was carried out and the models that are summarised in Table 5-8 were optimised to maximise sensitivity for lung cancer whilst having a minimum specificity of 80%. These new models and optimisations are summarised in Figure 5-15.

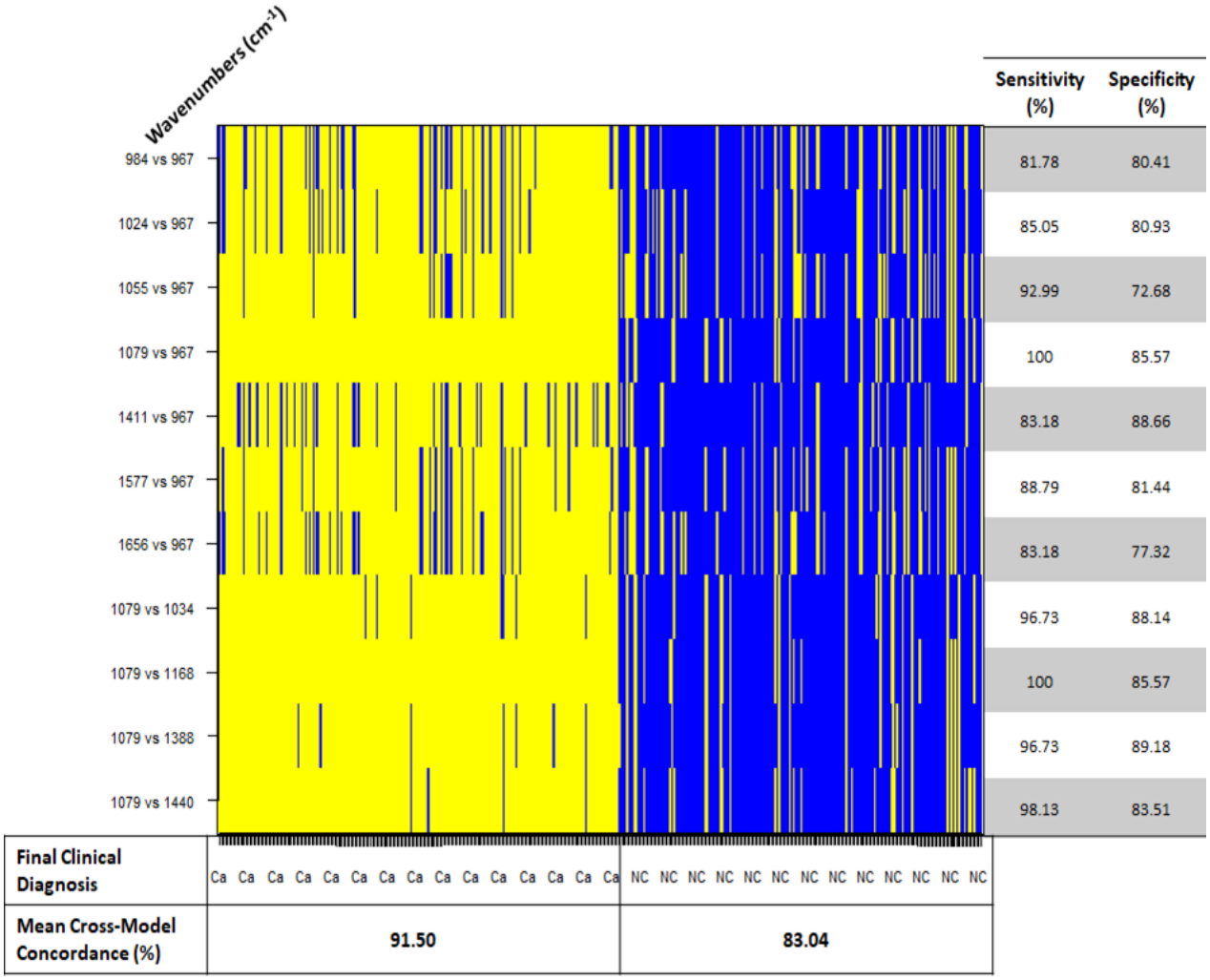


**Figure 5-15: Two-dimensional heatmap of predictions by each regression model, with sensitivities and specificities for each model shown. A prediction of cancer is shown as yellow, and a prediction of non-cancer is shown as blue. The cross-model concordance is >80%, with greater agreement between models observed when predicting a cancer diagnosis. This suggests a possible bias towards a cancer prediction output from the models.**

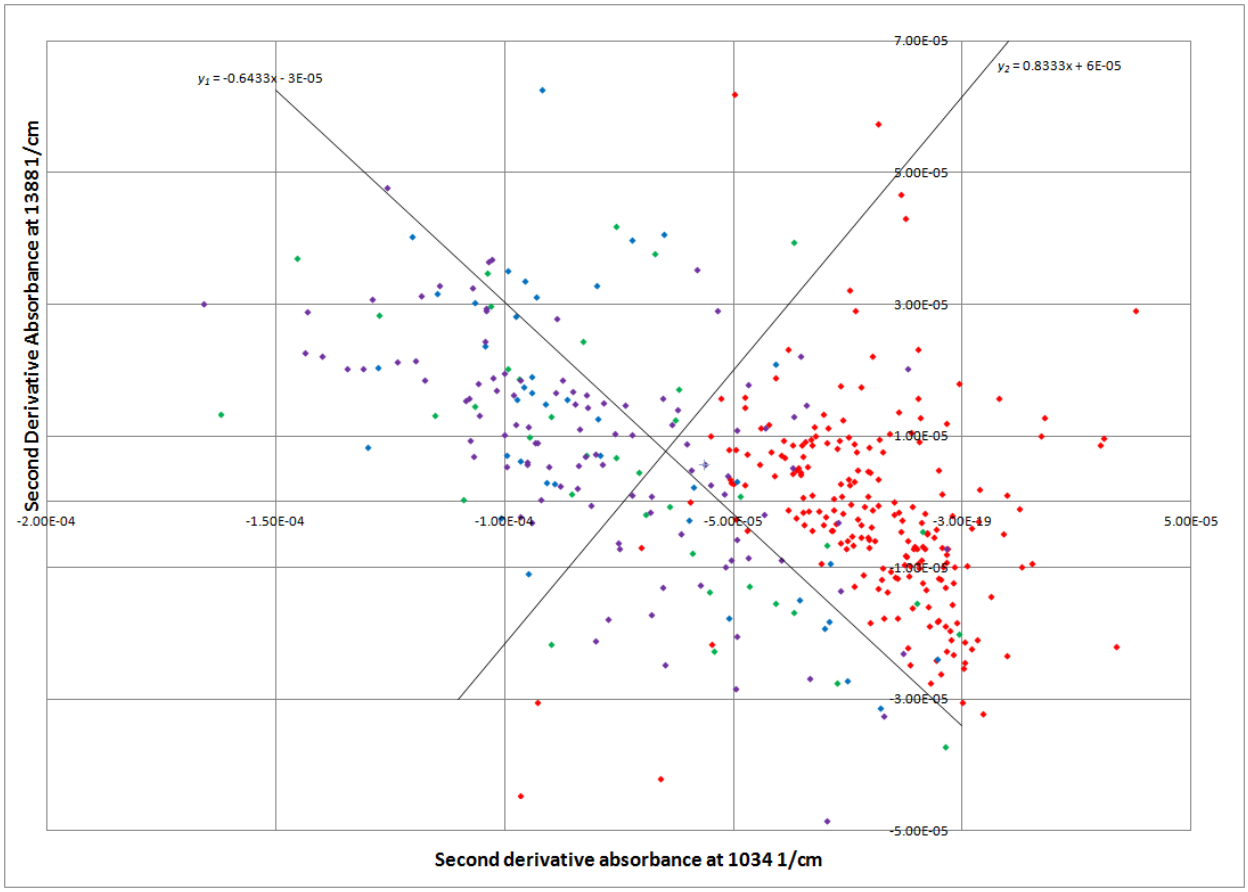


**Figure 5-16: Two-directional linear model of second-derivative absorbencies at 1034 $cm^{-1}$ and 1388$cm^{-1}$, showing good separation of cancer and non-cancer cohorts. The linear separator lines divide the space into 4 quadrants (Q1, Q2, Q3. and Q4) allowing the classification of "non-cancer" (Q1) "higher risk" (Q2 & Q3) and "cancer" (Q4) quadrants. rCOPD – random COPD; bCOPD – bronchoscopy COPD; NC – non-cancer-non-COPD**

The two-directional regression model shows a more complex model based on the appearance of peaks at 1034 & 1388$cm^{-1}$ in the second-derivative spectrum (Figure 5-16). Two regression lines are used in conjunction to improve the specificity of the model. The cancer cohort spectra show a strong grouping in Q4, whilst the spectra from the non-cancer cohorts show less strong clustering, but the majority of these are found in Q1.

The cancer samples have a tight cluster in Q4 that is surrounded by a few non-cancer cases in Q2 & Q3, whilst the main cluster of non-cancer is in Q1. There are also a few cancer cases in Q2 and Q3. These therefore could be described as "higher risk" quadrants, where the samples have absorbance patterns more similar to a typical cancer spectrum. As seen in Table 5-9, the sensitivity for each linear model is excellent, with 93.93% and 99.07% for $y_1$ and $y_2$, respectively however the specificities

for each are low, at 78.35% and 63.40%. By combining the two models the sensitivity is slightly reduced to 92.99%, but the specificity markedly improves to 89.18%.

| | $y_1$ | $y_2$ | $y_1$ & $y_2$ |
|---|---|---|---|
| **True Positives (Ca in Q4)** | 201 | 212 | *199* |
| **False Positives (NC in Q4)** | 42 | 71 | *21* |
| **True Negatives (NC in Q1, Q2 & Q3)** | 152 | 123 | *173* |
| **Overall False Negatives (Ca in Q1, Q2 & Q3)** | 13 | 2 | *15* |
| **False Negatives in "Non-cancer" Quadrant (Ca in Q1 only)** | 0 | 0 | *0* |
| **False Negatives in "Higher Risk" Quadrants (Ca in Q2 & Q3)** | 13 | 2 | *15* |
| **Sensitivity (%)** | **93.93** | **99.07** | ***92.99*** |
| **Specificity (%)** | **78.35** | **63.40** | ***89.18*** |
| **False Negatives Classified as "Higher Risk" (Q2 & Q3) (%)** | | | *100* |
| **True Negatives Classified as "Higher Risk" (Q2 & Q3) (%)** | | | *41.04* |

**Table 5-9: Sensitivity and specificity calculations of each single linear model ($y_1$, $y_2$ individually) and the two-directional model ($y_1$ & $y_2$ simultaneously). $Y_1$=-0.6433x-0.00003 $y_2$=0.8333x+0.00005; Ca - cancer; NC - non-cancer**

### 5.3.3 Squared Euclidean Distance

The models presented above generally show the cancer samples clustering closely with one another, whilst the COPD and NC cohorts are more dispersed. This suggests that cancer sputum spectra may be more similar to one-another than spectra from COPD and other respiratory disease samples. This also suggests that there may be an IR spectral signature for cancer. The geometric mean of absorbencies at 967 and 1076cm$^{-1}$ for each patient cohort was calculated and treated as the "cancer model" or "non-cancer model". These wavenumbers were chosen due to the high sensitivity and specificity with which they separate cancer from non-cancer sputum along with the strong clustering exhibited (Figure 5-14). The Squared Euclidean Distance (SED) of each sample spectrum was then calculated from this point to estimate the similarity of each spectrum to the cancer signature. SED calculation from both a cancer and non-cancer model allows the distance-value generation of the sample from each model. These can be compared against a threshold distance, with an SED value being smaller than a set cancer-threshold and larger than a set non-cancer threshold, implying that a sample should be classified as cancer. These distances were compared and the results are shown in Figure 5-17.

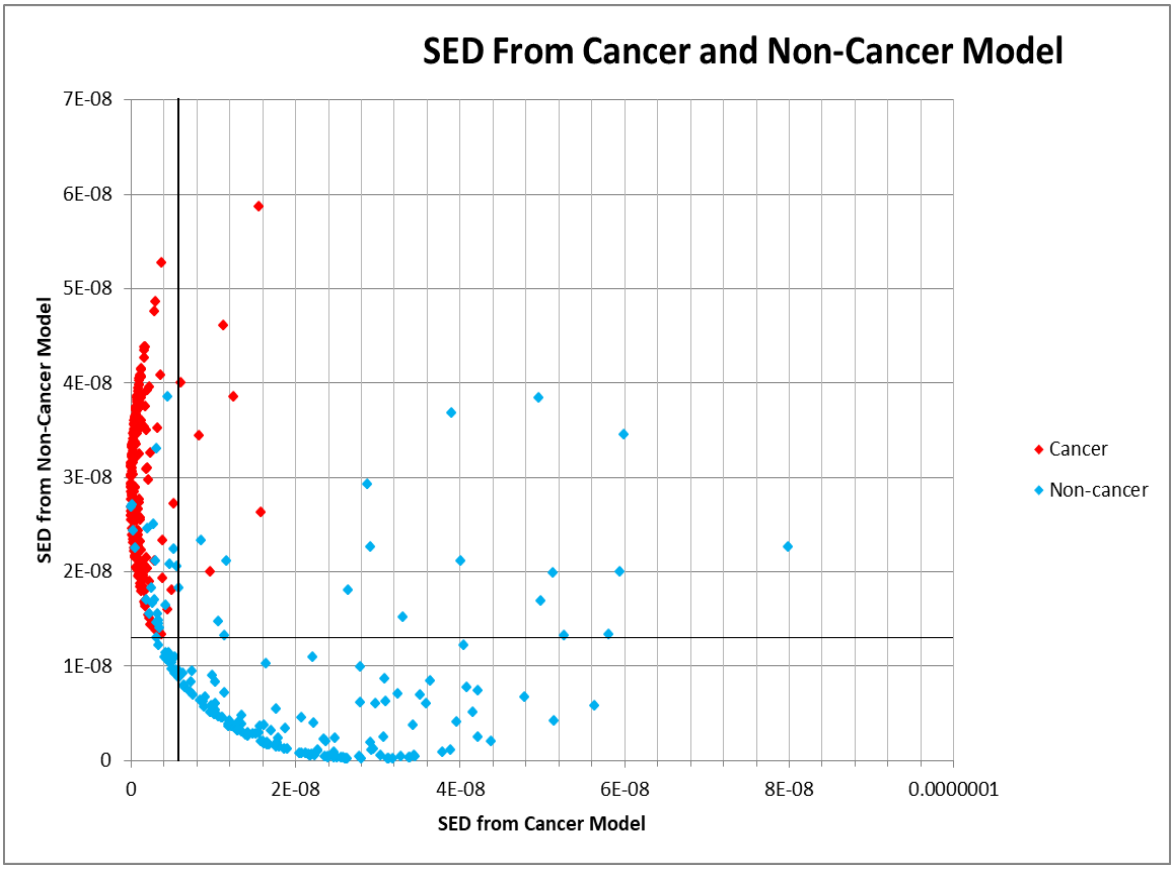


**Figure 5-17: SED values of all cancer and non-cancer spectra from the cancer and non-cancer model spectra, with threshold lines at $y=1.3*10^{-8}$ and $x=6*10^{-9}$. The thresholds were set so as to capture as many cancer cases within a quadrant as possible, whilst ensuring as few non-cancer cases as possible were also captured. A few false negative results can be observed but crucially these cases cannot be said to be closer to the non-cancer model than the cancer model. Therefore, it can be said that these cases are more similar to cancer model and thus a higher % likelihood of cancer should be returned.**

The plotted SED values show the distances of individual sample spectra from the cancer and non-cancer models. As can be seen, the cancer spectra generally have a low SED value from the cancer model with little variation within the group, whereas the SED of non-cancer spectra from the cancer model is generally much greater. There is also a greater amount of variation within the non-cancer samples, although most samples show a low SED from the non-cancer model. This higher variation is indicative of the mixed nature of this cohort of patients, being made up of COPD and non-COPD patients.

The low amount of observed variation in the SED of the confirmed cancer sample spectra from the cancer model implies that the spectra that build the cancer model are similar to each other, in that they are share similar spectral characteristics and therefore the calculated models have a good strength. Due to infrared spectra

directly relating to the molecular structure and configuration of the sample, this therefore means that the molecular structure of the cancer samples can be said to be similar to each other. This implies that there is a detectable molecular alteration to the structure of sputum from lung cancer patients.

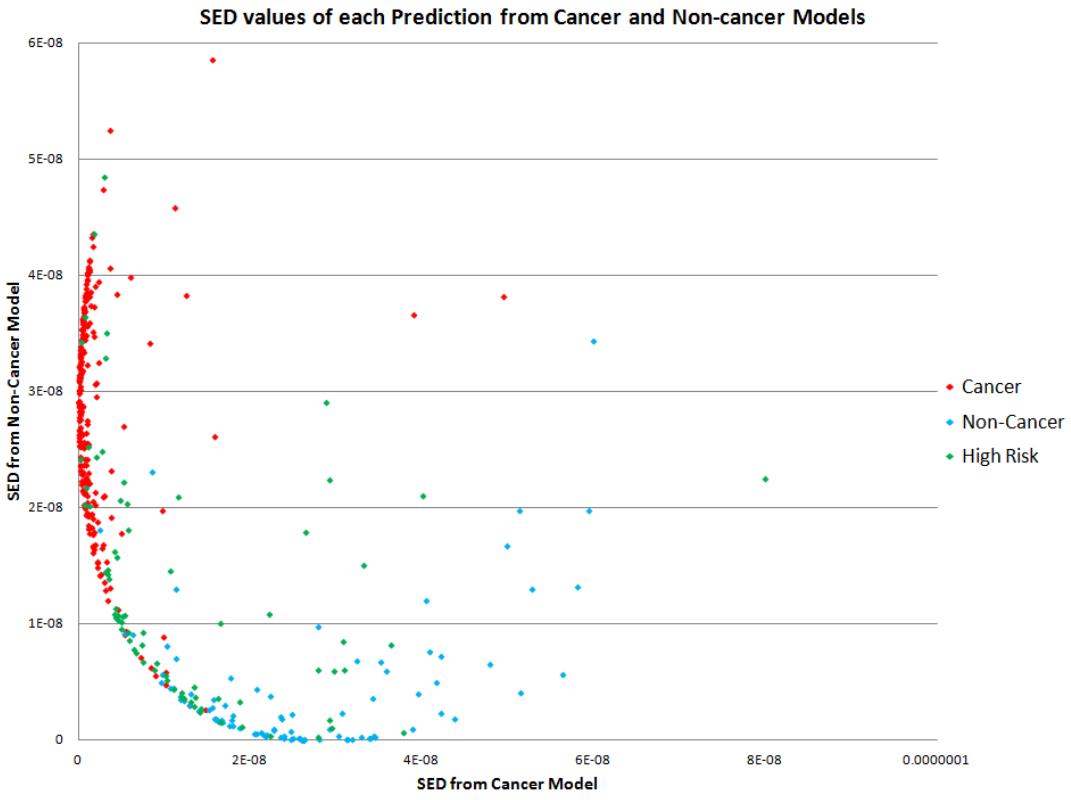


**Figure 5-18: SED values of each prediction made by the two-directional model in Figure 5-16 to the cancer and non-cancer model spectra. The high-risk prediction samples (green) are seen to group closely towards the cancer predictions (red) and the cancer model spectrum, whilst the non-cancer predictions (blue) are closest to the non-cancer model. This implies that the high-risk predictions made by the model in Figure 5-16 are not random and could therefore indicate that this could be used a method for identifying true higher-risk patients.**

There is a small overlap of groups, which occurs when the Euclidean distance of the sample spectrum from the non-cancer model increases whilst also decreasing the SED value from the cancer model. This means a sample spectrum found within this overlap is more similar to the cancer model than the non-cancer model, which directly implies the molecular structure of this particular sputum sample may be more similar to a cancer sample than a non-cancer sample. This overlap of SED values mirrors what can be seen in the linear separation model in Figure 5-14, with a small but important number of false negative and false positive results. It is possible that these cases could represent a subset of patients who have not received a diagnosis of

cancer due to no cancer being detectable at bronchoscopy, but who may be at risk of developing cancer in the short-term. This hypothesis is supported by the generally lower SED values of the predicted higher-risk patients towards the cancer model spectrum, than those observed in the predicted non-cancer patients Figure 5-18.

### 5.3.4 Distinguishing Lung Cancer Sub-types and Histology by FTIR

A secondary aim of this chapter was to evaluate the ability of FTIR spectroscopic analysis to determine lung cancer subtypes and histologies. A smaller subset of the lung cancer sample set was statistically interrogated for significant trends which may show spectral differences between adenocarcinoma, squamous cell carcinoma and SCLC. Twenty spectra for each histological group were analysed. Initial examination of the average second-derivative spectra shows a high degree of similarity between the histological groups.

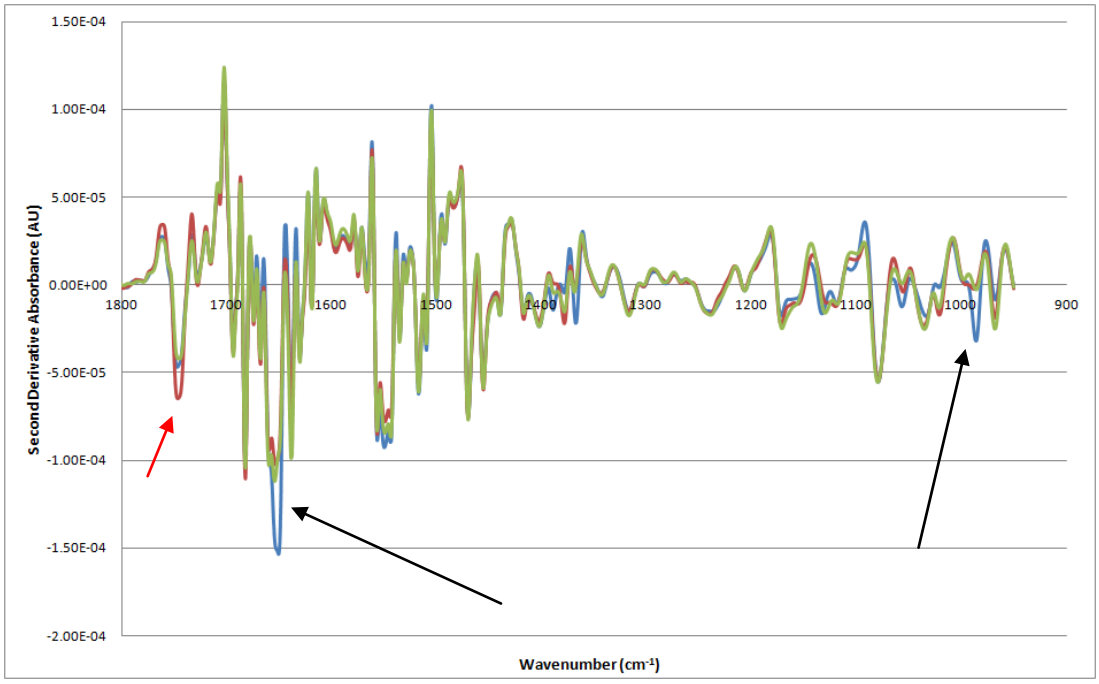


**Figure 5-19: Average second-derivative spectra of adenocarcinoma (blue), squamous cell carcinoma (red) and SCLC (green) lung cancer patient groups. The spectra show high similarity with few differences, although a promising difference in absorbance intensity can be seen at a few wavenumbers; namely a small increase in intensity at 1747cm$^{-1}$ in squamous cell carcinoma (red arrow), and an increase in intensity at 1649 and 986cm$^{-1}$ in adenocarcinoma (black arrows).**

Statistical tests were performed to assess the significance of the differences in second-derivative absorbance at the wavenumbers identified in Figure 5-19. ANOVA testing suggested that the variation between the histology groups was statistically significant at the 95% confidence level at only 986cm$^{-1}$ (Table 5-10). Subsequent T-

Test analysis with multiple hypothesis correction (α = 0.01667) found that the differences between the histology groups were not statistically significant (Table 5-11).

| | Wavenumber ($cm^{-1}$) | | |
|---|---|---|---|
| | 1747.38068 | 1649.018 | 986.5179 |
| **Between Groups Variation (p-value)** | 0.462188797 | 0.052109 | 0.035623 |

**Table 5-10: Single factor ANOVA test results for second-derivative absorbance at wavenumbers of interest for distinguishing histological groups, with statistically significant (p <0.05) variation between groups being shown at 986$cm^{-1}$, suggesting that second-derivative absorbance at 986$cm^{-1}$ may be key for determining histology.**

| | Squamous | SCLC |
|---|---|---|
| **Adenocarcinoma** | 0.0657 | 0.0706 |
| **Squamous** | | 0.9267 |

**Table 5-11: Two-sided p-value scores from significance testing of absorbance at 986$cm^{-1}$ between histological groups; multiple hypothesis testing set the α = 0.01667. Statistical significance was not found. SCLC – Small cell lung cancer**

Statistical analysis of the second-derivative peaks found within the glycogen-rich region also did not identify any significant peak shifts between the histological groups (Table 5-12).

| | Wavenumber ($cm^{-1}$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1130** | **1079** | **1055** | **1030** | **1020** | **998** | **985** | **967** |
| **Adeno-Sq** | 0.197651 | 0.938103 | 0.937318 | 0.011446 | 0.074654 | 0.013765 | 0.303451 | 0.855445 |
| **Adeno-SC** | 0.505636 | 0.866535 | 0.924569 | 0.440799 | 0.08236 | 0.189174 | 0.471057 | 0.802816 |
| **Sq-SC** | 0.386532 | 0.806994 | 0.985283 | 0.076731 | 0.795943 | 0.143944 | 0.667538 | 0.916255 |

**Table 5-12: Two-sided p-value scores from T-tests for significance of differences between detected peak positions in the second-derivative spectra. No statistical significance was found (p > 0.05). Adeno – adenocarcinoma; Sq – Squamous cell carcinoma; SC – Small cell lung cancer**

## 5.4 Discussion

The aim of this study was to assess the ability of FTIR to detect and diagnose lung cancer using raw sputum with no sample pre-processing and a simple analytical protocol. Currently there is a strong clinical need for an early detection method. We have demonstrated that high-throughput FTIR technology can be readily utilised to sensitively and specifically distinguish lung cancer from non-cancer respiratory disease. The linear separation models developed above show a strong degree of clustering, with a clear separation of groups and small overlap. The mean

absorbencies of cancer spectra at 1740, 1653, 1589, 1411, and 1076cm$^{-1}$ were shown to be statistically significantly different from all other non-cancer spectra. However, linear models of these wavenumber could only distinguish cancer from control at a maximum sensitivity and specificity of 77.78% and 81.96% respectively when plotting the absorbencies at 1740 and 1589cm$^{-1}$.

The initial results were encouraging but the sensitivity and specificity scores were not high enough for a potential clinical application. Deconvolution of the absorbance spectra by calculation of the 2$^{nd}$ derivative spectra was used to increase spectral resolution and subsequently linear regression models were developed from these 2$^{nd}$ derivative spectra. The work presented in this chapter has developed 11 distinct regression models based on second-derivative spectra which are all capable of distinguishing cancer from non-cancer with sensitivities and specificities greater than 80% and 70% respectively. Within these 11 models, there were five models which showed sensitivities in excess of 95%, whilst retaining >80% specificity for lung cancer.

Additionally, the models developed in this chapter produce highly concordant results, with an average cross-model concordance level of 87.48%. This cross-model agreement was shown to be higher for the cancer patient cohort, which reflects the sensitivity-weighting of the regression models.

In each model a small but significant number of false positives and false negatives were identified, due to a small overlap of clusters. This overlap of clusters suggests that FTIR may be able to identify patients who may have a higher risk of developing lung cancer in the near future, compared to other patients who attend bronchoscopy clinic. This could be utilised in a screening capacity, where patients who attend respiratory clinics could have their sputum routinely analysed and therefore avoid the need to undergo bronchoscopy. This is further supported by the high cross-model concordance indicating that these models could be combined into a battery of tests which could then be used to provide a "% likelihood" of cancer, with more models reporting a diagnosis of "cancer" contributing to a higher "% likelihood".

Many models show a close clustering of cancer samples, with few outliers, compared to the NC and COPD cohorts which show a more dispersed grouping. This

higher dispersion is indicative of the variability of the patients within these cohorts. The NC cohort shows the most variability with more diverse diagnoses, reflected in the clusters, thus suggesting a greater amount of variation between spectra within this cohort. In contrast, the cancer cohort tends to group more closely, implying that there is less variation between spectra within this cohort. Indeed, analysis of variance testing of the absorbencies used in these tests indicated a greater degree of variation within the NC cohort than in rCOPD or bCOPD, and all control cohorts showed more variation than the cancer cohort. This suggests that there may be an IR spectral signature for lung cancer, to which a patient's similarity through SED could be calculated. In this work the similarity of each non-cancer and cancer sample spectrum was calculated to the cancer signature for the 967 vs 1079cm$^{-1}$. The band around 967cm$^{-1}$ has been associated with phosphorylated proteins, nucleic acids and Le$^{X}$ antigens in sputum (A. T. Lewis *et al.*, 2013; Malins *et al.*, 2005; Maziak *et al.*, 2007).

Further work needs to be carried out to clarify if the models developed in this study could be used to identify lung cancer cases earlier than current diagnostic methods. As shown in this chapter, a number of potential diagnostic models for lung cancer have been developed, but the number of false positives identified in many of these models is not trivial. A raw sputum sample was predicted as being from a cancer patient if its spectral features were similar to that of many other cancer samples. It could be possible that a patient's sputum sample may give a spectral signature more similar to that of cancer before any histological affects could be detected by bronchoscopy. In this way, FTIR could be utilised as a potential early-warning system. In order to confirm this, a new longitudinal follow-up study must be carried out in which patients are monitored over an extended period of time (>5 years). Such a study would compare the similarity of their sputum spectral signatures to the cancer signature and then follow up with respect to any future cancer diagnosis.

The two-directional linear model (Figure 5-16) highlights an interesting possibility of using such a model to estimate a patient's likelihood of developing cancer within a short period of time, if a diagnosis of cancer is not possible through X-ray, CT scan or bronchoscopy. It could be possible that those patients who received a diagnosis of non-cancer, but whose spectra fall into Q2, Q3 or Q4 regions in

Cartesian space may have currently undetectable pre-malignancies, or be at an early-stage of cancer. A major challenge which reduces the effectiveness of lung cancer treatment is the difficulty involved in diagnosing early-stage lung cancer, when treatment outcomes are usually better: through early detection of lung cancer, overall survival rates can be improved. (Allemani *et al.*, 2015; Morgan & Wilkes, 2017). With respect to the two-directional linear model (Figure 5-16), an effective method for detecting pre-malignant or early-stage lung cancer cases may be possible. When a patient's sample falls within the "higher risk" (Q2, Q3) or "cancer" (Q4) quadrant the Euclidean distance of the sample from the geometric centre of the cancer model spectrum developed in this work could be calculated. Through a future study based on following up these patients, it may be possible to provide a quantifiable measure of potential risk of lung cancer that each patient may have.

A review of lung cancer diagnostic methods highlighted how sputum cytology screening programs have a higher chance of detecting early-stage lung cancer, and how sputum cytology was able to precede a radiological diagnosis by 18-36 months (Thunnissen, 2003). This suggests that premalignant/early-stage carcinomas may be occurring many months prior to diagnosis by radiological methods. Sputum cytology relies on observations of the morphology of cells found in the sputum. The work presented in this chapter shows that FTIR is capable of distinguishing between lung cancer and respiratory disease patient sputum samples based on spectral changes which can be related to molecular structural change. Given the observation that a sputum cytological diagnosis can precede a radiological diagnosis by many months, it is reasonable to hypothesise that molecular structural change indicative of lung cancer may be occurring within the lung tissue for an extended period of time prior to the appearance of symptoms or a diagnosis. FTIR may be capable of detecting structural changes to morphology of cells within sputum. However a large study comparing early-stage lung cancer detection rates in high-risk populations of multiple diagnostic modalities, such as sputum cytology, x-ray, bronchoscopy and CT scanning, to FTIR spectroscopy must be carried out in order to fully establish the usefulness of FTIR as a screening tool for early-stage lung cancer.

An analysis carried out on a limited subset of 20 adenocarcinoma, squamous-cell-carcinoma and SCLC lung cancer sputum samples did not find any statistically significant differences in second-derivative absorbance or peak position between the

histological groups of lung cancer sputum spectra. This is not surprising given that the previous analysis in this chapter showed extremely close clustering of the cancer spectra when analysing them using wavenumbers within the glycogen-rich region. These findings suggest that a larger study is necessary to truly determine whether molecular structural change specific to lung cancer histologies may be detectable by FTIR spectroscopy. However, these findings do support the universality of the major finding of this chapter: that FTIR spectroscopy is capable of determining lung cancer patient sputum from chronic respiratory disease patient sputum, and the histology of the cancer does not affect the ability to detect the cancer.

## 5.5 Conclusion

FTIR analysis has been shown to have potential as an effective tool for the detection of lung cancer in raw sputum, using a rapid protocol with no sample preparation and simple spectral analysis. The work presented in this chapter showed a panel of two-dimensional linear models capable of differentiating lung cancer sputum from COPD sputum with sensitivity and specificity scores of >80% each, and the best performing models showing 100% sensitivity with 85.6% specificity.

The possibility of using FTIR as a screening tool for early cancer detection in high-risk individuals is raised. However, the work presented in this thesis does not yet allow the absolute determination of a “cancer” or “non-cancer” diagnosis in a patient. Whilst the high sensitivity and specificity scores are highly encouraging, the study needs to be validated on a larger set of patients, over a longer period of time. A future study in which patients attending the diagnostic pathway for lung cancer are recruited and monitored long-term is needed to fully assess the diagnostic power of FTIR for early-stage lung carcinomas. It could be possible that molecular structural changes are detectable in sputum by FTIR spectroscopy some months prior to detection by X-ray, CT-scan or bronchoscopy.

# Chapter 6

# Utility of FTIR for Monitoring Retention and Interaction of a Novel Therapeutic with Sputum Mucins in Cystic Fibrosis Patient Sputum

## 6.1 Introduction

### 6.1.1 Cystic Fibrosis Background

Cystic fibrosis (CF) is a genetic disorder which primarily affects the airways, through airway-restriction by increased mucus load (Rose & Voynow, 2006). There are approximately 85,000 patients with CF worldwide and this number is projected to increase by up to 75% by 2025 (Burgel *et al.*, 2015; Kris De Boeck & Amaral, 2016). This is mainly due to an increasing life-expectancy for CF-patients, from 27 in 1986 to 38.3 in 2010 (Cystic Fibrosis Foundation, 2010). Despite this increase in life-expectancy, the leading cause of death for CF-patients is still respiratory failure (Elborn, 2016).

The increased mucus load restricts the airflow, reducing lung capacity and has a significant impact on the quality-of-life (QOL) for the patient, with patients generally reporting a decrease in QOL over time (Uchmanowicz *et al.*, 2015). Patients must undergo constant physical therapy and monitoring in order to appropriately manage their disease and improve their QOL (Button *et al.*, 2016; Feiten *et al.*, 2016).

CF is caused by mutations or deficiency in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which encodes for the CFTR transmembrane chloride and bicarbonate ion channel (Gustafsson *et al.*, 2012). Over 1900 mutations of CFTR have been reported, with the most common CF-causing mutuation being Phe108del (K. De Boeck, Zolin, Cuppens, Olesen, & Viviani, 2014b). Defects in this protein lead to dysregulation of epithelial fluid transport, decreasing mucus hydration and increasing the viscosity of CF airway mucus. As a direct consequence of this, mucociliary-clearance (MCC) is impaired in CF-patients (Lee & Foskett, 2014). Dysregulation of the CFTR channels can also lead to an elevation in salt levels in CF airway mucus, which has been shown to inhibit the activity of antimicrobial peptides (Zabner *et al.*, 1998). Due to the combination of reduced MCC activity and inhibition of antimicrobial peptides, CF patients are more susceptible to airway infection.

As a consequence of these inhibited activities, most adult CF patients are persistently colonized by biofilm-forming *Pseudomonas aeruginosa* strains (Wagner & Iglewski, 2008). A bacterial biofilm is formed when bacteria adhere to solid surfaces and secrete extracellular polysaccharides to form a matrix to which more bacteria cells can adhere. Biofilms also contribute to resistance to antimicrobial agents,

including antibiotics, further increasing the susceptibility of CF patients to persistent infection (Donlan, 2001).

### 6.1.2 Mucus and mucins

Mucus is the primary component of the airway surface liquid (ASL) barrier in the airways. The main functions of mucus are to protect the airways from pathogens & foreign particles, and to maintain hydration of the airways. The vast majority of respiratory-mucus is primarily composed of water – approximately 95% of mucus is water (Bansil *et al.*, 1995). Of the remaining 5%, mucin glycoproteins are the dominant molecules, namely the secretory mucins MUC5AC and MUC5B, although other mucins are also present (Rose & Voynow, 2006). These mucins are high-weight glycoproteins and are extremely heavily glycosylated along the length of the polypeptide chain. Mucins are responsible for the biological, chemical and physical properties of mucus, and provide for mucus the structural matrix through a strong mesh-like structure formed by a complex web of ionic, hydrophilic, hydrophobic and covalent bonds (Georgiades *et al.*, 2014). Therefore it is easy to see how alterations in the constituent parts of mucus, especially in mucin glycoprotein expression/secretion can have a significant impact on the properties of the mucus itself.

Evidence suggests that the CTFR channel is able to act as a bicarbonate channel, leading to suboptimal levels of bicarbonate being present in the CF lung (Borowitz, 2015). Bicarbonate is used to neutralise the pH, and remove the $Ca^{2+}$ to unpack the secreted mucin granules and form the mucus mesh. The reduced levels of bicarbonate in the CF lung prevents full mucin expansion, leading to denser more viscous airway mucus (Ambort, Johansson, Gustafsson, Ermund, & Hansson, 2012; Gustafsson *et al.*, 2012).

### 6.1.3 OligoG

OligoG is a low weight alginate oligosaccharide extracted and purified from the stem of brown seaweed (*Laminaria hyperborean*) and composed of α-L-guluronate subunits (Figure 6-1).

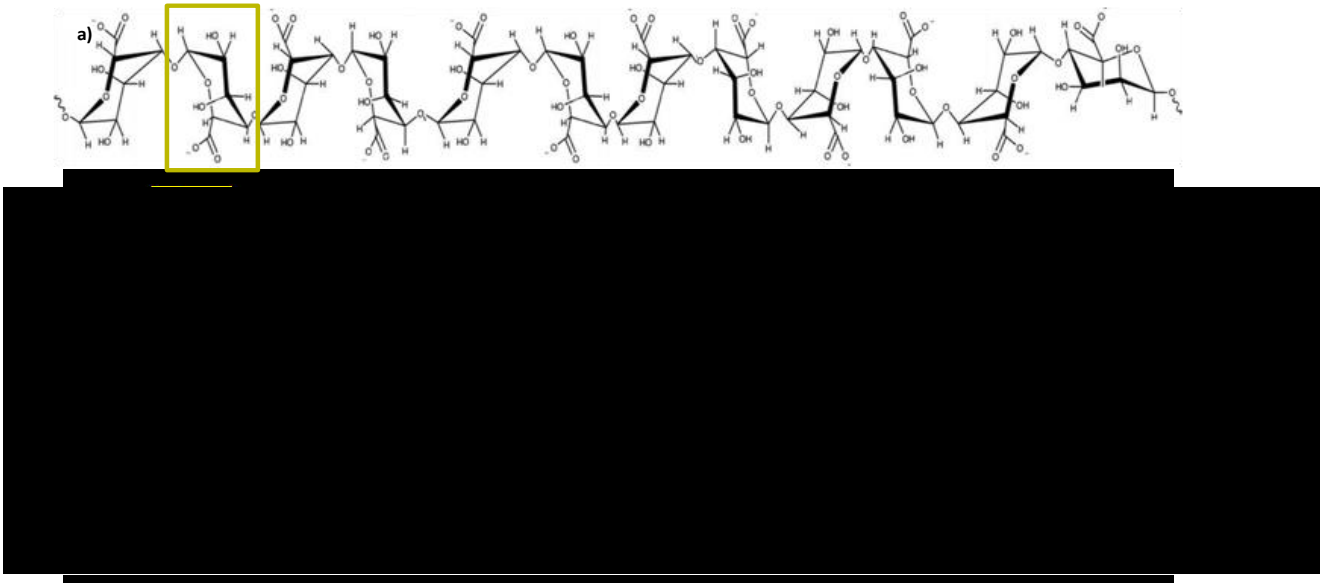


**Figure 6-1: (a) Fischer projections, (b) stick, and (c) space filling visualisations of 12-(a), and 16- (b & c) residue OligoG molecule. One of α-L-guluronate residue has been highlighted by the yellow box. CPK colouring has been used. Fischer projection adapted from Hengzhuang *et al.*, 2016.**

In vitro experimentation with multidrug resistant bacteria (including *Pseudomonas aeruginosa, Staphylococcus aureus and Streptococcus oralis*) has shown that OligoG possesses an antibacterial effect on such cells with it commonly being observed to visibly distort the outer LPS membrane (Khan *et al.*, 2012). It is currently thought that this is due to the interaction of OligoG with calcium ions found within the membrane (Pritchard *et al.*, 2017). This cell surface interaction has shown to be able to resist removal via hydrodynamic shear (Powell *et al.*, 2014). In addition to this when used in conjunction with antibiotics, such as triclosan and colistin, OligoG was seen to have a potentiating effect increasing the effectiveness of the drug several-fold (Hengzhuang *et al.*, 2016; Khan *et al.*, 2012; Roberts *et al.*, 2013).

Antifungal properties of OligoG have also been investigated against strains of pathogenic fungus (e.g. *Candida tropicalis* and *Aspergillus spp*), where at low concentrations (>0.5%) OligoG was shown to significantly inhibit hyphal growth, and at high concentrations (>6%) showed a significant inhibition of cell growth. It was also shown to have a potentiating effect on various antifungal drugs, with as high as 16-fold reduction in minimum inhibitory concentration being observed when used in conjunction with nystatin (Tøndervik *et al.*, 2014).

OligoG has also been found to have a dose dependant antibiofilm effect with biofilms being visibly weakened in its presence described as appearing more open and

porous, and containing less biomass. Biofilms treated with OligoG are also found to have their mechanical properties compromised, being removed with less force and also being less resistant to deformation under force than untreated biofilms (Khan *et al.*, 2012; Powell *et al.*, 2014; Roberts *et al.*, 2013; Tøndervik *et al.*, 2014). This antibiofilm effect has also been observed in murine lungs (Hengzhuang *et al.*, 2016), suggesting that the antibiofilm effect is not limited to bacterial and fungal strains found in human lungs. This effect is seen with regards to both bacterial and fungal biofilms. These antibacterial, antifungal and antibiofilm properties have identified OligoG as a potentially useful drug in the treatment of respiratory infections.

Another property of OligoG that further promotes its suitability as a treatment for CF is its ability to improve the state of the mucous system within the CF lung. As previously stated, individuals with CF have abnormally thick mucus that is difficult to clear from the lung, providing both an ideal environment for bacterial biofilms to become established, and posing as a barrier for drug delivery to the full lung. Evidence for the ability of OligoG to modify the viscoelastic properties of mucus comes from atomic force microscopy (AFM) imaging and rheological studies. Such analyses have shown that OligoG-treated CF sputum has a larger surface area and pore size, and a marked decrease in elasticity and viscosity (Pritchard *et al.*, 2016). In CF mice orally treated with OligoG this modification have been shown to improve intestinal transit, which is usually hindered by the viscous CF mucus which builds up in the intestinal system (Vitko *et al.*, 2016).

The method of action for this modulation of mucus viscosity appears to be its calcium chelating capabilities. As previously stated, secretion of bicarbonate is impaired in CF. In healthy individuals this bicarbonate would compete with the N-terminus of MUC2 for calcium ions. The removal of these calcium ions from the N-terminus then allows the mucin network to properly expand (Ambort *et al.*, 2012). OligoG appears to be able to act as a substitutional competitor for the calcium ions at these N-termini in the absence of bicarbonate, with a high affinity for $Ca^{2+}$ (Ermund *et al.*, 2017). Therefore, this would allow OligoG to alleviate the symptoms of CF associated with increased mucus viscosity.

### 6.1.4 Importance of continued drug presence/persistence in lung

It has repeatedly been shown that the effects of OligoG as a mucolytic, antifungal and antibacterial therapeutic are highly dependent on the dose with which

it is present (Khan *et al.*, 2012; Pritchard *et al.*, 2016; Tøndervik *et al.*, 2014). Therefore, the ability to quantify the amount of drug retained in the lung is necessary for ensuring the patients are receiving a therapeutic level of the drug. Animal studies have shown that ingested OligoG is largely excreted from the gastrointestinal tract within 24 hours, and intravenously administered OligoG is also rapidly removed in urine (Pritchard *et al.*, 2016). Measuring the volume of drug removed from the lung, or the volume still present within, however is markedly harder due to the nature of the organ.

Preclinical trials have shown that inhalation of OligoG into the lung can safely be undertaken daily without adverse effect. Dry powder inhalation was also seen to be a better inhalation route than nebulisation for whole lung deposition of the drug (Pritchard *et al.*, 2016). Currently relatively little is known about the elimination of OligoG from the lung, although most likely it is removed via a combination of the lungs regular mechanism of drug elimination such as MCC, alveolar clearance, and/or macrophage uptake (Ibrahim & Garcia-Contreras, 2013).

There is a need for a simple and non-invasive technique to monitor the presence of OligoG within the lung. Ideally, this technique would be performed multiple times over an extended period of time and would allow a clinician to optimise the dosing strategy to keep a patient within the effective dose range for as long as possible.

### 6.1.5 FTIR for Monitoring Disease States and Drug

Fourier-transform infrared (FTIR) spectroscopy is a fast, cheap-to-use and effective vibrational spectroscopy technique used to detect and monitor the molecular structure of a specific sample.

Interest in using FTIR as a clinical tool for patient-sample testing has been growing steadily for a number of years and FTIR continues to be shown as a viable tool for biofluid and patient sample analysis (Backhaus *et al.*, 2010; Baker *et al.*, 2014; Hands *et al.*, 2016; Paul D. Lewis *et al.*, 2010; Menzies *et al.*, 2014; Rodrigues *et al.*, 2017). The non-invasive nature of FTIR makes it an ideal candidate technology for monitoring drug levels in patient sputum samples without causing discomfort or inconvenience to the patient. Through analysing the spectra of patient samples, it is possible to identify specific patterns within the spectra that relate to specific molecular changes within the sample itself. By taking a reference sample spectrum of

OligoG, it is possible to screen patient samples for unique features that relate to the reference OligoG spectrum. In this way FTIR can be utilised to screen CF-patient samples for the presence of OligoG after treatment.

### 6.1.6 Chapter Aims

The aims of this chapter were to evaluate the use of FTIR analysis for successful use as a tool for routine screening of the alginate drug, OligoG, in CF sputum. Specifically, this study aims to use OligoG specific peaks, in order to show that the drug is present in patients' sputum, and provide insight into quantifying OligoG concentration through the generation of OligoG in sputum dosing curves.

Additionally, FTIR spectroscopy and FTIR-microspectroscopy will be utilised to interrogate CF-patient sputum in finer detail. Two dimensional maps of sputum will be generated to identify areas rich in mucin and OligoG, providing evidence for OligoG-mucin interactions. OligoG interactions with DNA and $Ca^{2+}$ will also be investigated by FTIR.

## 6.2 Methods

### 6.2.1 Patient Recruitment and Sample Collection

CF patients were recruited and sputum samples collected for FTIR analysis as described in Chapter 2, Section 2.1.3.

### 6.2.2 Spectrum Acquisition

A Bruker Vertex 70 with HTS-XT attachment with a DTGS detector (Bruker Optics), controlled by OPUS version 7.5, was used in transmission mode to generate transmission IR spectra of CF patient sputum samples from each stage in the study. Each spectrum was composed of an average of 32 scans and a background of 32 scans taken before each sample spectrum. Samples were pipetted (2µl) onto silicon 96-well plates (Bruker Optics) and allowed to dry in air for a minimum of one hour before measurements were taken. The plates were cleaned in 70% ethanol.

ATR-FTIR spectra were acquired using a Bruker Alpha (Bruker Optics) with diamond ATR module attachment, controlled by OPUS version 7.5. Each spectrum was composed of an average of 24 scans, and a background scan of 24 scans taken before each sample spectrum. The ATR sampling module was cleaned in 70% ethanol prior to each background scan. Samples were pipetted onto the sampling crystal (2µl) and dried in air before spectrum acquisition. Sample dryness was monitored using the

onboard live-scanning facility in OPUS. Sputum drying was expedited using a small USB-powered fan to provide airflow across the surface of the sample.

### 6.2.3 FTIR Microspectroscopy

FTIR microspectroscopy was performed using a Bruker Vertex 70 with a Hyperion 2000 IR microscope attachment, in reflectance mode. The spectrometer was fitted with a KBr beamsplitter and a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. The knife-edge aperture was set to 50x50μm. A single background spectrum of a gold standard plate was taken prior to spectrum acquisition. Sputum was pipetted (2μl) onto silicon plate and allowed to dry in air. 900 spectra were acquired over an area of approximately 1000x1000μm within the sputum spot.

### 6.2.4 Spectrum Processing

Prior to any analysis or processing, quality analysis was performed on all spectra on a case-by-case basis, where any spectra showing excessive noise or an arching baseline were rejected. Spectra were processed using the in-built tools and algorithms in OPUS 7.5 (Bruker). Spectra were vector-normalised and then baseline-corrected using the automatic rubberband correction. Second derivative spectra were calculated using the Savitizky-Golay method with 9 smoothing-points. The OPUS automatic peak picking algorithm was set to a threshold of 10%.

### 6.2.5 Statistical Analysis

Statistical tests were carried out using the programming environment R (R Core Team, 2016). Statistical significance was calculated using the Mann-Whitney Rank Sum Test, with significance shown with p-values of $< 0.05$. Correlation analysis was performed using the Spearman's Rho, with a two-tailed hypothesis.

## 6.3 OligoG Incubation

Incubation of OligoG with DNA and $Ca^{2+}$ ions was carried out as described in chapter 2, section 2.7.

| Stock conc. OligoG (% w/v) | Dilution factor (OligoG:Sputum) | Final OligoG conc. in Sputum (% w/v) |
|---|---|---|
| 0 | 1:10 | 0 |
| 0.2 | 1:10 | 0.02 |
| 0.5 | 1:10 | 0.05 |
| 1 | 1:10 | 0.1 |
| 2 | 1:10 | 0.2 |
| 5 | 1:10 | 0.5 |
| 10 | 1:10 | 1 |
| 15 | 1:10 | 1.5 |
| 20 | 1:10 | 2 |

**Table 6-1: OligoG concentrations incubated with CF-patient sputum at a 1:10 ratio for the final OligoG concentrations shown.**

CF-patient sputum screening (day 0) samples were incubated with OligoG of varying concentrations from 0% w/v to 20% w/v (Table 6-1) at a 1:10 (OligoG:sputum) ratio for 30 minutes at 37°C, prior to FTIR analysis.

OligoG solutions at a concentration of 2% (w/v) were dosed with 0mM (control), 1mM and 5mM $Ca^{2+}$ ions (Cole-Parmer) dissolved in $dH_2O$. Each mixture of OligoG and $Ca^{2+}$ ions were then dosed with 0mM (control), 1mM and 10mM whole human DNA (Promega), dissolved in $dH_2O$. Each final mixture of OligoG, $Ca^{2+}$ ions and DNA were then incubated for 30 minutes at 37°C in a water bath. Table 6-2 below shows all final dilutions of OligoG, DNA, and $Ca^{2+}$ ions made up prior to incubation and subsequent high-throughput T-FTIR analysis, as described earlier in section 2.2.2.

| Variable | Control | Experimental Solutions | |
|---|---|---|---|
| **DNA** | 2%OligoG + 0mM DNA + 0mM $Ca^{2+}$ (Control) | 2%OligoG + 1mM DNA + 0mM $Ca^{2+}$ | 2%OligoG + 10mM DNA + 0mM $Ca^{2+}$ |
| **$Ca^{2+}$** | | 2%OligoG + 0mM DNA + 1mM $Ca^{2+}$ | 2%OligoG + 0mM DNA + 5mM $Ca^{2+}$ |
| **$Ca^{2+}$ & DNA** | | 2%OligoG + 1mM DNA + 1mM $Ca^{2+}$ | 2%OligoG + 10mM DNA + 5mM $Ca^{2+}$ |

**Table 6-2: Concentrations of DNA and $Ca^{2+}$ ions incubated with 2% OligoG prior to FTIR analysis.**

## 6.3 Results

### 6.3.1 OligoG IR Spectrum and Peak Characterisation

Replicate (n=18) FTIR spectra of OligoG were acquired to establish the reproducibility of OligoG IR-spectra and to identify peaks associated with OligoG.

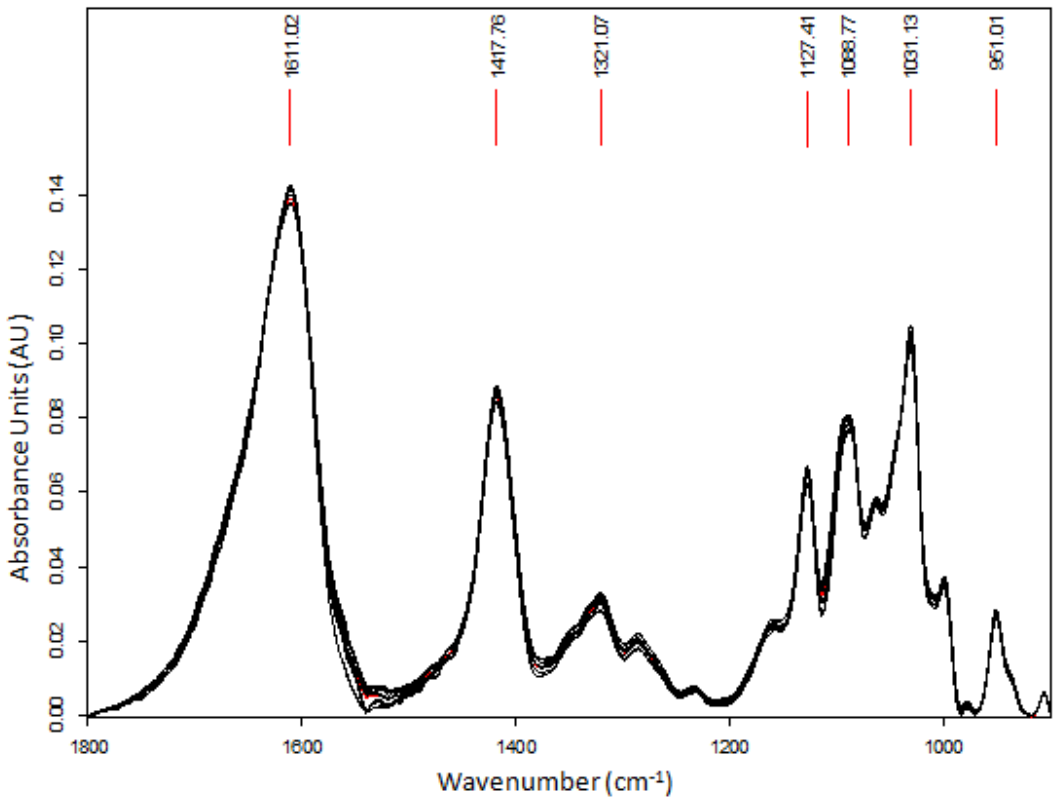


**Figure 6-2: Baseline-corrected absorbance spectra of OligoG (n=18) from 1800-900cm$^{-1}$. Average positions of major peaks are highlighted. Very little variation in peak position and absorbance can be observed across all replicate spectra.**

OligoG absorbance spectra (Figure 6-2 & Table 6-3) show many distinct and well-defined peaks within the wavenumber region of 1800 - 900cm$^{-1}$. The distribution of major peaks is mainly towards to glycogen-rich region (1200-900cm$^{-1}$), with four major peaks readily identifiable, compared to three in the rest of the spectrum (Figure 6-2). Peaks in this region are mainly associated with the C-O bonds within the pentose ring of each monomer unit. The glycosidic bond (O-O-C) linking each monomer unit, has been associated with bands around 987-993cm$^{-1}$, 1132-1136cm$^{-1}$, and 1141-1156cm$^{-1}$ (Kačuráková & Mathlouthi, 1996). A major peak is identified at 1127.413cm$^{-1}$ in the OligoG absorbance spectrum, which is close to the 1132-1136cm$^{-1}$ glycosidic bond peak range and could therefore represent the glycosidic linkages between each α-L-guluronate monomer.

The average characteristics of each major peak are described in Table 6-3, with standard deviations calculated from n=18 replicate spectra shown. For each peak, the standard deviations of the wavenumber and absorption intensity are small, suggesting little variation in OligoG molecular structure. The standard deviation of the peak width at 1031.132cm$^{-1}$ is much larger than the standard deviations of all other

peaks. Additionally, the average width of 50.351 indicates that the 1031.132cm$^{-1}$ peak is broad, and could be comprised of multiple smaller peaks corresponding to separate molecular species. Wide peaks are also observed at 1611.018, 1417.758, and 1321.069cm$^{-1}$, indicative of additional smaller absorbance bands being present within the large peaks.

| Wavenumber (cm$^{-1}$) | Absorption Intensity | Peak Width (FWHM) |
|---|---|---|
| 1611.018 (0.407) | 0.139 (0.0014) | 70.496 (2.361) |
| 1417.758 (0.414) | 0.086 (0.0014) | 37.0645 (0.262) |
| 1321.069 (0.334) | 0.03 (0.0013) | 41.8643 (1.181) |
| 1127.413 (0.127) | 0.064 (0.0016) | 15.3828 (0.281) |
| 1088.77 (0.859) | 0.079 (0.0016) | 21.426 (0.748) |
| 1031.132 (0.214) | 0.100 (0.0016) | 50.351 (16.717) |
| 951.0141 (0.159) | 0.027 (0.0008) | 16.0141 (0.302) |

**Table 6-3: OligoG absorbance spectra peak positions, absorbance intensities and peak widths with standard deviations shown in brackets. Low standard deviation in peak positions and absorbance intensity can be seen for all major peaks. The higher standard deviation for the peak width at the peak centred around 1031.132cm$^{-1}$, alongside the higher peak width suggests that this peak may be composed of multiple underlying absorption bands. These bands could be associated with molecular structures within OligoG. FWHM: Full Width at Half Maximum.**

To confirm this, second-derivative spectra can be calculated from the absorbance spectra. Calculation of second-derivative spectra determines the rate-of-change within a moving window of a defined size, which moves across the absorbance spectrum. Small peaks and shoulders within larger absorbance peaks cause directional change and are therefore detected as a difference in rate-of-change of absorbance by the second-derivative calculation.

The second-derivative of the OligoG absorbance spectra were calculated to increase the resolution of peaks obscured by large and/or broad peaks. The number of peaks detected within the glycogen-rich region increased from four in the absorbance spectrum, to eleven in the second-derivative spectrum. Figure 6-3 shows the second-derivative spectrum for the (a) whole fingerprint region (1800-900cm$^{-1}$) and, (b) the glycogen-rich region (1200-900cm) in closer detail.

| 2nd Derivative Peak Position ($cm^{-1}$) | 2nd Derivative Peak Width (FWHM) | Absorbance Peak Position ($cm^{-1}$) | Difference between Absorbance & 2nd Derivative Peak Position |
|---|---|---|---|
| **1126.751 (0.309)** | **13.0512 (0.513)** | **1127.413 (0.127)** | **0.662** |
| 1095.923 (0.387) | 20.1977 (6.285) | | |
| **1084.971 (0.179)** | **21.689 (0.907)** | **1088.77 (0.859)** | **3.799** |
| 1064.257 (0.531) | 10.8373 (1.097) | | |
| 1047.397 (0.712) | 9.0654 (0.999) | | |
| **1029.659 (0.190)** | **11.6004 (0.315)** | **1031.132 (0.214)** | **1.473** |
| 997.239 (0.182) | 10.3055 (0.263) | | |
| 977.964 (0.228) | 35.004 (7.725) | | |
| **951.4249 (0.158)** | **9.1402 (0.377)** | **951.0141 (0.159)** | **-0.4108** |
| 935.826 (0.331) | 10.3638 (27.642) | | |
| 905.5773 (0.256) | 3.3965 (0.459) | | |

**Table 6-4: Peak positions and corresponding widths with standard deviations in brackets of average OligoG second-derivative spectrum from 1200-900$cm^{-1}$, detected by the peak detection algorithm in OPUS. The positions of OligoG peaks in the absorbance spectrum are also shown for comparison, with difference in wavenumbers between the peaks. Peaks detectable in both absorbance and second-derivative spectra are bolded for clarity. FWHM: Full Width at Half Maximum.**

Small changes in wavenumber position of major peaks can be seen, for example from 1031.132$cm^{-1}$ in the absorbance spectrum to 1029.659$cm^{-1}$ in the second-derivative spectrum, with a smaller secondary peak becoming evident at 1047.397$cm^{-1}$. It is possible that the absorbance peak found at 1031.132$cm^{-1}$ may be the sum of these two second-derivative peaks; this is likely considering the large peak width of the peak centred around 1031.132$cm^{-1}$.

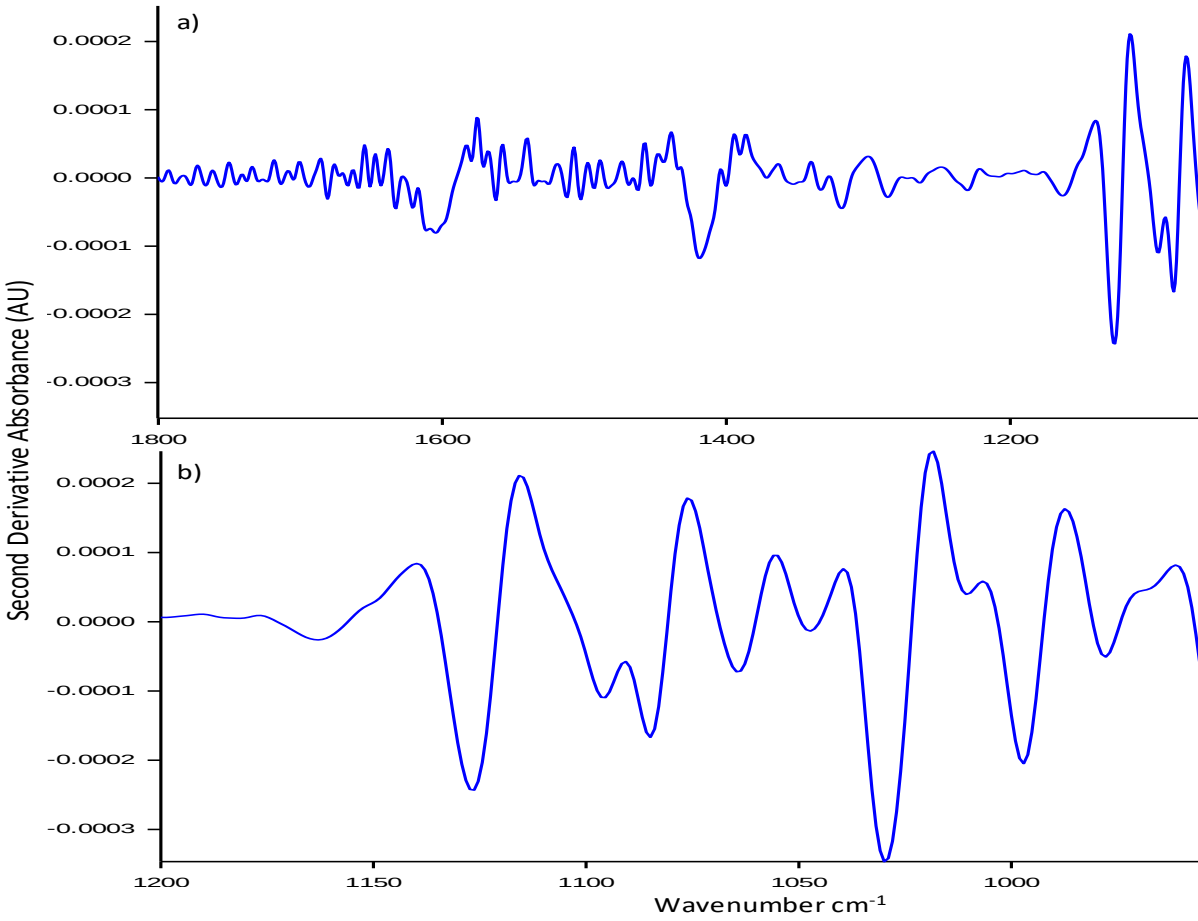


**Figure 6-3: Second derivative average spectrum of OligoG showing the (a) fingerprint region from 1800-900cm$^{-1}$ from 1200-900cm$^{-1}$.**

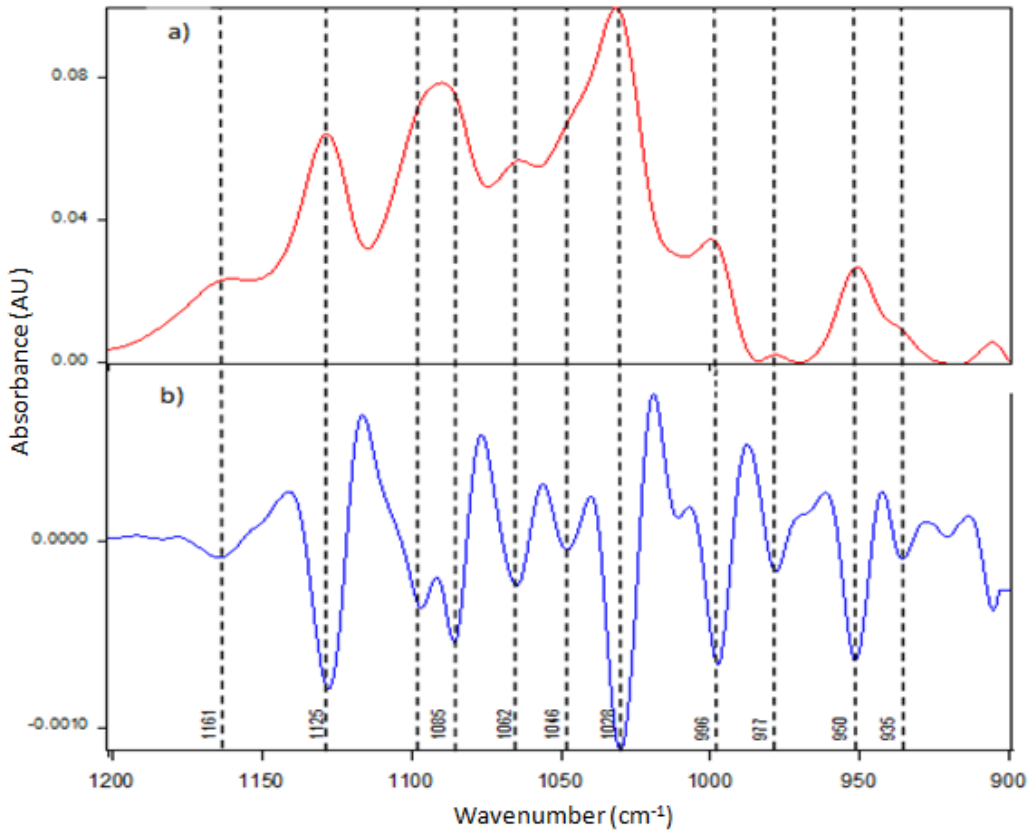


**Figure 6-4: Average FTIR (a) absorbance and (b) second-derivative spectra of OligoG from 1200-900cm-1, with dashed lines highlighting how the second-derivative peaks (pointing downwards) correspond with peaks and shoulders observed in the absorbance spectrum.**

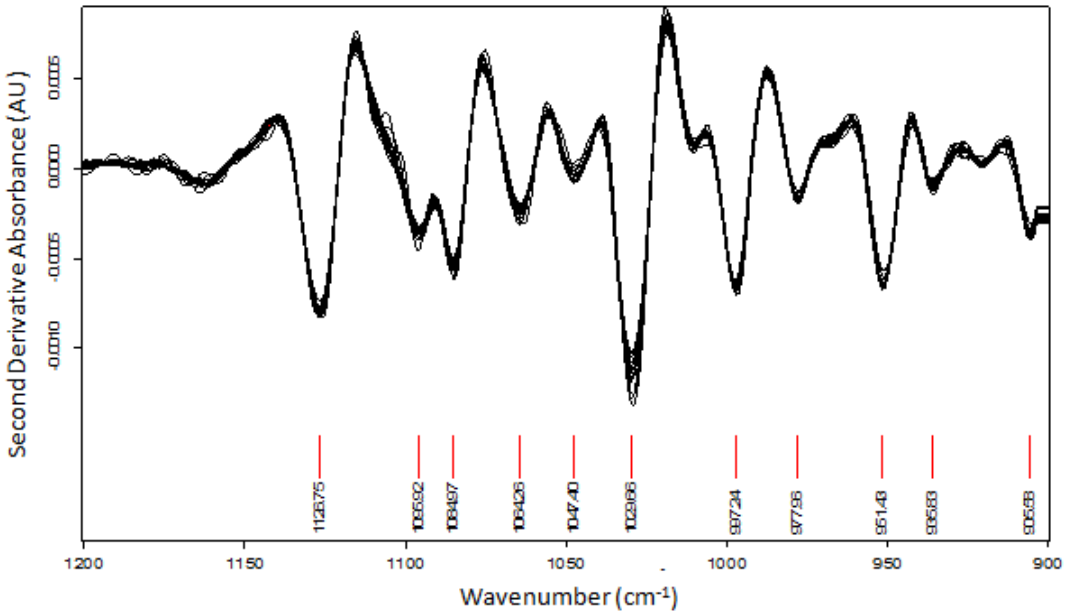


**Figure 6-5: Replicate (n=18) 2nd derivative spectra of OligoG from 1200-900cm-1 with average peak positions shown. As in Figure 6-2, the variability between replicate spectra is low, and many peaks are easily defined.**

### 6.3.2 Studying Interactions of CF sputum with OligoG by FTIR

CF patients, like COPD patients, produce a high volume of sputum although it is known that the viscoelastic properties of sputum produced by each group of patients are different and this contributes towards differing pathologies of the disease (Voynow, 2002; Voynow & Rubin, 2009). OligoG has been shown to have an alleviating effect on the symptoms of CF, possibly through $Ca^{2+}$ mediated interactions between (Ermund *et al.*, 2017). However there are uncertainties around OligoG, with respect in particular to its mode-of-action, but also to its persistence in the lung. FTIR spectroscopy is highly-efficient at detecting very low quantities of molecules within more complex biological mixtures, for example detection of environmental pollutants in animal tissues (Llabjani *et al.*, 2012). The work presented in this section seeks to evaluate the ability of FTIR to detect and quantify OligoG within the complex biofluid of sputum.

#### 6.3.2.1 FTIR Analysis of CF Sputum

Thus far this thesis has focussed heavily on sputum from COPD patients, and a good representation of the molecular structure of COPD sputum has been made. COPD and CF are both characterised by excessive mucus load, leading to restricted airways and reduced lung function. The successes of OligoG for treating CF patients described previously, coupled with the similarities in respiratory symptoms of both diseases, presents the possibility that OligoG could also have success for treating COPD patients. Therefore, before examining OligoG-dosed CF sputum, it is necessary to have an understanding of the molecular similarities and differences between COPD and CF sputum. Whilst excessive mucus load in the airways and lung tissue damage are common features of both diseases, they differ in the mechanisms which contribute to this increased load and damage. One important difference is an increased presence of bacterial DNA in CF sputum compared to other respiratory diseases, including COPD (Brandt, Breitenstein, von der Hardt, & Tümmler, 1995; Matthews, Spector, Lemm, & Potter, 1963; Riethmueller *et al.*, 2008; Shak, Capon, Hellmiss, Marsters, & Baker, 1990). Increased DNA presence in sputum creates a secondary polymeric structure, alongside the primary mucin polymer mesh formed during MUC5AC and MUC5B secretion. This secondary mesh is more rigid than the mucins mesh and contributes to increased viscoelasticity of CF sputum (Voynow & Rubin, 2009).

Both COPD and CF spectra (Figure 6-6) show dominant amide I (~1650cm$^{-1}$) and amide II (~1530cm$^{-1}$) peaks, as well as a series of peaks and shoulders in the glycogen-rich region (1200 - 900cm$^{-1}$), which are likely to be resultant from the high mucin glycoprotein content of sputum. A main difference between COPD and CF sputum appears around 1410cm$^{-1}$, where the CF spectrum shows a distinct peak, compared to the COPD sputum spectrum. This peak correlates well with the strong 1409cm$^{-1}$ peak seen in the DNA spectrum (Figure 6-6). Futher evidence for the elevated level of DNA affecting the CF sputum spectrum can be seen within the second-derivative spectra (Figure 6-7).

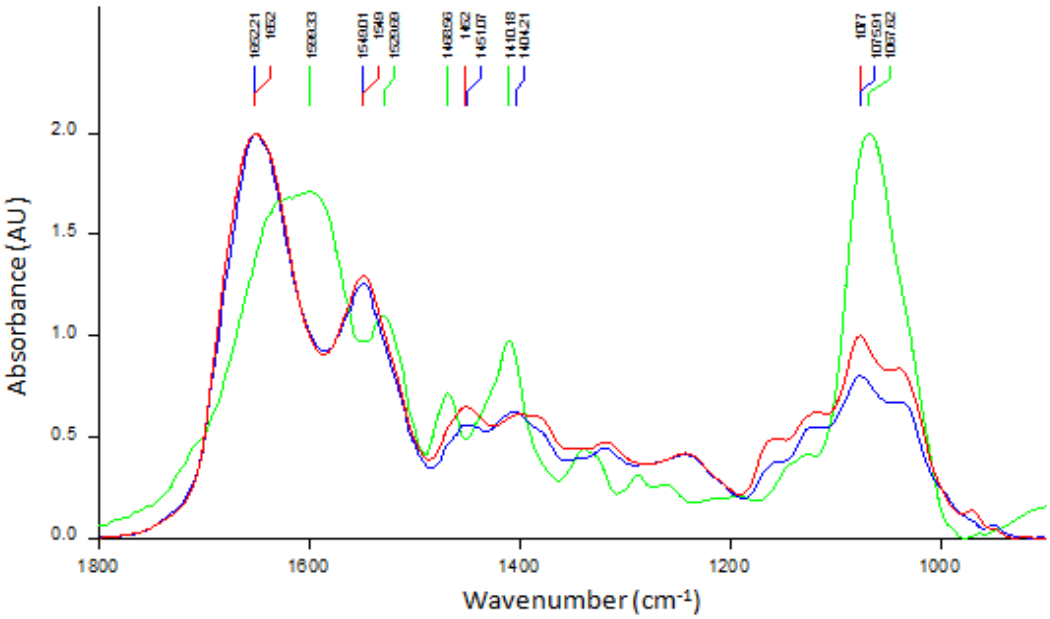


**Figure 6-6: Baseline-corrected, min-max-normalised FTIR spectra of sputum samples from randomly selected COPD (red), and CF (blue) patients from 1800-900cm$^{-1}$. A spectrum of human DNA (green) is also shown. Major peak positions of each spectrum are shown in the corresponding colours.**

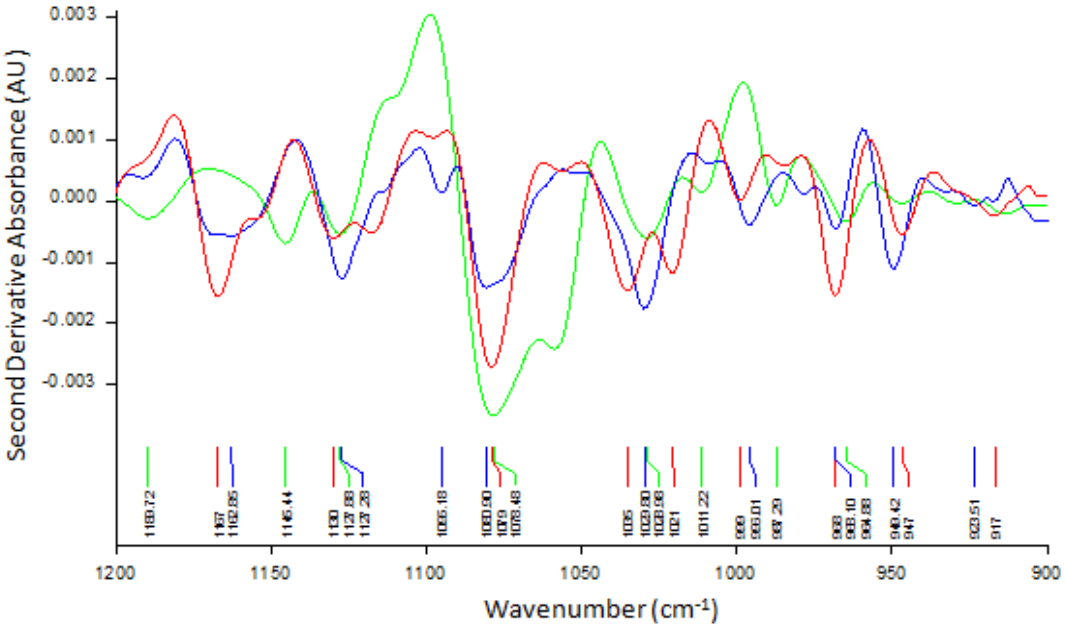


**Figure 6-7: Second derivative spectra of sputum samples from randomly-selected COPD (red), CF (blue) patients and DNA (green) from 1200-900cm$^{-1}$. Positions of detected peaks are highlighted.**

Both the CF and COPD sputum second-derivative spectra show similar overall patterns, however a notable difference is observable with the range of 1040-1020cm$^{-1}$. The COPD sputum shows two distinct peaks at 1035 and 1021cm$^{-1}$, whereas the CF sputum shows only one at 1029.80cm$^{-1}$ matching the DNA spectrum peak at 1028.98cm$^{-1}$. Peaks at these wavenumbers are associated with C-O bonds in pentose and hexose rings. The sugars found in mucin glycoprotein glycan chains are hexose ring structures whereas the sugars in DNA are deoxy-ribose pentose structures.

Recent work has identified potential IR bands which could be markers of many pentose and hexose sugars, where 1011cm$^{-1}$ is suggested as a potential marker for deoxy-ribose (Wiercigroch *et al.*, 2017). The authors did not identify a peak at or around 1028cm$^{-1}$ for deoxy-ribose, although phosphate is known to produce an absorption band within this region (Fredericks, Bennett, Williams, & Rogers, 2012), so it is possible that the peak seen at 1028.98cm$^{-1}$ in the DNA spectrum (Figure 6-7) could correspond to the phosphate backbone of the DNA molecule. Similarly, the peak identified at 1029.80cm$^{-1}$ in the CF sputum spectrum may contain contributions from the phosphate backbone of any DNA present in the sputum. A peak in the DNA spectrum is identified at 1011cm$^{-1}$, and a change in the shape of the CF spectrum is observable, although a peak is not identified (Figure 6-8). This change in peak shape

in the CF sputum spectrum may be due to the presence of the deoxy-ribose sugars present in the DNA backbone.

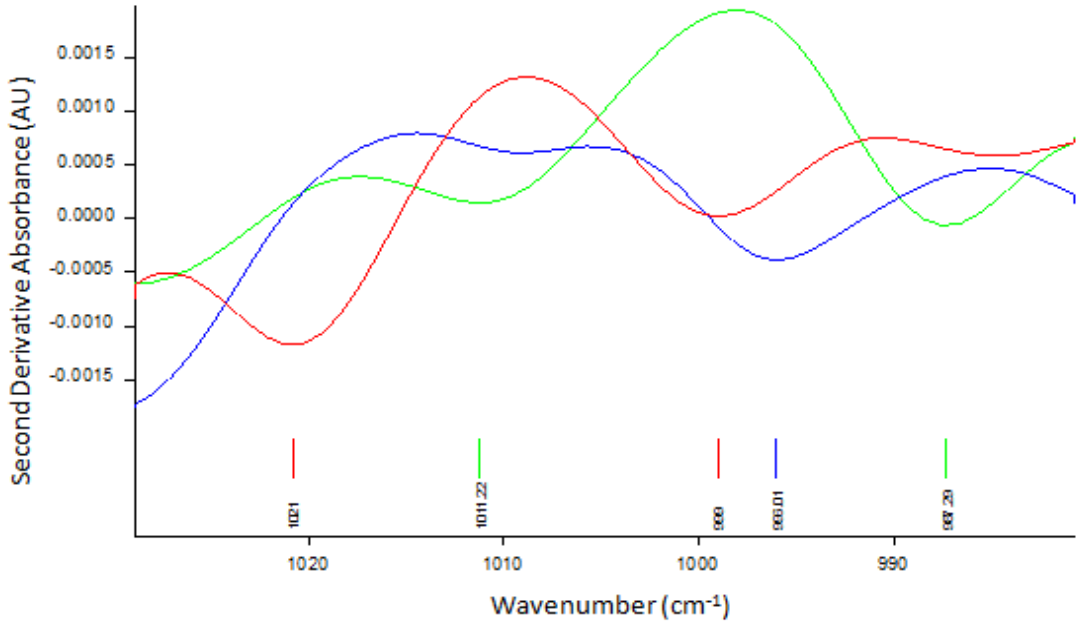


**Figure 6-8: Second derivative spectra of sputum samples from randomly selected COPD (red), and CF (blue) patients and DNA (green) from 1030-980cm$^{-1}$. Positions of detected peaks are highlighted.**

Accurate identification of the characteristics of CF sputum spectra is important for accurately characterising OligoG/mucin interactions by FTIR spectroscopy. Molecular dynamics simulations, FTIR spectroscopy and isothermal titration calorimetry experiments have suggested that OligoG does not interact with directly with DNA (Powell *et al.*, 2018), but DNA does produce some strong IR absorption bands which could influence the parameters of mucin and/or OligoG peaks. Whilst current results suggest that OligoG and DNA in sputum do not interact, the elevated levels of DNA in CF sputum are known to change the viscoelastic properties of sputum. Therefore, this change in physical properties may influence the interaction of OligoG with COPD sputum mucins and this should be taken into account before any potential studies of OligoG in COPD patients.

### 6.3.2.2 FTIR Spectra of OligoG-incubated CF Sputum Spectra

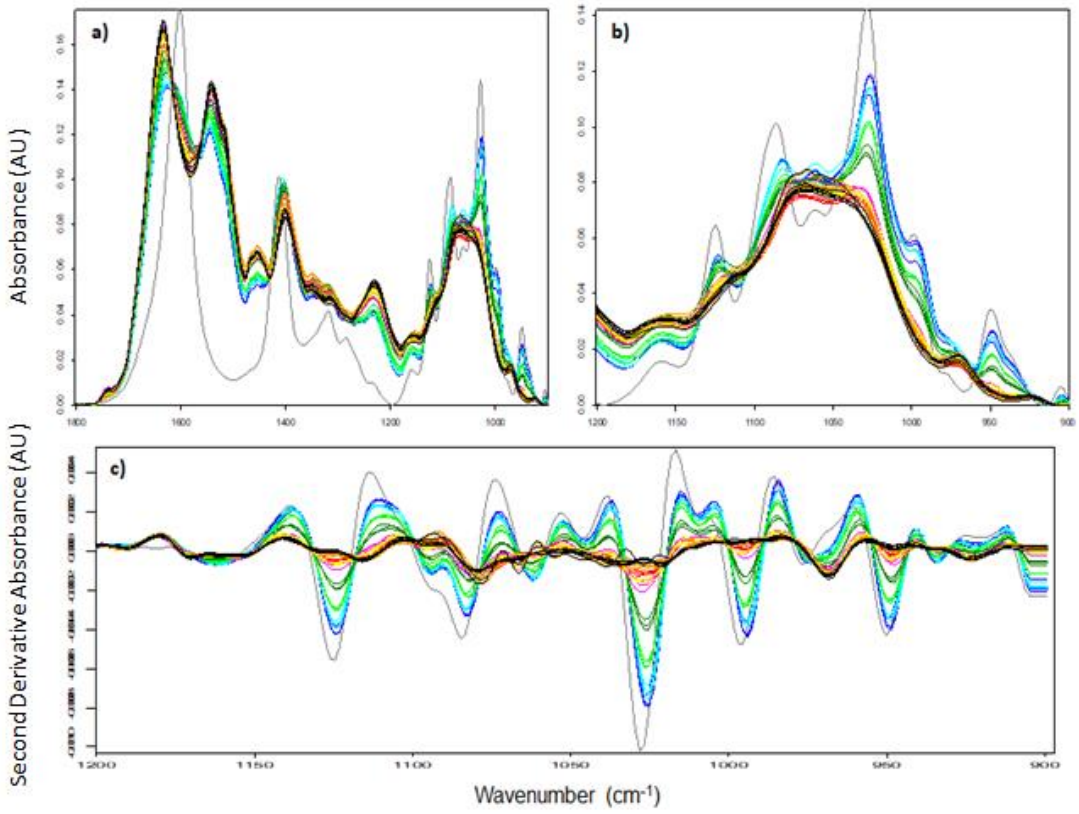


**Figure 6-9: Vector-normalised, baseline-corrected IR spectra from (a) 1800-900cm$^{-1}$ and (b) 1200-900cm$^{-1}$, and (c) second-derivative spectra from 1200-900cm$^{-1}$ of OligoG (grey), control CF-sputum (black), and OligoG-incubated-CF sputum in progressively increasing final concentrations from 0.02% to 20% (orange 0.02%; red 0.05%; yellow 0.1%; pink 0.2%, dark-green 0.5%; light-green 1.0%; light-blue 1.5%; dark-blue 2.0%). As the OligoG concentration in sputum increases, the sputum spectra show peaks which are more similar to those observed in the OligoG spectrum. This is especially evident in the glycogen-rich region of the spectrum (b and c).**

Sputum was obtained from a CF patient at time of screening/recruitment to the study, prior to receiving any dose of OligoG, and was incubated with OligoG solutions at varying concentrations from 0% (w/v) to 20% (w/v) at a 1 in 10 dilution (OligoG:Sputum) for final OligoG-in-sputum concentrations of 0:2% (w/v). Sputum obtained from the screening/recruitment stage was chosen as the patient had not yet been exposed to either the drug or placebo. Therefore, it was possible to certain that there was no residual OligoG presence in the sputum, so baseline levels of absorption at the wavenumbers detailed above (Table 6-3) could be established. This also allowed any change in absorption to be directly attributed to OligoG incubation. Sputum samples were incubated with OligoG as described in the methods section.

FTIR spectra of CF sputum incubated with OligoG at increasing concentrations were acquired and then analysed. As the concentration of OligoG in

sputum is increased, the absorbance intensity of wavenumbers corresponding to OligoG peaks shows a corresponding increase (Figure 6-9). This is most strongly evident within the glycogen-rich region (Figure 6-9b), where OligoG-associated peaks in the sputum spectra become evident at low concentrations of OligoG. For example, an increase in absorption at approximately 1030cm$^{-1}$ can be observed at 0.2% w/v OligoG in sputum, which is detectable as an additional peak to the sputum glycogen peak at approximately 1065cm$^{-1}$ (Table 6-5). As the concentration of OligoG is further increased, the absorption peak profile within the glycogen-rich region becomes more similar to that of OligoG (Table 6-5).

| OligoG | Control | 0% | 0.02% | 0.05% | 0.10% | 0.20% | 0.50% | 1.00% | 1.50% | 2.00% |
|---|---|---|---|---|---|---|---|---|---|---|
| 1127.413 (0.127) | | | | | | | | 1122.4861 (0.0608) | 1122.8391 (0.041) | 1123.0066 (0.0255) |
| 1088.77 (0.859) | | | | | | | | | 1082.1809 (0.0143) | 1082.2961 (0.0425) |
| | 1066.9449 (0.1719) | 1067.0956 (0.7342) | 1066.431 (0.1134) | 1069.5283 (0.1396) | 1065.1708 (1.8751) | 1065.2001 (0.2937) | | | | |
| | | | | 1038.3549 (0) | | | | | | |
| 1031.132 (0.214) | | | | | | 1032.4149 (0) | 1028.6331 (0.258) | 1027.5035 (0.0409) | 1026.675 (0.1782) | 1026.7871 (0.3189) |
| 951.0141 (0.159) | | | | | | | | 948.9743 (0.1039) | 948.7242 (0.2563) | 948.9345 (0.2694) |

**Table 6-5: Average wavenumber (cm$^{-1}$) positions of IR-absorbance peaks detected in control and OligoG-incubated CF sputum; standard deviations shown in brackets. Peaks associated with OligoG are highlighted in red.**

| OligoG | Control CF Sputum | Final OligoG concentration in Incubated CF Sput | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0% | 0.02% | 0.05% | 0.10% | 0.20% | 0.50% | 1.0 |
| | 1191.0853 (0.99918) | 1190.6669 (0.55083) | 1190.7585 (0.02175) | 1192.3993 (0.44497) | 1192.5308 (0.36911) | 1191.8661 (0.3437) | | |
| | 1170.0004 (0.5606) | 1170.6249 (0.18043) | 1166.7727 (2.77815) | | 1166.7367 (1.97165) | 1168.588 (0.1487) | 1163.9194 (0) | 1162. (0.51 |
| | 1155.4481 (2.10901) | 1153.0801 (0.43105) | 1154.0896 (0.92342) | 1154.9698 (1.02161) | 1156.4085 (0) | 1155.3805 (0.25215) | 1159.2292 (0.36065) | |
| 1126.751 (0.309) | | | 1127.3124 (0) | 1125.7938 (0.2803) | 1124.0108 (0.29666) | 1123.5928 (1.04382) | 1123.8721 (0.05495) | 1124. (0.10 |
| | 1116.1911 (0.45161) | 1116.1528 (0.42845) | 1115.0315 (1.0284) | 1118.6097 (0) | | | | |
| 1095.923 (0.387) | 1097.0973 (0) | | | 1095.8289 (0.5008) | 1094.9142 (0.00525) | 1096.4296 (0.2527) | 1094.6252 (0.43936) | 1094 (( |
| 1084.971 (0.179) | 1079.2034 (0.73885) | 1079.1527 (0.25292) | 1077.9319 (0.51803) | 1079.4059 (0.71975) | 1080.1017 (0.13945) | 1080.8203 (0.93451) | 1081.9415 (0.27913) | 1082. (0.31 |
| 1064.257 (0.531) | 1066.9644 (0) | 1066.3775 (0) | 1066.4709 (0.06174) | 1061.0571 (0) | 1064.6519 (2.60876) | 1065.3826 (0.28463) | 1062.5991 (0.08478) | 1062. (0.61 |
| | | 1055.5424 (0) | 1055.2209 (0.07104) | | | | | |
| 1047.397 (0.712) | | | | | 1043.9682 (0) | 1043.1787 (0.58235) | 1044.4327 (0.40601) | 1045. (0.0 |
| | 1035.3473 (2.02689) | 1034.5655 (2.8002) | 1037.5503 (0.45267) | | | | | |
| 1029.659 (0.190) | 1021.3348 (0.28405) | 1023.3314 (2.67252) | 1027.3264 (0.26962) | 1027.6543 (1.80979) | 1027.4842 (0.81041) | 1027.112 (0.2545) | 1026.1277 (0.08476) | 1026 (0.1 |
| | | 1018.3135 (0) | | 1008.1413 (0) | | 1010.1535 (0.246) | 1008.9127 (0) | 1009. (0.23 |
| 997.239 (0.182) | 993.5119 (5.81825) | | 996.3369 (0) | 996.3228 (1.08228) | 995.9859 (0.09168) | 995.3628 (0.55131) | 994.7463 (0.13209) | 994.8 (0.06 |
| 977.964 (0.228) | 968.329 (0.17396) | 968.7587 (0.18003) | 968.6719 (0.25908) | 968.5463 (0.04051) | 968.5094 (0.31091) | 969.0161 (0.15093) | 968.7541 (0.18761) | 971.8 (2.07 |
| 951.4249 (0.158) | 948.03 (0.79115) | 947.4837 (0.0641) | 947.2096 (1.47375) | 947.8127 (0.44371) | 948.2972 (0.5146) | 948.3355 (0.3151) | 948.5597 (0.19562) | 949. (0.02 |
| 935.826 (0.331) | | | | | | 934.4534 (0) | 933.7569 (0) | 934. (0.00 |
| | 923.8742 (1.21977) | 923.6613 (0.92118) | 923.285 (0.46075) | 922.9677 (0.18064) | 923.4774 (0.30518) | 922.8006 (0.35753) | 921.6212 (1.1278) | |
| 905.5773 (0.256) | | 907.3613 (0) | 907.7411 (0) | | | | | |

**Table 6-6: Average wavenumber ($cm^{-1}$) positions of second-derivative peaks detected in control and OligoG-incu deviations shown in brackets. Peaks corresponding with OligoG are highlighted in red.**

### 6.3.2.2 Absorption Curves

OligoG-incubated sputum IR-spectra were generated and the absorption at peak positions corresponding to major OligoG absorption peaks were plotted against the corresponding concentration of OligoG in sputum. Table 6-2 shows the wavenumbers of interest which were examined. The observed absorbance values were plotted against OligoG concentration in sputum. Linear and fourth-order polynomial trend lines were fitted to the observed data and the goodness-of-fit of trend lines were evaluated using Pearson's Chi Square. The equations of both linear and polynomial trend lines are shown on each plot.

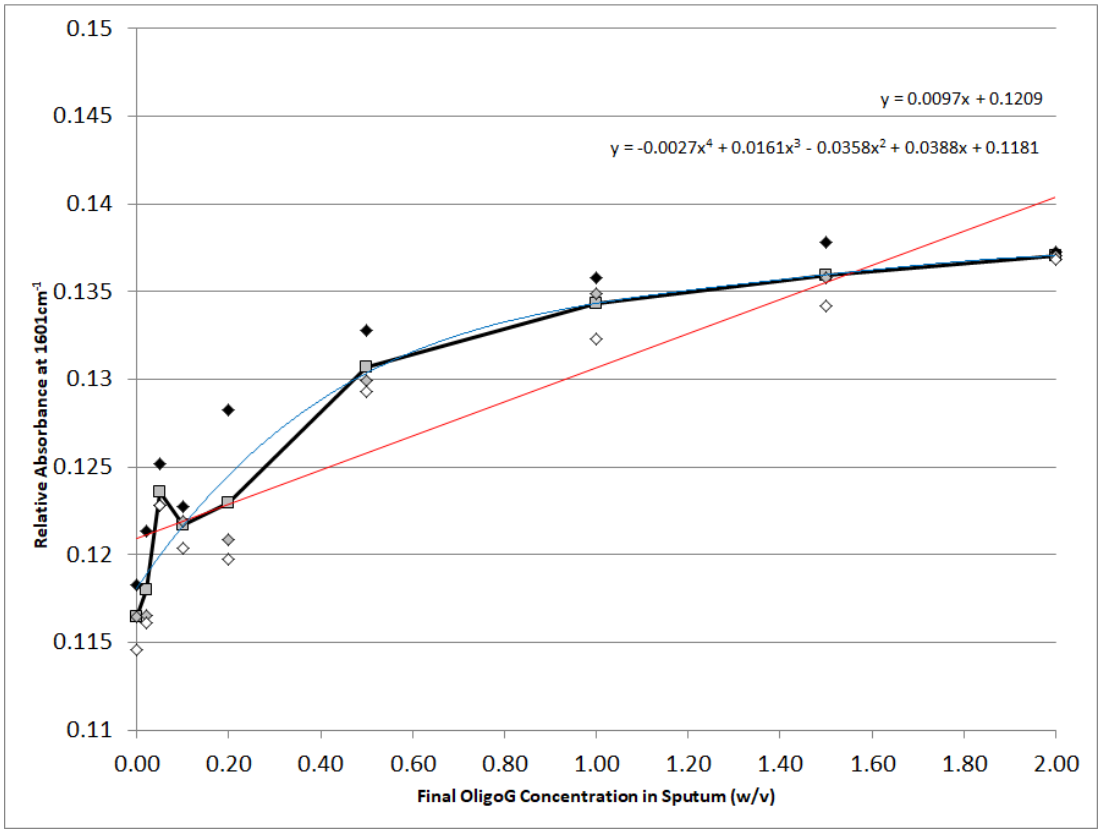


**Figure 6-10: Relative absorbance (AU) of 1601cm$^{-1}$, against final concentration of OligoG incubated in sputum (w/v) with linear (red) and polynomial (blue) trend lines fitted to the average absorbance at each concentration (black).**

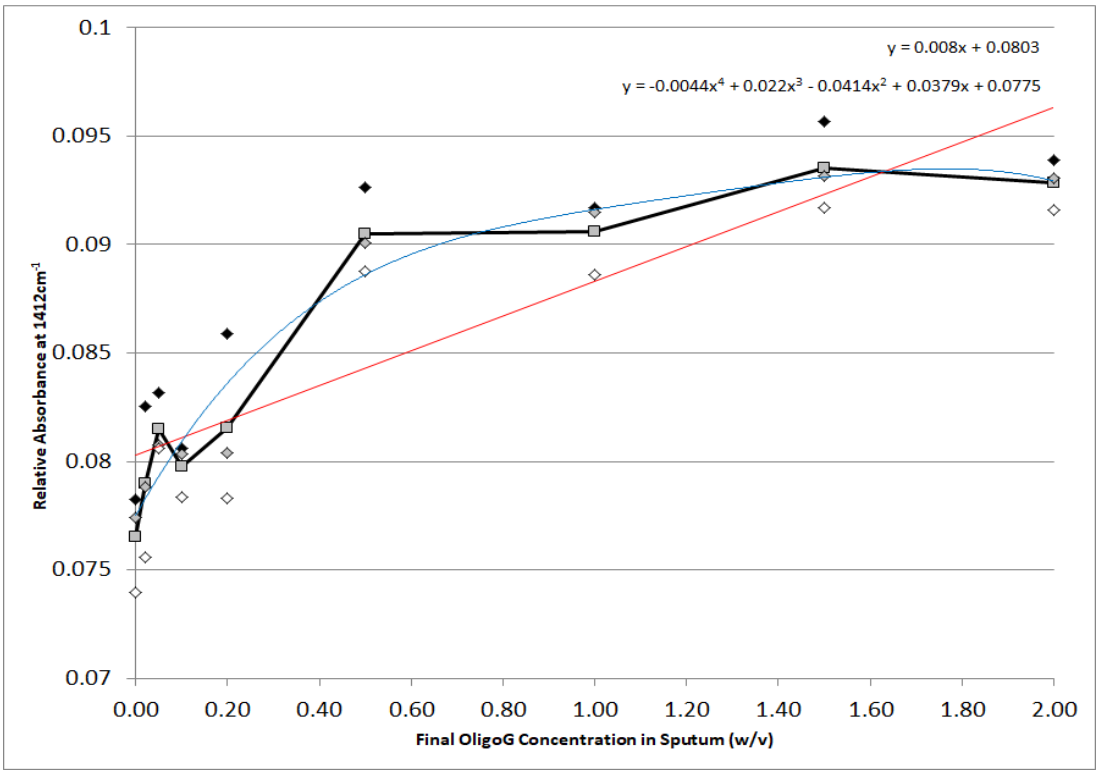


**Figure 6-11: Relative absorbance (AU) of 1412cm$^{-1}$, against final concentration of OligoG incubated in sputum (w/v) with linear (red) and polynomial (blue) trend lines fitted to the average absorbance at each concentration (black).**

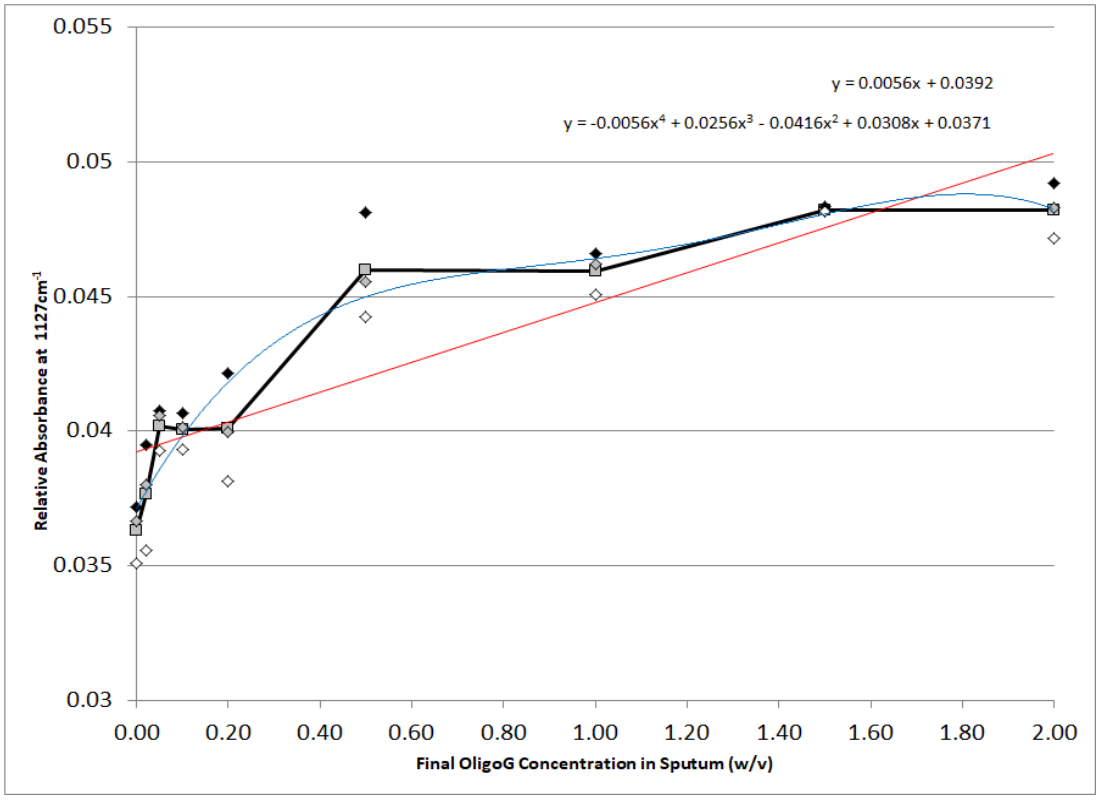


**Figure 6-12: Relative absorbance (AU) of 1125cm$^{-1}$, against final concentration of OligoG incubated in sputum (w/v) with linear (red) and polynomial (blue) trend lines fitted to the average absorbance at each concentration (black).**

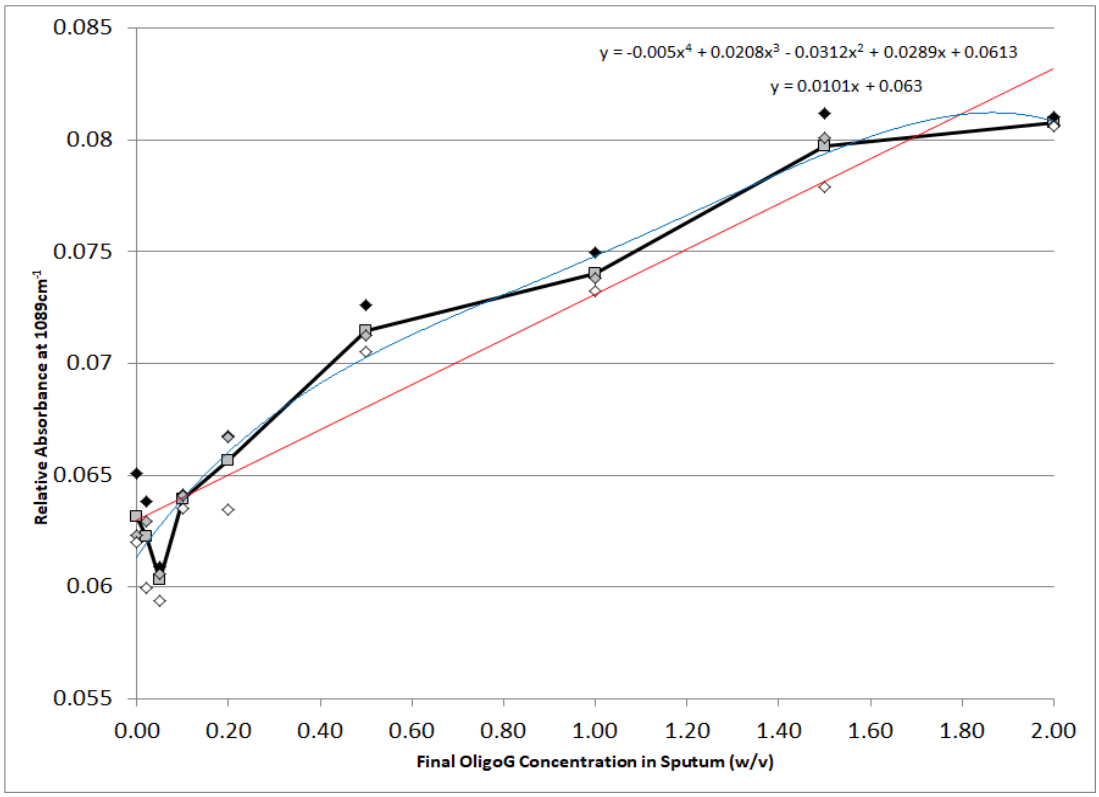


**Figure 6-13: Relative absorbance (AU) of 1089cm$^{-1}$, against final concentration of OligoG incubated in sputum (w/v) with linear (red) and polynomial (blue) trend lines fitted to the average absorbance at each concentration (black).**

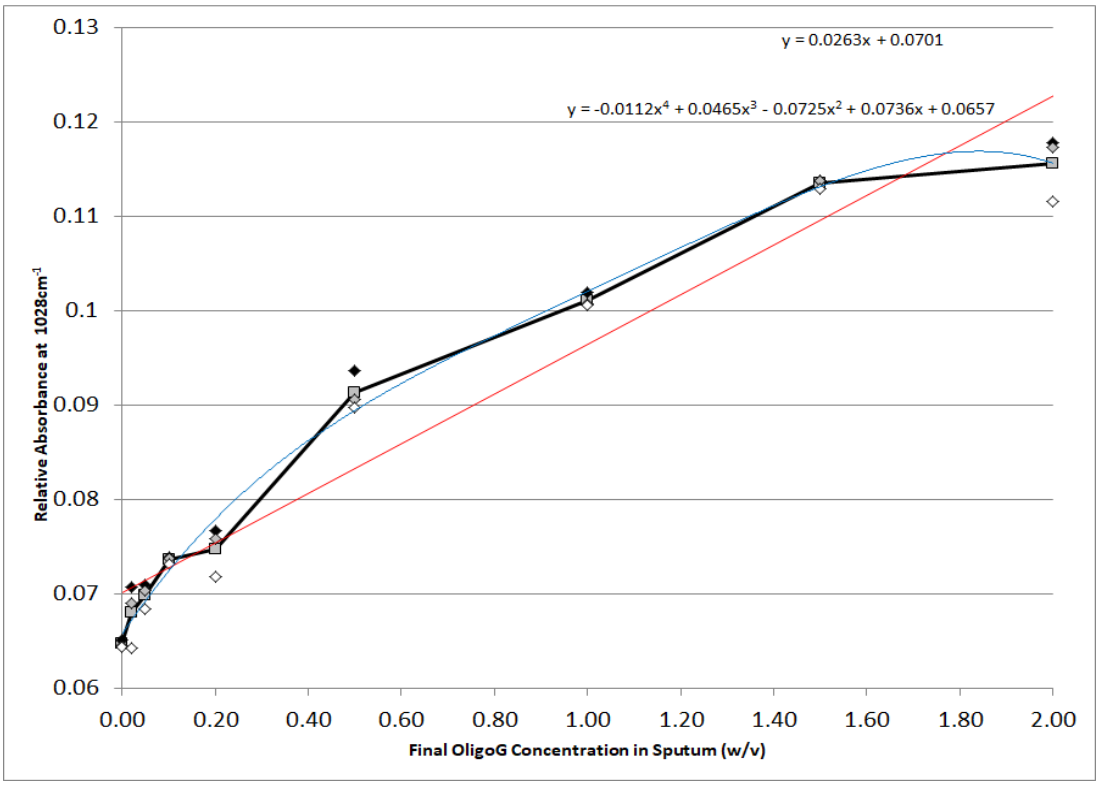


**Figure 6-14: Relative absorbance (AU) of 1028cm$^{-1}$, against final concentration of OligoG incubated in sputum (w/v) with linear (red) and polynomial (blue) trend lines fitted to the average absorbance at each concentration (black).**

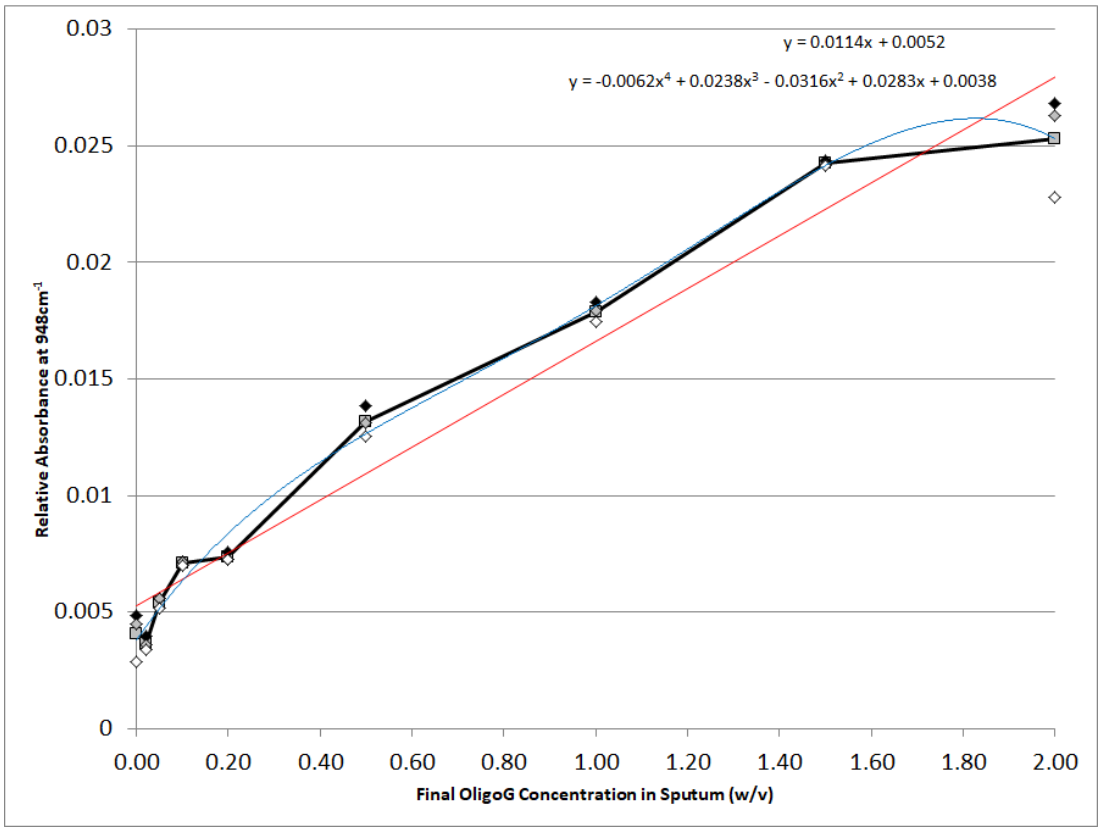


**Figure 6-15: Relative absorbance (AU) of 948cm$^{-1}$, against final concentration of OligoG incubated in sputum (w/v) with linear (red) and polynomial (blue) trend lines fitted to the average absorbance at each concentration (black).**

The absorption standard curves above all show an increase of absorption at OligoG-associated wavenumbers at increased concentrations of OligoG in CF-sputum.

| Wavenumber (cm$^{-1}$) | Pearson's Chi Square to Linear Fit | Pearson's Chi Square to Polynomial Fit | OligoG concentration statistically significantly different from control (% w/v) |
|---|---|---|---|
| **1601** | 0.91239 (p=0.000601) | 0.98004 (p=0.0000036) | 0.05 (p=0.0043) |
| **1412** | 0.88684 (p=0.001435) | 0.97608 (p=0.0000068) | 0.05 (p=0.0072) |
| **1125** | 0.89361 (p=0.001164) | 0.97643 (p=0.0000065) | 0.05 (p=0.0056) |
| **1086** | 0.96545 (0.00002441) | 0.98793 (p=0.00000063) | 0.10 (p=0.0271) |
| **1028** | 0.96923 (p=0.000016) | 0.99722 (p=0.000000004) | 0.05 (p=0.0117) |
| **948** | 0.9801 (p=0.0000036) | 0.99776 (p=0.000000002) | 0.10 (p=0.0277) |

**Table 6-7: Summary of standard curves of absorption at each OligoG-associated wavenumber (cm$^{-1}$), with Pearson's correlation coefficient for each fit to the observed data shown. The lowest final concentration (w/v) of OligoG in sputum which was**

**found to have statistically significantly ($p<0.05$) increased absorption compared to the control sputum sample is also shown for each wavenumber. Calculated p-values are shown in parenthesis.**

### 6.3.2.3 Correlation Analysis

Each IR-spectrum of OligoG-incubated and control CF-sputum was correlated to a reference OligoG spectrum using the non-parametric Spearman's Rho test of correlation. Both absorbance and second-derivative spectra were correlated to the corresponding OligoG absorbance or second-derivative spectrum. The most distinct peaks in the OligoG spectrum are found within the glycogen-rich region (Figure 6-2), from 1200-900cm$^{-1}$. For this reason, the correlation analysis was focussed within this region.

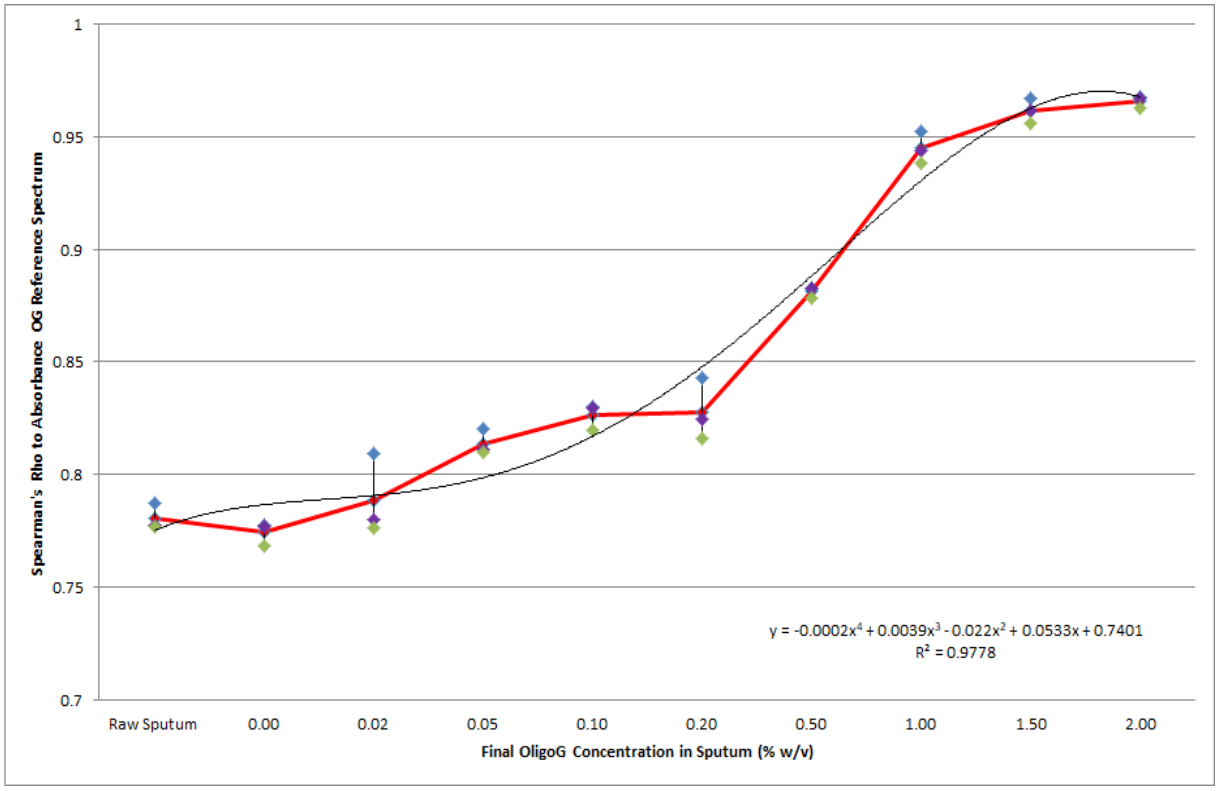


**Figure 6-16: Spearman's Rho of OligoG-incubated-sputum absorbance spectra to the reference OligoG absorbance spectrum against the final concentration of OligoG. As the concentration of OligoG in sputum is increased, the corresponding Spearman's Rho increases. The average Spearman's Rho coefficient is shown (red line) and a fourth-order polynomial trendline (black line) has been fitted, with an $R^2$ of 0.9778**

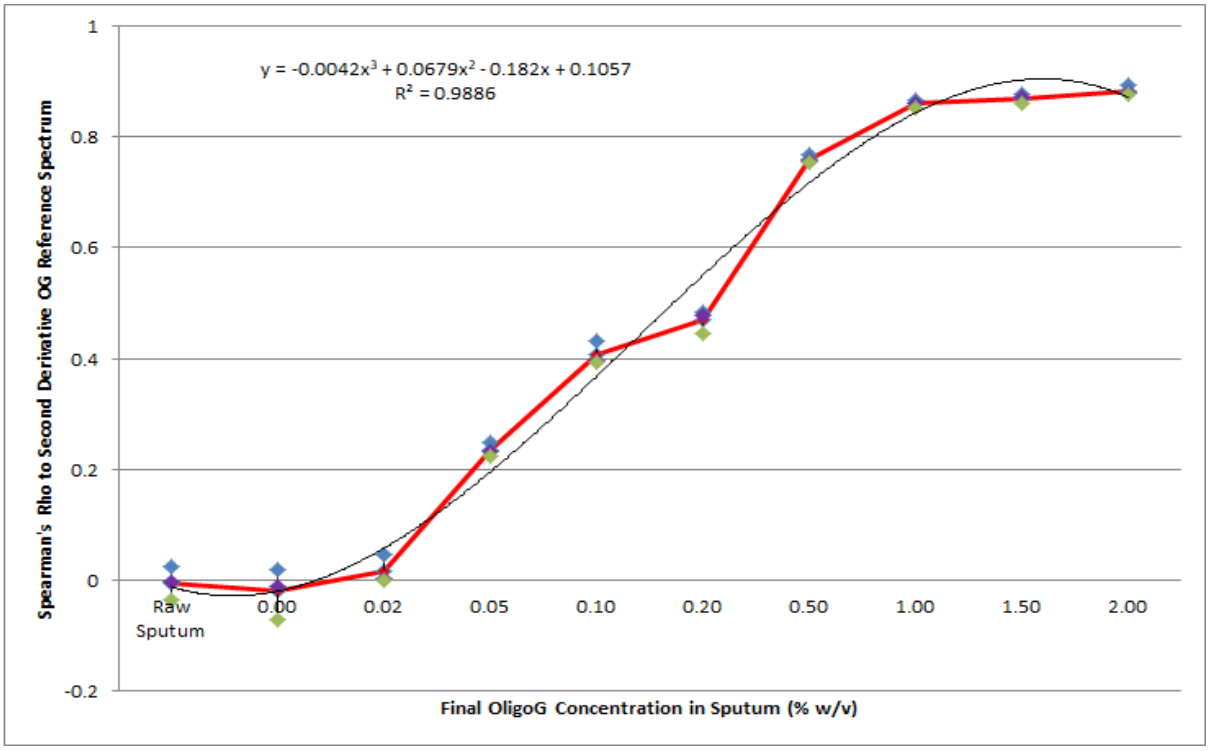


**Figure 6-17: Spearman's Rho of OligoG-incubated-sputum second-derivative spectra to the reference OligoG second-derivative spectrum against the final concentration of OligoG. As the concentration of OligoG in sputum is increased, the corresponding Spearman's Rho increases. The average Spearman's Rho coefficient is shown (red line) and a third-order polynomial trendline (black line) has been fitted, with an $R^2$ of 0.9886**

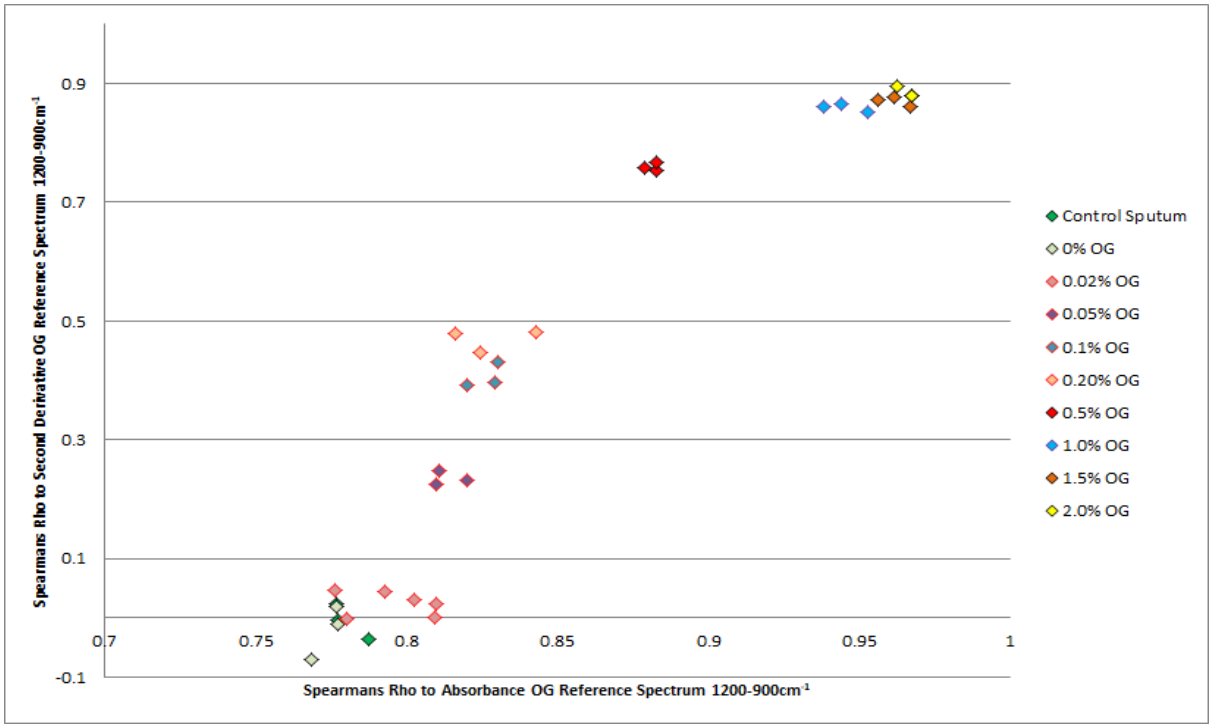


**Figure 6-18: The relationship between the Spearman's Rho coefficients for the absorbance and second-derivative spectra as the concentration of OligoG in sputum is**

**increased is shown. The correlation coefficient between the second-derivative sputum spectra and OligoG reference spectrum is low at concentrations of OligoG below 0.5% (w/v), but is statistically significantly increased at 0.02% (w/v). The correlation coefficients of the absorbance sputum spectra all show a high degree of similarity to the OligoG reference spectrum, with all coefficients being greater than 0.75.**

Correlation analysis shows how an increasing concentration of OligoG corresponds with an increase in Spearman's Rho coefficient when comparing the sputum spectrum to an OligoG reference spectrum. An increase in correlation can be observed even at 0.02% OligoG, which was found to be statistically significant at the 95% confidence level for both the absorbance spectra (one-tailed $p = 0.0142$), and second-derivative spectra (one-tailed $p = 0.0278$). Correlation coefficients for all other concentrations were also found to be statistically significantly different from the non-incubated control and 0%-OligoG incubated sputum ($p<0.05$). The correlation coefficients for the control and 0%-OligoG incubated sputum were not found to be statistically significantly different from each other at the 95% confidence level, for either the absorbance or second-derivative spectra (absorbance $p = 0.2261$, second-derivative $p = 0.6585$), showing that the differences observed were not due to the experimental conditions.

### 6.3.3 Detecting OligoG in Treated, Placebo and Control CF Patient Sputum Samples

Sputum samples were collected from CF patients at the time of screening, and during the OligoG- and placebo-treatment phases. Samples were analysed by FTIR to determine the capability FTIR to detect OligoG in sputum and distinguish it from placebo treatment.

#### 6.3.3.1 Correlation Analysis

In the previous section of this chapter, correlation analysis was shown to be effective for determining OligoG concentration in OligoG-incubated sputum. In this section, CF-patient sputum samples from screening and both treatment phases were analysed and the IR-spectra of sputum were compared to a reference average OligoG spectrum. Correlation coefficients of these sputum spectra to the reference OligoG spectrum were calculated using Spearman's Rho and a two-tailed hypothesis. Correlation coefficients were calculated for absorbance and second-derivative spectra, in the wavenumber region of 1200-900cm$^{-1}$. This region was identified as a potential region-of-interest because, as seen from Figure 6-2 to Figure 6-5, OligoG produces many strong and distinct bands within the region.

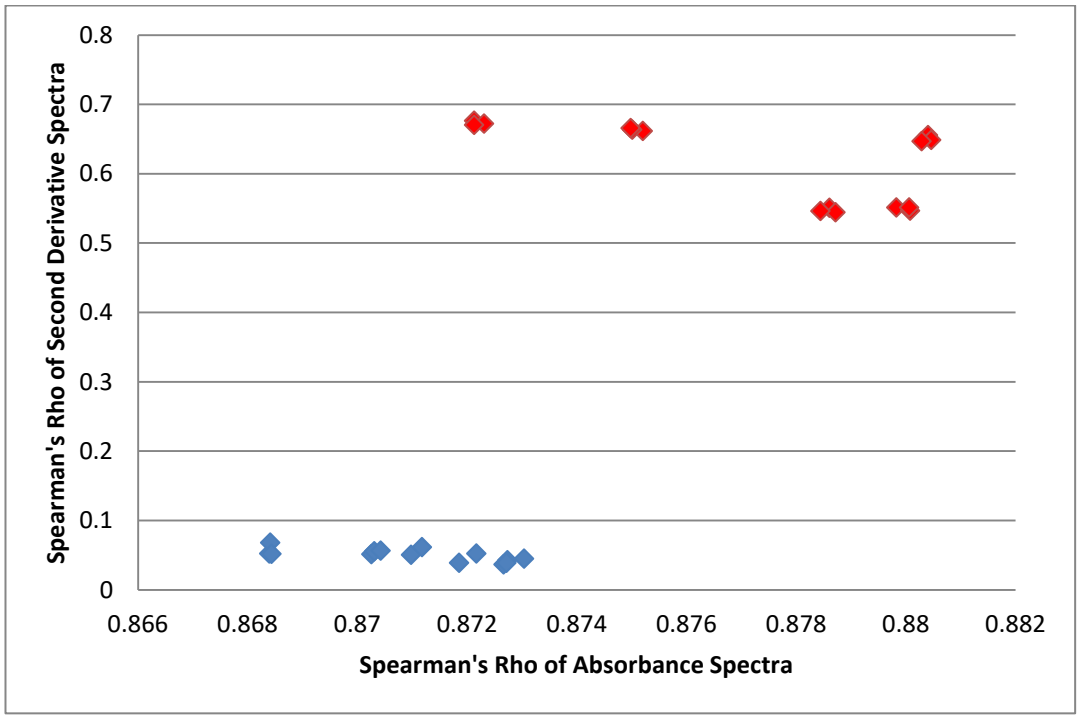


**Figure 6-19: Correlation coefficient scatterplot of absorbance and second-derivative CF patient sputum spectra to the absorbance and second-derivative OligoG spectrum for placebo-treated sputum (blue) and OligoG-treated sputum (red) from patient no. 82606003.**

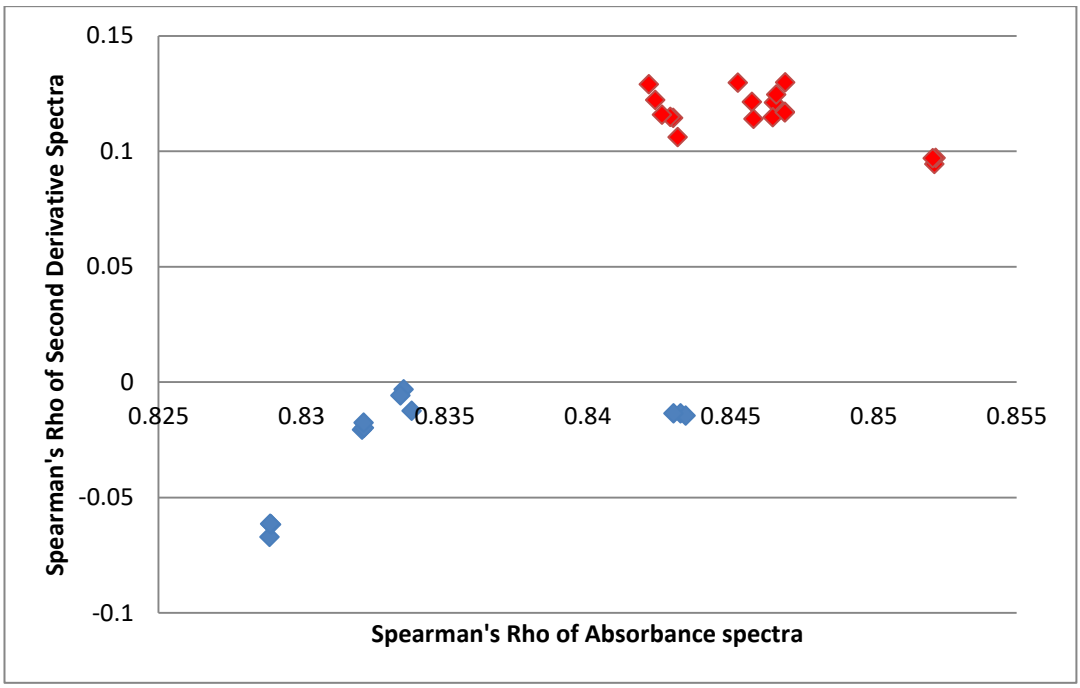


**Figure 6-20: Correlation coefficient scatterplot of absorbance and second-derivative CF patient sputum spectra to the absorbance and second-derivative OligoG spectrum for placebo-treated sputum (blue) and OligoG-treated sputum (red) from patient no. 82602002.**

The scatterplots above (Figure 6-19 & Figure 6-20) show that by calculating the correlation of the sputum absorbance or second-derivative spectrum to the OligoG absorbance or second-derivative spectrum respectively it is possible to distinguish between a pre-treatment sample which has not received any OligoG (blue) and a treatment phase sample which has received OligoG (red). Each group clusters together well, with few outliers and clear distinction between group boundaries. The variability seen in the plots could be due to the heterogeneous nature of sputum itself and the innate variability seen between patients. However despite this variation, this method is able to distinctly separate OligoG-treated sputum from non-treated sputum.

This analysis was then expanded to all patients recruited in the trial who had received both placebo and OligoG treatment. All of the correlation coefficients were plotted together to build a linear regression model for detection of OligoG in CF sputum (Figure 6-21). Table 6-8 shows the sensitivity and specificity of the model at 86 and 90%, respectively, for correct prediction of OligoG presence.

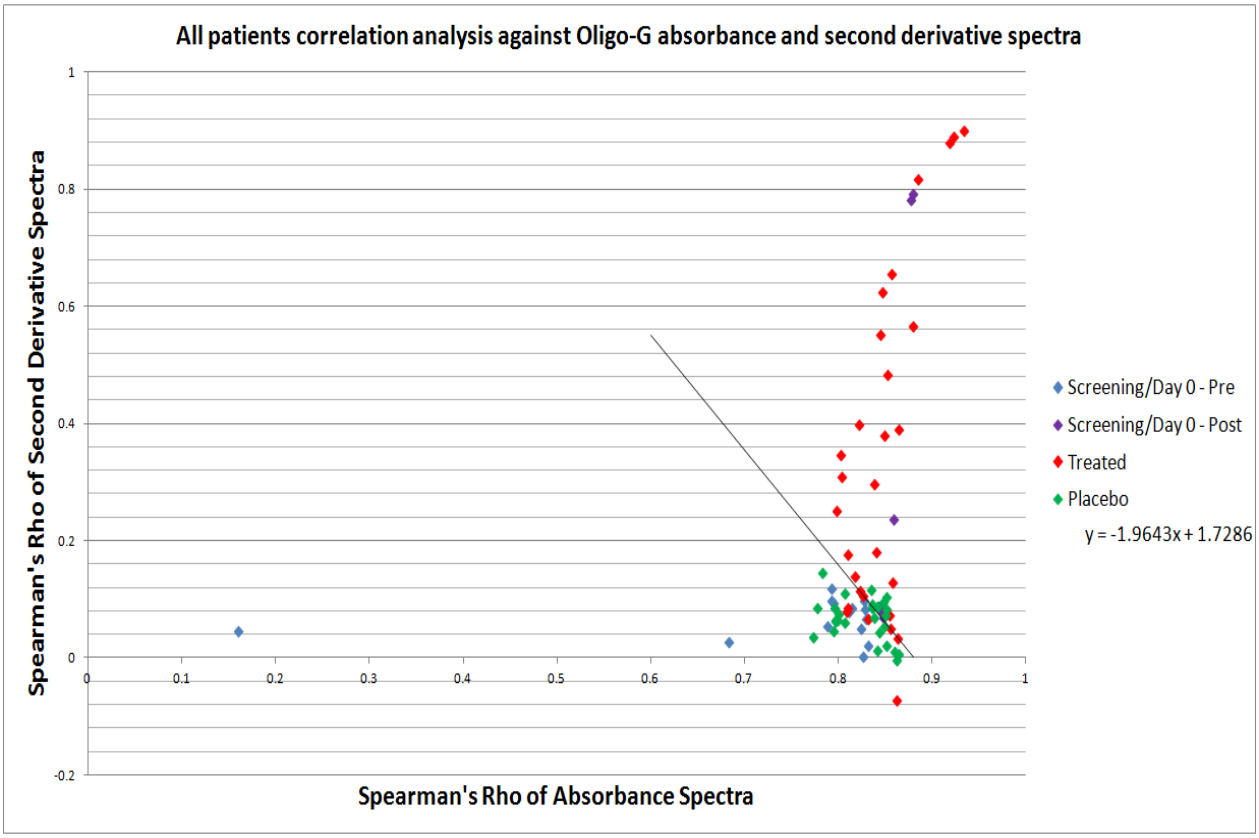


**Figure 6-21: Scatterplot plot of sputum-OligoG correlation coefficients of each CF patient sputum sample from each phase of the trial; placebo (green), OligoG**

**treatment (red), screening samples obtained pre-test OligoG dose (blue) and screening samples obtained post-test OligoG dose.**

| | Treated | Placebo | Screening/Day 0 | Placebo & Screening/Day 0 |
|---|---|---|---|---|
| **Total Samples** | 29 | 27 | 23 | 50 |
| **Correctly Predicted** | 25 | 22 | 23 | 45 |
| **Incorrectly Predicted** | 4 | 5 | 0 | 5 |
| **Sensitivity** | 0.86 | | | |
| **Specificity** | | 0.81 | 1.00 | 0.90 |

**Table 6-8: Sensitivity and specificity scores for prediction of OligoG presence in CF sputum, showing overall sensitivity and specificity scores of 86% and 90%, respectively.**

### 6.3.3 OligoG and Calcium Ion Interaction

OligoG is known to have calcium chelating properties (Ermund *et al.*, 2017), however the mode of action is unclear. This chapter sought to further characterise molecular interactions between OligoG and $Ca^{2+}$ using FTIR spectroscopy. OligoG at 2% w/v concentration was incubated with varying concentrations of $Ca^{2+}$ from omM to 5mM and FTIR analysis was carried out.

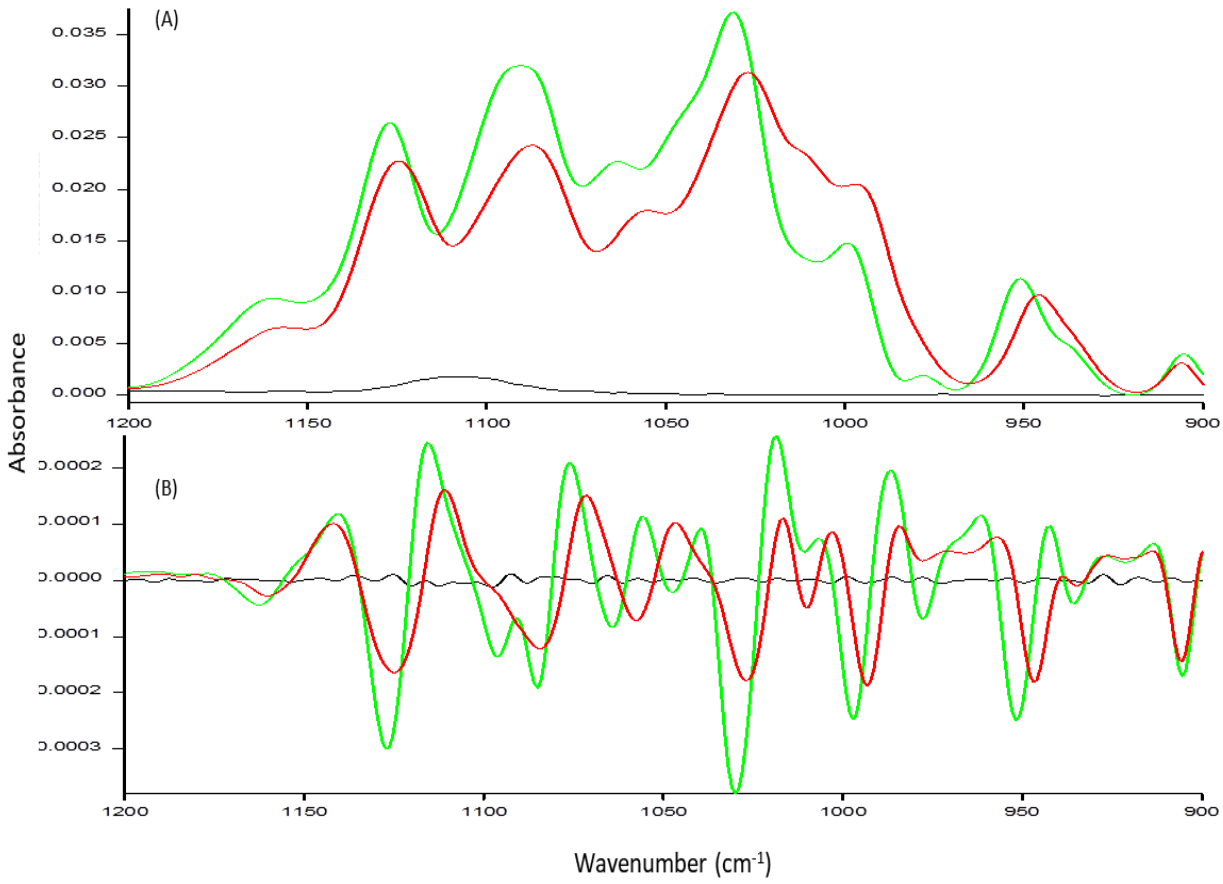


**Figure 6-22: (A) Absorbance and (B) second-derivative average spectra from 1200-900cm$^{-1}$ of 2% w/v OligoG (green), 2% w/v OligoG incubated with 5mM $Ca^{2+}$ (red), and 5mM $Ca^{2+}$ (black).**

The average IR-spectrum of OligoG incubated with 5mM $Ca^{2+}$ ions shows distinct blue shifts towards lower wavenumbers in the 1200-900cm$^{-1}$ region, compared to OligoG incubated with only $dH_2O$ (Figure 6-22). This is indicative of electrostatic interactions between the OligoG and $Ca^{2+}$ ions, causing the C-O and C-H bonds of the OligoG molecule to exist at a higher vibrational energy state. Marked peak shifts are seen from 951cm$^{-1}$ to 947cm$^{-1}$, and from 998cm$^{-1}$ to 990cm$^{-1}$ in $Ca^{2+}$-incubated OligoG. The large shift from 998cm$^{-1}$ to 900cm$^{-1}$ suggests interaction between the glycosidic linkage of OligoG and $Ca^{2+}$ ions. Another peak shift can be observed from 1031cm$^{-1}$ to 1025cm$^{-1}$, indicating that the $Ca^{2+}$ ions interact with the C-O bonds present in the pentose rings of the OligoG monosaccharide units. Notably, a small peak found around 1050cm$^{-1}$ in the non-incubated OligoG second-derivative average spectrum disappears altogether in the $Ca^{2+}$-incubated OligoG average second-derivative spectrum (Figure 6-22(b)). Glycosidic linkage bonds (CO-O-CO) are associated with IR-absorption from 1050-1040cm$^{-1}$. This could indicate that $Ca^{2+}$ chelation by OligoG

may influence the bending/stretching moments of the glycosidic bonds between OligoG monosaccharide units.

### 6.3.4 FTIR Microspectroscopy Imaging

FTIR microspectroscopy was carried out on pre-treatment screening CF sputum (AP114) and OligoG-treated CF sputum (AP148) samples produced by the same patient and 2D maps showing the intensity across specific spectral regions relating to biologically significant molecular bonds across the sputum were generated. The spectral regions investigated are the i) amide I peak (~1718-1650cm$^{-1}$), ii) glycogen-rich region (~1177-931cm$^{-1}$) and iii) an OligoG-specific peak (1029cm$^{-1}$). The peak at 1029cm$^{-1}$ was chosen for map generation as the second-derivative spectrum of OligoG shows the largest peak at 1029cm$^{-1}$ (Figure 6-3).

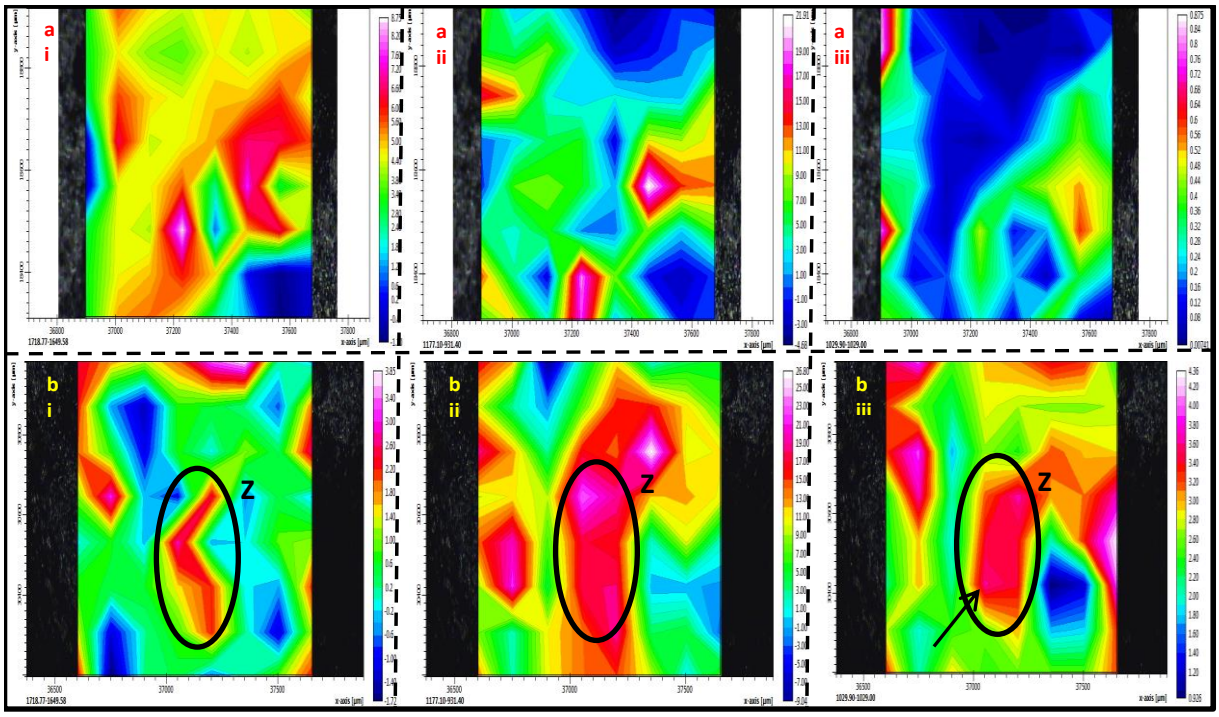


**Figure 6-23: FTIR microspectroscopy images of CF sputum from pre-treatment screening phase AP114 (a) and OligoG-treatment phase AP148 (b). The maps for each treatment phase are all from the same area of sputum and relate to the (i) amide I region, (ii) glycogen-rich region, and (iii) the area under 1029cm$^{-1}$. The region labelled 'Z' is highlighted as a potential hot spot for OligoG interaction with the mucus mucins. Taken together, the amide I region (b(i)) and glycogen-rich region (b(ii)) show a high concentration of mucin. This region also shows a high concentration of the OligoG-specific peak at 1029cm$^{-1}$. This same effect is not seen in the pre-treatment phase sputum, which show an overall low intensity of 1029cm$^{-1}$, despite higher levels of mucin across the sample.**

Area-under-the-curve (AUC) analysis within the amide I, glycogen-rich and OligoG-specific spectral regions highlighted how absorption of OligoG-specific peaks

is independent of absorption of mucin glycoproteins. This is demonstrated by the screening sample (Figure 6-23) showing a lower overall AUC value within the OligoG-specific region (Figure 6-23a(iii)) compared to the amide I (Figure 6-23a(i)) and glycogen-rich (Figure 6-23a(ii)) regions, whilst the OligoG-treatment phase sample (Figure 6-23b) shows AUC values for the OligoG-specific region (figure 6-19b(iii)) comparable to those observed in the amide I (Figure 6-23b(i)) and glycogen-rich regions (Figure 6-23b(ii)).

| Sample ID & Map | Spectral Region | Z min | Z max |
|---|---|---|---|
| **(a/i) Screening** | Amide I | -1.10 | 8.75 |
| **(a/ii) Screening** | Glycogen-rich | 4.68 | 21.91 |
| **(a/iii) Screening** | OligoG-specific 1029cm$^{-1}$ | 0.00741 | 0.875 |
| **(b/i) OligoG-treatment** | Amide I | -1.72 | 3.85 |
| **(b/ii) OligoG-treatment** | Glycogen-rich | 9.04 | 26.80 |
| **(b/iii) OligoG-treatment** | OligoG-specific 1029cm$^{-1}$ | 0.926 | 4.36 |

**Table 6-9: Range of spectral intensities for each spectral region, across each sample.**

## 6.4 Discussion

This chapter focussed on further evaluating FTIR spectroscopy as a tool for use in clinical settings; specifically in this case, the aim was to assess the capability of FTIR to detect the presence of a novel therapeutic drug, OligoG, in CF patients' sputum throughout the duration of a Phase IIb, placebo-controlled, crossover study.

CF patients require constant therapy and monitoring in order to appropriately manage their disease (Button *et al.*, 2016; Feiten *et al.*, 2016). There is currently no cure for CF, but the median age of survival has steadily improved in recent times (Cystic Fibrosis Foundation, 2016). This is mainly due to improvements in treating persistent airway infection and enhancing airway mucus clearance (Elborn, 2016). The alginate-based drug OligoG is thought to have promise as a novel therapy for CF patients, with multiple studies demonstrating its anti-microbial and mucolytic properties (Powell *et al.*, 2013; Pritchard *et al.*, 2016, 2017; Roberts *et al.*, 2013; Tøndervik *et al.*, 2014).

As part of on-going clinical studies into the efficacy of OligoG as an adjunct-therapy for CF, this thesis carried out FTIR analysis on a subset of sputum samples

from CF-patients who have received OligoG or placebo treatment over a 28 day period before crossing over into the next phase. This work aimed to assess whether it is possible to use FTIR technology to detect OligoG within CF patient sputum samples, following a simple protocol and involving minimal sample preparation. The overall goal of this work was to develop a protocol centred around FTIR spectroscopy to be used for drug screening in biofluids which could provide clinicians and healthcare providers with valuable information pertaining to drug uptake/adherence to therapy, and persistence of the drug within specific biofluids which could have direct implications on the required dose.

The results show that FTIR is a viable tool for detecting the presence/absence of OligoG within raw sputum, with minimal statistical and spectral analysis. Using multiple analyses this work was able to distinguish pre-treatment phase CF sputum from OligoG-treatment phase CF sputum on a patient-by-patient basis. The strongest technique explored in this study is the calculation of Spearman's Rho correlation coefficient of second-derivative and absorbance sputum spectra to second-derivative and absorbance OligoG spectra respectively. Dosing experiments showed that OligoG could be identified in sputum by statistically significant increases in absorbance at all key wavenumbers at concentrations of 0.1% (w/v) and higher (Table 6-7).

All patients recruited to the trial received an initial test-dose of OligoG at their screening visit. This test-dose was either given prior to, or after production of the patient's sputum sample. Review of the patient data made clear which patients had received a dose of OligoG prior to providing their screening sputum sample. This allowed the classification of screening samples based on the time of sputum production and OligoG test dosing. Subsequent calculation of overall specificity scores for correctly identifying OligoG in sputum was shown to be 90% (Table 6-8).

In all patients a relationship was identified indicating an increase in correlation between sputum and OligoG spectra after treatment. This suggests that OligoG at the therapeutic concentration of 0.2% changes the molecular environment of sputum significantly enough to be detected by FTIR. The minimum effective dose for OligoG to cause biofilm disruption is 2% (w/v) (Powell *et al.*, 2018), therefore an FTIR-based protocol could readily detect therapeutic concentrations of OligoG in patient sputum samples.

Data from FTIR microspectroscopy suggest that OligoG is primarily interacting with the glycan chains present on the mucin glycoprotein backbone because areas of higher concentration of OligoG correlate closely with areas of higher mucin concentration (Figure 6-23).

## 6.5 Conclusions

OligoG has been shown to be an exciting and novel potential therapy for CF patients. Studies such as this are important for establishing how long OligoG persists in respiratory mucus, but are also for showing how FTIR can be used to enhance the management of chronic disease. Using FTIR to monitor CF patients could allow clinicians and healthcare providers to tailor treatment with OligoG based on patient's specific needs. Patient adherence to CF therapy can be low, but electronic monitoring of adherence in CF patients has been suggested to be a driver to enhance treatment adherence (Narayanan, Mainz, Gala, Tabori, & Grossoehme, 2017). A simple protocol based on FTIR for detection of OligoG in sputum could help to ensure that patients are taking the optimum dosage of OligoG.

# Chapter 7

# General Discussion, Final Conclusions and Future Work

## 7.1 General Discussion

Chronic respiratory diseases, such as COPD, lung cancer and CF, are a major burden on healthcare systems in the UK and across the world, and a significant cause of patient mortality yearly. Patients typically have a poorer QOL and reduced life expectancy, compared to their healthy peers. The diseases discussed in this thesis are all characterised by damage to the respiratory system which is associated with reduced lung function, directly impacting on the patient's QOL (Rose & Voynow, 2006). Although the particular form of damage caused by each specific disease is unique, and requires bespoke treatment, these diseases do share certain aetiologies, such as a chronic inflammatory response within the airways, leading to increased mucin production and secretion.

Through enhanced management and/or screening of such diseases, it would be possible to see improvements in patient care and subsequent reductions in disease-related patient mortality. Indeed, much of the current advice and guidelines for COPD, CF and lung cancer focuses on enhanced monitoring and screening of patients to improve the efficiency of healthcare provision and understand disease progression more fully (Cystic Fibrosis Foundation, 2016; Global Initiative for Chronic Obstructive Lung Disease, 2018; Latimer & Mott, 2015; Stewart & Wild, 2014). The FTIR spectrometer is simple & inexpensive to use, IR spectra of sputum are non-invasive and rapid to generate. A vast amount of molecular structural information about the patient's sputum sample can be generated in a short timeframe. The work contained within this thesis clearly demonstrates the utility of FTIR in respiratory disease management through raw sputum analysis. Indeed, this thesis adds weight to the mounting body of evidence that applications involving FTIR for diagnosing diseases from biofluids could be viable for implementation into clinical pathways (Baker *et al.*, 2015; Bird *et al.*, 2012; Hands *et al.*, 2016; P. D. Lewis *et al.*, 2010; Menzies *et al.*, 2014; Rodrigues *et al.*, 2017; Smith *et al.*, 2016).

The workflow for generating IR spectra of a sputum sample is very simple, involving little-to-no sample pre-processing or preparation, and could be readily implemented within a pathology laboratory. Once a spectrum is generated it must first be analysed, before any conclusions can be made about the patient's current state-of-health. However, analysis of raw biological sample-spectra is complex, due to the presence of overlapping bands within the spectrum which may obscure peaks of

interest, and spectral artefacts which can arise from inappropriate sampling procedures (Baker *et al.*, 2015). Therefore it would be inadvisable to introduce FTIR spectroscopy into the clinic without first implementing a series of modifications to the spectrometer and controlling software in order to make the devices more suitable for the clinical environment, whether that is primary or secondary care focussed. Discussions must be held with primary and secondary care clinicians, patients, clinical research teams and pathology laboratories to identify the particular needs of each potential user-base for an FTIR-based screening/diagnostic technology. During such discussions, it would be important to factor-in the costs associated with training laboratory staff to use FTIR screening against the costs of training for other methodologies.

### 7.1.1 FTIR for Detection and Prediction of COPD Exacerbation

FTIR spectroscopy on COPD sputum samples to detect exacerbation and distinguish it from baseline disease proved to be problematic. The distributions of FTIR spectrum peaks and absorbencies were found to be highly similar for both baseline and exacerbating COPD patients. Two peak positions were found to be statistically significantly different between the disease states, but this was deemed unreliable due to the relatively low number of patient spectra in which both peaks are always detectable. It was thought that one of the main confounding factors was the uncertainty whether the patient had recently, or was soon to have an exacerbation, which could influence the biophysical properties of the sputum. Such ambiguity was thought to be the major reason many COPD baseline and exacerbation spectra shared numerous spectral characteristics.

Prediction of exacerbation through time-series analysis of COPD patient sputum sought to account for the uncertainty based around exacerbation timelines. A predictive algorithm based on inflammation-associated wavenumbers was generated and tested on the dataset. The algorithm was able to detect 60% of all individual exacerbation episodes in the study, but a very high number of false positive predictions were also made, limiting the effectiveness of the algorithm.

### 7.1.2 Lung Cancer Diagnosis by FTIR Spectroscopy

FTIR spectroscopy was carried out on sputum samples collected from lung cancer patients in order to create diagnostic algorithms which could be used to detect lung cancer. The work presented in this thesis showed how lung cancer has a distinct spectral signature, with little variation observed within the lung cancer cohort. A

series of predictive linear regression models was developed, capable of distinguishing lung cancer from non-cancer respiratory disease with overall sensitivity and specificity scores greater than 80%. The best performing single models showed sensitivity scores of 100% with >85% specificity.

Due to the low running costs and simplicity of the protocol used in this thesis, the work represents a major potential for FTIR to be used in the lung cancer diagnostics pathway as a potential screening tool for early lung cancer detection. FTIR spectrometers could be in place in primary care scenarios, for use by a clinician as a tool to reduce the uncertainty around potential lung cancer diagnosis when a patient arrives in primary care displaying symptoms which could suggest carcinoma presence. It could be possible that such a tool could reduce the numbers of patients being referred to X-ray, CT-scans or bronchoscopies, by rapid identification of patients who can be said, within a degree of certainty, to not currently have lung cancer, or a pre-malignancy. Further work would need to be carried out to fully characterise the ability of an FTIR-based protocol to detect early-stage lung cancer in raw sputum and to state with an extremely high degree of certainty that any given patient does not currently have lung cancer.

### 7.1.3 OligoG Monitoring in CF Sputum by FTIR

OligoG is an exciting novel therapeutic for CF patient treatment and management. It has antibacterial, antifungal and mucolytic properties, all of which work to alleviate the symptom burden on CF patients. However, the mode of action of OligoG in the CF lung was uncertain (Ermund *et al.*, 2017; Hengzhuang *et al.*, 2016; Pritchard *et al.*, 2017). It was unclear if OligoG was detectable in airway mucus after treatment is, so an FTIR protocol was developed to quantify the levels of OligoG present in the lung through comparisons with an OligoG reference spectrum. This method proved to be highly powerful, capable of sensitively detecting concentrations of OligoG in sputum as low as 0.05% (w/v), well below the minimum effective dosage of 2% (w/v) (Powell *et al.*, 2018). This represents an important finding for clinicians of CF patients; through rapid FTIR analysis an estimation of the current OligoG concentrations in the airways can be calculated, helping clinicians determine a CF patient's adherence to the therapy.

## 7.2 Spectral Biobank Libraries

Due to the low-cost, non-destructive and rapid nature of FTIR spectroscopy, performing IR spectroscopy on patient biofluids and tissues presents the opportunity for the creation of large spectral libraries. These libraries would contain IR spectra of patient samples collected from a multitude of diseases, and sample sources. An unknown sample could then be compared against the spectral biobank, allowing for qualitative identification of the sample. This principle could be extended from diagnosis of respiratory disease through sputum analysis, to detection and qualitative identification of bacterial and viral pathogens in sputum. Indeed, studies have shown that FTIR chemometric analysis can differentiate bacteria at the genus, species and clonal levels (Grewal, Jaiswal, & Jha, 2015; Zarnowiec, Lechowicz, Czerwonka, & Kaca, 2015).

Characterisation of individual absorbance bands in IR spectra of biofluids is highly complex, due to the varied chemical groups present within the sample (Baker *et al.*, 2015). Through collection of samples from thousands of patients, it could be possible to identify and account for the variation within patient subgroups that contributes to spectral variations. For example, the medication available to COPD patients is highly varied and is tailored to the particular patient's specific needs (Global Initiative for Chronic Obstructive Lung Disease, 2018). The use of spectral biobank libraries of COPD patients coupled with patient phenotype data, could eventually allow COPD patients to be rapidly phenotyped based on the biochemistry of their sputum.

Implementation of FTIR into the clinical setting would allow access to an extremely large set of patients of varying diseases and sample sources. Using cloud-based technology, IR spectra of patient biofluids could be uploaded to a central spectrum repository, into which the patient's clinical information is also loaded. Machine-learning and deep-learning algorithms could identify specific patterns within the patient groups that could be indicative of disease subtypes. An example of such classification in use could eventually be preliminary identification of lung cancer subtypes by FTIR spectroscopy. Spectral markers specific to early-stage carcinoma or specific histologies may be identifiable through use of large spectral banks. This application could help direct the diagnostic pathway for lung cancer most efficiently towards the final clinical diagnosis.

With respect to COPD, in this thesis it was found that some associations could be made between FTIR spectra and certain physiological variables such as FEV1 (%) predicted, smoking pack-years, and exposure to pollutants. Through collection of detailed COPD patient information and sputum samples, these associations could be used to build classification models to classify COPD patients into phenotypes based on those described in Mirza & Benzo (2017). Use of spectral libraries would allow the collection of data from many thousands of patients, which are required by machine-learning and deep-learning tools for optimal accuracy.

## 7.3 Final Conclusions

COPD exacerbation is a complex disease, influenced by many confounding variables, not least the patient themselves. FTIR analysis of COPD patient raw sputum was unable to develop a generalised, predictive model for COPD exacerbation. Associations between FTIR spectral features and physiological characteristics were made, suggesting that sputum does undergo molecular structural change which could be related to the patient's current state of health, but further work would be required to fully identify a link between COPD exacerbation status and FTIR spectra of raw sputum.

FTIR analysis of raw sputum samples has clearly shown that lung cancer patients can be readily distinguished from non-cancer-chronic-respiratory disease patients. The current work has built on an already solid foundation of evidence (P. D. Lewis *et al.*, 2010) suggesting that lung cancer patients are spectrally distinct from healthy controls. This thesis has expanded this work by increasing the numbers of patients analysed and distinguishing lung cancer patients from a clinically relevant group of respiratory disease non-cancer controls with up to 100% sensitivity and >85% specificity. Through this work, FTIR has been demonstrated to be a highly-effective tool for detecting lung cancer, using a simple protocol which can be readily translated to the clinic.

This work has also indicated how it is possible to utilise FTIR as a tool for monitoring drug retention in CF patients. OligoG has been previously shown to be an effective potential therapy for CF patients, but the mode of action and clearance of OligoG within the lung was unclear, leading to ambiguity around the maximum safe long-term therapeutic dosage. By showing that FTIR can sensitively and specifically identify OligoG presence in sputum with no sample pre-processing and a simple

algorithm, this thesis has highlighted the potential of FTIR to be used as a long-term screening tool for detecting OligoG in raw sputum with applications in CF patient management, notably in monitoring patient adherence to therapy.

## 7.4 Future Work

The work presented in this thesis is limited by the multifactorial nature of COPD, which is characterised by the many observed phenotypes described in the literature. These phenotypes are based on the physical and mental characteristics of the patients, such as lung function, frequency of exacerbation, mental and physical frailty, and presence of comorbidities (Mirza & Benzo, 2017). Future studies should focus on COPD's multifactorial nature, by recruiting a large cohort of patients to account for the inter-patient variability, and allow the production of generalised predictive models for exacerbation in each COPD phenotype.

Interest in the application of FTIR spectroscopic analysis to CF sputum samples has been raised. One question which has commonly arisen during this thesis is: "Could FTIR spectroscopy be used to detect CF exacerbation?" Pulmonary exacerbations of CF are strongly associated with mortality and decreased QOL, and the numbers of CF pulmonary exacerbations has not significantly decreased in recent years, despite overall improvement in CF patient health and pulmonary function (Cystic Fibrosis Foundation, 2016). The CF patient sputum samples used in this work were collected only once every two weeks. The work carried out in Chapter 3 in the SPEDIC study collected five samples per week, and only minimal evidence for a reliable exacerbation prediction protocol was gathered. In order to fully investigate the potential of FTIR spectroscopic analysis of raw sputum from CF patients to detect pulmonary exacerbation prior to the commencement of symptoms, a large longitudinal study collecting many more samples is required. Such a study would also have to tightly control the CF patients recruited to the study, to ensure variability between patients is minimal. An example of required control may be to ensure that all CF patients have the same CFTR mutation.

To ensure the efficacy of FTIR for lung cancer detection, a large, blinded, multi-centre study must be carried out, comparing the predictive power of FTIR to the clinical pathway. This should focus on the capability of FTIR spectroscopy for early-stage lung cancer detection, when chances of survival are greatest (McPhail *et al.*, 2015).

Miniaturised IR spectroscopy should continue to be developed towards eventual use in the clinical environment, either by clinicians in primary, secondary or tertiary care scenarios, or by patients in their own homes. A major barrier to entry however is the complexity of benchtop FTIR spectrometers. They require trained operators to apply samples and interpret the resultant data. For an IR spectrometer to be acceptable for use by a clinician or patient, a simplified user interface for spectrometer control is required. The user would place a sample on a sampling substrate and a clinical output pertaining to the patient's current state of health is given through the use of computer programs.

Another barrier is the cost of conventional FTIR spectrometers, which can be many thousands of pounds. A significant portion of this cost can be the ATR crystal. Silicon is highly transparent for wavenumbers below 10,000$cm^{-1}$, low-cost and is readily manufactured into thin wafers (Wolffenbuttel, 2005). It is therefore an excellent candidate as a substrate for applying sputum samples. Recent advances within this field have shown that silicon wafers are a viable substrate for use as ATR crystals. It is possible to manufacture these wafers at a lower cost than current ATR crystals, and a drop in performance in terms of spectrum acquisition when compared to a single-bounce diamond ATR crystal is not observed. Future developments of miniaturised or home- and/or clinic-based spectrometers could be feasibly based on the use of such crystals.

The low cost of these silicon ATR wafers raises another possible application for FTIR spectroscopy of sputum. Earlier in this thesis, the viability of short-term storage of sputum at low, or room temperatures was addressed. A review of the literature found that mucus-based samples are stable for a period of time, with no effect on the quantitative yield of bacterial colonies or mucin stability after short-term refrigeration (Gastaldi *et al.*, 2000; Kolwijck *et al.*, 2013; Panicker *et al.*, 2007). Taken together these two factors raise the possibility of supplying patients of chronic respiratory disease, for example CF patients, or patients who have been identified as higher-risk for developing lung cancer, with specialised sputum collectors and applicators which could apply the sputum to the ATR wafer. These could then be transported to a laboratory for analysis, with the results being used to inform the patient's care.

The efficacy with which a simple FTIR protocol and analytical procedure has been shown to detect the alginate-based CF therapy OligoG suggests that other inhaled therapies could be detected in raw sputum by a similar protocol. Adherence of CF patients to therapies is known to be a major influencing factor for the patient's QOL (Narayanan *et al.*, 2017). Future studies could be employed to develop procedures and algorithms for detection of other inhaled therapies in sputum, such as Symdeko or Orkambi, which could in turn help inform management of CF patients prescribed with these therapies.

To finally summarise, FTIR spectroscopy has been shown to have great potential in the diagnosis and on-going management of three serious respiratory diseases. Rapid diagnosis is known to improve prognosis and reduce mortality: should this technology be developed to its fullest, it has the potential to contribute to improved survival-rates and reduced costs. It is an exciting area of study which promises to increase accurate detection, and help manage the symptoms, of potentially life-limiting conditions.

## 8.0 References

Allemani, C., Weir, H. K., Carreira, H., Harewood, R., Spika, D., Wang, X.-S., … Group, C. W. (2015). Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet*, *385*(9972), 977–1010. https://doi.org/10.1016/S0140-6736(14)62038-9.

Ambort, D., Johansson, M. E. V, Gustafsson, J. K., Ermund, A., & Hansson, G. C. (2012). Perspectives on mucus properties and formation-lessons from the biochemical world. *Cold Spring Harbor Perspectives in Medicine*, *2*(11), 1–10. https://doi.org/10.1101/cshperspect.a014159

Andolfi, M., Potenza, R., Capozzi, R., Liparulo, V., Puma, F., & Yasufuku, K. (2016). The role of bronchoscopy in the diagnosis of early lung cancer: A review. *Journal of Thoracic Disease*, *8*(11), 3329–3337. https://doi.org/10.21037/jtd.2016.11.81

Audie, J. P., Janin, A., Porchet, N., Copin, M.-C., Gosselin, B., & Aubert, J.-P. (1993). Human Mucin Genes in Respiratory, Digestive, and Reproductive Tracts Ascertained by In Situ Hybridization. *The Journal of Histochemistry and Cytochemistry*, *41*(10), 1479–1481.

Aviel-Ronen, S., Blackhall, F. H., Shepherd, F. A., & Tsao, M. S. (2006). K-ras mutations in non-small-cell lung carcinoma: A review. *Clinical Lung Cancer*, *8*(1), 30–38. https://doi.org/10.3816/CLC.2006.n.030

Backhaus, J., Mueller, R., Formanski, N., Szlama, N., Meerpohl, H.-G., Eidt, M., & Bugert, P. (2010). Diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vibrational Spectroscopy*, *52*(2), 173–177.

Baker, M. J., Hussain, S. R., Lovergne, L., Untereiner, V., Hughes, C., Lukaszewski, R. A., … Sockalingum, G. D. (2015). Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.*, *45*(7), 1803–1818. https://doi.org/10.1039/C5CS00585J

Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., & Butler, H. J. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nat Protoc*, *9*(8), 1771–1791. https://doi.org/10.1038/nprot.2014.110.

Bansil, R., Stanley, E., & LaMont, J. T. (1995). Mucin biophysics. *Annual Review of Physiology*, *57*, 635–657. https://doi.org/10.1146/annurev.ph.57.030195.003223

Bansil, R., & Turner, B. S. (2006). Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid & Interface Science*, *11*(2–3), 164–170. https://doi.org/10.1016/j.cocis.2005.11.001

Barth, A. (2007). Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta - Bioenergetics*, *1767*(9), 1073–1101. https://doi.org/10.1016/j.bbabio.2007.06.004

Bautista, M. V., Chen, Y., Ivanova, V. S., Rahimi, M. K., Watson, A. M., & Rose, M. C. (2009). IL-8 regulates mucin gene expression at the post-transcriptional level in lung epithelial cells. *Journal of Immunology (Baltimore, Md. : 1950)*, *183*(3), 2159–2166. https://doi.org/10.4049/jimmunol.0803022

Bentini, G. G., Bianconi, M., Cerutti, A., Nubile, A., Cerabolini, P., Chiarini, M., … Guzzi, R. (2006). A new miniaturised optical system for chemical species spectroscopic detection based on a scanning integrated Mach-Zehnder microinterferometer on LiNbO3. *Origins of Life and Evolution of the Biosphere*, *36*(5–6), 597–603. https://doi.org/10.1007/s11084-006-9052-3

Bird, B., Miljković, M., Remiszewski, S., Akalin, A., Kon, M., & Diem, M. (2012). Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer. *Laboratory Investigation*, *92*(9), 1358–1373. https://doi.org/10.1038/labinvest.2012.101

Blanc, P. D., Iribarren, C., Trupin, L., Earnest, G., Katz, P. P., Balmes, J., … Eisner, M. D. (2009). Occupational exposures and the risk of COPD: dusty trades revisited. *Thorax*, *64*(1), 6–12. https://doi.org/10.1136/thx.2008.099390

Borowitz, D. (2015). CFTR, bicarbonate, and the pathophysiology of cystic fibrosis. *Pediatric Pulmonology*, *50*(April), S24–S30. https://doi.org/10.1002/ppul.23247

Bozinovski, S., Anthony, D., & Vlahos, R. (2014). Targeting pro-resolution pathways to combat chronic inflammation in COPD. *Journal of Thoracic Disease*, *6*(11), 1548–1556. https://doi.org/10.3978/j.issn.2072-1439.2014.08.08

Brandt, T., Breitenstein, S., von der Hardt, H., & Tümmler, B. (1995). DNA concentration and length in sputum of patients with cystic fibrosis during inhalation with recombinant human DNase. *Thorax*, *50*(8), 880–882. https://doi.org/10.1136/thx.50.8.880

Brøgger, J., Steen, V. M., Eiken, H. G., Gulsvik, A., & Bakke, P. (2006). Genetic association between COPD and polymorphisms in TNF, ADRB2 and EPHX1. *The European Respiratory Journal*, *27*(4), 682–688. https://doi.org/10.1183/09031936.06.00057005

Buisine, M.-P., Devisme, L., Copin, M.-C., Durand-réville, M., Gosselin, B., Aubert, J.-P., & Porchet, N. (1999). Developmental Mucin Gene Expression in the Human Respiratory Tract. *American Journal of Respiratory Cell and Molecular Biology*, *20*(13), 209–218.

Burgel, P. R., Bellis, G., Olesen, H. V., Viviani, L., Zolin, A., Blasi, F., … Solé, A. (2015). Future trends in cystic fibrosis demography in 34 European countries. *European Respiratory Journal*, *46*(1), 133–141. https://doi.org/10.1183/09031936.00196314

Button, B. M., Wilson, C., Dentice, R., Cox, N. S., Middleton, A., Tannenbaum, E., … Holland, A. E. (2016). Physiotherapy for cystic fibrosis in Australia and New Zealand: A clinical practice guideline. *Respirology*, *21*(4), 656–667. https://doi.org/10.1111/resp.12764

Cancer Research UK. (2017). *Lung Cancer Incidence Statistics*. Retrieved from http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence#heading-Three. Date retrieved 03/03/18.

Cancer Research UK. (2018). Cancer Incidence for Common Cancers.

Caramori, G., Casolari, P., Di Gregorio, C., Saetta, M., Baraldo, S., Boschetto, P., … Papi, A. (2009). MUC5AC expression is increased in bronchial submucosal

glands of stable COPD patients. *Histopathology*, *55*(3), 321–331. https://doi.org/10.1111/j.1365-2559.2009.03377.x

Chang, Y., Al-Alwan, L., Alshakfa, S., Audusseau, S., Mogas, A., Chouiali, F., … Eidelman, D. H. (2014). Upregulation of IL-17A/F from human lung tissue explants with cigarette smoke exposure: implications for COPD. *Respiratory Research*, *15*(1), 145. https://doi.org/10.1186/s12931-014-0145-7

Chen, X., Xu, X., & Xiao, F. (2013). Heterogeneity of chronic obstructive pulmonary disease: from phenotype to genotype. *Frontiers of Medicine*, *7*(4), 425–432. https://doi.org/10.1007/s11684-013-0295-x

Chen, Y., Thai, P., Zhao, Y.-H., Ho, Y.-S., DeSouza, M. M., & Wu, R. (2003). Stimulation of airway mucin gene expression by interleukin (IL)-17 through IL-6 paracrine/autocrine loop. *The Journal of Biological Chemistry*, *278*(19), 17036–17043. https://doi.org/10.1074/jbc.M210429200

Chillappagari, S., Preuss, J., Licht, S., Müller, C., Mahavadi, P., Sarode, G., … Henke, M. O. (2015). Altered protease and antiprotease balance during a COPD exacerbation contributes to mucus obstruction. *Respiratory Research*, *16*(1), 85. https://doi.org/10.1186/s12931-015-0247-x

Chin, C. L., Manzel, L. J., Lehman, E. E., Humlicek, A. L., Shi, L., Starner, T. D., … Look, D. C. (2005). *Haemophilus influenzae* from Patients with Chronic Obstructive Pulmonary Disease Exacerbation Induce More Inflammation than Colonizers. *American Journal of Respiratory and Critical Care Medicine*, *172*, 85–91. https://doi.org/10.1164/rccm.200412-1687OC

Choi, H. J., Chung, Y.-S., Kim, H. J., Moon, U. Y., Choi, Y. H., van Seuningen, I., … Yoon, J.-H. (2009). Signal pathway of 17beta-estradiol-induced MUC5B expression in human airway epithelial cells. *American Journal of Respiratory Cell and Molecular Biology*, *40*, 168–178. https://doi.org/10.1165/rcmb.2007-0377OC

Couraud, S., Zalcman, G., Milleron, B., Morin, F., & Souquet, P. J. (2012). Lung cancer in never smokers - A review. *European Journal of Cancer*, *48*(9), 1299–1311. https://doi.org/10.1016/j.ejca.2012.03.007

Crystal, R. G., Randell, S. H., Engelhardt, J. F., Voynow, J., & Sunday, M. E. (2008). Airway Epithelial Cells: Current Concepts and Challenges. *Proceedings of the American Thoracic Society*, *5*(7), 772–777. https://doi.org/10.1513/pats.200805-041HR

Cui, L., Butler, H. J., Martin-hirsch, P. L., & Martin, F. L. (2015). Analytical Methods Aluminium foil as a potential substrate for ATR- FTIR , transflection FTIR or Raman spectrochemical analysis of biological specimens. *Analytical Methods*, *00*(3), 1–7. https://doi.org/10.1039/C5AY02638E

Cystic Fibrosis Foundation. (2010). *Cystic Fibrosis Foundation Patient Registry, 2010 Annual Data Report*. *Cystic Fibrosis Foundation Publications*.

Cystic Fibrosis Foundation. (2015). *Cystic Fibrosis Foundation Patient Registry* 2014 Annual Data Report, 88. *Cystic Fibrosis Foundation Publications*.

Cystic Fibrosis Foundation. (2016). *Cystic Fibrosis Foundation patient registry. 2016 annual data report*. *Cystic Fibrosis Foundation Publications*.

https://doi.org/10.1039/b815899a

Davies, J. C., Moskowitzm, S. M., Brown, C., Horsley, A., Mall, M. A., Mckone, E. F., … Rowe, S. M. (2018). VX-445–Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *New England Journal of Medicine*, *379*(17), 1612–1620. https://doi.org/10.1056/NEJMoa1807120

Davril, M., Degroote, S., Humbert, P., Galabert, C., Dumur, V., Lafitte, J.-J., … Roussel, P. (1999). The sialylation of bronchial mucins secreted by patients suffering from cystic fibrosis or from chronic bronchitis is related to the severity of airway infection. *Glycobiology*, *9*(3), 311–321. https://doi.org/10.1093/glycob/9.3.311

De Boeck, K., & Amaral, M. D. (2016). Progress in therapies for cystic fibrosis. *The Lancet Respiratory Medicine*, *4*(8), 662–674. https://doi.org/10.1016/S2213-2600(16)00023-0

De Boeck, K., Zolin, A., Cuppens, H., Olesen, H. V., & Viviani, L. (2014a). The relative frequency of CFTR mutation classes in European patients with cystic fibrosis. *Journal of Cystic Fibrosis*, *13*(4), 403–409. https://doi.org/10.1016/j.jcf.2013.12.003

De Boeck, K., Zolin, A., Cuppens, H., Olesen, H. V., & Viviani, L. (2014b). The relative frequency of CFTR mutation classes in European patients with cystic fibrosis. *Journal of Cystic Fibrosis*, *13*(4), 403–409. https://doi.org/10.1016/j.jcf.2013.12.003

Degroote, S., Lo-Guidice, J.-M., Strecker, G., Ducourouble, M.-P., Roussel, P., & Lamblin, G. (1997). Characterization of an N-Acetylglucosamine-6-O-sulfotransferase from Human Respiratory Mucosa Active on Mucin Carbohydrate Chains. *Journal of Biological Chemistry*, *272*(47), 29493–29501. https://doi.org/10.1074/jbc.272.47.29493

Degroote, S., Maes, E., Humbert, P., Delmotte, P., Lamblin, G., & Roussel, P. (2003). Sulfated oligosaccharides isolated from the respiratory mucins of a secretor patient suffering from chronic bronchitis. *Biochimie*, *85*(3–4), 369–379. https://doi.org/10.1016/S0300-9084(03)00022-1

Delmotte, P., Degroote, S., Lafitte, J.-J., Lamblin, G., Perini, J.-M., & Roussel, P. (2002). Tumor necrosis factor alpha increases the expression of glycosyltransferases and sulfotransferases responsible for the biosynthesis of sialylated and/or sulfated Lewis x epitopes in the human bronchial mucosa. *The Journal of Biological Chemistry*, *277*(1), 424–431. https://doi.org/10.1074/jbc.M109958200

Desseyn, J.-L., Aubert, J.-P., Porchet, N., & Laine, A. (2000). Evolution of the large secreted gel-forming mucins. *Molecular Biology and Evolution*, *17*, 1175–1184.

Dhar, P., Ng, G. Z., Dunne, E. M., & Sutton, P. (2017). Mucin 1 protects against severe *Streptococcus pneumoniae* infection. *Virulence*, *0*(0), 1–12. https://doi.org/10.1080/21505594.2017.1341021

Dittrich, A. S., Kuhbandner, I., Gehrig, S., Rickert-Zacharias, V., Twigg, M., Wege, S., … Mall, M. A. (2018). Neutrophil elastase activity on the surface of sputum neutrophils is associated with severity of cystic fibrosis lung disease. *European Respiratory Journal*, *51*(3). https://doi.org/10.1183/13993003.01910-2017

Doll, R., & Hill, A. B. (1999). Smoking and carcinoma of the lung: Preliminary report. *Bulletin of the World Health Organization*, *77*(1), 84–93.

https://doi.org/10.1136/bmj.2.4682.739

Donlan, R. M. (2001). Biofilm formation: a clinically relevant microbiological process. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *33*(8), 1387–1392. https://doi.org/10.1086/322972

Easton, R. (2011). Glycosylation of Proteins - Structure, Function and Analysis. *Life Science - Technical Bulletin*.

Eisner, M. D., Anthonisen, N., Coultas, D., Kuenzli, N., Perez-Padilla, R., Postma, D. S., … Balmes, J. R. (2010). An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *182*(5), 693–718. https://doi.org/10.1164/rccm.200811-1757ST

Elborn, J. S. (2016). Cystic fibrosis. *The Lancet*, *388*(10059), 2519–2531. https://doi.org/10.1016/S0140-6736(16)00576-6

Emadi, A., Wu, H., De Graaf, G., & Wolffenbuttel, R. F. (2011). IR microspectrometers based on linear-variable optical filters. *Procedia Engineering*, *25*, 1401–1404. https://doi.org/10.1016/j.proeng.2011.12.346

Endo, C., Nakashima, R., Taguchi, A., Yahata, K., Kawahara, E., Shimagaki, N., … Sato, M. (2015). Inter-Rater Agreement of Sputum Cytology for Lung Cancer Screening in Japan. *Diagnostic Cytopathology*, *43*(7), 545–550. https://doi.org/https://doi.org/10.1002/dc.23253

Enomoto, Y., Takada, K., Hagiwara, E., & Kojima, E. (2013). Distinct features of distant metastasis and lymph node stage in lung adenocarcinoma patients with epidermal growth factor receptor gene mutations. *Respiratory Investigation*, *51*(3), 153–157. https://doi.org/10.1016/j.resinv.2013.02.004

Ermund, A., Recktenwald, C. V., Skjåk-Bræk, G., Meiss, L. N., Onsøyen, E., Rye, P. D., … Hansson, G. C. (2017). OligoG CF-5/20 normalizes cystic fibrosis mucus by chelating calcium. *Clinical and Experimental Pharmacology and Physiology*, *44*(6), 639–647. https://doi.org/10.1111/1440-1681.12744

Feiten, T. D. S., Flores, J. S., Farias, B. L., Rovedder, P. M. E., Camargo, E. G., Dalcin, P. de T. R., & Ziegler, B. (2016). Respiratory therapy: a problem among children and adolescents with cystic fibrosis. *Jornal Brasileiro de Pneumologia*, *42*(1), 29–34. https://doi.org/10.1590/S1806-37562016000000068

Forey, B. A., Thornton, A. J., & Lee, P. N. (2011). Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. *BMC Pulmonary Medicine*, *11*(1), 36. https://doi.org/10.1186/1471-2466-11-36

Fossiez, F., Djossou, O., Chomarat, P., Flores-Romo, L., Ait-Yahia, S., Maat, C., … Lebecque, S. (1996). T cell interleukin-17 induces stromal cells to produce proinflammatory and hematopoietic cytokines. *Journal of Experimental Medicine*, *183*(6), 2593–2603. https://doi.org/10.1084/jem.183.6.2593

Fredericks, J. D., Bennett, P., Williams, A., & Rogers, K. D. (2012). FTIR spectroscopy: A new diagnostic tool to aid DNA analysis from heated bone. *Forensic Science International: Genetics*, *6*(3), 375–380. https://doi.org/10.1016/j.fsigen.2011.07.014

Gao, Y.-H., Guan, W.-J., Liu, Q., Wang, H.-Q., Zhu, Y.-N., Chen, R.-C., & Zhang, G.-J. (2016). Impact of COPD and emphysema on survival of patients with lung cancer: A meta-analysis of observational studies. *Respirology*, *21*(2), 269–279. https://doi.org/10.1111/resp.12661

Garcia-Aymerich, J., Tobías, A., Antó, J. M., & Sunyer, J. (2000). Air pollution and mortality in a cohort of patients with chronic obstructive pulmonary disease: a time series analysis. *Journal of Epidemiology and Community Health*, *54*(1), 73–74. https://doi.org/10.1136/jech.54.1.73

Garczewska, B., Jarzynka, S., Kuś, J., Skorupa, W., & Augustynowicz-Kopeć, E. (2016). Fungal infection of cystic fibrosis patients - single center experience. *Pneumonologia i Alergologia Polska*, *84*(3), 151–159. https://doi.org/10.5603/PiAP.2016.0017

Gastaldi, A., Jardim, J., & King, M. (2000). The influence of temperature and length of time of storage of frog mucus samples. *Biorheology*, *37*(3), 203–211.

Gensch, E., Gallup, M., Sucher, A., Li, D., Gebremichael, A., Lemjabbar, H., … Basbaum, C. (2004). Tobacco smoke control of mucin production in lung cells requires oxygen radicals AP-1 and JNK. *The Journal of Biological Chemistry*, *279*(37), 39085–39093. https://doi.org/10.1074/jbc.M406866200

Georgiades, P., Pudney, P. D. A., Thornton, D. J., & Waigh, T. A. (2014). Particle Tracking Microrheology of Purified Gastrointestinal Mucins. *Biopolymers*, *101*(4), 366–377. https://doi.org/10.1002/bip.22372

Global Burden of Disease Collaborators. (2015). Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* (Vol. 385). Elsevier Ltd. https://doi.org/10.1016/S0140-6736(14)61682-2

Global Initiative for Chronic Obstructive Lung Disease. (2018). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease 2018 Report. *Global Obstructive Lung Disease*. https://doi.org/10.1097/00008483-200207000-00004

Gooptu, B., Ekeowa, U. I., & Lomas, D. A. (2009). Mechanisms of emphysema in α1-antitrypsin deficiency: molecular and cellular insights. *The European Respiratory Journal*, *34*(2), 475–488. https://doi.org/10.1183/09031936.00096508

Goormaghtigh, E., Gasper, R., Bénard, A., Goldsztein, A., & Raussens, V. (2009). Protein secondary structure content in solution, films and tissues: Redundancy and complementarity of the information content in circular dichroism, transmission and ATR FTIR spectra. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1794*(9), 1332–1343. https://doi.org/10.1016/j.bbapap.2009.06.007

Gosalia, N., Leir, S.-H., & Harris, A. (2013). Coordinate regulation of the gel-forming mucin genes at chromosome 11p15.5. *The Journal of Biological Chemistry*, *288*(9), 6717–6725. https://doi.org/10.1074/jbc.M112.437400

Grewal, M. K., Jaiswal, P., & Jha, S. N. (2015). Detection of poultry meat specific bacteria using FTIR spectroscopy and chemometrics. *Journal of Food Science and Technology*, *52*(6), 3859–3869. https://doi.org/10.1007/s13197-014-1457-9

Guevera, M. T., & Mccolley, S. A. (2017). The Safety of Lumacaftor and Ivacaftor for the Treatment of Cystic Fibrosis. *Expert Opinion on Drug Safety*, *16*(11), 1305–1311. https://doi.org/10.1080/14740338.2017.1372419. The

Guha, S., Hariharan, P., & Myers, M. R. (2014). Enhancement of ICRP's Lung Deposition Model for Pathogenic Bioaerosols. *Aerosol Science and Technology*, *48*(12), 1226–1235. https://doi.org/10.1080/02786826.2014.975334

Gustafsson, J. K., Ermund, A., Ambort, D., Johansson, M. E. V., Nilsson, H. E., Thorell, K., … Hansson, G. C. (2012). Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *The Journal of Experimental Medicine*, *209*(7), 1263–1272. https://doi.org/10.1084/jem.20120562

Han, M. K., Postma, D. S., Mannino, D. M., Giardino, N. D., Buist, S., Curtis, J. L., & Martínez, F. J. (2007). Gender and chronic obstructive pulmonary disease: why it matters. *American Journal of Respiratory and Critical Care Medicine*, *176*(12), 1179–1184. https://doi.org/10.1164/rccm.200704-553CC

Hands, J. R., Clemens, G., Stables, R., Ashton, K., Brodbelt, A., Davis, C., … Baker, M. J. (2016). Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy. *Journal of Neuro-Oncology*, *127*(3), 463–472. https://doi.org/10.1007/s11060-016-2060-x

Haq, I., Irving, J. A., Saleh, A. D., Dron, L., Regan-Mochrie, G. L., Motamedi-Shad, N., … Lomas, D. A. (2016). Deficiency mutations of alpha-1 antitrypsin: Effects on folding, function, and polymerization. *American Journal of Respiratory Cell and Molecular Biology*, *54*(1), 71–80. https://doi.org/10.1165/rcmb.2015-0154OC

Hasleton, P. S. (1972). The internal surface area of the adult human lung. *Journal of Anatomy*, *112*(3), 391–400.

Health and Safety Executive. (2014). *Chronic Obstructive Pulmonary Disease (COPD) in Great Britain in 2014*. *Health and Safety Executive*.

Hegewald, M. J., Gallo, H. M., & Wilson, E. L. (2016). Accuracy and Quality of Spirometry in Primary Care Offices. *Annals of the American Thoracic Society*, *13*(12), 2119–2124. https://doi.org/10.1513/AnnalsATS.201605-418OC

Hengzhuang, W., Song, Z., Ciofu, O., Onsøyen, E., Rye, P. D., & Høiby, N. (2016). OligoG CF-5/20 Disruption of Mucoid *Pseudomonas aeruginosa* Biofilm in a Murine Lung Infection Model. *Antimicrobial Agents and Chemotherapy*, *60*(5), 2620–2626. https://doi.org/10.1128/AAC.01721-15.Address

Hoffman, R. M., & Sanchez, R. (2017). Lung Cancer Screening. *Medical Clinics of North America*, *101*(4), 769–785. https://doi.org/10.1016/j.mcna.2017.03.008

Hogg, J. C. (2004). Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet*, *364*(9435), 709–721. https://doi.org/10.1016/S0140-6736(04)16900-6

Houghton, A. M., Mouded, M., & Shapiro, S. D. (2008). Common Origins of Lung Cancer and COPD. *Nature Medicine*, *14*(10), 1023–1024. https://doi.org/10.1164/rccm.200802-336OC

Hovenberg, H. W., Davies, J. R., & Carlstedt, I. (1996). Different mucins are produced by the surface epithelium and the submucosa in human trachea: identification of MUC5AC as a major mucin from the goblet cells. *The Biochemical Journal*, *318 ( Pt 1*, 319–324.

Huang, Y. J., & Boushey, H. A. (2015). The sputum microbiome in chronic obstructive pulmonary disease exacerbations. *Annals of the American Thoracic Society*, *12*(November), S176–S180. https://doi.org/10.1513/AnnalsATS.201506-319AW

Hunninghake, G. M., Cho, M. H., Tesfaigzi, Y., Soto-Quiros, M. E., Avila, L., Lasky-Su, J., … Celedon, J. C. (2009). MMP12, Lung Function, and COPD in High-Risk Populations. *The New England Journal of Medicine*, *361*(27), 2599–2608. https://doi.org/10.1056/NEJMoa0904006.MMP12

Hurst, J. R., Donaldson, G. C., Quint, J. K., Goldring, J. J. P., Baghai-Ravary, R., & Wedzicha, J. A. (2009). Temporal clustering of exacerbations in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *179*(5), 369–374. https://doi.org/10.1164/rccm.200807-1067OC

Ibrahim, M., & Garcia-Contreras, L. (2013). Mechanisms of absorption and elimination of drugs administered by inhalation. *Therapeautic Delivery*, *4*(8), 1027–1045. https://doi.org/10.4155/tde.13.67

Ishibashi, Y., Takayama, G., Inouye, Y., & Taniguchi, A. (2010). Carbocisteine normalizes the viscous property of mucus through regulation of fucosylated and sialylated sugar chain on airway mucins. *European Journal of Pharmacology*, *641*(2–3), 226–228. https://doi.org/10.1016/j.ejphar.2010.05.045

Iwamoto, H., Gao, J., Koskela, J., Kinnula, V., Kobayashi, H., Laitinen, T., & Mazur, W. (2014). Differences in plasma and sputum biomarkers between COPD and COPD-asthma overlap. *European Respiratory Journal*, *43*(2), 421–429. https://doi.org/10.1183/09031936.00024313

Kačuráková, M., & Mathlouthi, M. (1996). FTIR and laser-Raman spectra of oligosaccharides in water: Characterization of the glycosidic bond. *Carbohydrate Research*, *284*(2), 145–157. https://doi.org/10.1016/0008-6215(95)00412-2

Kato, K., Hanss, A. D., Zemskova, M. A., Morgan, N. E., Kim, M., Knox, K. S., … Chul, K. (2017). *Pseudomonas aeruginosa* increases MUC1 expression in macrophages through the TLR4-p38 pathway. *Biochemical and Biophysical Research Communications*, 1–5. https://doi.org/dx.doi.org/10.1016/j.bbrc.2017.08.056

Kawabata, K., Hagio, T., & Matsuoka, S. (2002). The role of neutrophil elastase in acute lung injury. *European Journal of Pharmacology*, *451*(1), 1–10. https://doi.org/10.1016/S0014-2999(02)02182-9

Khajehpour, M., Dashnau, J. L., & Vanderkooi, J. M. (2006). Infrared spectroscopy used to evaluate glycosylation of proteins. *Analytical Biochemistry*, *348*(1), 40–48. https://doi.org/10.1016/j.ab.2005.10.009

Khan, S., Tøndervik, A., Sletta, H., Klinkenberg, G., Emanuel, C., Onsøyen, E., … Thomas, D. W. (2012). Overcoming drug resistance with alginate oligosaccharides able to potentiate the action of selected antibiotics. *Antimicrobial Agents and Chemotherapy*, *56*(10), 5134–5141.

https://doi.org/10.1128/AAC.00525-12

Kim, K. C., McCracken, K., Lee, B. C., Shin, C. Y., Jo, M. J., Lee, C. J., & Ko, K. H. (1997). Airway goblet cell mucin: its structure and regulation of secretion. *European Respiratory Journal*, *10*(11), 2644–2649. https://doi.org/10.1183/09031936.97.10112644

Kim, Y. S., & Ho, S. B. (2010). Intestinal goblet cells and mucins in health and disease: Recent insights and progress. *Current Gastroenterology Reports*, *12*(5), 319–330. https://doi.org/10.1007/s11894-010-0131-2

Kirkham, S., Kolsum, U., Rousseau, K., Singh, D., Vestbo, J., & Thornton, D. J. (2008). MUC5B is the major mucin in the gel phase of sputum in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *178*(10), 1033–1039. https://doi.org/10.1164/rccm.200803-391OC

Kirkham, S., Sheehan, J. K., Knight, D., Richardson, P. S., & Thornton, D. J. (2002). Heterogeneity of airways mucus: variations in the amount and glycoforms of the major oligomeric mucins MUC5AC and MUC5B. *Biochemical Journal*, *361*, 537–546. https://doi.org/10.1042/0264-6021:3610537

Knowles, M. R., & Boucher, R. C. (2002). Mucus clearance as a primary innnate defense mechanism for mammalian airways. *The Journal of Clinical Investigation*, *109*(5), 571–577. https://doi.org/10.1172/JCI200215217.The

Koehler, K. F., Helguero, L. A., Haldosén, L.-A., Warner, M., & Gustafsson, J.-A. (2005). Reflections on the discovery and significance of estrogen receptor beta. *Endocrine Reviews*, *26*(3), 465–478. https://doi.org/10.1210/er.2004-0027

Kolwijck, E., Mitchell, M., Venter, A., Friedrich, S. O., Dawson, R., & Diacon, A. H. (2013). Short-term storage does not affect the quantitative yield of mycobacterium tuberculosis in sputum in early-bactericidal-activity studies. *Journal of Clinical Microbiology*, *51*(4), 1094–1098. https://doi.org/10.1128/JCM.02751-12

Kong, J., & Yu, S. (2007). Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochimica et Biophysica Sinica*, *39*(8), 549–559. https://doi.org/10.1111/j.1745-7270.2007.00320.x

Latimer, K. M., & Mott, T. F. (2015). Lung Cancer: Diagnosis, Treatment Principles, and Screening. *Am Fam Physician*, *91*(4), 250–256.

Lee, R. J., & Foskett, J. K. (2014). Ca2+ signaling and fluid secretion by secretory cells of the airway epithelium. *Cell Calcium*, *55*(6), 325–336. https://doi.org/10.1016/j.ceca.2014.02.001

Lemjabbar, H., Li, D., Gallup, M., Sidhu, S., Drori, E., & Basbaum, C. (2003). Tobacco smoke-induced lung cell proliferation mediated by tumor necrosis factor alpha-converting enzyme and amphiregulin. *The Journal of Biological Chemistry*, *278*(28), 26202–26207. https://doi.org/10.1074/jbc.M207018200

Lewis, A. T., Jones, K., Lewis, K. E., Jones, S., & Lewis, P. D. (2013). Detection of Lewis antigen structural change by FTIR spectroscopy. *Carbohydrate Polymers*, *92*(2), 1294–1301. https://doi.org/10.1016/j.carbpol.2012.09.078

Lewis, P. D., Lewis, K. E., Ghosal, R., Bayliss, S., Lloyd, A. J., Wills, J., … Mur, L. A. J. (2010). Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer*, *10*(1), 640. https://doi.org/10.1186/1471-2407-10-640

Lewis, P. D., & Menzies, G. E. (2015). Vibrational spectra, principal components analysis and the horseshoe effect. *Vibrational Spectroscopy*, *81*, 62–67. https://doi.org/10.1016/j.vibspec.2015.10.002

Lewis, S. P., Lewis, A. T., & Lewis, P. D. (2013). Prediction of glycoprotein secondary structure using ATR-FTIR. *Vibrational Spectroscopy*, *69*, 21–29. https://doi.org/10.1016/j.vibspec.2013.09.001

Lhermitte, M., Rahmoune, H., Lamblin, G., Roussel, P., Strang, A.-M., & van Halbeek, H. (1991). Structures of neutral oligosaccharides isolated from the respiratory mucins of a non-secretor (O, Le a+b-) patient suffering from chronic bronchitis. *Glycobiology*, *1*(3), 277–293.

Li, J.-D., Feng, W., Gallup, M., Kim, J.-H., Gum, J., Kim, Y., & Basbaum, C. (1998). Activation of NF-κB via a Src-dependent Ras-MAPK-pp90rsk pathway is required for *Pseudomonas aeruginosa*-induced mucin overproduction in epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 5718–5723.

Li, X., Hu, Z., Qu, X., Zhu, J., Li, L., Ring, B. Z., & Su, L. (2011). Putative EPHX1 enzyme activity is related with risk of lung and upper aerodigestive tract cancers: A comprehensive meta-analysis. *PLoS ONE*, *6*(3), 1–12. https://doi.org/10.1371/journal.pone.0014749

Liang, R., Zhang, W., & Song, Y. M. (2013). Levels of leptin and IL-6 in lungs and blood are associated with the severity of chronic obstructive pulmonary disease in patients and rat models. *Molecular Medicine Reports*, *7*(5), 1470–1476. https://doi.org/10.3892/mmr.2013.1377

Liu, W., Wei, H., Gao, Z., Chen, G., Liu, Y., Gao, X., … Xiao, J. (2017). COL5A1 May Contribute the Metastasis of Lung Adenocarcinoma. *Gene*, (2018). https://doi.org/doi:10.1016/j.gene.2018.04.066 GENE

Llabjani, V., Malik, R. N., Trevisan, J., Hoti, V., Ukpebor, J., Shinwari, Z. K., … Martin, F. L. (2012). Alterations in the infrared spectral signature of avian feathers reflect potential chemical exposure: A pilot study comparing two sites in Pakistan. *Environment International*, *48*, 39–46. https://doi.org/10.1016/j.envint.2012.06.019

Lo-Guidice, J.-M., Herz, H., Lamblin, G., Plancke, Y., Roussel, P., & Lhermitte, M. (1997). Structures of sulfated oligosaccharides isolated from the respiratory mucins of a non-secretor ( O , Lea+b- ) patient suffering from chronic bronchitis. *Glycoconjugate Journal*, *14*, 113–125.

Lu, Y. C., Yeh, W. C., & Ohashi, P. S. (2008). LPS/TLR4 signal transduction pathway. *Cytokine*, *42*(2), 145–151. https://doi.org/10.1016/j.cyto.2008.01.006

Malins, D. C., Gilman, N. K., Green, V. M., Wheeler, T. M., Barker, E. A., & Anderson, K. M. (2005). A cancer DNA phenotype in healthy prostates, conserved in tumors and adjacent normal cells, implies a relationship to carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*,

*102*(52), 19093–19096. https://doi.org/10.1073/pnas.0509630102

Manser, R., Lethaby, A., Lb, I., Stone, C., Byrnes, G., Mj, A., & Campbell, D. (2013). Screening for lung cancer Review. *Cochrane Database of Systematic Reviews*, (6). https://doi.org/10.1002/14651858.CD001991.pub3.www.cochranelibrary.com

Matthews, L. W., Spector, S., Lemm, J., & Potter, J. L. (1963). Studies on pulmonary secretions 1. The Over-all Chemical Composition of Pulmonary Secretions from Patients with Cystic Fibrosis, Bronchiectasis, and Laryngectomy. *American Review of Respiratory Disease*, *88*(2), 199–204.

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. https://doi.org/10.1016/j.cdp.2007.03.003

McLean, S., Hoogendoorn, M., Hoogenveen, R. T., Feenstra, T. L., Wild, S., Simpson, C. R., … Sheikh, A. (2016). Projecting the COPD population and costs in England and Scotland: 2011 to 2030. *Scientific Reports*, *6*, 31893. https://doi.org/10.1038/srep31893

McPhail, S., Johnson, S., Greenberg, D., Peake, M., & Rous, B. (2015). Stage at diagnosis and early mortality from cancer in England. *British Journal of Cancer*, *112*(s1), S108–S115. https://doi.org/10.1038/bjc.2015.49

Meldrum, O. W., Yakubov, G. E., Bonilla, M. R., Deshmukh, O., McGuckin, M. A., & Gidley, M. J. (2018). Mucin gel assembly is controlled by a collective action of non-mucin proteins, disulfide bridges, Ca2+-mediated links, and hydrogen bonding. *Scientific Reports*, *8*(1), 5802. https://doi.org/10.1038/s41598-018-24223-3

Menzies, G. E., Fox, H. R., Marnane, C., Pope, L., Prabhu, V., Winter, S., … Lewis, P. D. (2014). Fourier transform infrared for noninvasive optical diagnosis of oral, oropharyngeal, and laryngeal cancer. *Translational Research*, *163*(1), 19–26. https://doi.org/10.1016/j.trsl.2013.09.006

Miller, M. R., Crapo, R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., … Wagner, J. (2005). General considerations for lung function testing. *European Respiratory Journal*, *26*(1), 153–161. https://doi.org/10.1183/09031936.05.00034505

Mirza, S., & Benzo, R. (2017). Chronic Obstructive Pulmonary Disease Phenotypes : Implications for Care. *Mayo Clinic Proceedings*, *92*(July), 1104–1112. https://doi.org/10.1016/j.mayocp.2017.03.020

Molfino, N. A., & Jeffery, P. K. (2007). Chronic obstructive pulmonary disease: histopathology, inflammation and potential therapies. *Pulmonary Pharmacology & Therapeutics*, *20*(5), 462–472. https://doi.org/10.1016/j.pupt.2006.04.003

Mollerup, S., Ryberg, D., Hewer, A., Phillips, D. H., & Haugen, A. (1999). Advances in Brief Sex Differences in Lung CYP1A1 Expression and DNA Adduct Levels among Lung Cancer Patients 1. *Cancer Research*, *59*, 3317–3320.

Monso, E., Rosell, A., Bonet, G., Manterola, J., Cardona, P. J., Ruiz, J., & Morera, J. (1999). Risk factors for lower airway bacterial colonization in chronic bronchitis. *European Respiratory Journal*, *13*, 338–342.

Montalbano, A. M., Riccobono, L., Siena, L., Chiappara, G., Di Sano, C., Anzalone, G., … Profita, M. (2015). Cigarette Smoke affects IL-17A, IL17-F and IL-17 receptor expression in the lung tissue: *Ex vivo* and *in vitro* studies. *Cytokine*, *76*(2015), 391–402. Retrieved from https://www.niehs.nih.gov/health/topics/agents/allergens/smoke/index.cfm

Morgan, I., & Wilkes, S. (2017). Improving early diagnosis of cancer in UK general practice. *British Journal of General Practice*, *67*(659), 276–277. https://doi.org/https://dx.doi.org/10.3399/bjgp17X691265

Mulshine, J. L., & D'Amico, T. A. (2014). Issues with implementing a high-quality lung cancer screening program. *CA: A Cancer Journal for Clinicians*, *64*(5), 352–363. https://doi.org/10.3322/caac.21239

Narayanan, S., Mainz, J. G., Gala, S., Tabori, H., & Grossoehme, D. (2017). Adherence to therapies in cystic fibrosis: a targeted literature review. *Expert Review of Respiratory Medicine*, *11*(2), 129–145. https://doi.org/10.1080/17476348.2017.1280399

National Institute for Health and Care Excellence. (2018). *Chronic Obstructive Pulmonary Disease in over 16s: Diagnosis and Management*. Retrieved from https://www.nice.org.uk/guidance/ng115. Date retrieved 15/12/18

Niewoehner, D. E., Kleinerman, J., & Rice, D. B. (1974). Pathologic changes in the peripheral airways of young cigarette smoker. *New England Journal of Medicine*, *10*(291), 755–758.

Noguera, A., Batle, S., Miralles, C., Iglesias, J., Busquets, X., MacNee, W., & Agustí, A. G. N. (2001). Enhanced neutrophil response in chronic obstructive pulmonary disease. *Thorax*, *56*(6), 432–437. https://doi.org/10.1136/thorax.56.6.432

Oberg, K. A., & Fink, A. L. (1998). A New Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy Method for the Study of Proteins in Solution. *Analytical Biochemistry*, *256*, 92–106.

Oliver, K. V, Vilasi, A., Maréchal, A., Moochhala, S. H., Unwin, R. J., & Rich, P. R. (2016). Infrared vibrational spectroscopy: a rapid and novel diagnostic and monitoring tool for cystinuria. *Scientific Reports*, *6*(July), 34737. https://doi.org/10.1038/srep34737

Onishi, R. M., & Gaffen, S. L. (2010). Interleukin-17 and its target genes: Mechanisms of interleukin-17 function in disease. *Immunology*, *129*(3), 311–321. https://doi.org/10.1111/j.1365-2567.2009.03240.x

Packer, N. H., Lawson, M. A., Jardine, D. R., Sanchez, J. C., & Gooley, A. A. (1998). Analyzing glycoproteins separated by two-dimensional gel electrophoresis. *Electrophoresis*, *19*, 981–988. https://doi.org/10.1002/elps.1150190613

Panicker, G., Meadows, K. S., Lee, D. R., Nisenbaum, R., & Unger, E. R. (2007). Effect of storage temperatures on the stability of cytokines in cervical mucous. *Cytokine*, *37*(2), 176–179. https://doi.org/10.1016/j.cyto.2007.03.006

Papakonstantinou, E., Roth, M., Klagas, I., Karakiulakis, G., Tamm, M., & Stolz, D. (2015). COPD exacerbations are associated with pro-inflammatory degradation of hyaluronic acid. *Chest*, *148*(6), 1497–1507. https://doi.org/10.1378/chest.15-0153

Papi, A., Luppi, F., Franco, F., & Fabbri, L. M. (2006). Pathophysiology of exacerbations of chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, *3*(3), 245–251. https://doi.org/10.1513/pats.200512-125SF

Parker, D., & Prince, A. (2011). Innate Immunity in the Respiratory Epithelium. *American Journal of Respiratory Cell and Molecular Biology*, *45*(2), 189–201. https://doi.org/10.1165/rcmb.2011-0011RT

Parr, D. G., White, A. J., Bayley, D. L., Guest, P. J., & Stockley, R. A. (2006). Inflammation in sputum relates to progression of disease in subjects with COPD: A prospective descriptive study. *Respiratory Research*, *7*(136), 1–11. https://doi.org/10.1186/1465-9921-7-136

Pauwels, R. A., Buist, A. S., Calverley, P. M., Jenkins, C. R., & Hurd, S. S. (2001). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *American Journal of Respiratory and Critical Care Medicine*, *163*(5), 1256–1276. https://doi.org/10.1164/ajrccm.163.5.2101039

Payne, J. E., Dubois, A. V., Ingram, R. J., Weldon, S., Taggart, C. C., Elborn, J. S., & Tunney, M. M. (2017). Activity of innate antimicrobial peptides and ivacaftor against clinical cystic fibrosis respiratory pathogens. *International Journal of Antimicrobial Agents*, *50*(3), 427–435. https://doi.org/10.1016/j.ijantimicag.2017.04.014

Pillai, S. G., Ge, D., Zhu, G., Kong, X., Shianna, K. V., Need, A. C., … Goldstein, D. B. (2009). A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci. *PLoS Genetics*, *5*(3), e1000421. https://doi.org/10.1371/journal.pgen.1000421

Pilling, M., & Gardner, P. (2016). Fundamental developments in infrared spectroscopic imaging for biomedical applications. *Chem. Soc. Rev.*, *45*(7), 1935–1957. https://doi.org/10.1039/C5CS00846H

Powell, L. C., Pritchard, M. F., Emanuel, C., Onsøyen, E., Rye, P. D., Wright, C. J., … Thomas, D. W. (2014). A nanoscale characterization of the interaction of a novel alginate oligomer with the cell surface and motility of *Pseudomonas aeruginosa*. *American Journal of Respiratory Cell and Molecular Biology*, *50*(3), 483–492. https://doi.org/10.1165/rcmb.2013-0287OC

Powell, L. C., Pritchard, M. F., Ferguson, E. L., Powell, K. A., Patel, S. U., Rye, P. D., … Thomas, D. W. (2018). Targeted disruption of the extracellular polymeric network of *Pseudomonas aeruginosa* biofilms by alginate oligosaccharides. *Npj Biofilms and Microbiomes*, *4*(May). https://doi.org/10.1038/s41522-018-0056-3

Powell, L. C., Sowedan, A., Khan, S., Wright, C. J., Hawkins, K., Onsøyen, E., … Thomas, D. W. (2013). The effect of alginate oligosaccharides on the mechanical properties of Gram-negative biofilms. *Biofouling*, *29*(4), 413–421. https://doi.org/10.1080/08927014.2013.777954

Pritchard, M. F., Powell, L. C., Khan, S., Griffiths, P. C., Mansour, O. T., Schweins, R., … Ferguson, E. L. (2017). The antimicrobial effects of the alginate oligomer

OligoG CF-5/20 are independent of direct bacterial cell membrane disruption. *Scientific Reports*, *7*(March), 44731. https://doi.org/10.1038/srep44731

Pritchard, M. F., Powell, L. C., Menzies, G. E., Lewis, P. D., Hawkins, K., Wright, C., … Thomas, D. W. (2016). A New Class of Safe Oligosaccharide Polymer Therapy to Modify the Mucus Barrier of Chronic Respiratory Disease. *Molecular Pharmaceutics*, *13*(3), 863–872. https://doi.org/10.1021/acs.molpharmaceut.5b00794

Qiu, F., Liang, C.-L., Liu, H., Zeng, Y.-Q., Hou, S., Huang, S., … Dai, Z. (2015). Impacts of cigarette smoking on immune responsiveness: Up and down or upside down? *Oncotarget*, *8*(1), 268–284. https://doi.org/10.18632/oncotarget.13613

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. https://doi.org/ISBN 3-900051-07-0

Rabe, K. F., Hurd, S. S., Anzueto, A., Barnes, P. J., Buist, S. A., Calverley, P., … Zielinski, J. (2007). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American Journal of Respiratory and Critical Care Medicine*, *176*(6), 532–555. https://doi.org/10.1164/rccm.200703-456SO

Rackley, C. R., & Stripp, B. R. (2012). Building and Maintaining the Epithelium of the Lung. *The Journal of Clinical Investigation*, *122*(8), 2724–2730. https://doi.org/10.1172/JCI60519; 10.1172/JCI60519

Ram, B., Young, S., Wook, D., Kook, H., & Hyock, J. (2017). Influence of Socioeconomic Status, Comorbidity, and Disability on Late-stage Cancer Diagnosis. *Osong Public Health Res Perspect*, *8*(4), 264–270.

Ramsey, B. W., Davies, J., McElvaney, N. G., Tullis, E., Bell, S. C., Dřevínek, P., … Elborn, J. S. (2011). A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation. *The New England Journal of Medicine*, *365*(18), 1663–1672. https://doi.org/10.1056/NEJMoa1411087

Raynal, B. D. E., Hardingham, T. E., Sheehan, J. K., & Thornton, D. J. (2003). Calcium-dependent protein interactions in MUC5B provide reversible cross-links in salivary mucus. *Journal of Biological Chemistry*, *278*(31), 28703–28710. https://doi.org/10.1074/jbc.M304632200

Reid, C. J., Gould, S., & Harris, A. (1997). Developmental Expression of Mucin Genes in the Human Respiratory Tract. *American Journal of Respiratory Cell and Molecular Biology*, *17*, 592–598.

Reidel, B., Radicioni, G., Clapp, P., Ford, A. A., Abdelwahab, S., Rebuli, M. E., … Kesimer, M. (2018). E-Cigarette Use Causes a Unique Innate Immune Response in the Lung, Involving Increased Neutrophilic Activation and Altered Mucin Secretion. *AJRCCM*, *197*(4). https://doi.org/10.1164/rccm.201708-1590OC

Reznikov, L. R., Abou Alaiwa, M. H., Dohrn, C. L., Gansemer, N. D., Diekema, D. J., Stoltz, D. A., & Welsh, M. J. (2014). Antibacterial properties of the CFTR potentiator ivacaftor. *Journal of Cystic Fibrosis*, *13*(5), 515–519. https://doi.org/10.1016/j.jcf.2014.02.004

Ridley, C., Kouvatsos, N., Raynal, B. D., Howard, M., Collins, R. F., Desseyn, J.-L., … Thornton, D. J. (2014). Assembly of the respiratory mucin MUC5B: a new model for a gel-forming mucin. *The Journal of Biological Chemistry*, *289*(23), 16409–16420. https://doi.org/10.1074/jbc.M114.566679

Riethmueller, J., Vonthein, R., Borth-Bruhns, T., Grassm, H., Eyrich, M., Schilbach, K., … Gulbins, E. (2008). DNA quantification and fragmentation in sputum after inhalation of recombinant human deoxyribonuclease. *Cellular Physiology and Biochemistry*, *22*(1–4), 347–352. https://doi.org/10.1159/000149813

Rivera, M. P., Mehta, A. C., & Wahidi, M. M. (2013). Establishing the diagnosis of lung cancer. Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*, *143*(5 SUPPL), e142S–e165S. https://doi.org/10.1378/chest.12-2353

Roberts, J. L., Khan, S., Emanuel, C., Powell, L. C., Pritchard, M. F., Onsøyen, E., … Hill, K. E. (2013). An in vitro study of alginate oligomer therapies on oral biofilms. *Journal of Dentistry*, *41*(10), 892–899. https://doi.org/10.1016/j.jdent.2013.07.011

Rodrigues, L. M., Magrini, T. D., Lima, C. F., Scholz, J., da Silva Martinho, H., & Almeida, J. D. (2017). Effect of smoking cessation in saliva compounds by FTIR spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *174*, 124–129. https://doi.org/10.1016/j.saa.2016.11.009

Rohde, G., Wiethege, A., Borg, I., Kauth, M., Bauer, T. T., Gillissen, A., … Schultze-Werninghaus, G. (2003). Respiratory viruses in exacerbations of chronic obstructive pulmonary disease requiring hospitalisation: a case-control study. *Thorax*, *58*(1), 37–42. https://doi.org/10.1136/thorax.58.1.37

Roos, A. B., Sandén, C., Mori, M., Bjermer, L., Stampfli, M. R., & Erjefält, J. S. (2015). IL-17A is elevated in end-stage chronic obstructive pulmonary disease and contributes to cigarette smoke-induced lymphoid neogenesis. *American Journal of Respiratory and Critical Care Medicine*. https://doi.org/10.1164/rccm.201410-1861OC

Rose, M. C., & Voynow, J. A. (2006). Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiological Reviews*, *86*(1), 245–278. https://doi.org/10.1152/physrev.00010.2005

Saetta, M., Turato, Graziella, Baraldo, Simonetta, Zanin, A., Braccioni, F., Mapp, C. E., … Fabbri, L. M. (2000). Goblet Cell Hyperplasia and Epithelial Inflammation in Peripheral Airways of Smokers with Both Symptoms of Chronic Bronchitis and Chronic Airflow Limitation. *American Journal of Respiratory and Critical Care Medicine*, *161*, 1016–1021.

Schneider, A., Gindner, L., Tilemann, L., Schermer, T., Dinant, G., Meyer, F. J., & Szecsenyi, J. (2009). Diagnostic accuracy of spirometry in primary care. *BMC Pulmonary Medicine*, *9*(31), 1–10. https://doi.org/10.1186/1471-2466-9-31

Schwarz, C., Brandt, C., Whitaker, P., Sutharsan, S., Skopnik, H., Gartner, S., … Röhmel, J. F. (2018). Invasive Pulmonary Fungal Infections in Cystic Fibrosis. *Mycopathologia*, *183*(1), 33–43. https://doi.org/10.1007/s11046-017-0199-4

Seatta, M., Stefano, A. Di, Turato, G., Facchini, F. M., Corbino, L., Mapp, C. E., … Fabbri, L. M. (1998). CD8+ T-lymphocytes in peripheral airways of smokers with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *157*, 822–826. Retrieved from http://www.atsjournals.org/doi/pdf/10.1164/ajrccm.157.3.9709027

Seemungal, T. A. R., Donaldson, G. C., Bhowmik, A., Jeffries, D. J., & Wedzicha, J. A. (2000). Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *161*(5), 1608–1613. https://doi.org/10.1164/ajrccm.161.5.9908022

Seemungal, T. A. R., Donaldson, G. C., Paul, E. A., Bestall, J. C., Jeffries, D. J., & Wedzicha, J. A. (1998). Effect of Exacerbation on Quality of Life in Patients with Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine*, *157*, 1418–1422.

Seemungal, T. A. R., Harper-Owen, R., Bhowmik, A., Jeffries, D. J., & Wedzicha, J. A. (2000). Detection of rhinovirus in induced sputum at exacerbation of chronic obstructive pulmonary disease. *The European Respiratory Journal*, *16*(4), 677–683. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11106212

Seemungal, T. A. R., Harper-Owen, R., Bhowmik, A., Moric, I., Sanderson, G., Message, S., … Wedzicha, J. a. (2001). Respiratory viruses, symptoms, and inflammatory markers in acute exacerbations and stable chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, *164*(9), 1618–1623. https://doi.org/10.1164/rccm2105011

Seemungal, T. A. R., Hurst, J. R., & Wedzicha, J. A. (2009). Exacerbation rate , health status and mortality in COPD – a review of potential interventions. *International Journal of COPD*, *4*, 203–223.

Seibold, M. A., Wise, A. L., Speer, M. C., Steele, M. P., Brown, K. K., Loyd, J. E., … Schwartz, D. a. (2011). A Common MUC5B Promoter Polymorphism and Pulmonary Fibrosis. *The New England Journal of Medicine*, *346*(16), 1503–1512.

Shak, S., Capon, D. J., Hellmiss, R., Marsters, S. A., & Baker, C. L. (1990). Recombinant human DNase I reduces the viscosity of cystic fibrosis sputum. *Proceedings of the National Academy of Sciences*, *87*(23), 9188–9192. https://doi.org/10.1073/pnas.87.23.9188

Shao, M. X. G., Ueki, I. F., & Nadel, J. A. (2003). Tumor necrosis factor α-converting enzyme mediates MUC5AC mucin expression in cultured human airway epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11618–11623.

Sheehan, J. K., Howard, M., Richardson, P. S., Longwill, T., & Thornton, D. J. (1999). Physical characterization of a low-charge glycoform of the MUC5B mucin comprising the gel-phase of an asthmatic respiratory mucous plug. *Biochemical Journal*, *338*, 507–513.

Sheehan, J. K., Kirkham, S., Howard, M., Woodman, P., Kutay, S., Brazeau, C., … Thornton, D. J. (2004). Identification of molecular intermediates in the assembly pathway of the MUC5AC mucin. *The Journal of Biological Chemistry*, *279*(15), 15698–15705. https://doi.org/10.1074/jbc.M313241200

Shental-Bechor, D., & Levy, Y. (2008). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(24), 8256–8261.

Sherratt, F. C., Field, J. K., & Marcus, M. W. (2017). Association between smoking and health outcomes in an economically deprived population: the Liverpool Lung Project. *Journal of Epidemiology and Community Health*, jech-2016-208730. https://doi.org/10.1136/jech-2016-208730

Singh, P. K., & Hollingsworth, M. A. (2006). Cell surface-associated mucins in signal transduction. *Trends in Cell Biology*, *16*(9), 467–476. https://doi.org/10.1016/j.tcb.2006.07.006

Skillrud, D. M., Offord, K. P., & Miller, R. D. (1986). Higher Risk of Lung Cancer in Chronic Obstructive Pulmonary Disease A Prospective, Matched, Controlled Study. *Annals of Internal Medicine*, *105*(4), 503–507. Retrieved from http://dx.doi.org/10.7326/0003-4819-105-4-503

Smith, B. R., Ashton, K. M., Brodbelt, A., Dawson, T., Jenkinson, M. D., Hunt, N. T., … Baker, M. J. (2016). Combining random forest and 2-D correlation analysis to identify serum spectral signatures for neuro-oncology. *The Analyst*, *141*(12), 3668–3678. https://doi.org/10.1039/C5AN02452H

Song, J.-S., Cho, K.-S., Yoon, H.-K., Moon, H.-S., & Park, S.-H. (2005). Neutrophil Elastase Causes MUC5AC Mucin Synthesis Via EGF Receptor, ERK and NF-kB Pathways in A549 Cells. *The Korean Journal of Internal Medicine*, *20*(4), 275–283.

Song, K. S., Lee, W. J., chung, K. C., Koo, J. S., Yang, E. J., Choi, J. Y., & Yoon, J. H. (2003). Interleukin-1β and tumor necrosis factor-α induce MUC5AC overexpression through a mechanism involving ERK/p38 mitogen-activated protein kinases-MSK1-CREB activation in human airway epithelial cells. *Journal of Biological Chemistry*, *278*(26), 23243–23250. https://doi.org/10.1074/jbc.M300096200

Stapley, S., Sharp, D., & Hamilton, W. (2006). Negative chest X-rays in primary care patients with lung cancer. *British Journal of General Practice*, *56*(529), 570–573.

Statistics on smoking England: 2017. (2017). *National Statistics*. https://doi.org/10.1038/1811181a0

Stewart, B. W., & Wild, C. P. (2014). World cancer report 2014. *World Health Organization*, 1–2. https://doi.org/9283204298

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., … Hogenesch, J. B. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 6062–6067. https://doi.org/10.1073/pnas.0400782101

Sutedja, G. (2003). New techniques for early detection of lung cancer. *European Respiratory Journal*, *21*(Suppl. 39), 57s–66s. https://doi.org/10.1183/09031936.03.00405303

Svanes, C., Sunyer, J., Plana, E., Dharmage, S., Heinrich, J., Jarvis, D., … Antó, J. M. (2010). Early life origins of chronic obstructive pulmonary disease. *Thorax*, *65*(1), 14–20. https://doi.org/10.1136/thx.2008.112136

Takeyama, K., Jung, B., Shim, J. J., Burgel, P. R., Dao-Pick, T., Ueki, I. F., … Nadel, J. A. (2001). Activation of epidermal growth factor receptors is responsible for mucin synthesis induced by cigarette smoke. *Am.J Physiol Lung Cell Mol.Physiol*, *280*(1), L165–L172.

Tam, A., Morrish, D., Wadsworth, S., Dorscheid, D., Man, S.-F. P., & Sin, D. D. (2011). The role of female hormones on lung function in chronic lung diseases. *BMC Women's Health*, *11*(1), 24. https://doi.org/10.1186/1472-6874-11-24

Tam, A., Wadsworth, S., Dorscheid, D., Man, S.-F. P., & Sin, D. D. (2014). Estradiol increases mucus synthesis in bronchial epithelial cells. *PloS One*, *9*(6), e100633. https://doi.org/10.1371/journal.pone.0100633

Taylor-Cousar, J. L., Munck, A., McKone, E. F., van der Ent, C. K., Moeller, A., Simard, C., … Elborn, J. S. (2017). Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del. *New England Journal of Medicine*, *377*(21), 2013–2023. https://doi.org/10.1056/NEJMoa1709846

Thai, P., Loukoianov, A., Wachi, S., & Wu, R. (2008). Regulation of airway mucin gene expression. *Annual Review of Physiology*, *70*, 405–429. https://doi.org/10.1146/annurev.physiol.70.113006.100441

Thorley, A. J., Grandolfo, D., Lim, E., Goldstraw, P., Young, A., & Tetley, T. D. (2011). Innate immune responses to bacterial ligands in the peripheral human Lung - Role of alveolar epithelial TLR expression and signalling. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0021827

Thornton, D. J., Rousseau, K., & McGuckin, M. A. (2008). Structure and function of the polymeric mucins in airways mucus. *Annual Review of Physiology*, *70*, 459–486. https://doi.org/10.1146/annurev.physiol.70.113006.100702

Thunnissen, F. B. J. M. (2003). Sputum examination for early detection of lung cancer. *J Clin Pathol*, *56*(11), 805–810. https://doi.org/10.1136/jcp.56.11.805

Titz, B., Sewer, A., Schneider, T., Elamin, A., Martin, F., Dijon, S., … Peitsch, M. C. (2015). Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. *Journal of Proteomics*, *128*, 306–320. https://doi.org/10.1016/j.jprot.2015.08.009

Tockman, M. S., Anthonisen, N., Wright, E., & Donithan, M. (1987). Airways Obstruction and the Risk for Lung Cancer. *Annals of Internal Medicine*, *106*(4), 512. https://doi.org/10.7326/0003-4819-106-4-512

Tøndervik, A., Sletta, H., Klinkenberg, G., Emanuel, C., Powell, L. C., Pritchard, M. F., … Hill, K. E. (2014). Alginate oligosaccharides inhibit fungal cell growth and potentiate the activity of antifungals against *Candida* and *Aspergillus spp*. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0112518

Toyoda, Y., Nakayama, T., Kusunoki, Y., Iso, H., & Suzuki, T. (2008). Sensitivity and specificity of lung cancer screening using chest low-dose computed tomography. *British Journal of Cancer*, *98*(10), 1602–1607. https://doi.org/10.1038/sj.bjc.6604351

Triebner, K., Matulonga, B., Johannessen, A., Suske, S., Benediktsdottir, B., Demoly, P., … Real, F. G. (2017). Menopause Is Associated with Accelerated Lung Function Decline. *American Journal of Respiratory and Critical Care Medicine*, *195*(8), 1058–

1065. https://doi.org/10.1164/rccm.201605-0968OC

Uchmanowicz, I., Jankowska-Polanska, B., Rosinczuk, J., & Wleklik, M. (2015). Health-related quality of life of patients suffering from cystic fibrosis. *Advances in Clinical and Experimental Medicine*, *24*(1), 147–152. https://doi.org/10.17219/ACEM/38147

Unger, M., Pfeifer, F., & Siesler, H. W. (2016). Comparative Variable Temperature Studies of Polyamide II with a Benchtop Fourier Transform and a Miniature Hand-held Near-Infrared Spectrometer Using 2-D-COS and PCMW-2-D Analysis. *Applied Spectroscopy*, *70*(7), 1202–1208. https://doi.org/10.1177/0003702816652365

Vahur, S., Teearu, A., Peets, P., Joosu, L., & Leito, I. (2016). ATR-FT-IR spectral collection of conservation materials in the extended region of 4000-80 cm–1. *Analytical and Bioanalytical Chemistry*, *408*(13), 3373–3379. https://doi.org/10.1007/s00216-016-9411-5

van Halbeek, H., Strang, A.-M., Lhermitte, M., Rahmoune, H., Lamblin, G., & Roussel, P. (1994). Structures of neutral oligosaccharides isolated from the respiratory mucins of a non-secretor (O, Le a+b-) patient suffering from chronic bronchitis. *Glycobiology*, *4*(2), 203–219.

Vareille, M., Kieninger, E., Edwards, M. R., & Regamey, N. (2011). The airway epithelium: Soldier in the fight against respiratory viruses. *Clinical Microbiology Reviews*, *24*(1), 210–229. https://doi.org/10.1128/CMR.00014-10

Vitko, M., Valerio, D. M., Rye, P. D., Onsøyen, E., Myrset, A. H., Dessen, A., … Hodges, C. A. (2016). A novel guluronate oligomer improves intestinal transit and survival in cystic fibrosis mice. *Journal of Cystic Fibrosis*, *15*(6), 745–751. https://doi.org/10.1016/j.jcf.2016.06.005

Voynow, J. A. (2002). What does mucin have to do with lung disease? *Paediatric Respiratory Reviews*, *3*(2), 98–103. https://doi.org/10.1016/S1526-0550(02)00007-0

Voynow, J. A., Gendler, S. J., & Rose, M. C. (2006). Regulation of mucin genes in chronic inflammatory airway diseases. *American Journal of Respiratory Cell and Molecular Biology*, *34*(6), 661–665. https://doi.org/10.1165/rcmb.2006-0035SF

Voynow, J. A., & Rubin, B. K. (2009). Mucins, mucus, and sputum. *Chest*, *135*(2), 505–512. https://doi.org/10.1378/chest.08-0412

Wagner, V. E., & Iglewski, B. H. (2008). *P. aeruginosa* Biofilms in CF Infection. *Clinical Reviews in Allergy {&} Immunology*, *35*(3), 124–134. https://doi.org/10.1007/s12016-008-8079-9

Wainwright, C. E., Elborn, J. S., Ramsey, B. W., Marigowda, G., Huang, X., Cipolli, M., … Boyle, M. P. (2015). Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. *New England Journal of Medicine*, *373*(3), 220–231. https://doi.org/10.1056/NEJMoa1409547.

Walters, S., Maringe, C., Coleman, M. P., Peake, M. D., Butler, J., Young, N., … Rachet, B. (2013). Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004–2007. *Thorax*, *68*(6), 551–564. https://doi.org/10.1136/thoraxjnl-2012-202297

Wedzicha, J. A., & Seemungal, T. A. R. (2007). COPD exacerbations: defining their cause and prevention. *Lancet*, *370*(9589), 786–796. https://doi.org/10.1016/S0140-6736(07)61382-8

Welsh, K. G., Rousseau, K., Fisher, G., Bonser, L. R., Bradding, P., Brightling, C. E., … Gaillard, E. A. (2017). MUC5AC and a Glycosylated Variant of MUC5B Alter Mucin Composition in Children With Acute Asthma. *Chest*, *152*(4), 771–779. https://doi.org/10.1016/j.chest.2017.07.001

Whiteman, S. C., Yang, Y., Jones, J. M., & Spiteri, M. A. (2008). FTIR spectroscopic analysis of sputum: preliminary findings on a potential novel diagnostic marker for COPD. *Therapeutic Advances in Respiratory Disease*, *2*(1), 23–31. https://doi.org/10.1177/1753465807087972

Wickström, C., Davies, J. R., Eriksen, G. V., Veerman, E. C. I., & Carlstedt, I. (1998). MUC5B is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and endocervix: identification of glycoforms and C-terminal cleavage. *The Biochemical Journal*.

Wiercigroch, E., Szafraniec, E., Czamara, K., Pacia, M. Z., Majzner, K., Kochan, K., … Malek, K. (2017). Raman and infrared spectroscopy of carbohydrates: A review. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *185*(May), 317–335. https://doi.org/10.1016/j.saa.2017.05.045

Wilkinson, T. M. A., Donaldson, G. C., Hurst, J. R., Seemungal, T. A. R., & Wedzicha, J. A. (2004). Early Therapy Improves Outcomes of Exacerbations of Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine*, *169*(12), 1298–1303. https://doi.org/10.1164/rccm.200310-1443OC

Wolffenbuttel, R. F. (2005). MEMS-based optical mini- and microspectrometers for the visible and infrared spectral range. *Journal of Micromechanics and Microengineering*, *15*(7). https://doi.org/10.1088/0960-1317/15/7/021

World Health Organization. (2008). World Health Statistics. *World Health Organization*. Retrieved from http://www.who.int/whosis/whostat/EN_WHS08_Full.pdf (Retrieved on Nov. 4th 2014).

World Health Organization. (2018). *World Health Statistics 2018- Monitoring Health for the SDG's (Sustainable development goals)*.

Yates, B., Braschi, B., Gray, K., Seal, R., Tweedie, S., & Bruford, E. (2017). Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Research*, *4*(45(D1)), D619-625.

Yuan, S., Hollinger, M., Lachowicz-Scroggins, M. E., Kerr, S. C., Dunican, E. M., Daniel, B. M., … Fahy, J. V. (2015). Oxidation increases mucin polymer cross-links to stiffen airway mucus gels. *Science Translational Medicine*, *7*(276), 276ra27. https://doi.org/10.1126/scitranslmed.3010525

Zabner, J., Smith, J. J., Karp, P. H., Widdicombe, J. H., & Welsh, M. J. (1998). Loss of CFTR chloride channels alters salt absorption by cystic fibrosis airway epithelia in vitro. *Molecular Cell*, *2*(3), 397–403. https://doi.org/10.1016/S1097-2765(00)80284-1

Zarnowiec, P., Lechowicz, L., Czerwonka, G., & Kaca, W. (2015). Fourier Transform Infrared Spectroscopy (FTIR) as a Tool for the Identification and Differentiation of Pathogenic Bacteria. *Current Medicinal Chemistry*, *22*(14), 1710–1718.

Zheng, Z., Qi, Y., Xu, X., Jiang, H., Li, Z., Yang, Q., … Lu, W. (2017). Sputum mucin 1 is increased during the acute phase of chronic obstructive pulmonary disease exacerbation. *Journal of Thoracic Disease*, *9*(7), 1873–1882. https://doi.org/10.21037/jtd.2017.06.63

# 9.0 Appendices

# Appendix 1: Patient Clinical Data, Grouped by Recruiting Study

## A1.1 SPEDIC COPD Patients Details

**Table A1-1: Details of patients recruited through SPEDIC, whose samples were used as part of this thesis: Ref – r known, CAT – COPD assessment test; MRC – Medical Research Council; FEV – Forced expiratory volume; FVC –**

| Patient ID. No. | Sex | Year of Birth | CAT Score | MRC breathlessness score | Ever Smoked - cigarettes | Estimated Pack-years | Date of Lung function | FE (Lit |
|---|---|---|---|---|---|---|---|---|
| **SP01** | M | 1950 | 32 | 4 | No | 0 | 08/22/16 | 0. |
| **SP02** | M | 1938 | 27 | 5 | Yes | 50 | Ref | R |
| **SP03** | F | 1949 | 33 | 5 | Yes | 108 | 09/05/16 | 0. |
| **SP04** | F | 1957 | 29 | 4 | Yes | 40 | ? | 1. |
| **SP05** | M | 1946 | 25 | 3 | Yes | 110 | 11/14/16 | 3. |
| **SP06** | M | 1946 | 41 | 5 | Yes | 55 | 09/01/16 | 1 |
| **SP07** | F | 1956 | 27 | 5 | Yes | 38 | 09/01/16 | 0 |
| **SP08** | F | 1947 | 14 | 3 | Yes | 65 | 01/18/17 | 1. |
| **SP09** | M | 1952 | 33 | 5 | Yes | 45 | 05/01/15 | 0. |
| **SP10** | F | 1967 | 57 | 4 | Yes | 22 | 09/12/16 | 1. |
| **SP11** | M | 1953 | 20 | 4 | Yes | 53 | W/D | W |
| **SP12** | F | 1951 | 38 | 4 | Yes | 99 | Ref | R |
| **SP13** | M | 1950 | 40 | 4 | Yes | 36 | W/D | W |
| **SP14** | F | 1938 | 24 | 4 | Yes | 39 | 2015 | 0. |
| **SP15** | M | 1952 | 33 | 5 | Yes | 104 | W/D | W |
| **SP16** | F | 1947 | 20 | 2 | Yes | 40 | 04/18/15 | 1. |
| **SP17** | M | 1956 | 21 | 3 | Yes | 54 | 10/06/16 | 1. |
| **SP18** | F | 1949 | 32 | 3 | Yes | 49 | 07/02/15 | 1. |
| **SP19** | M | 1941 | 27 | 4 | Yes | 18 | 04/29/16 | 1. |
| **SP20** | F | 1942 | 28 | 4 | Yes | 60 | 01/01/16 | 1. |
| **SP21** | M | 1939 | 37 | 5 | Yes | 35 | 05/01/16 | 1. |
| **SP22** | M | 1978 | 34 | 5 | Yes | 28 | 09/08/15 | 0 |
| **SP23** | M | 1939 | 18 | 4 | Yes | 8 | 01/16/16 | 1. |

| Patient ID. No. | Sex | Year of Birth | CAT Score | MRC breathlessness score | Ever Smoked - cigarettes | Estimated Pack-years | Date of Lung function | FE (Lit |
|---|---|---|---|---|---|---|---|---|
| **SP24** | M | 1949 | 33 | 4 | Yes | 74 | 06/15/17 | 0. |
| **SP25** | M | 1945 | 33 | 5 | Yes | 45 | 10/03/16 | 1. |
| **SP26** | F | 1965 | 0 | 3 | Yes | 6 | 10/07/16 | 1. |
| **SP27** | F | 1959 | 33 | 4 | Yes | 86 | 10/27/15 | 1. |
| **SP28** | F | 1949 | 31 | 5 | Yes | NK | Ref | R |
| **SP29** | F | 1949 | 32 | 5 | Yes | 49 | W/D | W |
| **SP30** | F | 1945 | 13 | 4 | Yes | 56 | 09/2015 | 0. |
| **SP31** | F | 1953 | 36 | 5 | Yes | 100 | 09/26/16 | 0. |
| **SP32** | M | 1957 | 38 | 5 | Yes | 156 | 01/14/16 | 2. |
| **SP33** | M | 1951 | 32 | 5 | Yes | 94 | Ref | R |
| **SP34** | F | 1962 | 32 | 2 | Yes | NK | 12/03/15 | 0. |
| **SP35** | M | 1957 | 30 | 3 | Yes | 17 | Ref | R |
| **SP36** | M | 1956 | 38 | 5 | Yes | NK | 10/01/16 | 1. |
| **SP37** | M | 1937 | 33 | 5 | Yes | 44 | W/D | W |
| **SP38** | F | 1947 | 37 | 5 | Yes | 47 | 08/21/14 | 0. |
| **SP39** | M | 1944 | 15 | 4 | Yes | 62 | 11/11/16 | 1. |
| **SP40** | M | 1937 | 30 | 4 | Yes | 14 | 11/14/16 | 1. |
| **SP41** | F | 1944 | 35 | 5 | No | 0 | W/D | W |
| **SP42** | F | 1936 | 37 | 5 | Yes | 49 | W/D | W |
| **SP43** | M | 1943 | 34 | 4 | Yes | 47 | 11/23/16 | 1. |
| **SP44** | M | 1942 | 33 | 3 | Yes | 90 | 08/01/16 | 1. |
| **SP45** | M | 1933 | 26 | 4 | Yes | 22 | 11/23/16 | 1. |
| **SP46** | F | 1965 | 28 | 4 | No | 0 | Ref | R |
| **SP47** | F | 1966 | 33 | 2 | Yes | 31 | 05/01/16 | 1. |
| **SP48** | M | 1946 | 27 | 4 | Yes | 0 | 09/29/16 | 2. |
| **SP49** | M | 1942 | 25 | 4 | Yes | 52 | 06/17/16 | 1. |
| **SP50** | M | 1944 | 40 | 5 | Yes | 8 | 09/07/16 | 1. |
| **SP51** | M | 1951 | 8 | 2 | Yes | 52 | W/D | W |
| **SP52** | M | 1935 | 31 | 4 | Yes | 362 | 04/05/15 | 2. |
| **SP53** | F | 1953 | 17 | 2 | Yes | 52 | 10/12/16 | 1. |
| **SP54** | M | 1948 | 35 | 5 | Yes | 177 | 02/09/15 | 0. |
| **SP55** | F | 1944 | 25 | 4 | Yes | 11 | 05/24/17 | 0. |

## A1.2 SPEDIC Non-COPD Patient Details

**Table A1-2: Details on non-COPD patients recruited during SPEDIC whose sputum samples were used in this th expiratory volume, FVC – forced vital capacity**

| Patient ID. No. | MRC breathlessness score | Diagnosis | Chronic Bronchitis | Smoking Status | Estimated Pack-years | FEV1 (Litres) | FEV1 (% Predicted) |
|---|---|---|---|---|---|---|---|
| **SPCC003** | 1 | asthma | No | Yes | 5 | 3.75 | 88 |
| **SPCC004** | 2 | N/K | No | No | 0 | 1.2 | . |
| **SPCC007** | 1 | N/K | No | Yes | 41 | 1.96 | 84 |
| **SPCC008** | 2 | Glucoma | No | No | 0 | 1.71 | 135 |
| **SPCC009** | 2 | N/K | No | Yes | 27 | 3.16 | |
| **SPCC010** | 3 | N/K | No | Yes | 42 | 2.3 | |
| **SPCC011** | 3 | Diabetes | No | Yes | 14 | Awaiting | |
| **SPCC012** | 4 | N/K | No | Yes | 1 | n/a | n/a |
| **SPCC014** | 3 | N/K | No | Yes | 0 | 2.08 | |
| **SPCC015** | 1 | N/K | No | No (passive) | 0 | 2.41 | |
| **SPCC016** | 3 | Asthma | No | No | 0 | 1.29 | |
| **SPCC017** | 4 | heart surgery | No | No | 0 | n/a | n/a |
| **SPCC018** | 2 | N/K | No | No | 0 | 2.37 | |
| **SPCC020** | 1 | chronic asthma | Yes | Yes | 14 | 1.78 | |
| **SPCC025** | 2 | asthma | No | No | 0 | | |
| **SPCC026** | 4 | sleep apnoea | No | Yes | 58 | | |
| **SPCC027** | 2 | asthma | No | No | 0 | 1.19 | 62 |
| **SPCC028** | 3 | asthma | No | Yes | 20 | 2.25 | |
| **SPCC029** | 3 | asthma/MI | No | Yes | 14 | | |
| **SPCC031** | 2 | asthma | No | Yes | 14 | | |
| **SPCC034** | 4 | asthma | No | Yes | 24 | 1.66 | 43 |
| **SPCC035** | 4 | asthma | No | Yes | 8 | 0.97 | 23 |
| **SPCC036** | 1 | asthma | No | Yes | 6 | 2.79 | 92 |
| **SPCC037** | 4 | asthma | No | Yes | 150 | 2.02 | 91 |
| **SPCC038** | 2 | asthma | N/K | Yes | 45 | 1.41 | 65 |

| Patient ID. No. | MRC breathlessness score | Diagnosis | Chronic Bronchitis | Smoking Status | Estimated Pack-years | FEV1 (Litres) | FEV1 (% Predicted) |
|---|---|---|---|---|---|---|---|
| **SPCC039** | 1 | ?asthma | | Yes | 20 | 2.88 | 88 |
| **SPCC040** | 3 | asthma/ bronchiectasis | Yes | No (passive) | 0 | 0.82 | 34 |
| **SPCC041** | 1 | asthma/ bronchiectasis | Yes | No | 0 | | 80 |
| **SPCC042** | 4 | asthma | No | No | 0 | 1.87 | 93 |
| **SPCC045** | 1 | asthma | No | Yes | 19 | 3.13 | 100 |
| **SPCC046** | 3 | asthma | No | Yes | 11 | 1.68 | 57 |
| **SPCC047** | 3 | asthma | No | Yes | 45 | 2.42 | 0 |
| **SPCC048** | 4 | asthma | Yes | Yes | 6 | 1.59 | 61 |
| **SPCC049** | 4 | asthma | No | Yes | 20 | 1.47 | |
| **SPCC050** | 3 | asthma | No | No | 0 | 2.21 | 93 |
| **SPCC052** | 2 | asthma | No | Yes | 6 | 2.71 | 106 |
| **SPCC053** | 4 | asthma | No | No | 0 | 1.92 | 74 |
| **SPCC056** | 3 | asthma | No | No (passive) | 0 | 1.93 | |
| **SPCC063** | 1 | asthma | No | Yes | 0 | 3.2 | |
| **SPCC065** | 2 | asthma | No | No | 0 | | |
| **SPCC070** | 4 | Chronic cough | No | No | 0 | 2.14 | |
| **SPCC071** | 4 | asthma | No | Yes | 2 | | |
| **SPCC074** | 4 | Chronic cough | No | Yes | 60 | 0.7 | 36 |
| **SPCC075** | 2 | asthma | No | No | 0 | 2.4 | |
| **SPCC078** | 3 | Cough | No | Yes | 48 | Declined | |
| **SPCC080** | 4 | Asthma | No | Yes | 81 | Never done | |

### A1.3 MEDLUNG Patient Details

**Table A1-3: Details of patients recruited to MEDLUNG whose samples were used during this thesis: NK – not kn cancer; SCLC – small cell lung cancer; FEV – Forced expiratory volume**

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **BC0002** | Birmingham City | M | 12/12/2012 | 68 | Ex | 5 | 2.17 | N |
| **BC0003** | Birmingham City | M | 01/03/2012 | 61 | Ex | 94 | 1.9 | N |
| **BC0005** | Birmingham City | M | 13/03/2012 | 75 | Ex | 80 | 41 | N |
| **BC0007** | Birmingham City | F | 19/06/2012 | 65 | Ex | 0.07 | NK | N |
| **BC0009** | Birmingham City | M | 25/09/2012 | 79 | Ex | 60 | 1.64 | N |
| **BC0010** | Birmingham City | F | 10/02/2012 | 67 | Current | 100 | 1.19 | N |
| **BGHLC0016** | Bronglais | M | 10/10/2012 | 68 | Current | 56 | NK | N |
| **BQE0023** | Birmingham, Queen Elizabeth | F | 09/05/2012 | 89 | Ex | nk | 80 | |
| **BQE0032** | Birmingham, Queen Elizabeth | M | 11/01/2012 | 74 | Ex | 132 | 74.1 | Squa ca |
| **BQE0033** | Birmingham, Queen Elizabeth | M | 11/01/2012 | 73 | Ex | 33 | 77.5 | Adeno |
| **BQE0036** | Birmingham, Queen Elizabeth | M | 12/06/2012 | 59 | Ex | 30 | 78 | Squa ca |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **CW0001** | Chelsea Westminster | M | 10/10/2012 | 69 | Current | 50 | 101 | Adeno |
| **CW0003** | Chelsea Westminster | M | 17/10/2012 | 75 | Ex | nk | nk | Adeno |
| **CW0005** | Chelsea Westminster | F | 18/10/2012 | 51 | | nk | nk | |
| **CW0011** | Chelsea Westminster | M | 16/01/2013 | 62 | Current | 70 | nk | N |
| **CW0012** | Chelsea Westminster | M | 16/01/2013 | 71 | Ex | 40 | nk | N |
| **CW0018** | Chelsea Westminster | M | 21/02/2013 | 69 | Current | 90 | 124 | N |
| **CW0020** | Chelsea Westminster | M | 05/03/2013 | 76 | Current | 120 | 39 | |
| **CW0030** | Chelsea Westminster | M | 10/04/2014 | 58 | Current | 40 | nk | N |
| **CW0036** | Chelsea Westminster | F | 16/04/2014 | 82 | Ex | | nk | |
| **CW0040** | Chelsea Westminster | M | 07/05/2014 | 75 | Ex | nk | nk | N |
| **CW0044** | Chelsea Westminster | M | 12/08/2014 | 46 | Never | 0 | nk | N |
| **CW0046** | Chelsea Westminster | M | 18/09/2014 | 65 | Ex | 100 | nk | N |
| **CW0050** | Chelsea Westminster | M | 25/09/2014 | 90 | Ex | 50 | nk | |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **DE0012** | Derby | F | 22/05/2013 | 84 | Ex | 70 | 107 | N |
| **DE0015** | Derby | M | 07/01/2013 | 80 | Ex | 25 | 70 | N |
| **DE0030** | Derby | M | 10/09/2013 | 83 | Current | 15 | NK | N |
| **DE0036** | Derby | M | 18/12/2013 | 77 | Never | 0 | NK | N |
| **DE0050** | Derby | M | 28/12/2012 | 75 | Ex | NK | 62 | N |
| **DE0051** | Derby | F | 01/09/2013 | 65 | Ex | NK | 63 | N |
| **DE0052** | Derby | M | 13/02/2013 | 62 | Current | 48 | 129 | N |
| **DE0053** | Derby | F | 20/02/2013 | 61 | Current | 40 | 97 | N |
| **DE0054** | Derby | F | 28/02/2013 | 67 | Ex | 41 | NK | N |
| **KMN0002** | Kings Mill, Nottingham | F | 27/06/12 | 64 | Ex | 49 | 64.00% | COP |
| **KMN0005** | Kings Mill, Nottingham | M | 23/07/12 | 64 | Ex | 92 | 46.00% | COP |
| **KMN0006** | Kings Mill, Nottingham | M | 25/07/12 | 76 | Ex | 45 | 50.00% | COP |
| **KMN0010** | Kings Mill, Nottingham | M | 15/08/12 | 59 | Ex | 34 | | COP |
| **KMN0013** | Kings Mill, Nottingham | M | 17/08/12 | 75 | Current | 57 | 64.00% | COP |
| **KMN0014** | Kings Mill, Nottingham | M | 22/08/12 | 50 | Current | 15 | 29.00% | COP |
| **KMN0015** | Kings Mill, Nottingham | M | 22/08/12 | 67 | Current | 24 | | COP |
| **KMN0016** | Kings Mill, Nottingham | M | 22/08/12 | 76 | Ex | 5 | | COP |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **KMN0017** | Kings Mill, Nottingham | M | 22/08/12 | 80 | Ex | 8 | 31.00% | COP |
| **KMN0018** | Kings Mill, Nottingham | F | 22/08/12 | 67 | Ex | 2 | 94.00% | COP |
| **KMN0020** | Kings Mill, Nottingham | M | 22/08/12 | 69 | Ex | 38 | 66.00% | COP |
| **KMN0021** | Kings Mill, Nottingham | M | 22/08/12 | 68 | eX | 165 | 88.00% | COP |
| **KMN0023** | Kings Mill, Nottingham | M | 22/08/12 | 71 | Ex | 65 | | COP |
| **KMN0024** | Kings Mill, Nottingham | M | 22/08/12 | 68 | Ex | 30 | NK | COP |
| **KMN0028** | Kings Mill, Nottingham | M | 18/09/2012 | 84 | Ex | 40 | NK | COP |
| **KMN0030** | Kings Mill, Nottingham | M | 28/09/2012 | 86 | Ex | 24 | 115.00% | COP |
| **KMN0031** | Kings Mill, Nottingham | F | 10/01/2012 | 78 | Current | 25 | 49.00% | COP |
| **KMN0033** | Kings Mill, Nottingham | F | 10/01/2012 | 52 | Current | 90 | 31.00% | COP |
| **KMN0034** | Kings Mill, Nottingham | M | 15/10/2012 | 57 | EX | 10 | 37.00% | COP |
| **KMN0035** | Kings Mill, Nottingham | M | 17/10/2012 | 71 | Ex | 88 | 39.00% | COP |
| **KMN0036** | Kings Mill, Nottingham | F | 17/10/2012 | 70 | Ex | 30 | NK | COP |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **KMN0037** | Kings Mill, Nottingham | M | 17/10/2012 | 76 | Current | 10 | NK | COP |
| **KMN0038** | Kings Mill, Nottingham | M | 17/10/2012 | 73 | Ex | 14 | 57.00% | COP |
| **KMN0039** | Kings Mill, Nottingham | F | 29/10/2012 | 64 | Ex | 50 | 77.00% | COP |
| **KMN0040** | Kings Mill, Nottingham | M | 11/09/2012 | 67 | Ex | 35 | 29.00% | COP |
| **KMN0042** | Kings Mill, Nottingham | M | 11/12/2012 | 68 | Never | 0 | NK | COP |
| **KMN0043** | Kings Mill, Nottingham | M | 11/12/2012 | 81 | EX | 50 | NK | COP |
| **KMN0044** | Kings Mill, Nottingham | M | 19/11/2012 | 80 | Never | 0 | NK | COP |
| **KMN0046** | Kings Mill, Nottingham | M | 21/11/2012 | 77 | Current | 90 | 30.00% | COP |
| **KMN0047** | Kings Mill, Nottingham | M | 21/11/2012 | 81 | Ex | 6 | 76.00% | COP |
| **KMN0049** | Kings Mill, Nottingham | F | 12/05/2012 | 67 | Current | 25 | 50.00% | C |
| **KMN0050** | Kings Mill, Nottingham | F | 12/05/2012 | 67 | Current | 50 | 60.00% | |
| **KMN0052** | Kings Mill, Nottingham | M | 12/10/2012 | 67 | Never | 0 | 47.00% | COP |
| **KMN0053** | Kings Mill, Nottingham | M | 14/12/2012 | 82 | Ex | 45 | 97.00% | COP |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **KMN0055** | Kings Mill, Nottingham | M | 21/12/2012 | 67 | Current | 50 | 44.00% | COP |
| **KMN0056** | Kings Mill, Nottingham | M | 16/01/2013 | 70 | Ex | 50 | 33.00% | COP |
| **KMN0057** | Kings Mill, Nottingham | M | 16/01/2013 | 65 | Current | 40 | 44.00% | COP |
| **KMN0058** | Kings Mill, Nottingham | M | 16/01/2013 | 69 | Current | 90 | NK | COP |
| **KMN0059** | Kings Mill, Nottingham | F | 28/01/2013 | 48 | Current | 20 | 39.00% | COP |
| **KMN0061** | Kings Mill, Nottingham | M | 02/08/2013 | 61 | Current | 23 | NK | COP |
| **KMN0062** | Kings Mill, Nottingham | M | 02/11/2013 | 74 | Current | 60 | 31.00% | COP |
| **KMN0063** | Kings Mill, Nottingham | M | 15/02/2013 | 65 | Never | 0 | 43.00% | COP |
| **KMN0064** | Kings Mill, Nottingham | M | 15/02/2013 | 74 | Ex | 56 | 73.00% | COP |
| **KMN0066** | Kings Mill, Nottingham | F | 18/02/2013 | 77 | Current | 60 | 68.00% | COP |
| **KMN0068** | Kings Mill, Nottingham | F | 25/02/2013 | 66 | Ex | 63 | 34.00% | COP |
| **KMN0069** | Kings Mill, Nottingham | M | 25/02/2013 | 67 | Ex | 100 | 48.00% | COP |
| **KMN0071** | Kings Mill, Nottingham | M | 03/01/2013 | 64 | Ex | 14 | 43.00% | COP |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **KMN0072** | Kings Mill, Nottingham | M | 03/01/2013 | 64 | Ex | 75 | 67.00% | C bron |
| **KMN0074** | Kings Mill, Nottingham | M | 18/03/2013 | 78 | EX | 20 | 65.00% | COP |
| **KMN0075** | Kings Mill, Nottingham | M | 18/03/2013 | 75 | Current | 90 | 70.00% | COP |
| **KMN0077** | Kings Mill, Nottingham | M | 04/08/2013 | 78 | Ex | 42 | 19.00% | COP |
| **KMN0078** | Kings Mill, Nottingham | M | 04/08/2013 | 77 | Ex | 45 | 61.00% | COP |
| **KMN0079** | Kings Mill, Nottingham | M | 04/08/2013 | 74 | Current | 31 | 37.00% | COP |
| **KMN0080** | Kings Mill, Nottingham | M | 04/08/2013 | 60 | Ex | 53 | 49.00% | COP |
| **KMN0081** | Kings Mill, Nottingham | M | 04/12/2013 | 81 | Ex | 54 | 64.00% | COP |
| **KMN0083** | Kings Mill, Nottingham | M | 04/12/2013 | 83 | Ex | 57 | 58.00% | COP |
| **KMN0084** | Kings Mill, Nottingham | M | 04/12/2013 | 68 | Ex | 35 | 35.00% | COP |
| **KMN0085** | Kings Mill, Nottingham | M | 04/12/2013 | 65 | Never | 0 | 68.00% | COP |
| **KMN0086** | Kings Mill, Nottingham | F | 15/04/2013 | 77 | Ex | 34 | 58.00% | COP |
| **KMN0088** | Kings Mill, Nottingham | F | 17/04/2013 | 64 | Current | 25 | 73.00% | COP |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **KMN0089** | Kings Mill, Nottingham | M | 17/04/2013 | 75 | Ex | PIPE | NK | COP |
| **KMN0091** | Kings Mill, Nottingham | M | 17/04/2013 | 66 | Ex | 50 | 51.00% | COP |
| **KMN0093** | Kings Mill, Nottingham | F | 17/04/2013 | 68 | Current | 11 | 67.00% | COP |
| **KMN0094** | Kings Mill, Nottingham | F | 17/04/2013 | 72 | Ex | 23 | 55.00% | COP |
| **KMN0095** | Kings Mill, Nottingham | M | 19/04/2013 | 83 | Ex | 40 | 46.00% | COP |
| **KMN0097** | Kings Mill, Nottingham | M | 22/04/2013 | 72 | Ex | 54 | 58.00% | COP |
| **KMN0098** | Kings Mill, Nottingham | M | 22/04/2013 | 73 | Ex | 75 | 39.00% | COP |
| **KMN0099** | Kings Mill, Nottingham | F | 23/04/2013 | 58 | Ex | 40 | 58.00% | COP |
| **KMN0102** | Kings Mill, Nottingham | F | 13/05/2013 | 57 | Ex | 30 | 42.00% | COP |
| **KMN0104** | Kings Mill, Nottingham | M | 13/05/2013 | 76 | EX | 30 | 29.00% | COP |
| **KMN0105** | Kings Mill, Nottingham | F | 13/05/2013 | 70 | EX | 49 | 27.00% | COP |
| **KMN0106** | Kings Mill, Nottingham | F | 13/05/2013 | 50 | Current | 20 | 20.00% | COP |
| **KMN0110** | Kings Mill, Nottingham | F | 20/05/2013 | 57 | Ex | 35 | 48.00% | COP |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **KMN0111** | Kings Mill, Nottingham | M | 20/05/2013 | 56 | Ex | 25 | 38.00% | COP |
| **KMN0112** | Kings Mill, Nottingham | F | 20/05/2013 | 71 | Current | 40 | 48.00% | COP |
| **KMN0114** | Kings Mill, Nottingham | M | 22/05/2013 | 72 | Ex | 15 | 76.00% | COP |
| **KMN0116** | Kings Mill, Nottingham | M | 22/05/2013 | 65 | EX | 175 | 20.00% | COP |
| **KMN0117** | Kings Mill, Nottingham | F | 29/05/2013 | 76 | Ex | 20 | 59.00% | COP |
| **KMN0118** | Kings Mill, Nottingham | M | 06/03/2013 | 63 | Ex | 45 | 60.00% | COP |
| **KMN0119** | Kings Mill, Nottingham | M | 06/10/2013 | 58 | Ex | 66 | 40.00% | COP |
| **LC0161** | Prince Phillip, Llanelli | F | 06/07/2012 | 61 | Current | Nk | 87.00% | N |
| **MK0007** | Milton Keynes | F | 09/10/2012 | 65 | Never | 0 | NK | Non Ca |
| **MK0010** | Milton Keynes | F | 14/11/2012 | 56 | EX | 20 | 91.00% | Metast |
| **MK0011** | Milton Keynes | F | 19/11/2012 | 68 | Ex | 42 | Nk | Non Carc Br |
| **MK0012** | Milton Keynes | M | 20/11/2012 | 80 | Ex | 43 | 72.00% | diffe Squa Lun |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **MK0014** | Milton Keynes | M | 18/12/2012 | 79 | Ex | 41 | nk | Me Adeno |
| **MK0019** | Milton Keynes | M | 16/04/2013 | 49 | Ex | 25 | nk | Small C |
| **MK0020** | Milton Keynes | M | 26/04/2013 | 63 | Current | 45 | 75.00% | Meta Smal C |
| **MK0021** | Milton Keynes | M | 30/04/2013 | 73 | Ex | 35 | 66.00% | M Adeno diffe |
| **MK0028** | Milton Keynes | F | 02/07/2013 | 69 | Ex | 30 | 43.00% | Lun |
| **MK0029** | Milton Keynes | F | 27/08/2013 | 69 | Ex | 58 | 65.00% | I Diffe Adeno o |
| **MK0030** | Milton Keynes | M | 27/08/2013 | 64 | Current | 100 | 65.00% | I Diffe Squa Ca |
| **MK0033** | Milton Keynes | M | 27/09/2013 | 70 | Current | 60 | 93.00% | Me Adeno of t |
| **MK0034** | Milton Keynes | F | 01/10/2013 | 64 | Ex | 15 | nk | Small C |
| **MK0035** | Milton Keynes | M | 01/10/2013 | 65 | EX | 20 | nk | I Diffe Squa Ca |
| **MK0036** | Milton Keynes | M | 08/10/2013 | 66 | EX | 35 | nk | Me Squa Ca |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **MK0038** | Milton Keynes | M | 22/10/2013 | 49 | Current | 35 | nk | Meta c |
| **MK0039** | Milton Keynes | F | 05/11/2013 | 55 | Current | 40 | 43.00% | Adeno c |
| **MK0041** | Milton Keynes | F | 12/03/2014 | 59 | Ex | 40 | 64.00% | Squa Ca |
| **MK0042** | Milton Keynes | M | 08/04/2014 | 60 | Ex | 45 | 54.00% | Squa Ca |
| **MK0043** | Milton Keynes | M | 17/06/2014 | 65 | Ex | 30 | 106.00% | Invasiv cell c (squa |
| **MK0044** | Milton Keynes | F | 18/06/2014 | 67 | Ex | 20 | 55.00% | Small C |
| **MK0047** | Milton Keynes | M | 29/07/2014 | 56 | Current | 40 | 63.00% | Metast ell C |
| **MK0054** | Milton Keynes | F | 15/10/2014 | 71 | Ex | 20 | 74% | N |
| **MK0055** | Milton Keynes | M | 19/11/2014 | 83 | Ex | 20 | nk | Adeno c |
| **MK0056** | Milton Keynes | M | 09/12/2014 | 81 | Ex | 50 | 61% | N |
| **MK0058** | Milton Keynes | F | 22/01/2015 | 62 | Ex | 32 | 57 | N |
| **MK0060** | Milton Keynes | M | 14/04/2015 | 72 | Ex | nk | 88% | N |
| **MK0061** | Milton Keynes | F | 15/04/2015 | 59 | Current | 60 | 84% | Squa Ca |
| **MK0062** | Milton Keynes | M | 15/04/2015 | 68 | Current | 50 | 40% | N |
| **MK0063** | Milton Keynes | F | 15/04/2015 | 51 | Current | 41 | nk | N |
| **MK0064** | Milton Keynes | M | 22/04/2015 | 56 | Ex | nk | nk | N |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **MK0067** | Milton Keynes | M | 15/07/2015 | 74 | ex | 17 | nk | N |
| **MK0069** | Milton Keynes | F | 09/03/2015 | 66 | Ex | 40 | nk | N |
| **MK0070** | Milton Keynes | F | 09/04/2015 | 72 | Ex | 38 | nk | N |
| **NCW0001** | New Cross, Wolverhampton | F | 24/01/2013 | 44 | EX | 20 | 63 | Adeno |
| **NCW0002** | New Cross, Wolverhampton | M | 29/01/2013 | 71 | EX | 20 | 69 | Squa Ca |
| **NCW0004** | New Cross, Wolverhampton | F | 02/06/2013 | 35 | EX | NK | NK | Adeno |
| **NCW0005** | New Cross, Wolverhampton | M | 15/02/2013 | 44 | EX | 16 | 66 | Adeno |
| **NCW0007** | New Cross, Wolverhampton | M | 26/02/2013 | 63 | EX | 61 | 57 | Squa |
| **NCW0009** | New Cross, Wolverhampton | M | 03/01/2013 | 55 | EX | 60 | 42 | Smal c |
| **NCW0010** | New Cross, Wolverhampton | F | 03/06/2013 | 49 | EX | 39 | 76 | Sq |
| **NCW0011** | New Cross, Wolverhampton | M | 13.03.2013 | 48 | NEVER | 0 | 48 | Adeno |
| **NCW0013** | New Cross, Wolverhampton | F | 05/02/2013 | 71 | EX | 20 | 97 | Smal c |
| **NCW0017** | New Cross, Wolverhampton | M | 07/04/2013 | 60 | EX | 40 | NK | Lun |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| | n | | | | | | | |
| **NCW0018** | New Cross, Wolverhampton | F | 17/07/2013 | 72 | never | 0 | 72 | me adeno |
| **NCW0019** | New Cross, Wolverhampton | F | 30/7/2013 | 77 | EX | 10 | 73 | N |
| **NCW0020** | New Cross, Wolverhampton | M | 31/07/2013 | 59 | current | 18 | 53 | Squa c |
| **NCW0022** | New Cross, Wolverhampton | M | 09/12/2013 | 73 | EX | 50 | 89 | Squa |
| **NCW0024** | New Cross, Wolverhampton | M | 13/09/2013 | 73 | current | 55 | NK | Squa Ca |
| **NCW0026** | New Cross, Wolverhampton | M | 17/09/2013 | 76 | EX | NK | 97 | Squa Ca |
| **NCW0027** | New Cross, Wolverhampton | M | 17/09/2013 | 71 | EX | 50 | NK | Squa Ca |
| **NCW0028** | New Cross, Wolverhampton | M | 24/09/2013 | 53 | current | 40 | 63 | N |
| **NCW0029** | New Cross, Wolverhampton | F | 24/09/2013 | 60 | current | 30 | 74 | N |
| **NCW0034** | New Cross, Wolverhampton | F | 29/10/2013 | 64 | current | 40 | | Lung |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **NCW0035** | New Cross, Wolverhampton | F | 29/10/2013 | 79 | EX | 20 | 42 | Adeno |
| **NCW0036** | New Cross, Wolverhampton | M | 11/04/2013 | 83 | EX | 5 | 64 | Squa |
| **NCW0039** | New Cross, Wolverhampton | M | 19/11/2013 | 78 | EX | 70 | 73 | Squa |
| **NCW0040** | New Cross, Wolverhampton | M | 12/04/2013 | 72 | EX | 45 | 59 | Squa |
| **NCW0044** | New Cross, Wolverhampton | F | 31/12/2013 | 42 | nk | nk | NK | Sm Ca |
| **NCW0049** | New Cross, Wolverhampton | F | 21/01/2014 | 51 | current | 20 | 99 | Adeno |
| **NCW0051** | New Cross, Wolverhampton | F | 02/04/2013 | 73 | EX | NK | NK | Adeno |
| **NCW0052** | New Cross, Wolverhampton | M | 02/06/2014 | 82 | EX | 40 | 79 | N |
| **NCW0053** | New Cross, Wolverhampton | F | 02/10/2014 | 64 | EX | 5 | 103 | Adeno |
| **NCW0054** | New Cross, Wolverhampton | M | 02/10/2014 | 73 | current | 60 | NK | Sm Ca |
| **NCW0055** | New Cross, Wolverhampton | M | 18/02/2014 | 71 | EX | 25 | 98 | Adeno |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **NCW0057** | New Cross, Wolverhampton | M | 03/11/2014 | 59 | EX | NK | 77 | Adeno |
| **NCW0058** | New Cross, Wolverhampton | F | 24/03/2014 | 67 | EX | 17 | NK | Squa |
| **NCW0062** | New Cross, Wolverhampton | M | 04/07/2014 | 72 | Current | 15 | NK | Squa Ca |
| **NCW0063** | New Cross, Wolverhampton | M | 15/04/2014 | 78 | Never | 0 | 71 | Adeno |
| **NCW0064** | New Cross, Wolverhampton | M | 16/04/2014 | 58 | current | 20 | NK | Squa |
| **NCW0065** | New Cross, Wolverhampton | M | 29/04/2014 | 74 | EX | NK | NK | Sm |
| **NCW0070** | New Cross, Wolverhampton | M | 21/05/2014 | 72 | EX | NK | NK | Adeno |
| **NCW0073** | New Cross, Wolverhampton | M | 16/06/2014 | 78 | Current | 40 | NK | Non |
| **NCW0074** | New Cross, Wolverhampton | M | 16/06/2014 | 84 | EX | 40 | NK | Adeno |
| **NCW0075** | New Cross, Wolverhampton | M | 19/06/2014 | 90 | EX | 15 | 55 | Sm |
| **NCW0076** | New Cross, Wolverhampton | M | 01/07/2014 | 77 | EX | 50 | 81 | Adeno |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **NCW0077** | New Cross, Wolverhampton | M | 07/07/2014 | 59 | EX | 40 | NK | Squa |
| **NCW0078** | New Cross, Wolverhampton | M | 10/07/2014 | 60 | EX | NK | 28 | Adeno |
| **NCW0080** | New Cross, Wolverhampton | F | 11/07/2014 | 71 | Current | 28 | NK | Sm |
| **NCW0081** | New Cross, Wolverhampton | M | 15/07/2014 | 68 | Current | 108 | 56 | Squa |
| **NCW0082** | New Cross, Wolverhampton | M | 22/07/2014 | 64 | Current | 52 | 96 | Adeno |
| **NCW0085** | New Cross, Wolverhampton | M | 11/08/2014 | 78 | EX | 40 | 58 | Squa |
| **NCW0086** | New Cross, Wolverhampton | M | 12/08/2014 | 54 | EX | 20 | 86 | Adeno |
| **NCW0087** | New Cross, Wolverhampton | M | 18/08/2014 | 66 | Current | 5 a day | 79 | Squa |
| **NCW0088** | New Cross, Wolverhampton | M | 21/08/2014 | 66 | Never | 0 | NK | Adeno |
| **NCW0089** | New Cross, Wolverhampton | M | 21/08/2014 | 70 | Never | 0 | NK | Adeno |
| **NCW0092** | New Cross, Wolverhampton | F | 04/09/2014 | 68 | Current | NK | 49 | Adeno |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **NCW0094** | New Cross, Wolverhampton | M | 08/09/2014 | 66 | EX | 20 | NK | Adeno |
| **NCW0095** | New Cross, Wolverhampton | F | 09/09/2014 | 70 | EX | NK | ND | Non |
| **NCW0097** | New Cross, Wolverhampton | M | 30/09/2014 | 68 | Current | NK | NK | Adeno |
| **NCW0098** | New Cross, Wolverhampton | F | 02/10/2014 | 71 | EX | 50 | NK | Squa |
| **NCW0099** | New Cross, Wolverhampton | M | 10/10/2014 | 70 | EX | 40 | ND | Adeno |
| **NCW0102** | New Cross, Wolverhampton | M | 24/10/2014 | 71 | Ex | NK | NK | Squa |
| **NCW0103** | New Cross, Wolverhampton | M | 24/10/2014 | 77 | Ex | nk | 33 | Adeno |
| **NCW0108** | New Cross, Wolverhampton | M | 01/12/2014 | 76 | EX | 20 | 36 | Sm |
| **NCW0110** | New Cross, Wolverhampton | M | 23/12/2014 | 79 | EX | 47 | 79 | Adeno |
| **NCW0113** | New Cross, Wolverhampton | F | 05/02/2015 | 61 | Current | 60 | NK | Non |
| **NCW0117** | New Cross, Wolverhampton | F | 05/03/2015 | 68 | EX | 40 | NK | Adeno |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **NCW0118** | New Cross, Wolverhampton | F | 09/03/2015 | 70 | Current | 15 | nk | Squa |
| **NCW0121** | New Cross, Wolverhampton | F | 02/06/2015 | 76 | EX | 8 | 123 | Adeno |
| **NCW0122** | New Cross, Wolverhampton | F | 09/06/2015 | 82 | EX | 2 | 45 | Adeno |
| **NCW0123** | New Cross, Wolverhampton | M | 20/10/2015 | 69 | EX | 70 | 54 | Adeno |
| **NCW0124** | New Cross, Wolverhampton | F | 29/10/2015 | 61 | EX | 135 | 82 | Adeno |
| **NCW0125** | New Cross, Wolverhampton | F | 29/10/2015 | 42 | EX | 35 | 48 | Sm |
| **NCW0126** | New Cross, Wolverhampton | M | 11/10/2015 | 48 | EX | NK | 72 | Adeno |
| **NCW0127** | New Cross, Wolverhampton | M | 23/11/2015 | 69 | EX | 30 | 62 | Sm |
| **NCW0131** | New Cross, Wolverhampton | F | 22/12/2015 | 72 | Current | 50 | 88 | Sm |
| **NCW0132** | New Cross, Wolverhampton | F | 14/01/2016 | 74 | EX | 50 | 95 | Sm |
| **NCW0133** | New Cross, Wolverhampton | M | 28/01/2016 | 74 | EX | 20 | 73 | Adeno |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **SGUL01** | Saint George's University Hospital, London | F | 15/04/2013 | 76 | Ex | 1.5 | NK | C exac |
| **SGUL02** | Saint George's University Hospital | F | 17/4/2013 | 86 | Ex | 8 | 72 | C exac |
| **SGUL03** | Saint George's University Hospital | M | 18/04/2013 | 90 | ex | 56 | 45.00% | C exac |
| **SGUL04** | Saint George's University Hospital | M | 23-04-2013 | NK | Ex | 30 | NK | pneu an Exac |
| **SGUL05** | Saint George's University Hospital | F | 24-04-2013 | 85 | Never | 0 | NK | C exacer |
| **SGUL06** | Saint George's University Hospital | M | 25-04-2013 | NK | Current | NK | 31.00% | C exac |
| **SGUL08** | Saint George's University Hospital | M | 26-04-2013 | 60 | EX | 40 | 17.00% | IE |
| **SGUL09** | Saint George's University Hospital | F | 29-4-2013 | NK | EX | 79 | NK | C exac |
| **SGUL10** | Saint George's University Hospital | M | 30-04-2013 | 55 | EX | 112 | NK | C exac |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **SGUL11** | Saint George's University Hospital | F | 05/02/2013 | 76 | EX | 52 | NK | exac |
| **SGUL12** | Saint George's University Hospital | M | 05/08/2013 | 70 | Ex | 56 | 47.00% | Exace |
| **SGUL13** | Saint George's University Hospital | M | 14-5-2013 | 50 | EX | 78 | 17% (Feb 2012) | exac |
| **SGUL14** | Saint George's University Hospital | M | 30/05/2013 | 57 | EX | 50 | 65% (Nov 2012) | CO exac Recru out-pa |
| **SGUL16** | Saint George's University Hospital | M | 06/05/2013 | 54 | current | 72 | 61% (dec 2012) | CO exac Recru out-pa |
| **SGUL17** | Saint George's University Hospital | M | 06/10/2013 | 53 | current | 45 | 49% (June 2013) | CO exac Recru out-pa |
| **SGUL18** | Saint George's University Hospital | F | 13/06/2013 | 58 | current | 33 | 58% March 2011 | CO exac Recru Che |
| **SGUL19** | Saint George's University Hospital | M | 17/07/2013 | 72 | EX | 20 | 35% March 2012 | IE |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **SGUL20** | Saint George's University Hospital | M | 17/07/2013 | 47 | Current | 30 | NK | |
| **SGUL21** | Saint George's University Hospital | M | 26/06/2013 | 64 | | 60 | 22% 13/10/2011 | In exace |
| **SGUL22** | Saint George's University Hospital | F | 25-7-2013 | 57 | EX | 81 | 60% pred 12/2012 | IE |
| **SGUL23** | Saint George's University Hospital | M | 16-10-2013 | 72 | Ex | 69 | 77% 16-10-2013 | CO exac recru Che |
| **SGUL24** | Saint George's University Hospital | F | 14/11/2013 | 58 | current | 68 | | In Exace |
| **SGUL25** | Saint George's University Hospital | M | 10/09/2013 | 78 | EX ( stopped 50 years ago) | 60 | 135.00% (22-08-2013) | exace hae seco bron |
| **SGUL26** | Saint George's University Hospital | M | 10/09/2013 | 55 | current | 32 | NK | Collap metast |
| **SGUL27** | Saint George's University Hospital | M | 19-11-2013 | 61 | current | 59 | NK | Asp Pneu |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **SGUL28** | Saint George's University Hospital | F | 20-11-2013 | 54 | ex | 42 | NK | Co ac pn |
| **SGUL29** | Saint George's University Hospital | M | 21-1-2014 | 58 | ex | 42 | 62% 17-8-2009 | Non exace |
| **SGUL30** | Saint George's University Hospital | F | 03/10/2014 | 53 | EX | 40 | 48% Sept 2013 | In exace |
| **SGUL32** | Saint George's University Hospital | M | 18-03-2014 | 71 | ex | 76 | 61% July 2011 | Co ac pn |
| **SGUL33** | Saint George's University Hospital | F | 18-3-2014 | 58 | ex | 56 | NK | IE |
| **SGUL34** | Saint George's University Hospital | M | 23-4-2014 | NK | ex | 50 | NK | IE |
| **SGUL35** | Saint George's University Hospital | F | 24-4-2014 | 63 | ex | 39 | 65% July 2012 | IE |
| **SGUL36** | Saint George's University Hospital | M | 01/05/2014 | 75 | ex | 50 | NK | IE |
| **SGUL37** | Saint George's University Hospital | F | 02/05/2014 | 64 | ex | 29 | NK | IE |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **SGUL38** | Saint George's University Hospital | M | 02/05/2014 | 84 | ex | 15 | 79% 3-2-2012 | IE |
| **TGW0011** | The Great Western, Swindon | M | 22/8/2011 | 67 | Current | 30 | 2.62 -102% | N |
| **TGW0043** | The Great Western, Swindon | M | 28/08/2012 | 68 | Ex | 80 | 82.00% | N |
| **TGW0044** | The Great Western, Swindon | M | 31/10/2012 | 64 | Current | 100 | 49.00% | |
| **TGW0046** | The Great Western, Swindon | M | 11/07/2012 | 46 | nk | nk | nk | |
| **TGW0047** | The Great Western, Swindon | M | 11/07/2012 | 58 | Ex | 40 | 82.00% | N |
| **TGW0048** | The Great Western, Swindon | M | 21/11/2012 | 64 | Ex | 100 | 93.00% | N |
| **TGW0049** | The Great Western, Swindon | F | 21/11/2012 | 46 | Ex | nk | 63.00% | |
| **TGW0050** | The Great Western, Swindon | F | 27/11/2012 | 69 | Ex | 40 | 86.00% | N |
| **TGW0051** | The Great Western, Swindon | M | 28/11/2012 | 53 | Never | 0 | nk | N |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TGW0053** | The Great Western, Swindon | F | 12/06/2012 | 68 | Ex | nk | nk | N |
| **TGW0054** | The Great Western, Swindon | M | 12/06/2012 | 48 | Current | nk | 76.00% | N |
| **TGW0055** | The Great Western, Swindon | F | 12/12/2012 | 65 | Ex | 30 | 66.00% | N |
| **TGW0057** | The Great Western, Swindon | F | 01/02/2013 | 67 | Current | 55 | 79.00% | N |
| **TGW0059** | The Great Western, Swindon | M | 01/10/2013 | 77 | Ex | nk | 78.00% | N |
| **TGW0062** | The Great Western, Swindon | F | 17/1/2013 | 65 | Current | nk | 90.00% | N |
| **TGW0063** | The Great Western, Swindon | F | 17/1/2013 | 70 | Never | 0 | 110.00% | N |
| **TGW0064** | The Great Western, Swindon | M | 23/1/2013 | 79 | Ex | 40 | 55.00% | |
| **TGW0065** | The Great Western, Swindon | F | 31/1/2013 | 60 | Ex | 50 | 47.00% | |
| **TGW0067** | The Great Western, Swindon | F | 02/07/2013 | 65 | Ex | 10 | 113.00% | N |
| **TGW0068** | The Great Western, Swindon | F | 14/2/2013 | 75 | Never | 0 | nk | N |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TGW0069** | The Great Western, Swindon | M | 21/2/2013 | 83 | Ex | 60 | 69.00% | N |
| **TGW0070** | The Great Western, Swindon | M | 26/2/2013 | 73 | Ex | 60 | 40.00% | N |
| **TGW0073** | The Great Western, Swindon | M | 03/05/2013 | 66 | Ex | 80 | 62.00% | N |
| **TGW0075** | The Great Western, Swindon | M | 03/12/2013 | 64 | Ex | nk | 91.00% | |
| **TGW0080** | The Great Western, Swindon | M | 18/6/2013 | 74 | Ex | nk | 63.00% | N |
| **TGW0082** | The Great Western, Swindon | M | 25/6/2013 | 77 | Current | nk | 65.00% | N |
| **TGW0083** | The Great Western, Swindon | M | 27/6/2013 | 77 | Ex | nk | nk | N |
| **TGW0085** | The Great Western, Swindon | F | 27/6/2013 | 67 | Never | 0 | nk | N |
| **TGW0091** | The Great Western, Swindon | F | 07/02/2013 | 74 | Ex | 20 | 76.00% | N |
| **TGW0092** | The Great Western, Swindon | F | 07/02/2013 | 60 | Current | 40 | nk | N |
| **TGW0095** | The Great Western, Swindon | NK | 17/7/2013 | NK | Ex | nk | 76.00% | |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TGW0097** | The Great Western, Swindon | M | 29/7/2013 | 69 | Current | 50 | nk | |
| **TGW0099** | The Great Western, Swindon | M | 28/8/2013 | 57 | Current | 40 | 45.00% | |
| **TGW0100** | The Great Western, Swindon | M | 17/9/2013 | 67 | Current | 25 | 52.00% | |
| **TGW0101** | The Great Western, Swindon | F | 20/9/2013 | 59 | Ex | nk | nk | |
| **TGW0106** | The Great Western, Swindon | M | 10/03/2013 | 58 | Current | 40 | 111.00% | N |
| **TGW0107** | The Great Western, Swindon | F | 10/08/2013 | 78 | Ex | 40 | 86.00% | N |
| **TGW0112** | The Great Western, Swindon | M | 16/10/2013 | 73 | Ex | nk | nk | N |
| **TGW0113** | The Great Western, Swindon | M | 24/10/2013 | 86 | Ex | 15 | nk | |
| **TGW0114** | The Great Western, Swindon | M | 31/10/2013 | 60 | Current | 30 | 74.00% | N |
| **TGW0118** | The Great Western, Swindon | M | 19/11/2013 | 73 | Current | 100 | 58.00% | |
| **TGW0121** | The Great Western, Swindon | M | 21/11/2013 | 62 | Current | 40 | nk | N |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TGW0125** | The Great Western, Swindon | F | 12/04/2013 | 74 | Ex | 20 | 127.00% | N |
| **TGW0126** | The Great Western, Swindon | M | 12/05/2013 | 74 | Current | 120 | 106.00% | |
| **TGW0127** | The Great Western, Swindon | F | 12/06/2013 | 66 | Ex | 100 | 69.00% | |
| **TGW0128** | The Great Western, Swindon | F | 01/06/2014 | 62 | Current | 40 | 100.00% | N |
| **TGW0129** | The Great Western, Swindon | M | 16/1/2014 | 68 | Ex | 20 | 95.00% | N |
| **TGW0130** | The Great Western, Swindon | M | 16/1/2014 | 83 | Ex | nk | nk | |
| **TGW0132** | The Great Western, Swindon | F | 20/06/2014 | 65 | Current | 20 | 96.00% | N |
| **TGW0134** | The Great Western, Swindon | M | 15/07/2014 | 63 | Current | 50 | nk | |
| **TGW0136** | The Great Western, Swindon | M | 18/07/2014 | 65 | Current | 50 | nk | |
| **TGW0137** | The Great Western, Swindon | F | 08/08/2014 | 67 | Current | 50 | nk | N |
| **TGW0140** | The Great Western, Swindon | M | 22/08/2014 | 69 | nk | nk | nk | N |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TGW0141** | The Great Western, Swindon | F | 22/08/2014 | 71 | Current | 20 | nk | |
| **TGW0142** | The Great Western, Swindon | M | 26/08/2014 | 70 | Ex | nk | 89.00% | N |
| **TGW0147** | The Great Western, Swindon | M | 04/11/2014 | 75 | Ex | 30 | 35% | N |
| **TGW0148** | The Great Western, Swindon | F | 19/11/2014 | 69 | Current | 50 | 116% | |
| **TGW0149** | The Great Western, Swindon | M | 19/11/2014 | 79 | Ex | nk | 73% | |
| **TGW0153** | The Great Western, Swindon | M | 16/01/2015 | 69 | Ex | 30 | nk | |
| **TGW0155** | The Great Western, Swindon | M | 23/02/2015 | 67 | Ex | nk | nk | N |
| **TGW0156** | The Great Western, Swindon | M | 05/03/2015 | 73 | Ex | 140 | 81 | N |
| **TGW0159** | The Great Western, Swindon | M | 09/06/2015 | 66 | Ex | 40 | 91 | N |
| **TGW0160** | The Great Western, Swindon | F | 09/06/2015 | 69 | Ex | 50 | 60 | |
| **TGW0161** | The Great Western, Swindon | M | 25/06/2015 | 74 | Ex | 75 | nk | N |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TGW0164** | The Great Western, Swindon | F | 04/08/2015 | 72 | Ex | nk | nk | N |
| **TGW0167** | The Great Western, Swindon | F | 19/08/2015 | 78 | Ex | nk | nk | N |
| **TGW0168** | The Great Western, Swindon | M | 01/09/2015 | 57 | Current | 40 | nk | N |
| **TGW0170** | The Great Western, Swindon | F | 10/09/2015 | 60 | Ex | 23 | nk | N |
| **TGW0172** | The Great Western, Swindon | F | 08/10/2015 | 64 | Ex | 40 | nk | |
| **TGW0173** | The Great Western, Swindon | M | 12/10/2015 | 69 | Ex | 25 | nk | |
| **TGW0174** | The Great Western, Swindon | F | 20/10/2015 | 74 | Ex | 5 | nk | |
| **TR01** | Prince Phillip, Llanelli | M | 22/11/2012 | nk | Current | nk | 36 | C Exa |
| **TR02** | Prince Phillip, Llanelli | F | 22/11/2012 | nk | Ex | 30 | 38 | C Exa |
| **TR04** | Prince Phillip, Llanelli | M | 12/12/2012 | NK | Current | 60 | 41 | C Exa |
| **TR05** | Prince Phillip, Llanelli | F | 12/12/2012 | NK | Never | 0 | 61 | C Exa |
| **TR06** | Prince Phillip, Llanelli | M | 23/01/2013 | NK | Ex | nk | 52 | C Exa |

| Patient ID | Recruiting Hospital | Patient Sex | Date Recruited | Age at Diagnosis | Smoking Status | Pack-years | FEV1 (%) of Predicted | Fina Di |
|---|---|---|---|---|---|---|---|---|
| **TR08** | Prince Phillip, Llanelli | M | 20/02/2013 | NK | Ex | 25 | 48 | C Exa |
| **TR09** | Prince Phillip, Llanelli | F | 25/02/2013 | NK | Ex | 34 | 69 | C Exa |
| **TR11** | Prince Phillip, Llanelli | M | 07/03/2013 | NK | Ex | nk | 72 | C Exa |
| **TR12** | Prince Phillip, Llanelli | F | 11/03/2013 | NK | Ex | 40 | 47 | C Exa |
| **TR13** | Prince Phillip, Llanelli | F | 11/03/2013 | NK | Ex | 40 | 48 | COP |
| **TR16** | Prince Phillip, Llanelli | F | 26/03/2013 | NK | Ex | 40 | 49 | COP |
| **TR17** | Prince Phillip, Llanelli | F | 10/04/2013 | NK | Current | 25 | 71 | COP |
| **TR18** | Prince Phillip, Llanelli | M | 10/04/2013 | NK | Current | 50 | 47 | C Exa |
| **TR20** | Prince Phillip, Llanelli | F | 03/05/2013 | NK | Ex | 40 | 77 | COP |
| **TR21** | Prince Phillip, Llanelli | M | 03/05/2013 | NK | Current | 50 | 52 | COP |
| **TR22** | Prince Phillip, Llanelli | M | 28/05/2015 | NK | Ex | 30 | 38 | C Exa |

## A 1.4 Cystic Fibrosis Inhaled Therapeutics Patient List

**Table A1-4: Treatment phase assignments for each CF patient.**

| | Phase I | Phase II |
|---|---|---|
| **20801-002** | OligoG | Placebo |
| **27601-005** | Placebo | OligoG |
| **27604-003** | OligoG | Placebo |
| **27605-003** | OligoG | Placebo |
| **27606-003** | OligoG | Placebo |
| **57801-003** | Placebo | OligoG |
| **75202-001** | OligoG | Placebo |
| **75202-003** | OligoG | Placebo |
| **82601-002** | Placebo | OligoG |
| **82602-002** | OligoG | Placebo |
| **82602-005** | OligoG | Placebo |
| **82602-006** | OligoG | Placebo |
| **82603-003** | OligoG | Placebo |
| **82604-008** | Placebo | OligoG |
| **82604-009** | Placebo | OligoG |
| **82606-003** | Placebo | OligoG |

# Appendix 2: Medlung Patient Spectra

## A2.1 COPD Patient Sputum Average Spectra

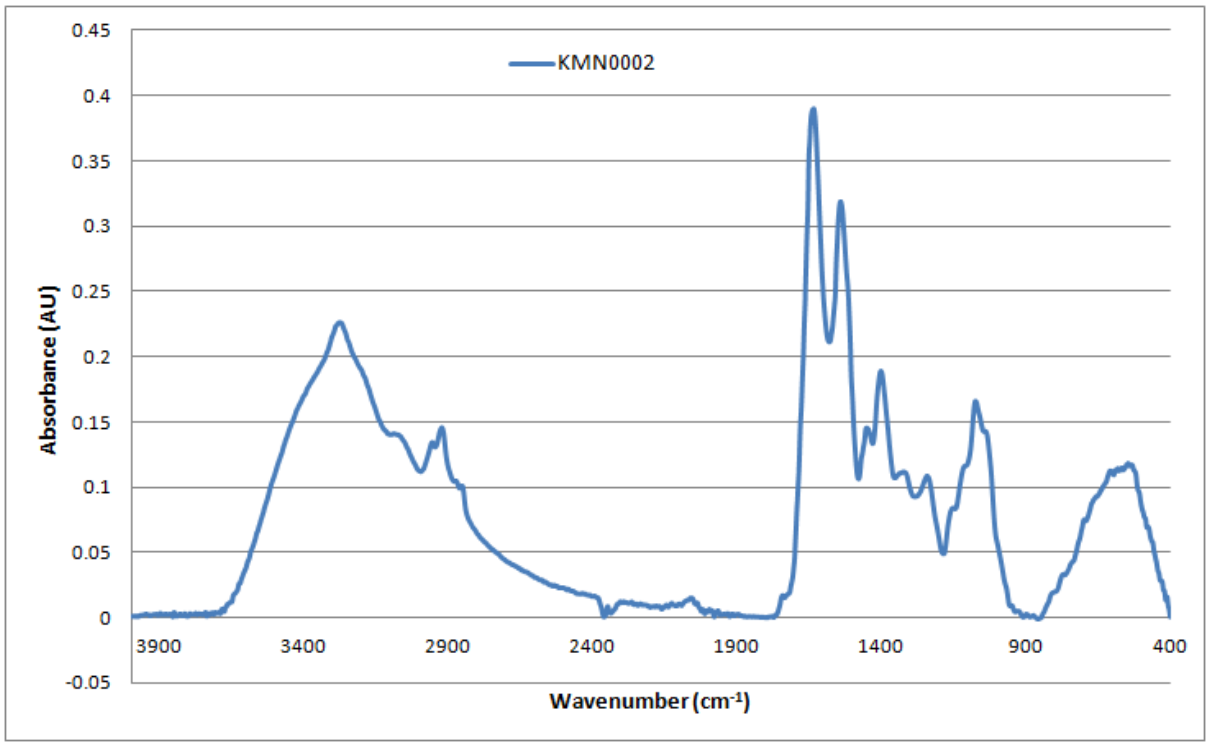


**Figure A2-1: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0002 from 4000-400cm$^{-1}$**

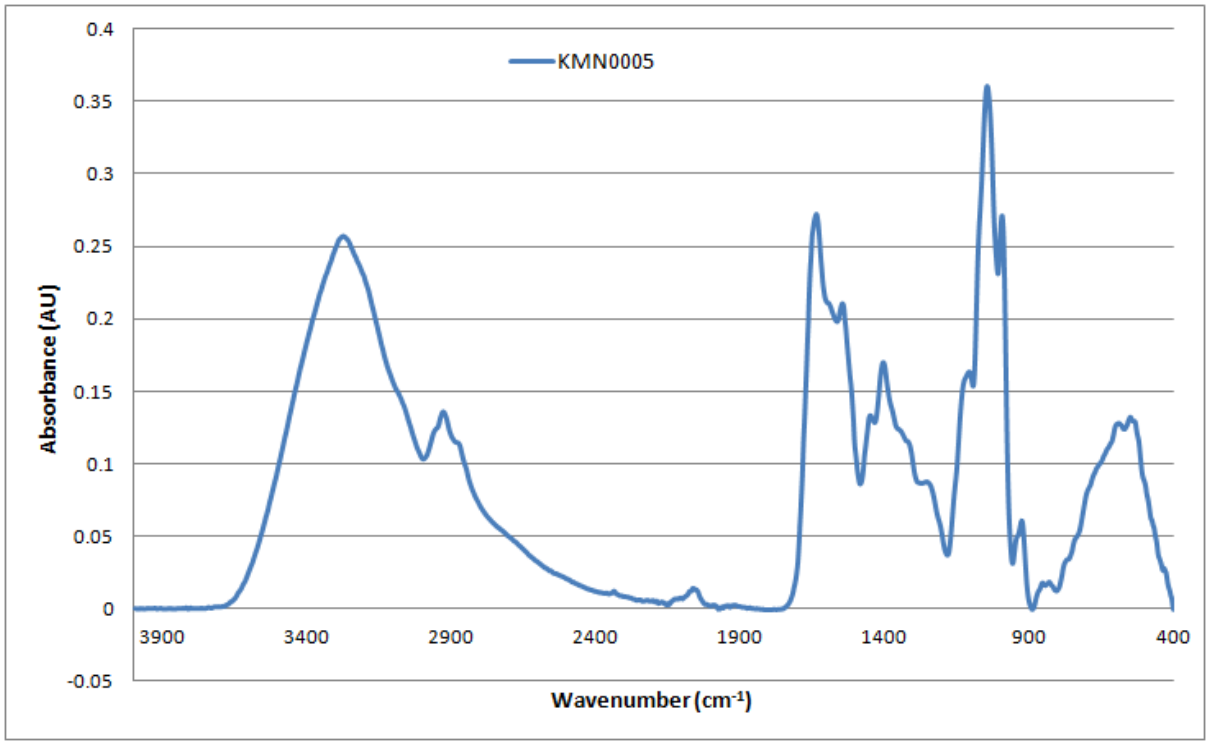


**Figure A2-2: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0005 from 4000-400cm$^{-1}$**

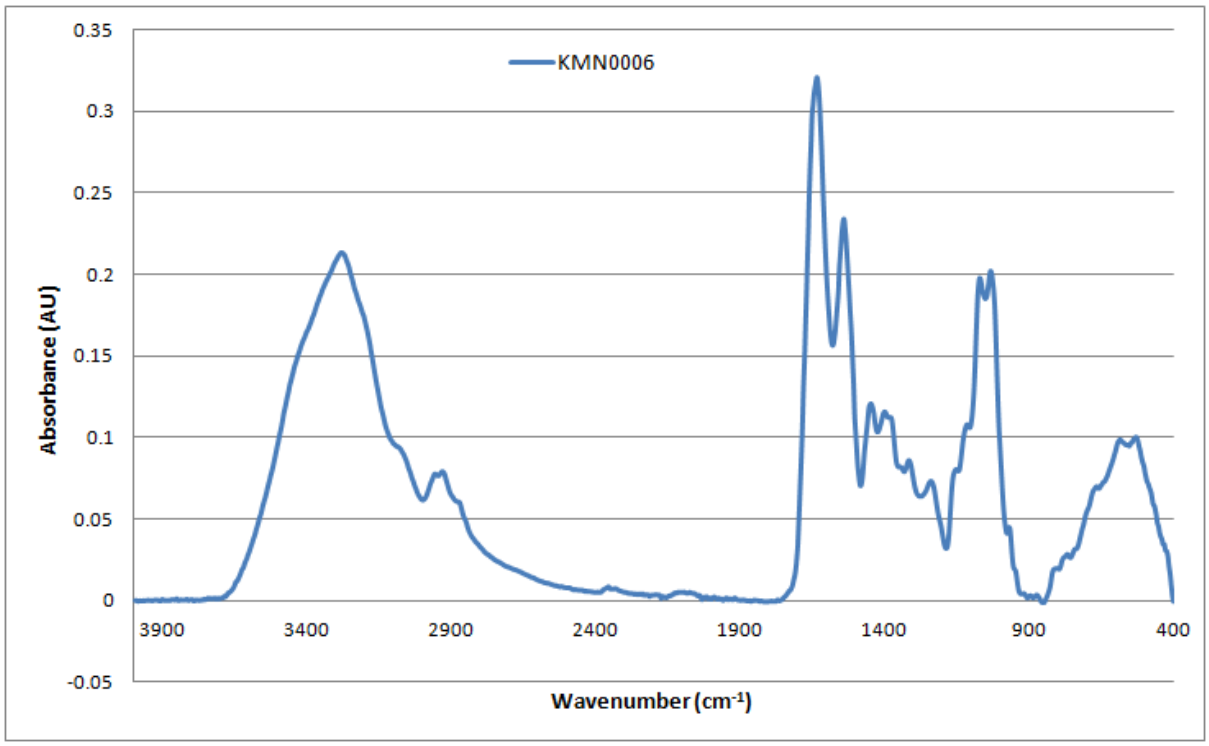


**Figure A2-3: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0006 from 4000-400cm$^{-1}$**

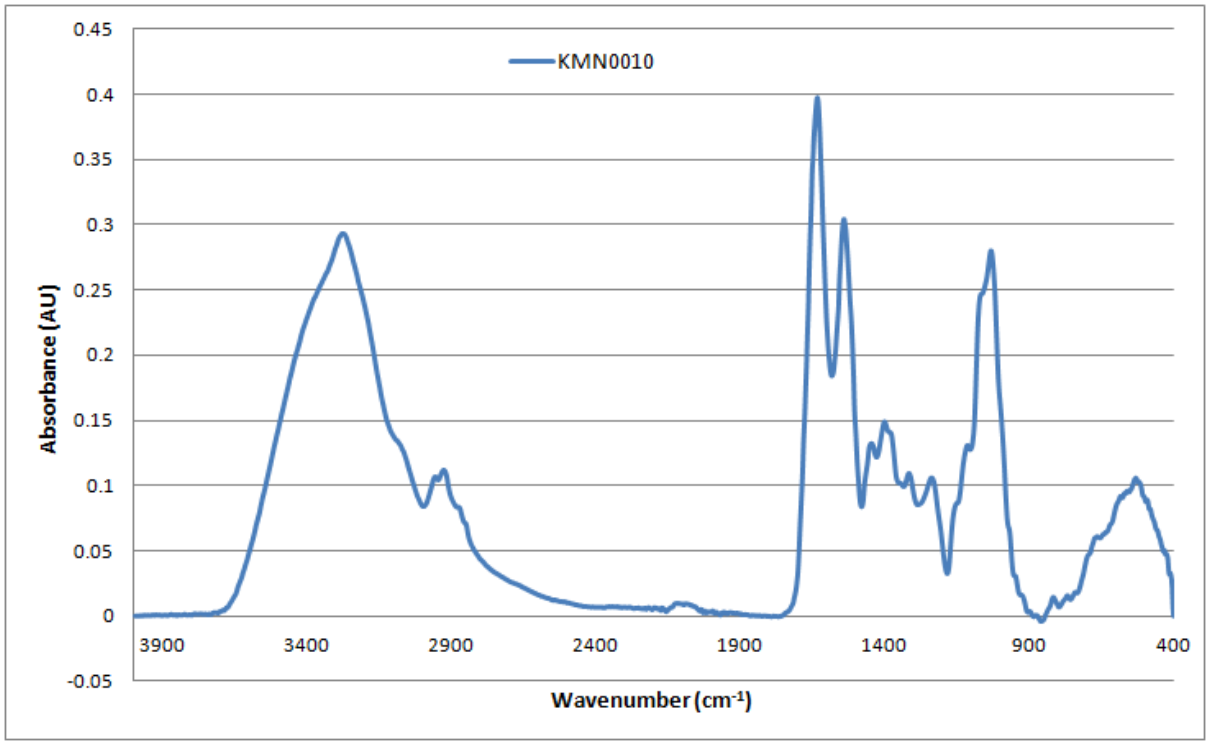


**Figure A2-4: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0010 from 4000-400cm$^{-1}$**

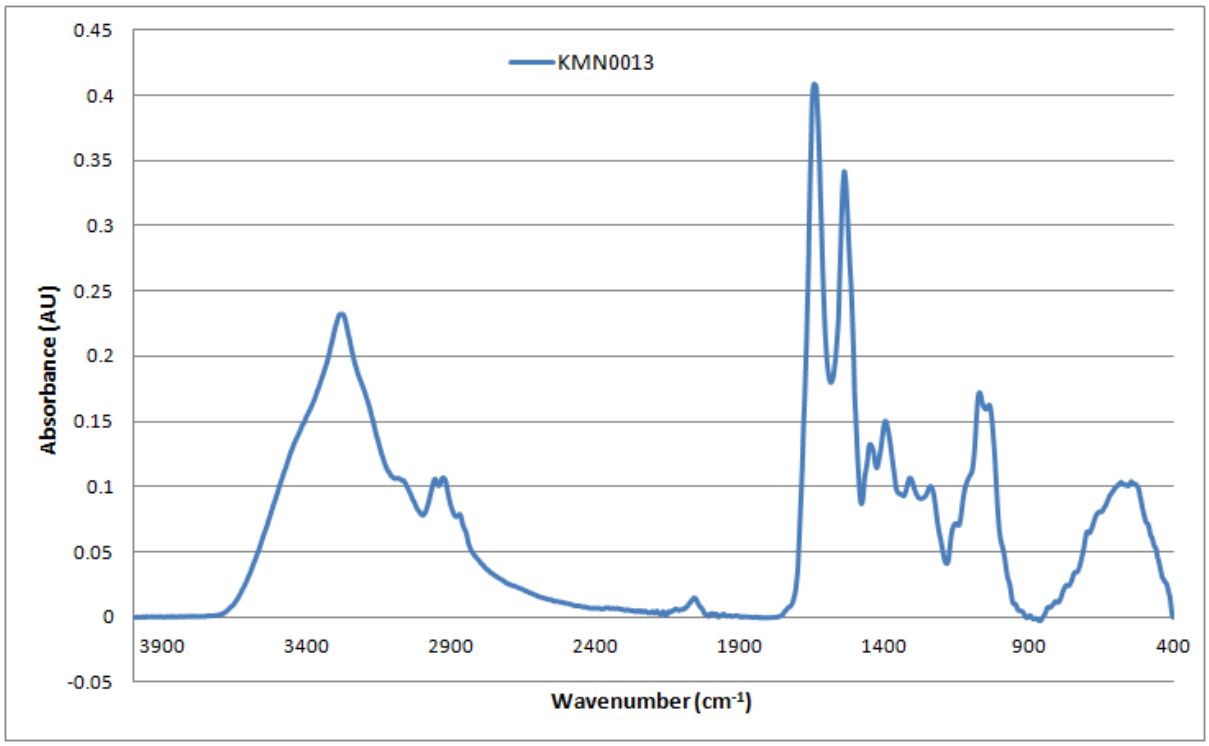


**Figure A2-5: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0013 from 4000-400cm$^{-1}$**

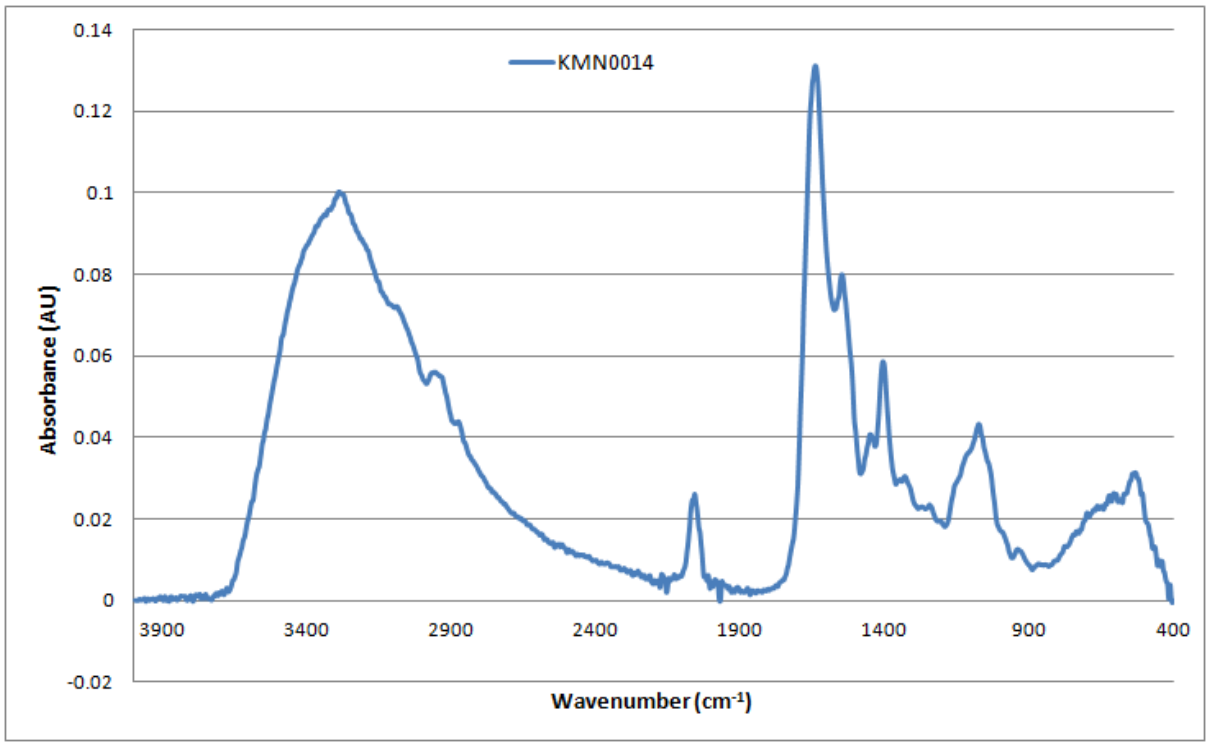


**Figure A2-6: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0014 from 4000-400cm$^{-1}$**

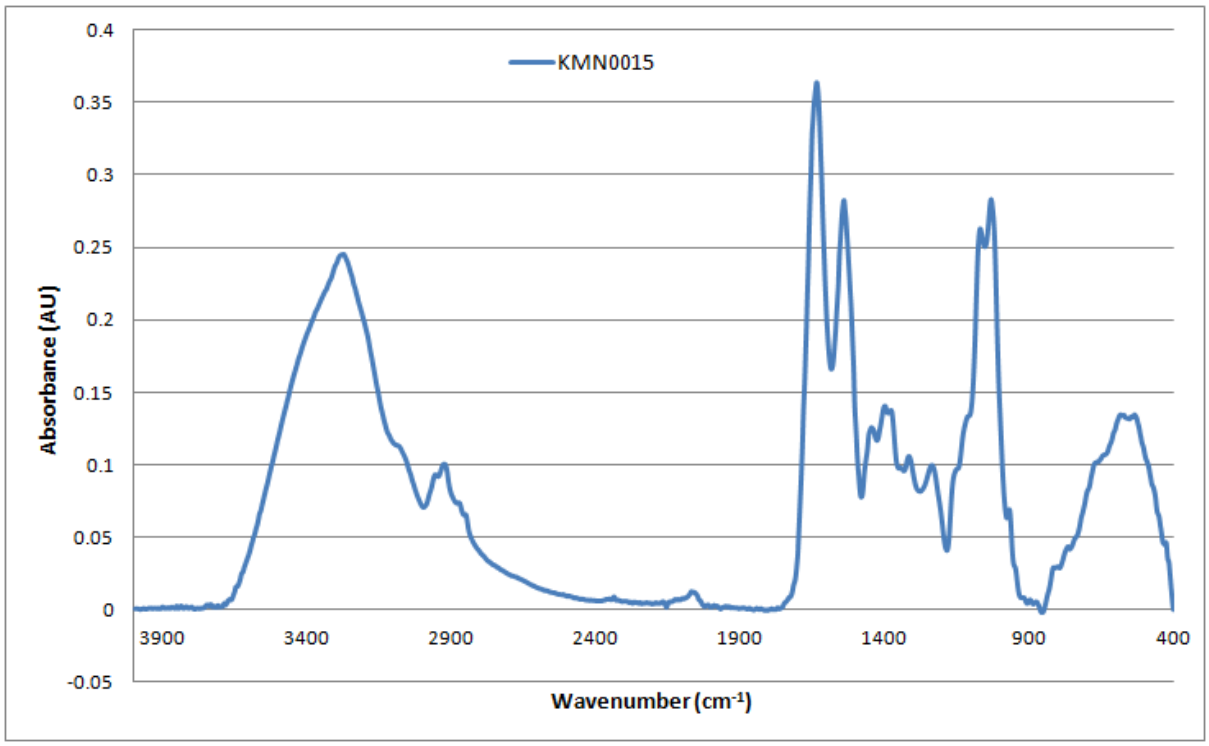


**Figure A2-7: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0015 from 4000-400cm$^{-1}$**

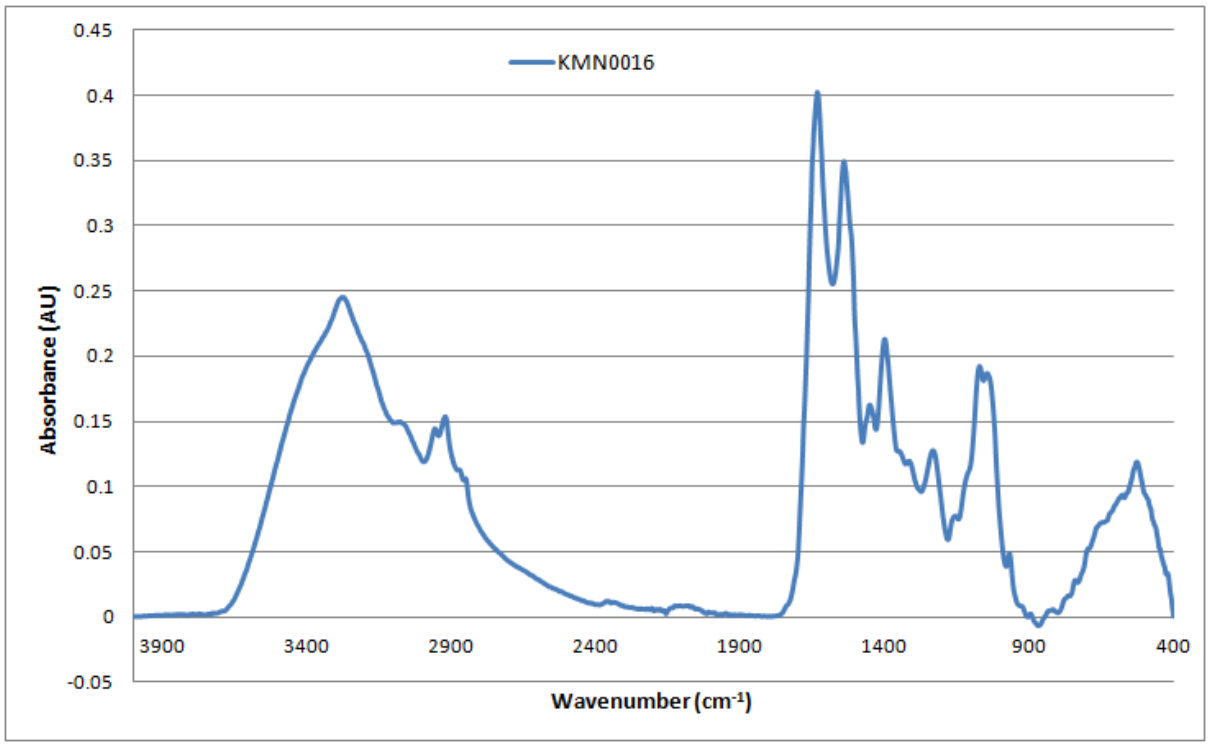


**Figure A2-8: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0016 from 4000-400cm$^{-1}$**

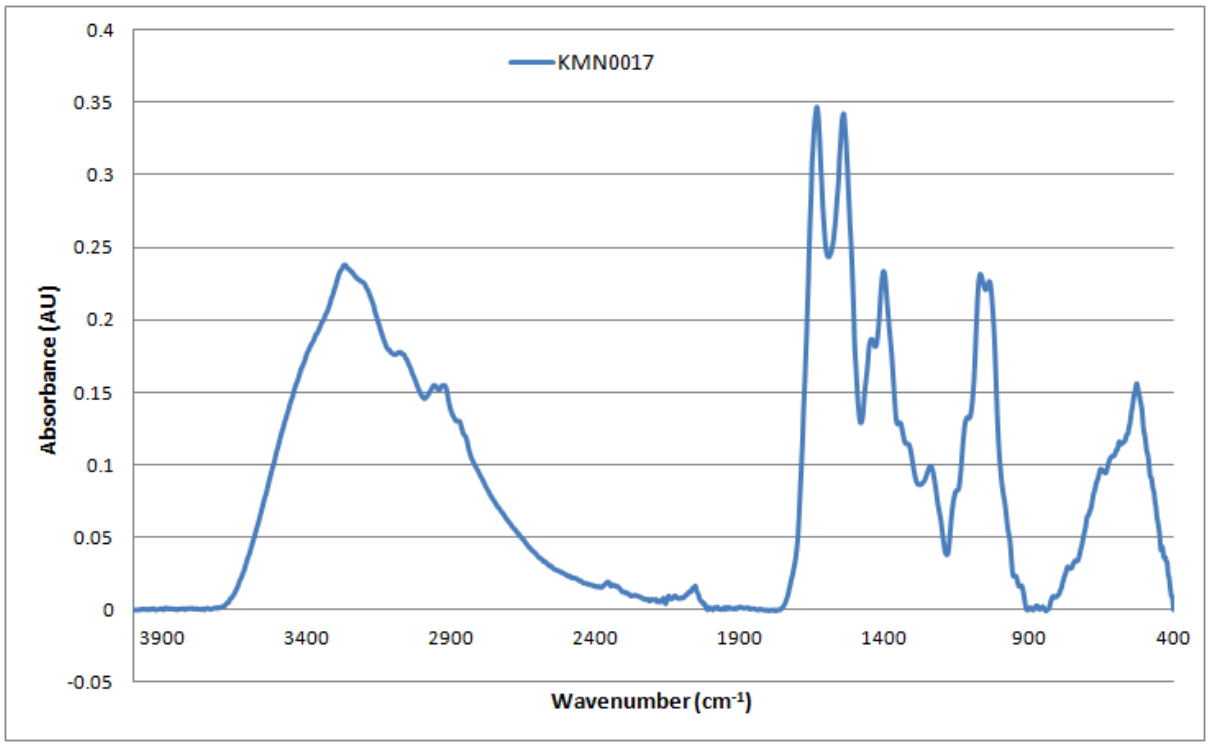


**Figure A2-9: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0017 from 4000-400cm$^{-1}$**

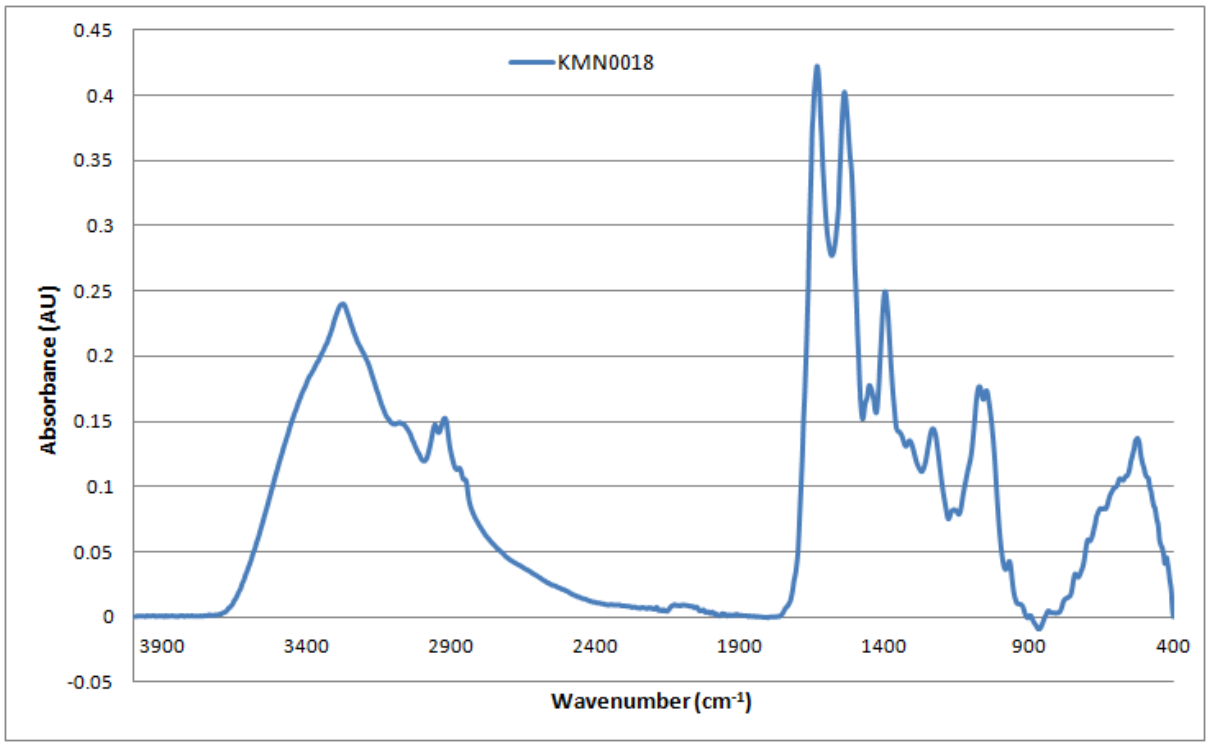


**Figure A2-10: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0018 from 4000-400cm$^{-1}$**

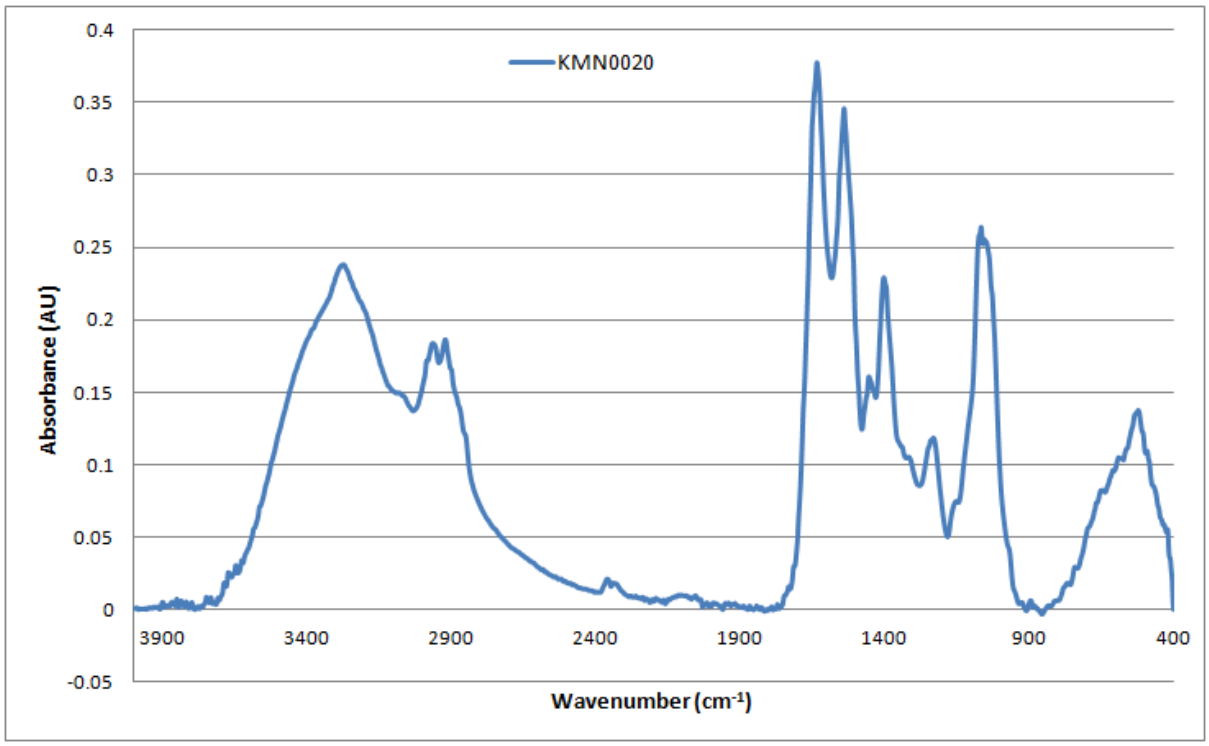


**Figure A2-11: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0020 from 4000-400cm$^{-1}$**

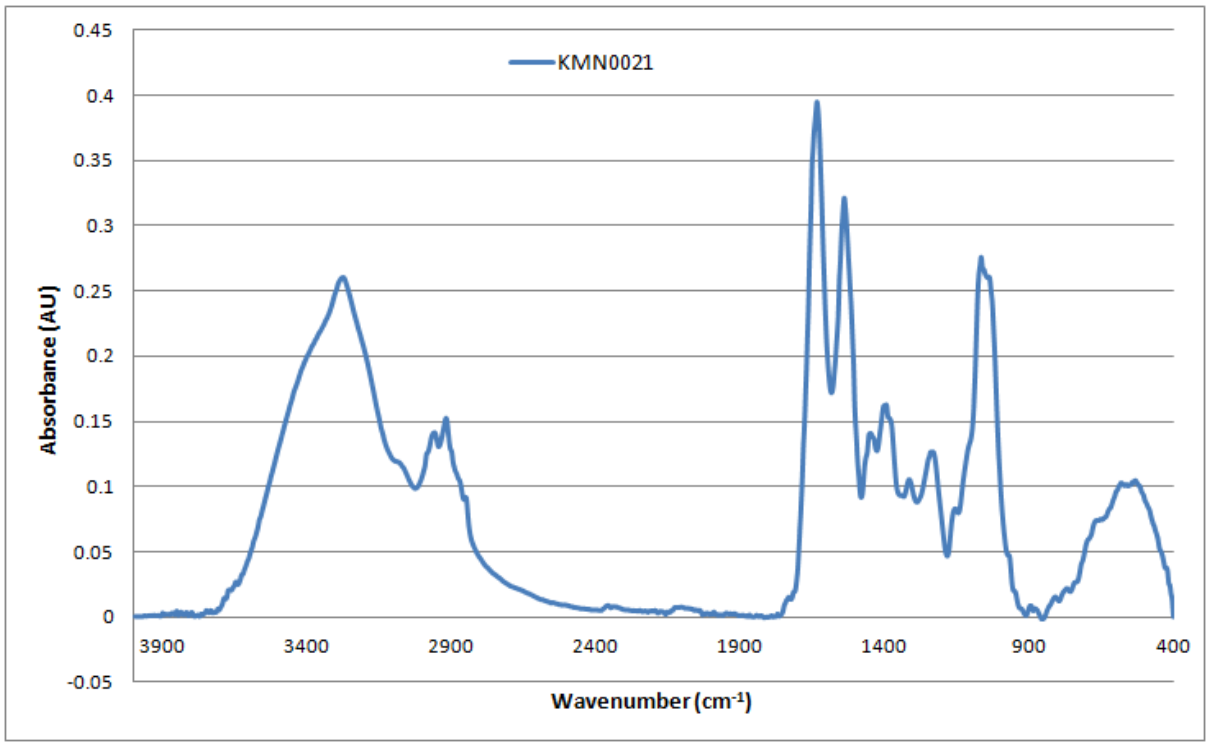


**Figure A2-12: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0021 from 4000-400cm$^{-1}$**

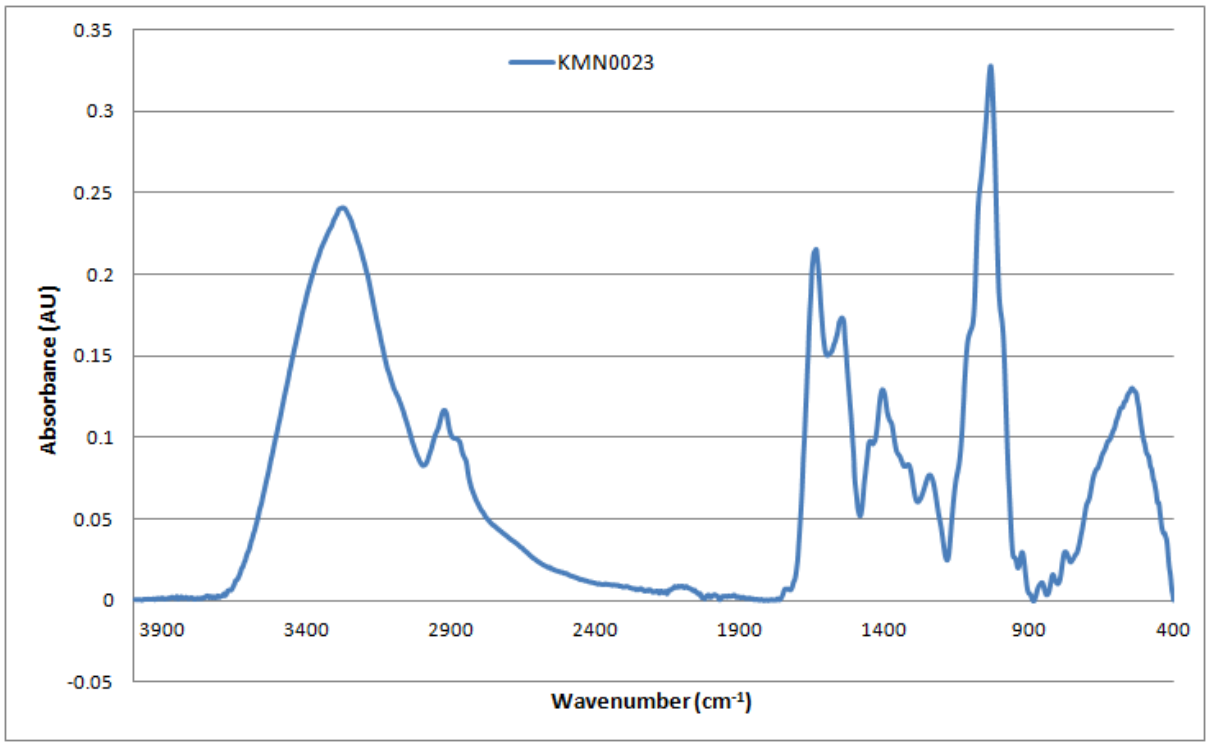


**Figure A2-13: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0023 from 4000-400cm$^{-1}$**

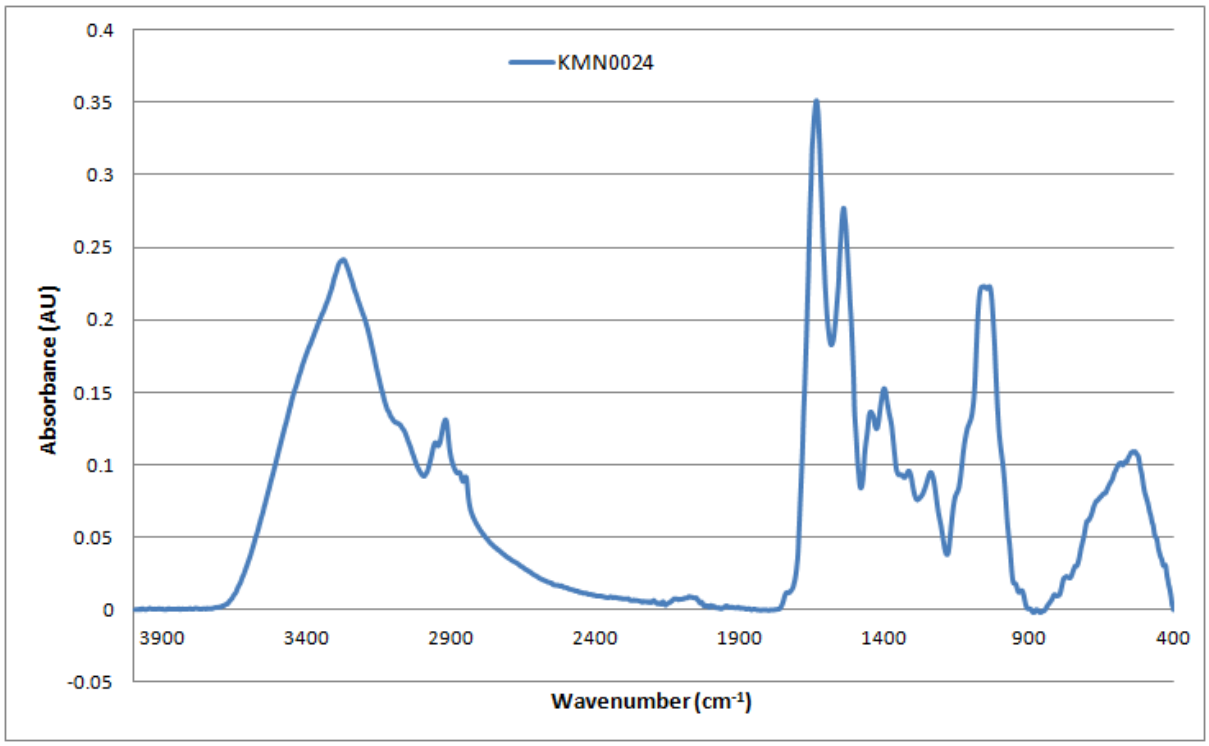


**Figure A2-14: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0024 from 4000-400cm$^{-1}$**

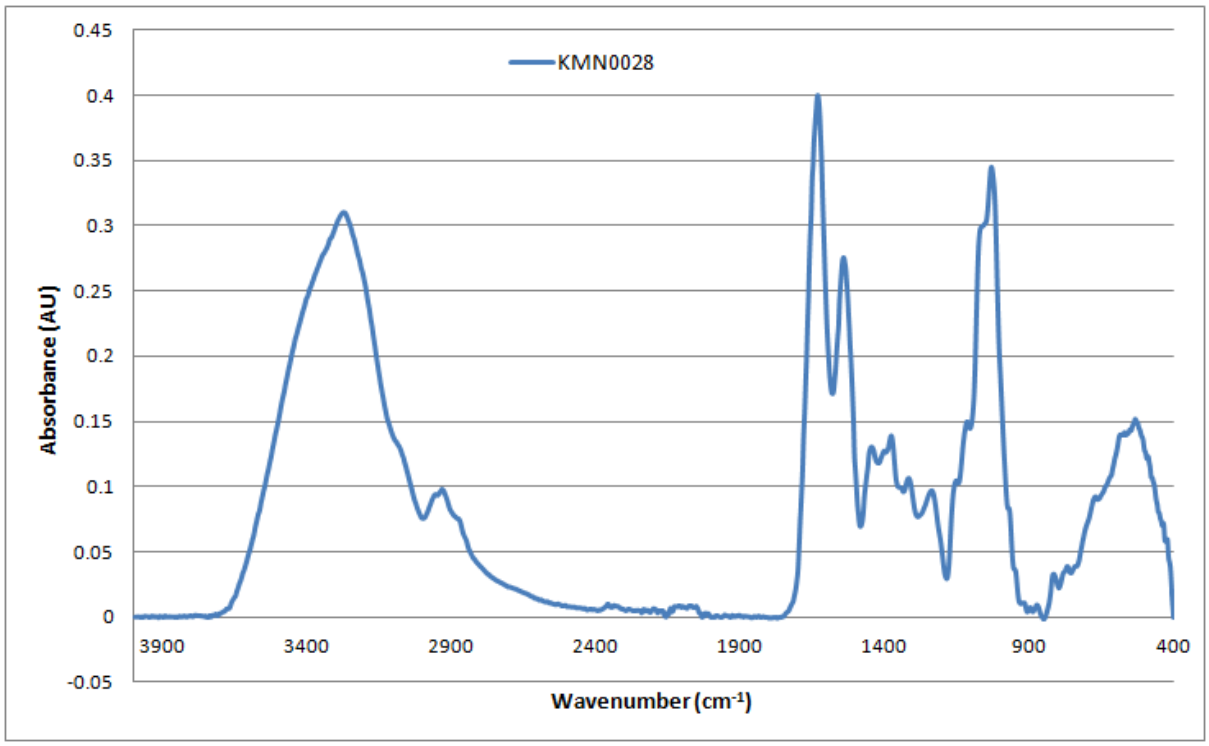


**Figure A2-15: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0028 from 4000-400cm$^{-1}$**

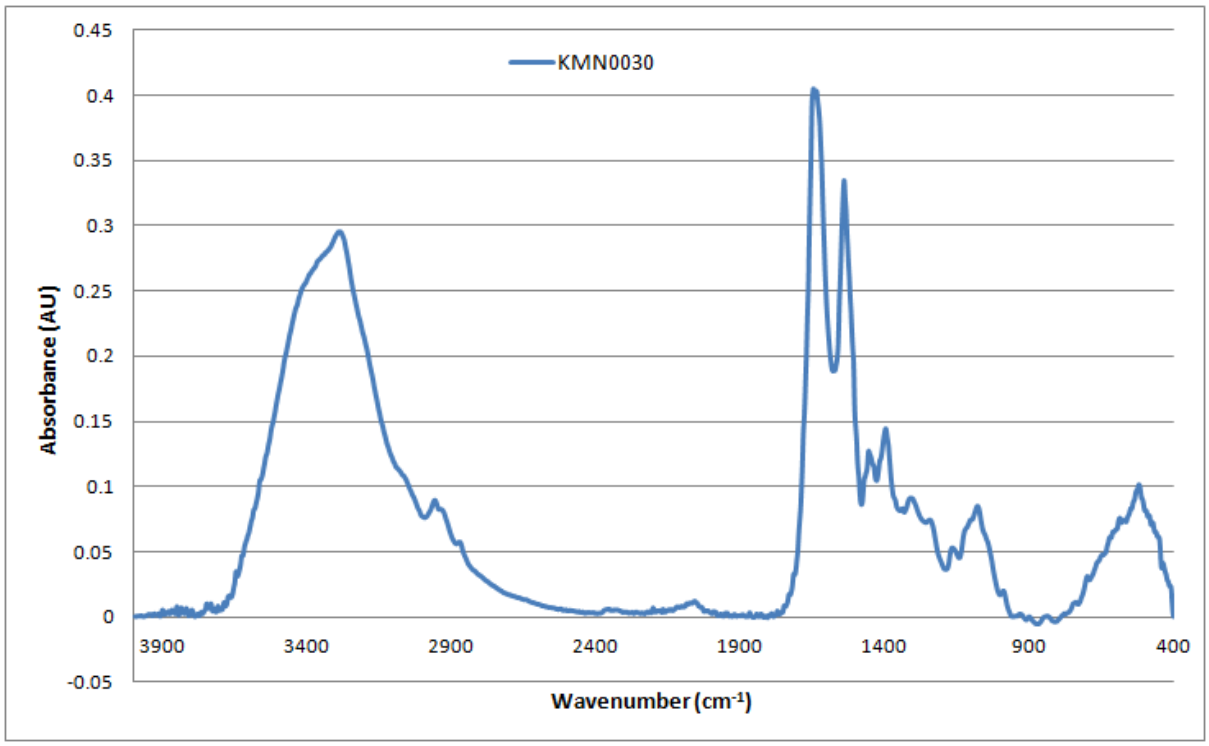


**Figure A2-16: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0030 from 4000-400cm$^{-1}$**

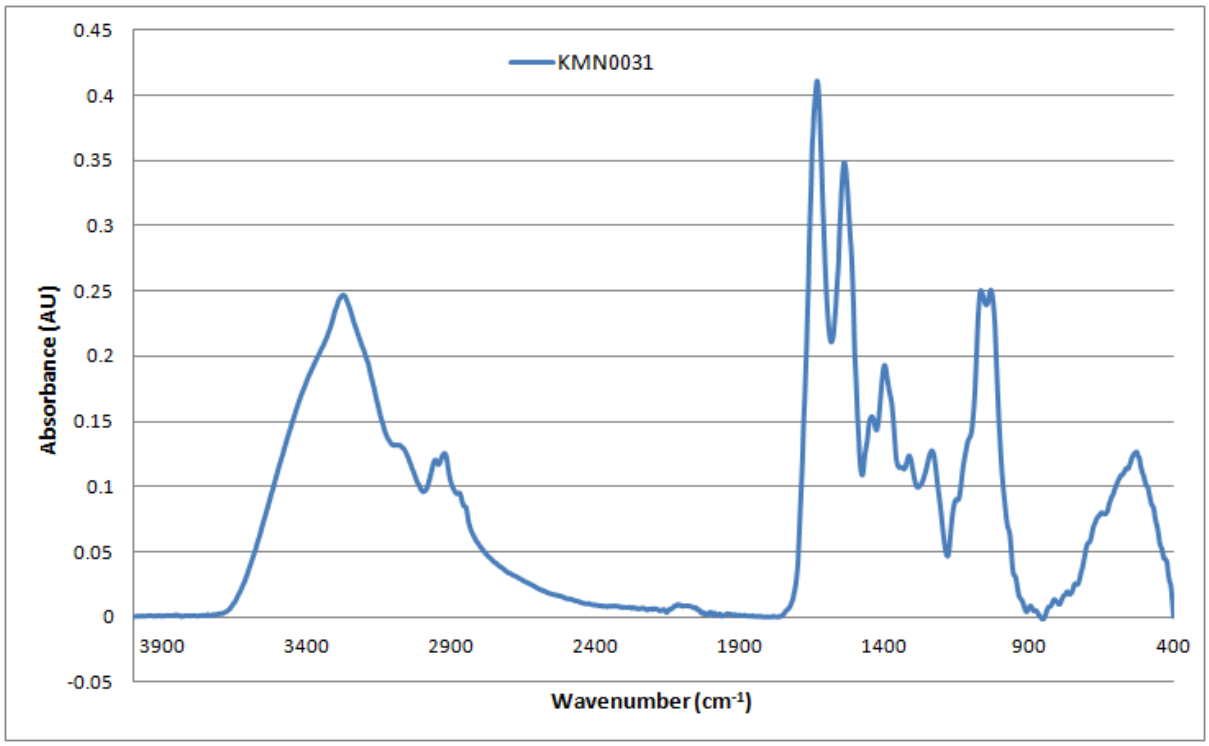


**Figure A2-17: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0031 from 4000-400cm$^{-1}$**

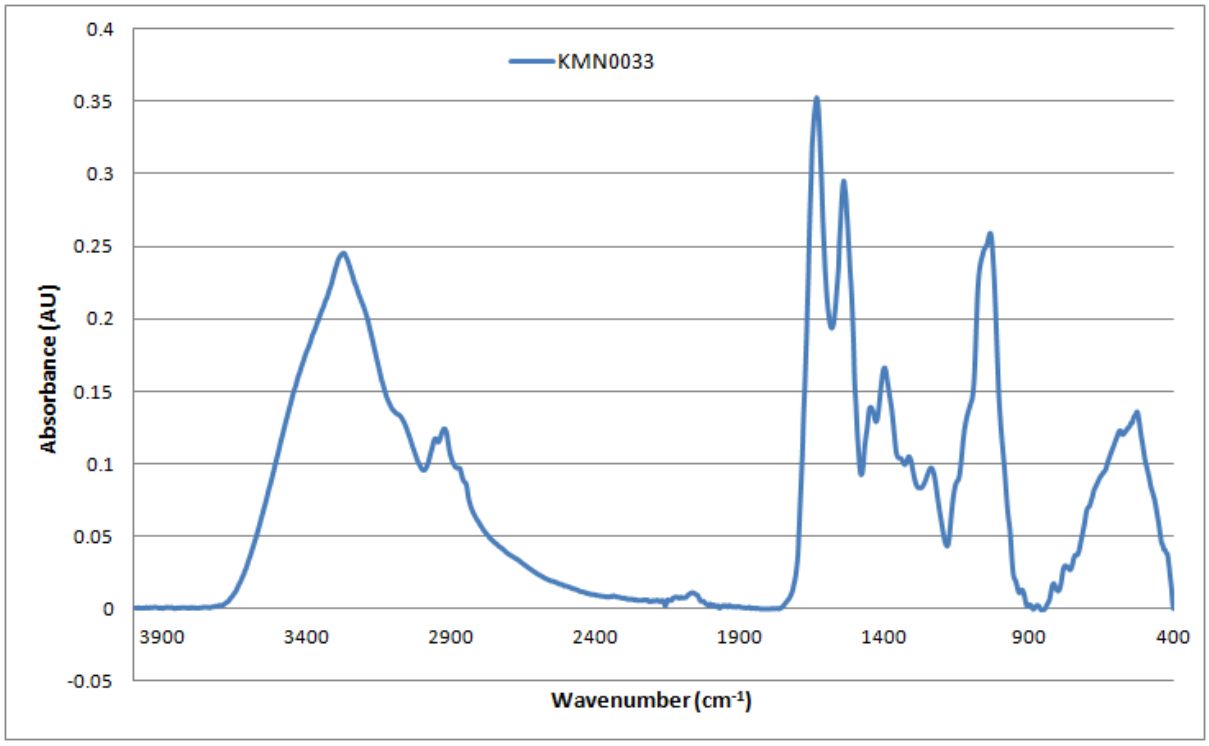


**Figure A2-18: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0033 from 4000-400cm$^{-1}$**

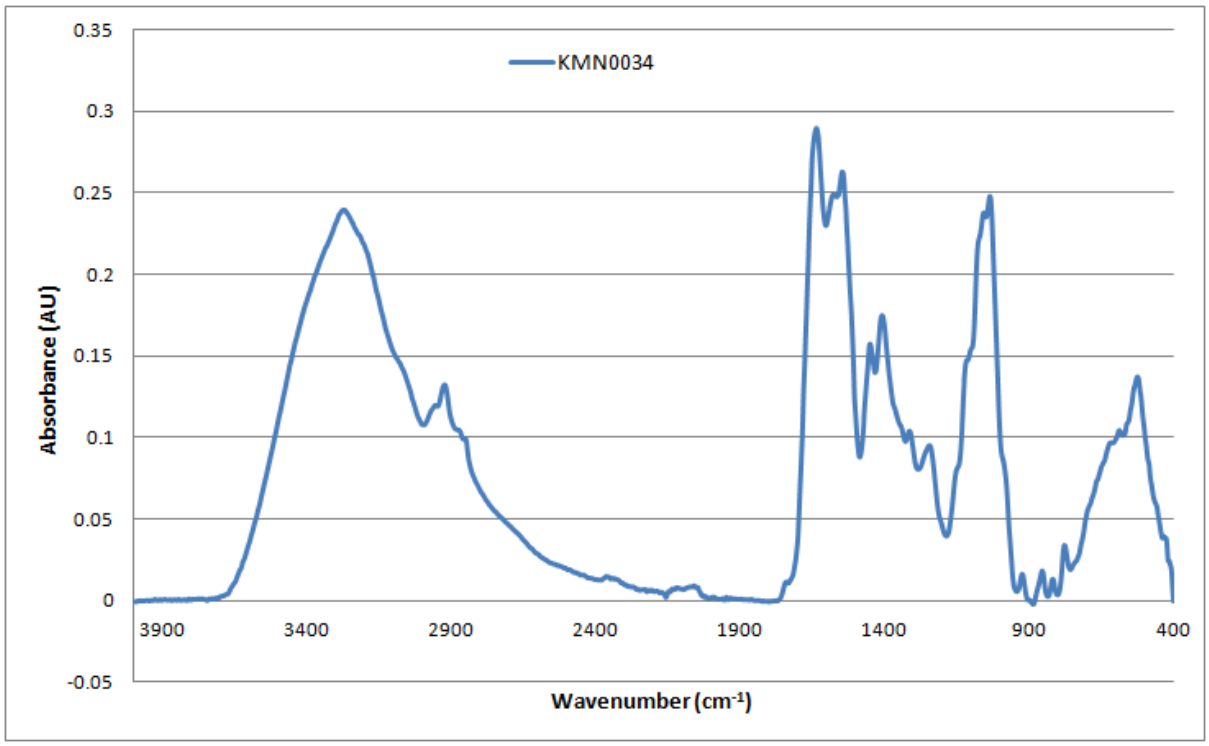


**Figure A2-19: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0034 from 4000-400cm$^{-1}$**

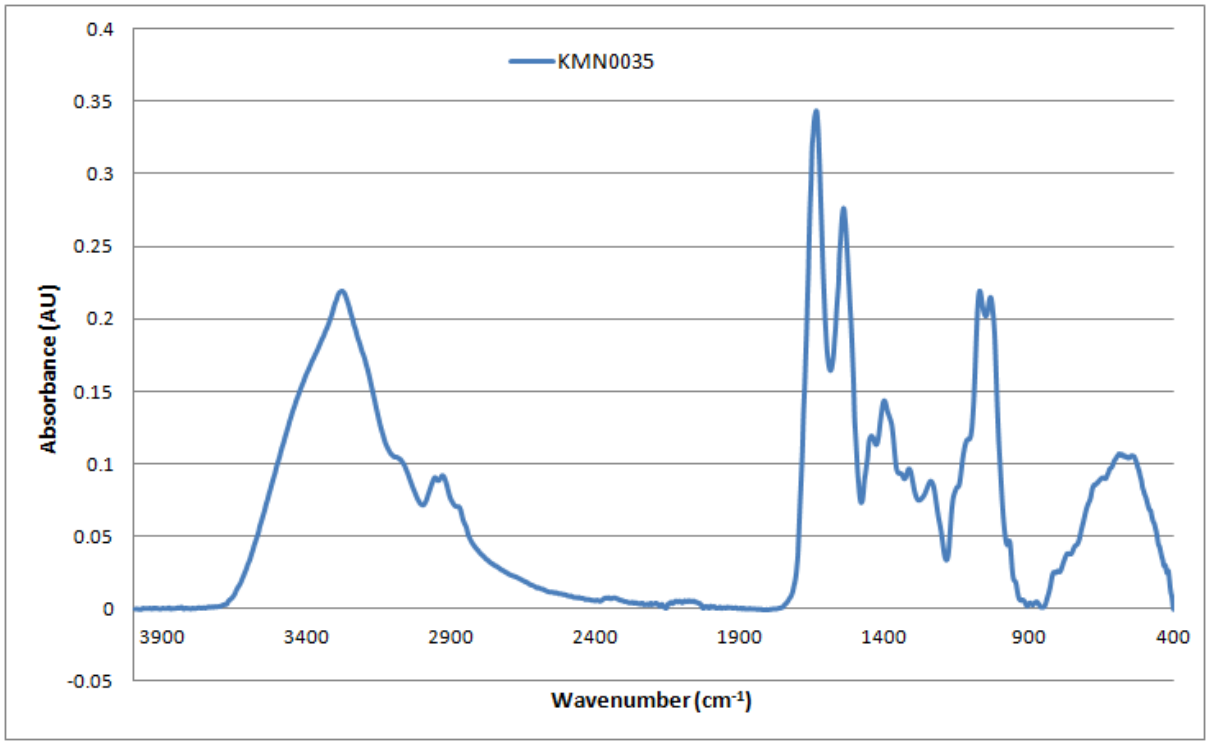


**Figure A2-20: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0035 from 4000-400cm$^{-1}$**

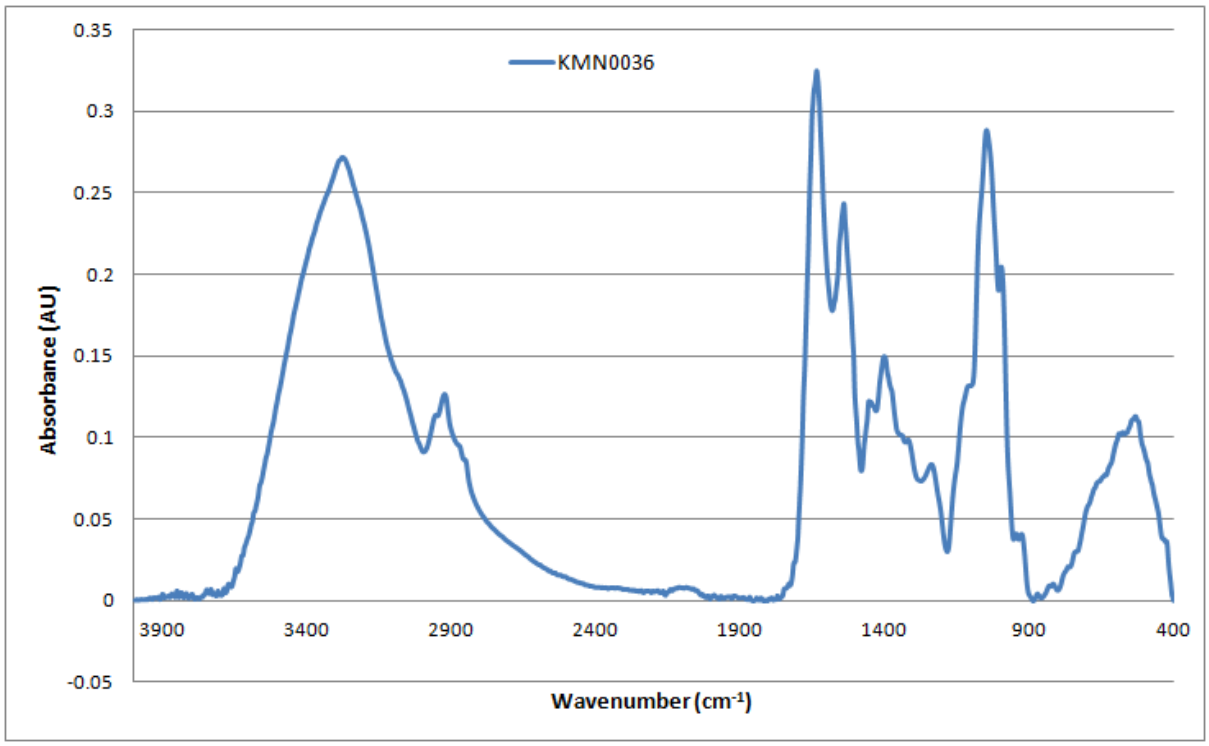


**Figure A2-21: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0036 from 4000-400cm$^{-1}$**

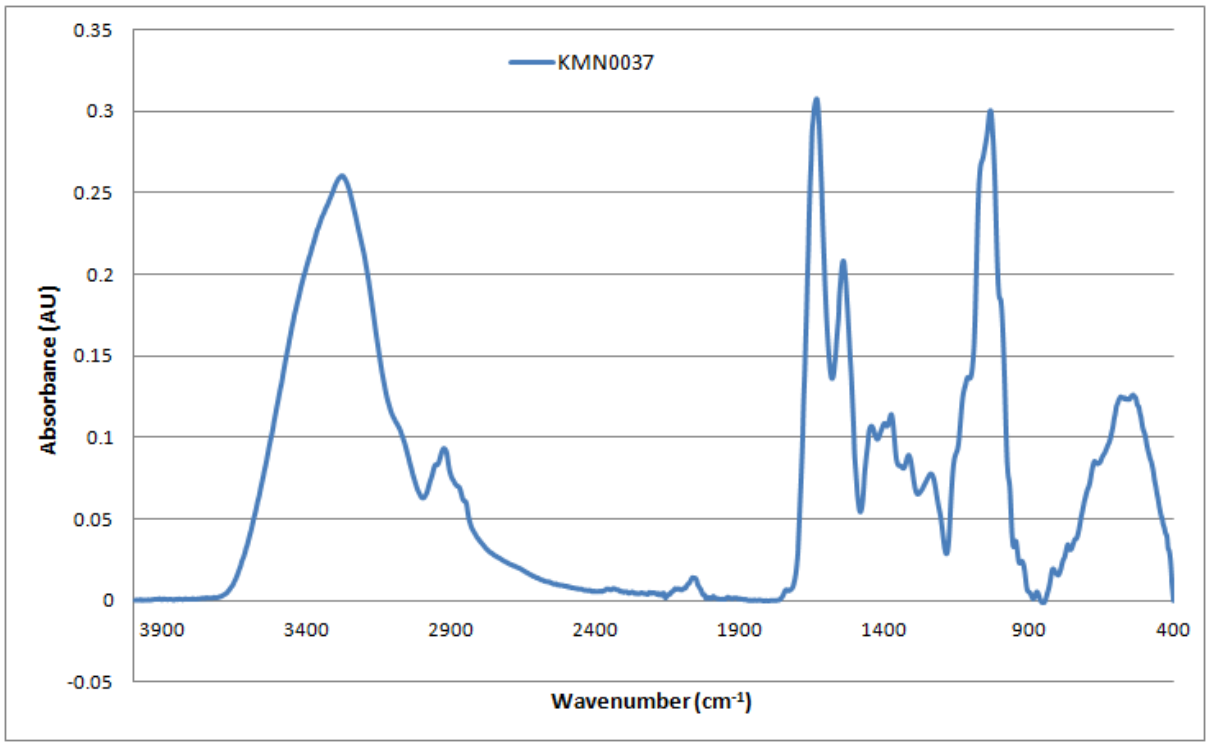


**Figure A2-22: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0037 from 4000-400cm$^{-1}$**

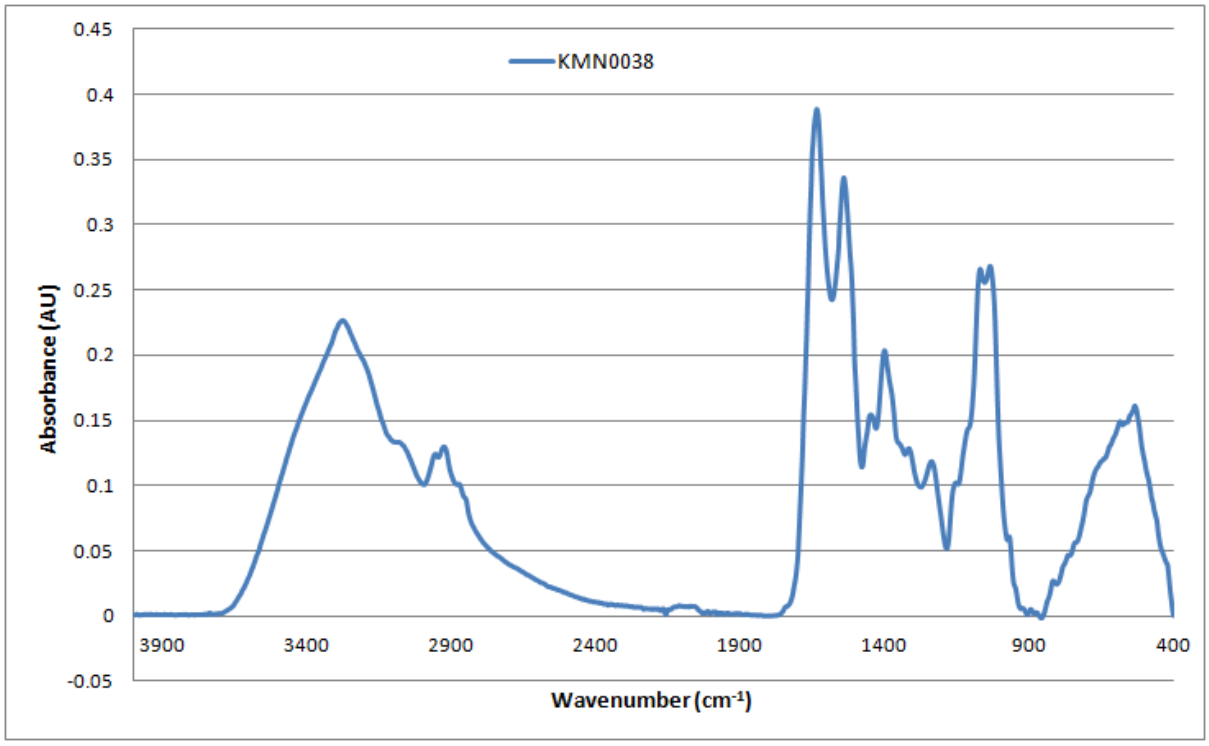


**Figure A2-23: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0038 from 4000-400cm$^{-1}$**

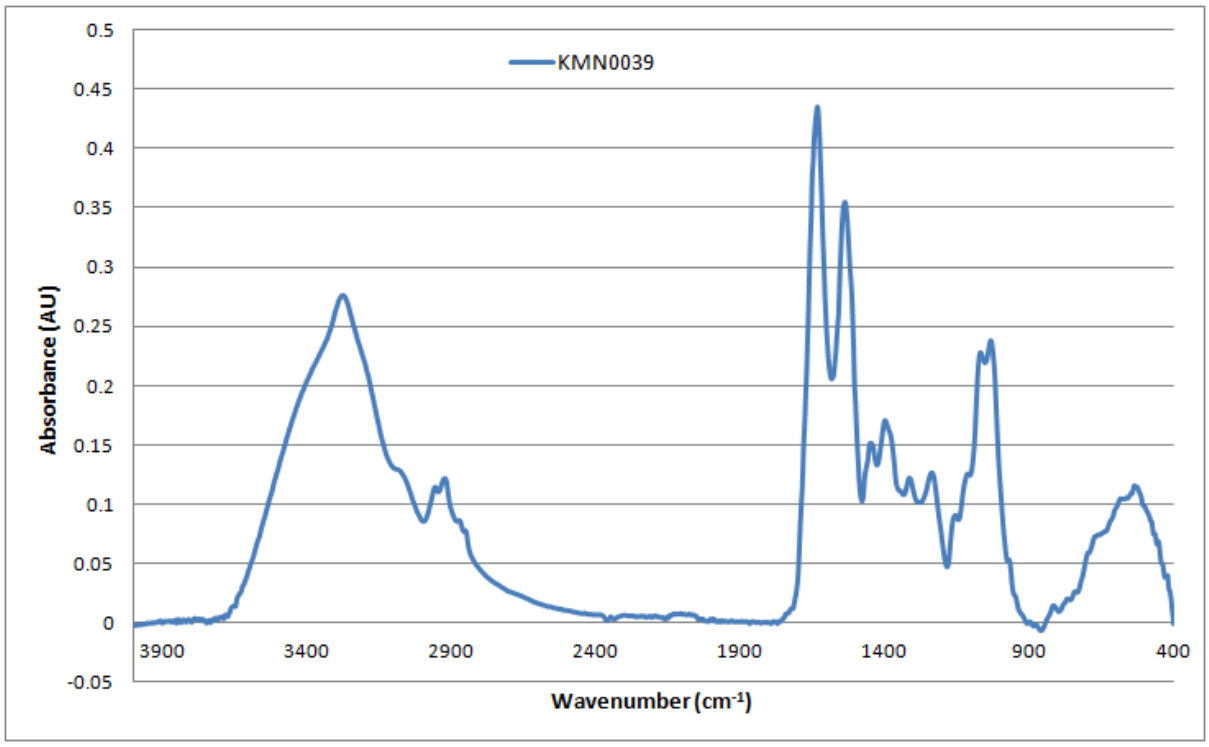


**Figure A2-24: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0039 from 4000-400cm$^{-1}$**

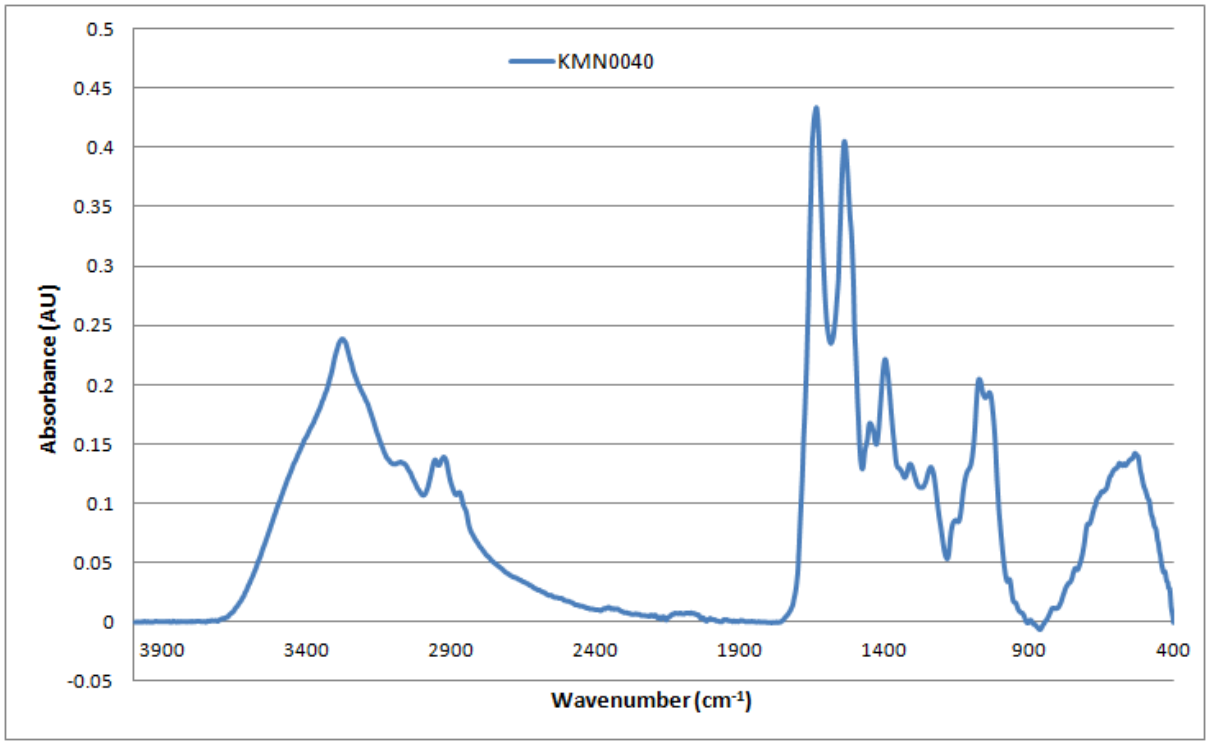


**Figure A2-25: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0040 from 4000-400cm$^{-1}$**

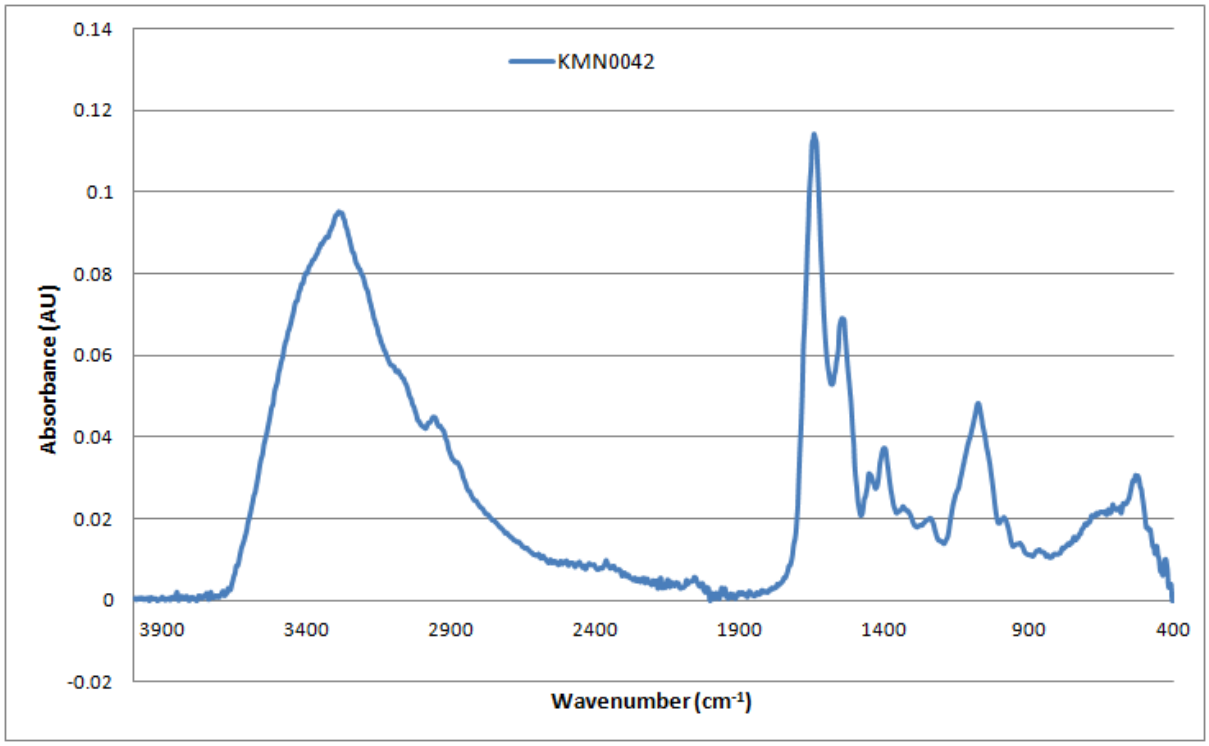


**Figure A2-26: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0042 from 4000-400cm$^{-1}$**

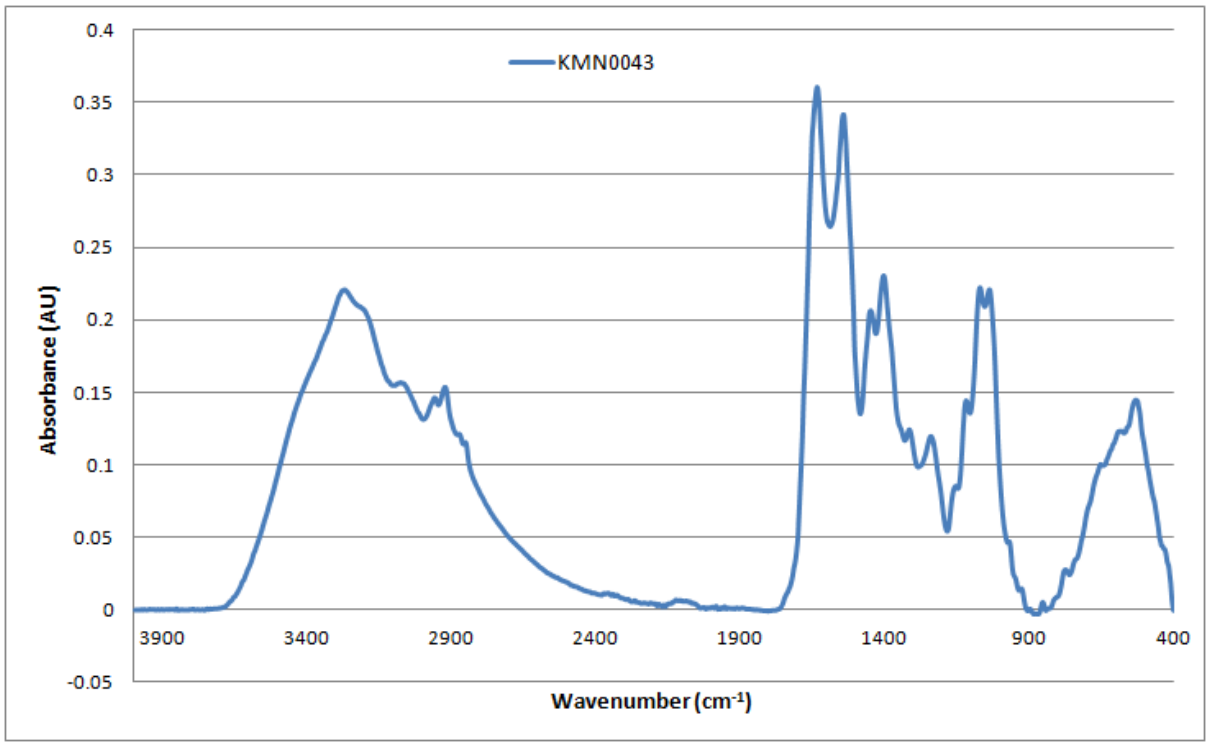


**Figure A2-27: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0043 from 4000-400cm$^{-1}$**

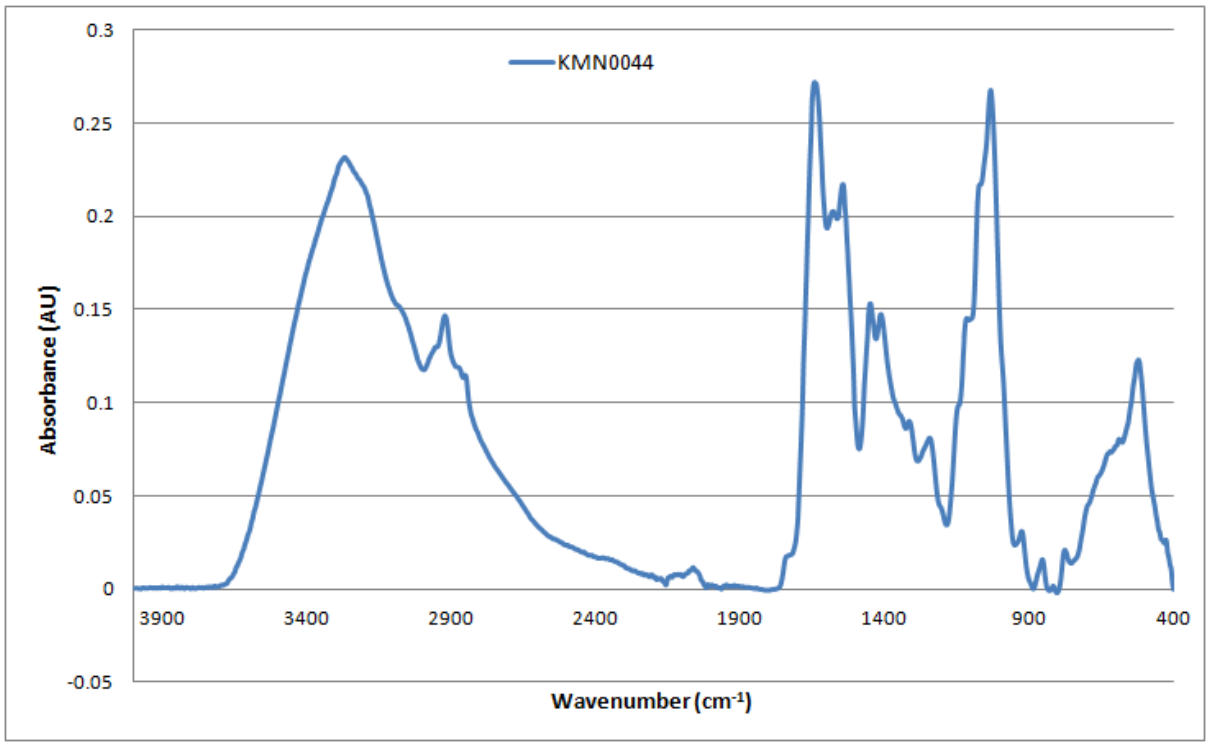


**Figure A2-28: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0044 from 4000-400cm$^{-1}$**

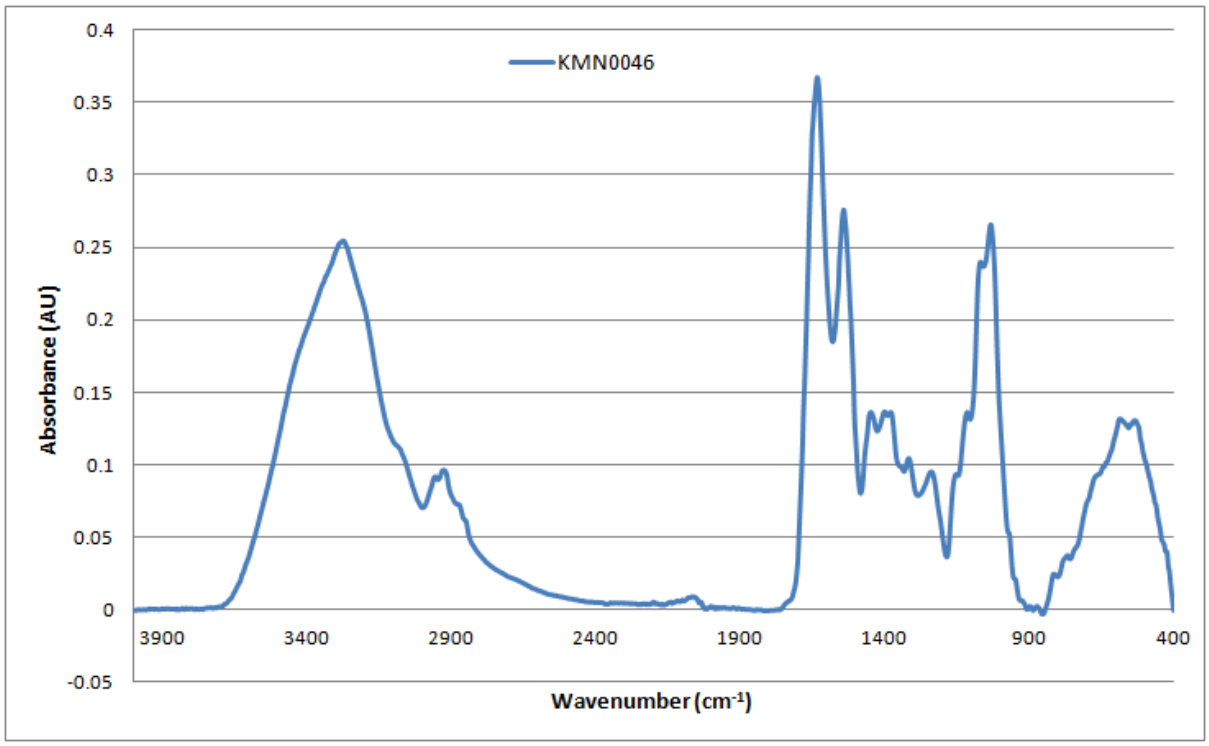


**Figure A2-29: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0046 from 4000-400cm$^{-1}$**

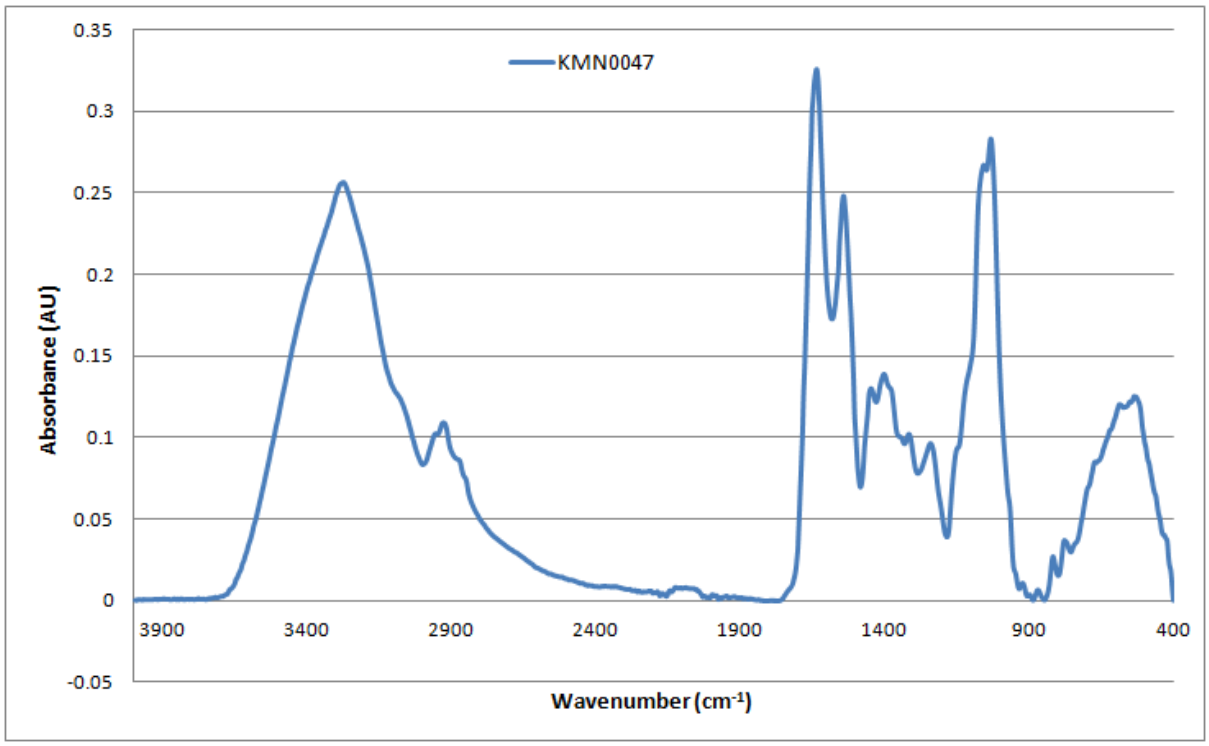


**Figure A2-30: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0047 from 4000-400cm$^{-1}$**

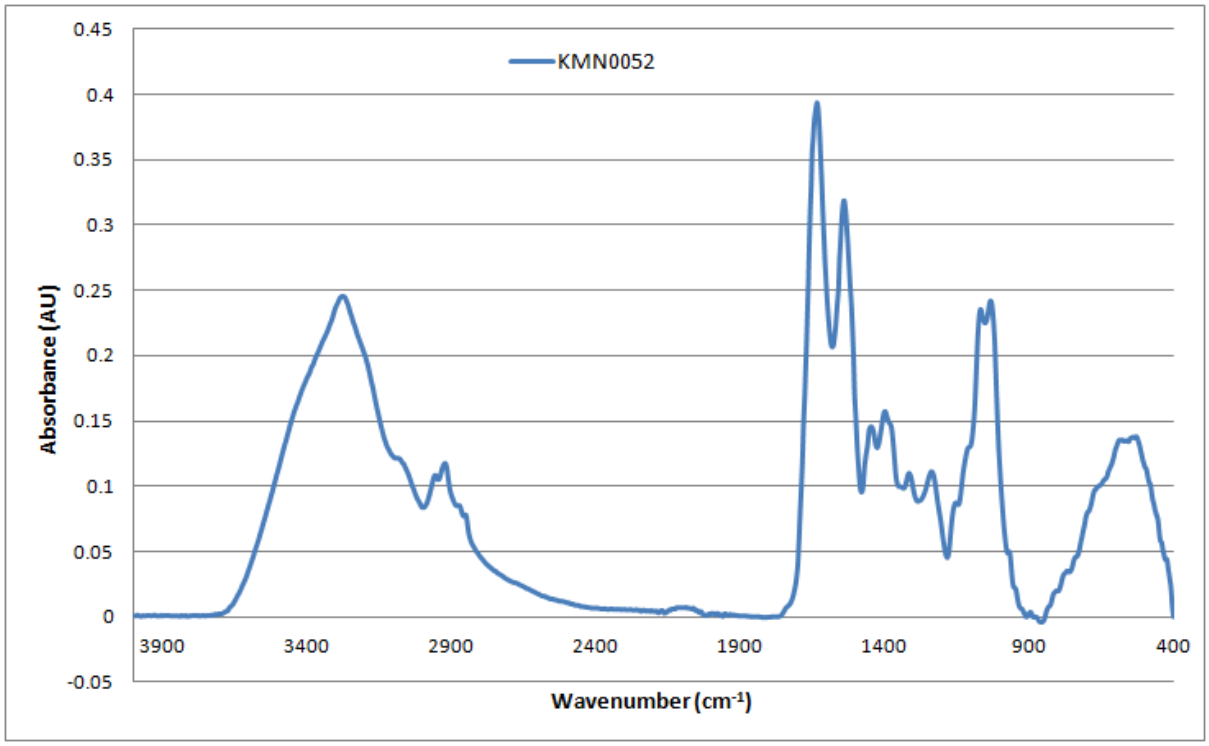


**Figure A2-31: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0052 from 4000-400cm$^{-1}$**

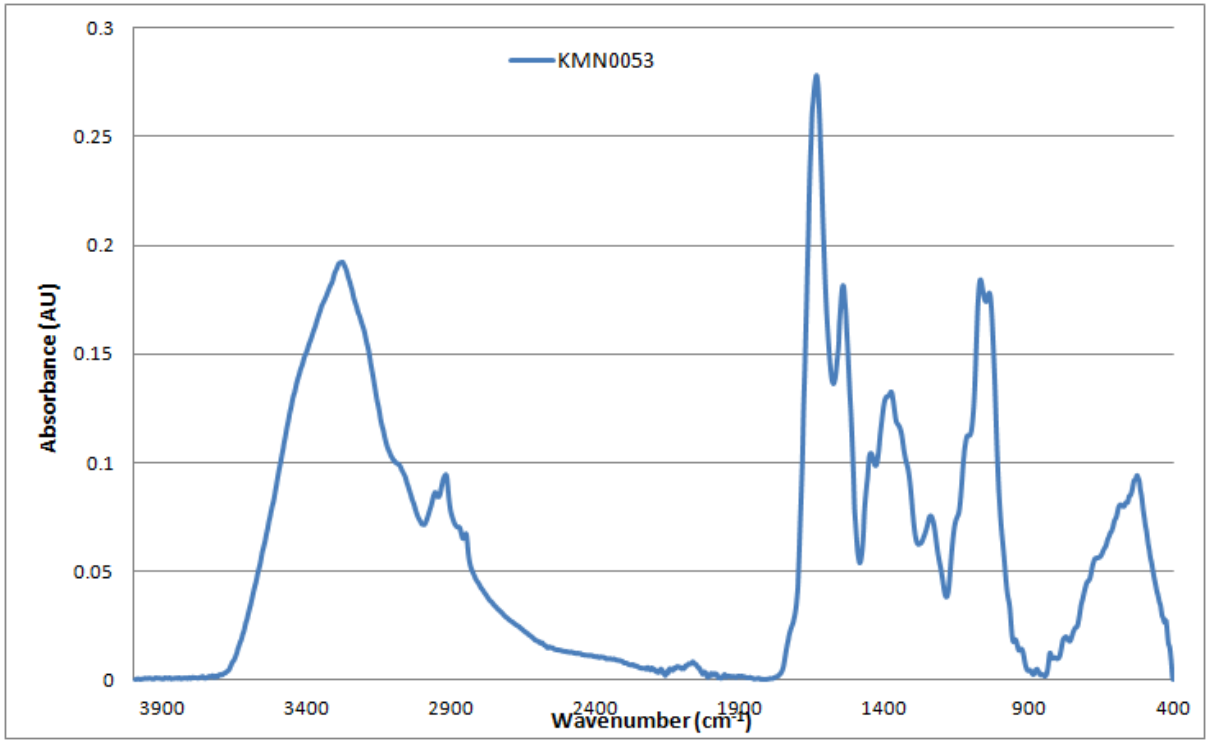


**Figure A2-32: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0053 from 4000-400cm$^{-1}$**

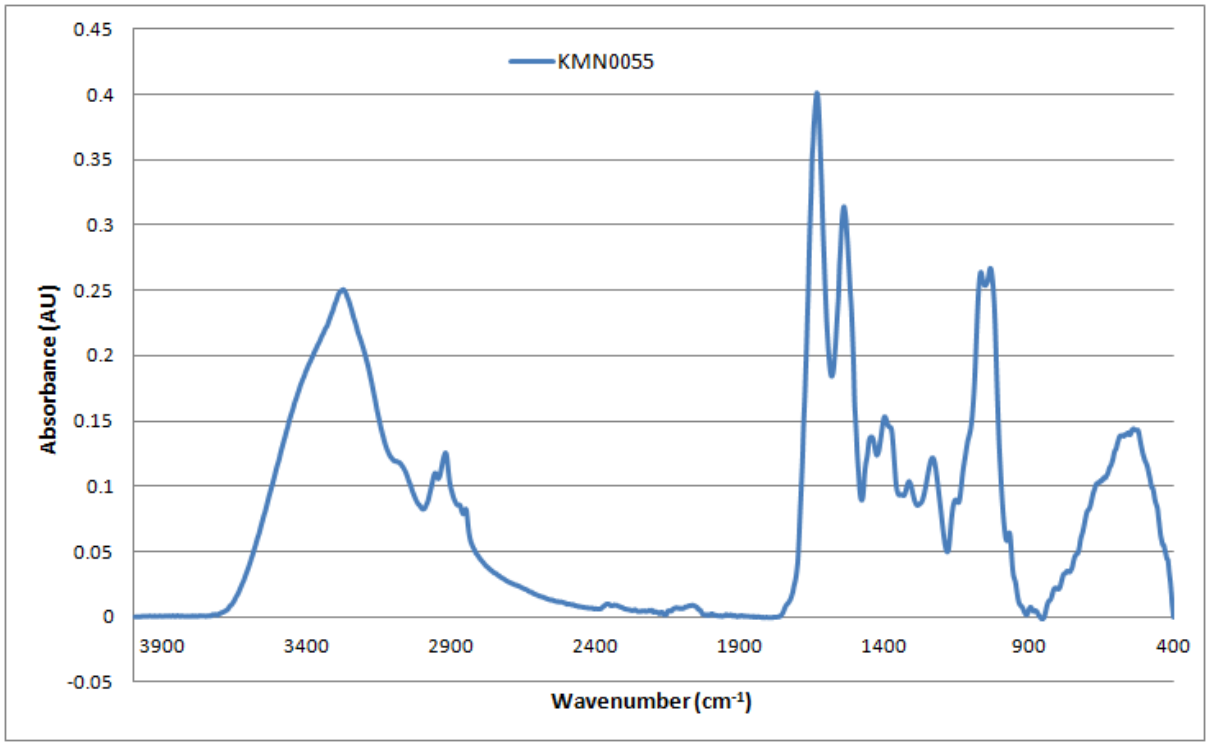


**Figure A2-33: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0055 from 4000-400cm$^{-1}$**

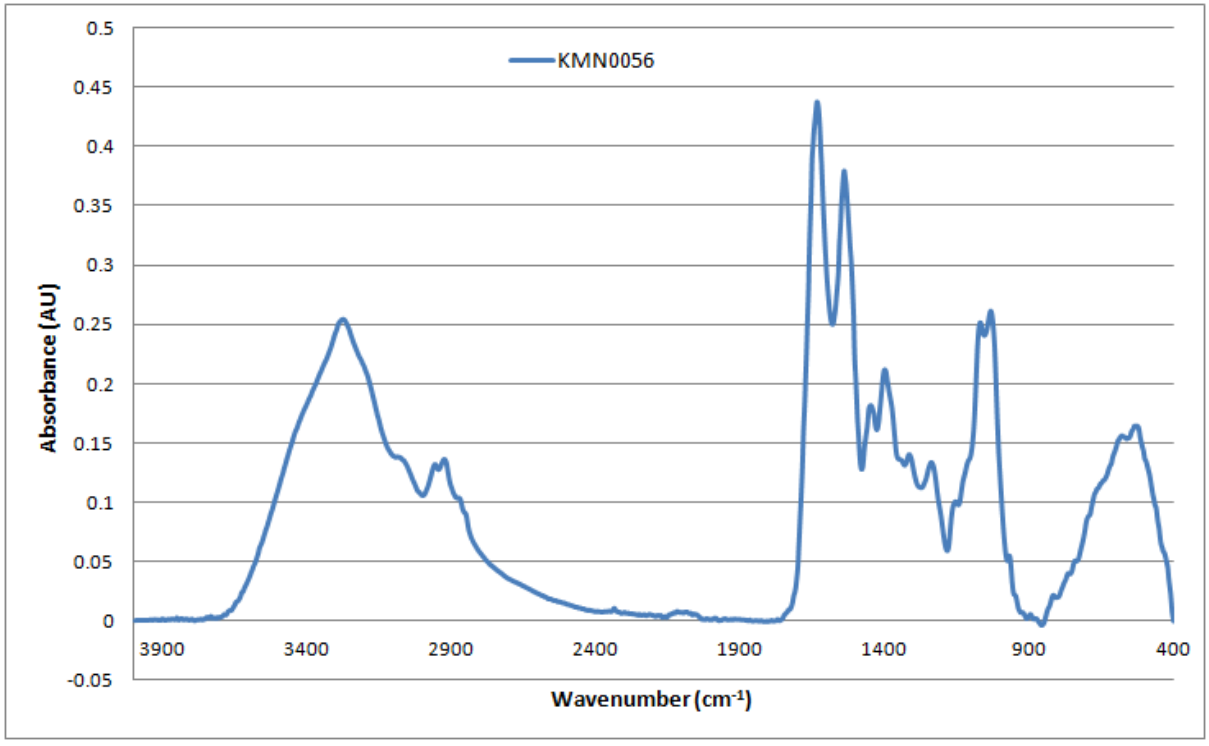


**Figure A2-34: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0056 from 4000-400cm$^{-1}$**

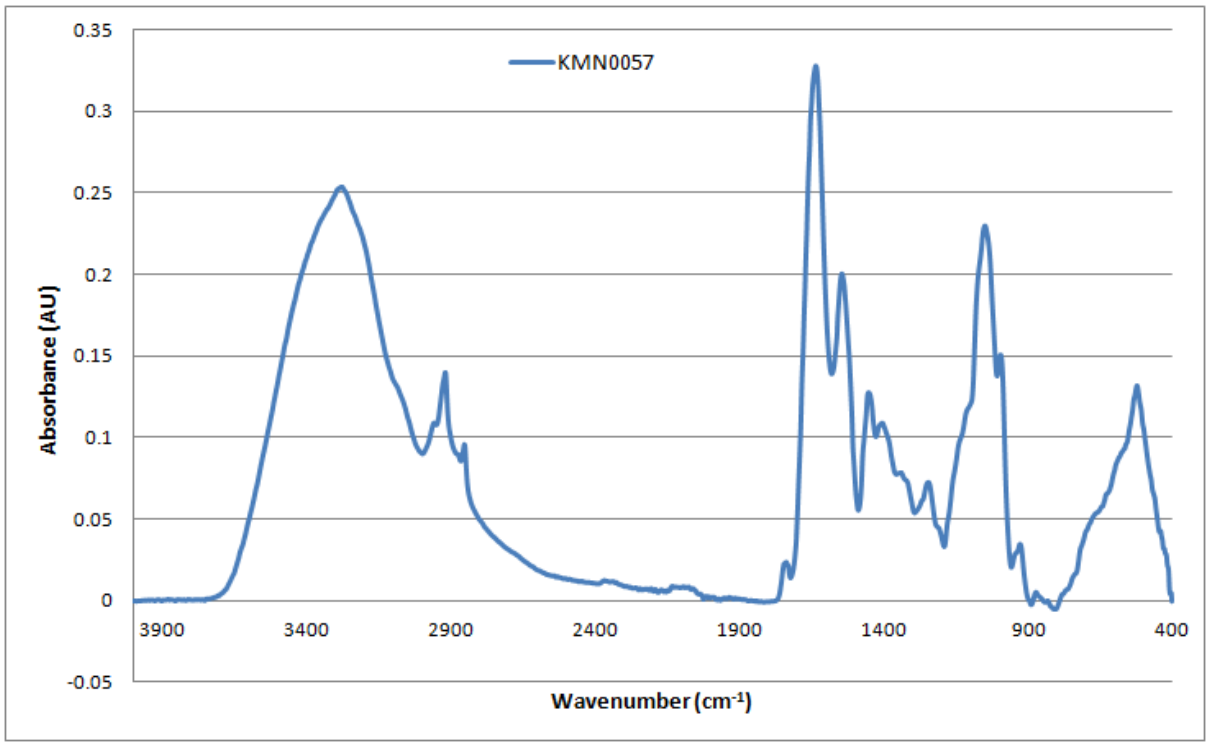


**Figure A2-35: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0057 from 4000-400cm$^{-1}$**

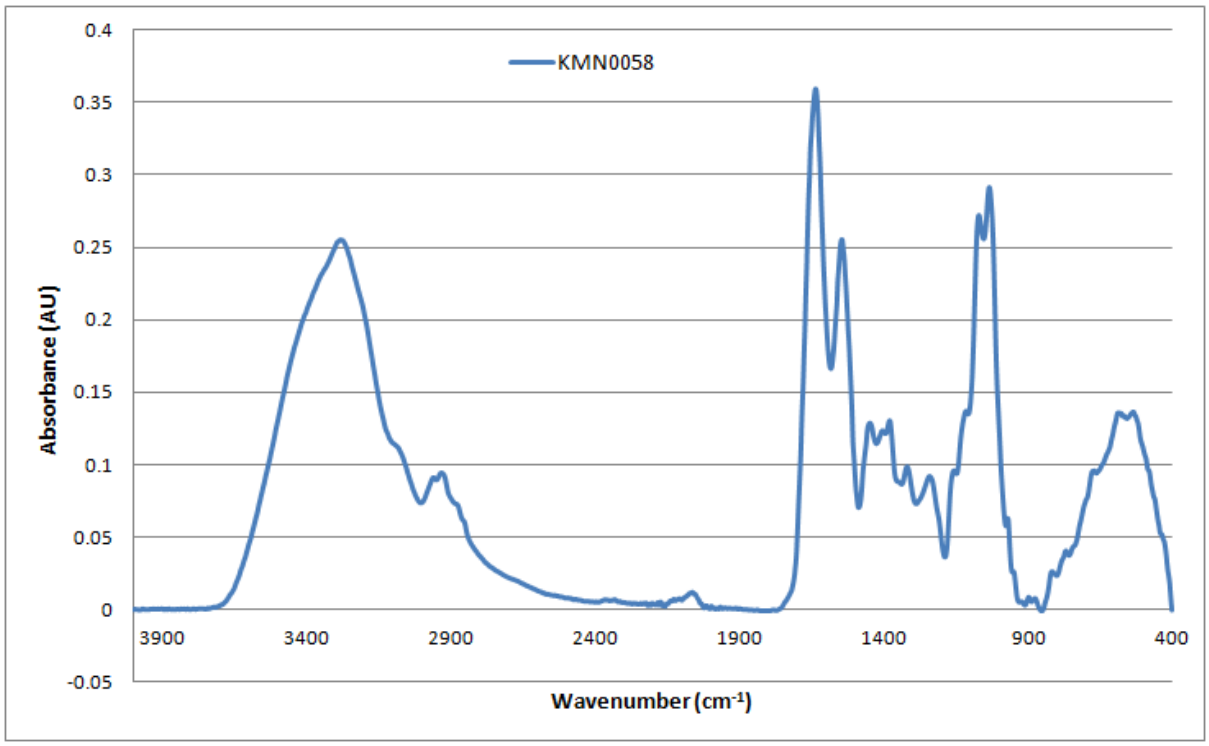


**Figure A2-36: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0058 from 4000-400cm$^{-1}$**

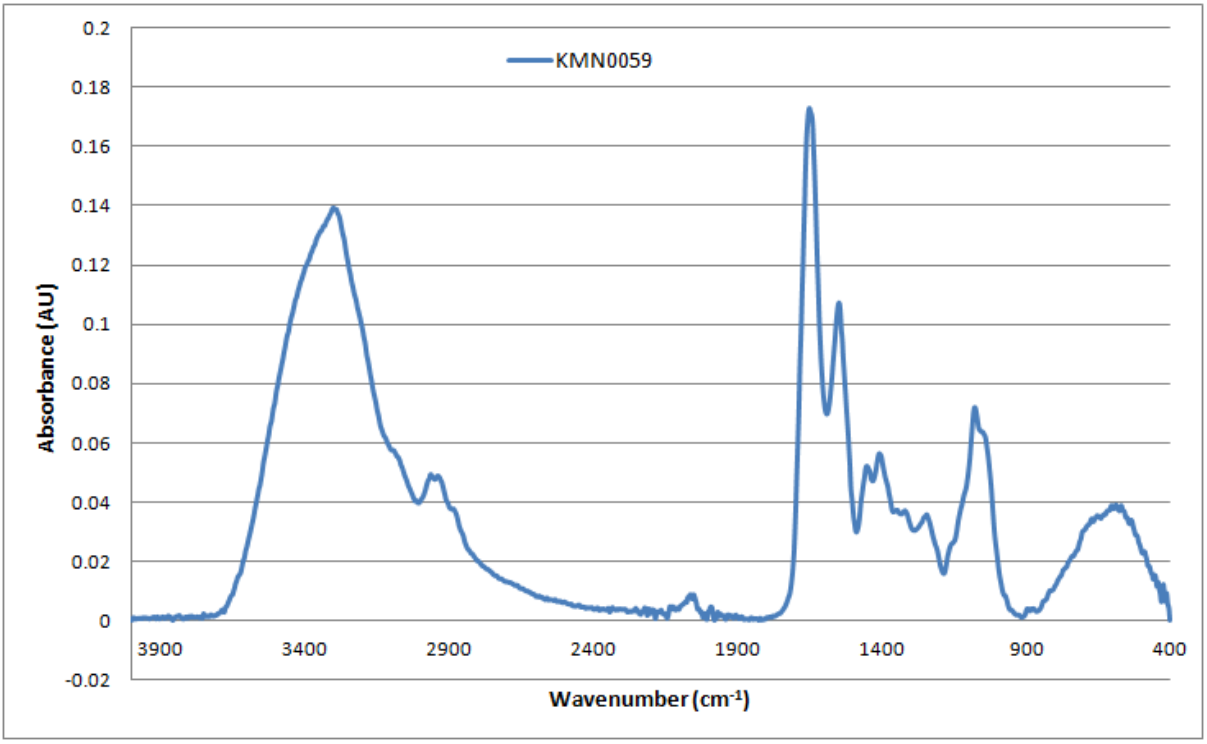


**Figure A2-37: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0059 from 4000-400cm$^{-1}$**

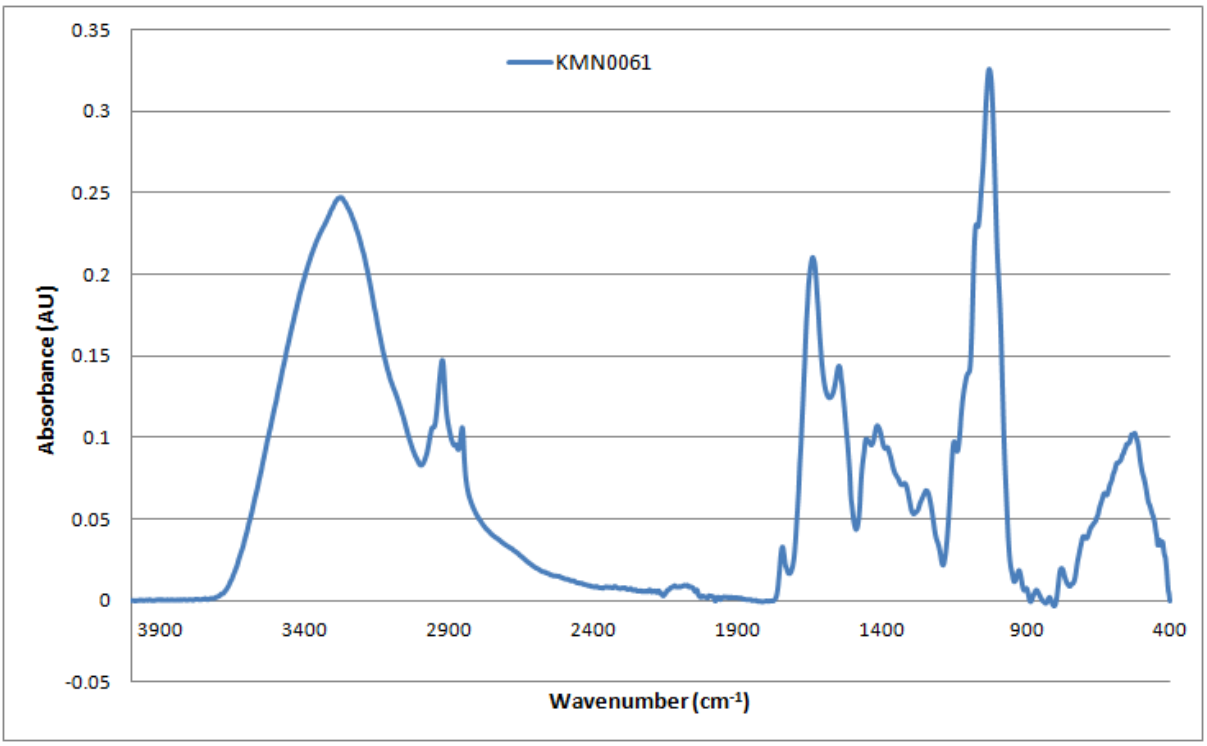


**Figure A2-38: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0061 from 4000-400cm$^{-1}$**

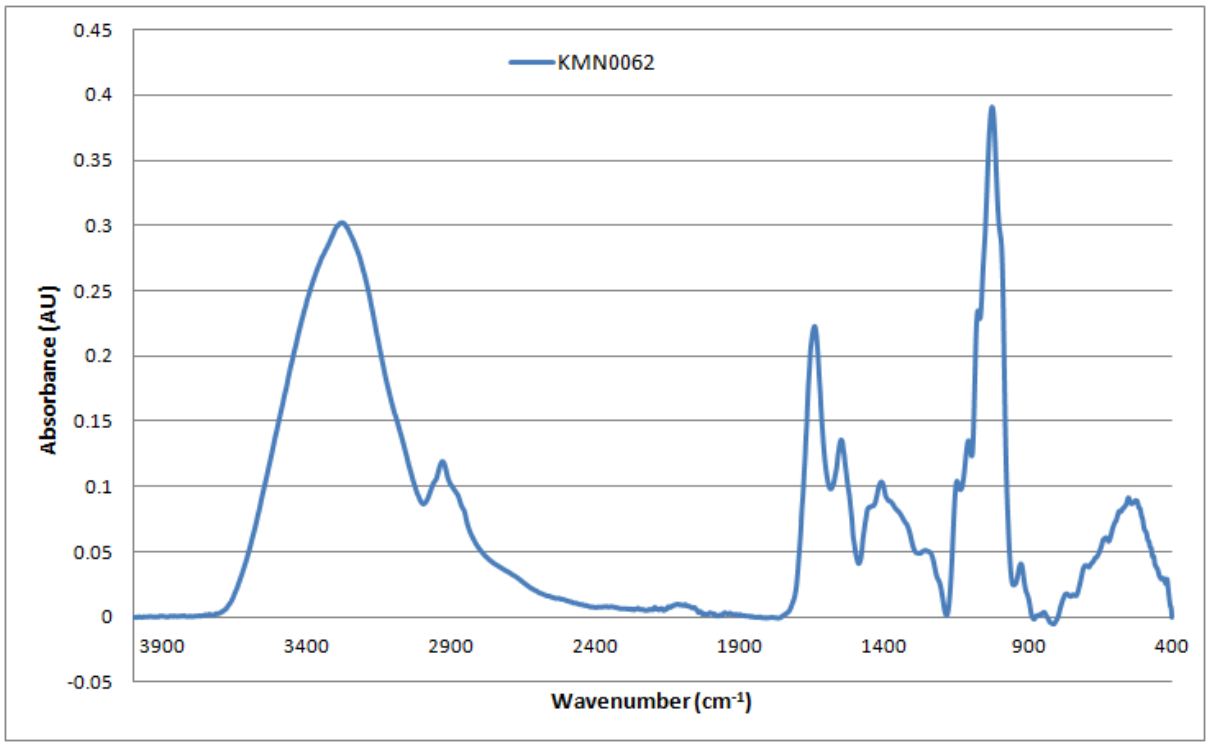


**Figure A2-39: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0062 from 4000-400cm$^{-1}$**

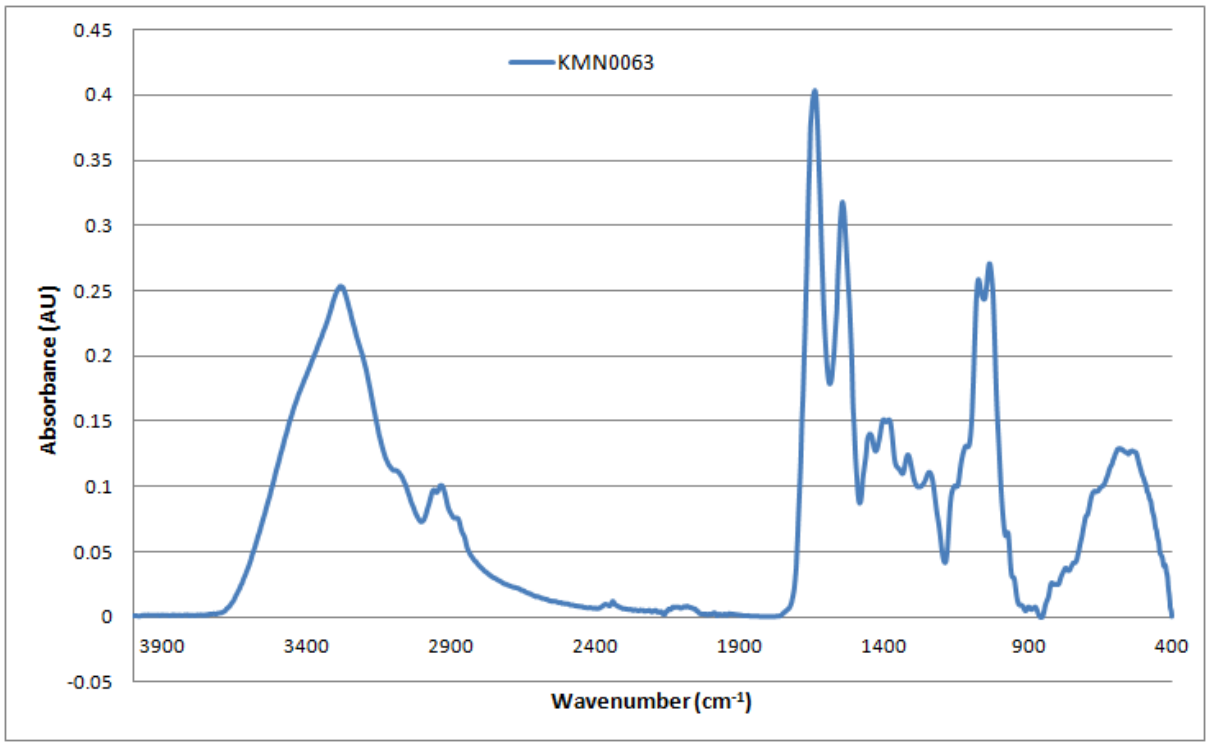


**Figure A2-40: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0063 from 4000-400cm$^{-1}$**

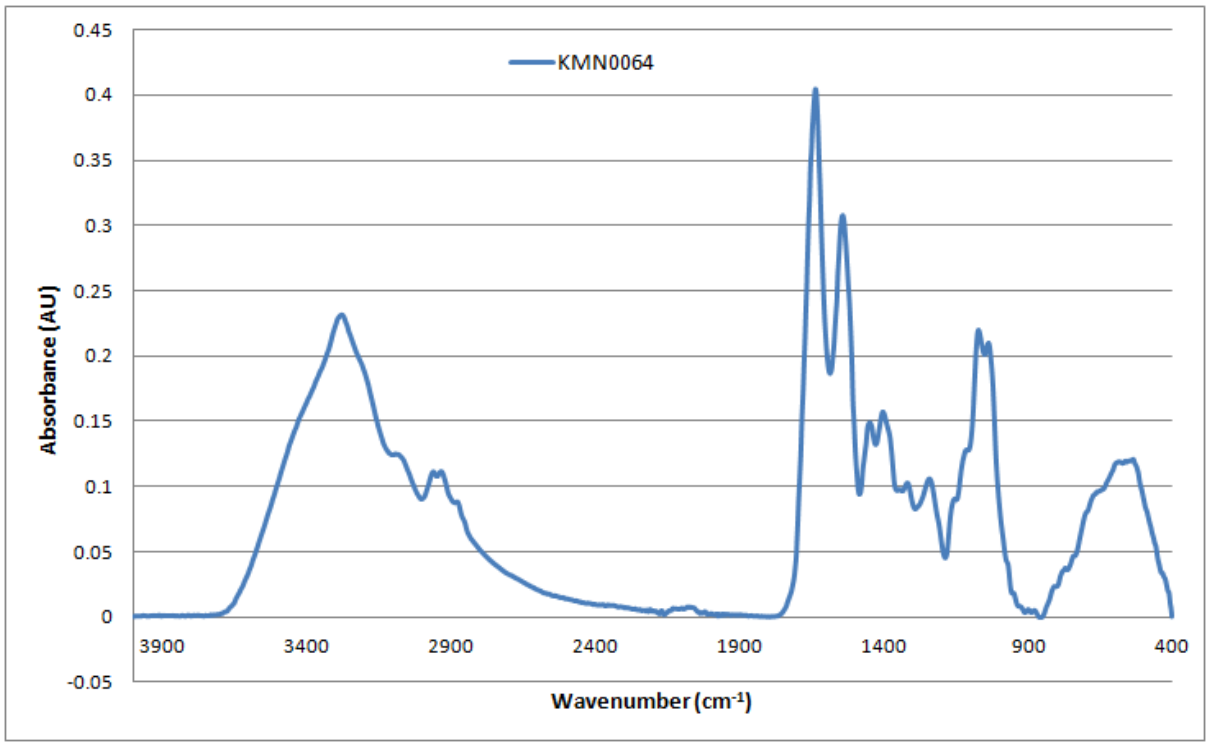


**Figure A2-41: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0064 from 4000-400cm$^{-1}$**

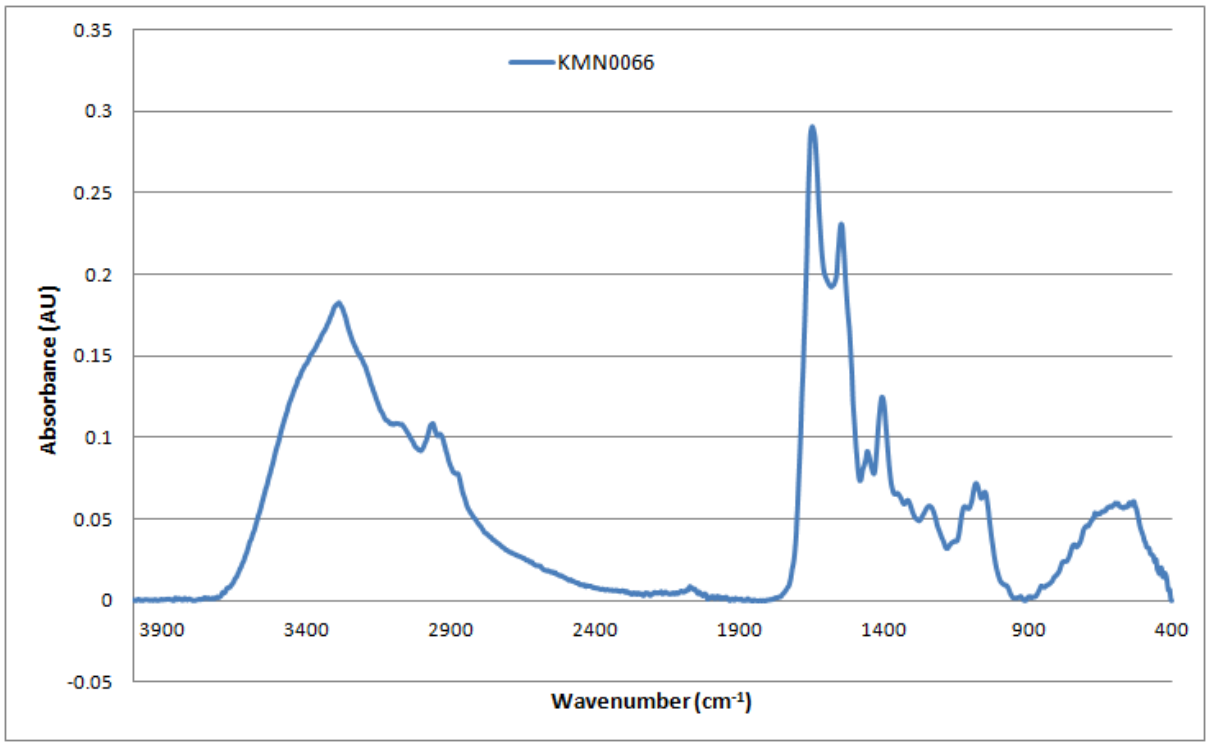


**Figure A2-42: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0066 from 4000-400cm$^{-1}$**

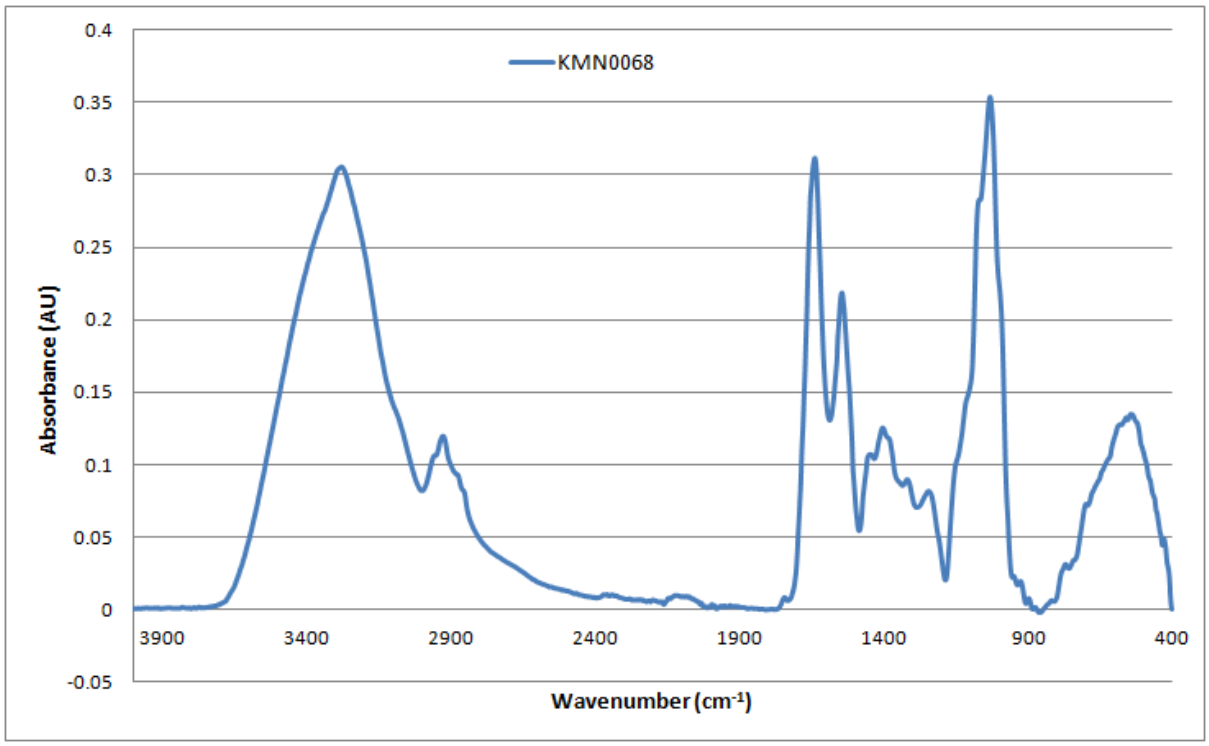


**Figure A2-43: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0068 from 4000-400cm$^{-1}$**

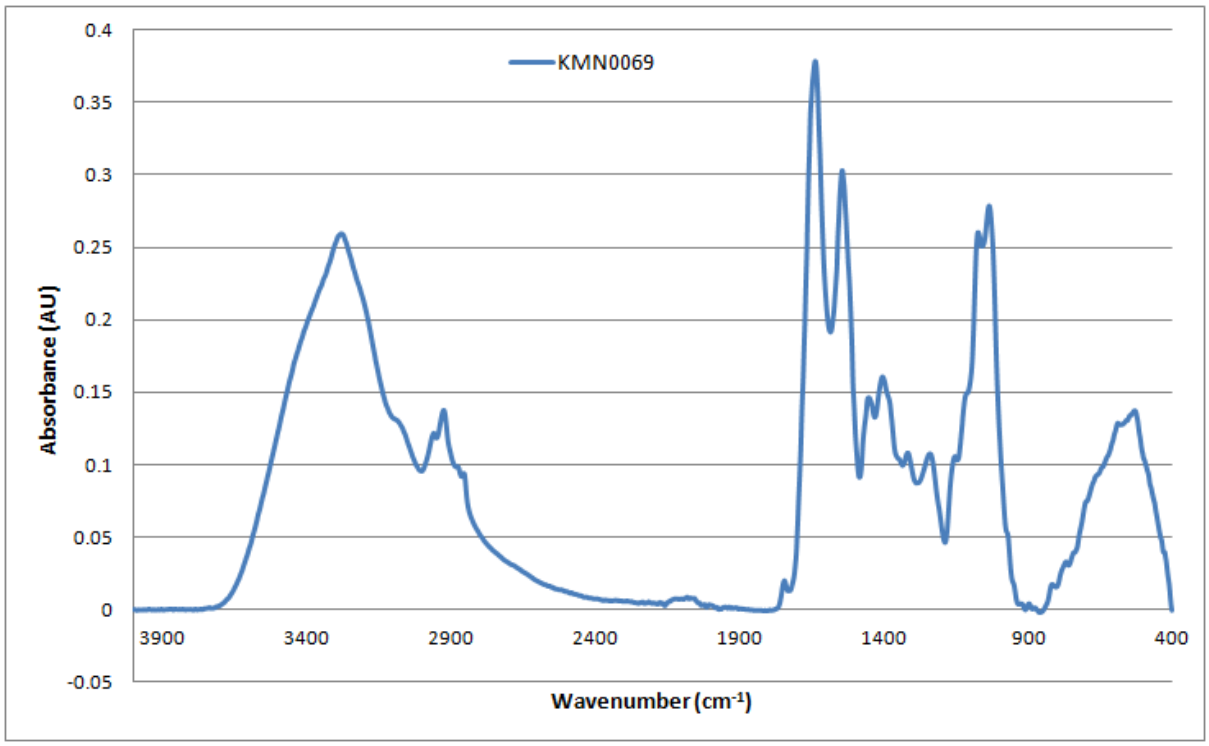


**Figure A2-44: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0069 from 4000-400cm$^{-1}$**

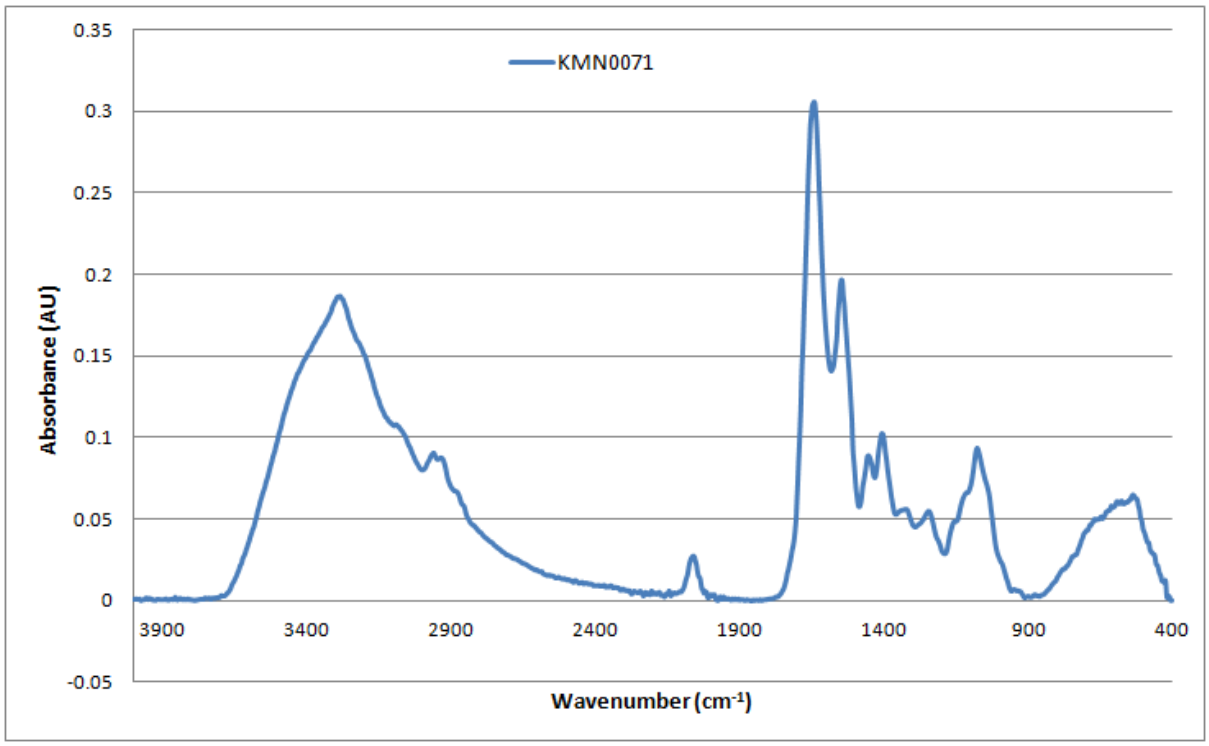


**Figure A2-45: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0071 from 4000-400cm$^{-1}$**

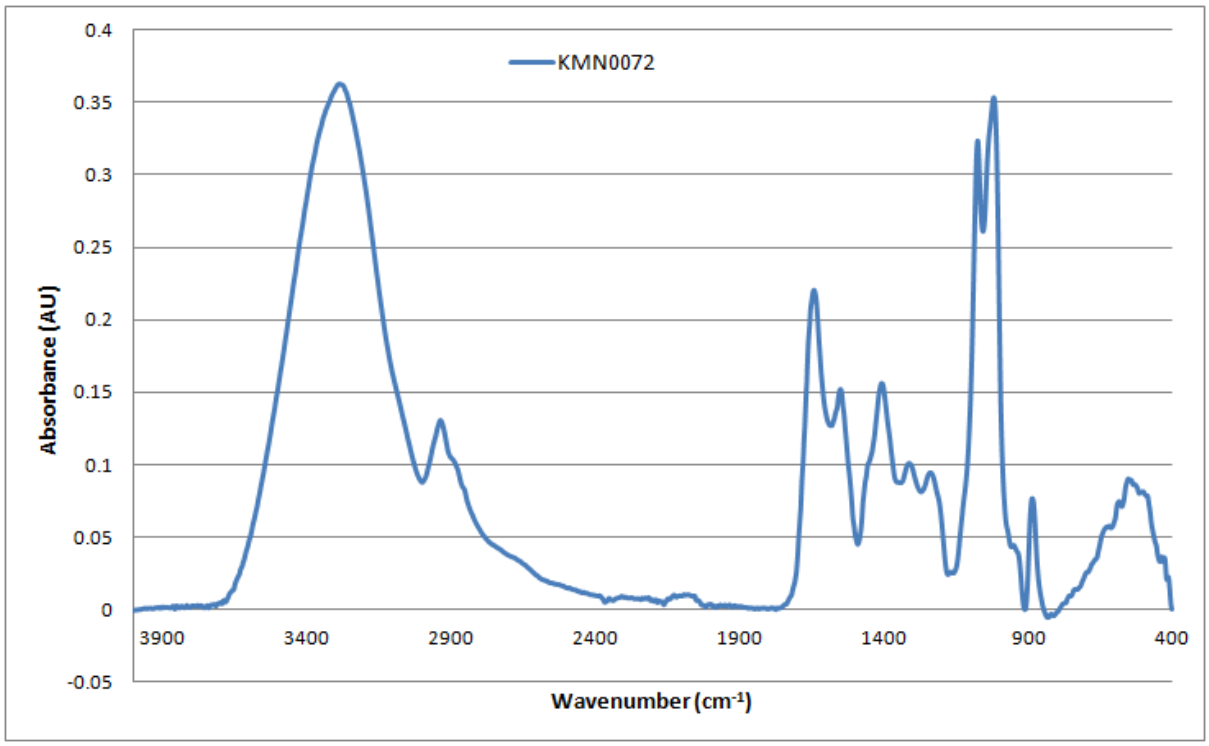


**Figure A2-46: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0072 from 4000-400cm$^{-1}$**

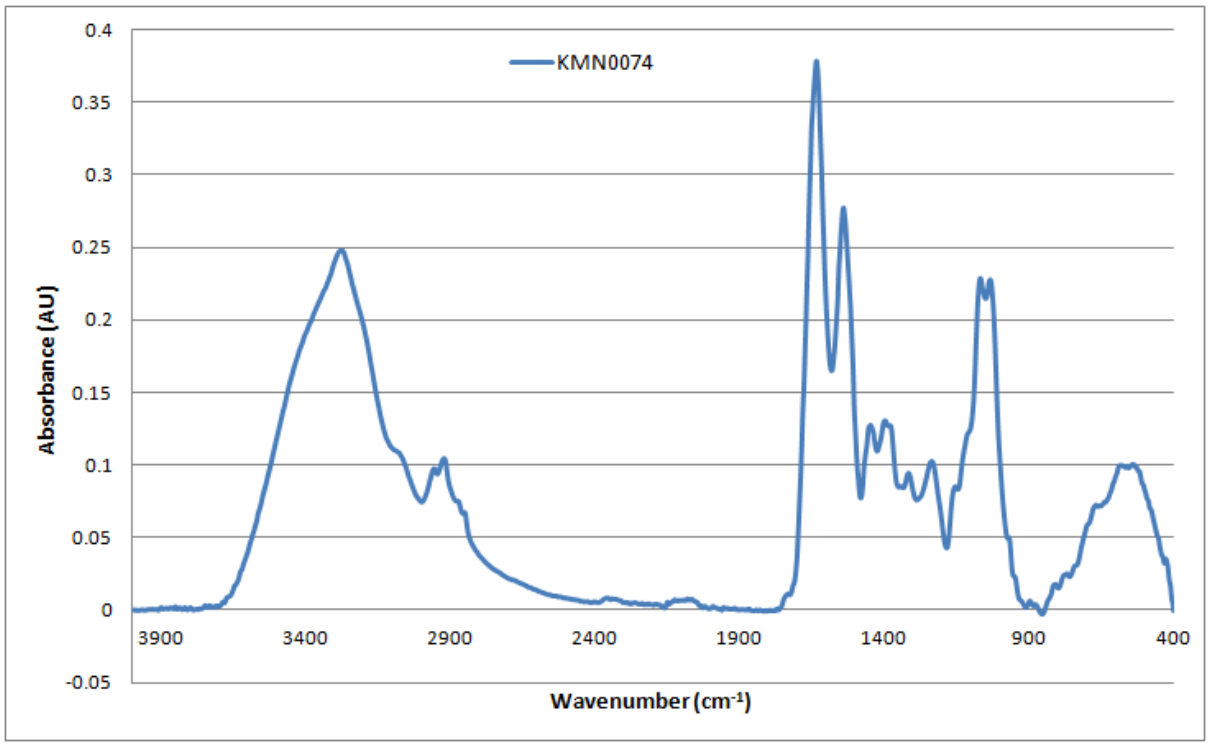


**Figure A2-47: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0074 from 4000-400cm$^{-1}$**

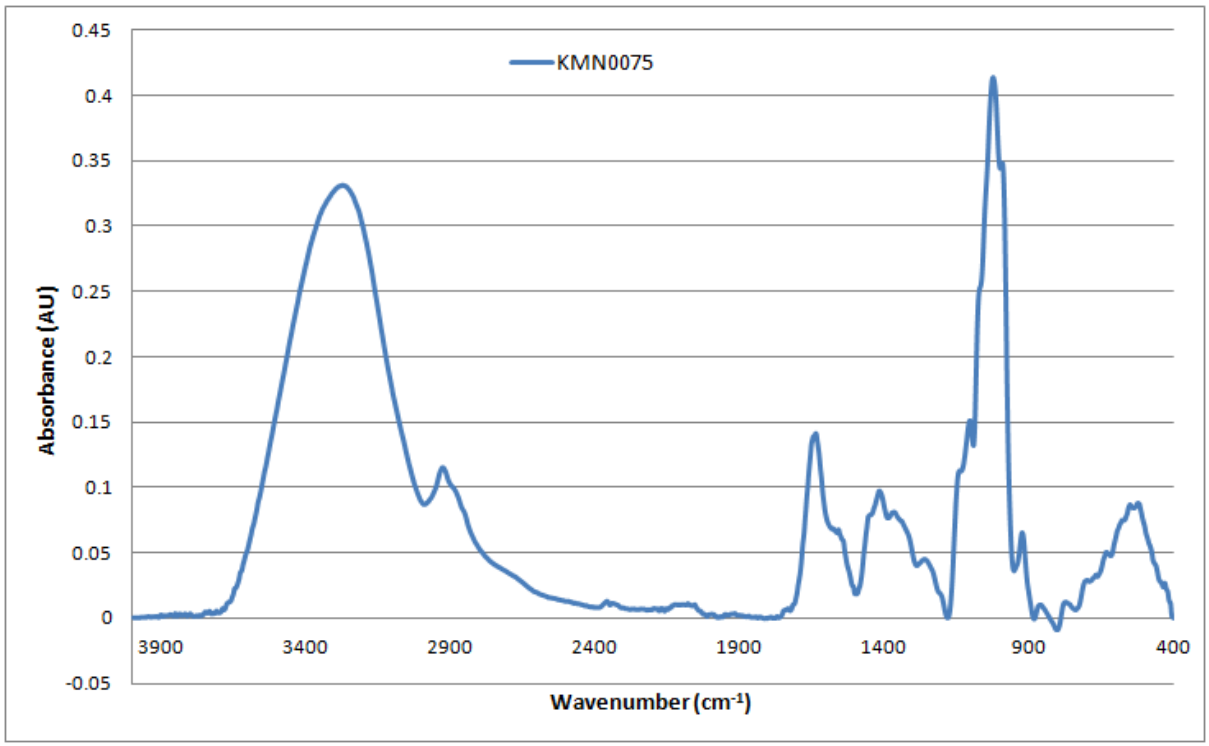


**Figure A2-48: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0075 from 4000-400cm$^{-1}$**

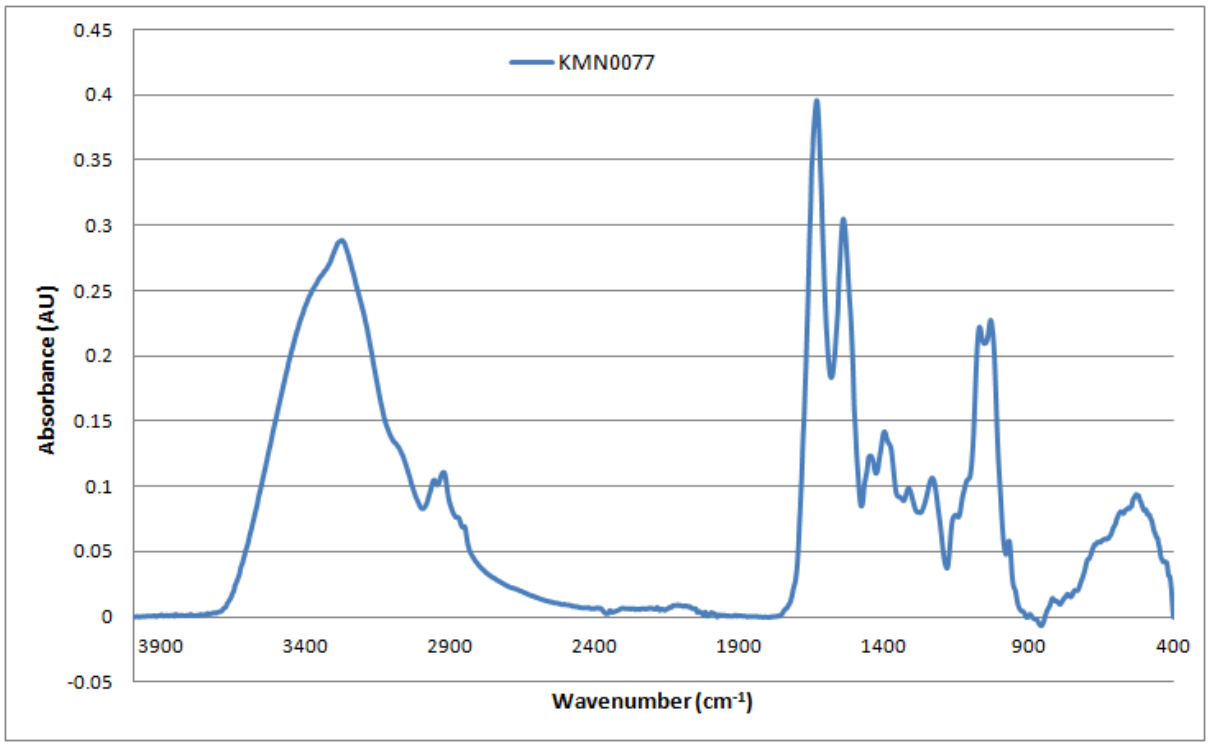


**Figure A2-49: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0077 from 4000-400cm$^{-1}$**

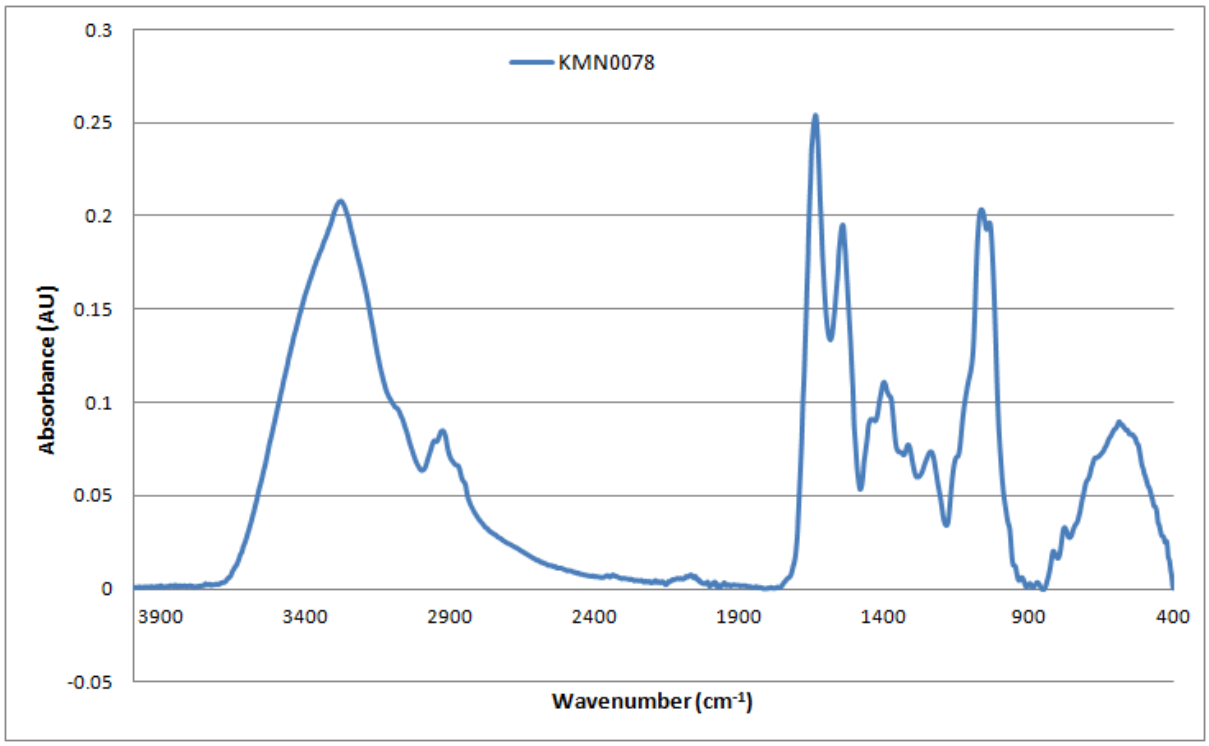


**Figure A2-50: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0078 from 4000-400cm$^{-1}$**

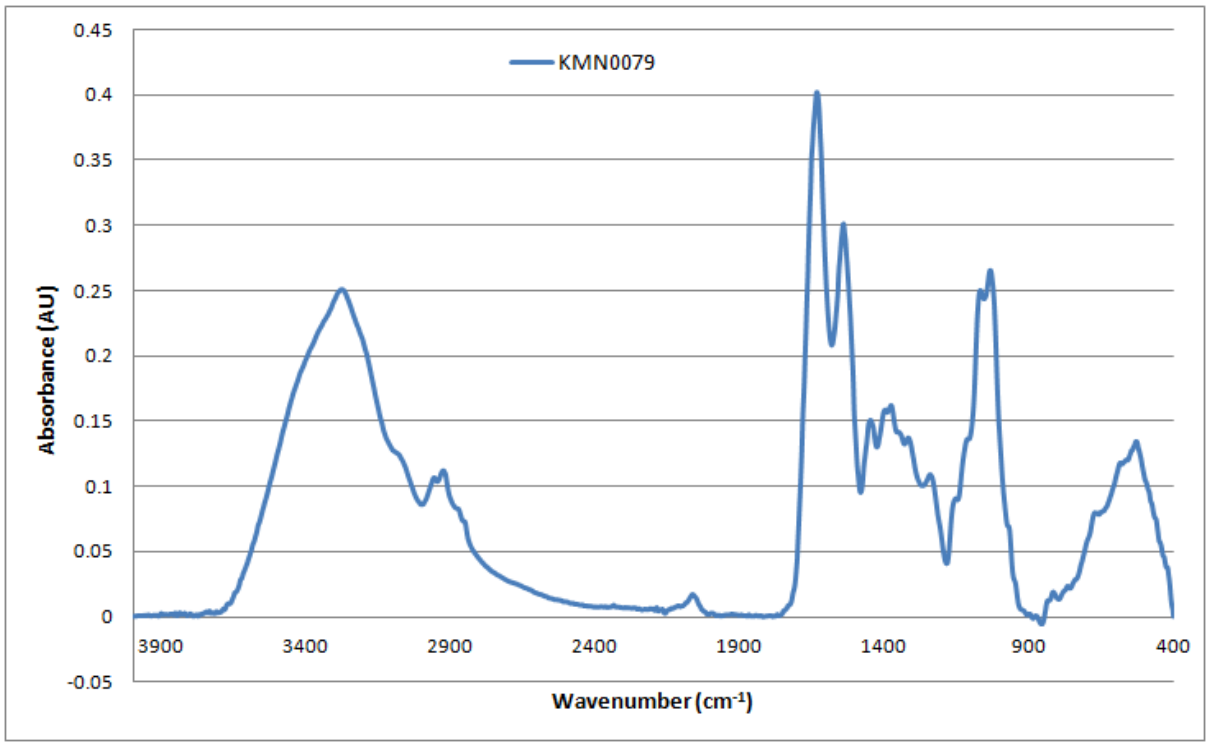


**Figure A2-51: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0079 from 4000-400cm$^{-1}$**

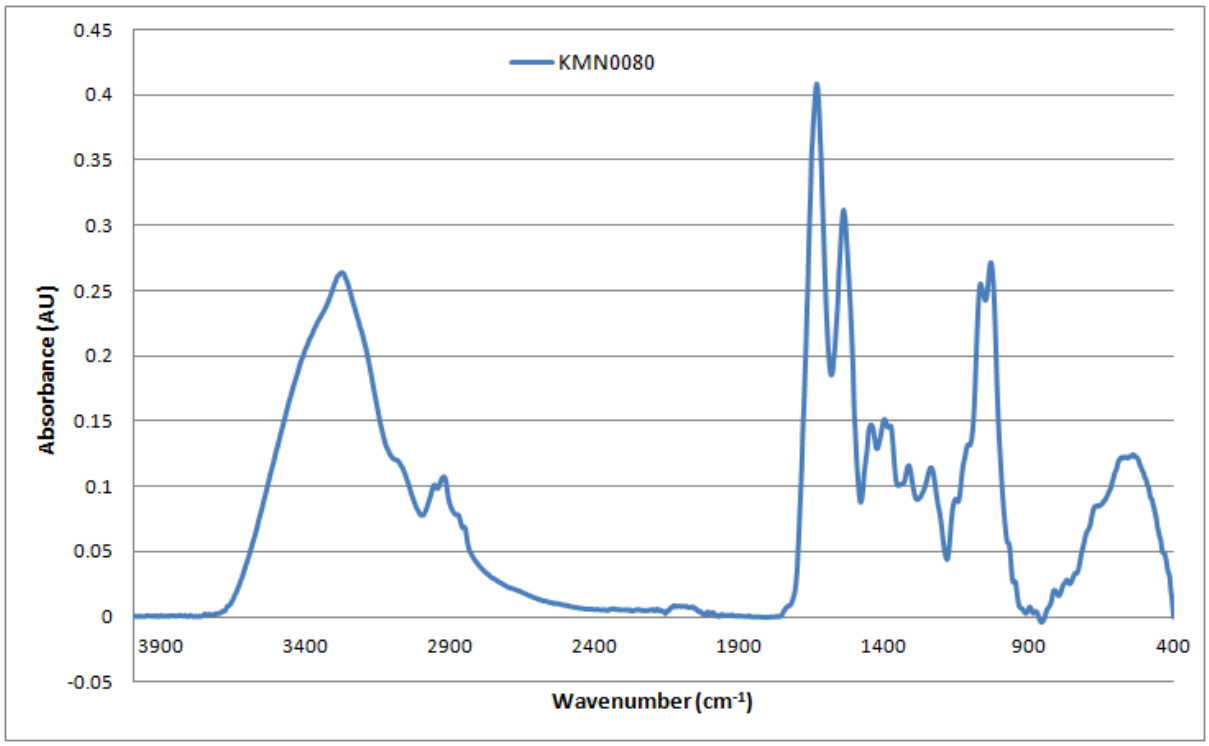


**Figure A2-52: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0080 from 4000-400cm$^{-1}$**

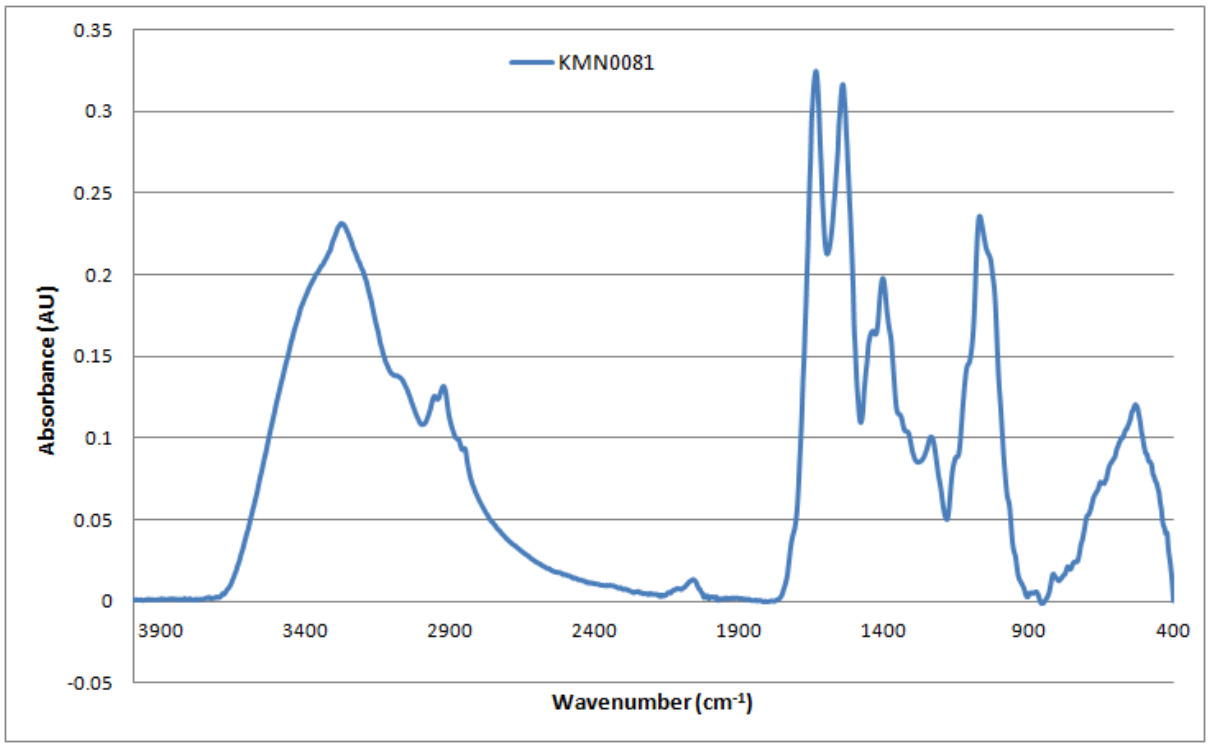


**Figure A2-53: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0081 from 4000-400cm$^{-1}$**

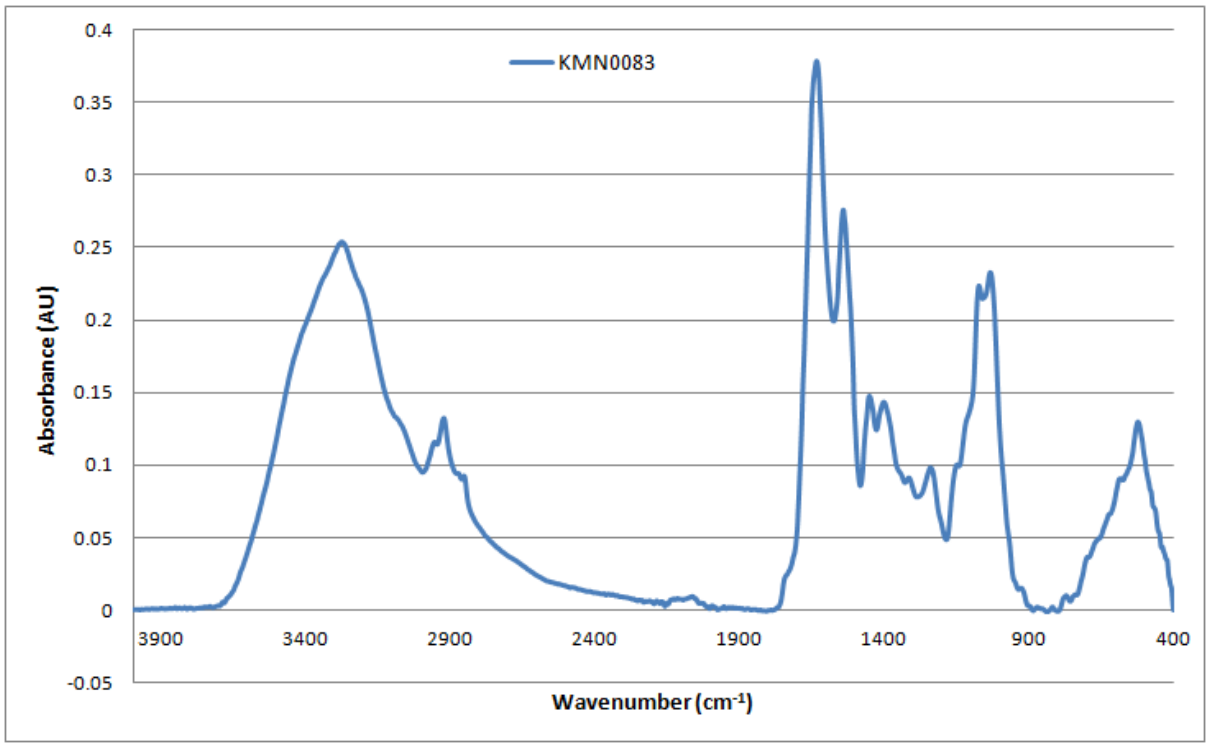


**Figure A2-54: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0083 from 4000-400cm$^{-1}$**

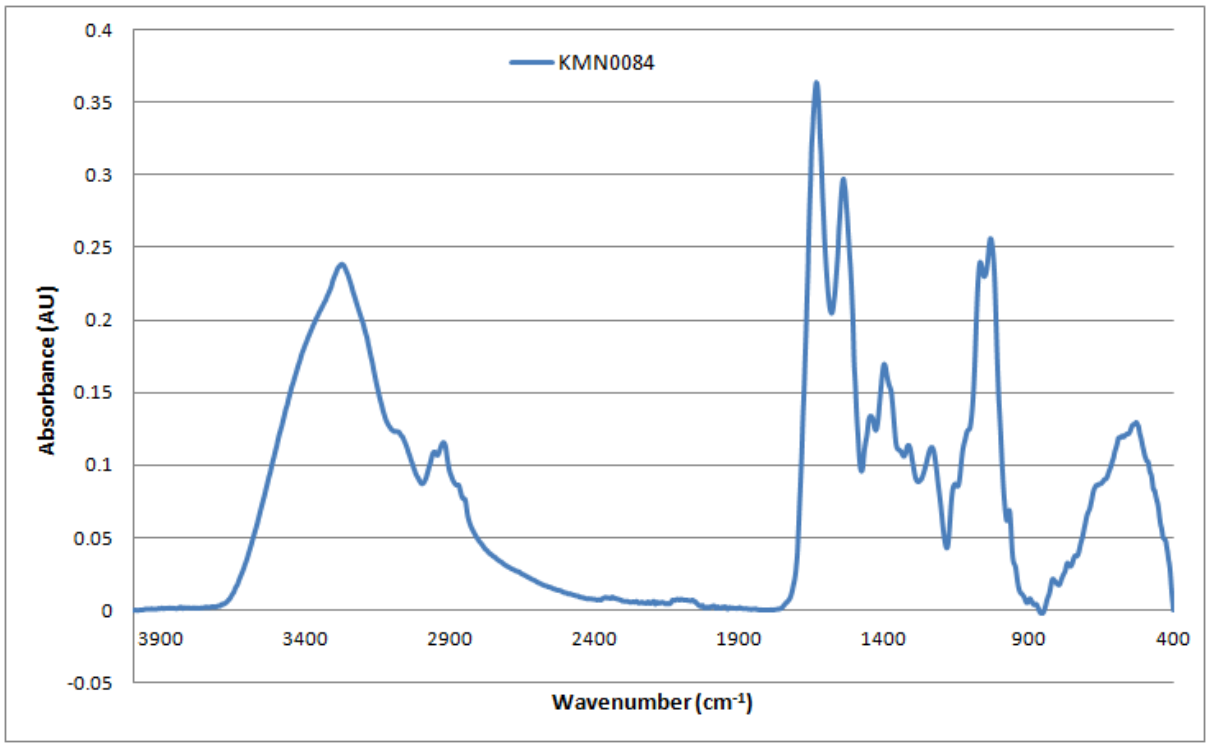


**Figure A2-55: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0084 from 4000-400cm$^{-1}$**

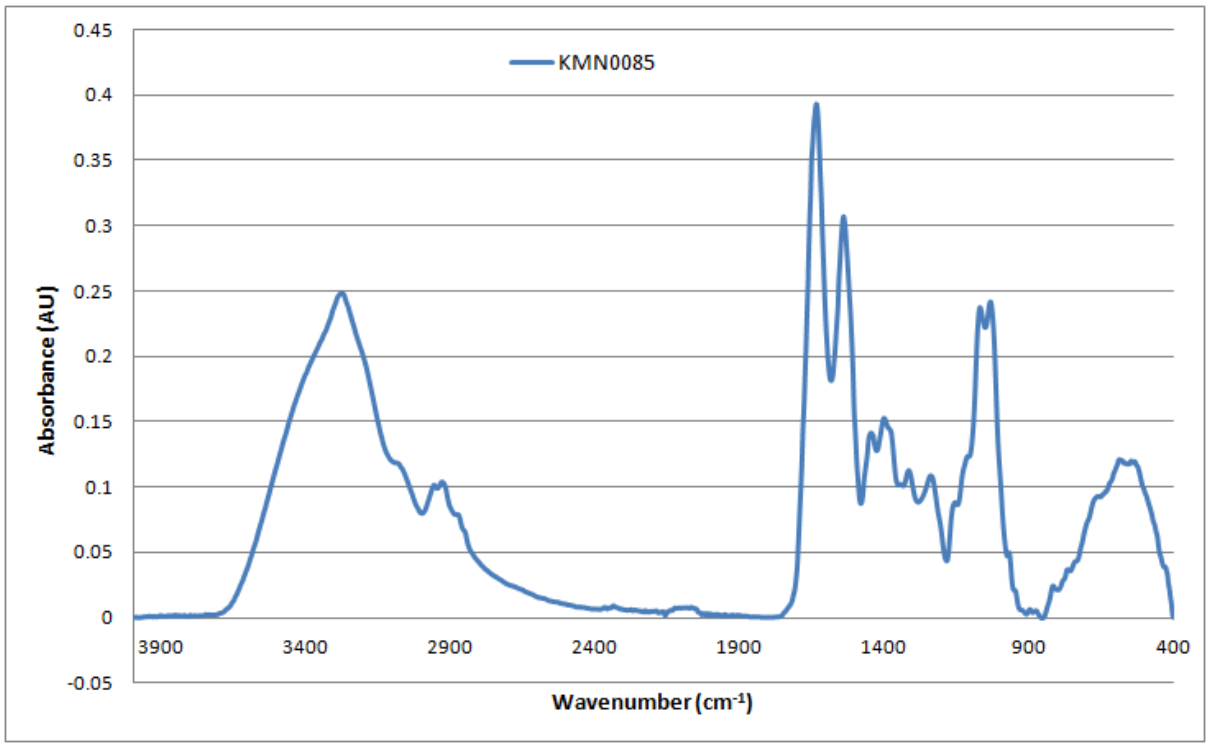


**Figure A2-56: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0085 from 4000-400cm$^{-1}$**

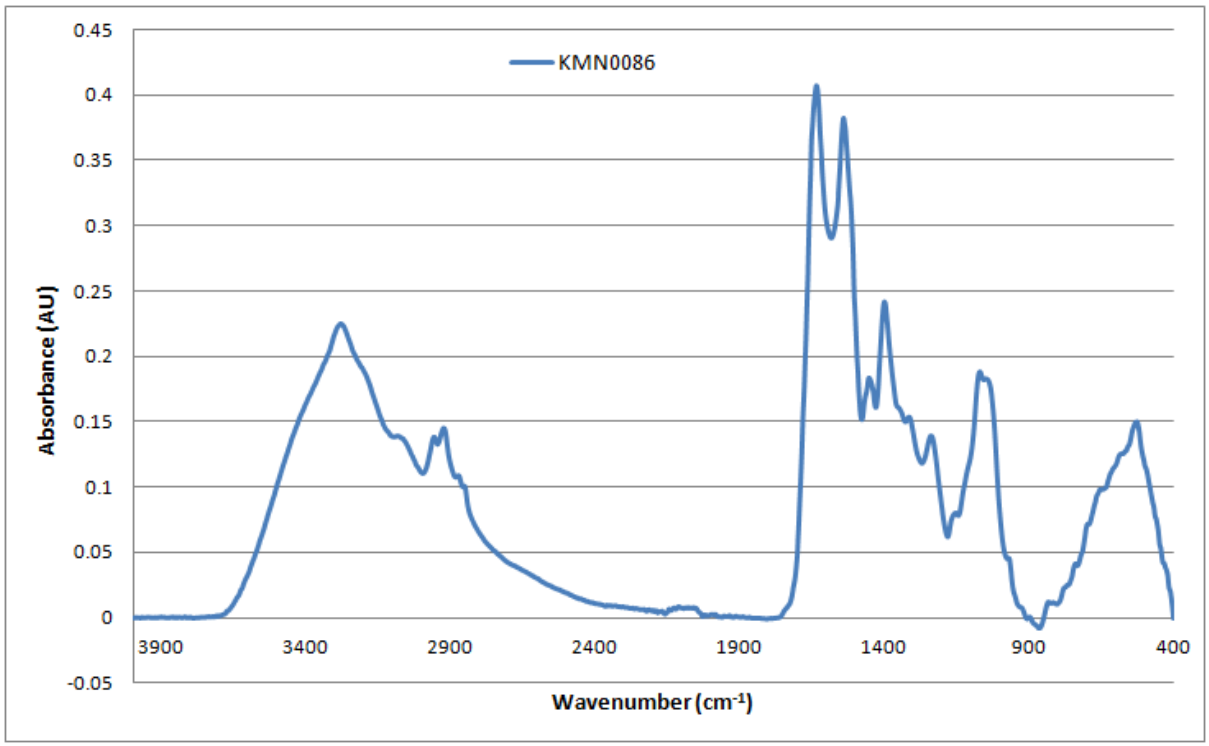


**Figure A2-57: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0086 from 4000-400cm$^{-1}$**

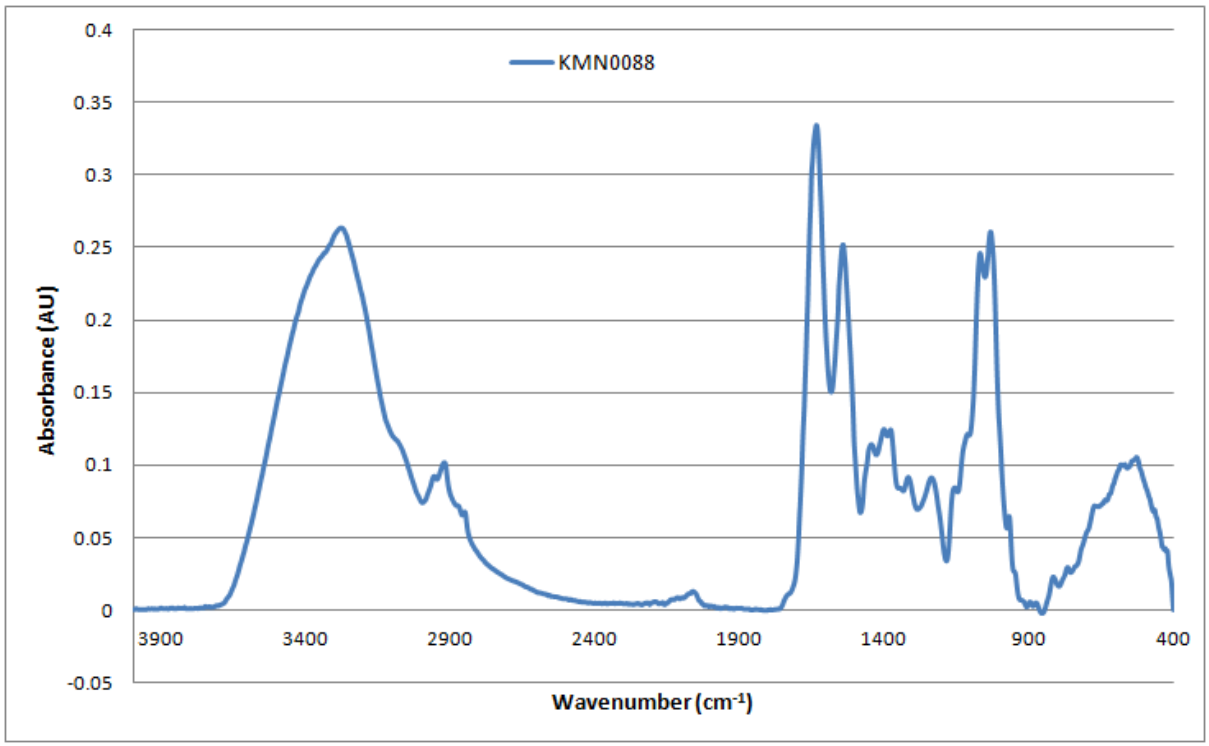


**Figure A2-58: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0088 from 4000-400cm$^{-1}$**

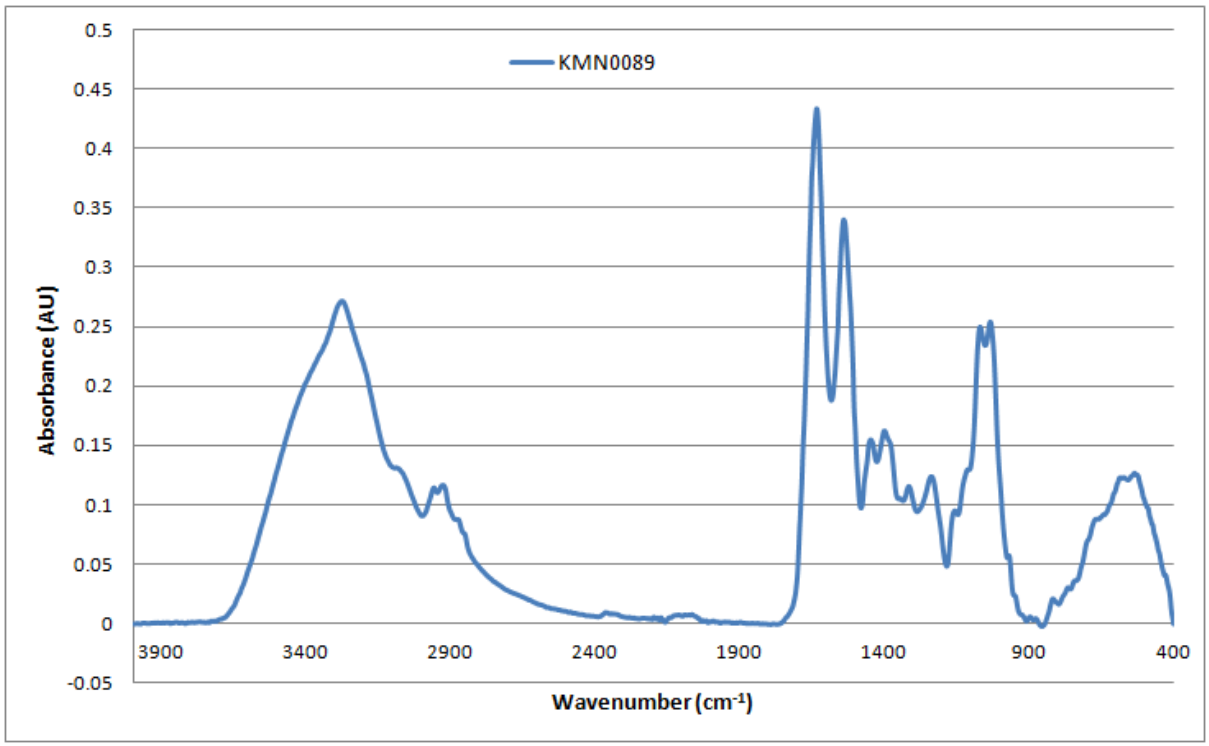


**Figure A2-59: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0089 from 4000-400cm$^{-1}$**

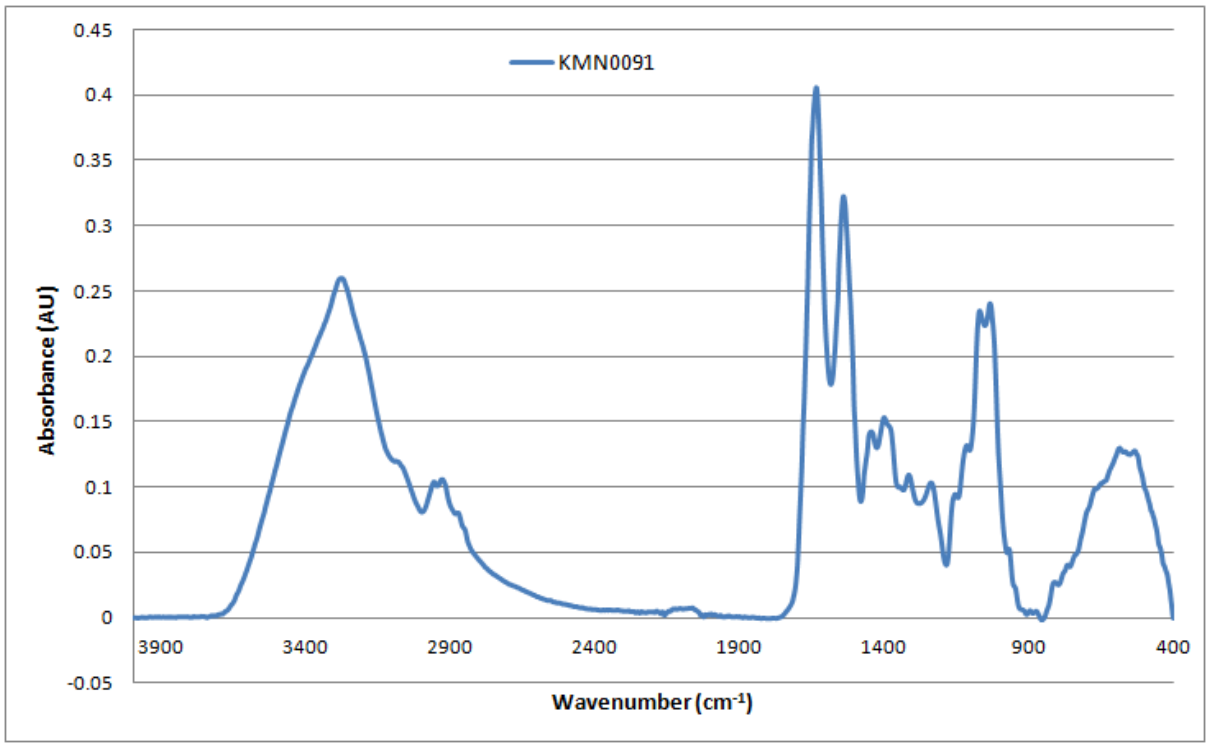


**Figure A2-60: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0091 from 4000-400cm$^{-1}$**

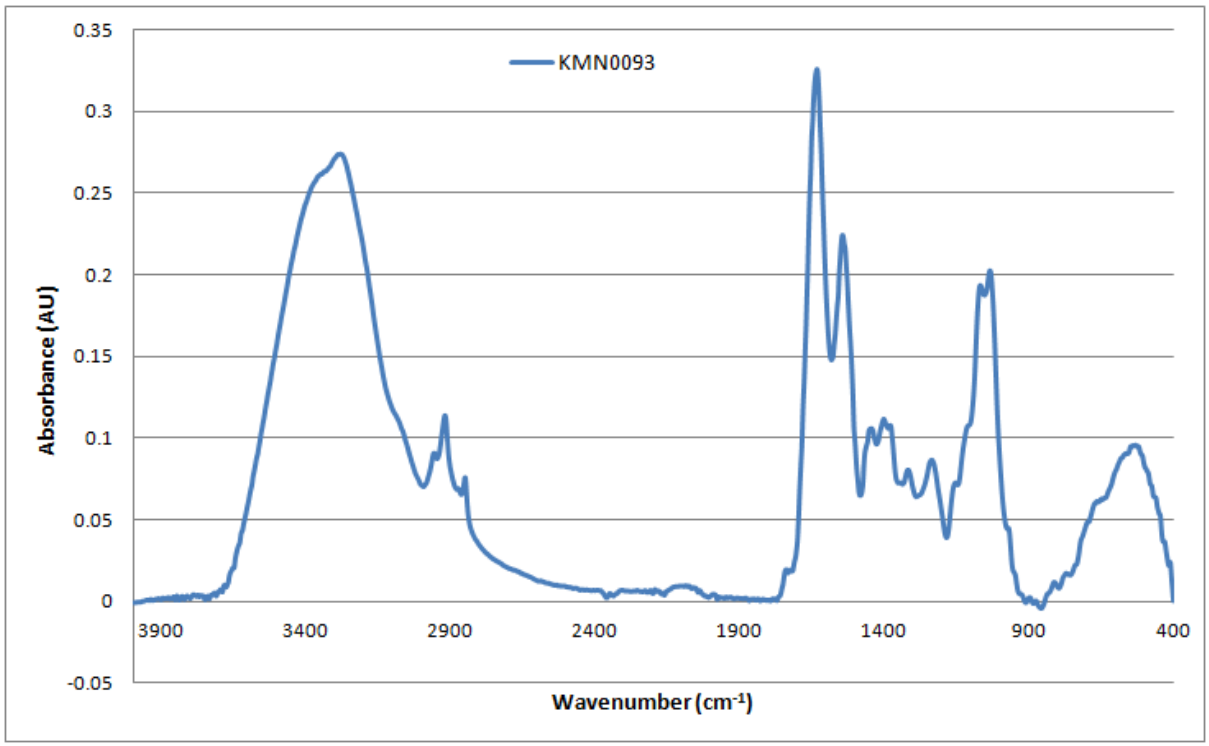


**Figure A2-61: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0093 from 4000-400cm$^{-1}$**

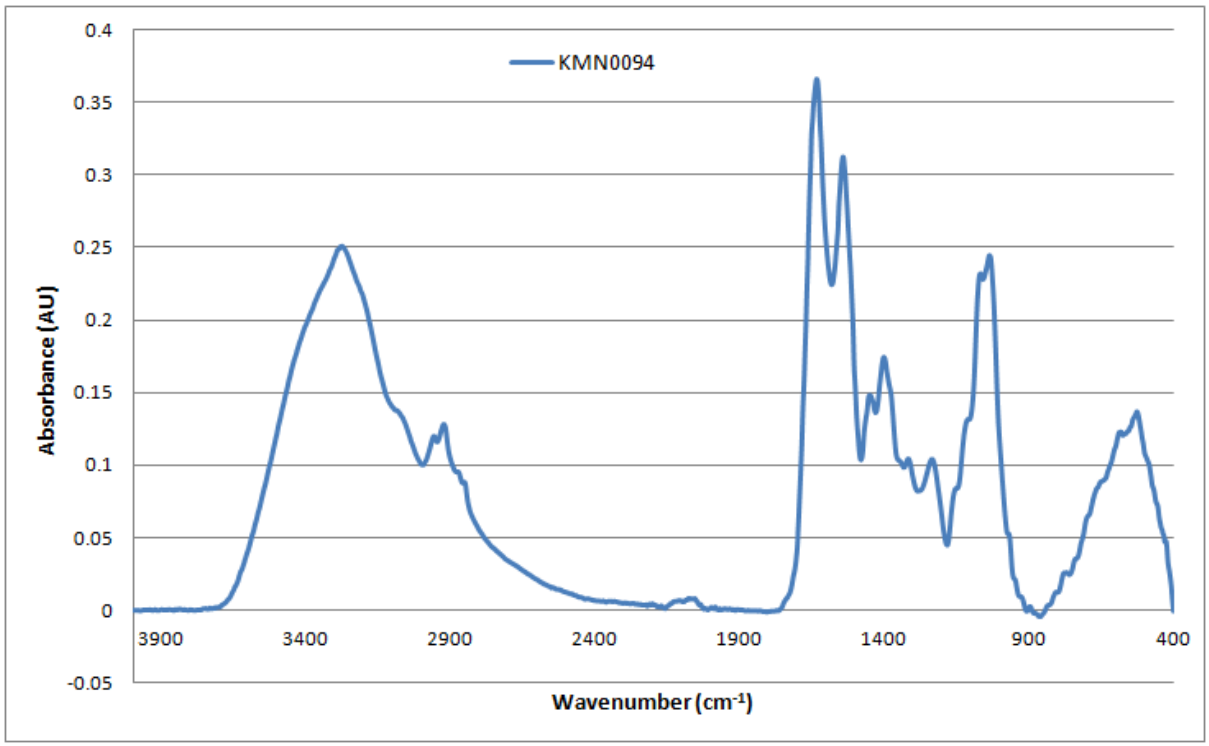


**Figure A2-62: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0094 from 4000-400cm$^{-1}$**

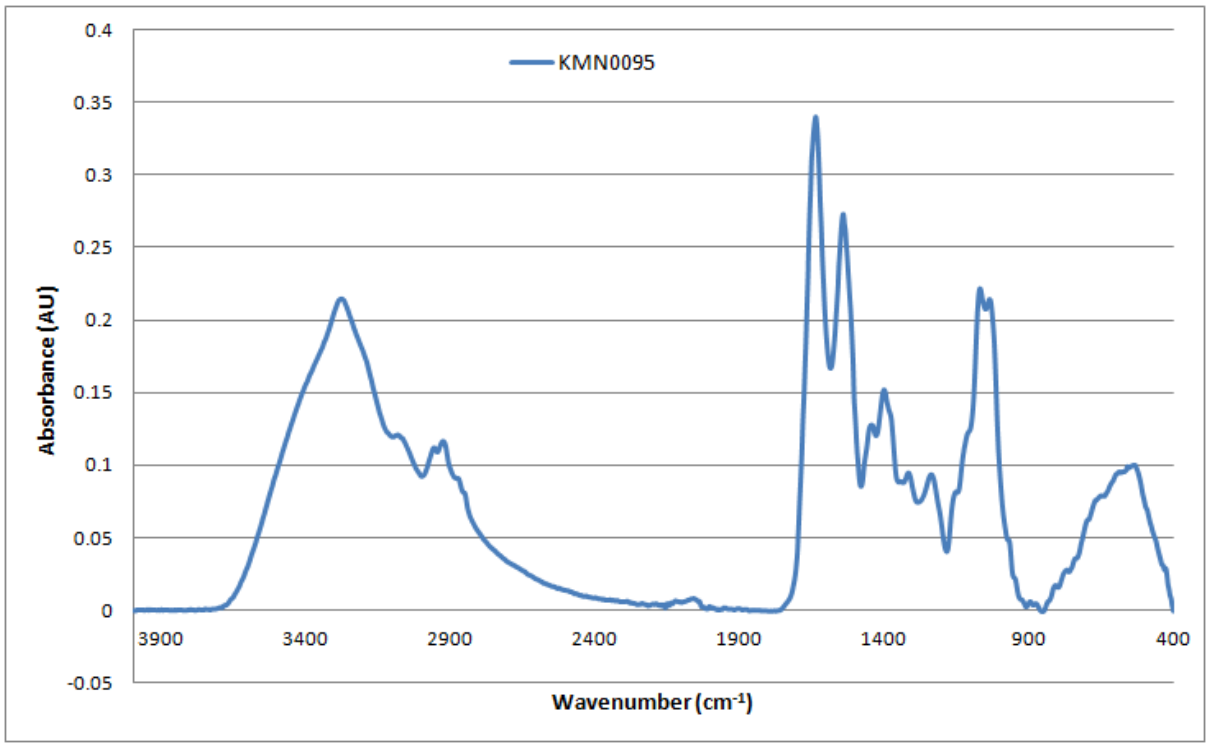


**Figure A2-63: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0095 from 4000-400cm$^{-1}$**

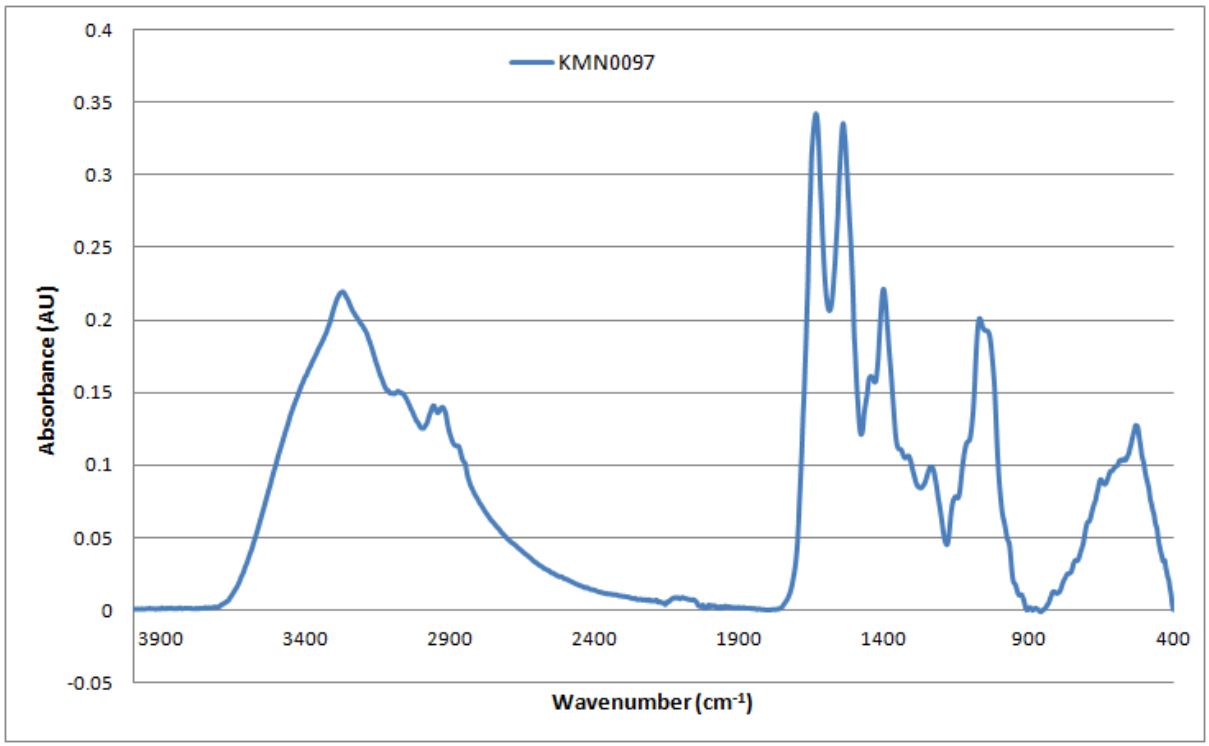


**Figure A2-64: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0097 from 4000-400cm$^{-1}$**

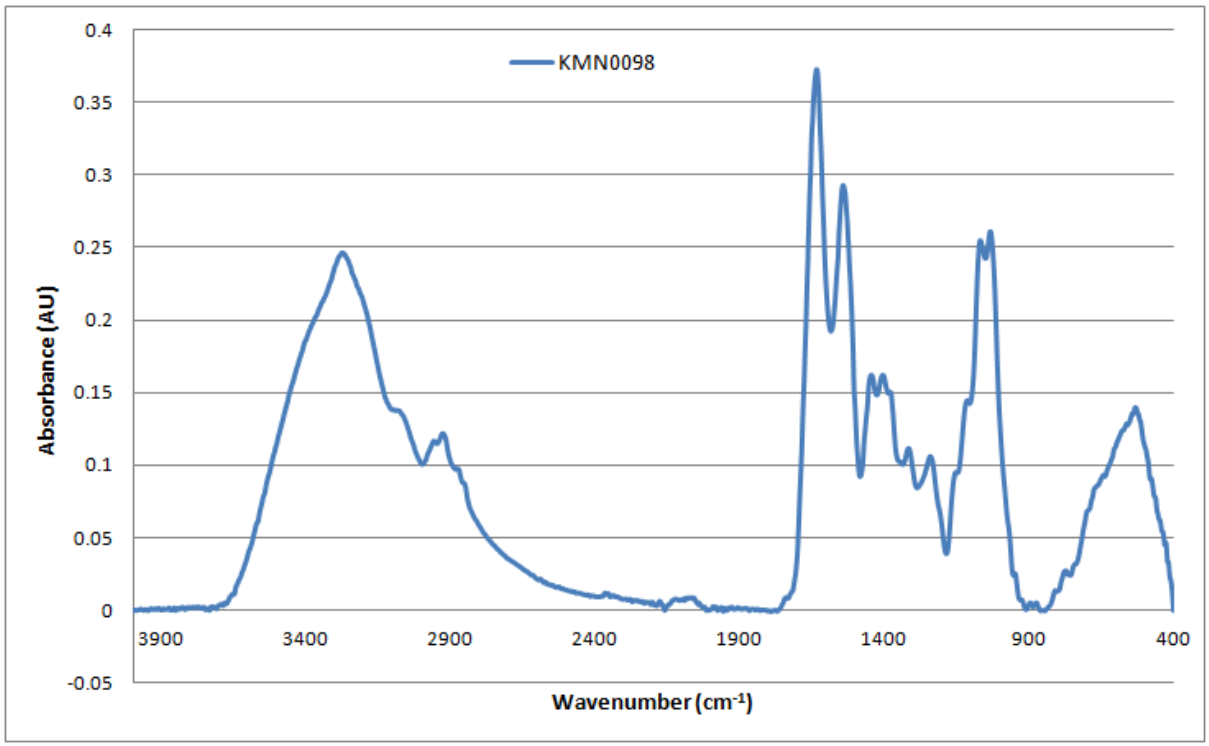


**Figure A2-65: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0098 from 4000-400cm$^{-1}$**

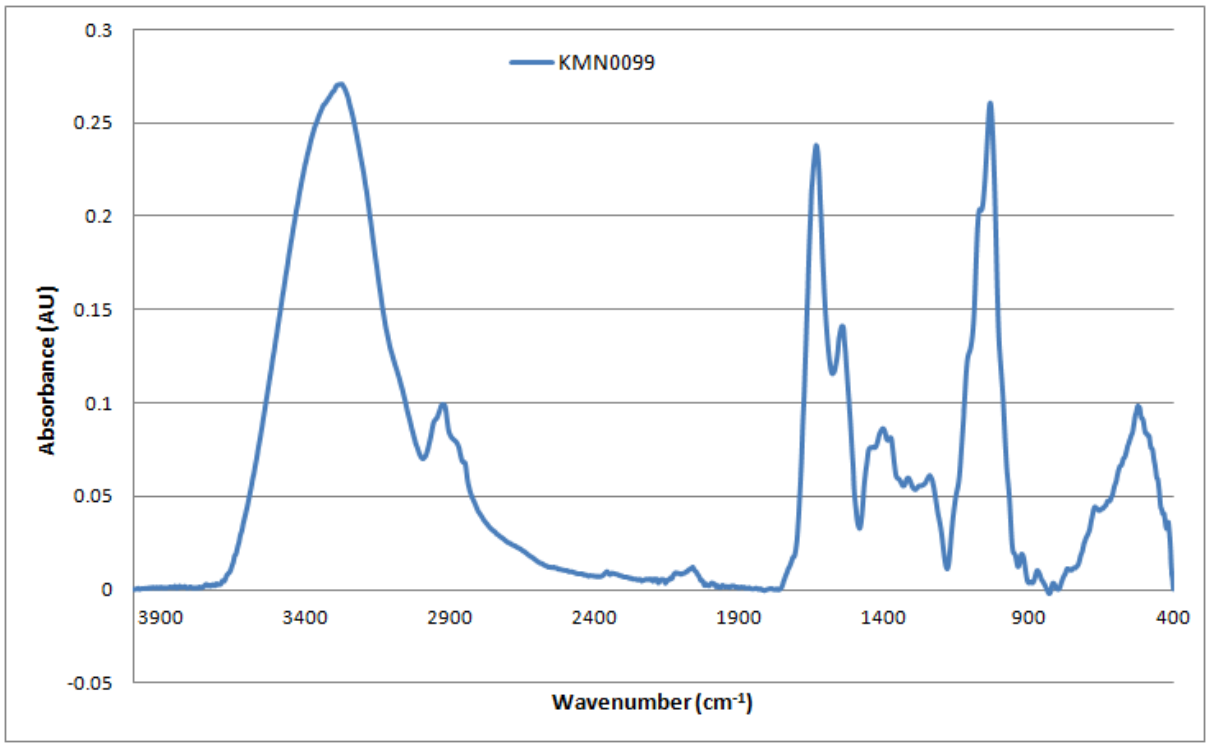


**Figure A2-66: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0099 from 4000-400cm$^{-1}$**

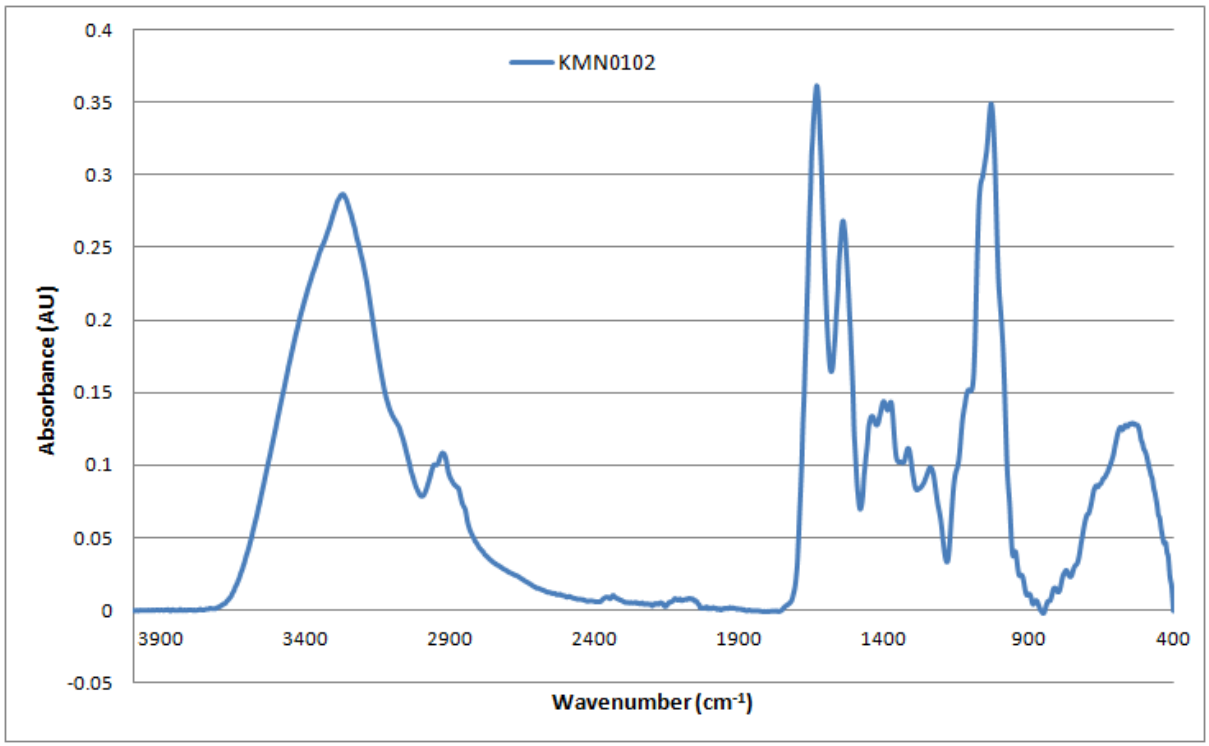


**Figure A2-67: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0102 from 4000-400cm$^{-1}$**

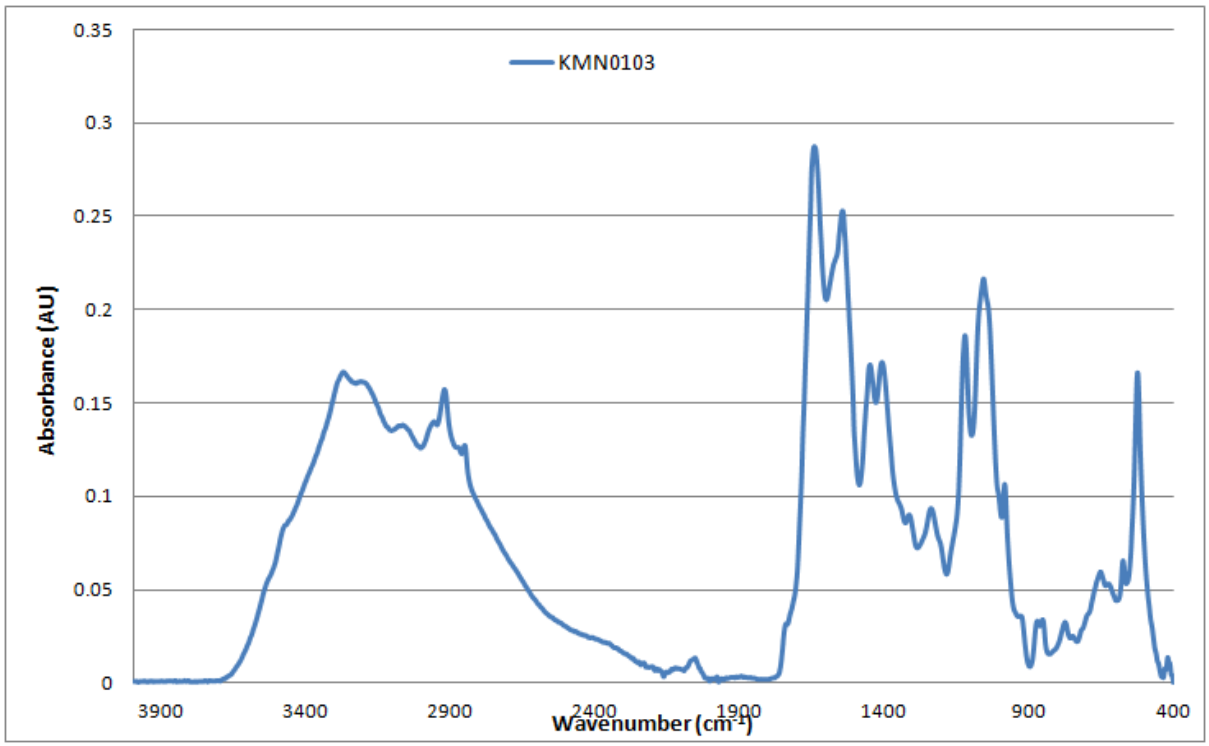


**Figure A2-68: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0103 from 4000-400cm$^{-1}$**

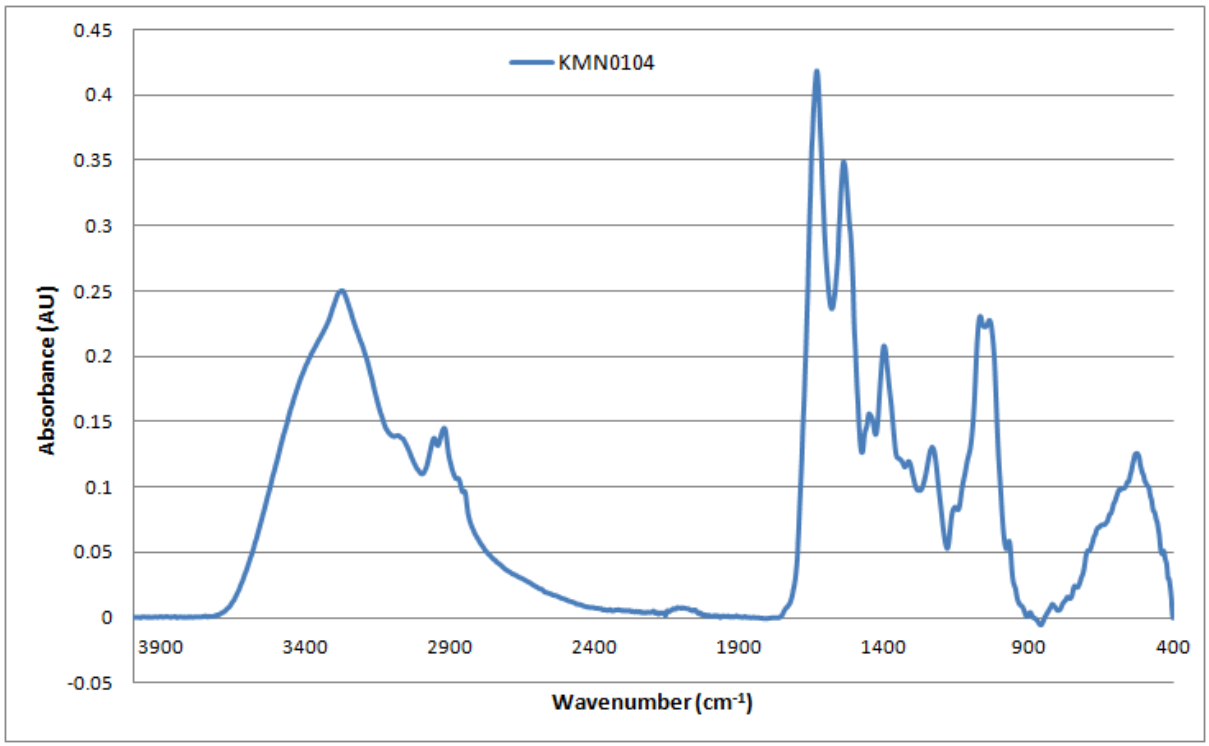


**Figure A2-69: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0104 from 4000-400cm$^{-1}$**

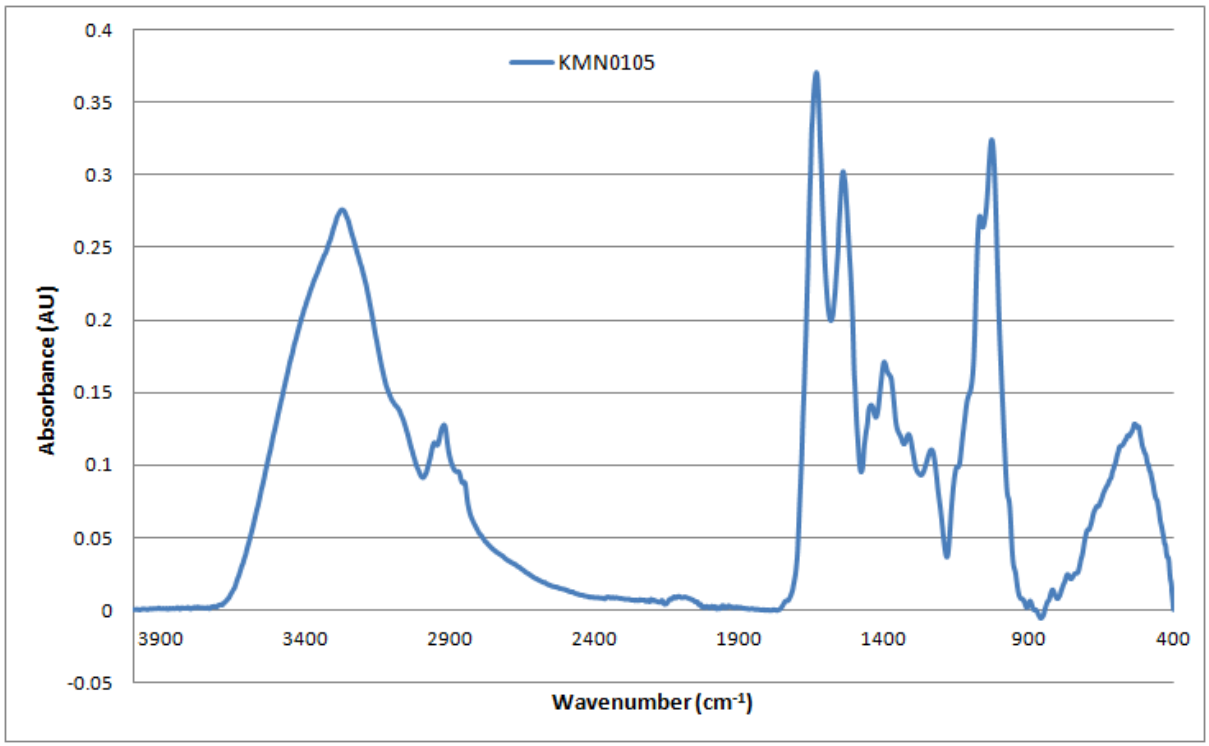


**Figure A2-70: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0105 from 4000-400cm$^{-1}$**

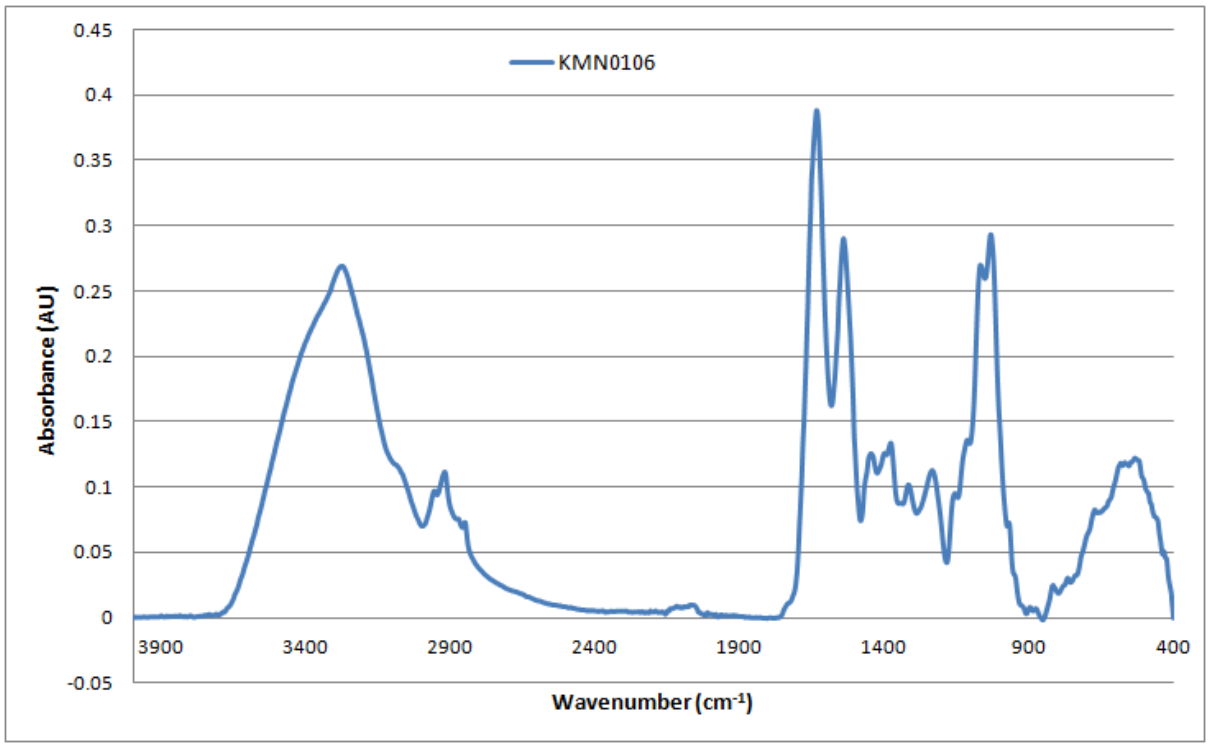


**Figure A2-71: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0106 from 4000-400cm$^{-1}$**

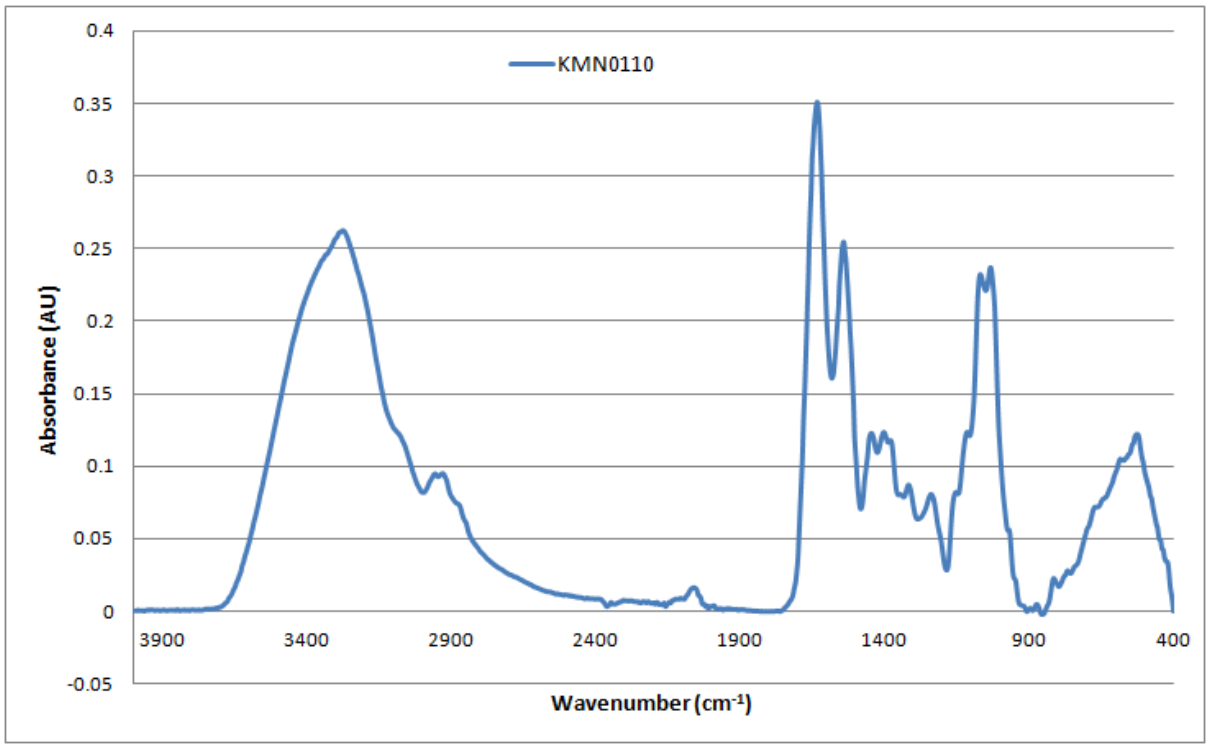


**Figure A2-72: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0110 from 4000-400cm$^{-1}$**

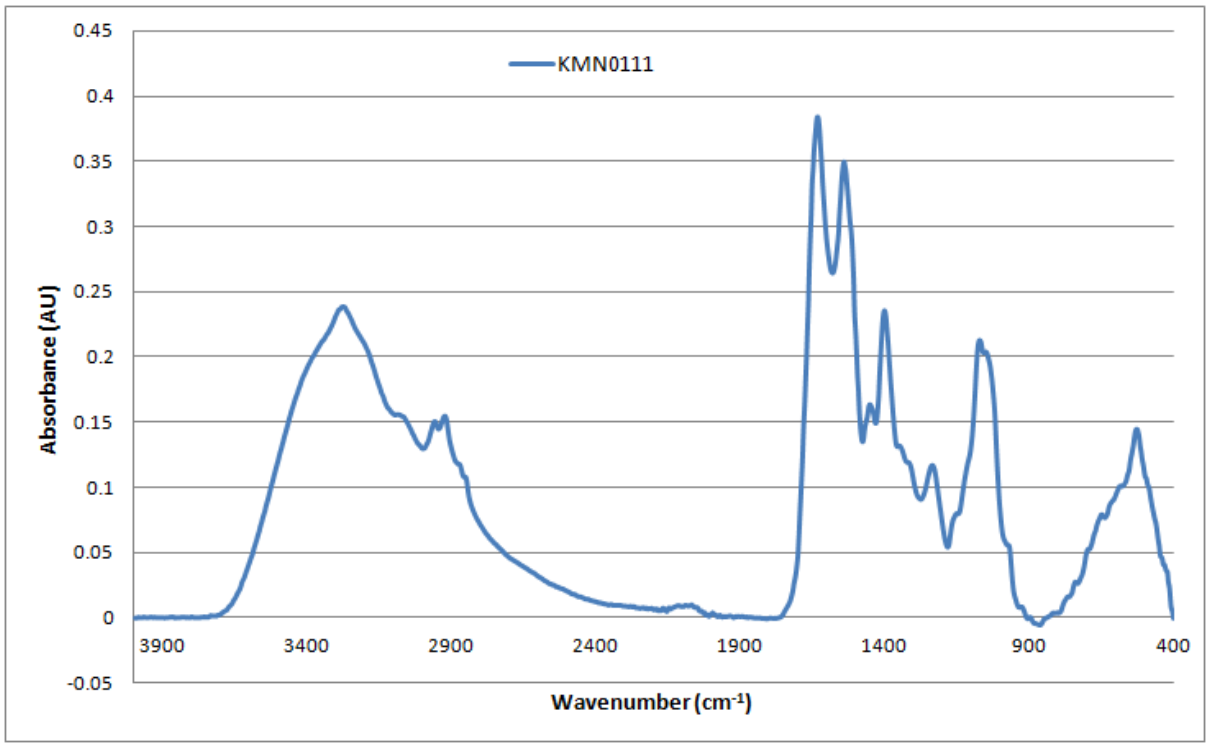


**Figure A2-73: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0111 from 4000-400cm$^{-1}$**

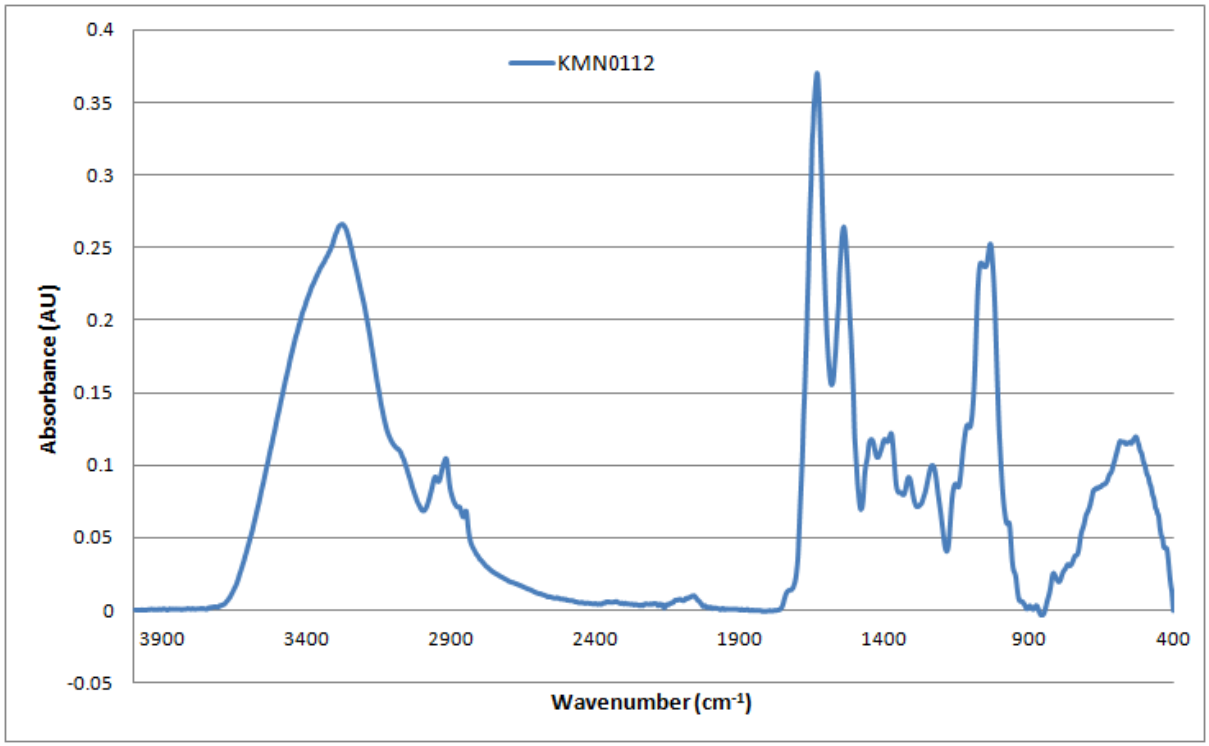


**Figure A2-74: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0112 from 4000-400cm$^{-1}$**

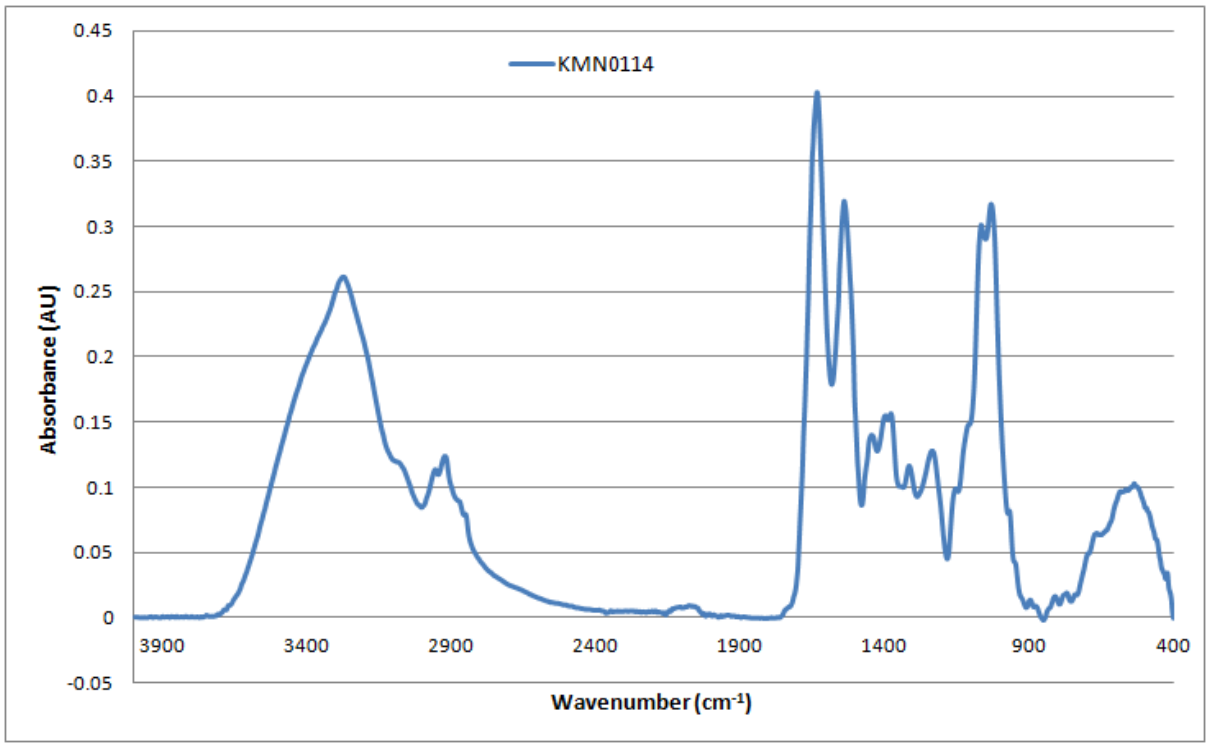


**Figure A2-75: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0114 from 4000-400cm$^{-1}$**

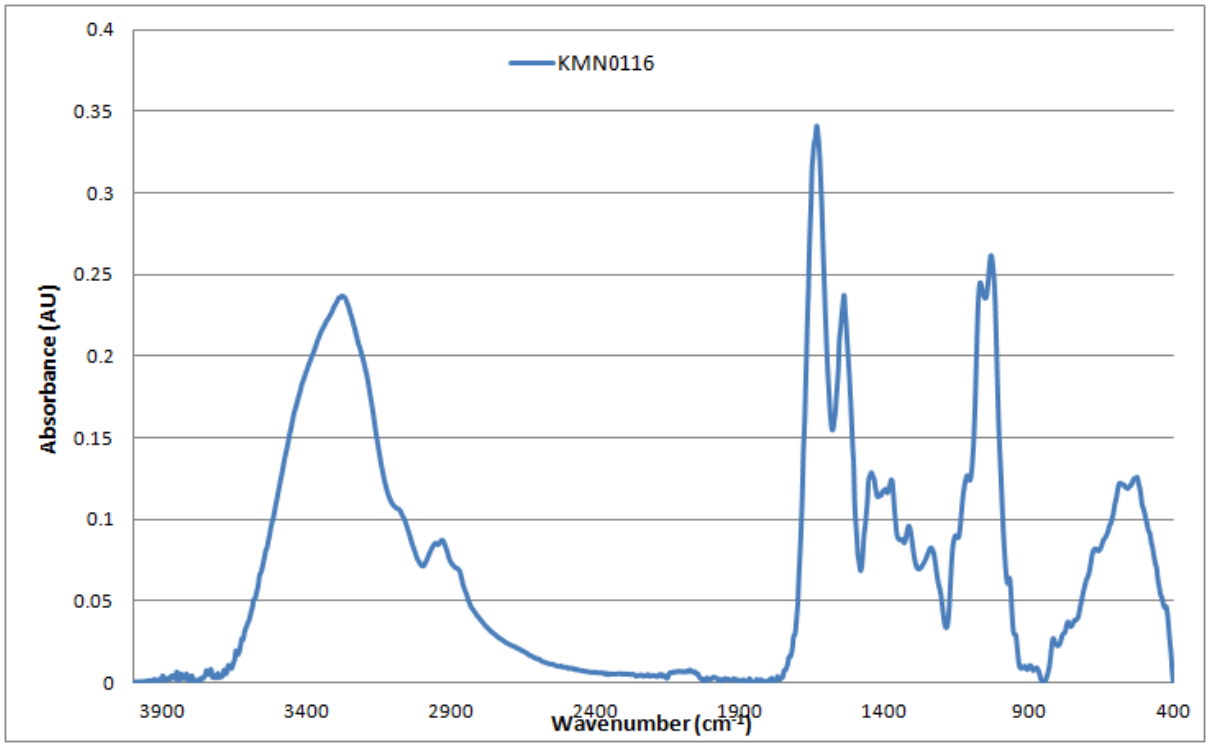


**Figure A2-76: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0116 from 4000-400cm$^{-1}$**

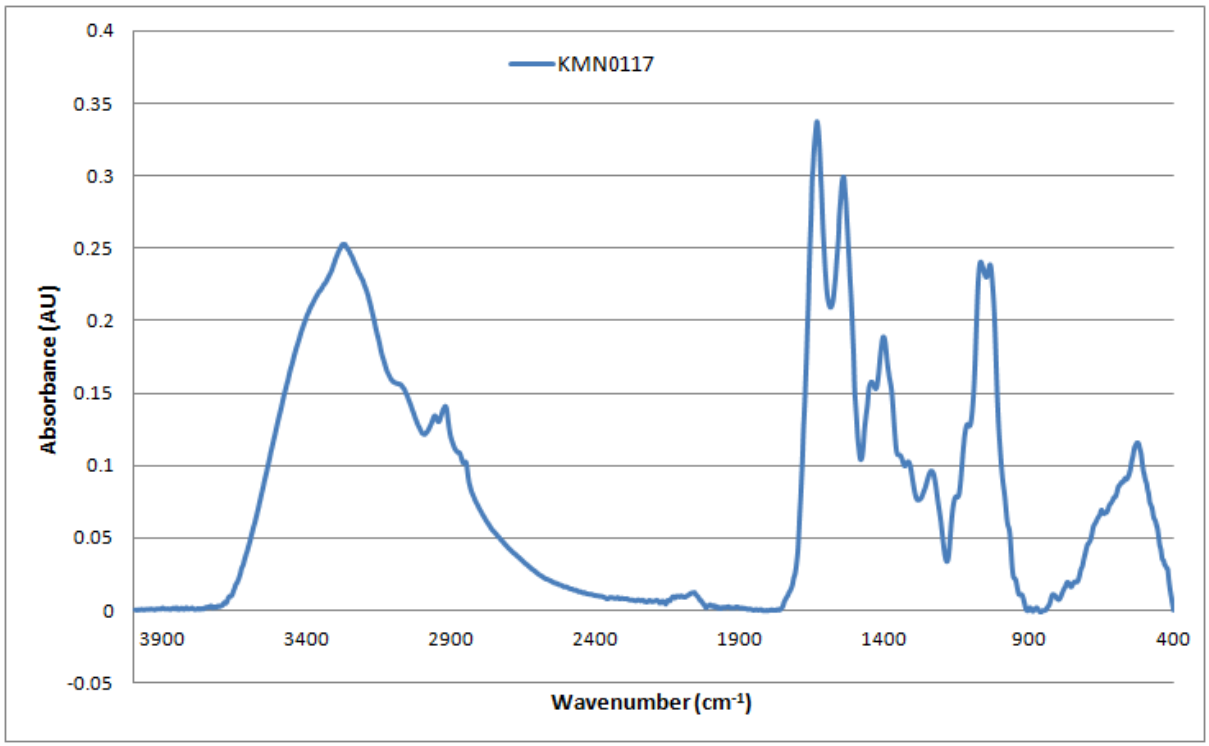


**Figure A2-77: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0117 from 4000-400cm$^{-1}$**

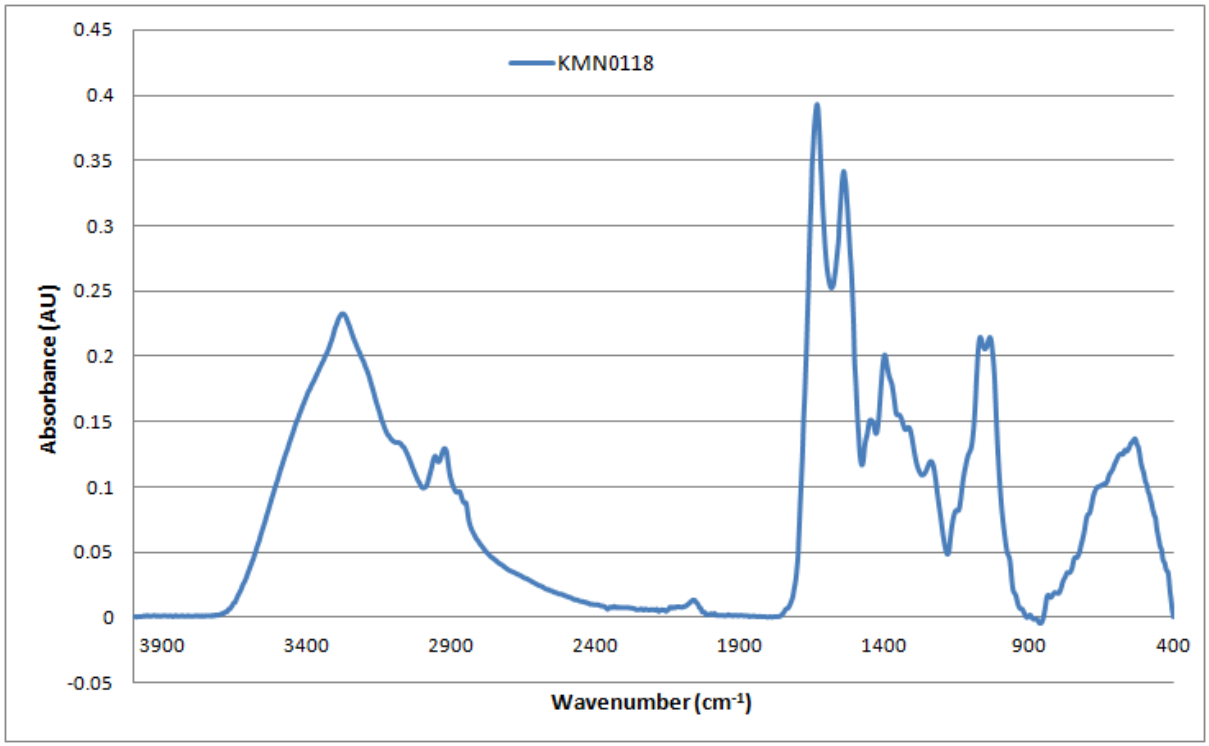


**Figure A2-78: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0118 from 4000-400cm$^{-1}$**

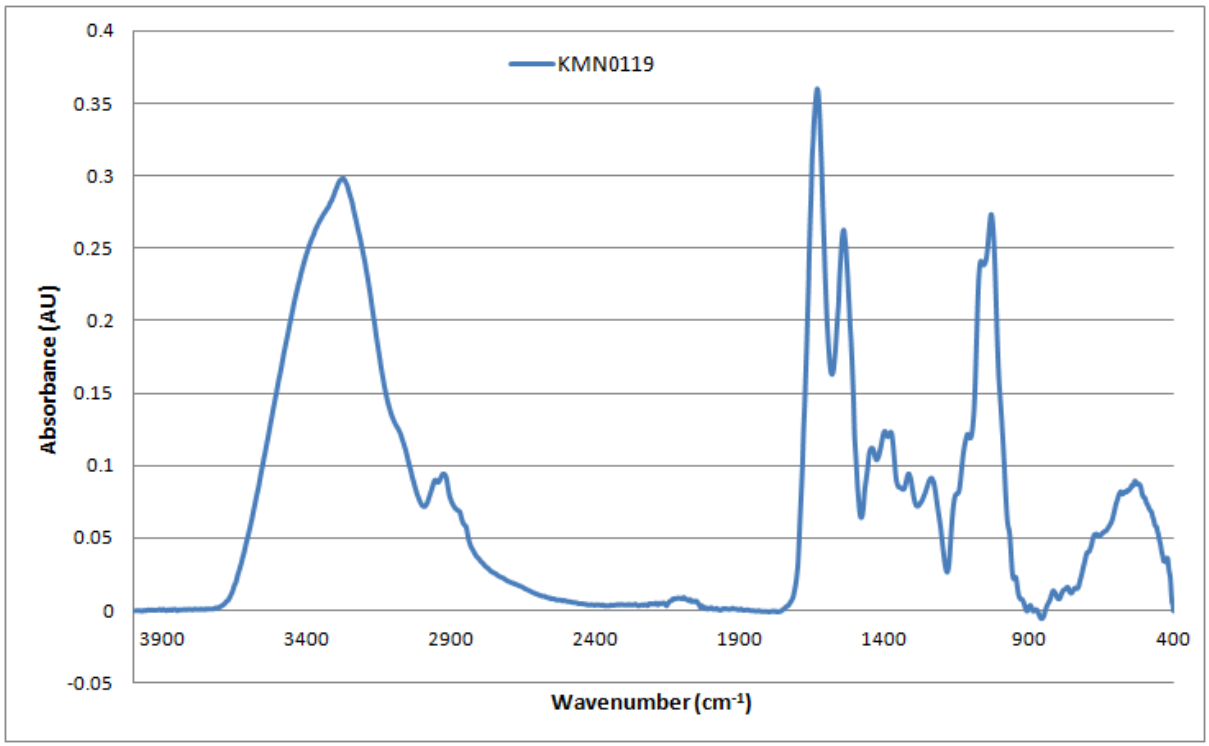


**Figure A2-79: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient KMN0119 from 4000-400cm$^{-1}$**

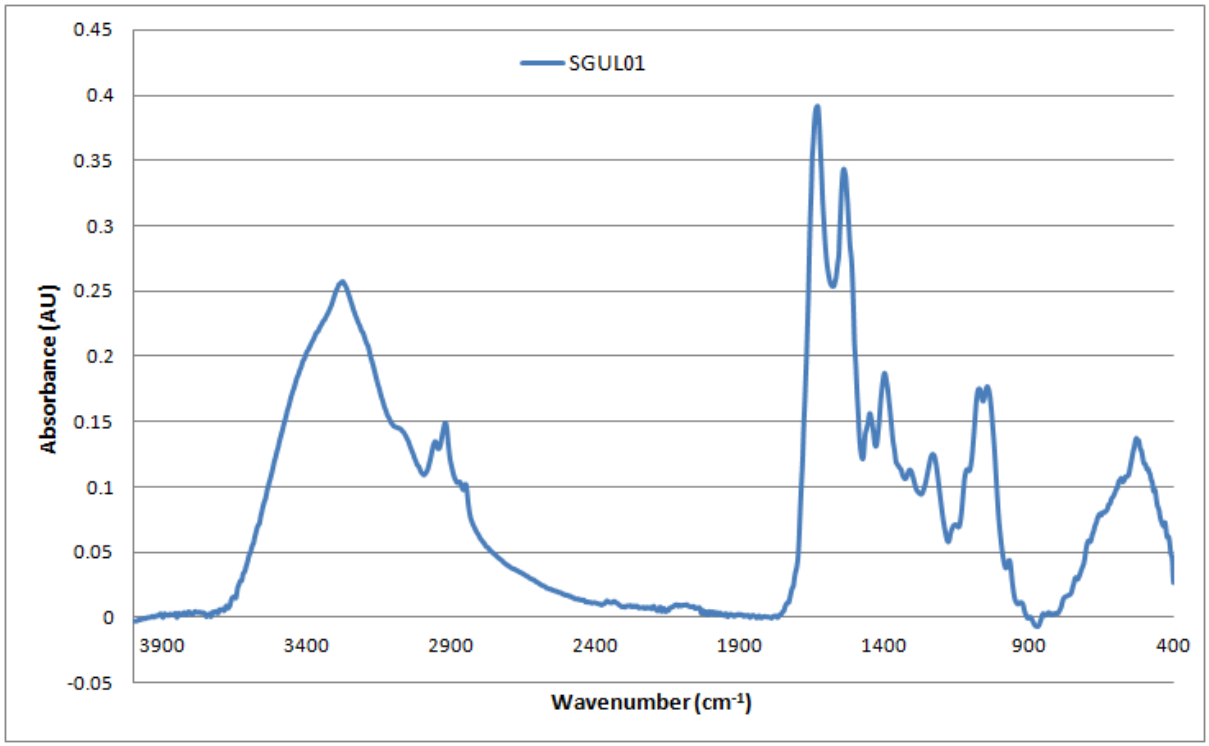


**Figure A2-80: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL01 from 4000-400cm$^{-1}$**

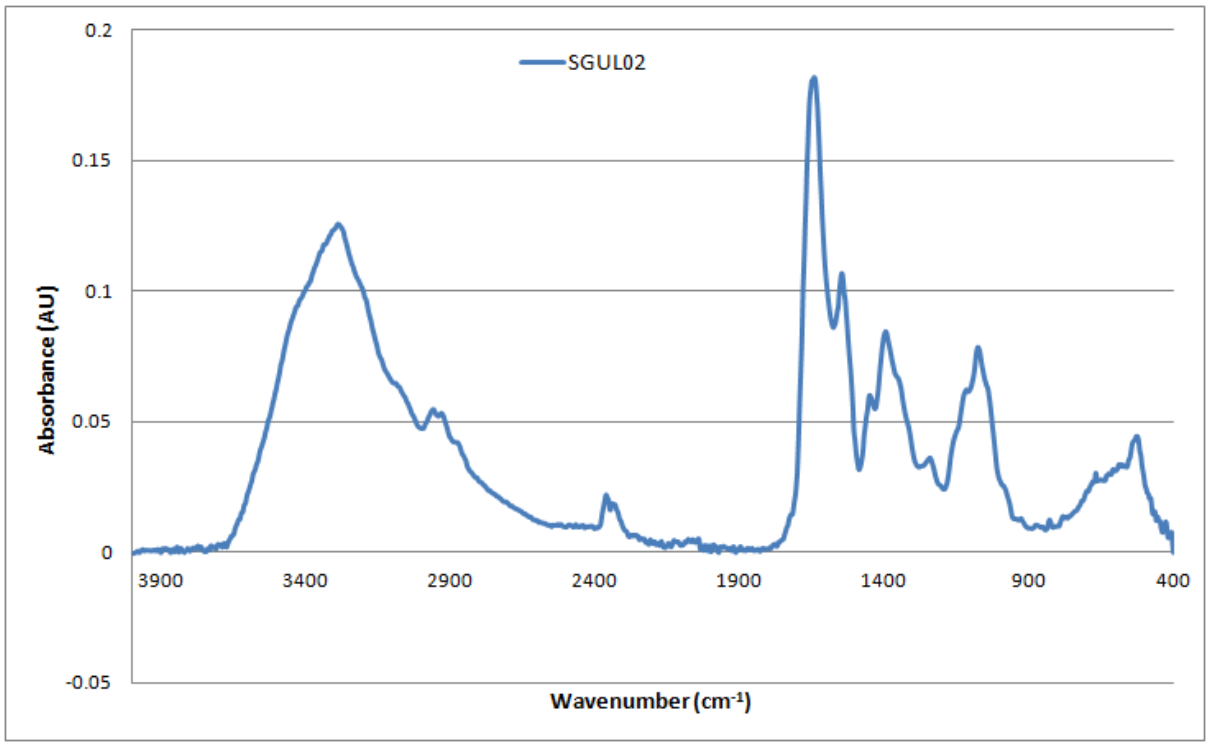


**Figure A2-81: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL02 from 4000-400cm$^{-1}$**

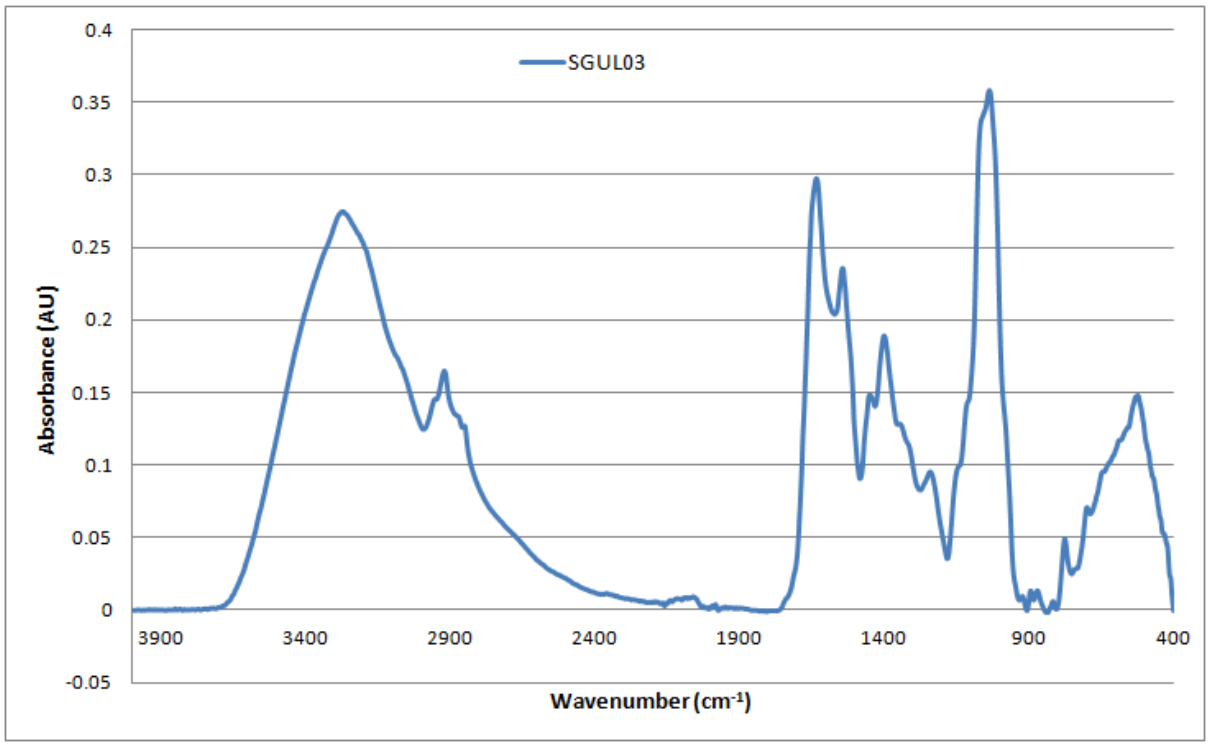


**Figure A2-82: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL03 from 4000-400cm$^{-1}$**

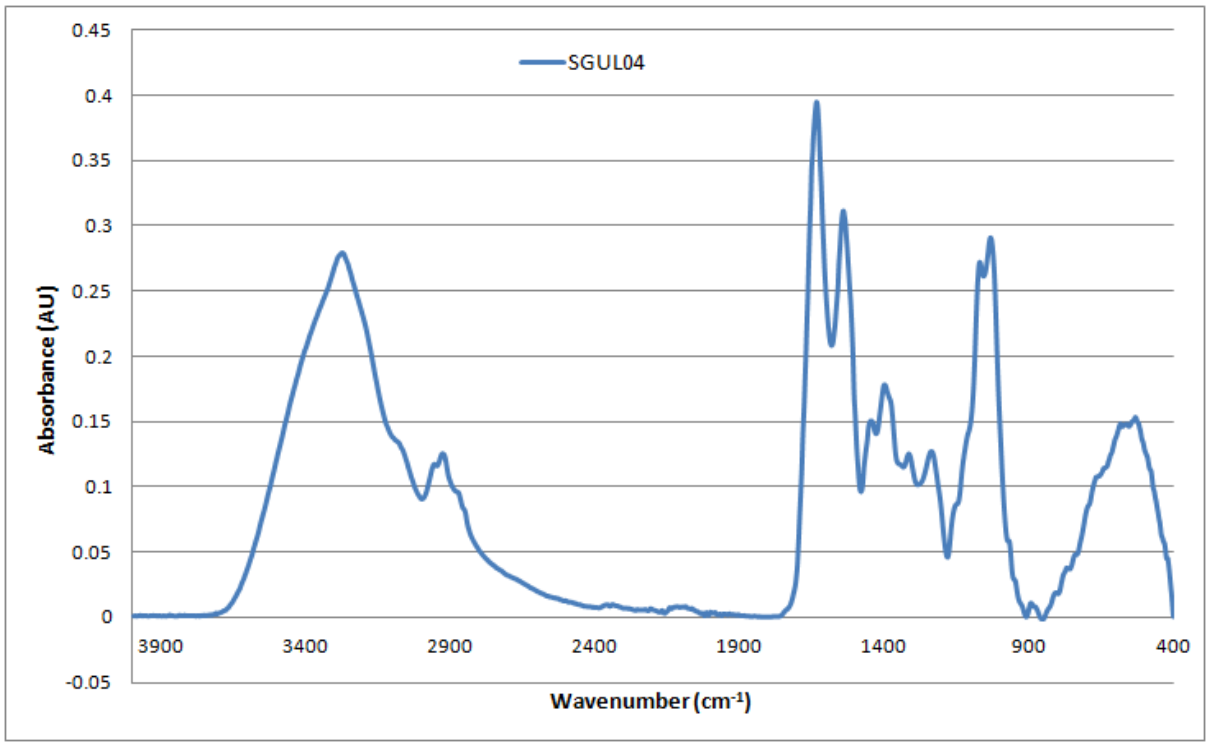


**Figure A2-83: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL04 from 4000-400cm$^{-1}$**

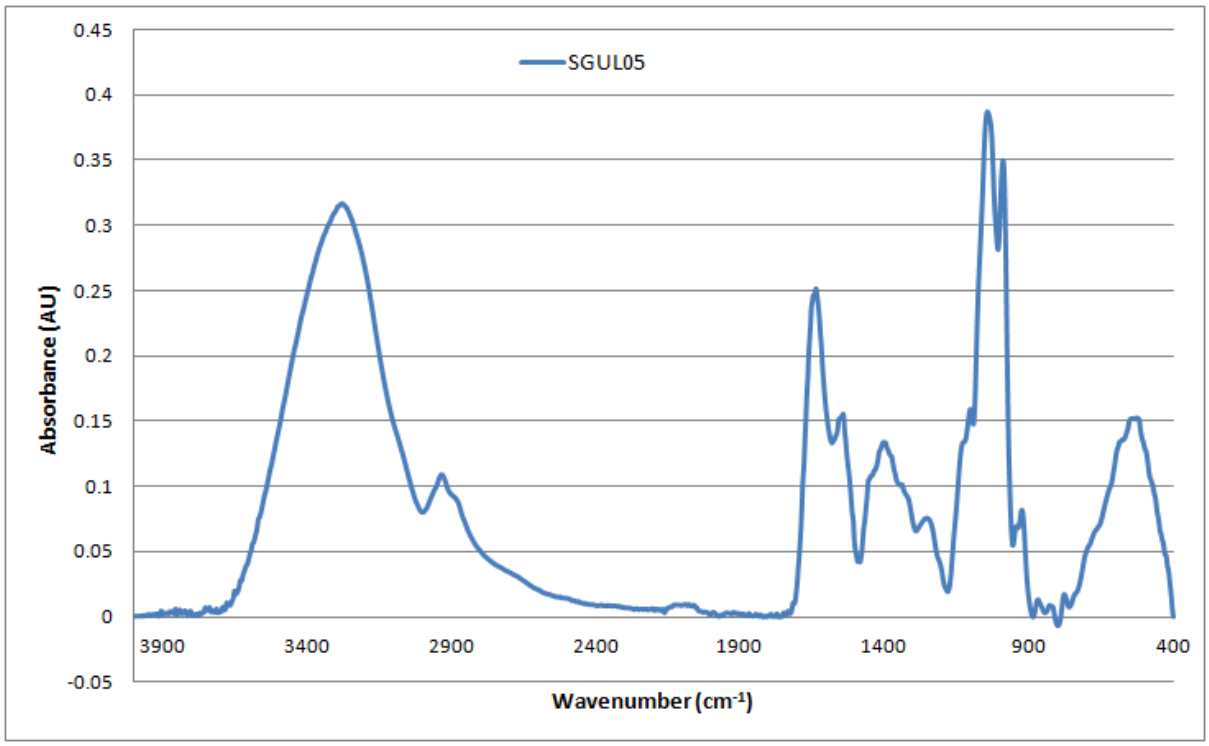


**Figure A2-84: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL05 from 4000-400cm$^{-1}$**

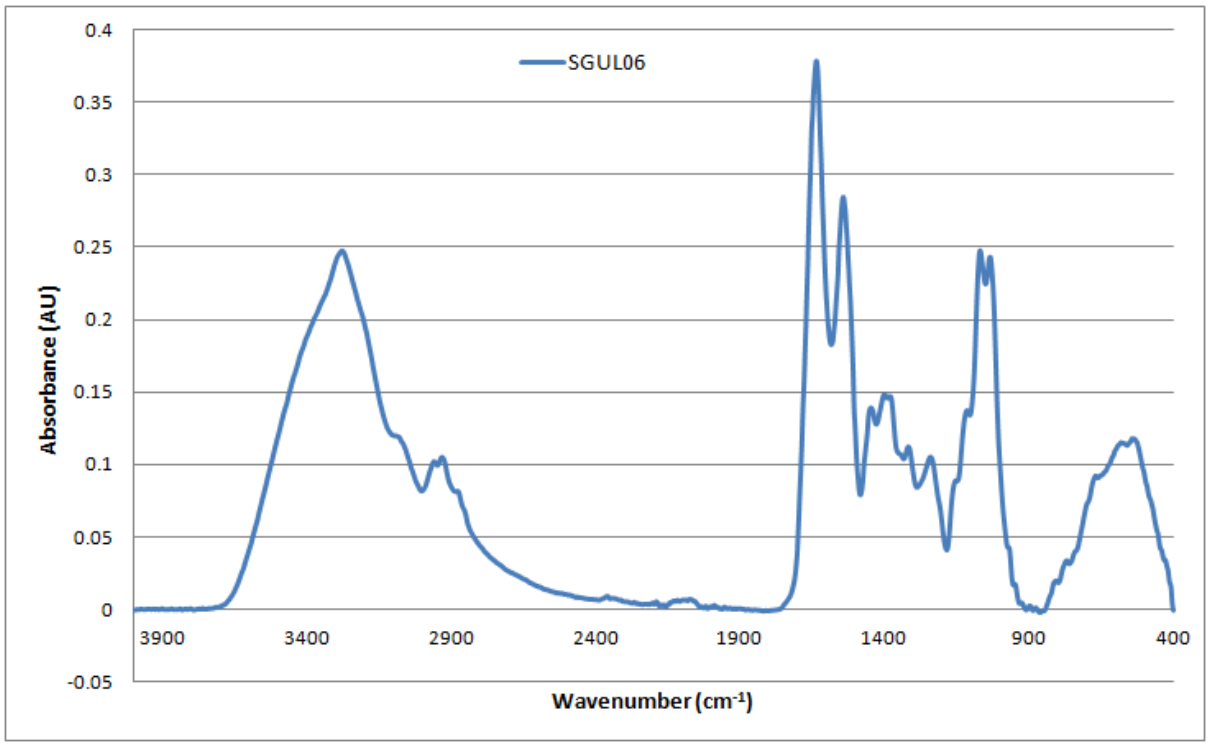


**Figure A2-85: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL06 from 4000-400cm$^{-1}$**

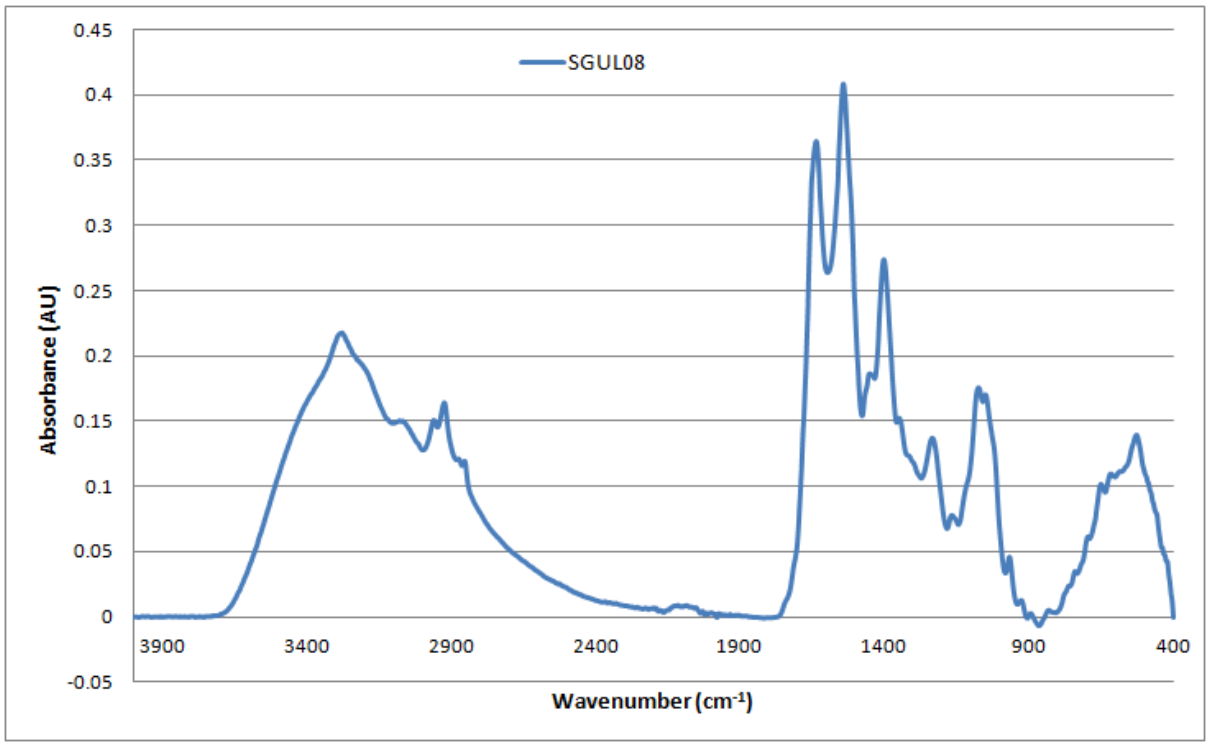


**Figure A2-86: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL08 from 4000-400cm$^{-1}$**

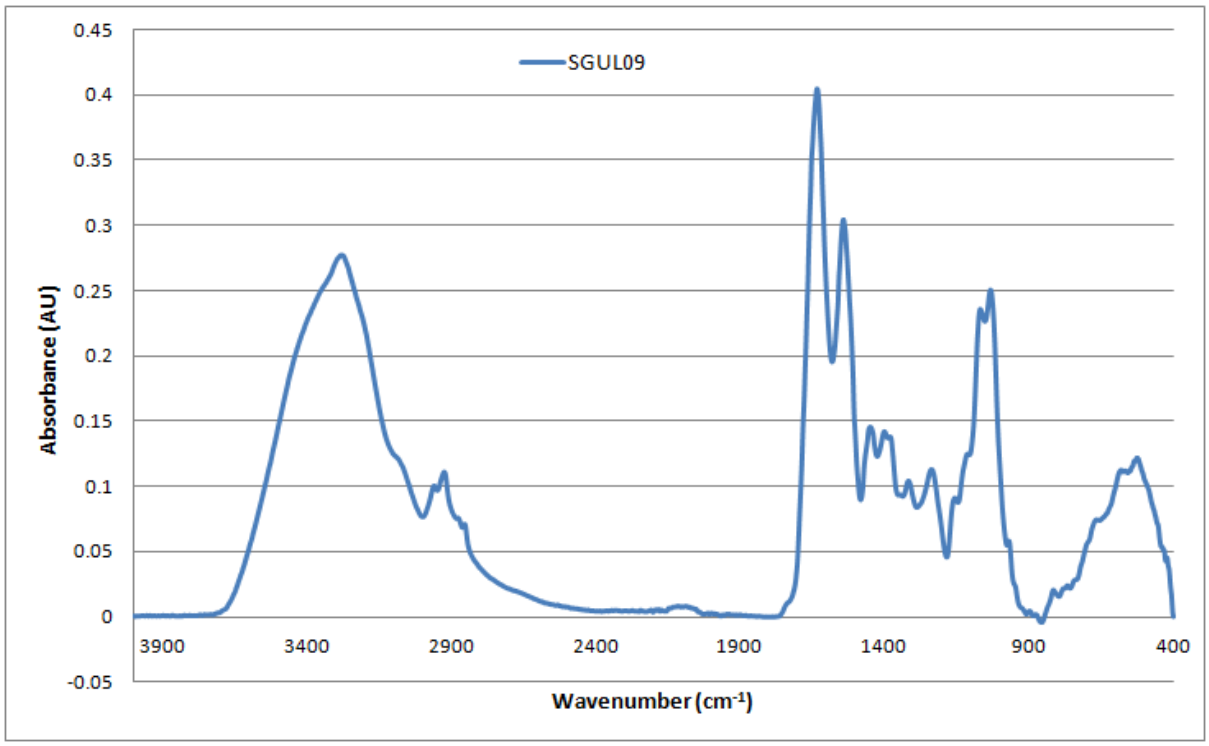


**Figure A2-87: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL09 from 4000-400cm$^{-1}$**

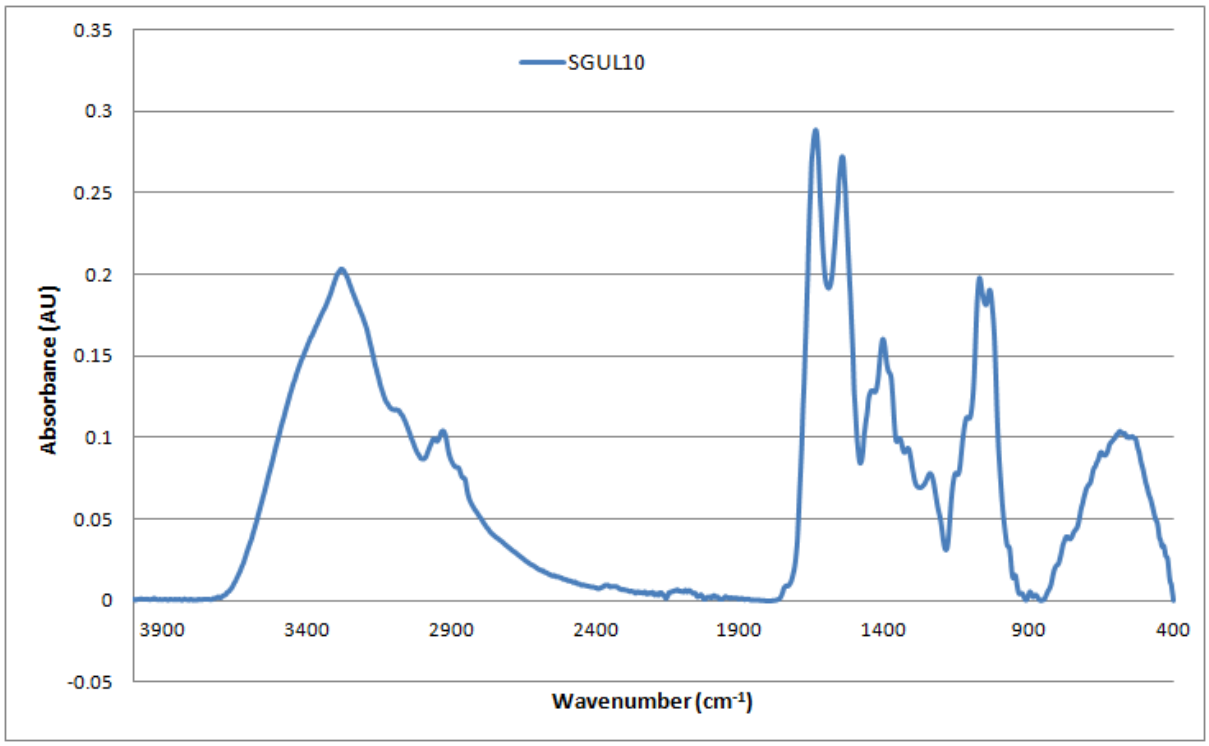


**Figure A2-88: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL10 from 4000-400cm$^{-1}$**

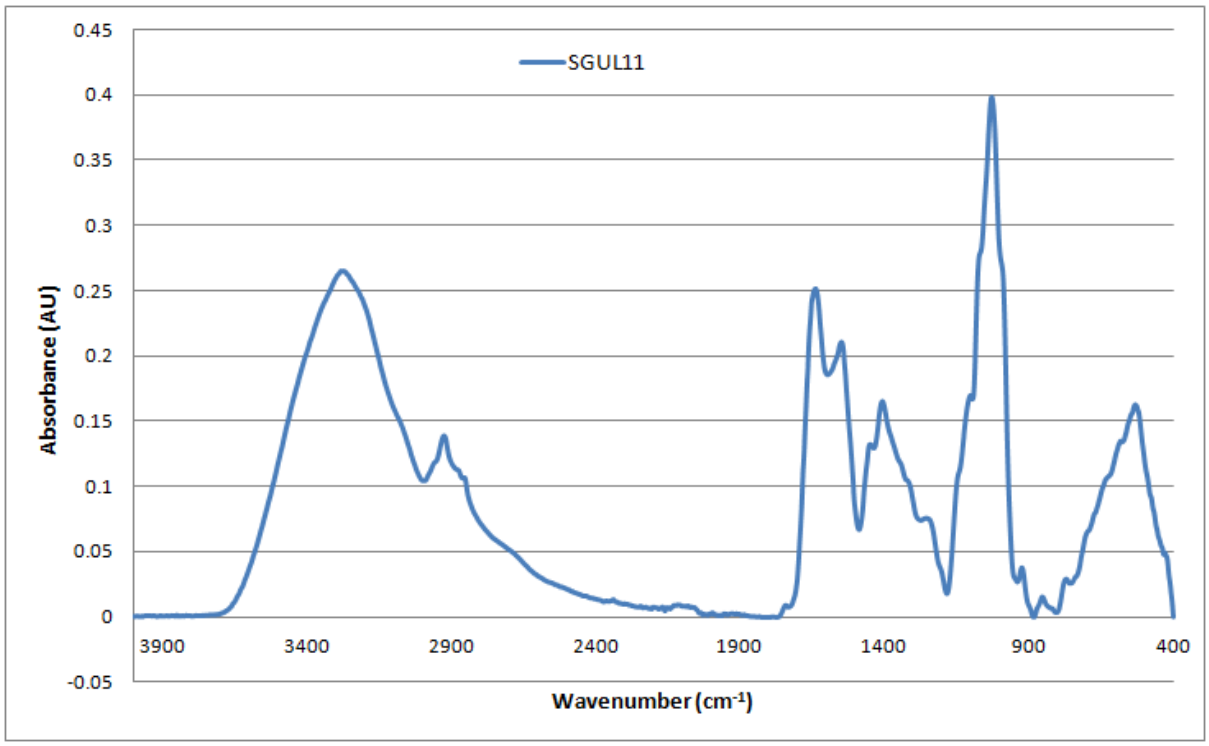


**Figure A2-89: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL11 from 4000-400cm$^{-1}$**

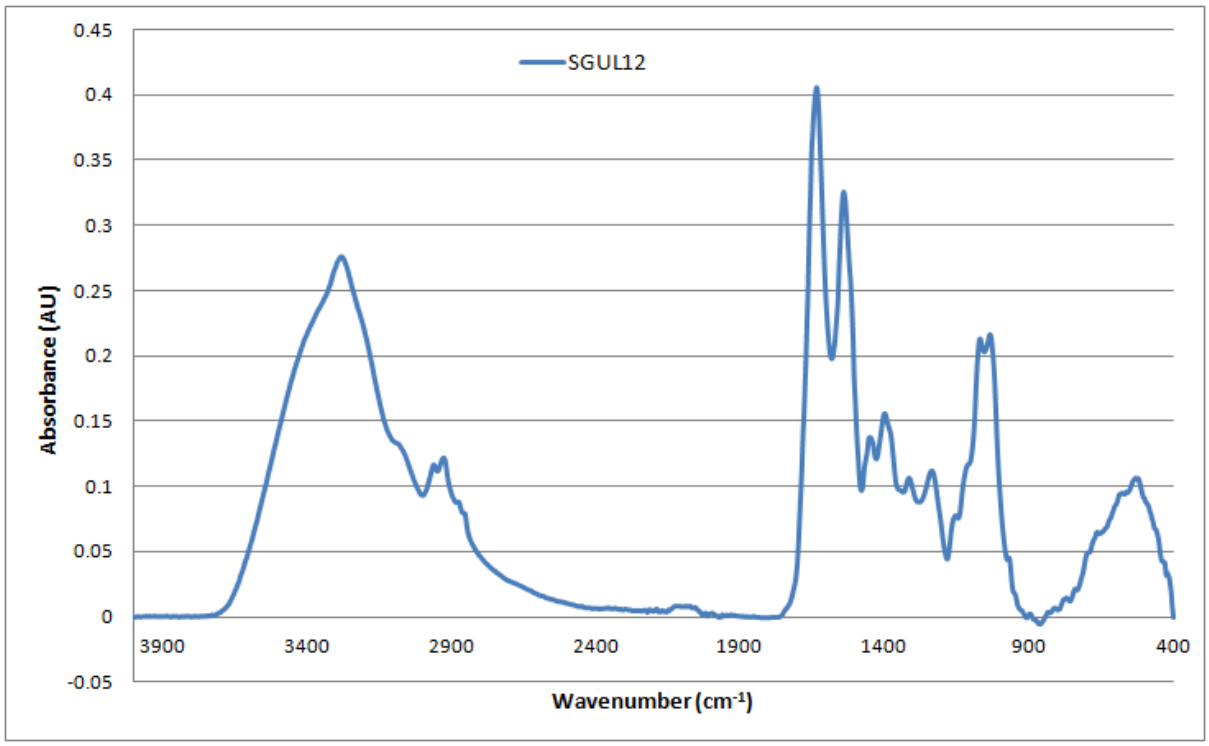


**Figure A2-90: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL12 from 4000-400cm$^{-1}$**

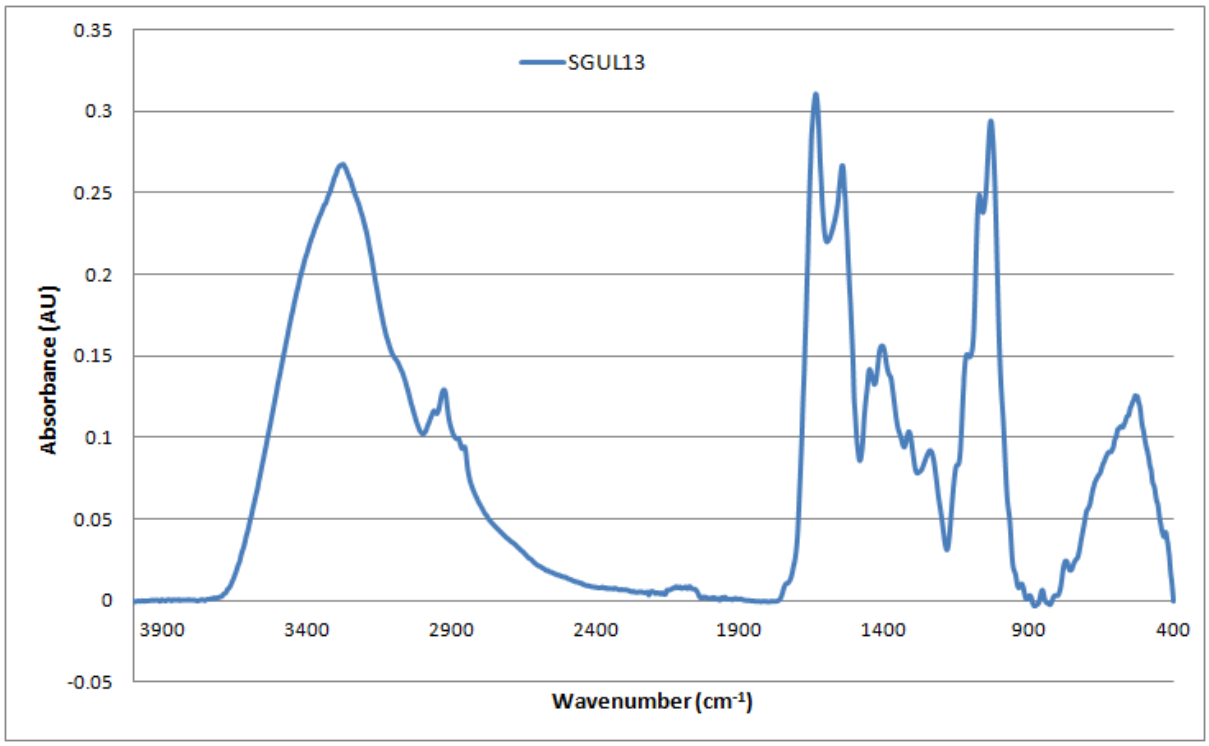


**Figure A2-91: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL13 from 4000-400cm$^{-1}$**

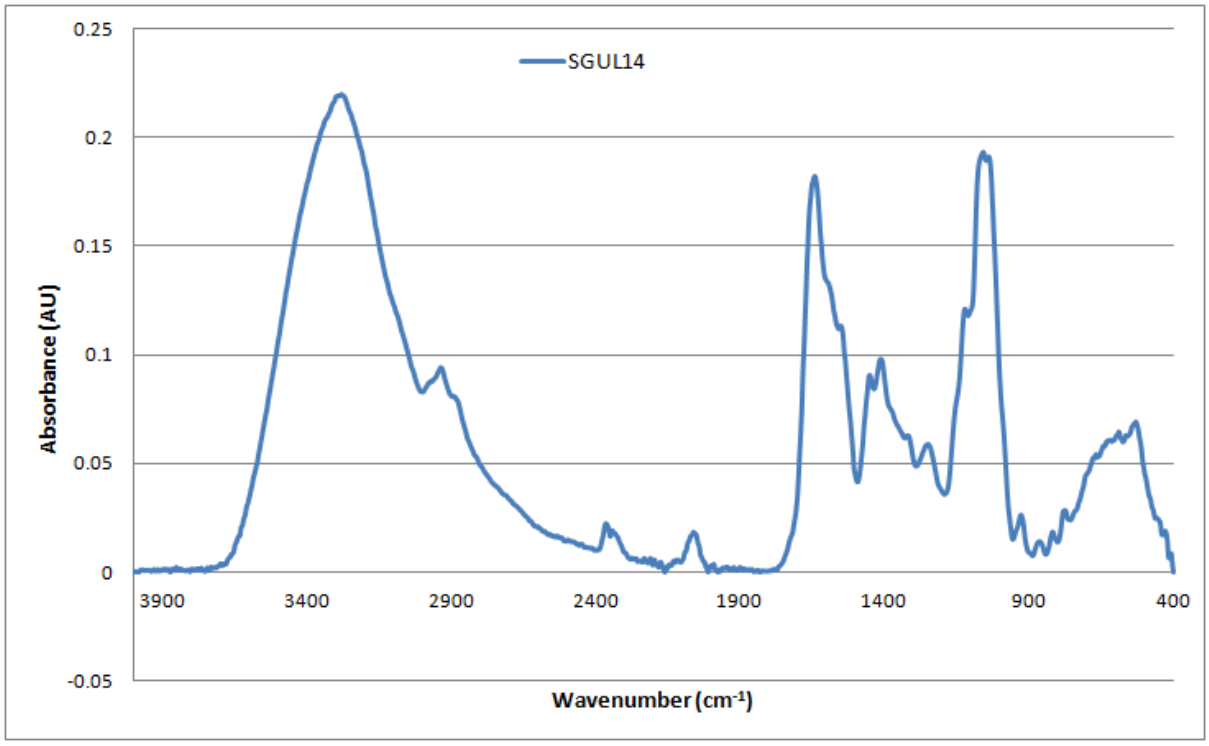


**Figure A2-92: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient SGUL14 from 4000-400cm$^{-1}$**

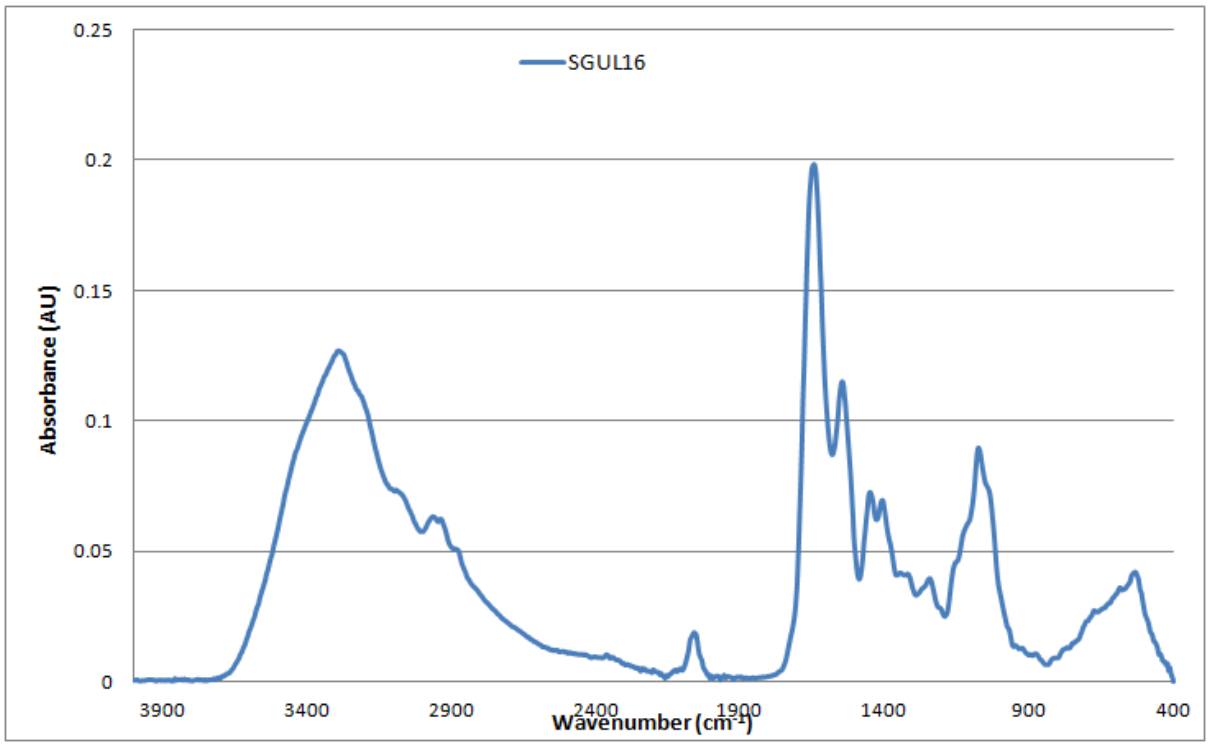


**Figure A2-93: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient SGUL16 from 4000-400cm$^{-1}$**

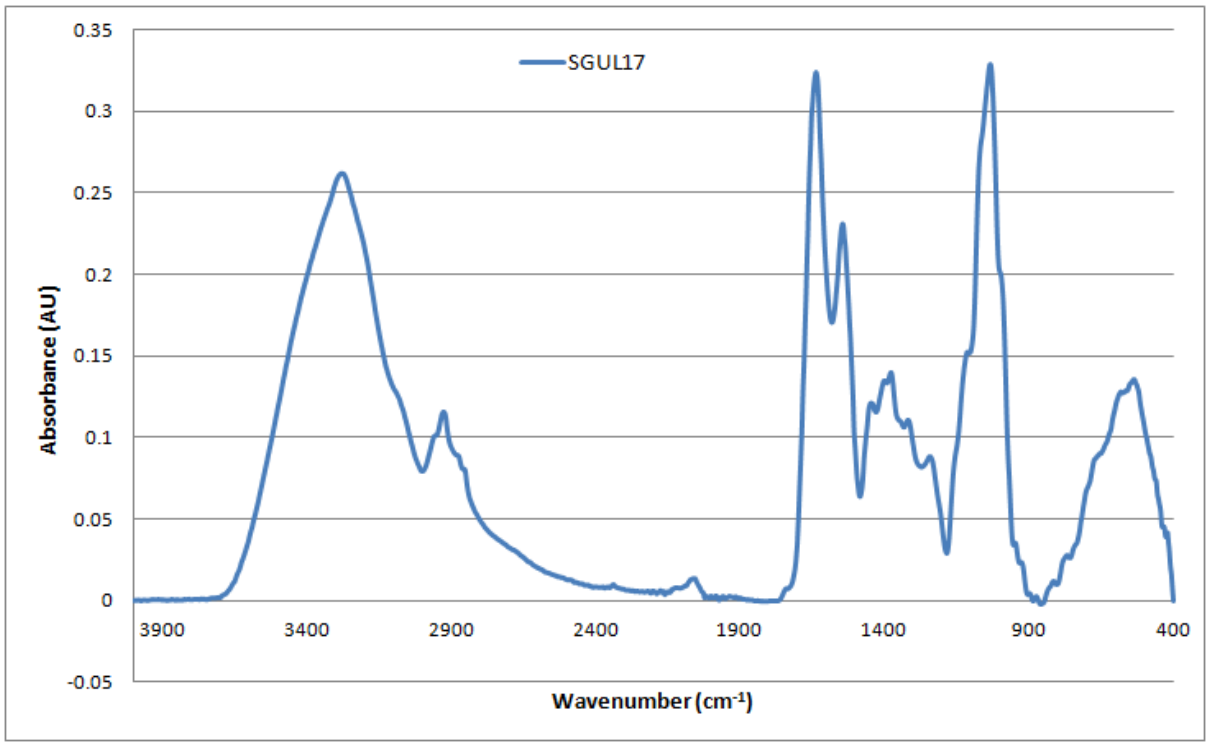


**Figure A2-94: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient SGUL17 from 4000-400cm$^{-1}$**

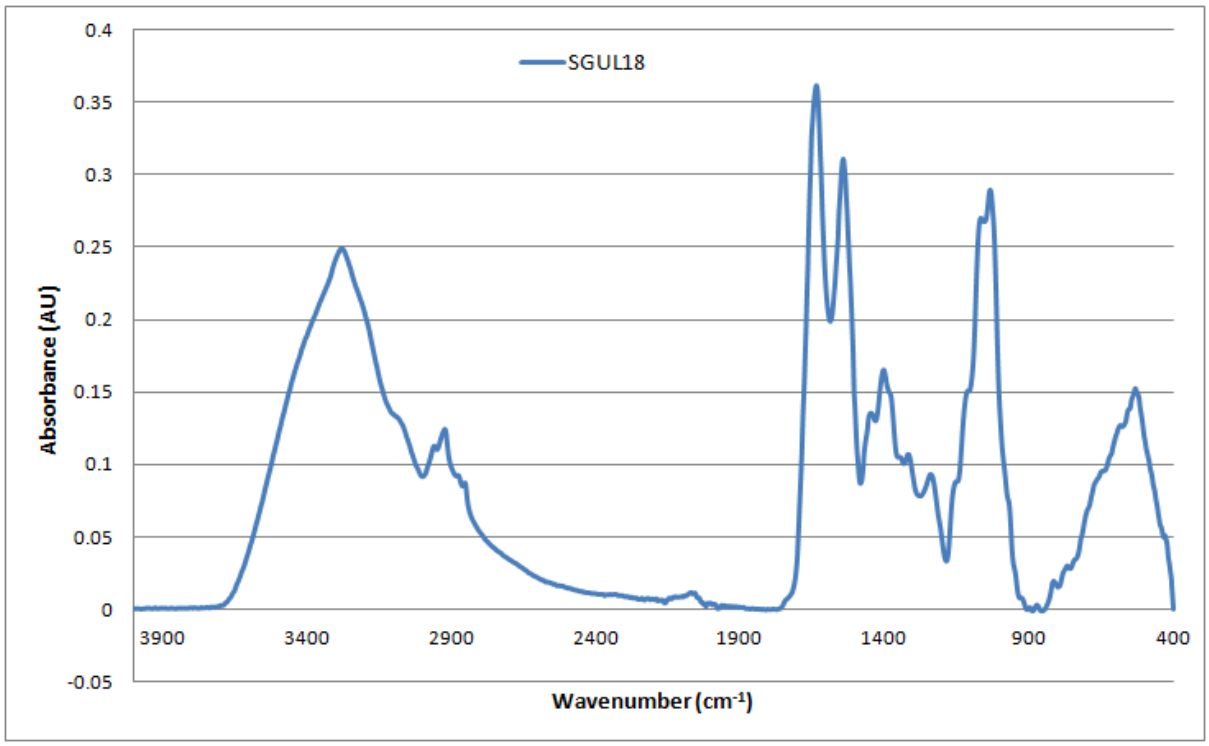


**Figure A2-95: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient SGUL18 from 4000-400cm$^{-1}$**

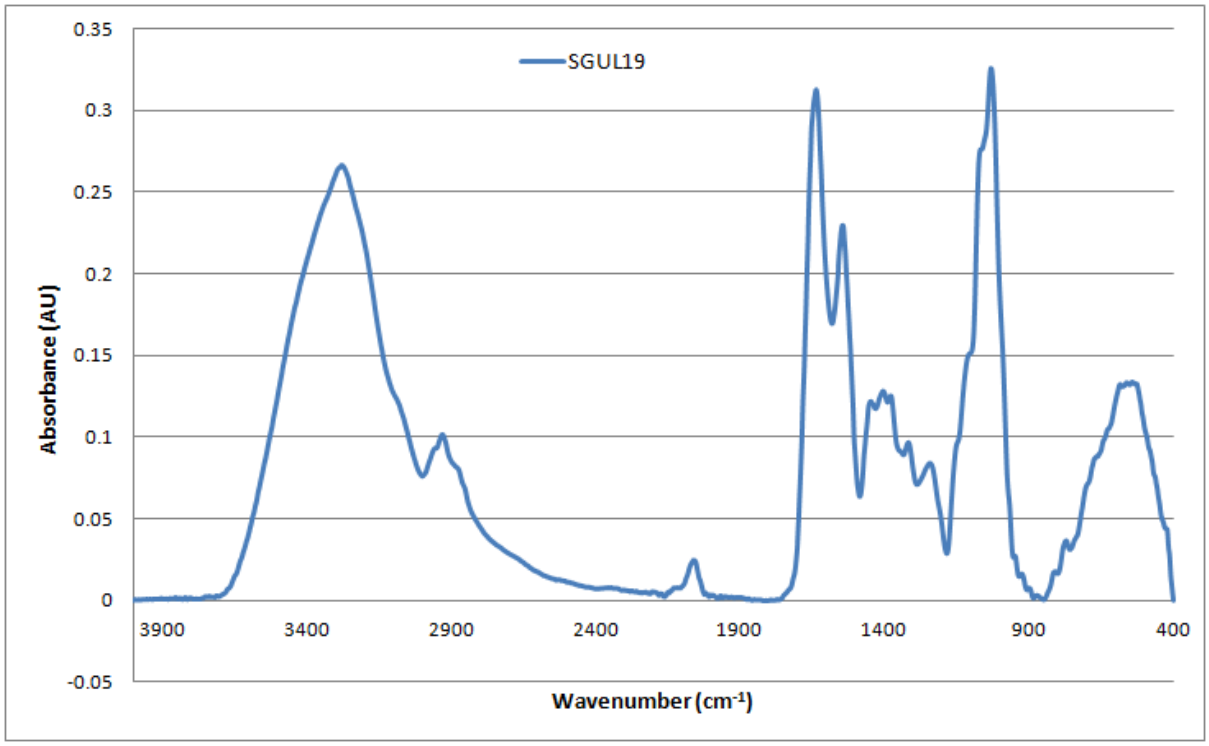


**Figure A2-96: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL19 from 4000-400cm$^{-1}$**

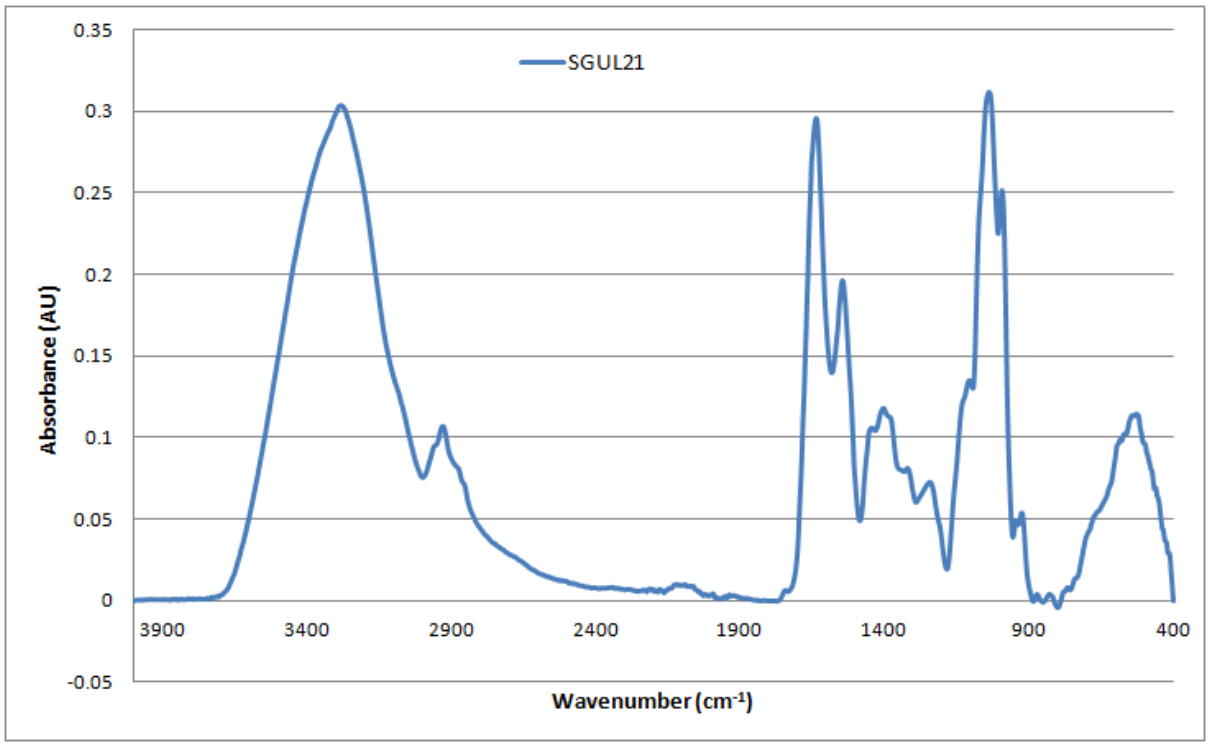


**Figure A2-97: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL21 from 4000-400cm$^{-1}$**

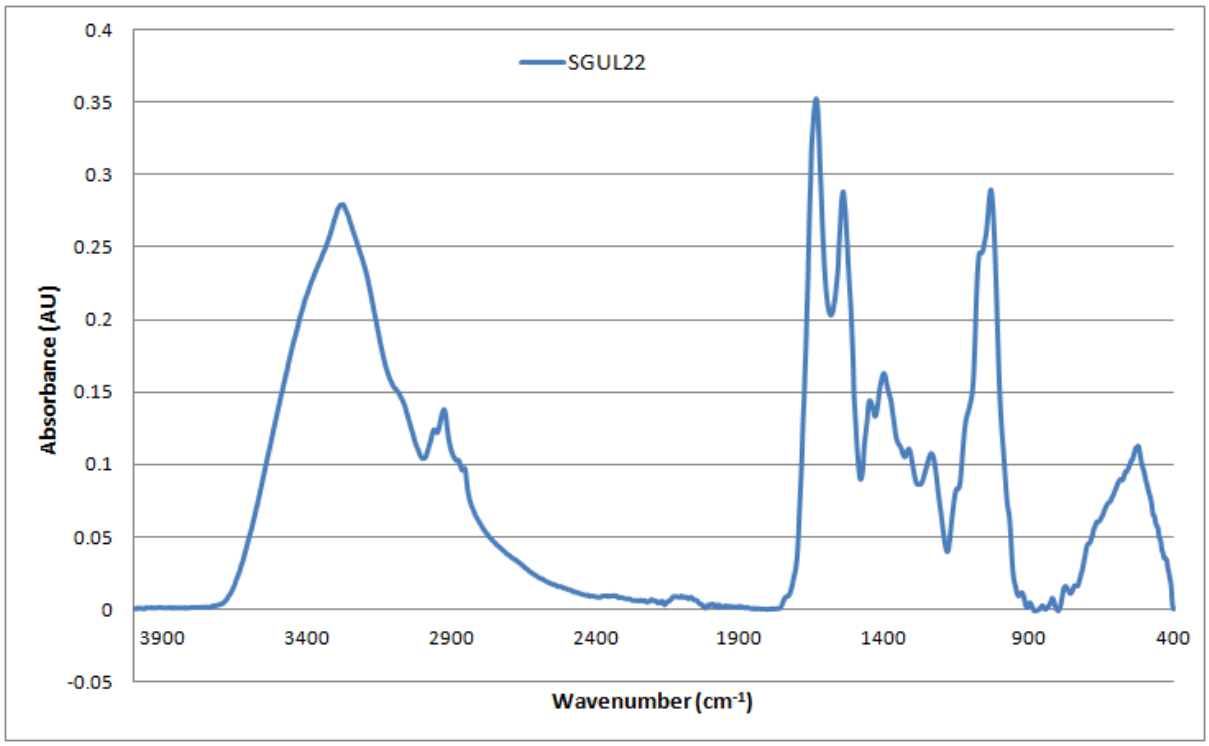


**Figure A2-98: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL22 from 4000-400cm$^{-1}$**

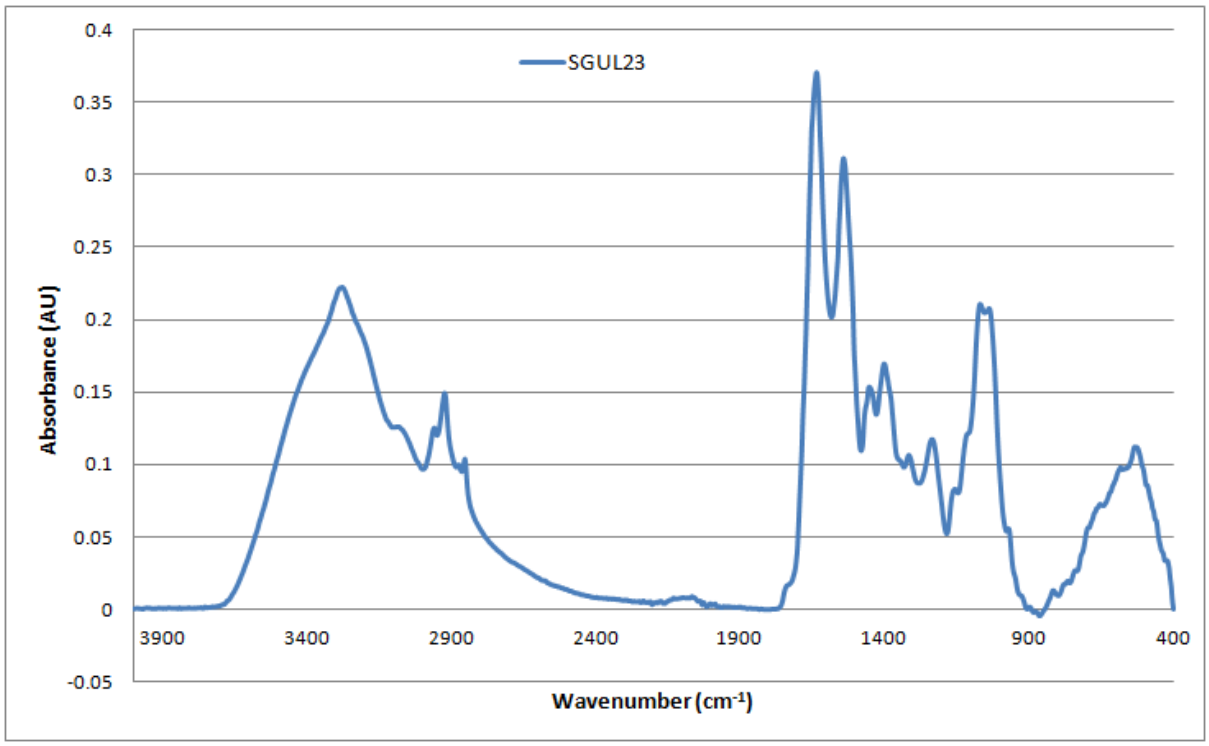


**Figure A2-99: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient SGUL23 from 4000-400cm$^{-1}$**

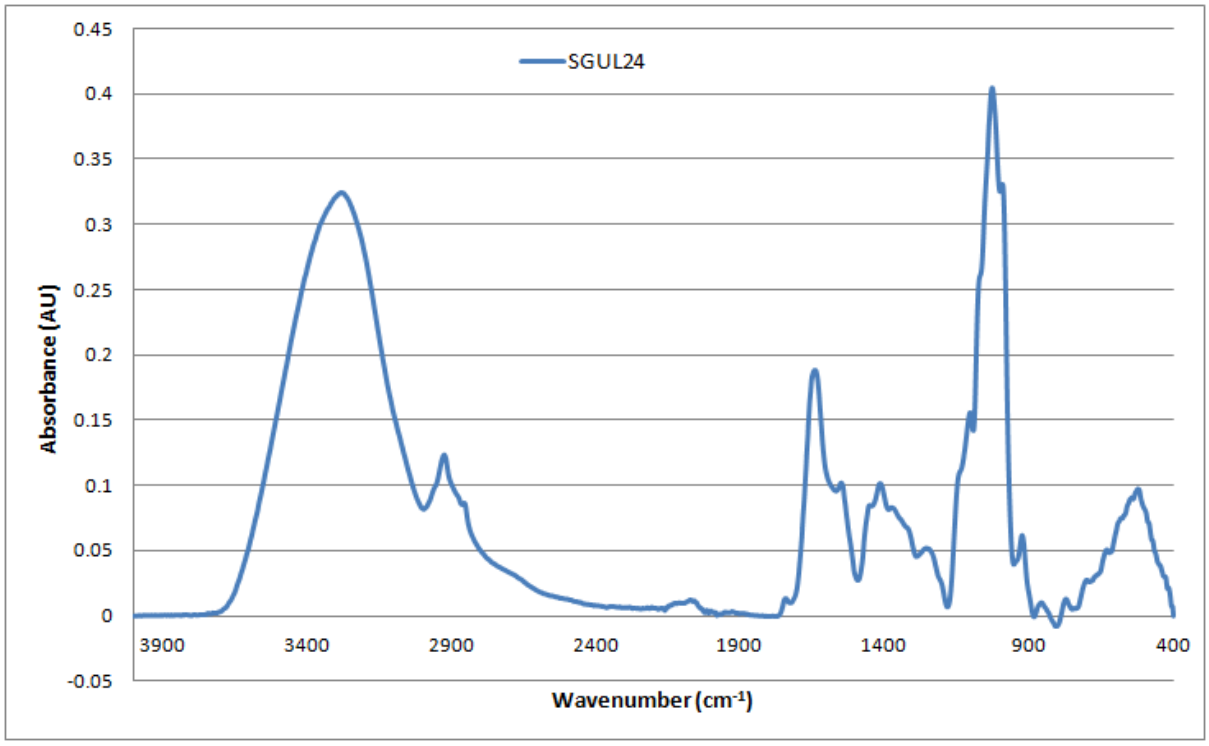


**Figure A2-100: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL24 from 4000-400cm$^{-1}$**

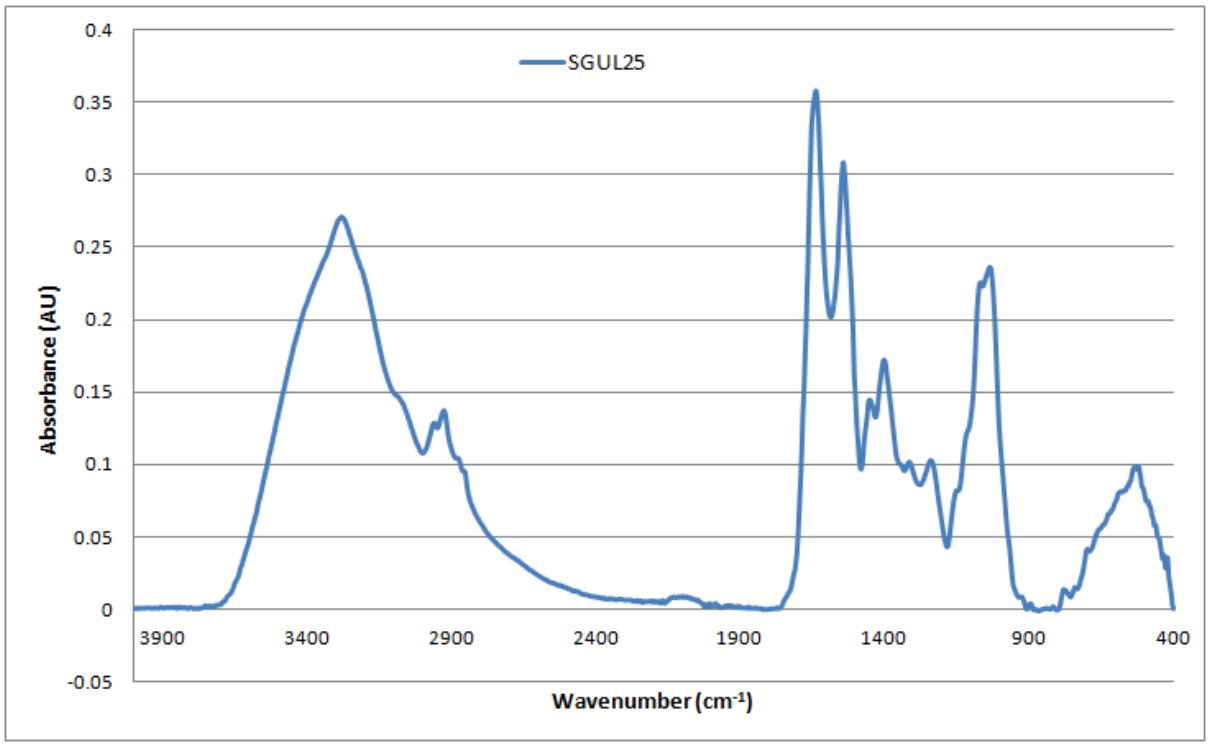


**Figure A2-101: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL24 from 4000-400cm$^{-1}$**

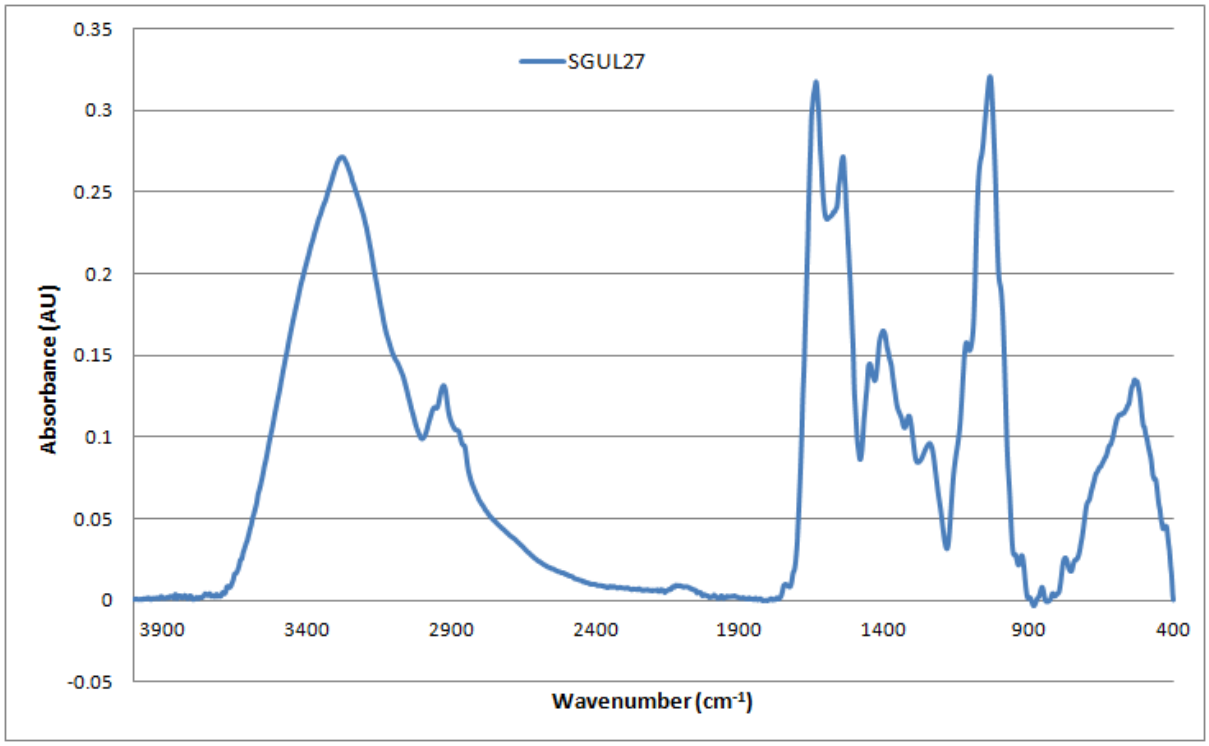


**Figure A2-102: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient SGUL27 from 4000-400cm$^{-1}$**

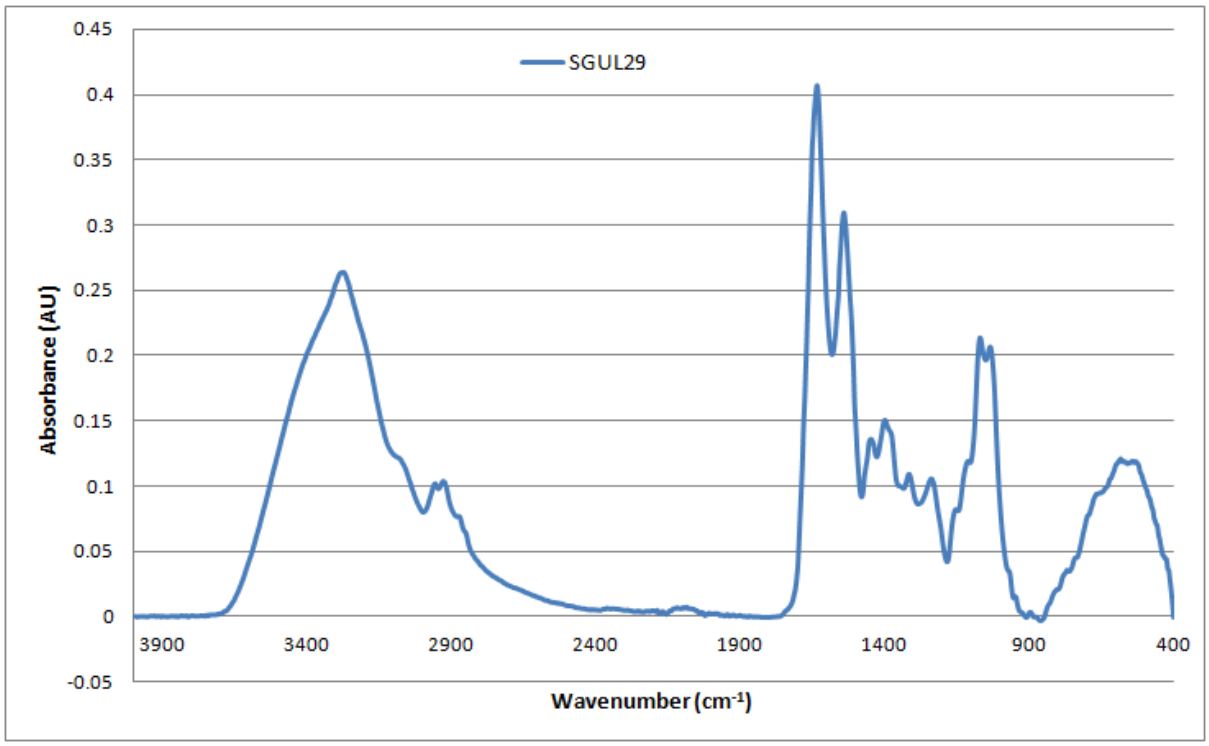


**Figure A2-103: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL29 from 4000-400cm$^{-1}$**

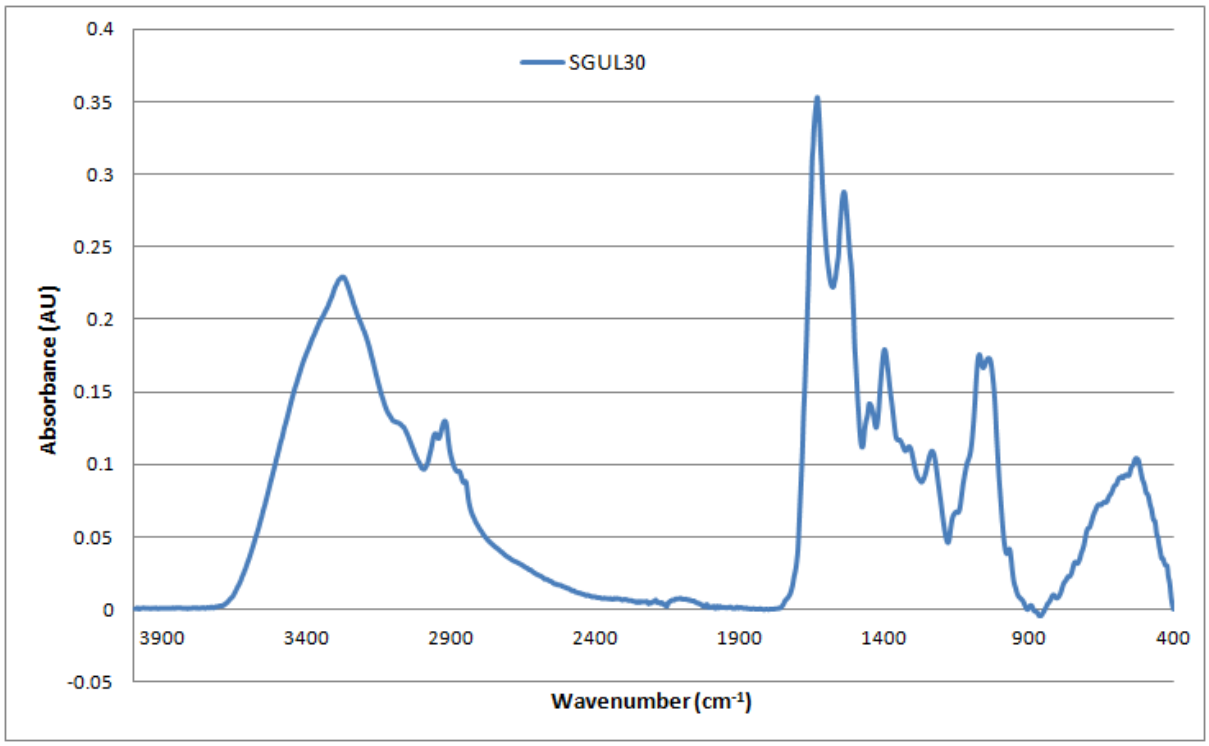


**Figure A2-104: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL30 from 4000-400cm$^{-1}$**

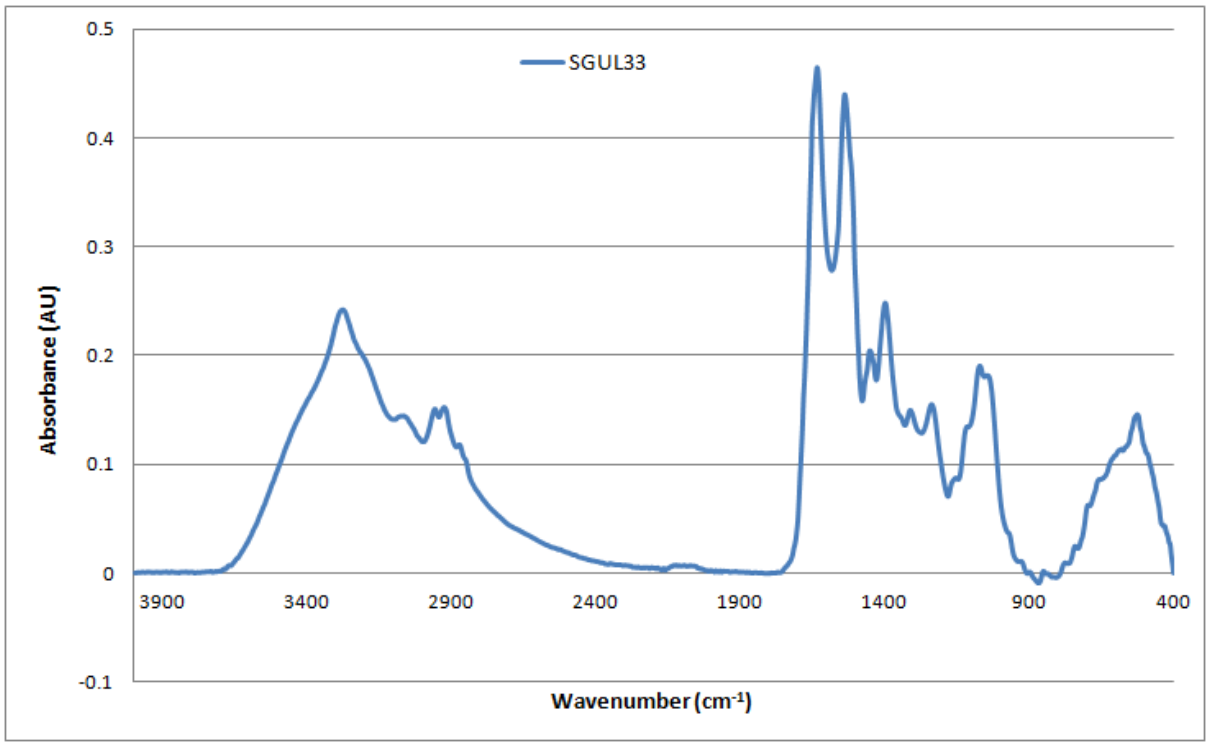


**Figure A2-105: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL33 from 4000-400cm$^{-1}$**

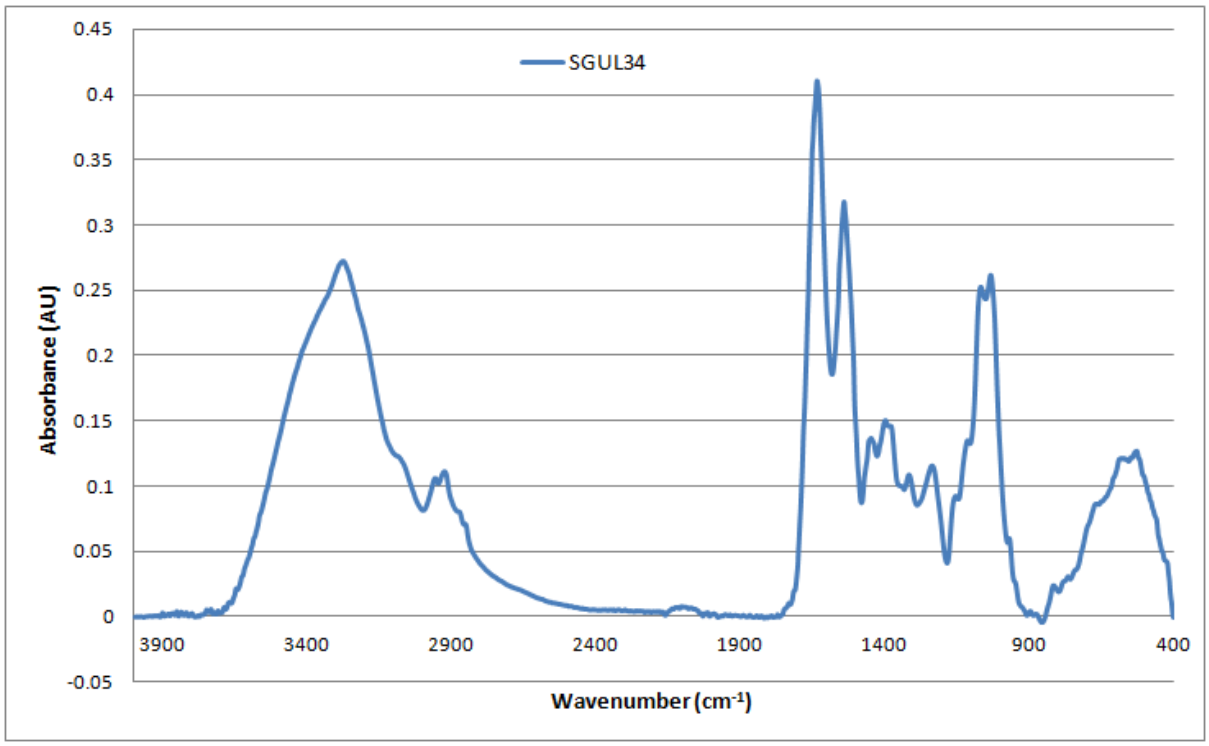


**Figure A2-106: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL34 from 4000-400cm$^{-1}$**

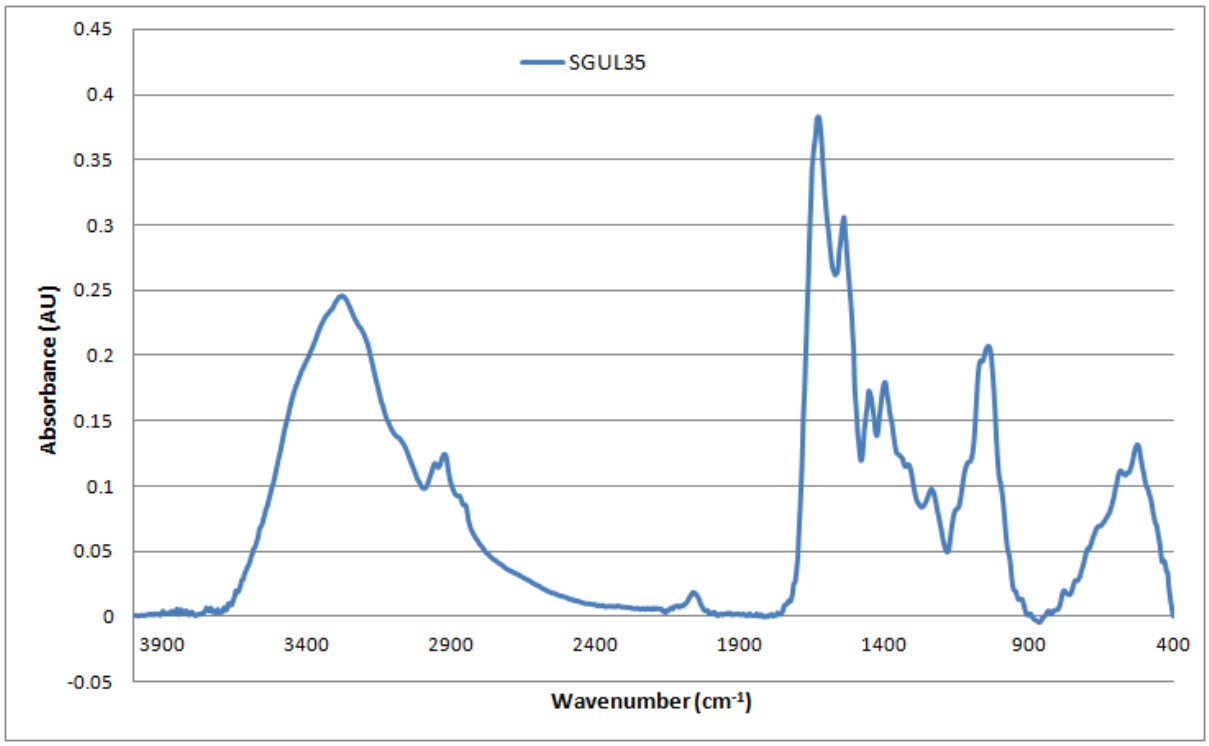


**Figure A2-107: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL35 from 4000-400cm$^{-1}$**

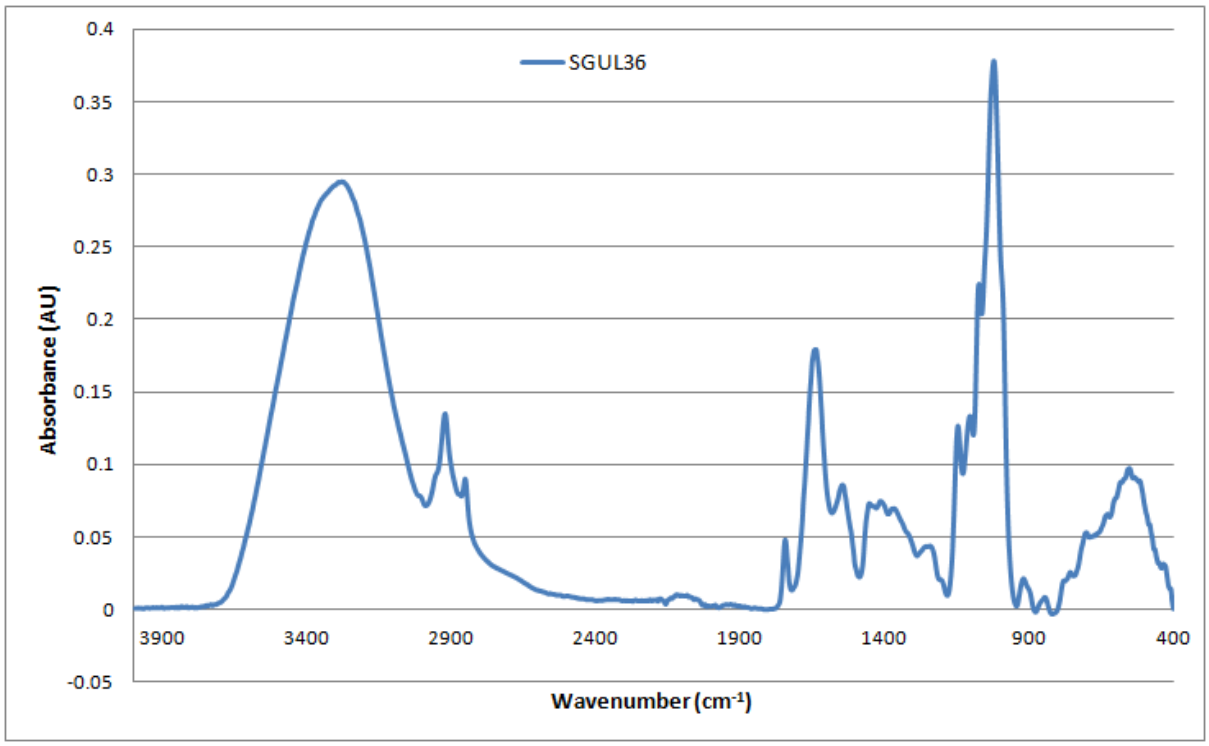


**Figure A2-108: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL36 from 4000-400cm$^{-1}$**

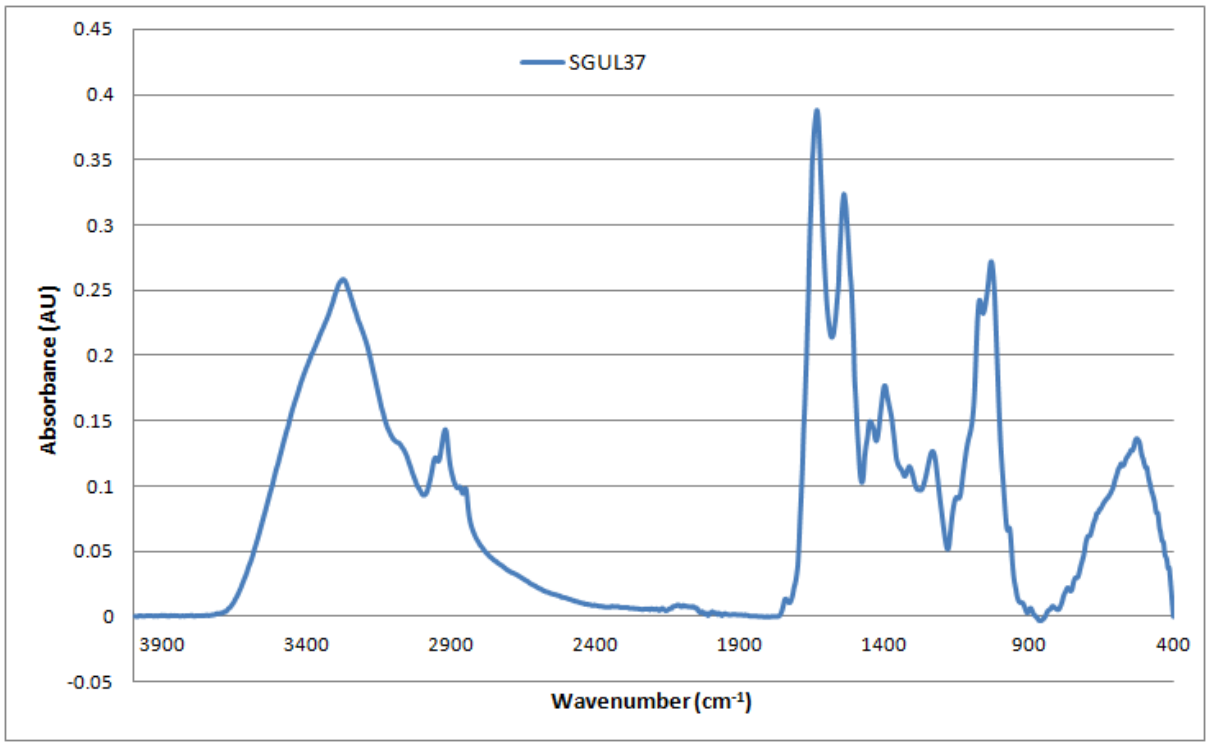


**Figure A2-109: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL37 from 4000-400cm$^{-1}$**

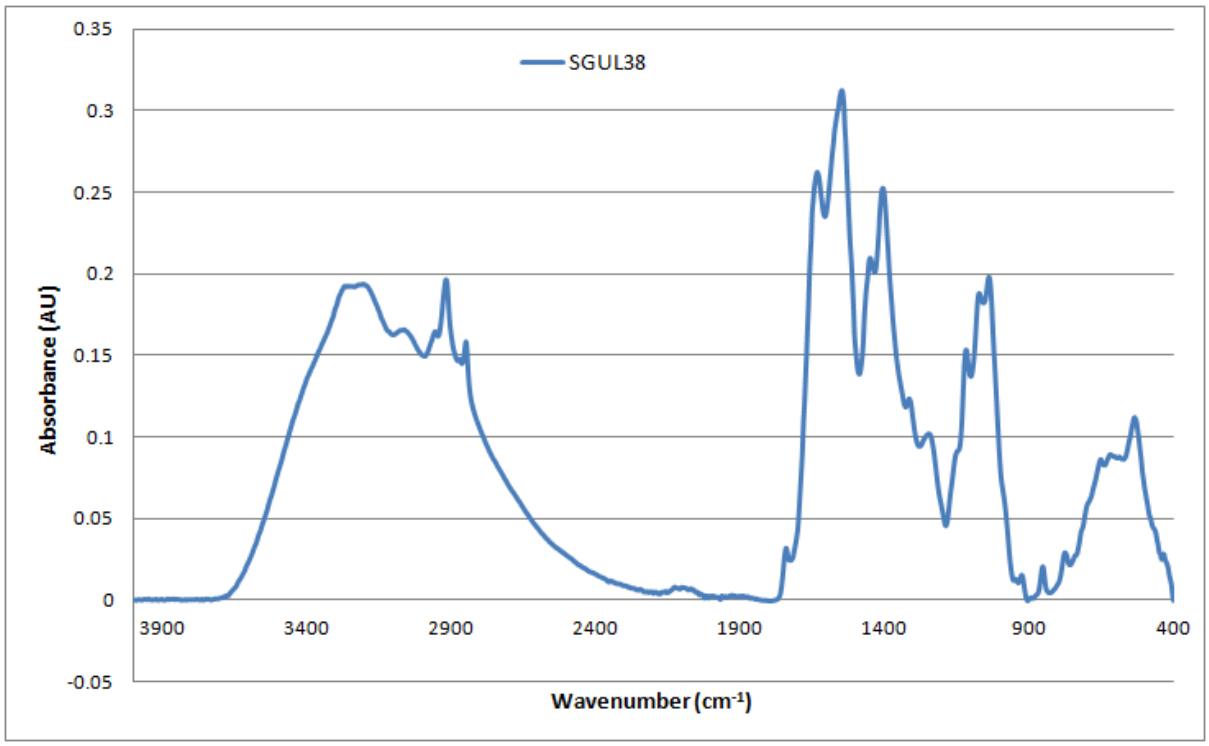


**Figure A2-110: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient SGUL38 from 4000-400cm$^{-1}$**

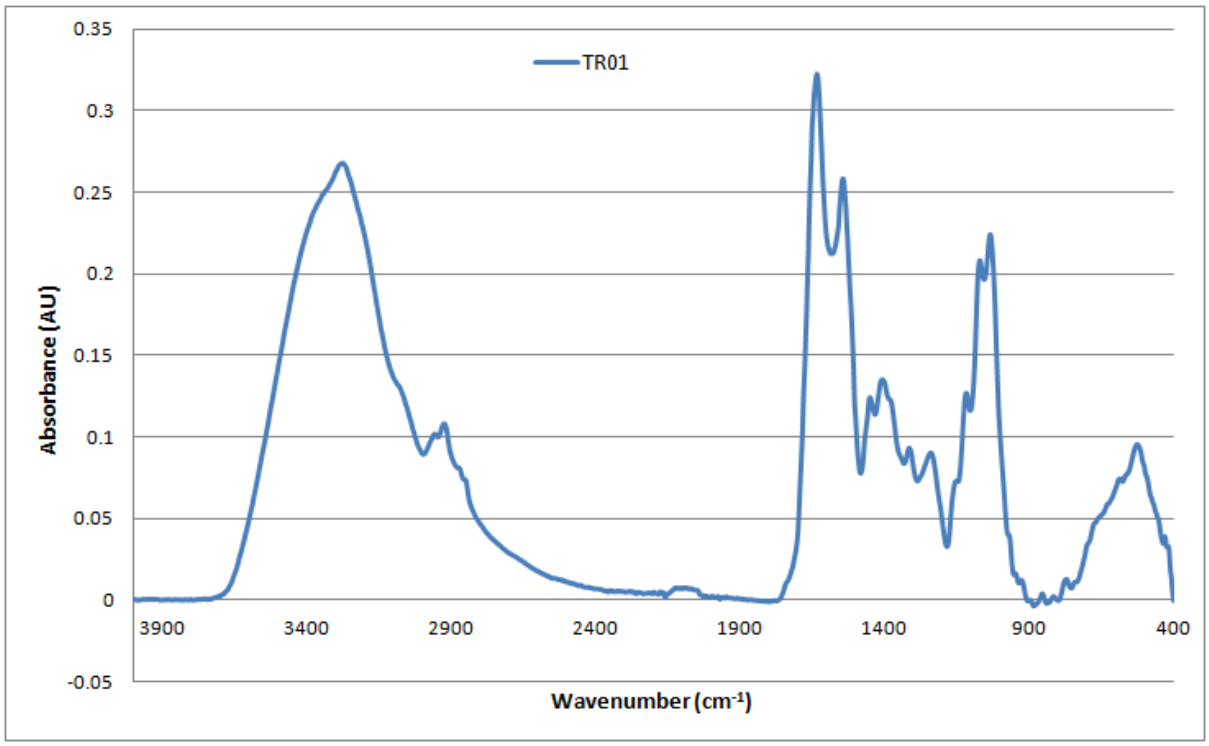


**Figure A2-111: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR01 from 4000-400cm$^{-1}$**

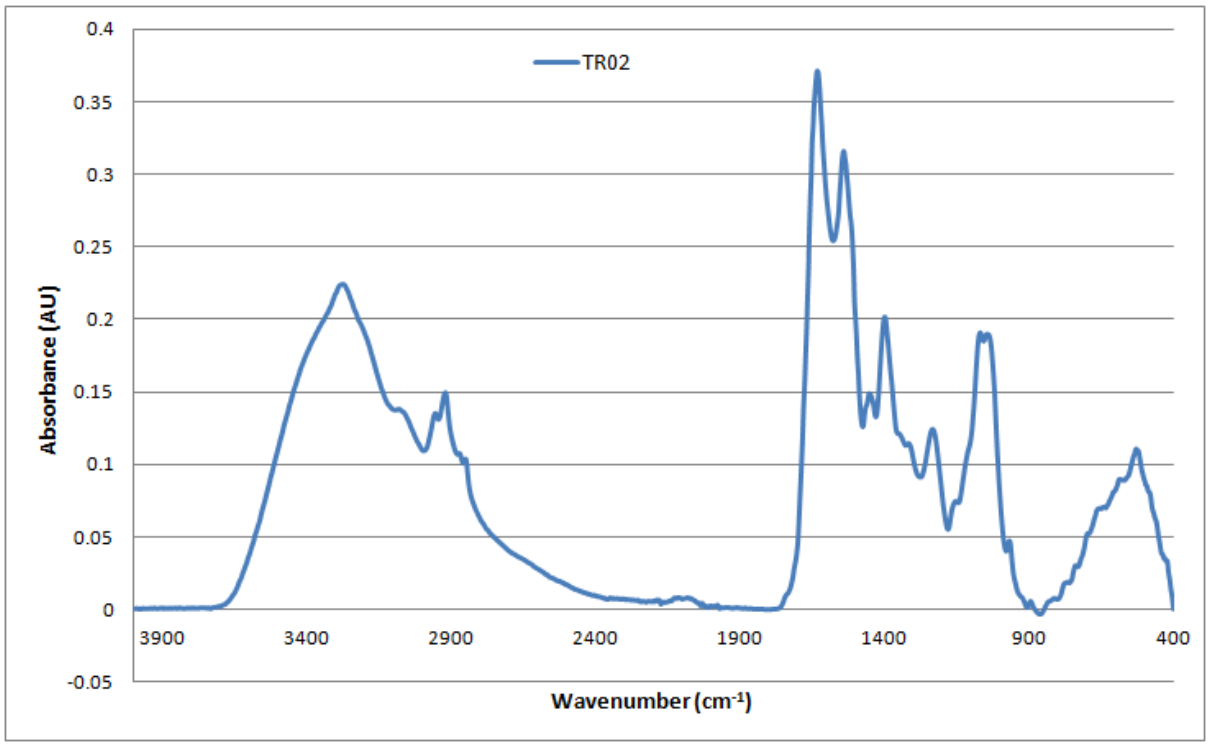


**Figure A2-112: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR02 from 4000-400cm$^{-1}$**

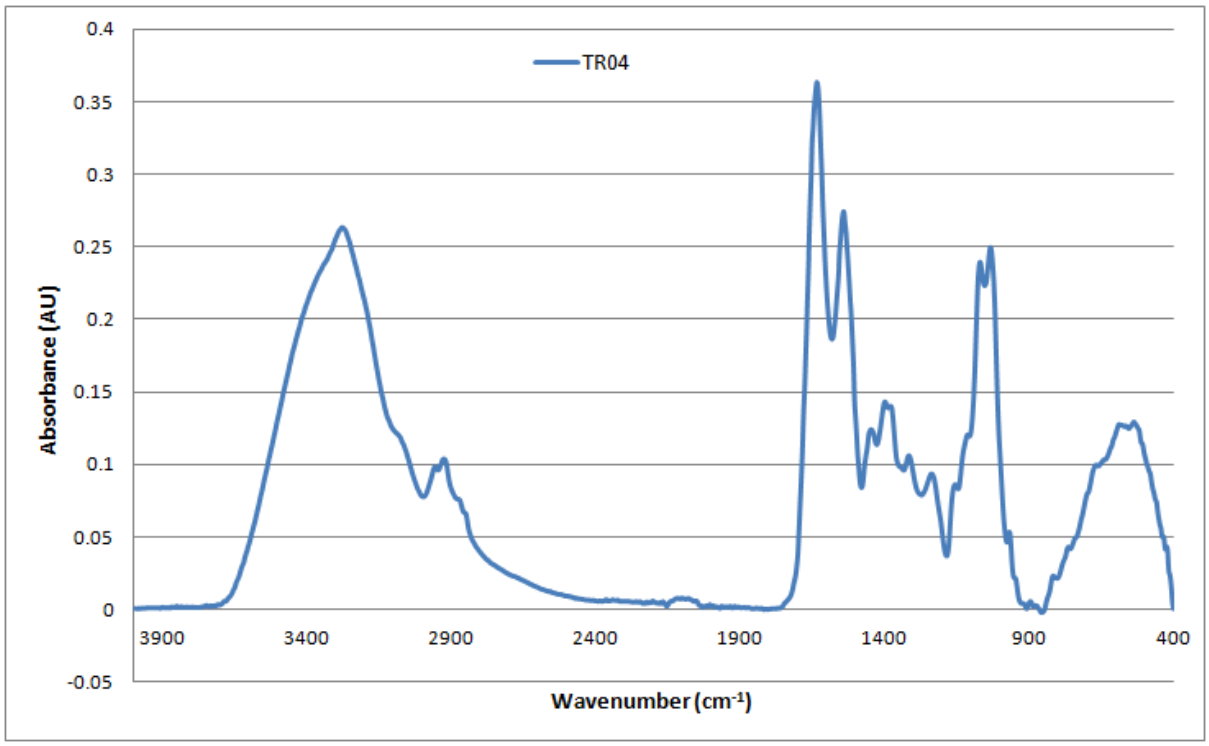


**Figure A2-113: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR04 from 4000-400cm$^{-1}$**

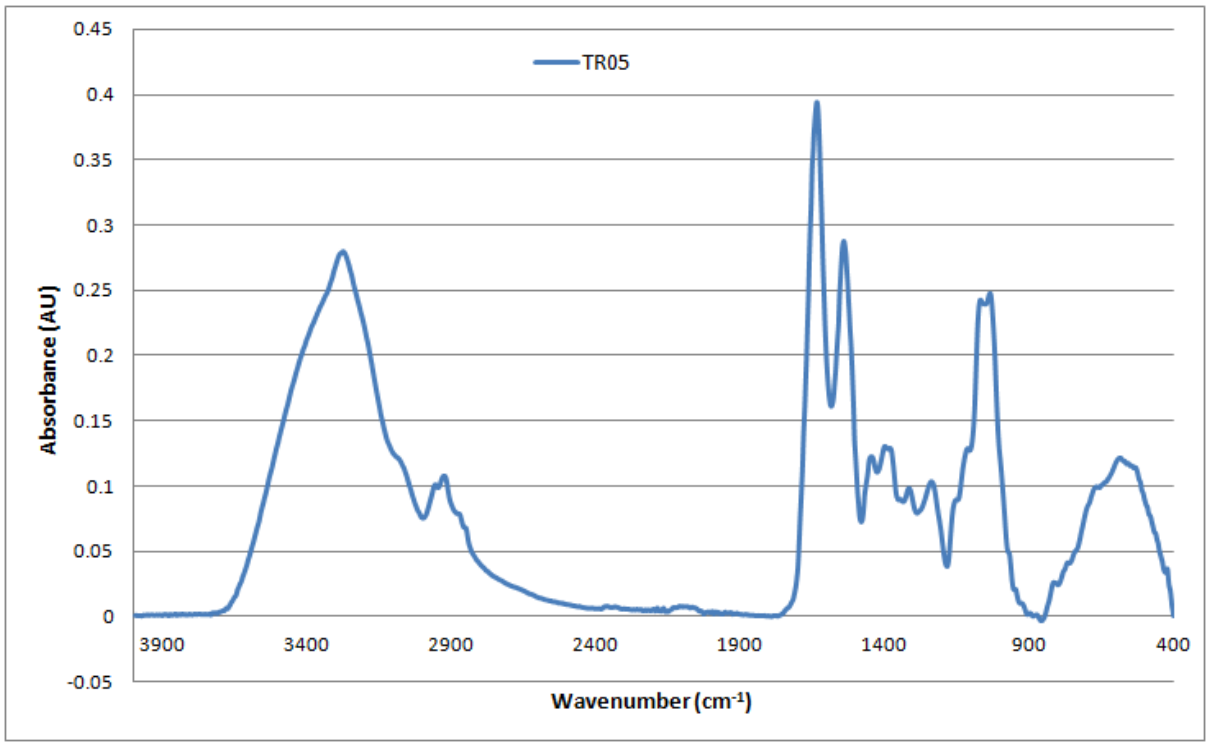


**Figure A2-114: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR05 from 4000-400cm$^{-1}$**

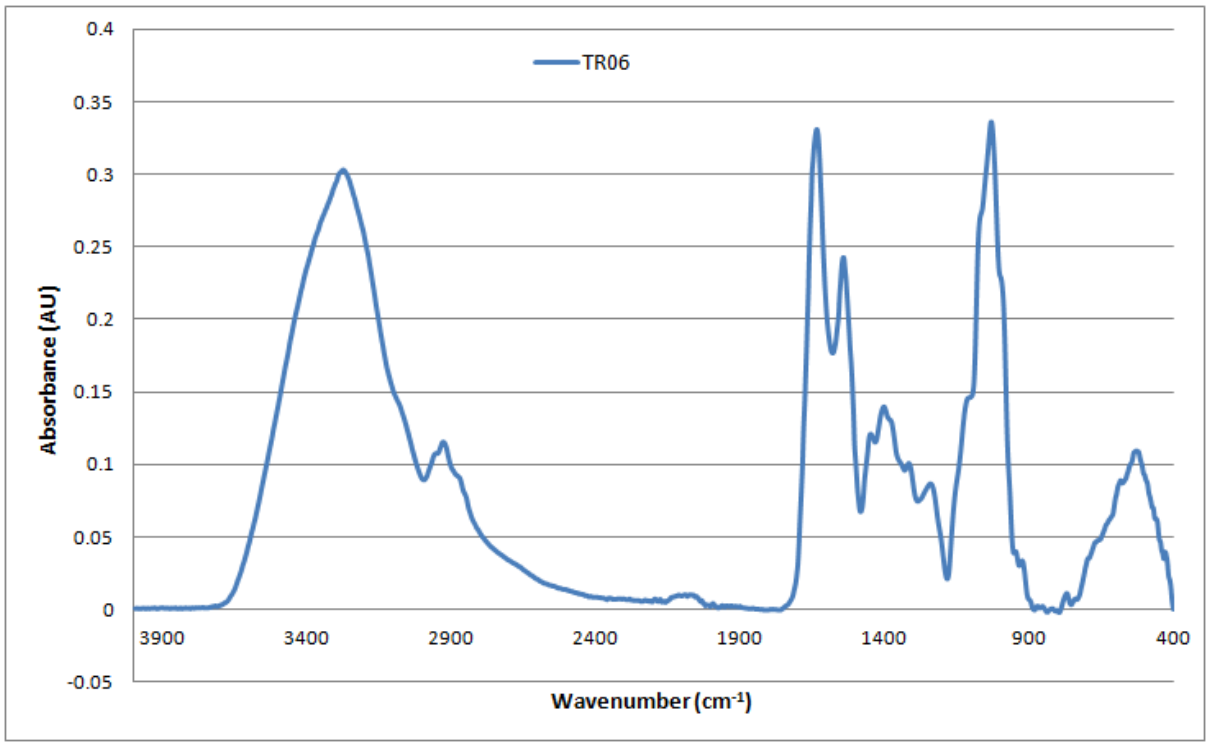


**Figure A2-115: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR06 from 4000-400cm$^{-1}$**

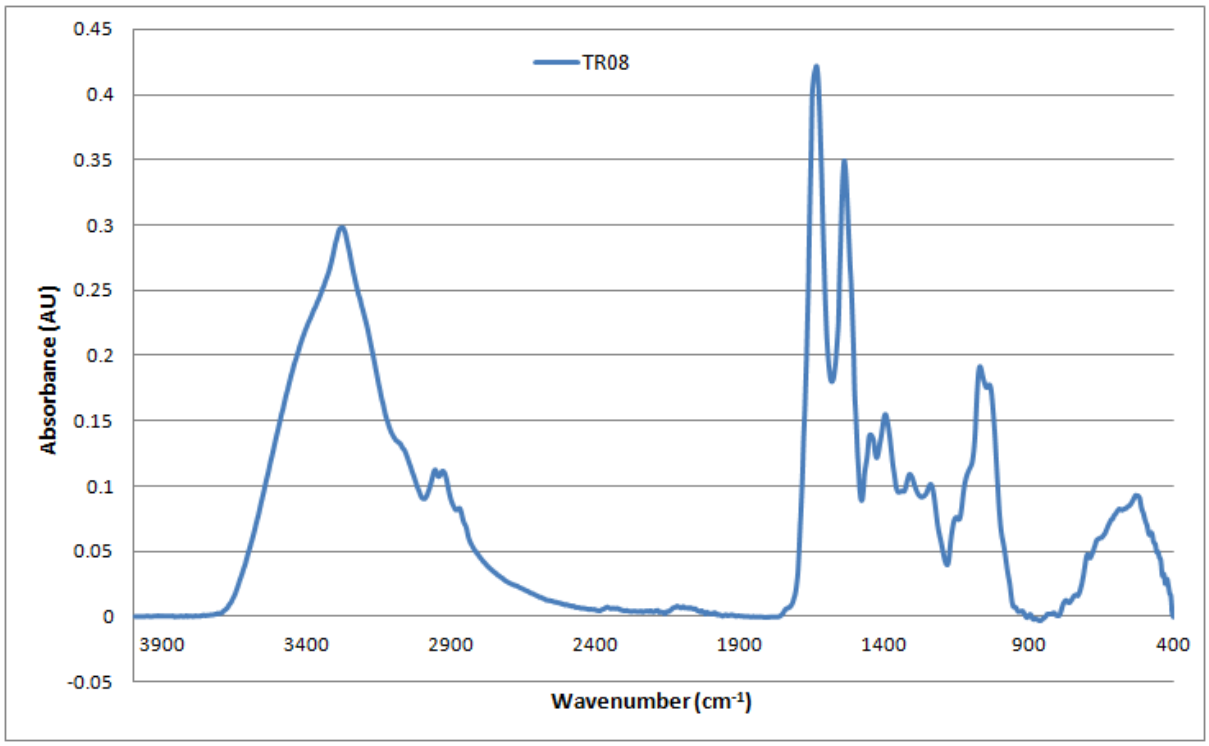


**Figure A2-116: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR08 from 4000-400cm$^{-1}$**

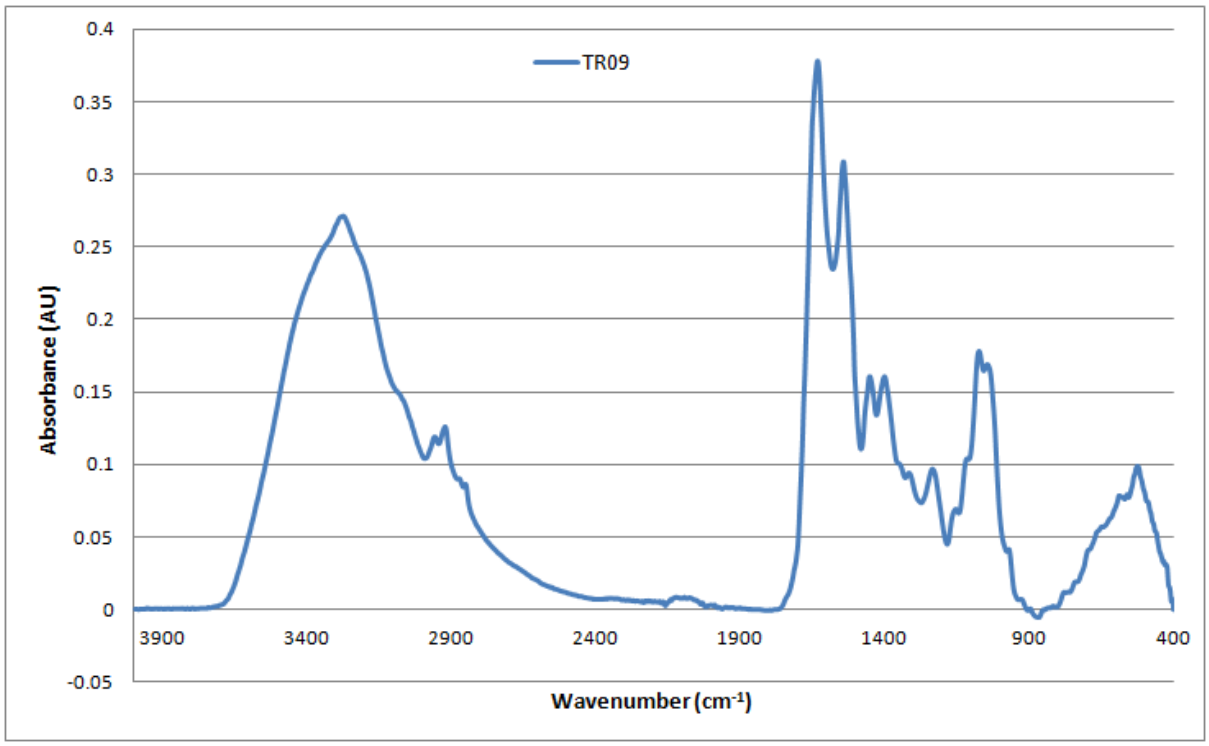


**Figure A2-117: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR09 from 4000-400cm$^{-1}$**

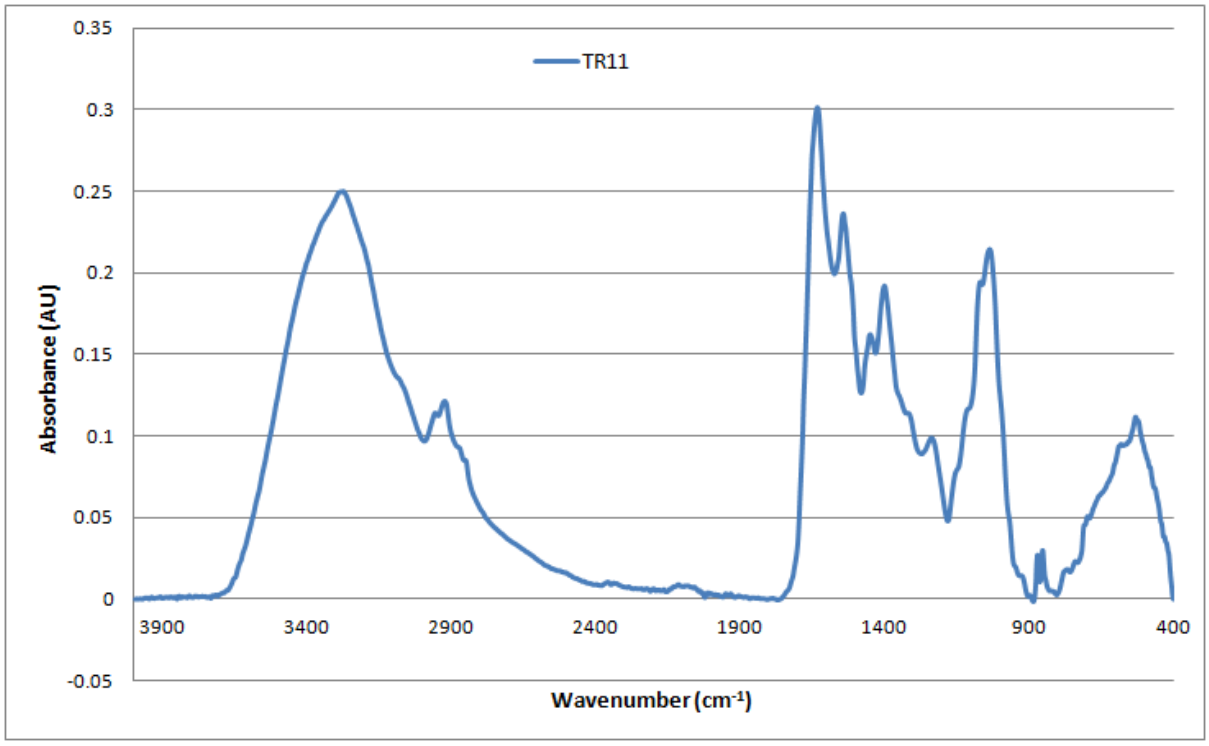


**Figure A2-118: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR11 from 4000-400cm$^{-1}$**

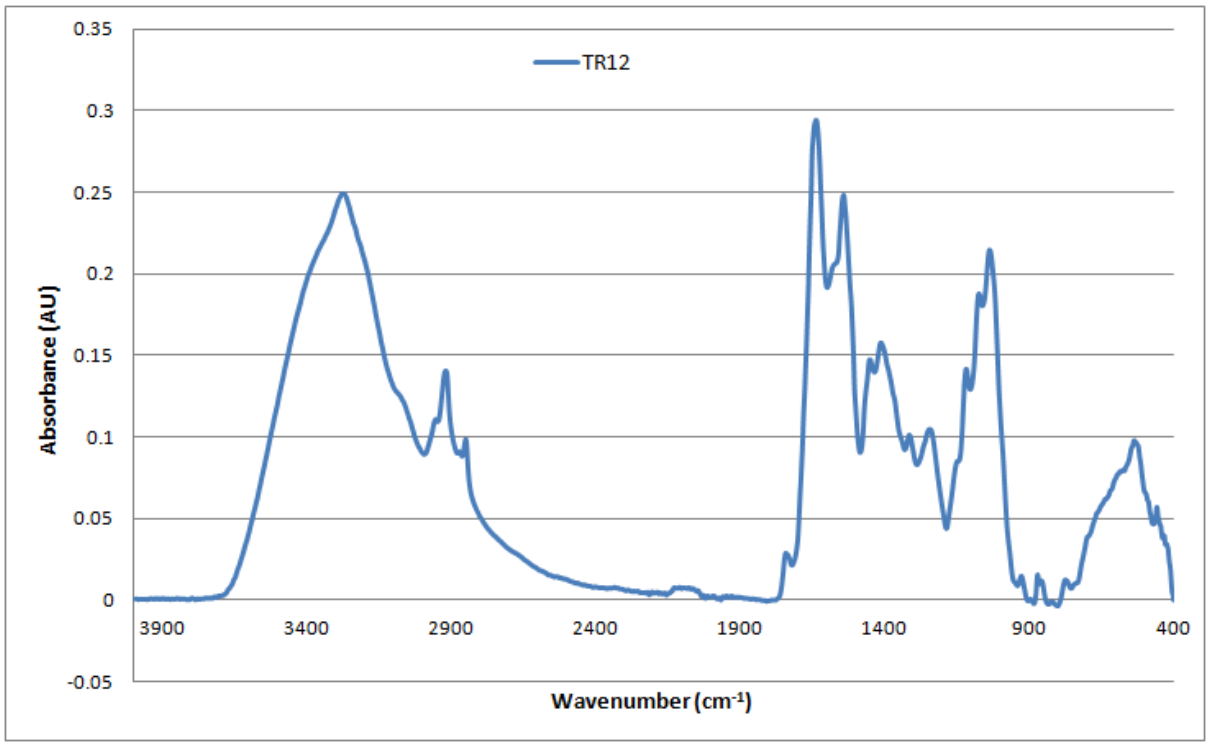


**Figure A2-119: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient TR12 from 4000-400cm$^{-1}$**

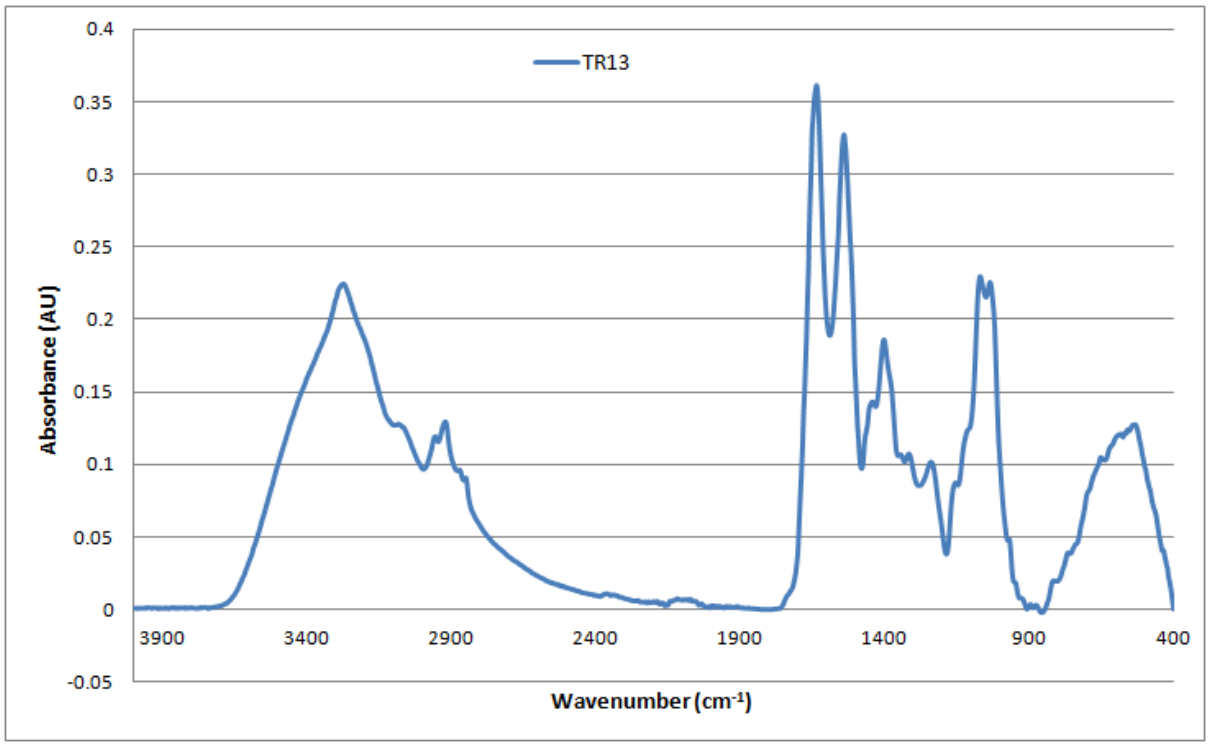


**Figure A2-120: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient TR13 from 4000-400cm$^{-1}$**

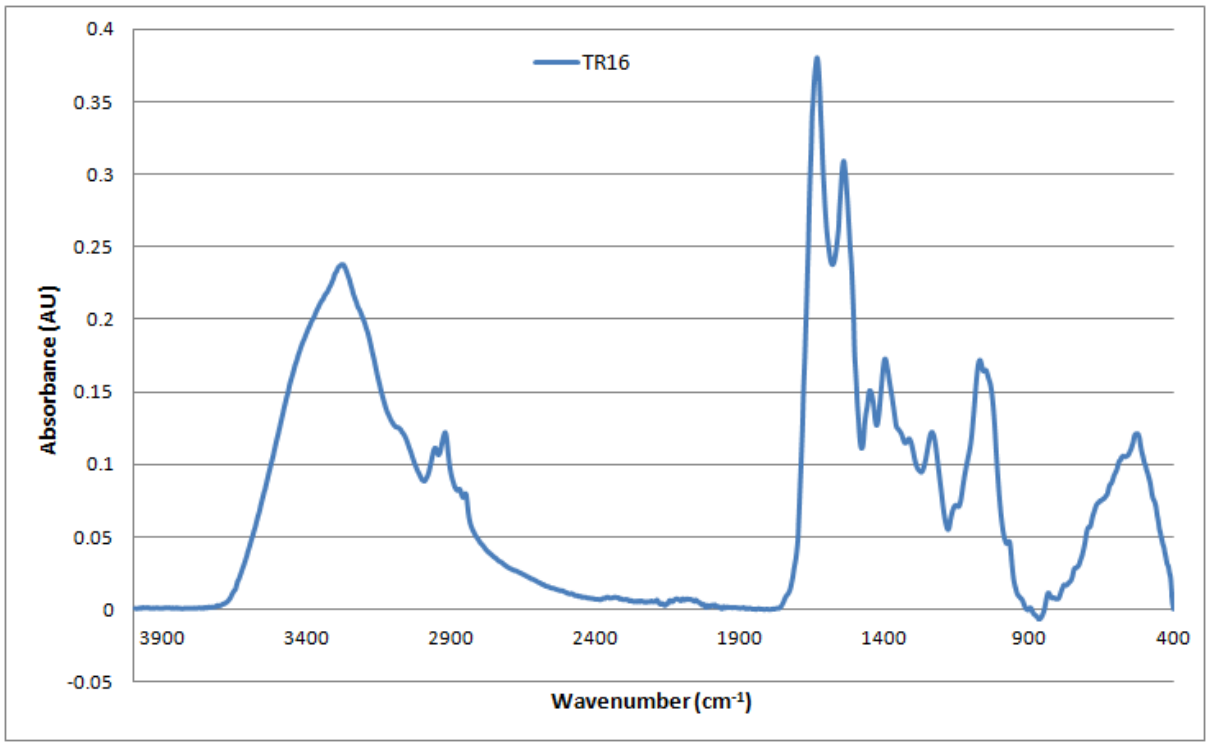


**Figure A2-121: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient TR16 from 4000-400cm$^{-1}$**

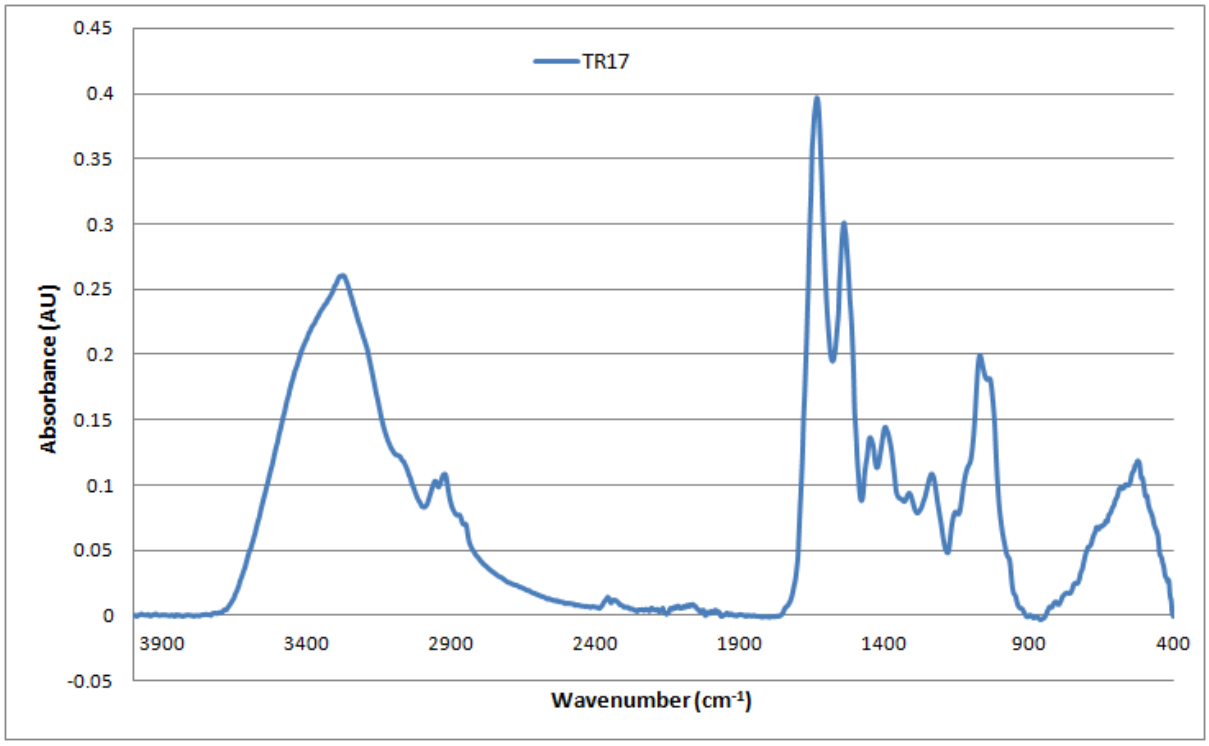


**Figure A2-122: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR17 from 4000-400cm$^{-1}$**

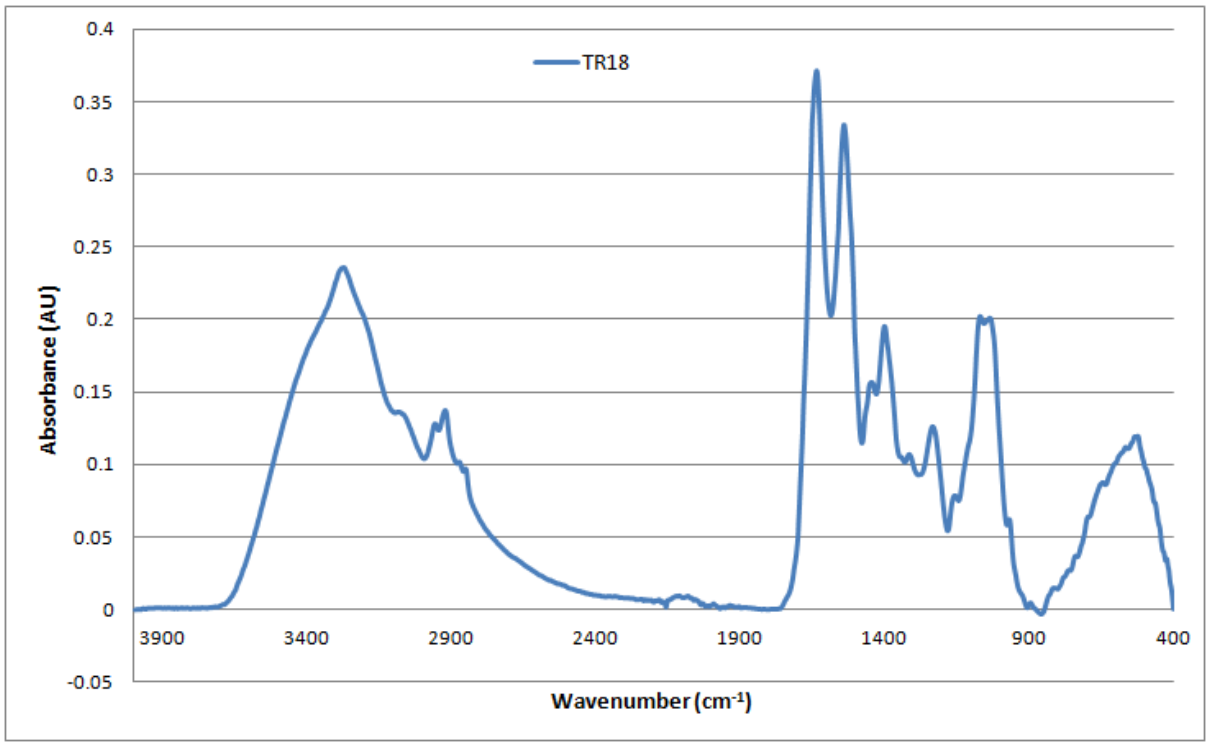


**Figure A2-123: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient TR18 from 4000-400cm$^{-1}$**

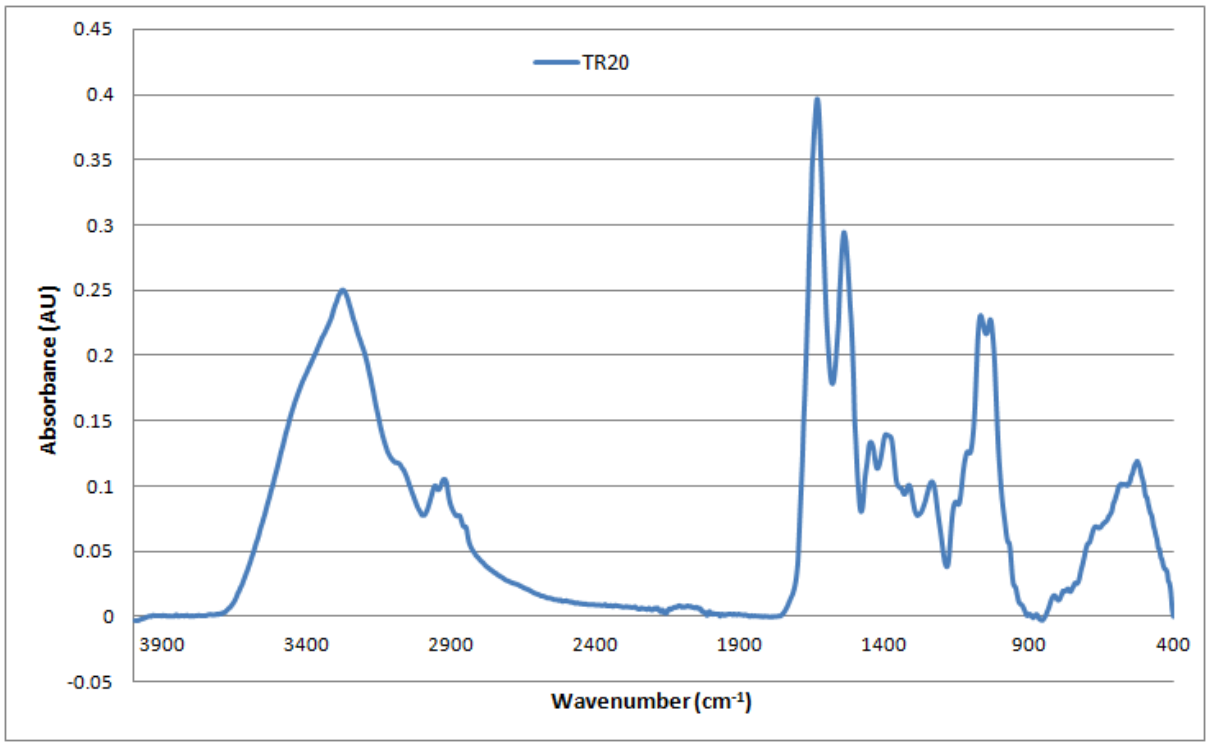


**Figure A2-124: Baseline-corrected absorbance spectrum of MEDLUNG COPD baseline patient TR20 from 4000-400cm$^{-1}$**

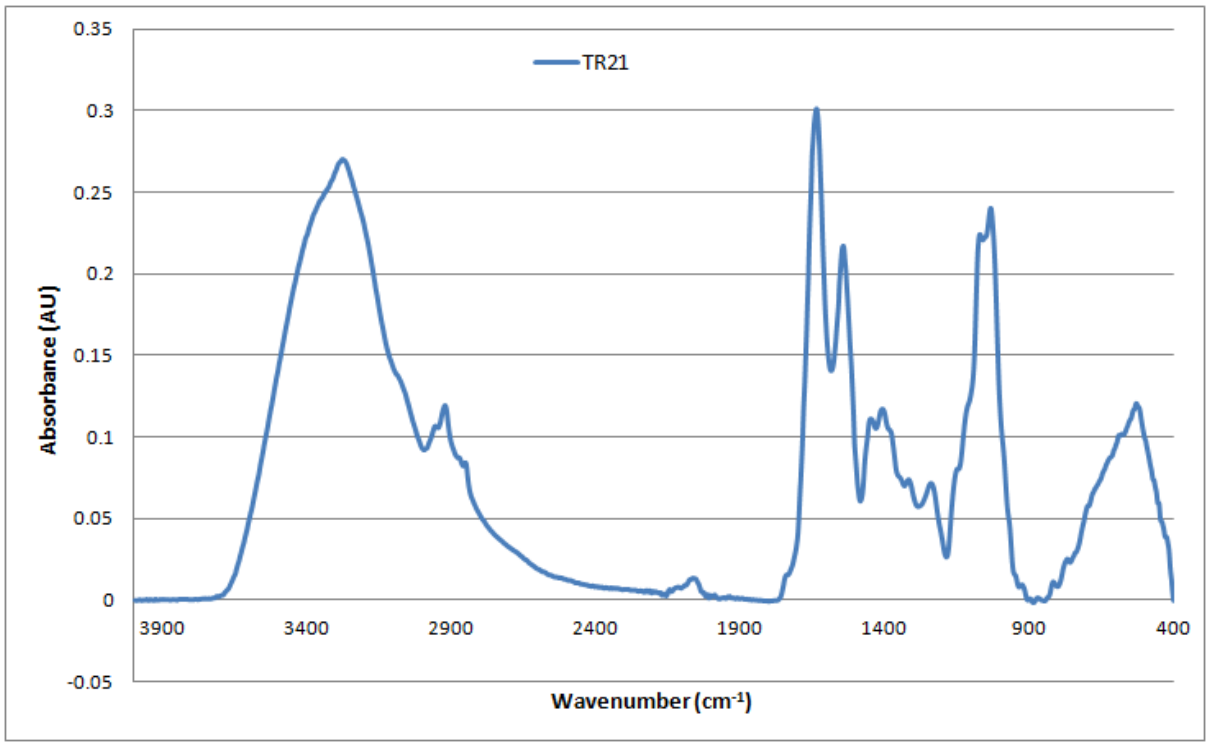


**Figure A2-125: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR21 from 4000-400cm$^{-1}$**

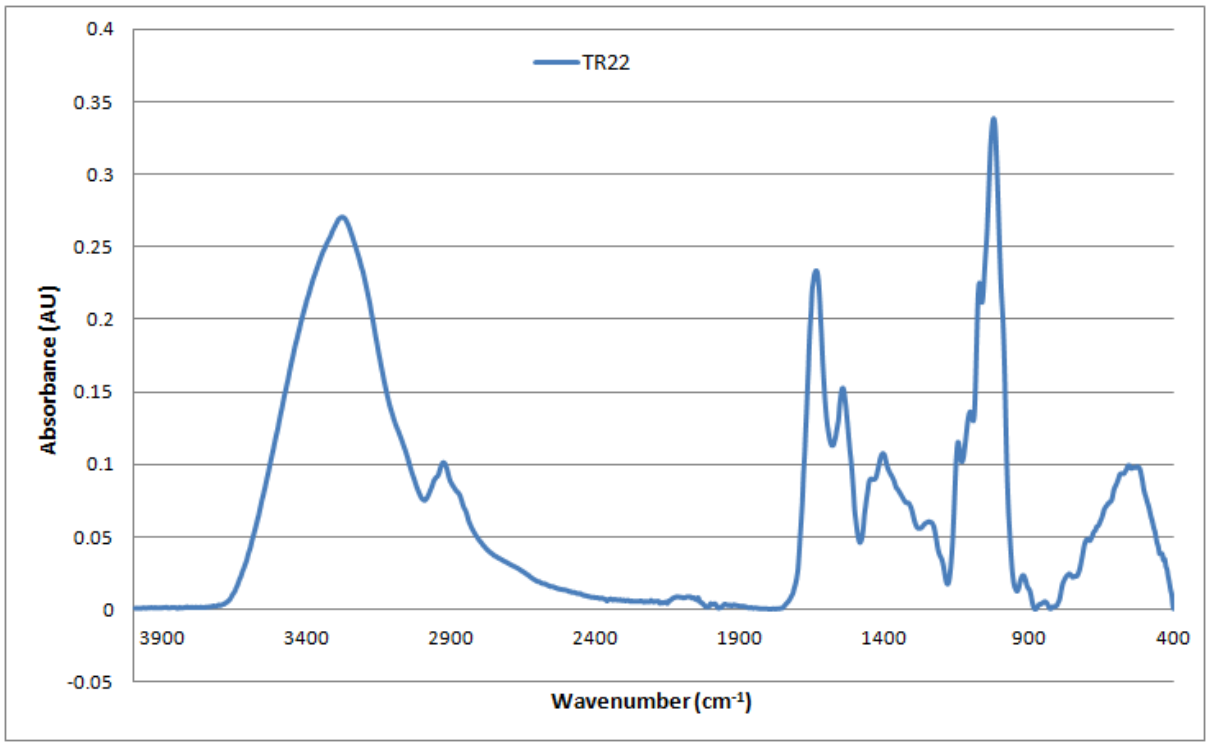


**Figure A2-126: Baseline-corrected absorbance spectrum of MEDLUNG COPD exacerbator patient TR22 from 4000-400cm$^{-1}$**

## A2.2 Lung Cancer Patient Sputum Average Spectra

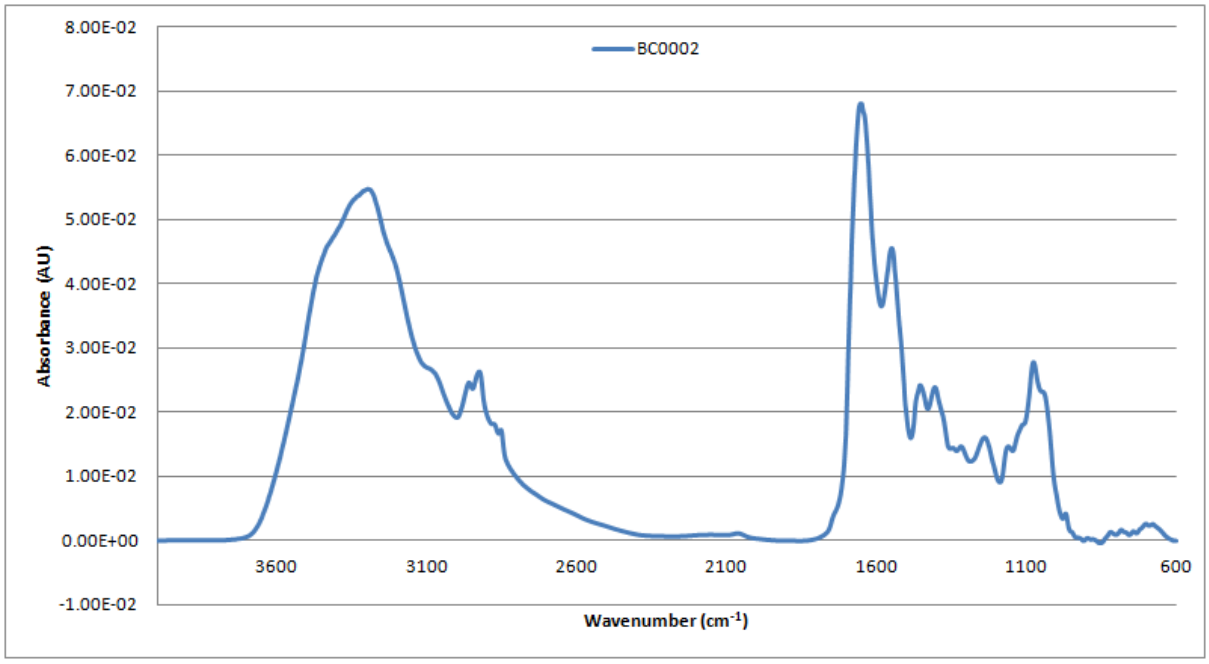


**Figure A2-127: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BC0002 from 4000-600cm$^{-1}$**

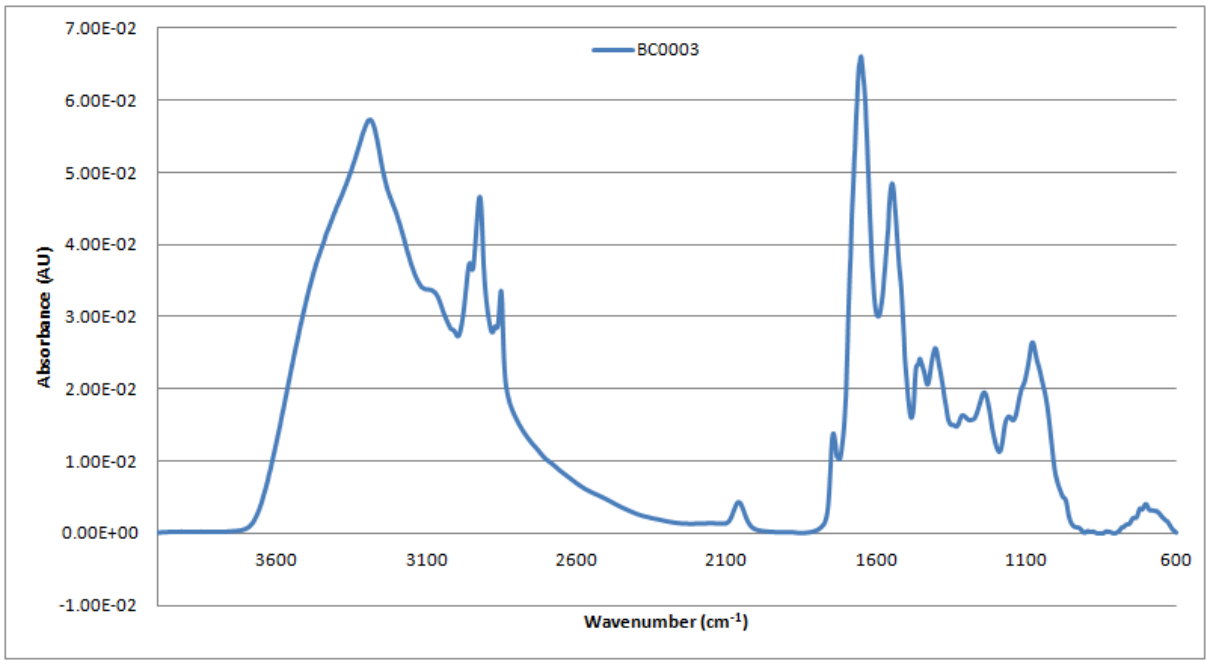


**Figure A2-128: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BC0003 from 4000-600cm$^{-1}$**

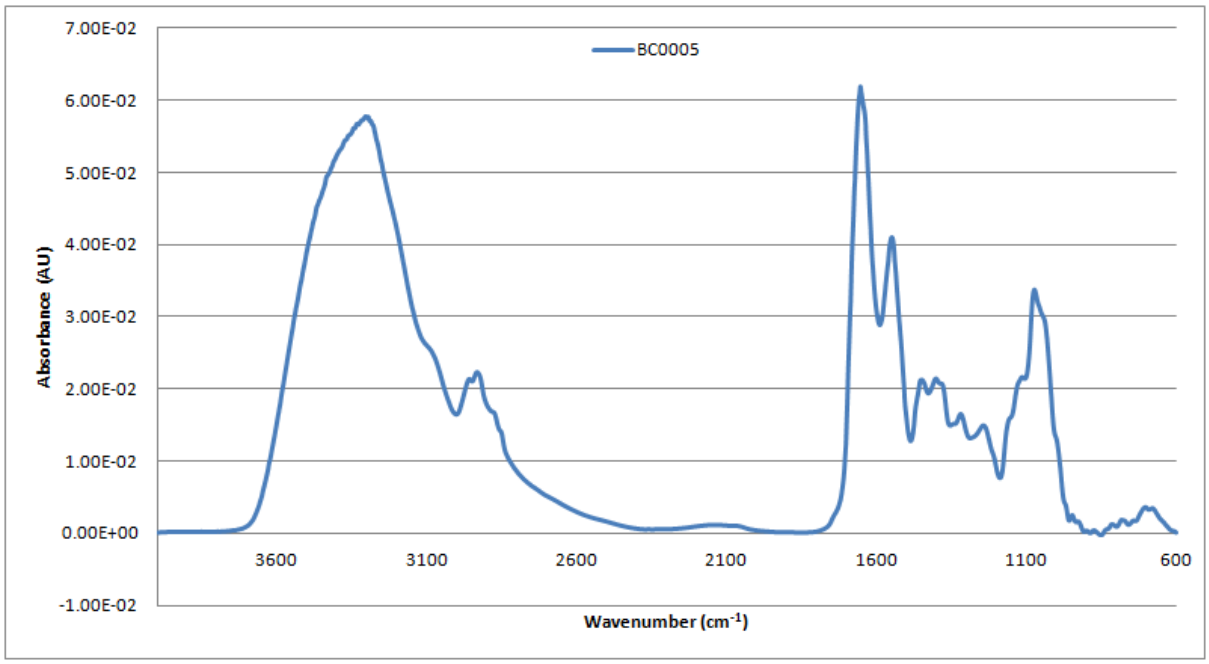


**Figure A2-129: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BC0005 from 4000-600cm$^{-1}$**

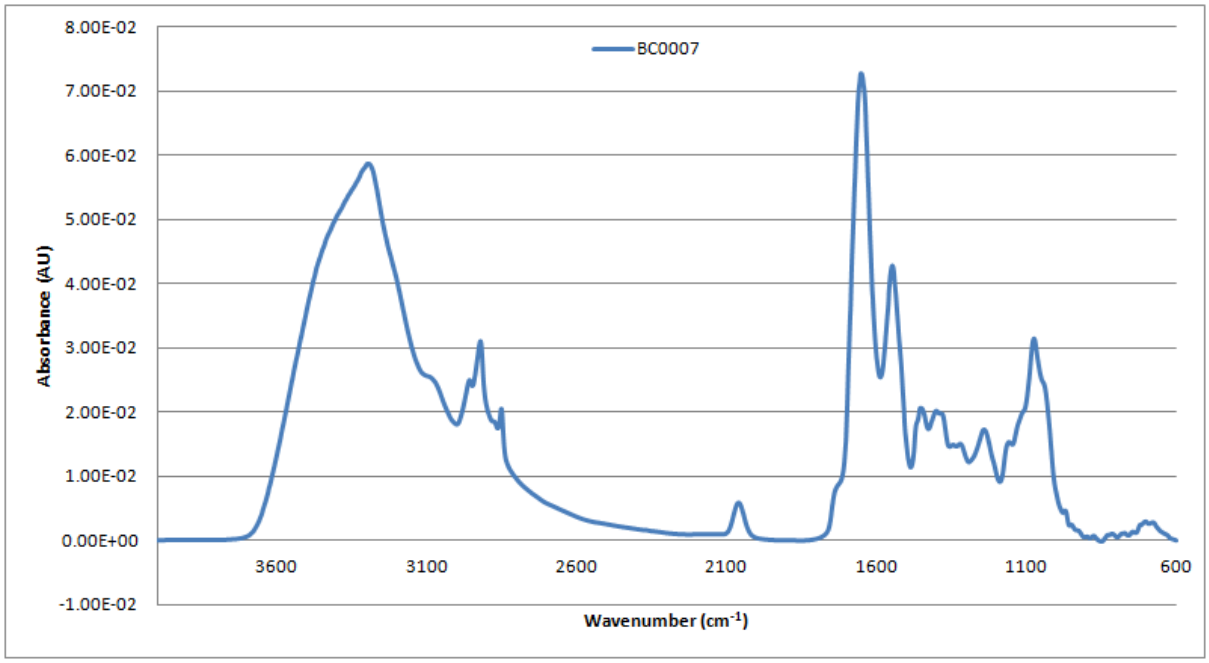


**Figure A2-130: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BC0007 from 4000-600cm$^{-1}$**

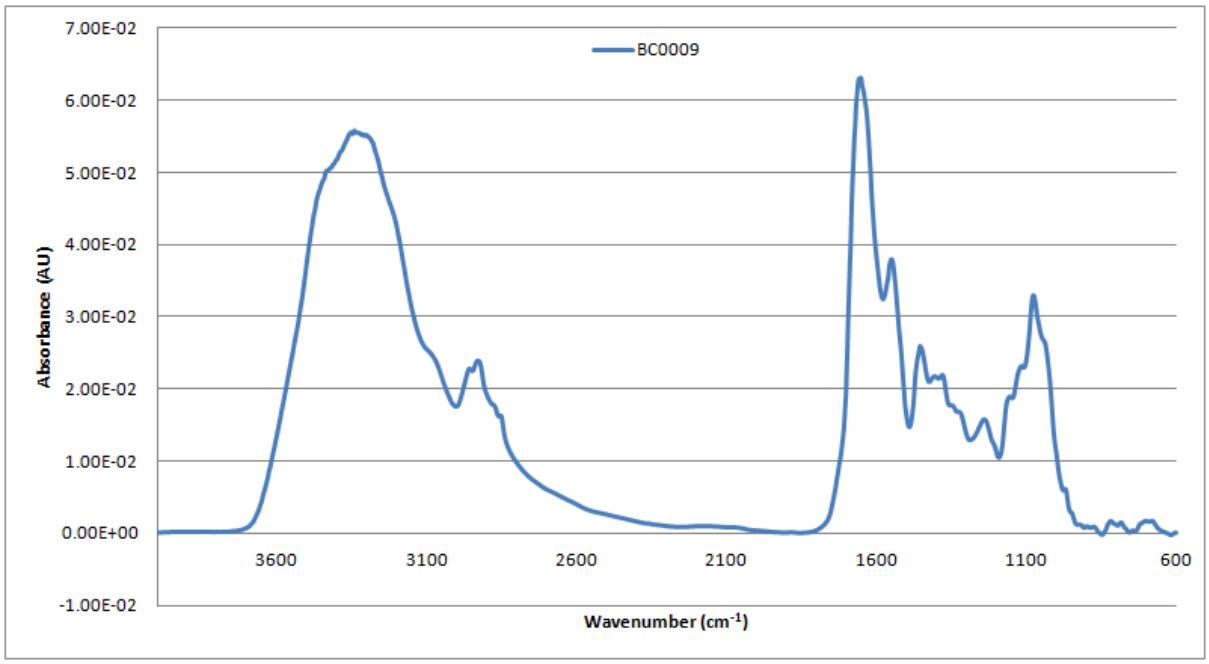


**Figure A2-131: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BC0009 from 4000-600cm$^{-1}$**

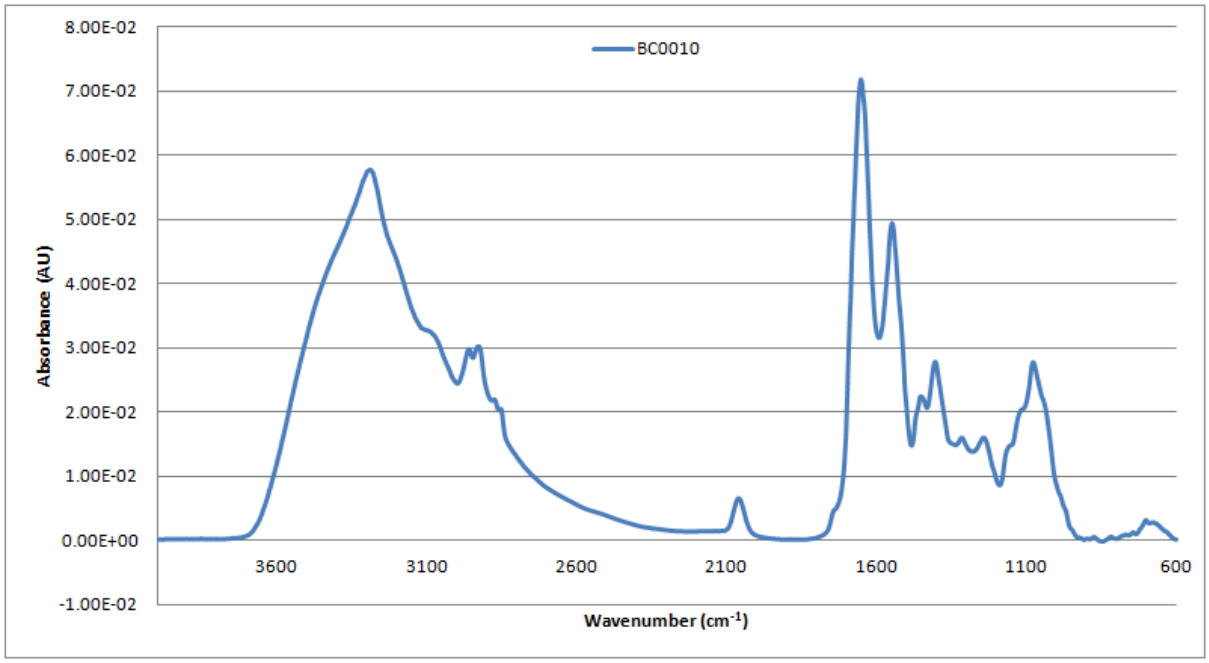


**Figure A2-132: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BC0010 from 4000-600cm$^{-1}$**

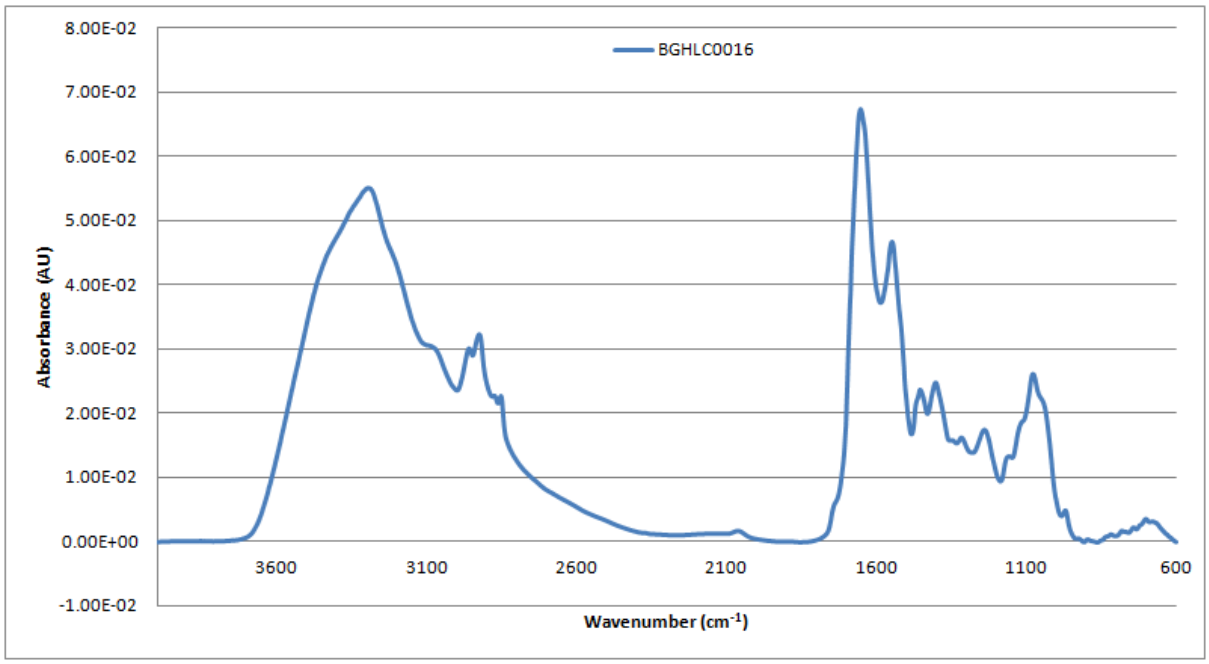


**Figure A2-133: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BGHLC0016 from 4000-600cm$^{-1}$**

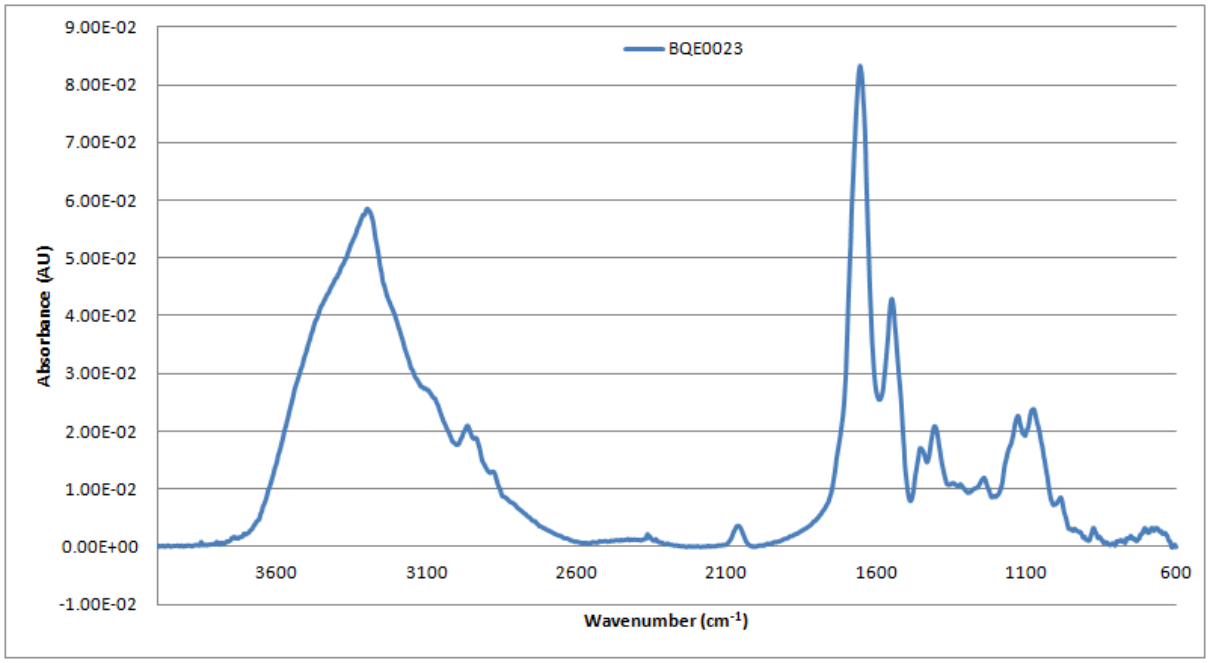


**Figure A2-134: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BQE0023 from 4000-600cm$^{-1}$**

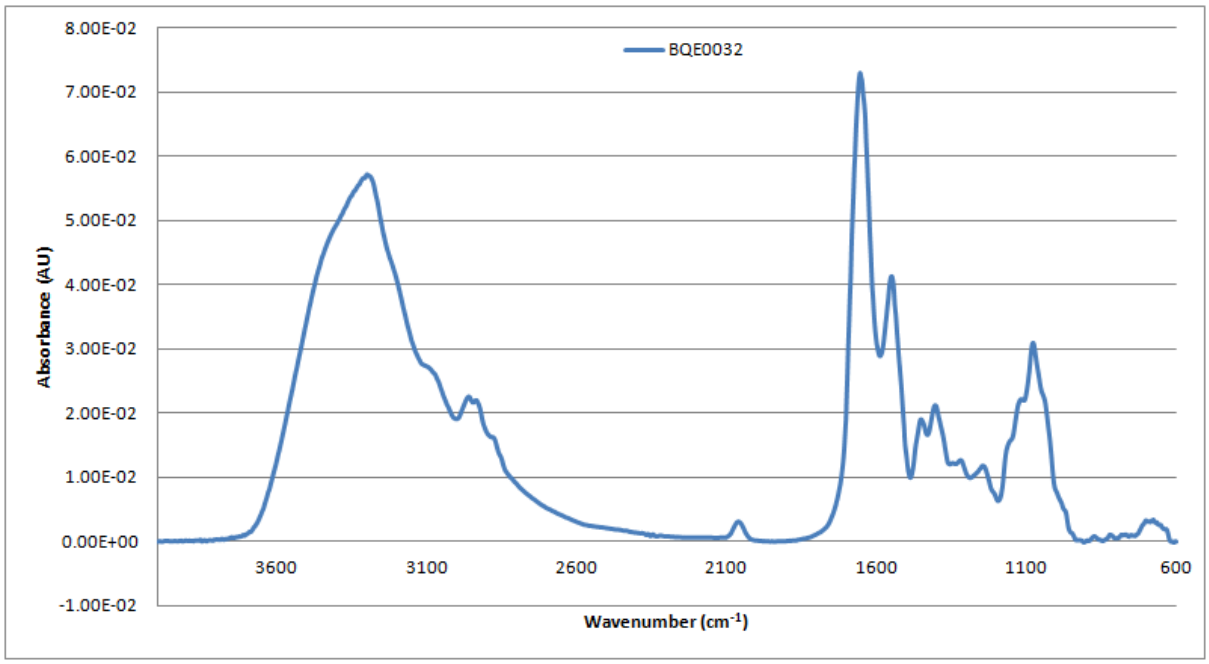


**Figure A2-135: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BQE0032 from 4000-600cm$^{-1}$**

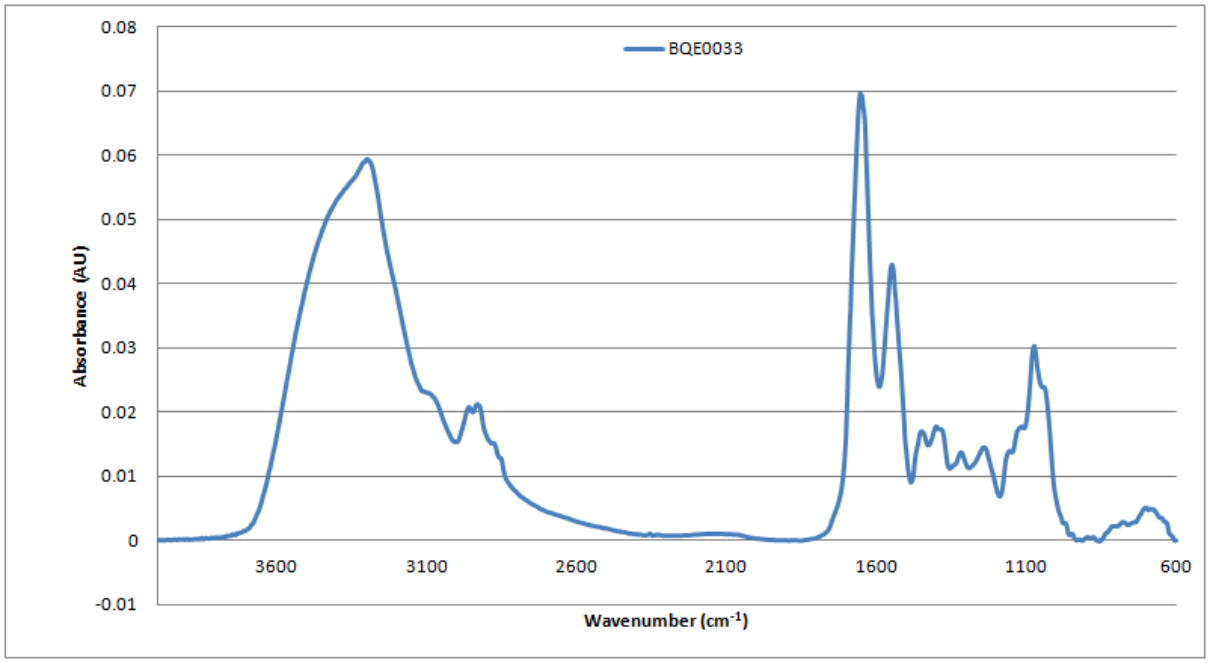


**Figure A2-136: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BQE0033 from 4000-600cm$^{-1}$**

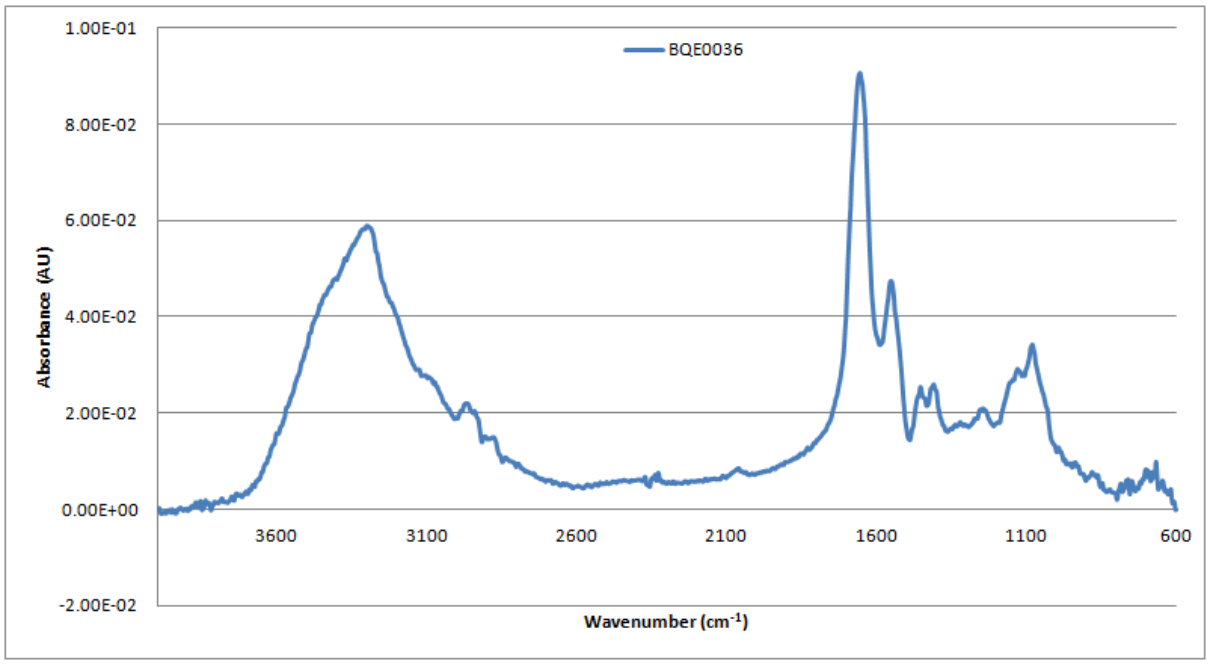


**Figure A2-137: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient BQE0036 from 4000-600cm$^{-1}$**

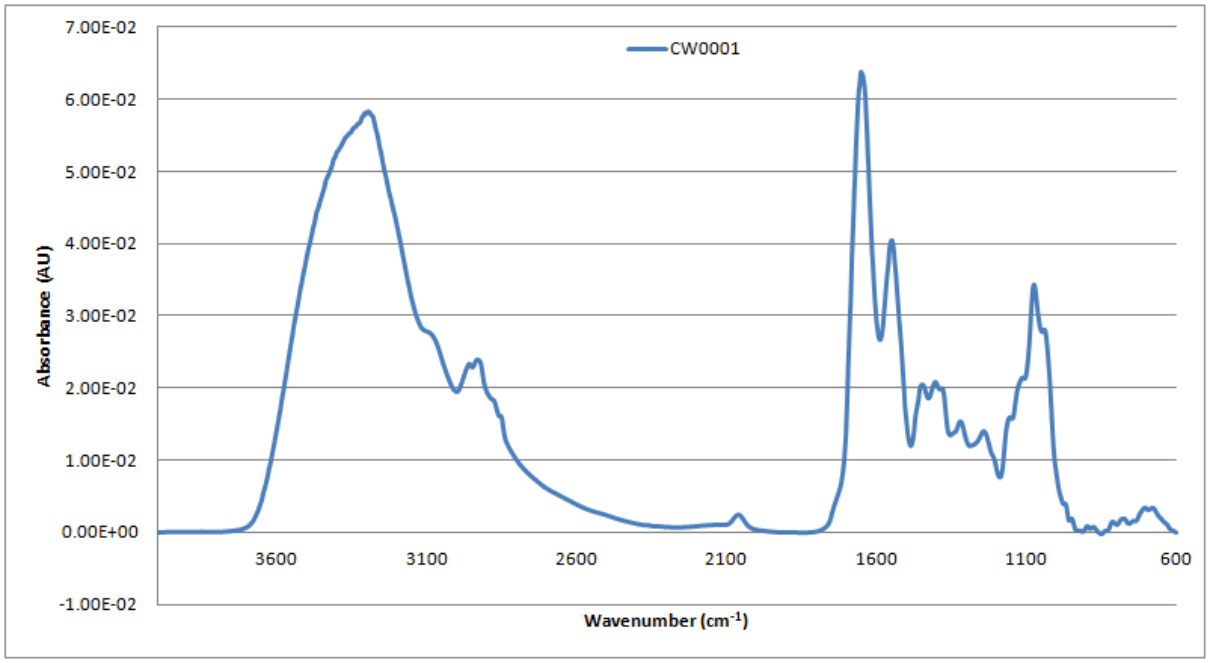


**Figure A2-138: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0001 from 4000-600cm$^{-1}$**

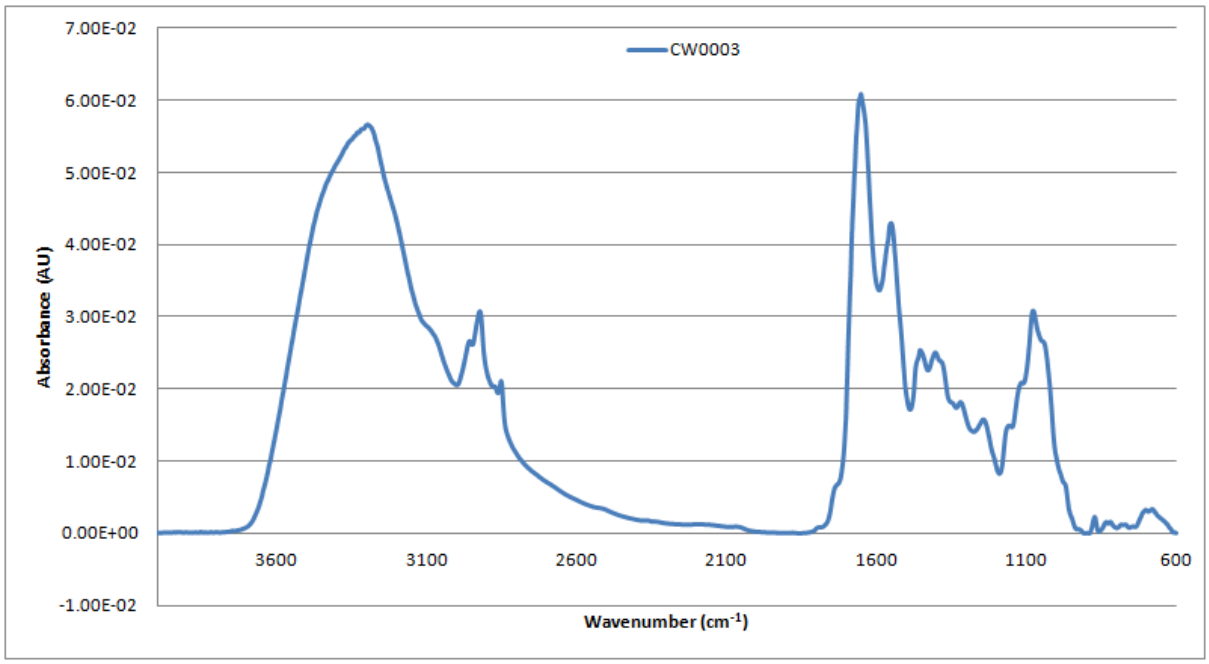


**Figure A2-139: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0003 from 4000-600cm$^{-1}$**

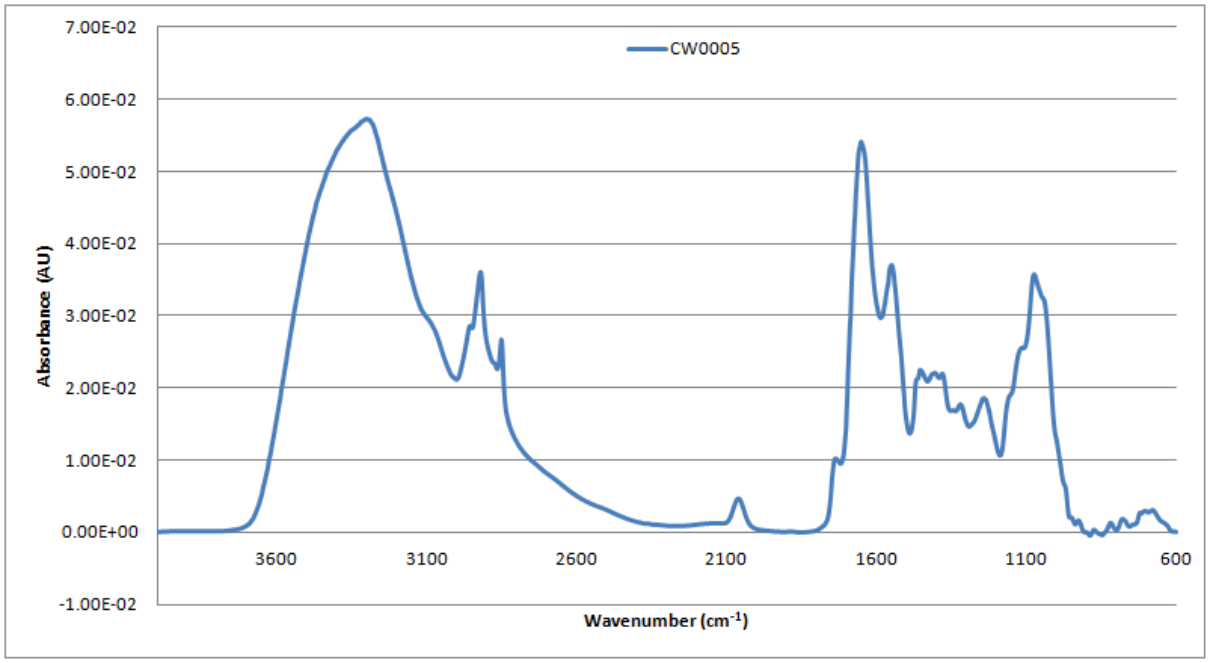


**Figure A2-140: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0005 from 4000-600cm$^{-1}$**

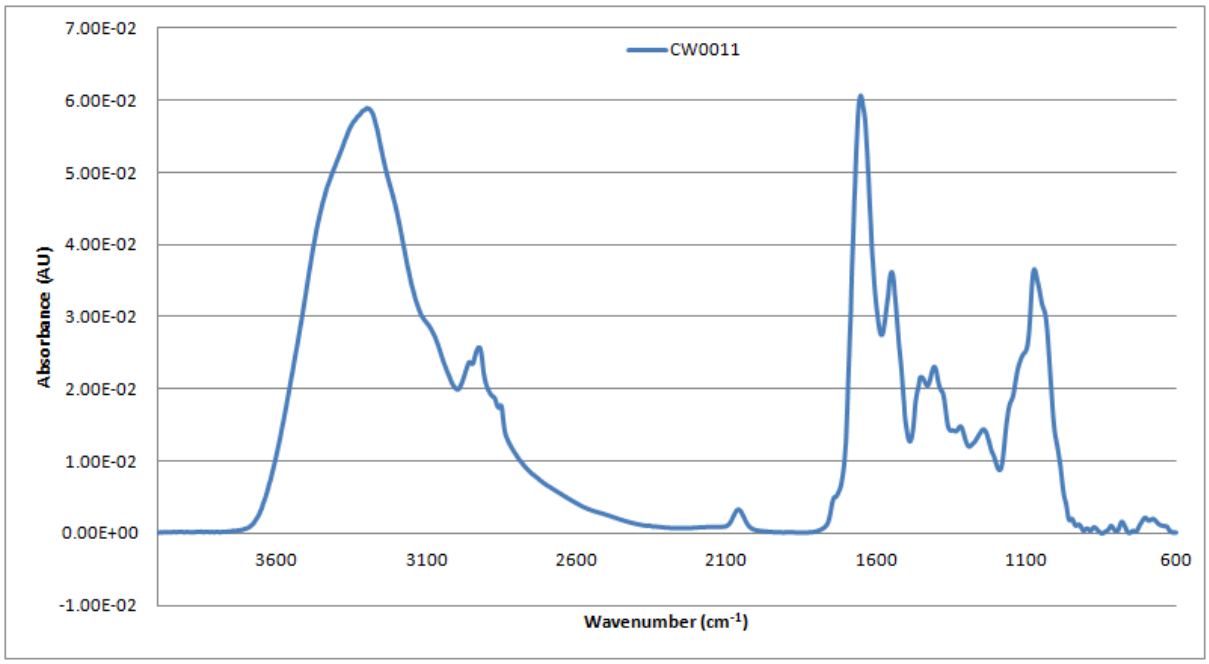


**Figure A2-141: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0011 from 4000-600cm$^{-1}$**

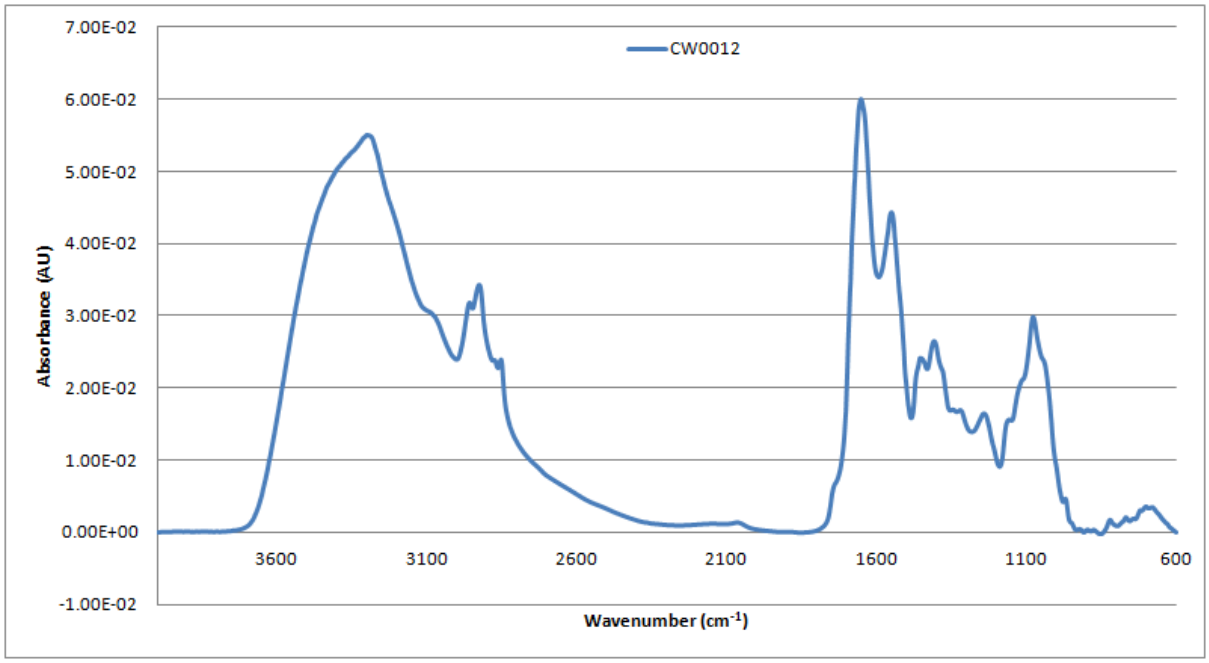


**Figure A2-142: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0012 from 4000-600cm$^{-1}$**

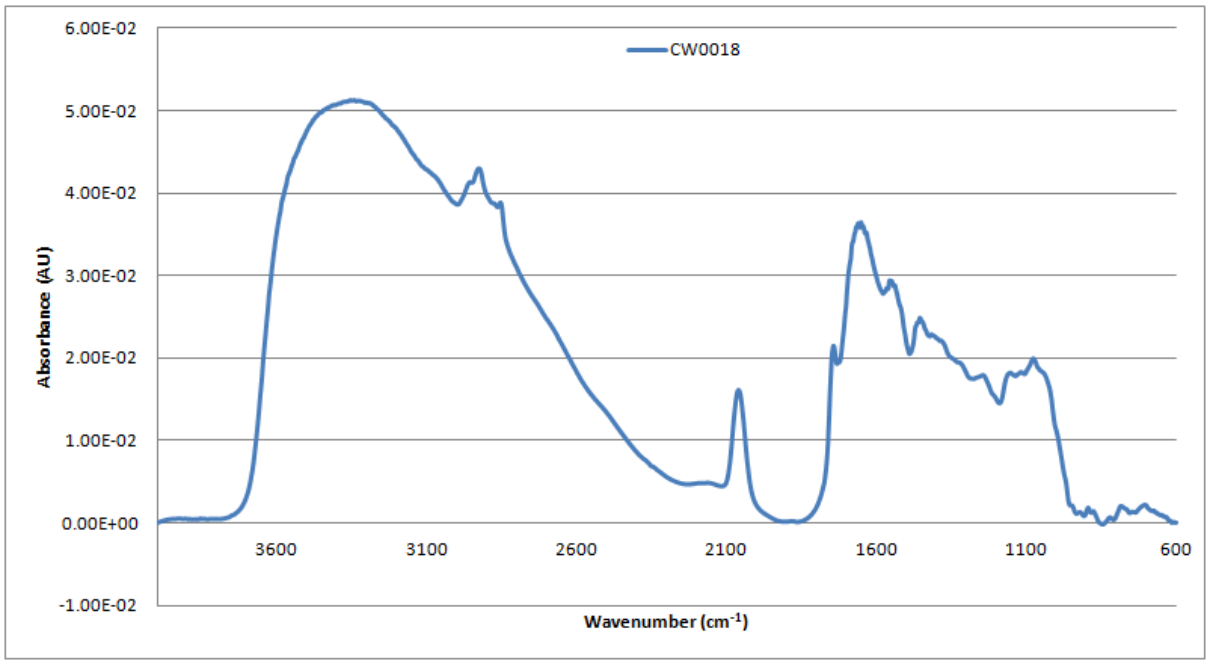


**Figure A2-143: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0018 from 4000-600cm$^{-1}$**

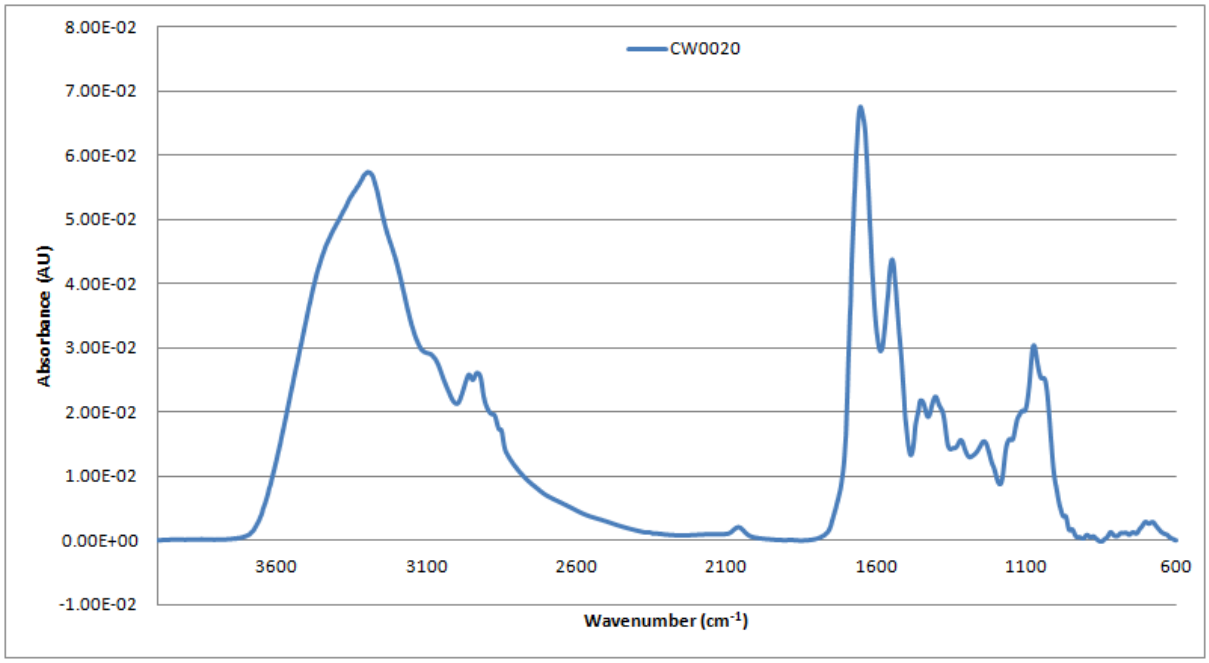


**Figure A2-144: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0020 from 4000-600cm$^{-1}$**

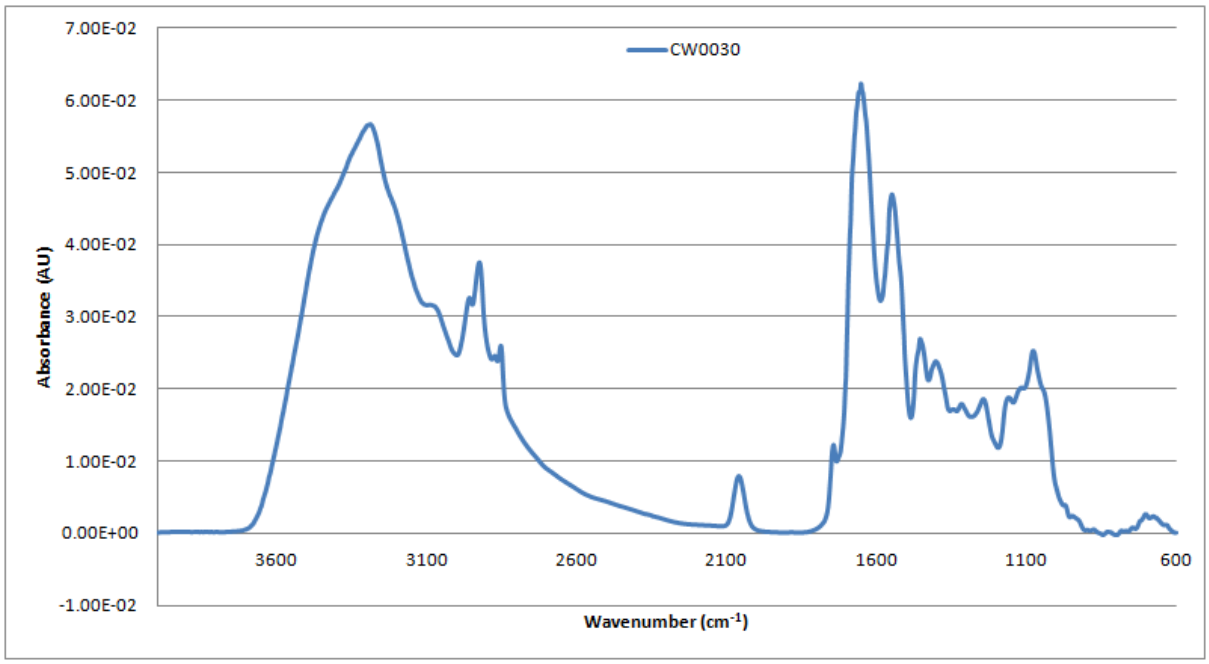


**Figure A2-145: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0030 from 4000-600cm$^{-1}$**

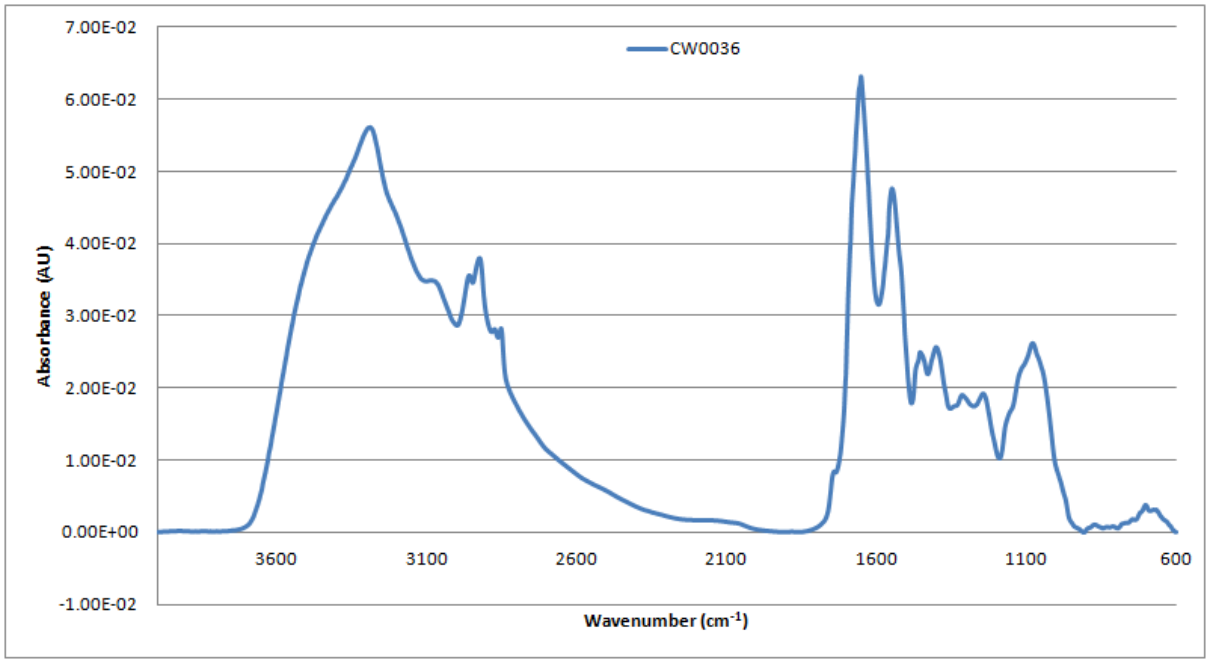


**Figure A2-146: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0036 from 4000-600cm$^{-1}$**

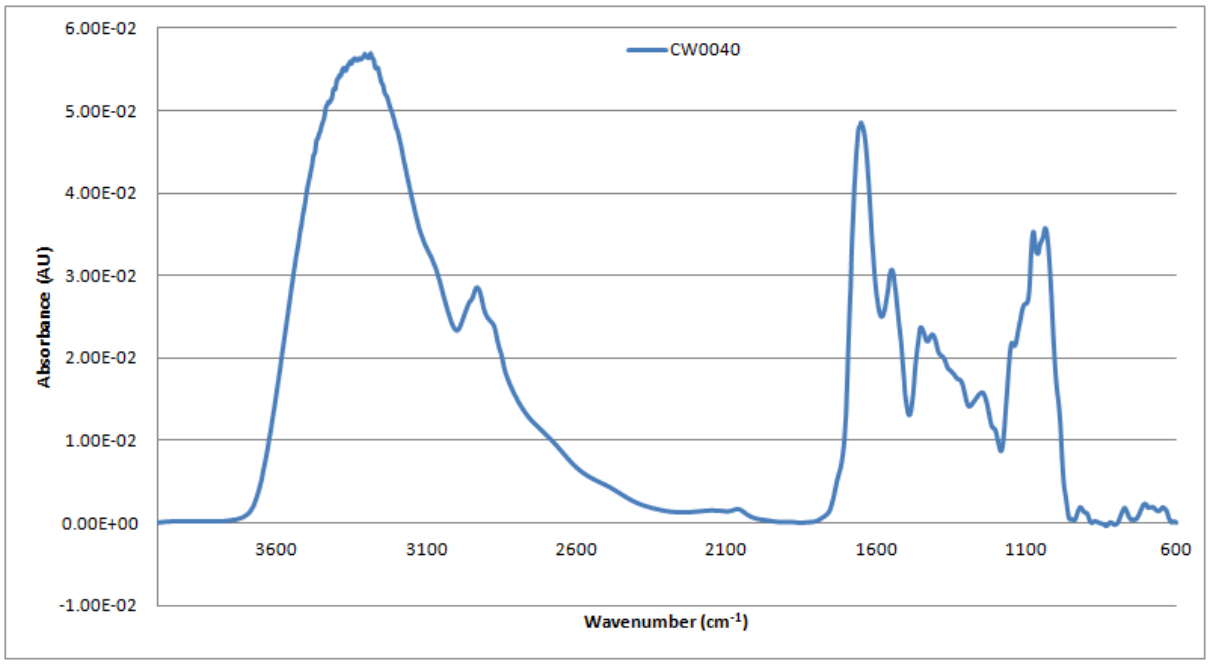


**Figure A2-147: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0040 from 4000-600cm$^{-1}$**

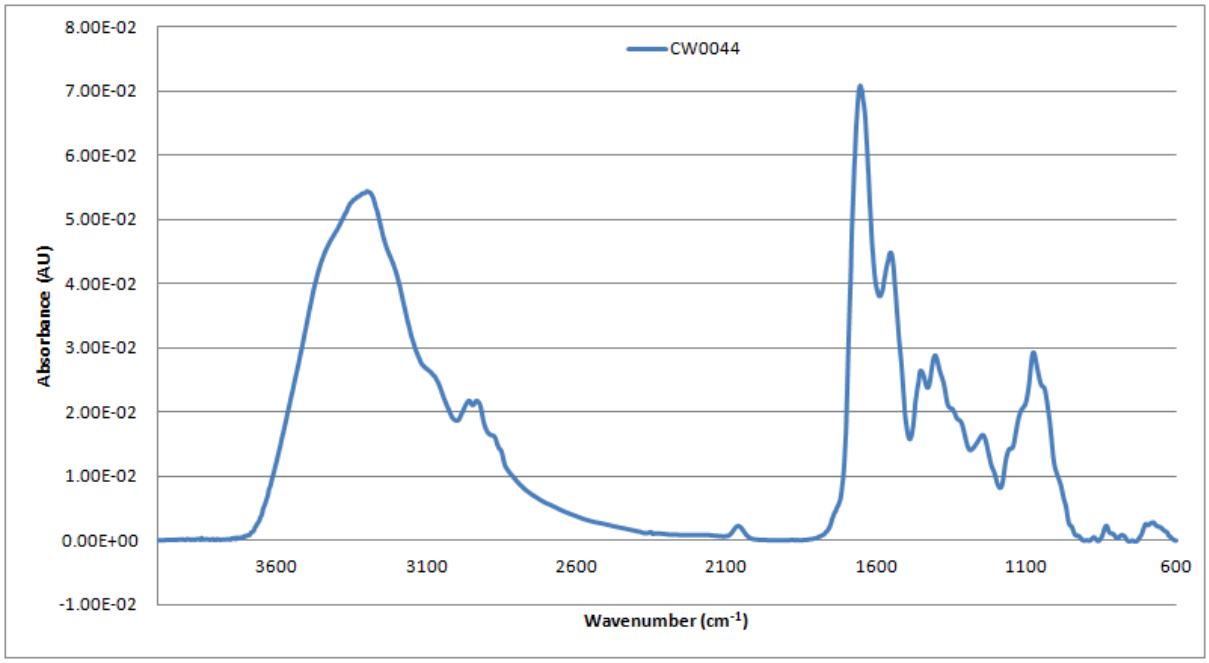


**Figure A2-148: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0044 from 4000-600cm$^{-1}$**

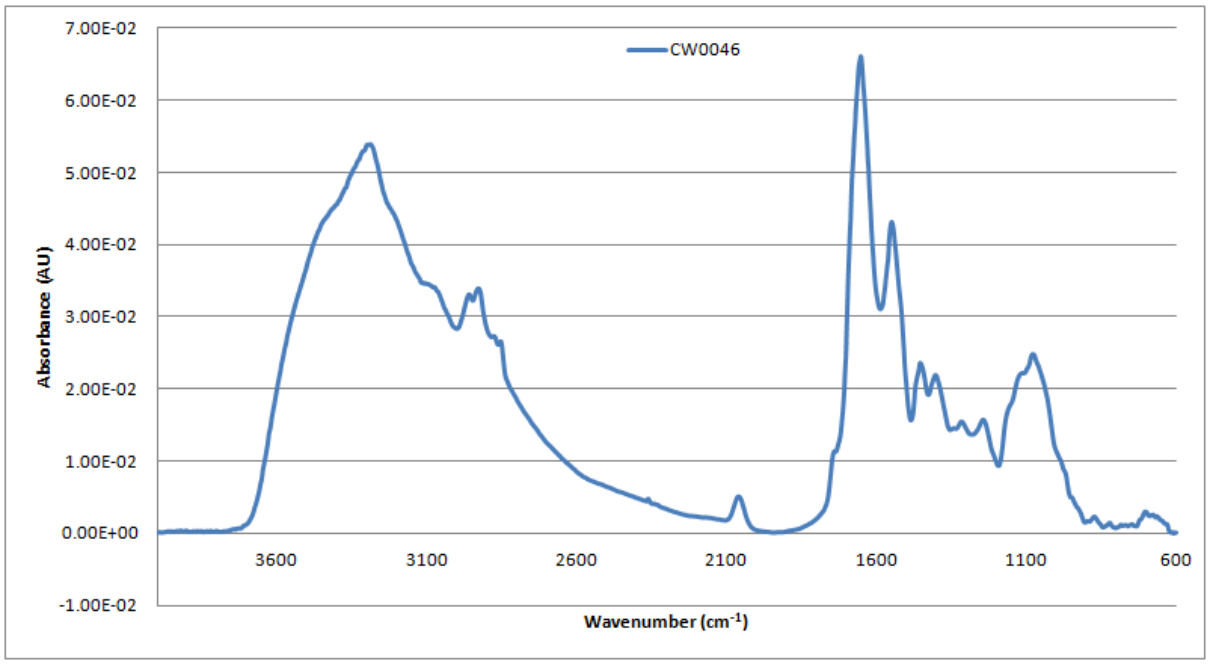


**Figure A2-149: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0046 from 4000-600cm$^{-1}$**

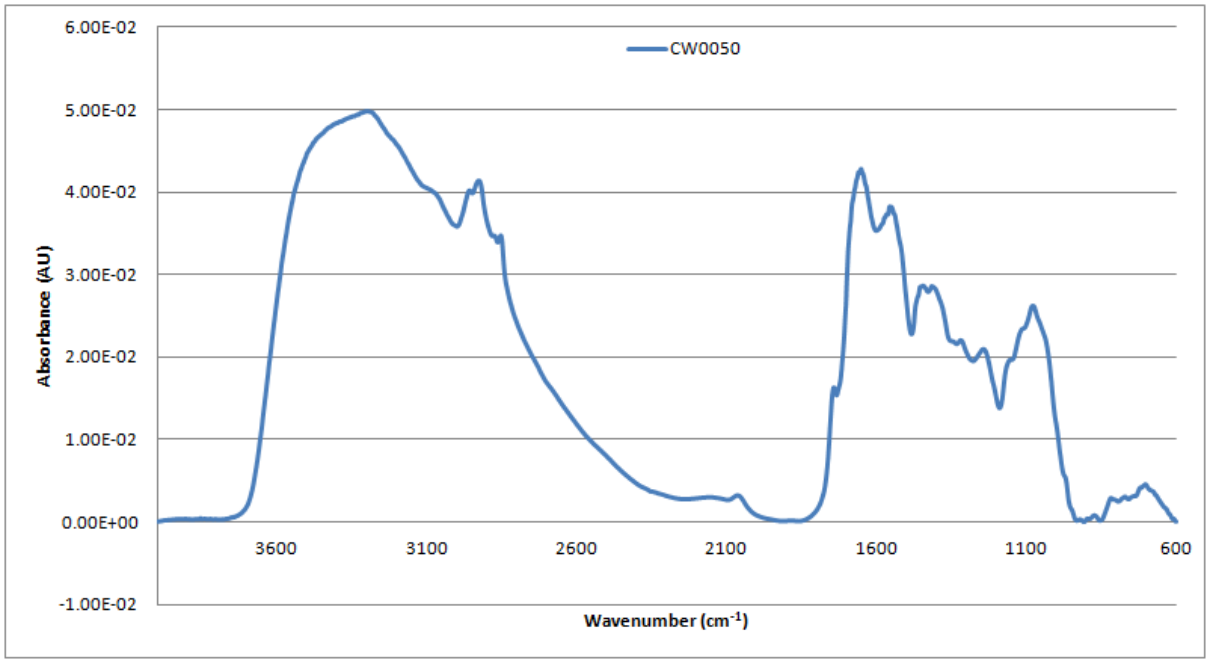


**Figure A2-150: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient CW0050 from 4000-600cm$^{-1}$**

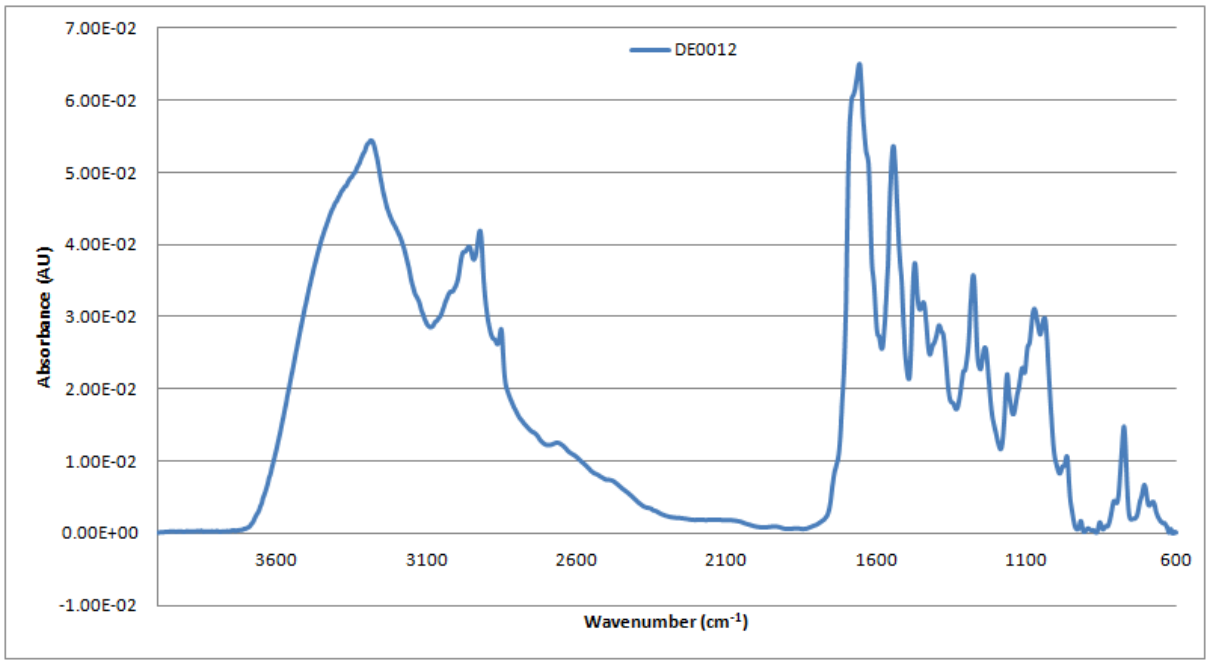


**Figure A2-151: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0012 from 4000-600cm$^{-1}$**

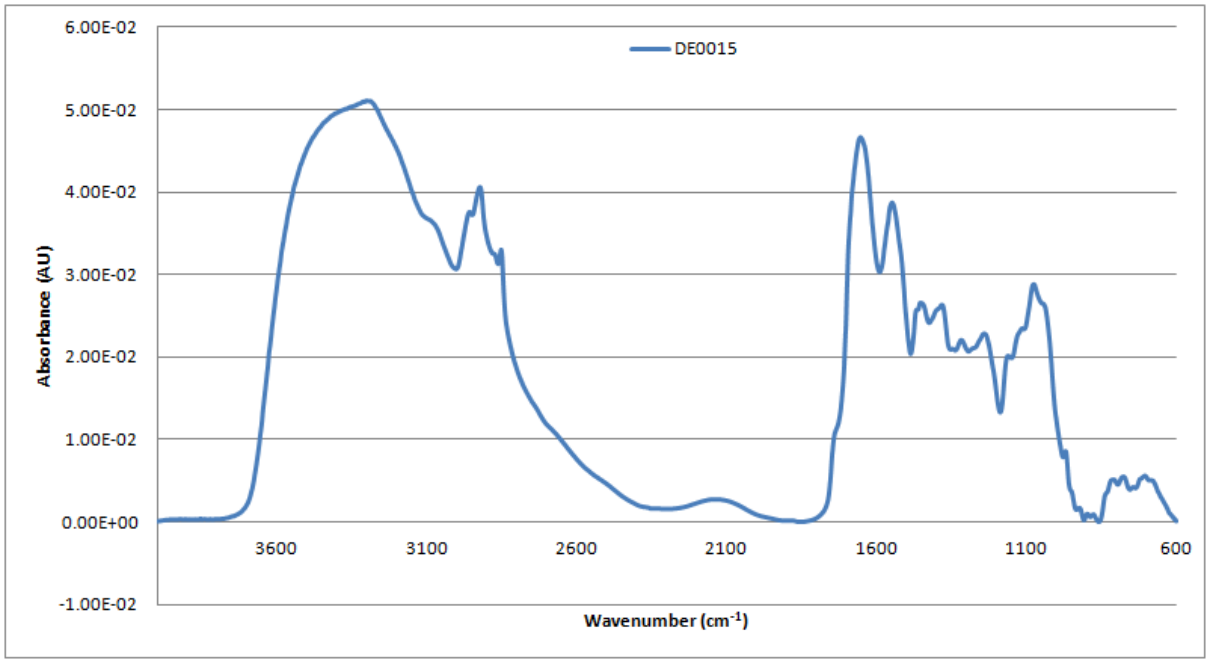


**Figure A2-152: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0015 from 4000-600cm$^{-1}$**

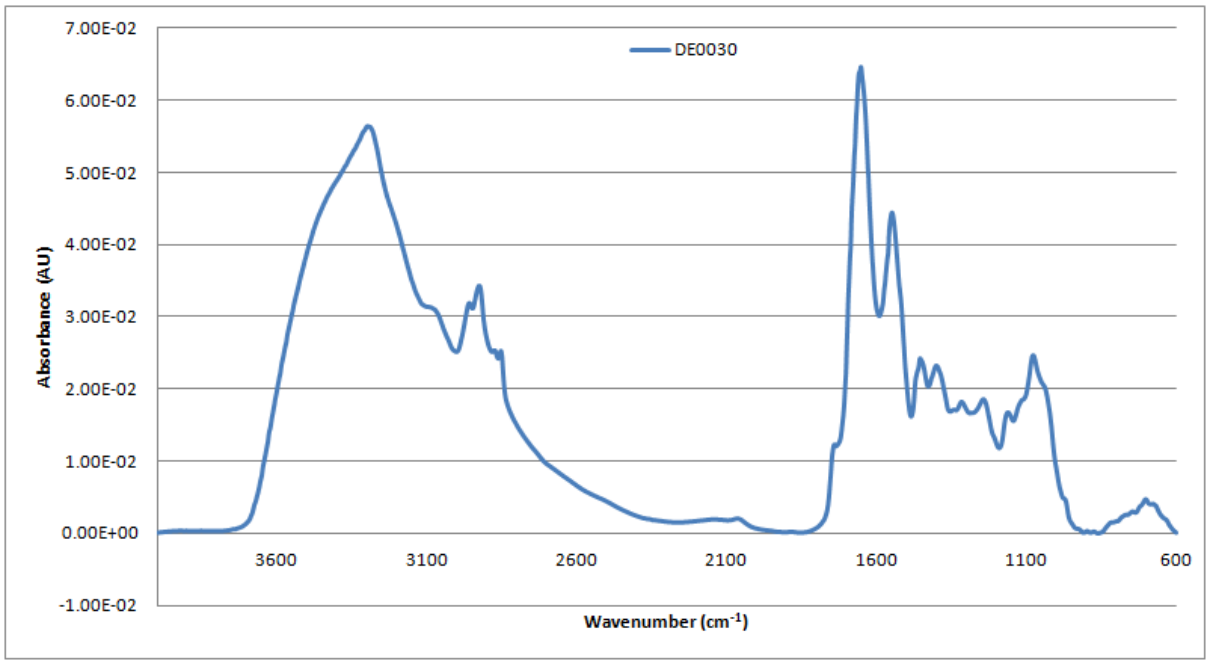


**Figure A2-153: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0030 from 4000-600cm$^{-1}$**

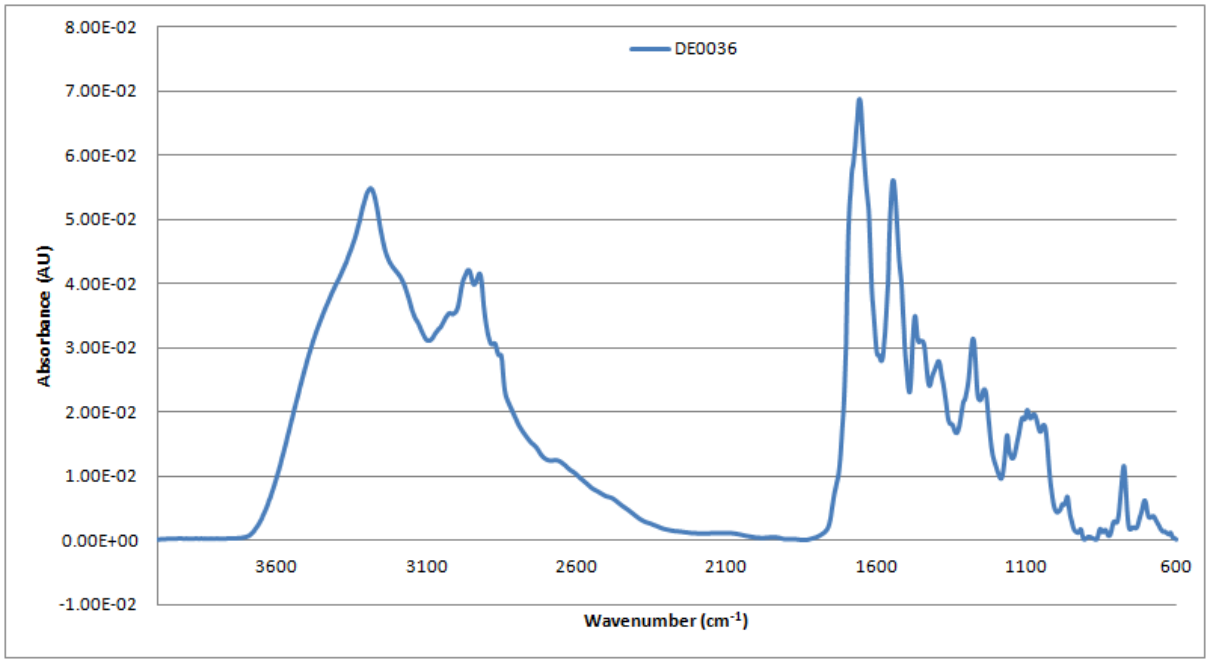


**Figure A2-154: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0036 from 4000-600cm$^{-1}$**

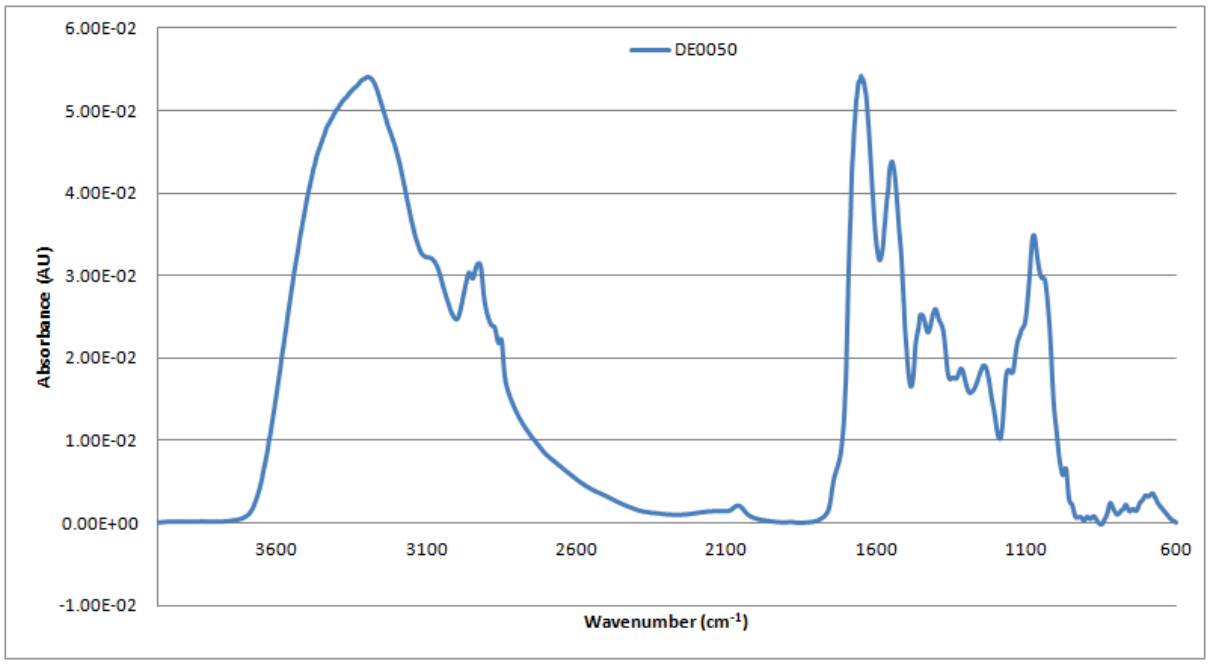


**Figure A2-155: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0050 from 4000-600cm$^{-1}$**

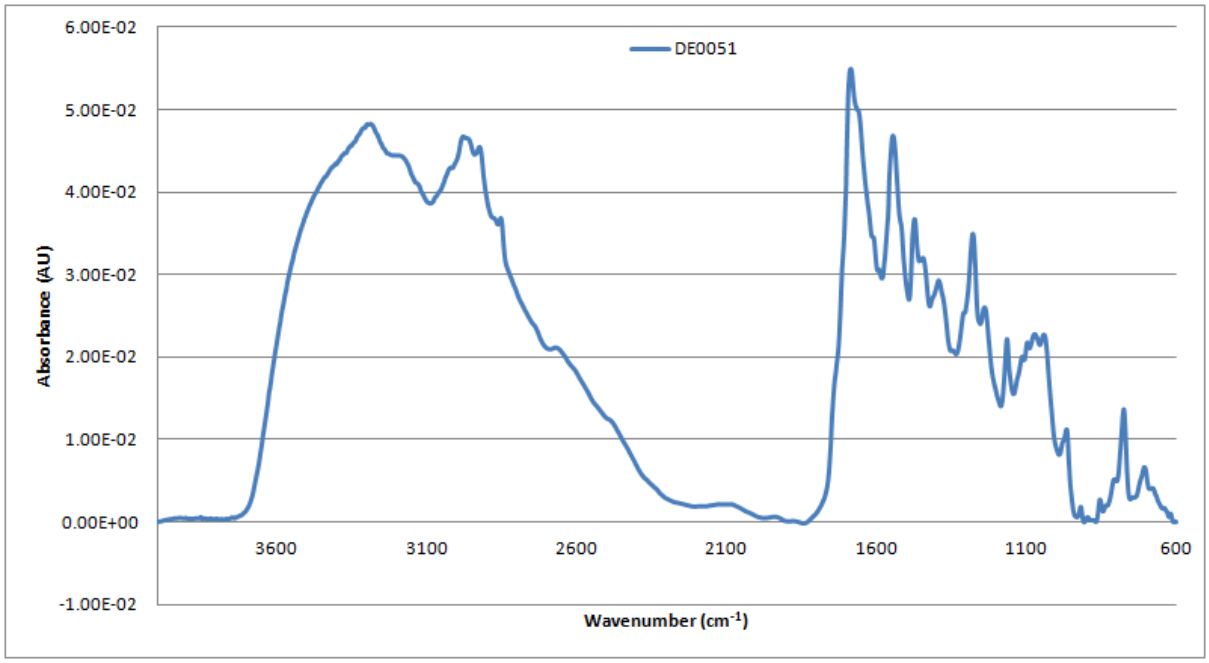


**Figure A2-156: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0051 from 4000-600cm$^{-1}$**

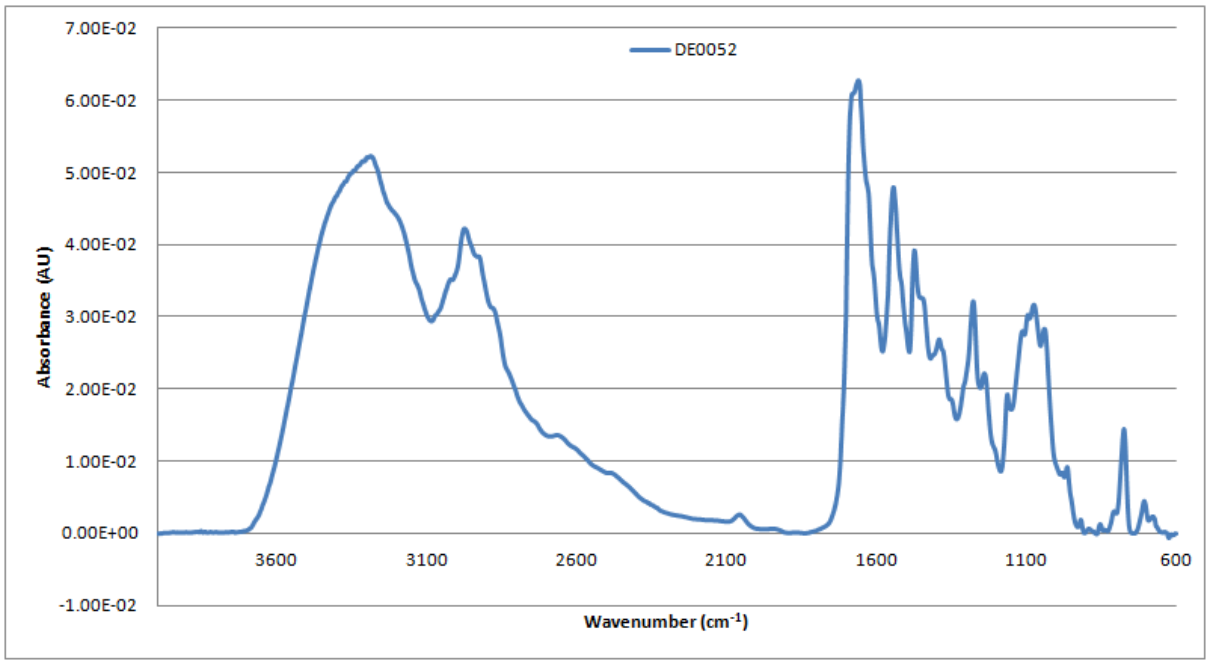


**Figure A2-157: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0052 from 4000-600cm$^{-1}$**

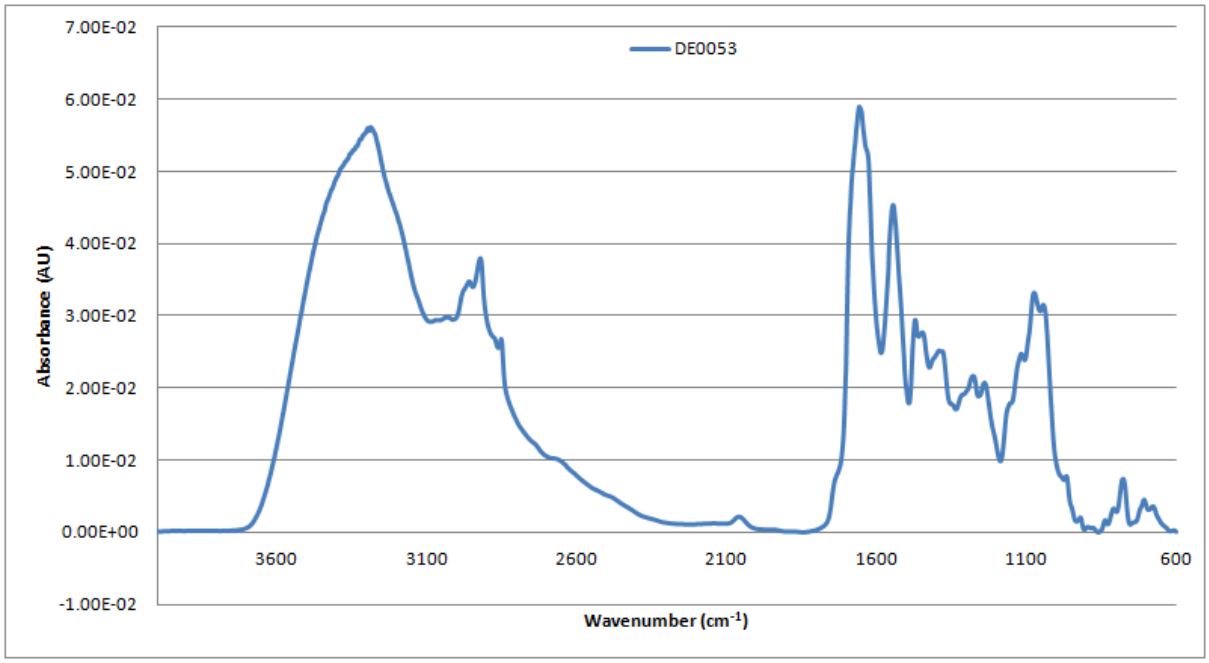


**Figure A2-158: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0053 from 4000-600cm$^{-1}$**

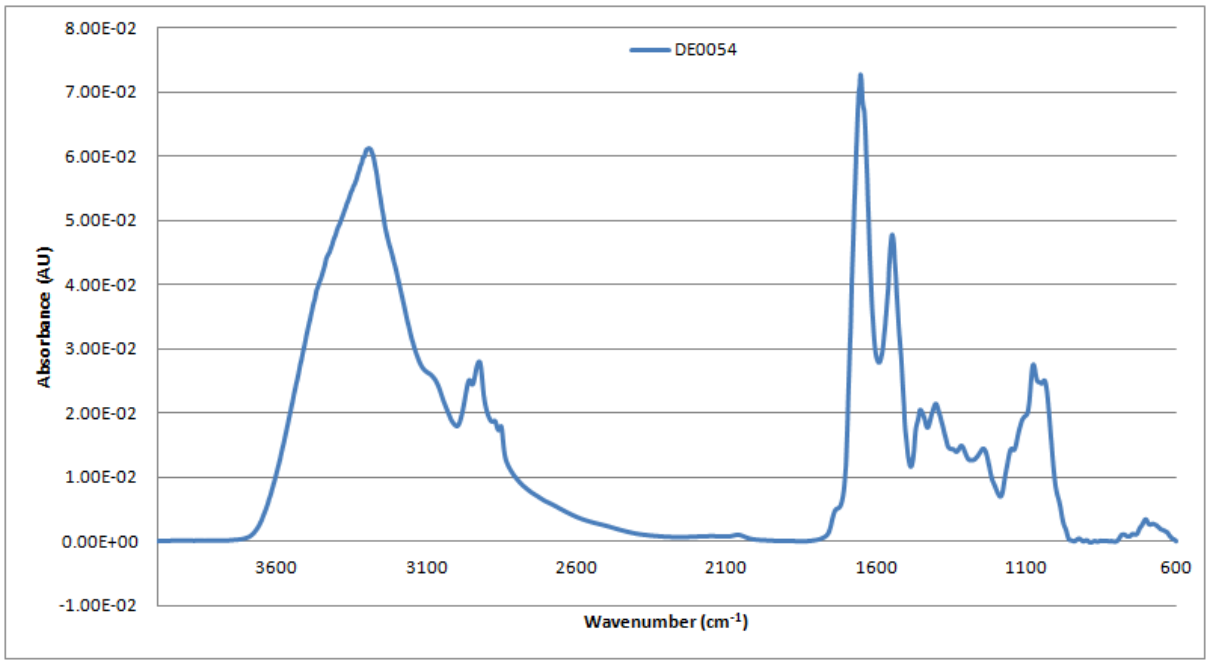


**Figure A2-159: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient DE0054 from 4000-600cm$^{-1}$**

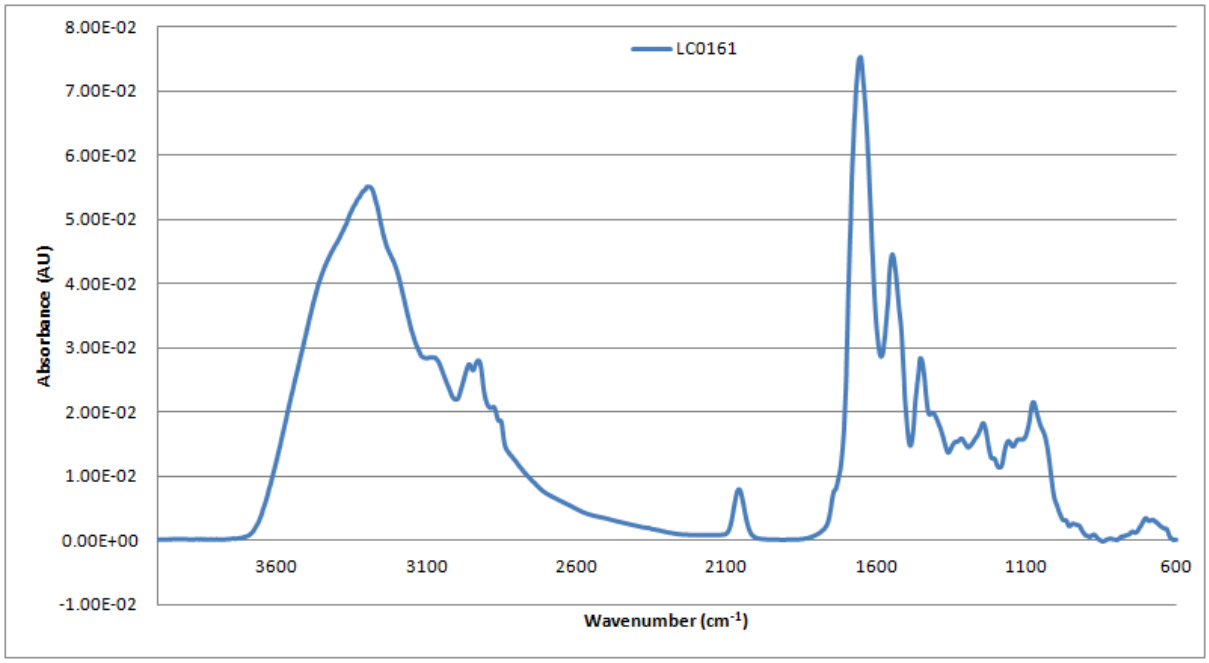


**Figure A2-160: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient LC0161 from 4000-600cm$^{-1}$**

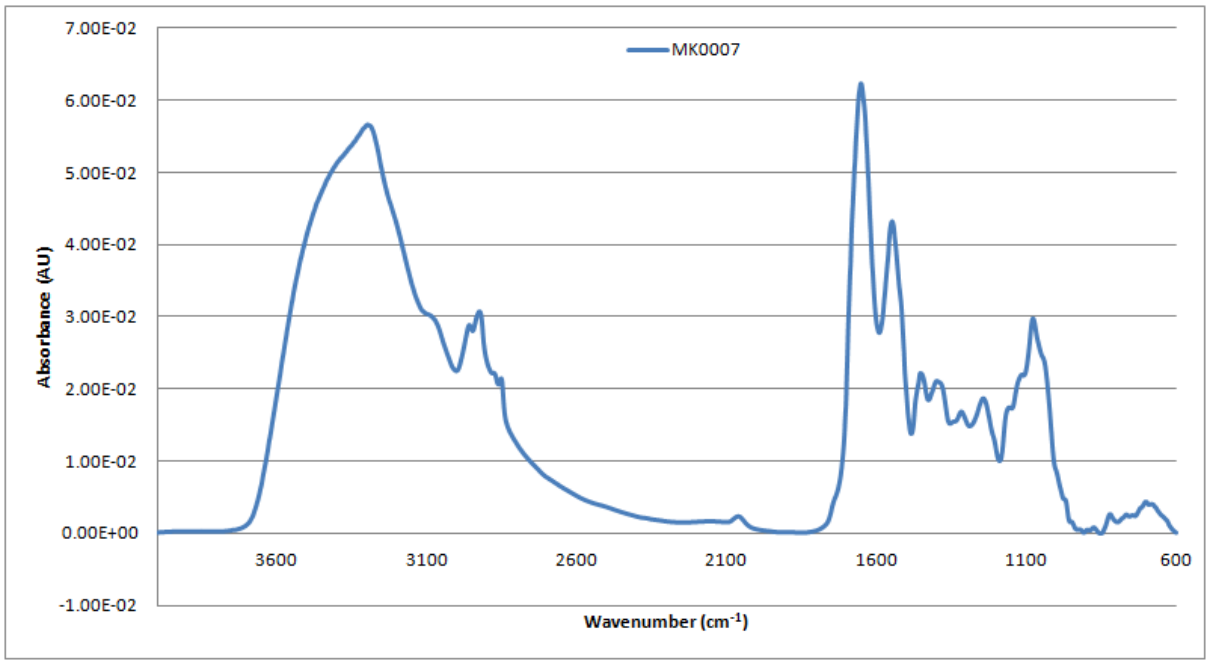


**Figure A2-161: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0007 from 4000-600cm$^{-1}$**

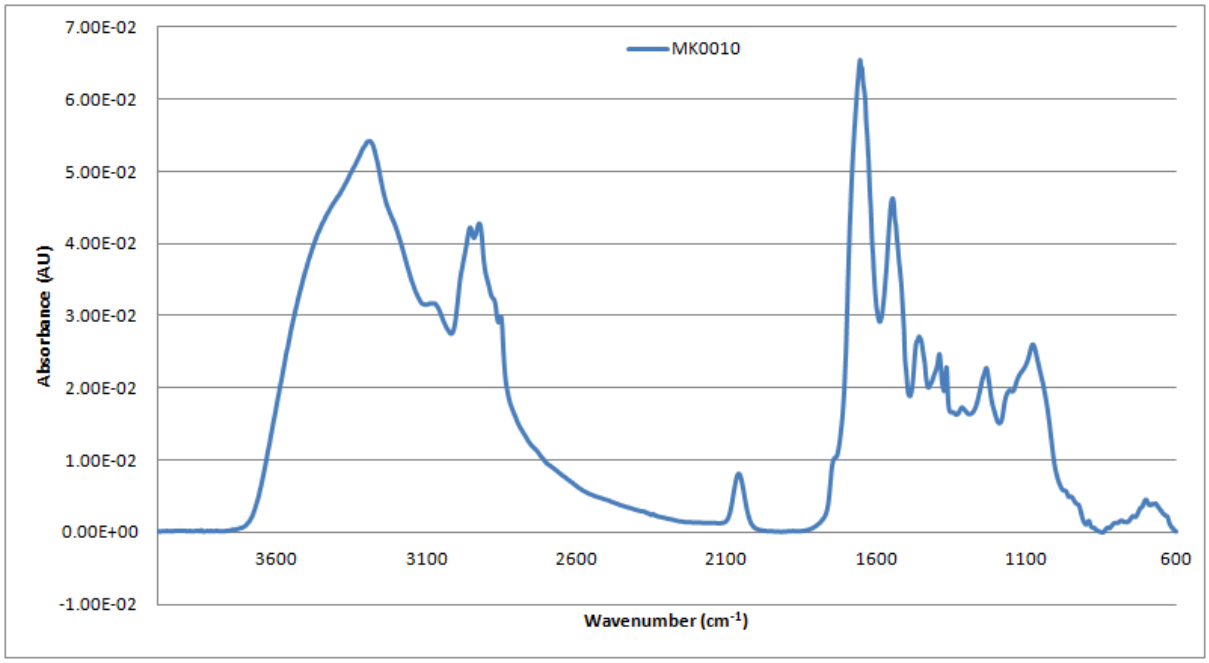


**Figure A2-162: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0010 from 4000-600cm$^{-1}$**

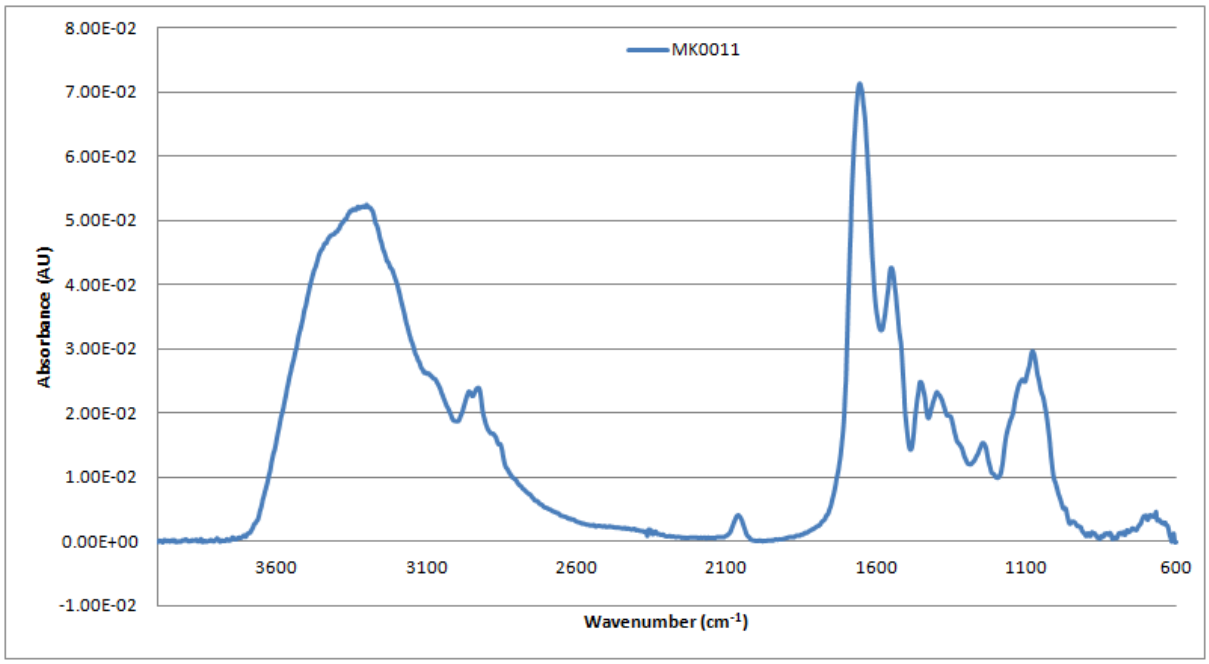


**Figure A2-163: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0011 from 4000-600cm$^{-1}$**

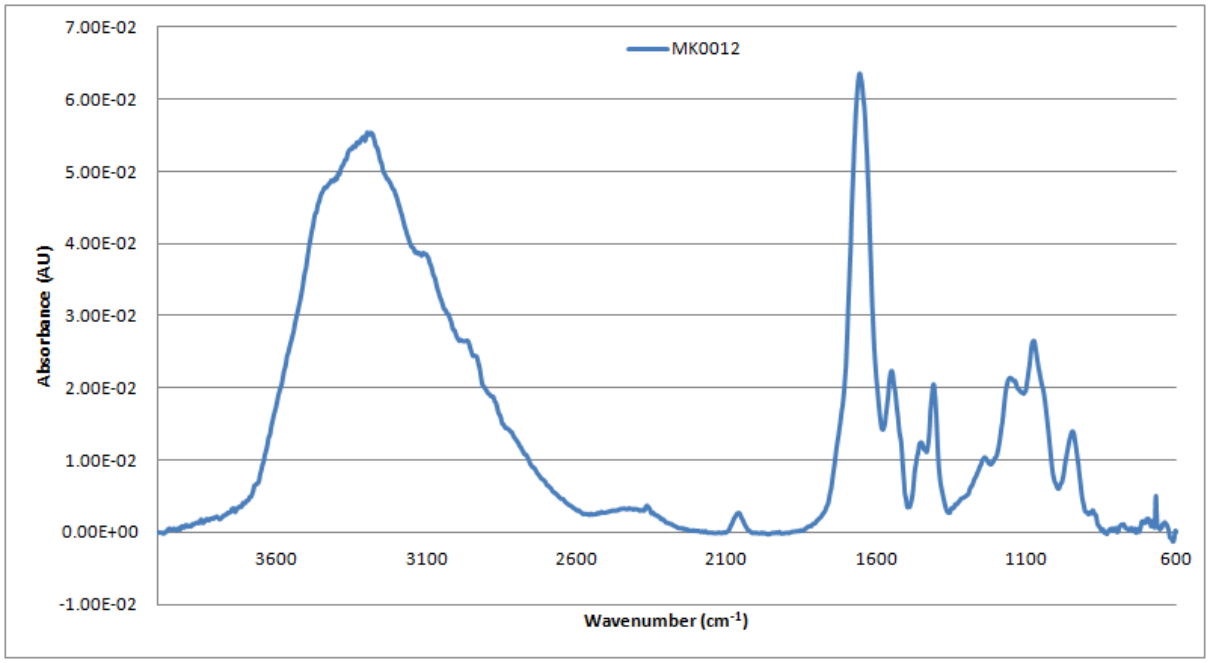


**Figure A2-164: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0012 from 4000-600cm$^{-1}$**

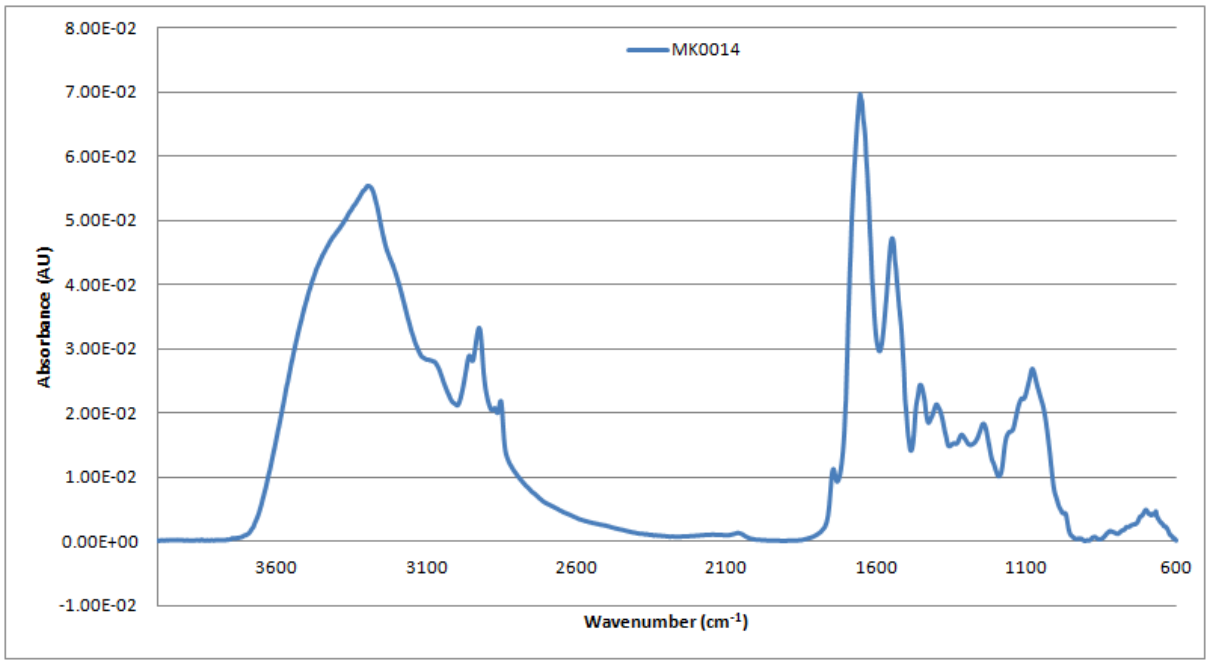


**Figure A2-165: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0014 from 4000-600cm$^{-1}$**

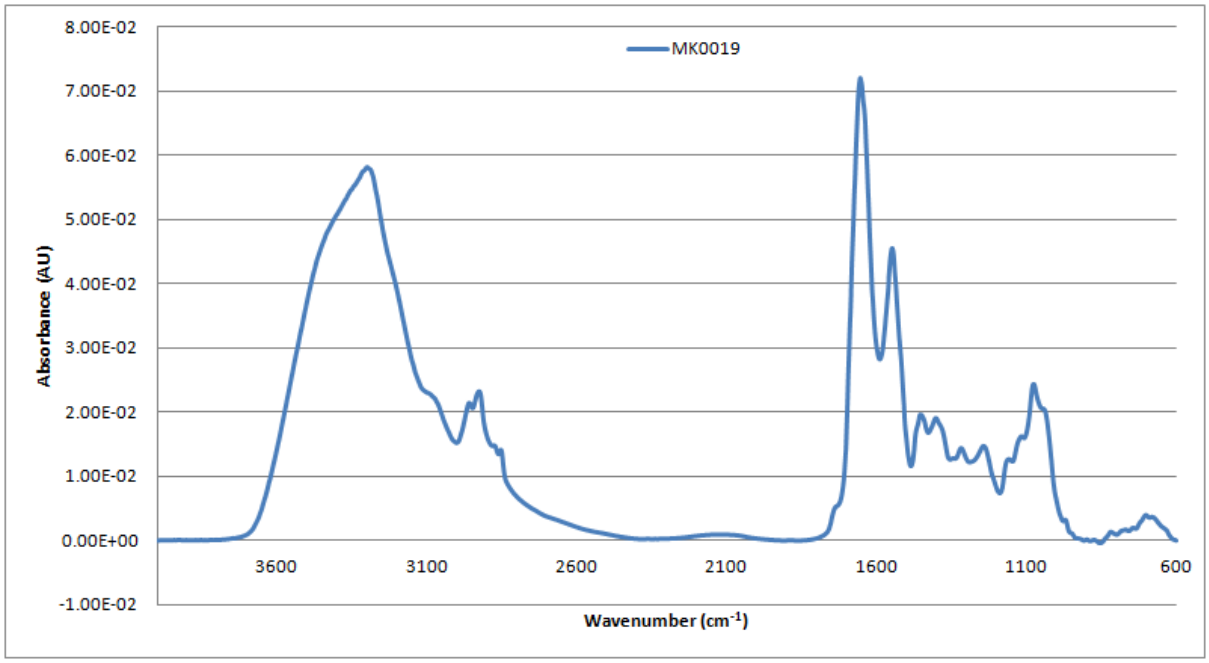


**Figure A2-166: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0019 from 4000-600cm$^{-1}$**

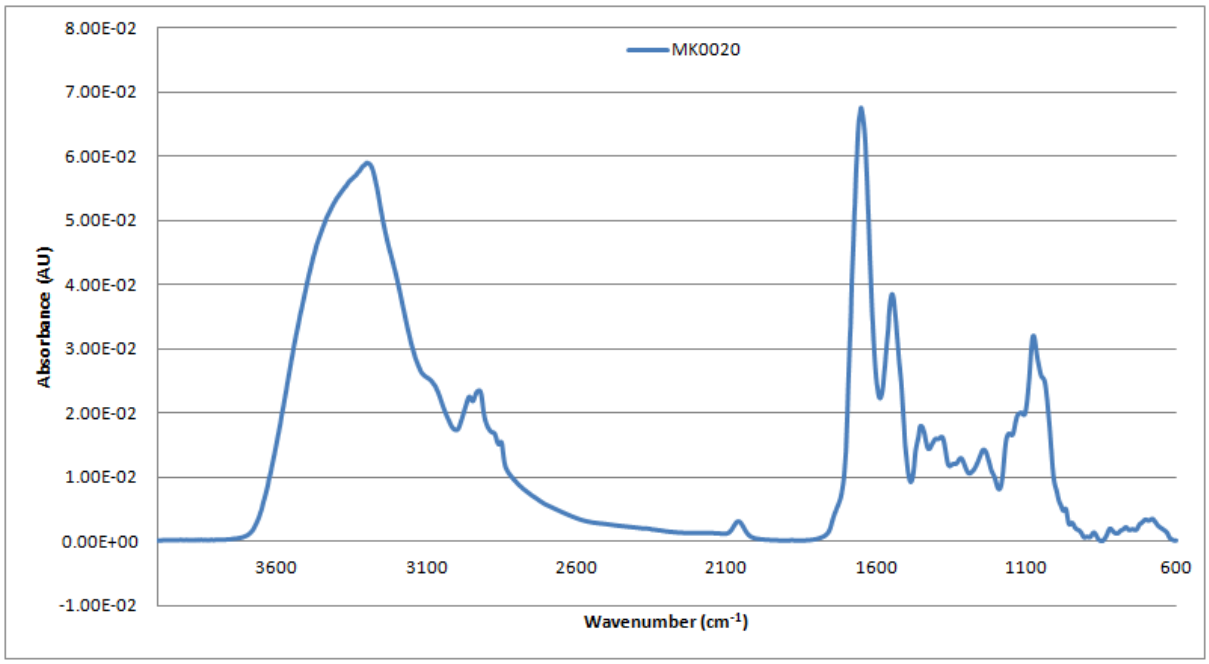


**Figure A2-167: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0020 from 4000-600cm$^{-1}$**

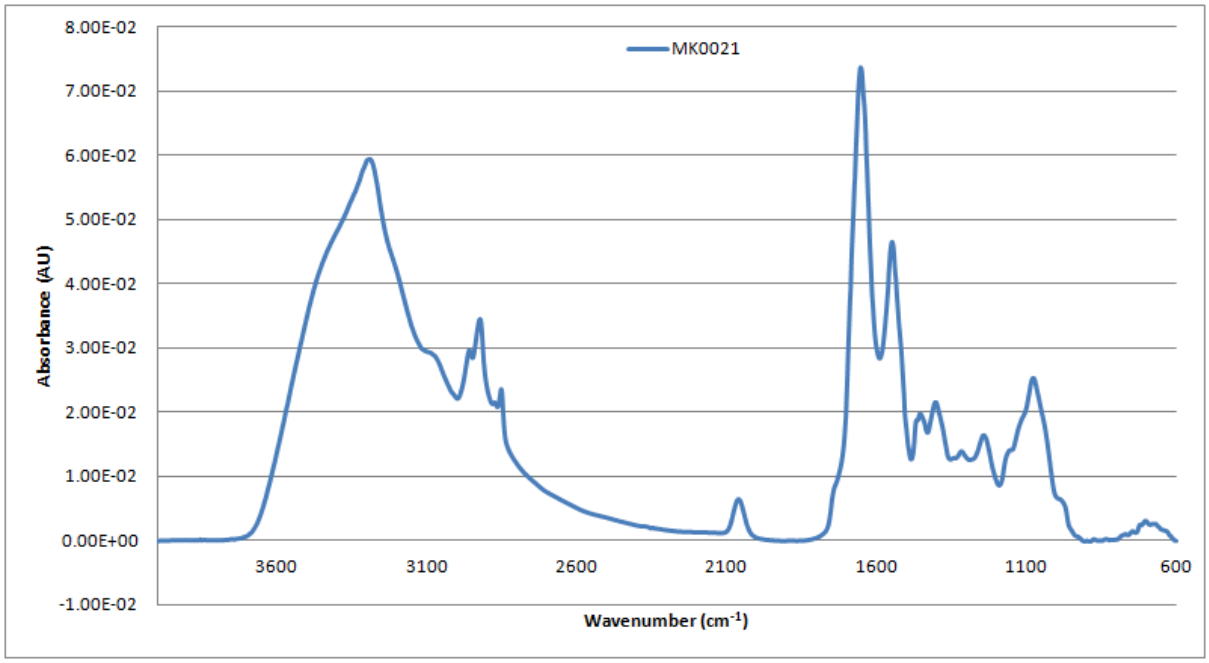


**Figure A2-168: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0020 from 4000-600cm$^{-1}$**

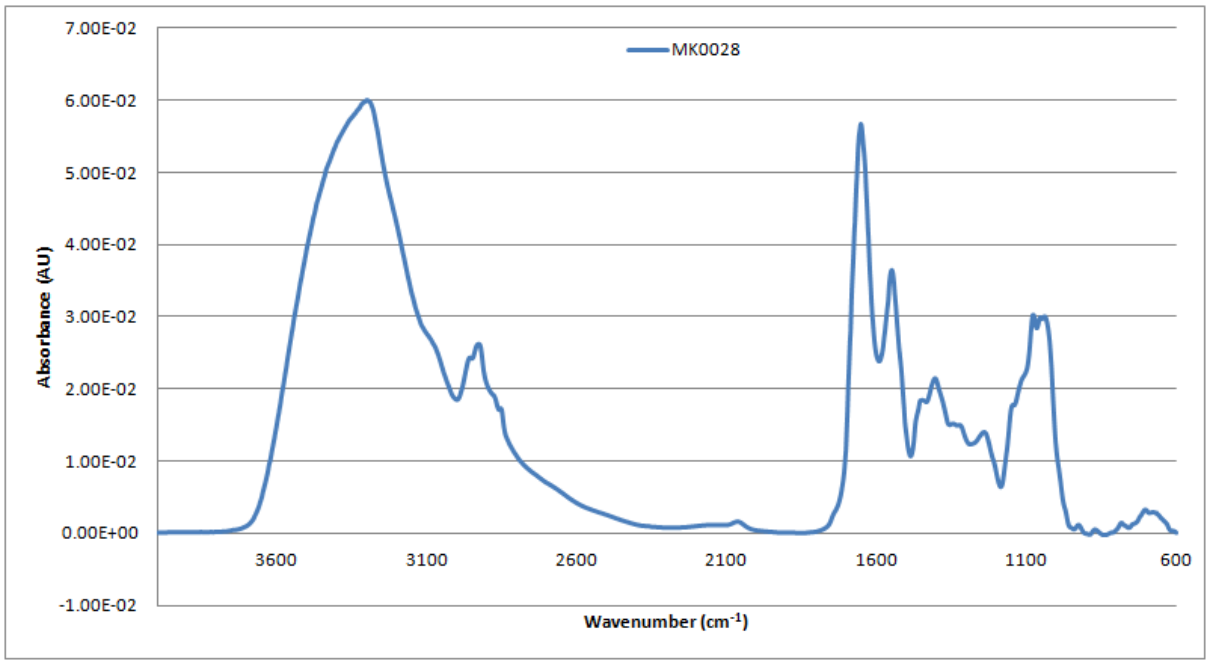


**Figure A2-169: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0028 from 4000-600cm$^{-1}$**

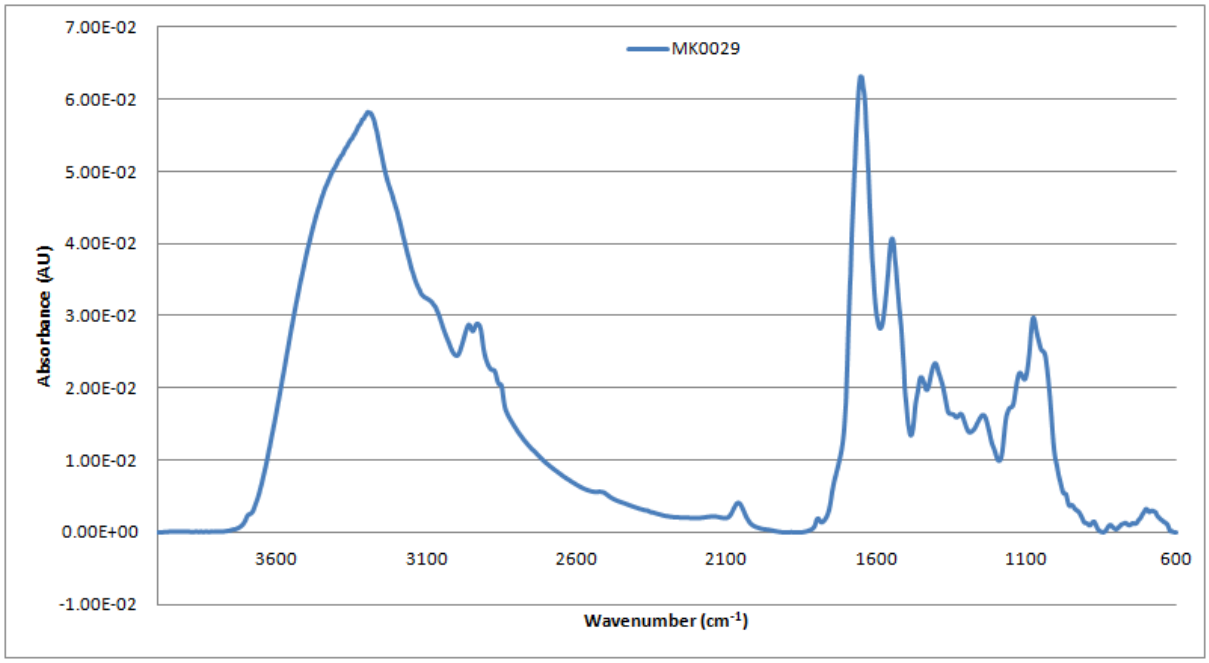


**Figure A2-170: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0029 from 4000-600cm$^{-1}$**

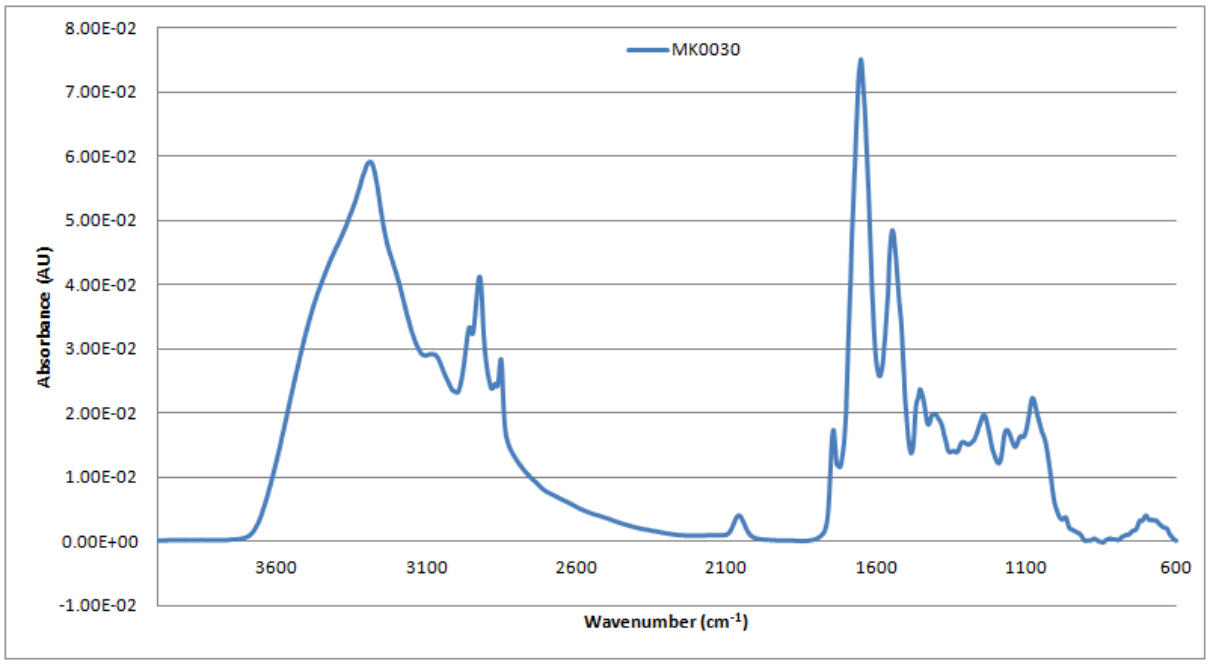


**Figure A2-171: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0030 from 4000-600cm$^{-1}$**

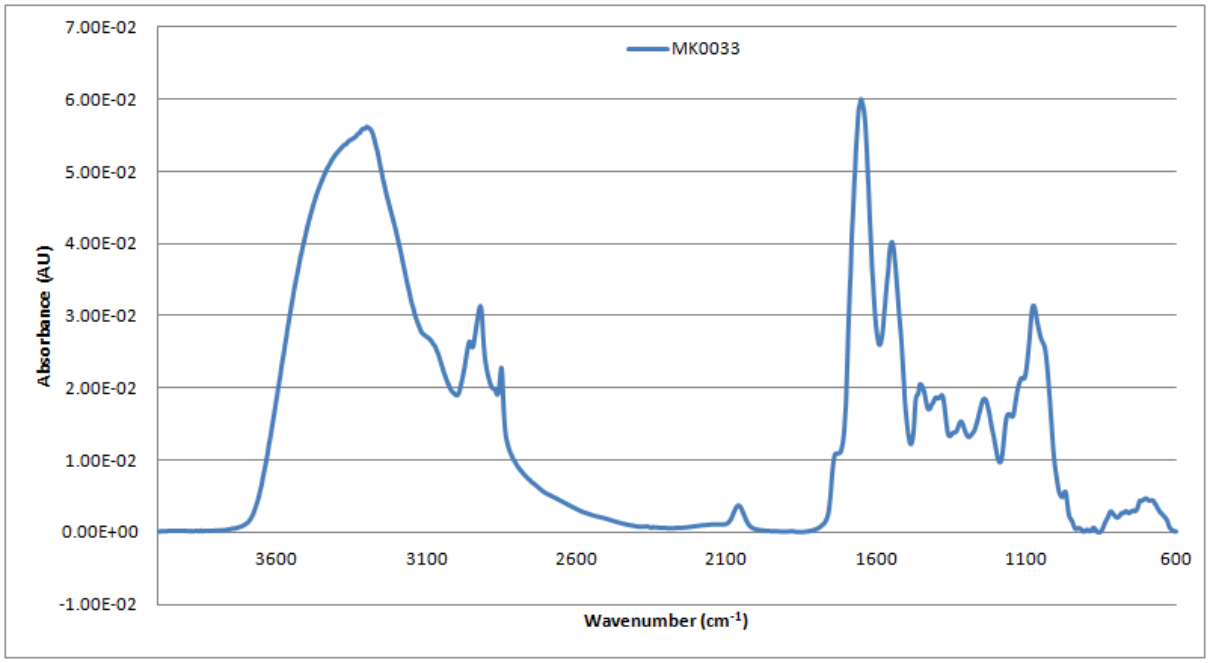


**Figure A2-172: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0033 from 4000-600cm$^{-1}$**

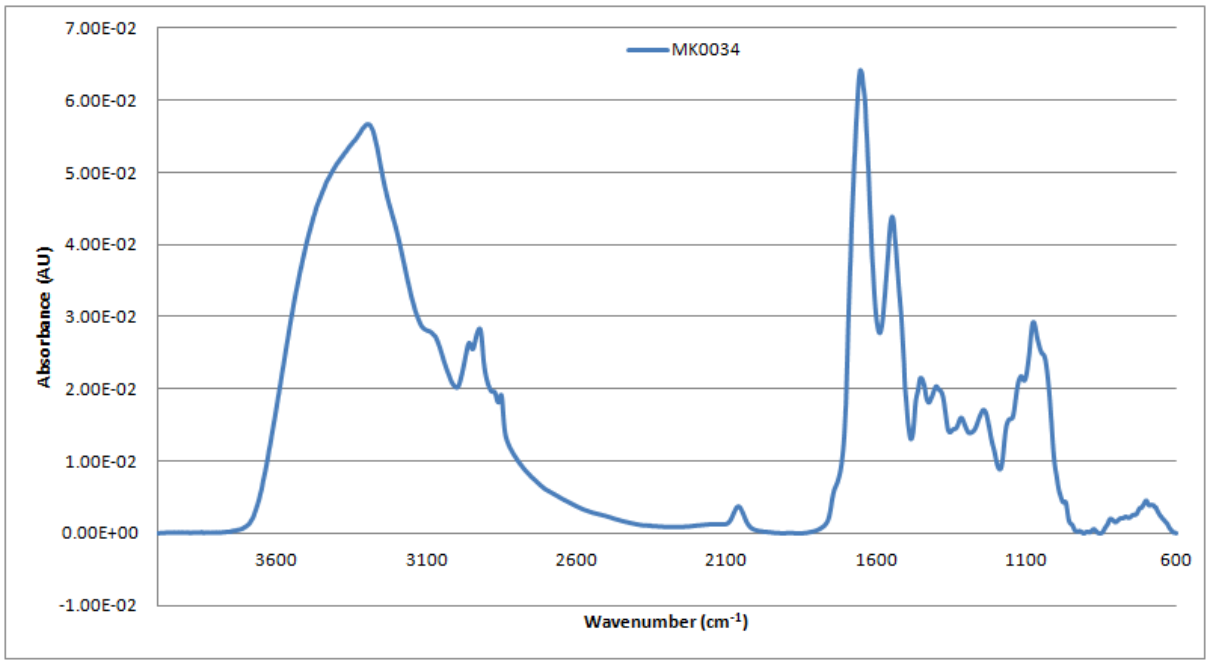


**Figure A2-173: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0034 from 4000-600cm$^{-1}$**

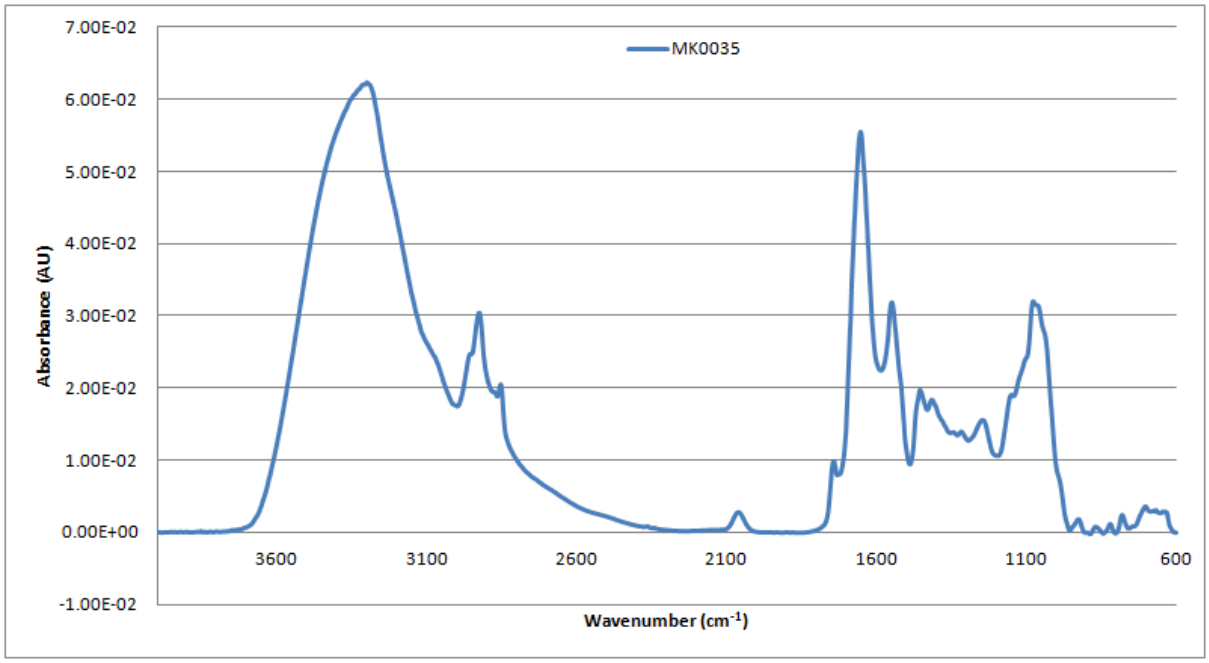


**Figure A2-174: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0035 from 4000-600cm$^{-1}$**

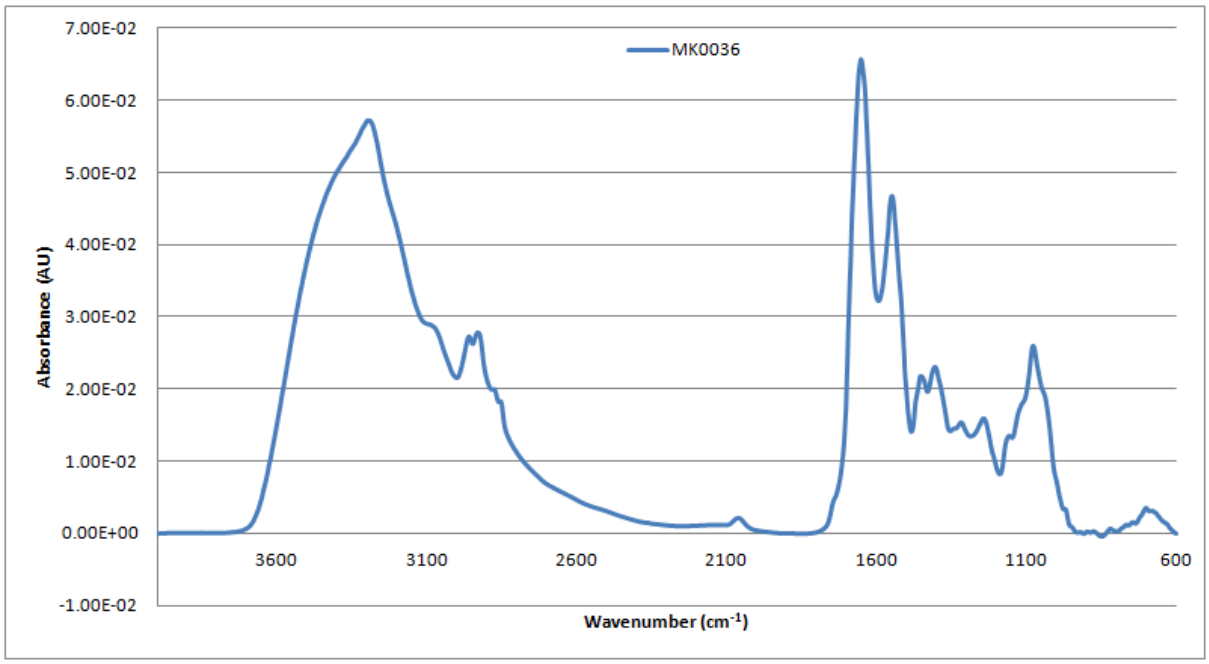


**Figure A2-175: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0036 from 4000-600cm$^{-1}$**

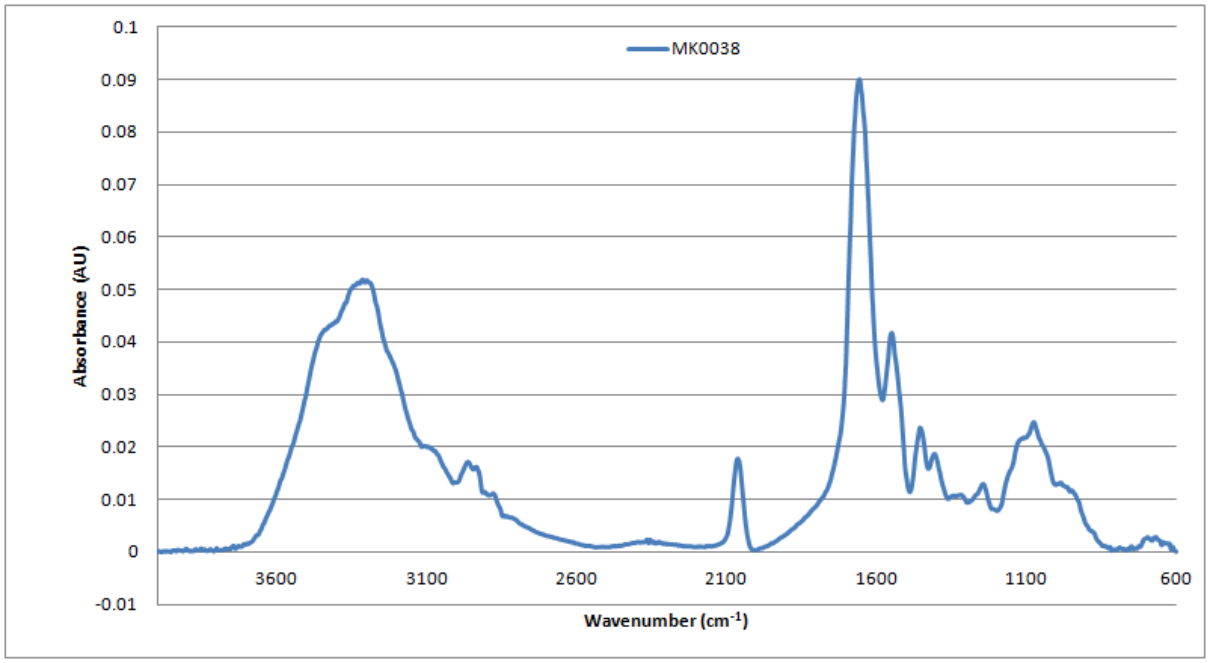


**Figure A2-176: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0038 from 4000-600cm$^{-1}$**

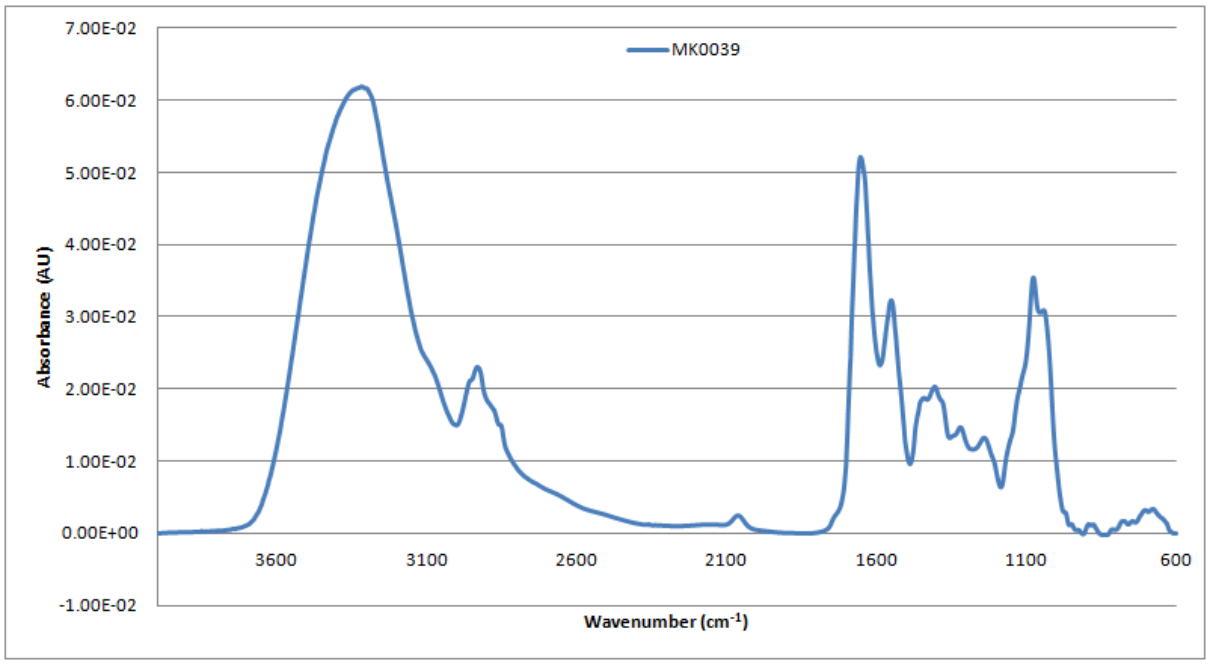


**Figure A2-177: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0039 from 4000-600cm$^{-1}$**

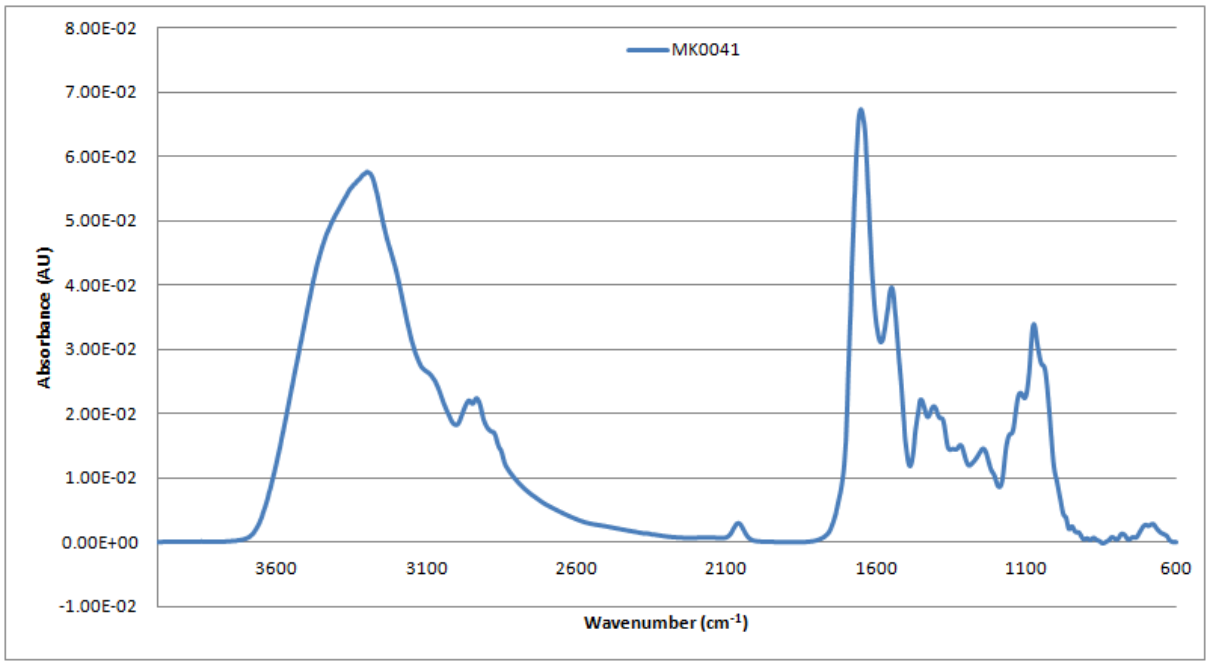


**Figure A2-178: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0041 from 4000-600cm$^{-1}$**

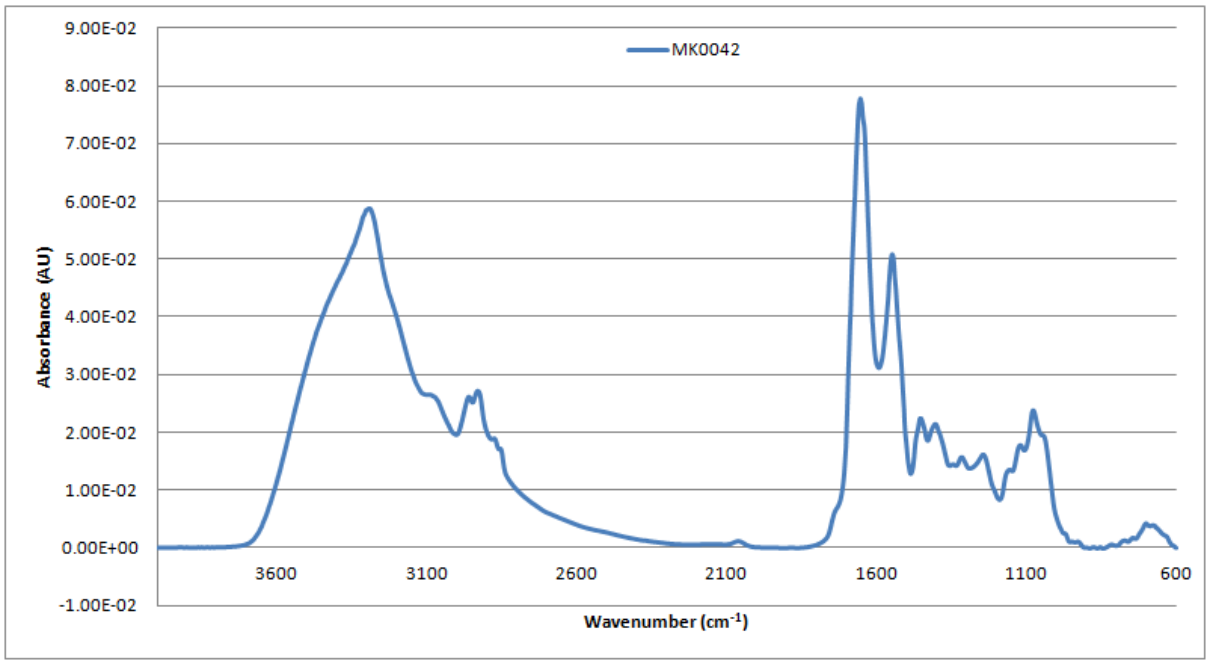


**Figure A2-179: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0042 from 4000-600cm$^{-1}$**

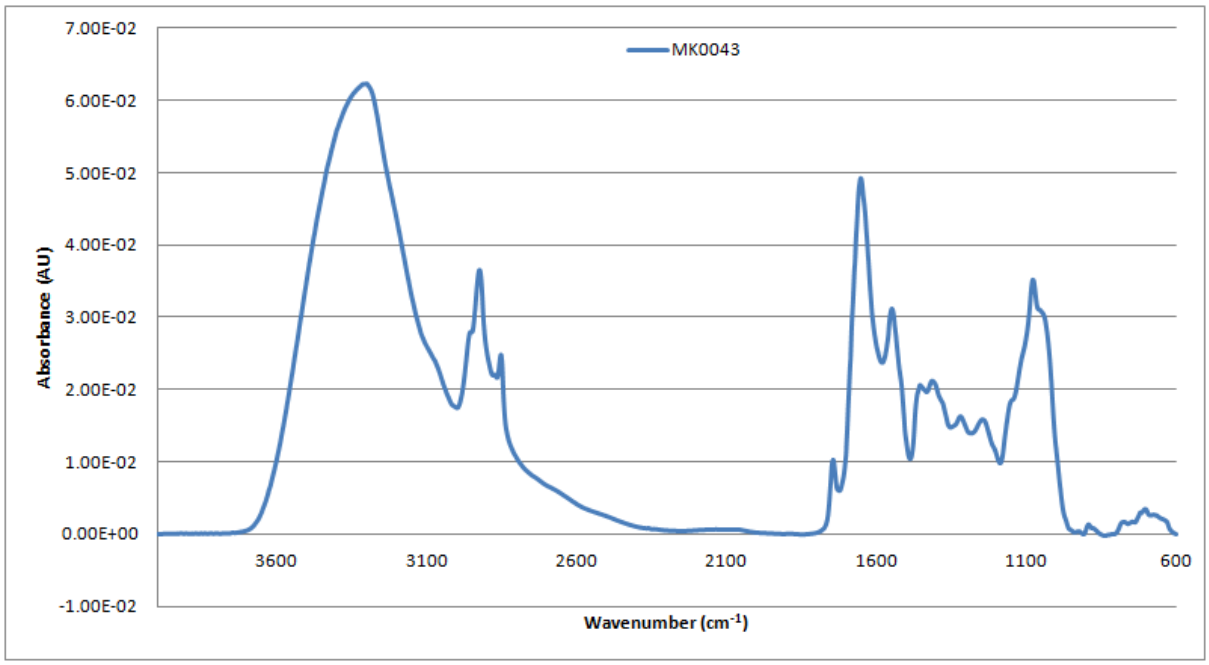


**Figure A2-180: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0043 from 4000-600cm$^{-1}$**

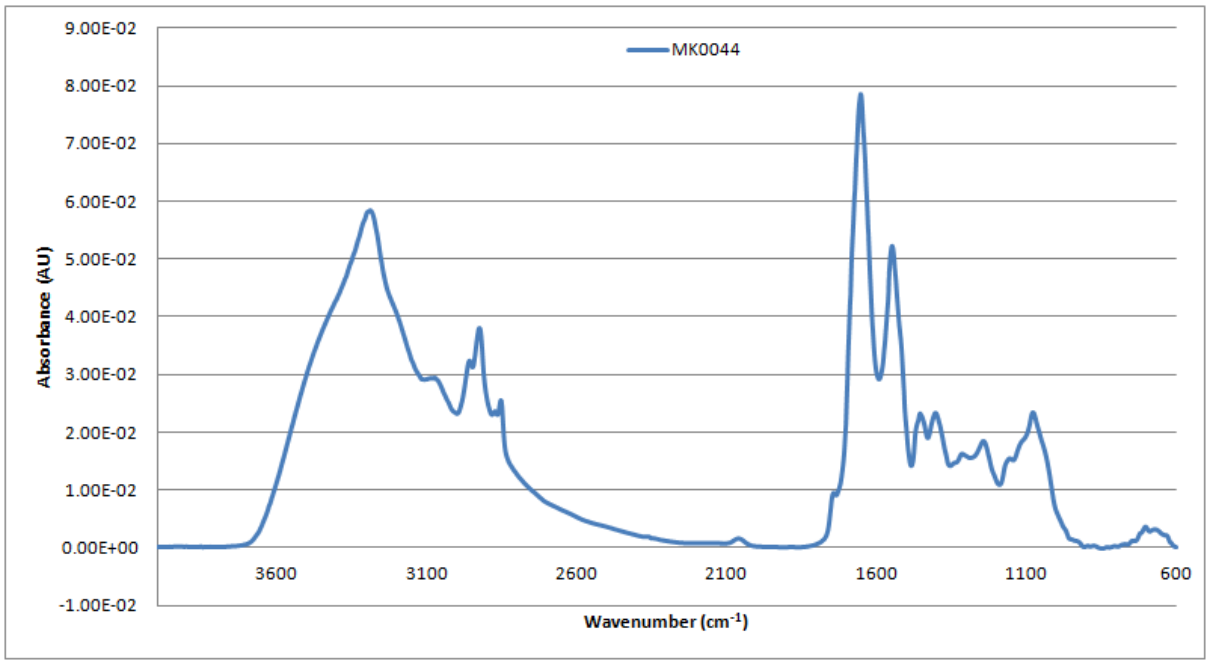


**Figure A2-181: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0044 from 4000-600cm$^{-1}$**

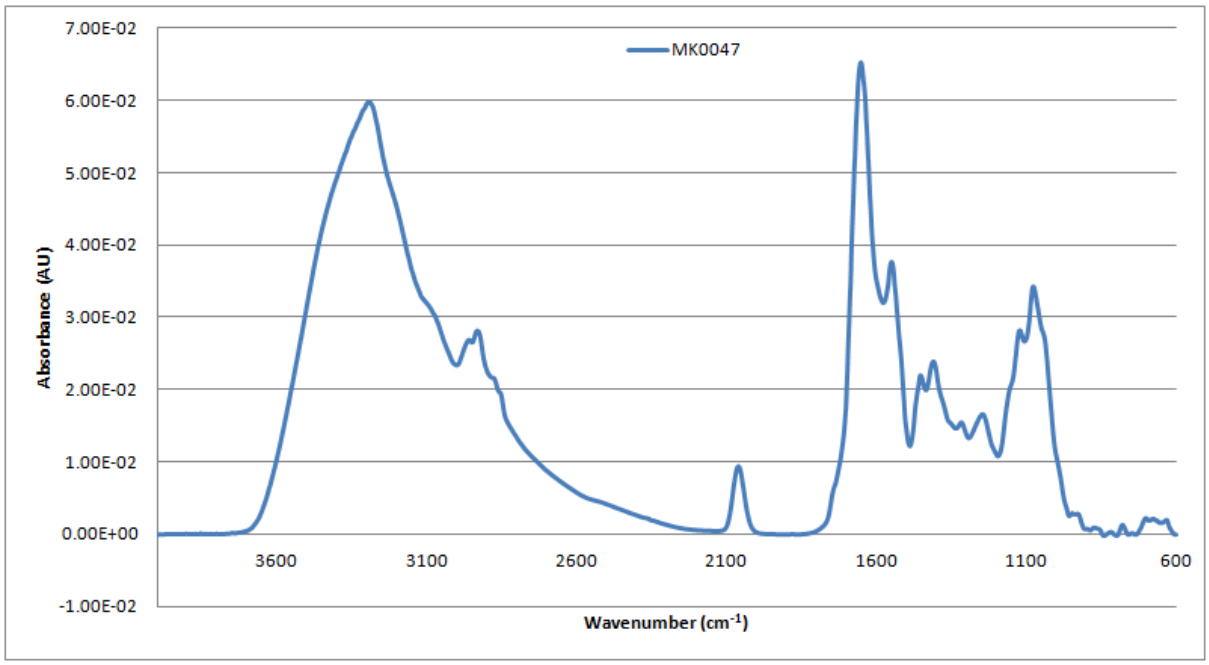


**Figure A2-182: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0047 from 4000-600cm$^{-1}$**

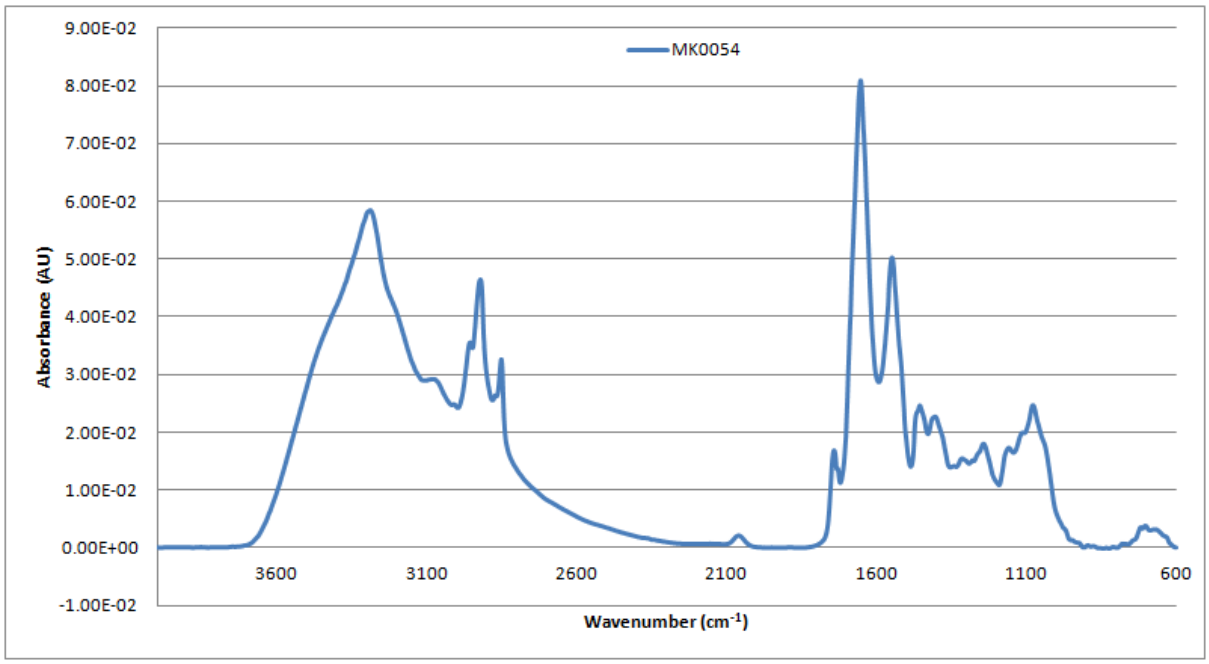


**Figure A2-183: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0054 from 4000-600cm$^{-1}$**

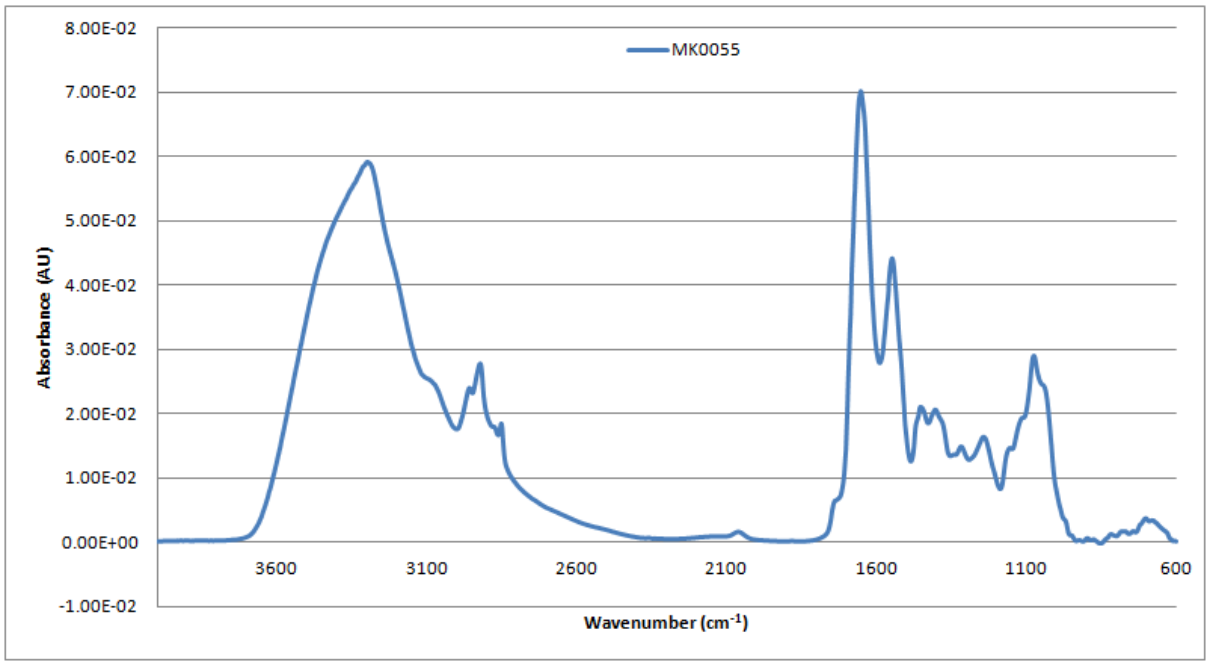


**Figure A2-184: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0055 from 4000-600cm$^{-1}$**

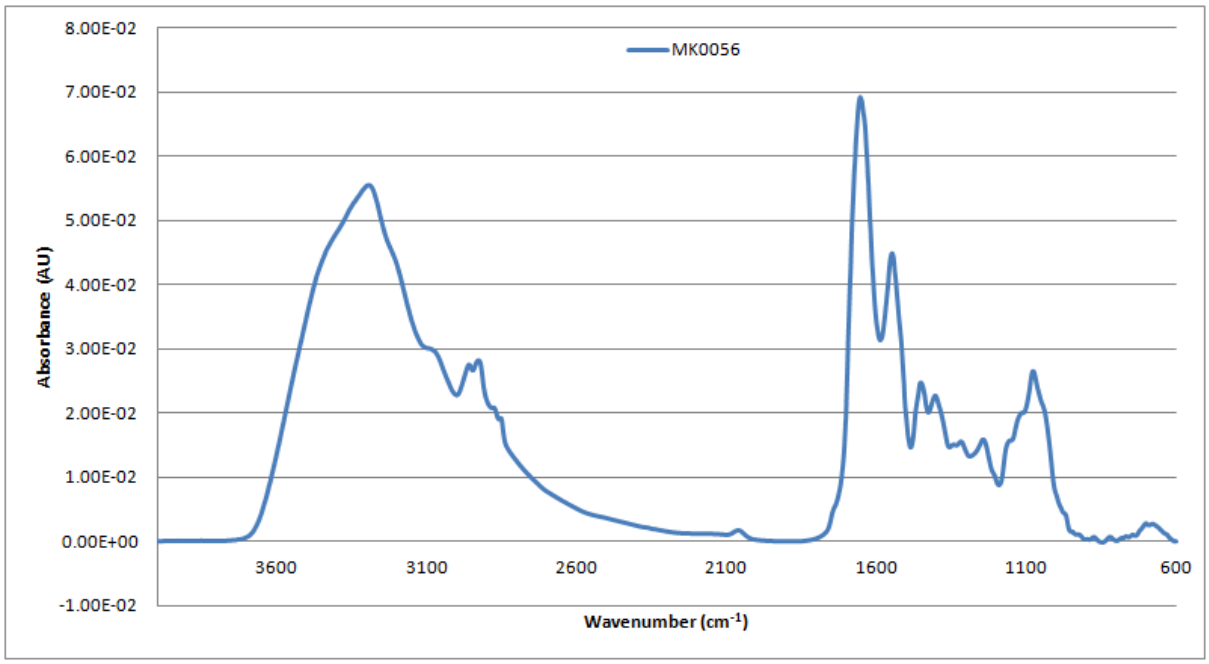


**Figure A2-185: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0056 from 4000-600cm$^{-1}$**

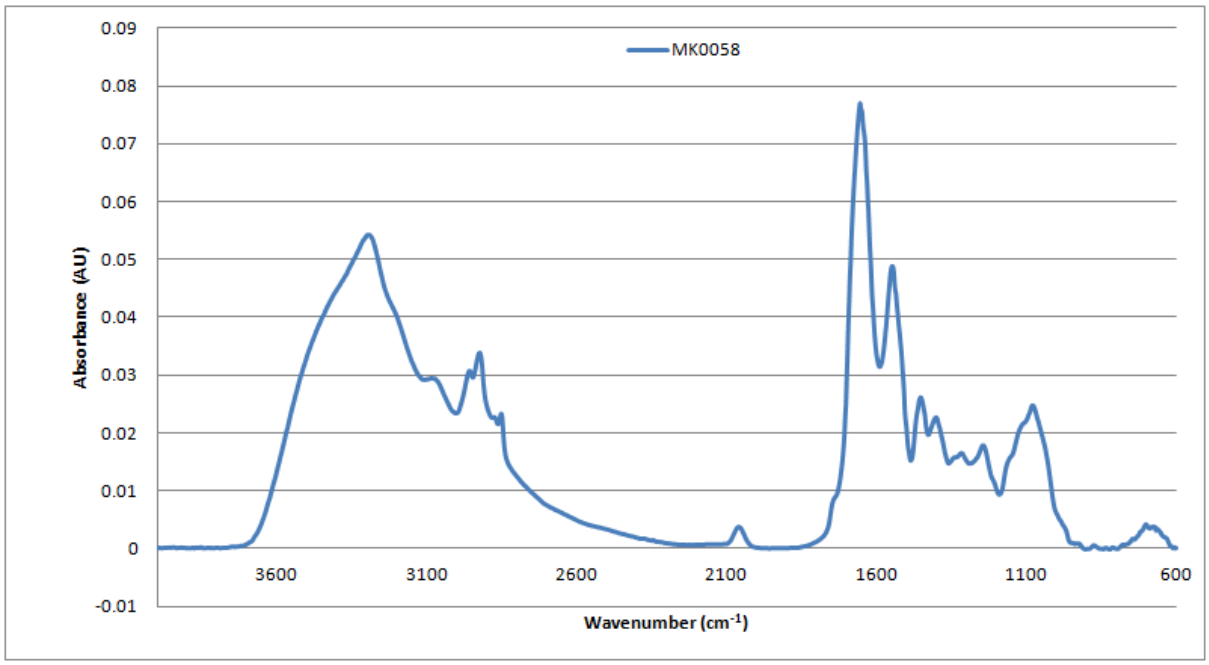


**Figure A2-186: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0058 from 4000-600cm$^{-1}$**

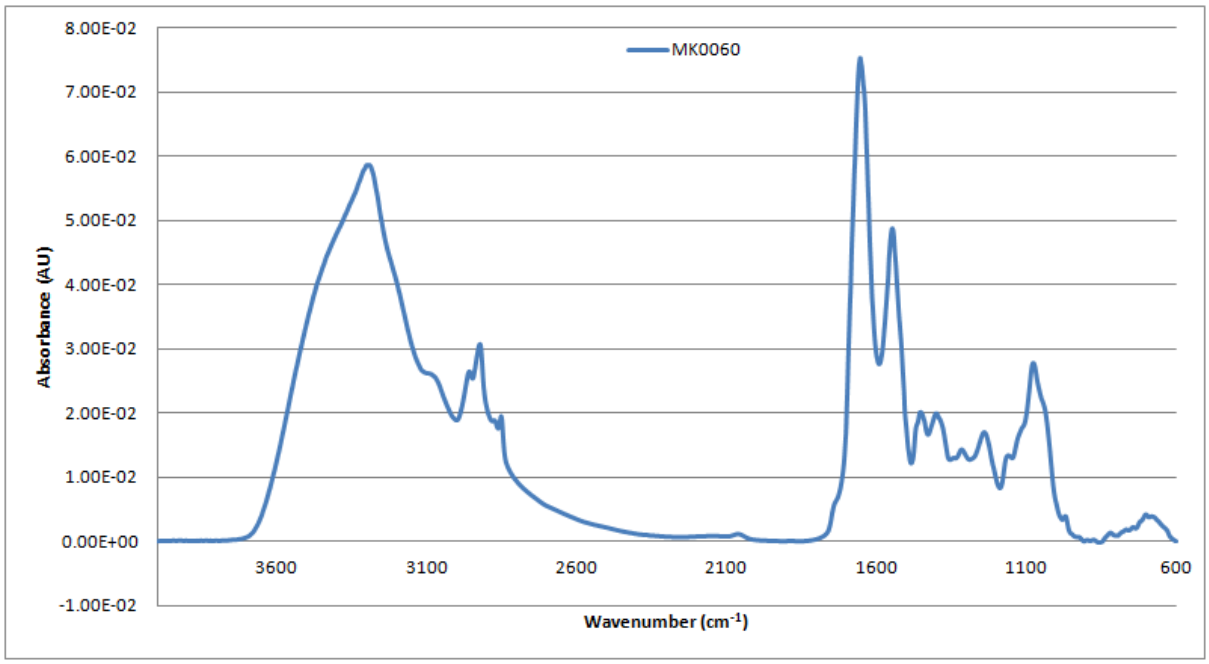


**Figure A2-187: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0060 from 4000-600cm$^{-1}$**

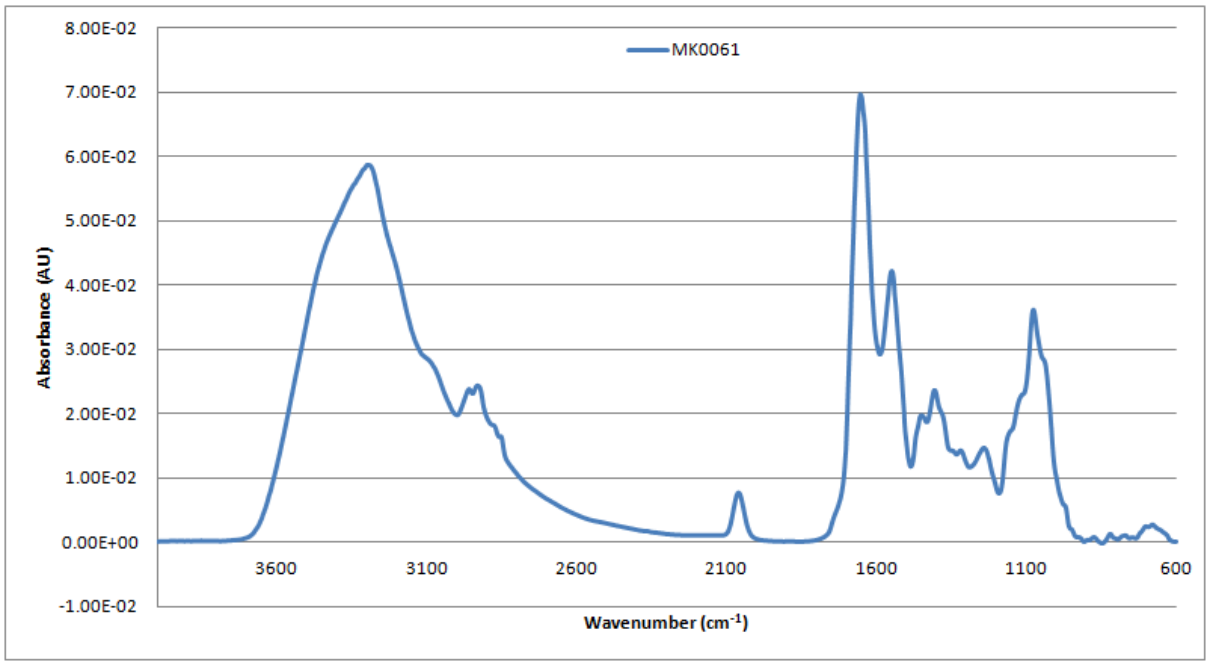


**Figure A2-188: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0061 from 4000-600cm$^{-1}$**

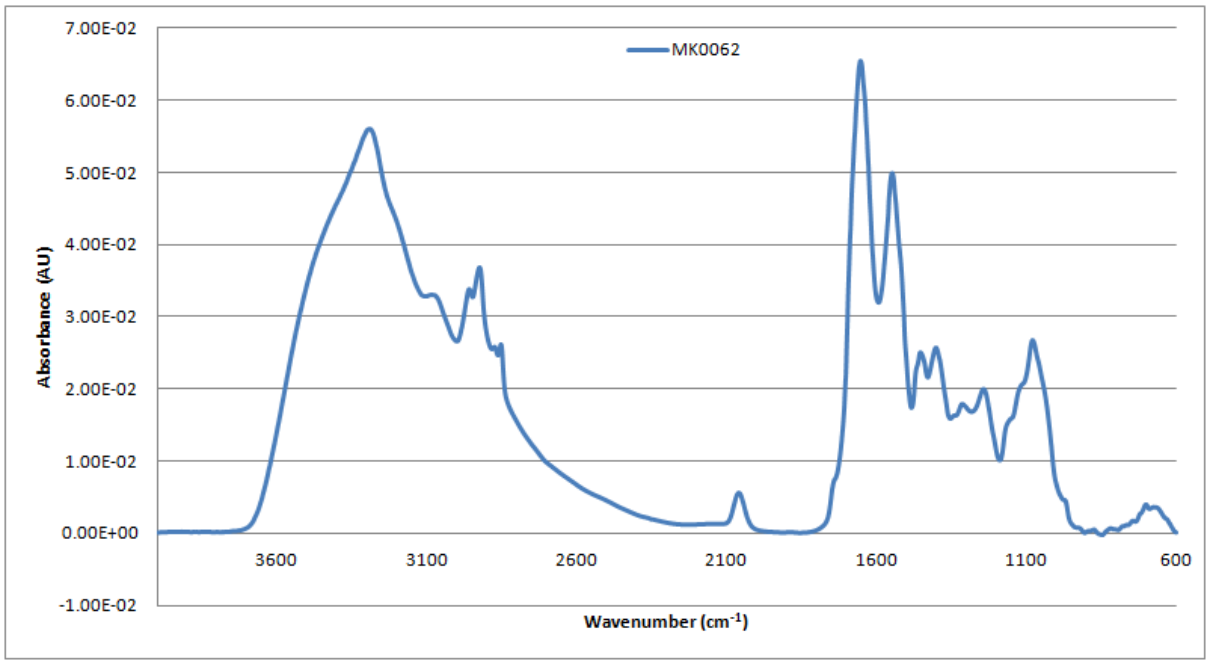


**Figure A2-189: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0062 from 4000-600cm$^{-1}$**

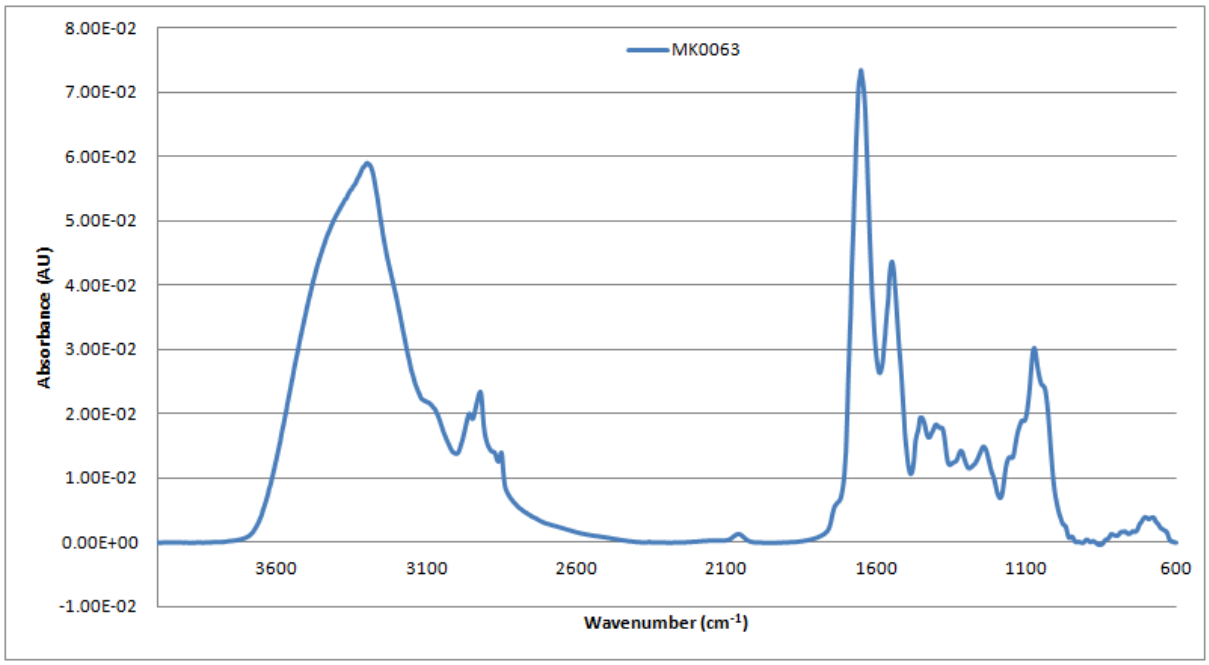


**Figure A2-190: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0063 from 4000-600cm$^{-1}$**

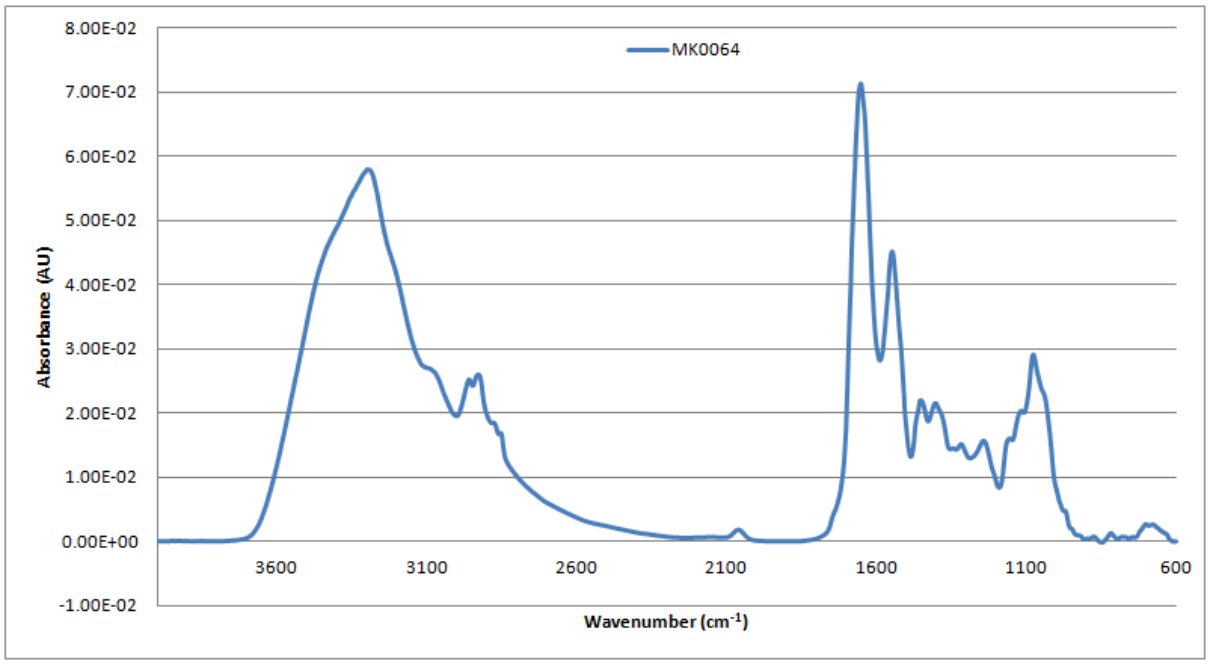


**Figure A2-191: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0064 from 4000-600cm$^{-1}$**

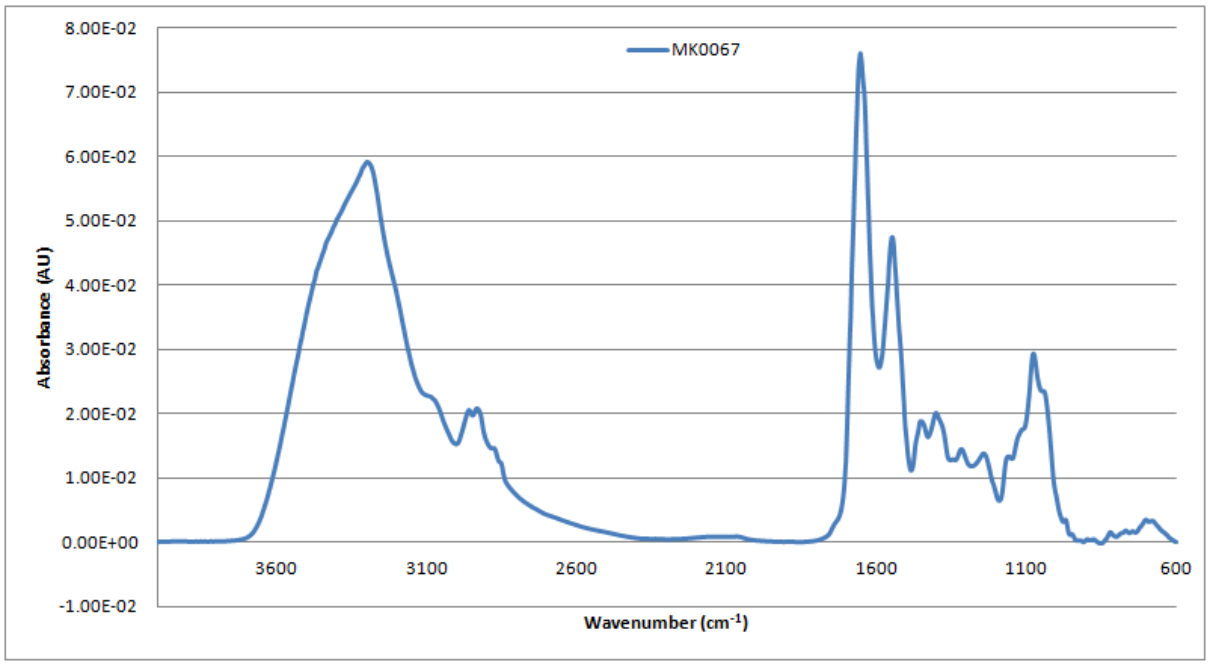


**Figure A2-192: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0067 from 4000-600cm$^{-1}$**

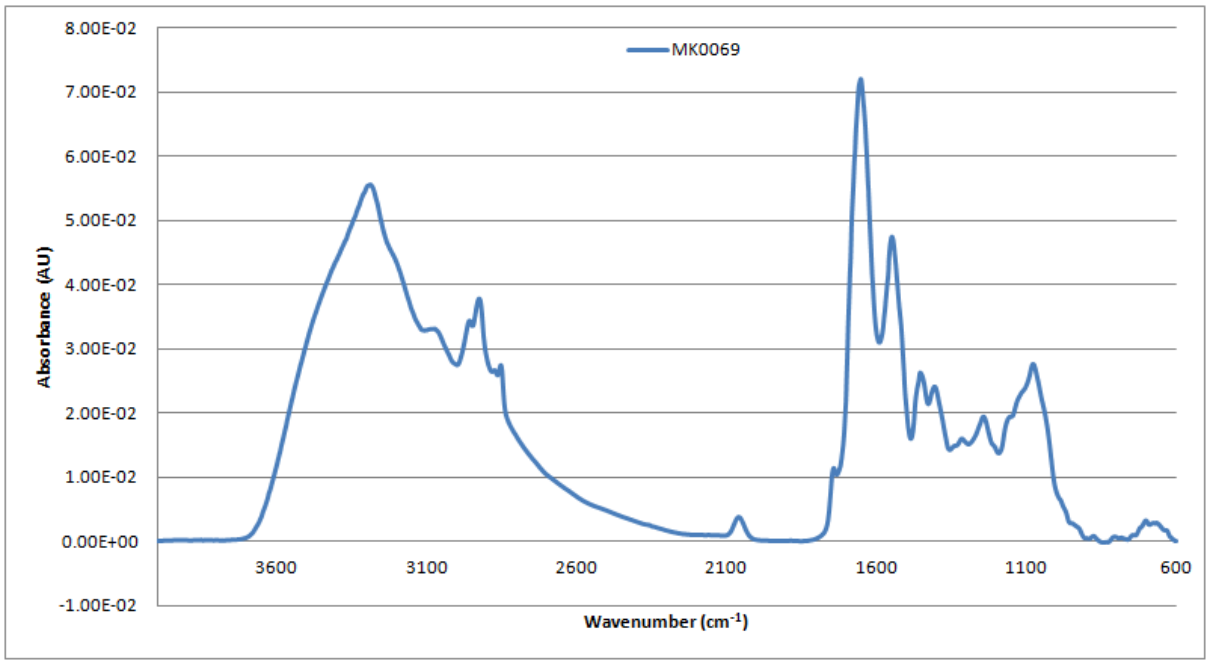


**Figure A2-193: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0069 from 4000-600cm$^{-1}$**

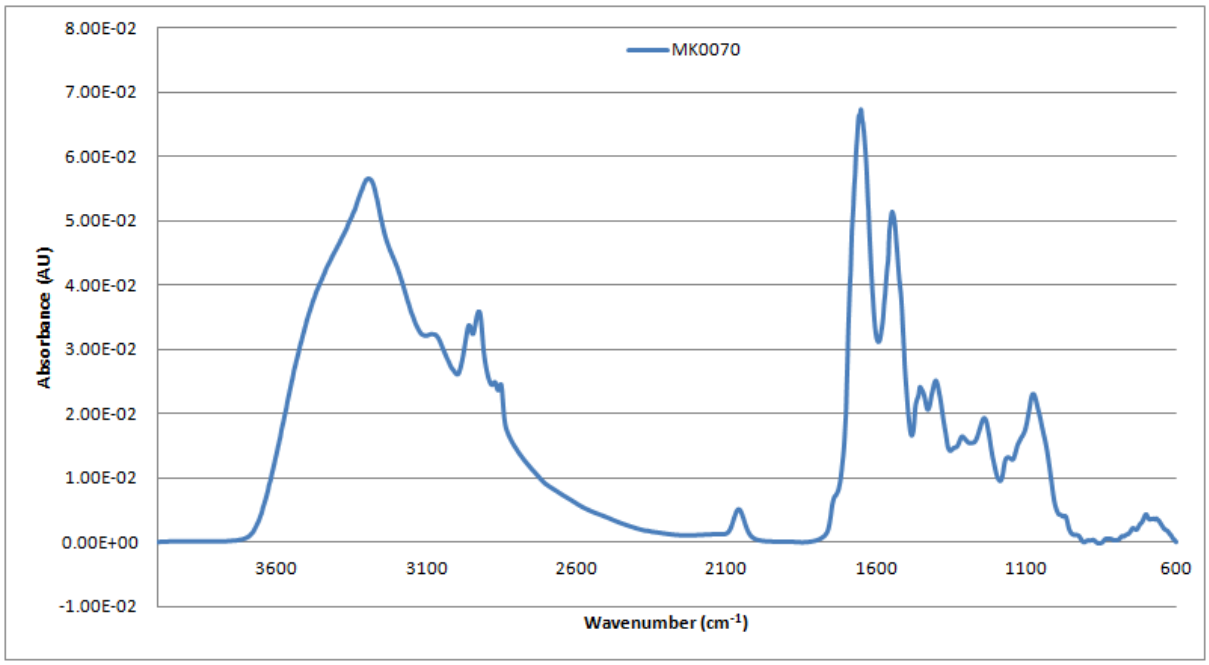


**Figure A2- 194: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient MK0070 from 4000-600cm$^{-1}$**

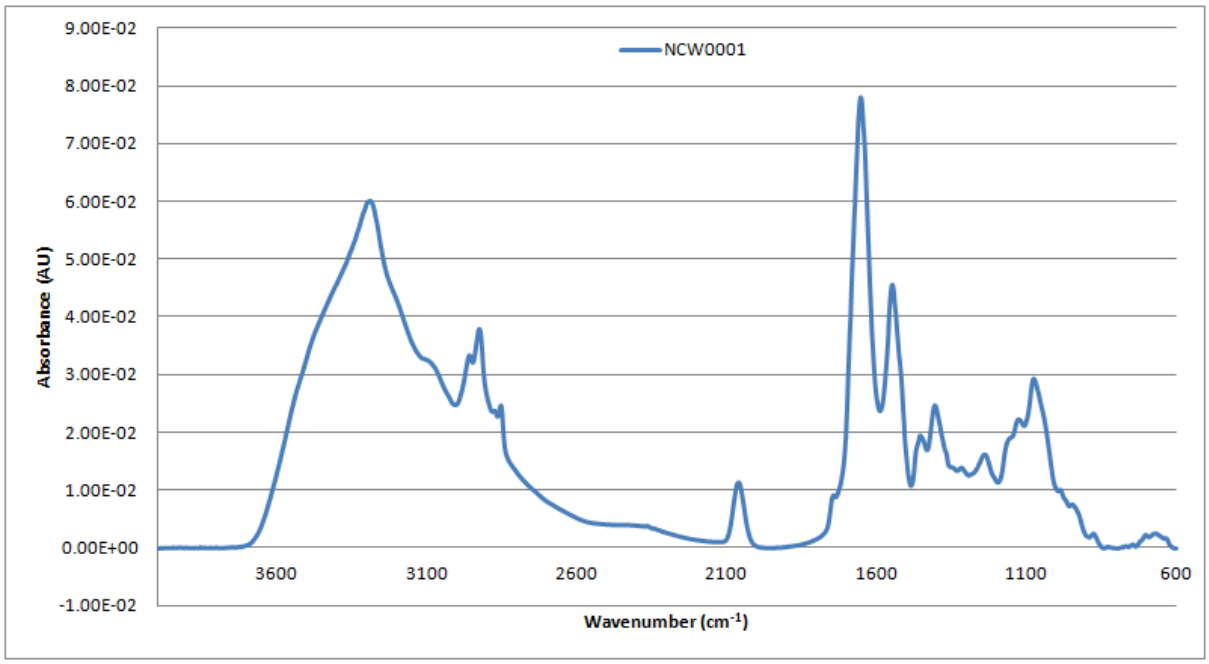


**Figure A2- 195: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0001 from 4000-600cm$^{-1}$**

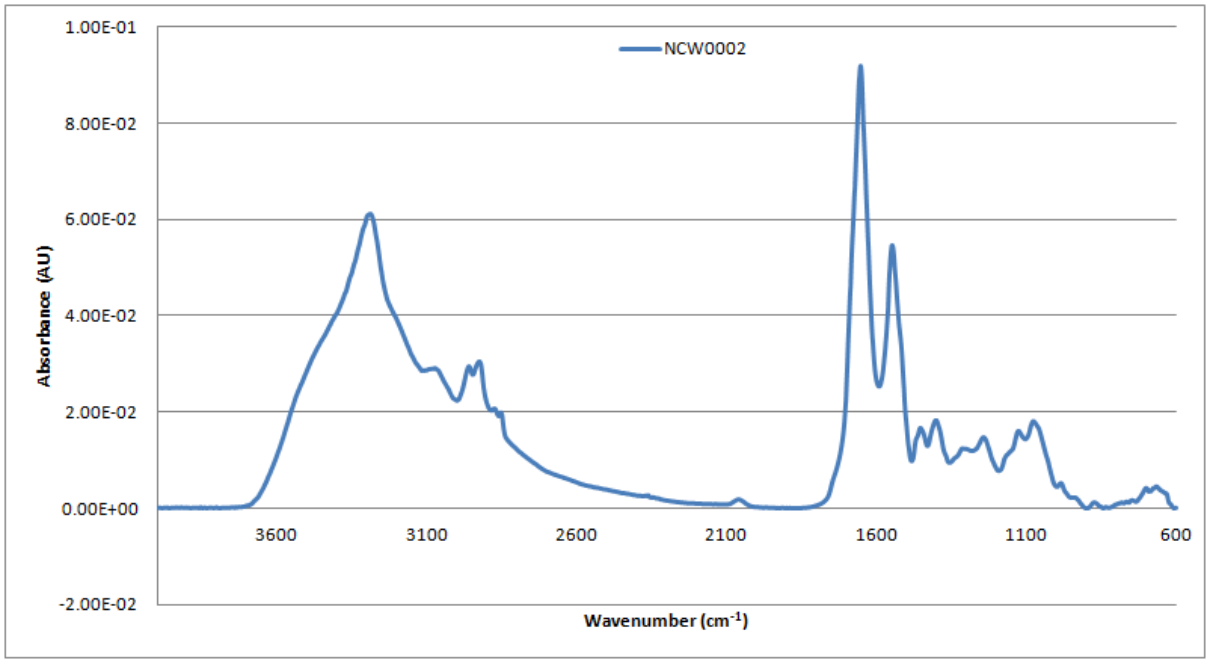


**Figure A2- 196: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0002 from 4000-600cm$^{-1}$**

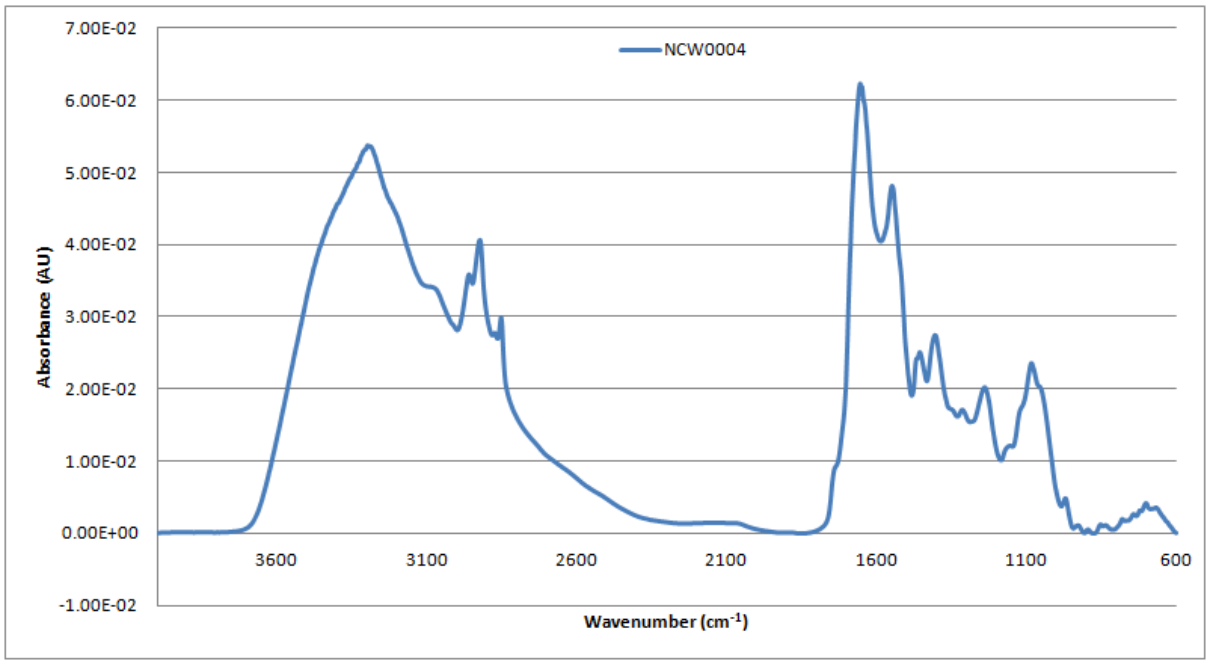


**Figure A2- 197: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0004 from 4000-600cm$^{-1}$**

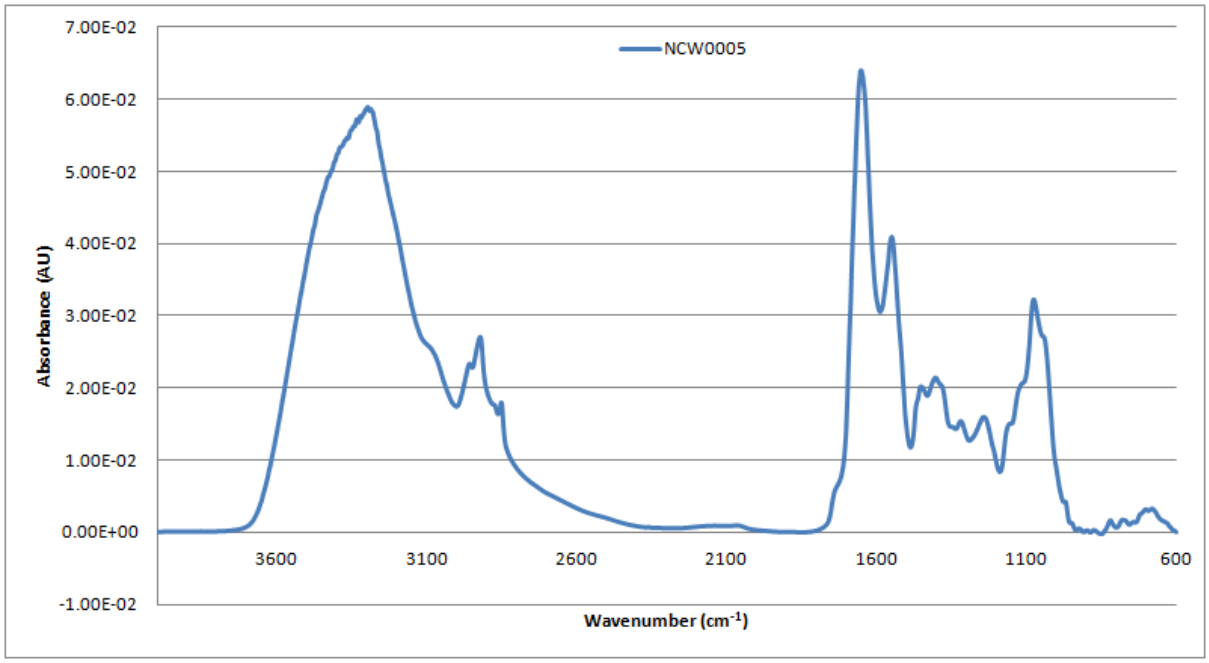


**Figure A2- 198: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0005 from 4000-600cm$^{-1}$**

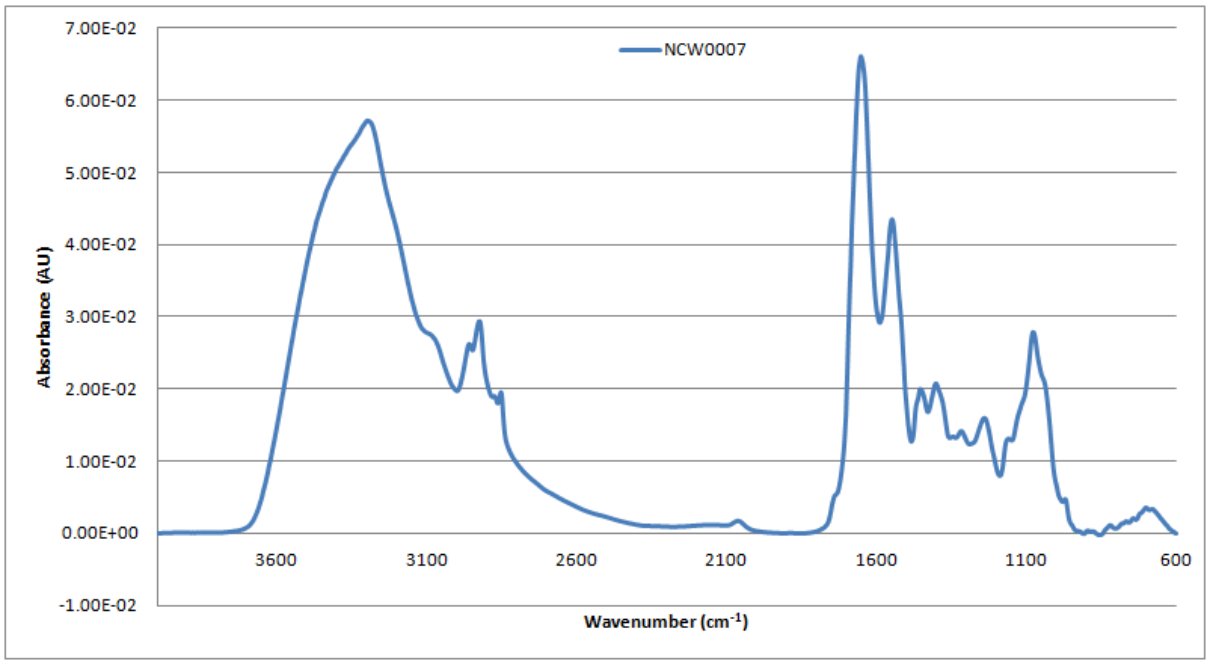


**Figure A2- 199: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0007 from 4000-600cm$^{-1}$**

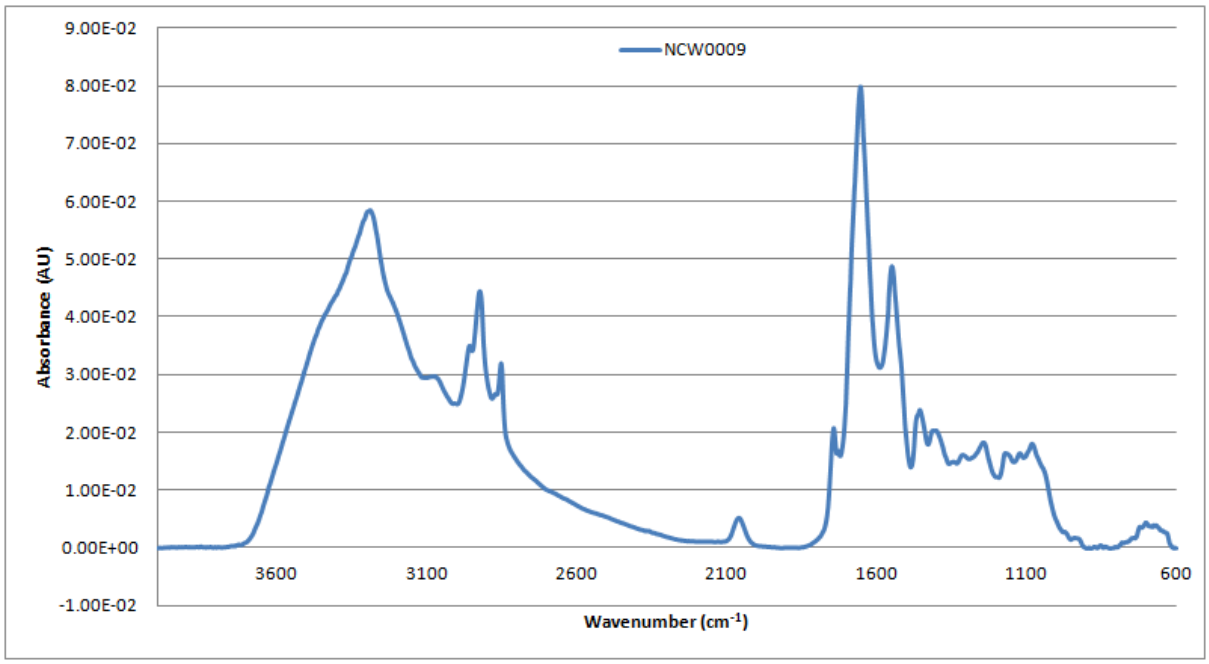


**Figure A2- 200: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0009 from 4000-600cm$^{-1}$**

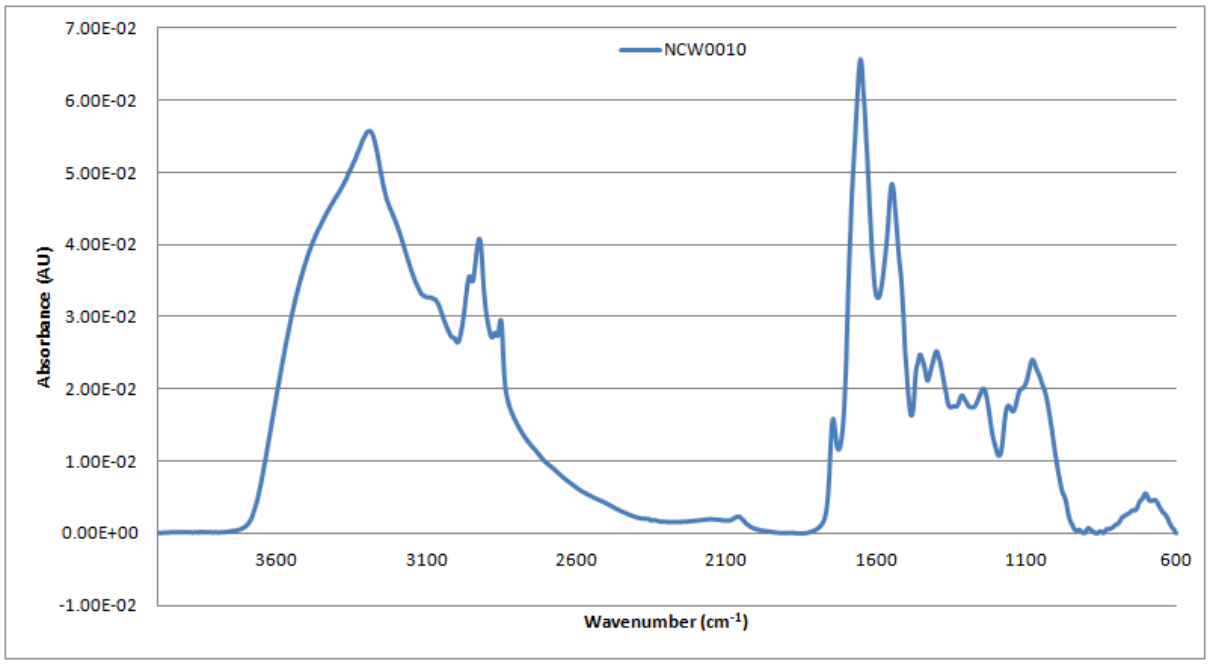


**Figure A2- 201: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0010 from 4000-600cm$^{-1}$**

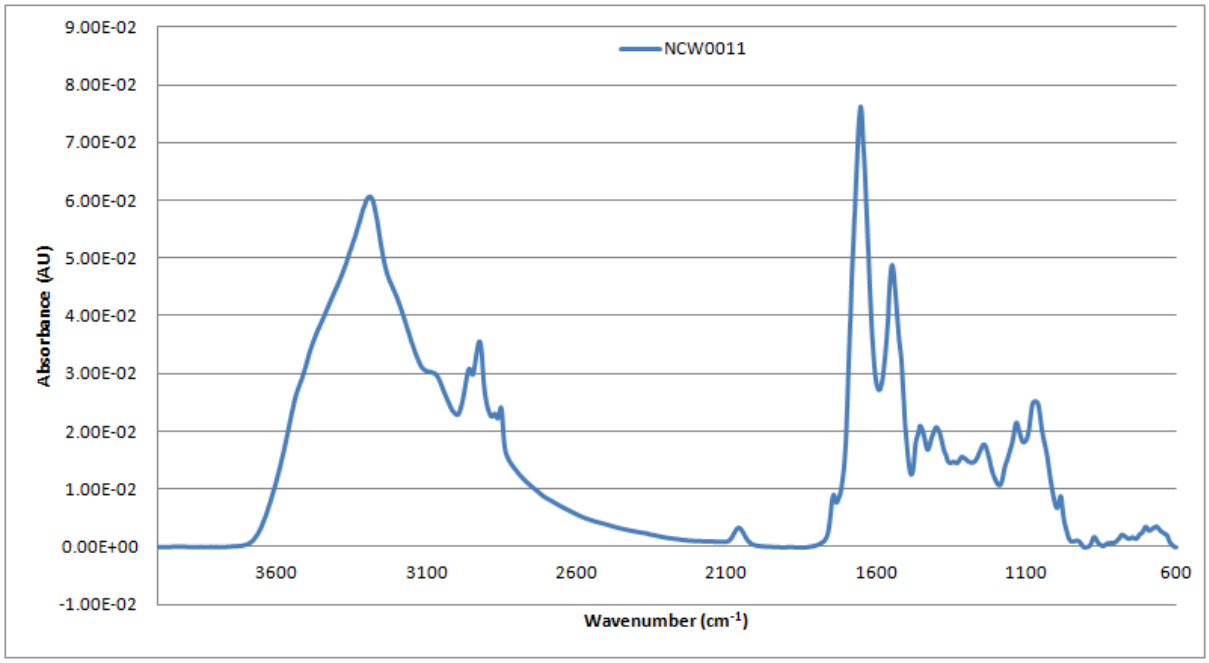


**Figure A2- 202: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0011 from 4000-600cm$^{-1}$**

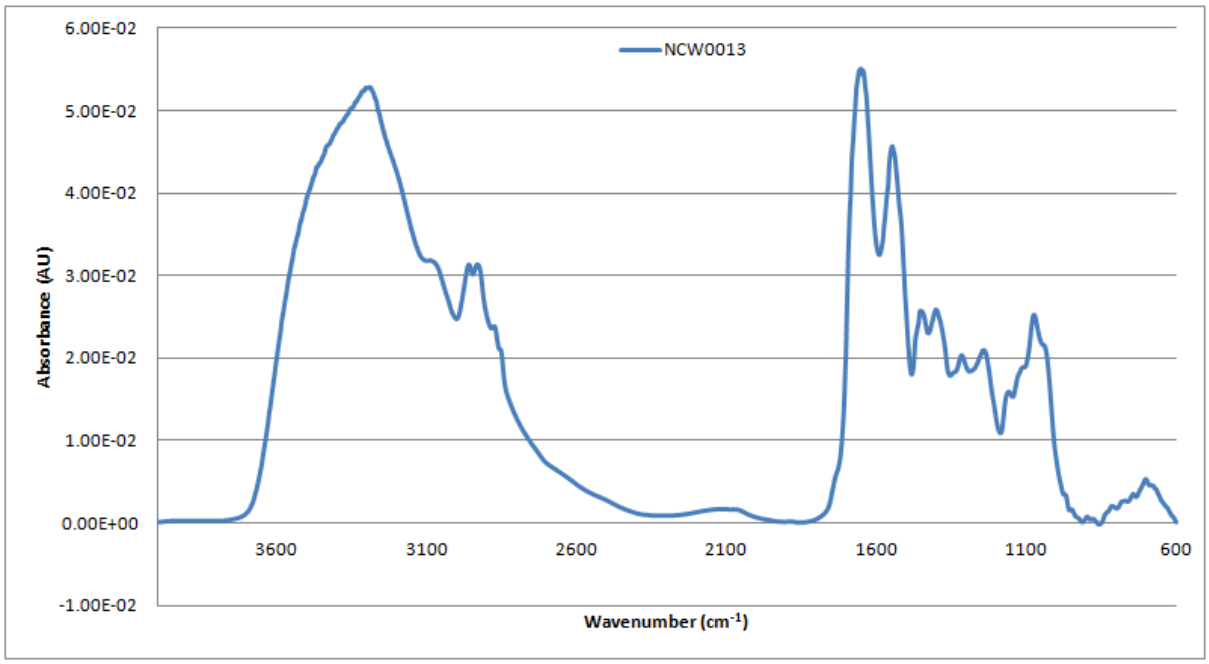


**Figure A2- 203: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0013 from 4000-600cm$^{-1}$**

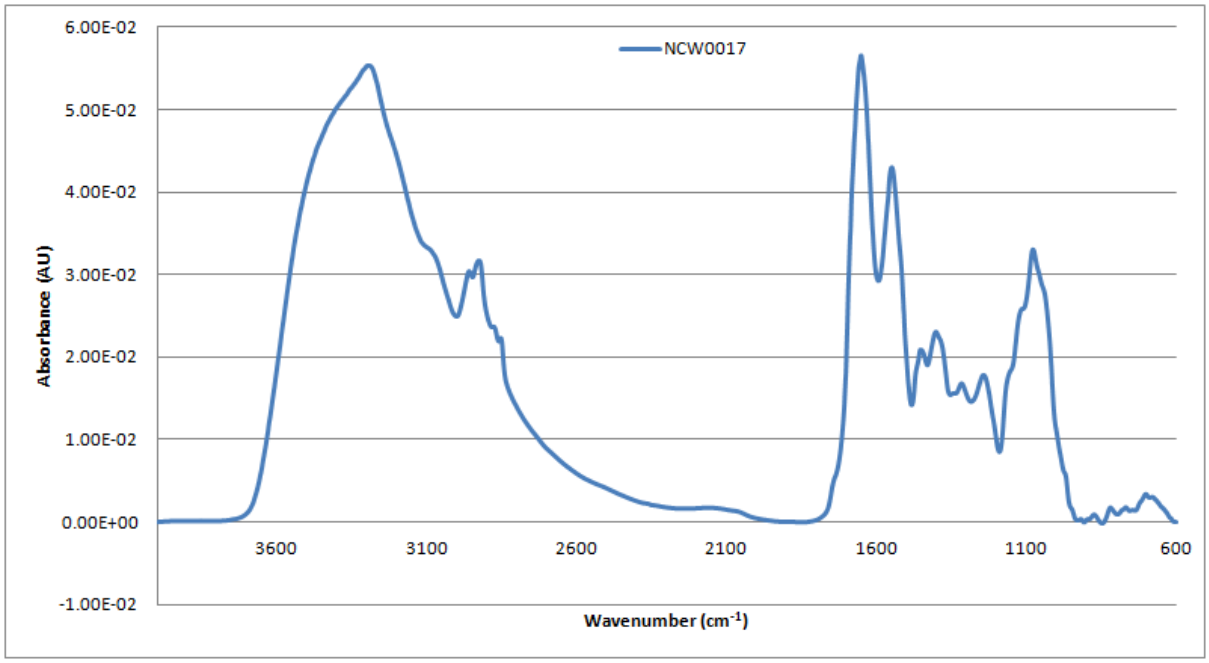


**Figure A2- 204: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0017 from 4000-600cm$^{-1}$**

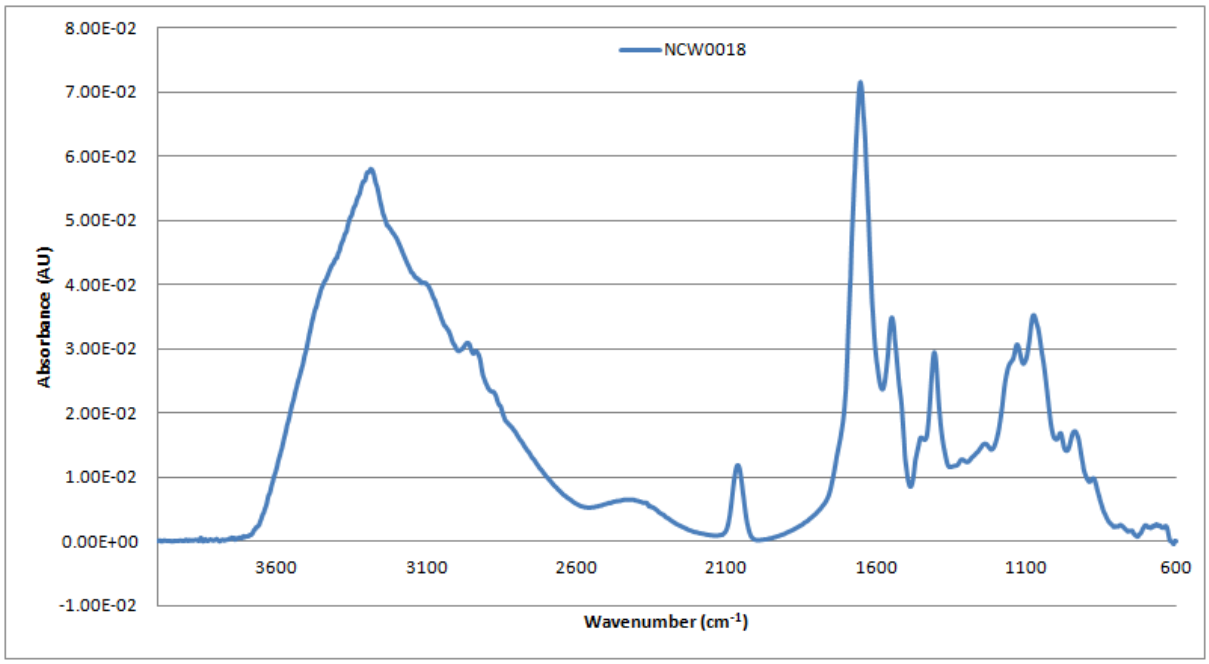


**Figure A2- 205: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0018 from 4000-600cm$^{-1}$**

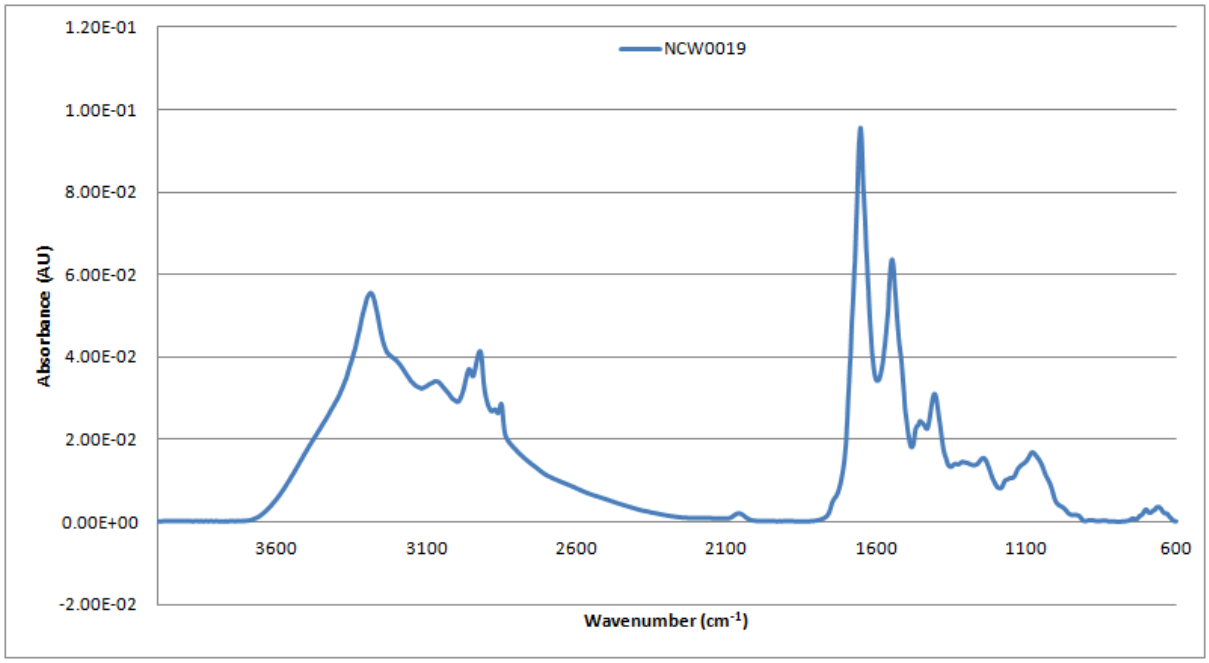


**Figure A2- 206: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0019 from 4000-600cm$^{-1}$**

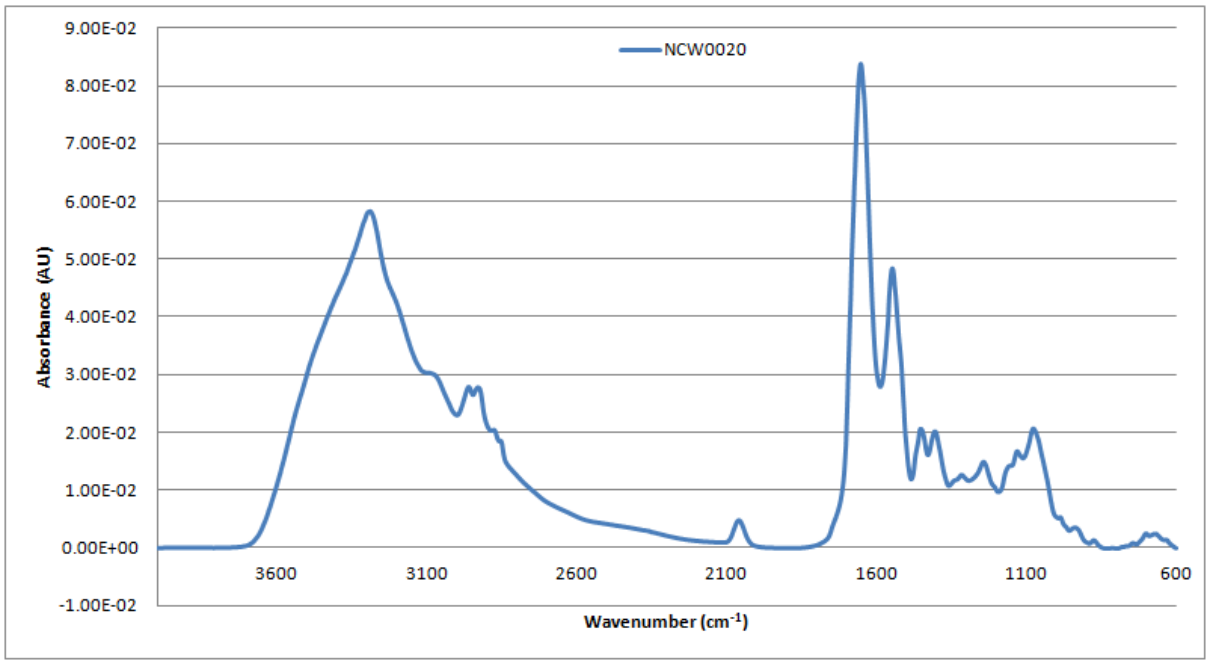


**Figure A2- 207: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0020 from 4000-600cm$^{-1}$**

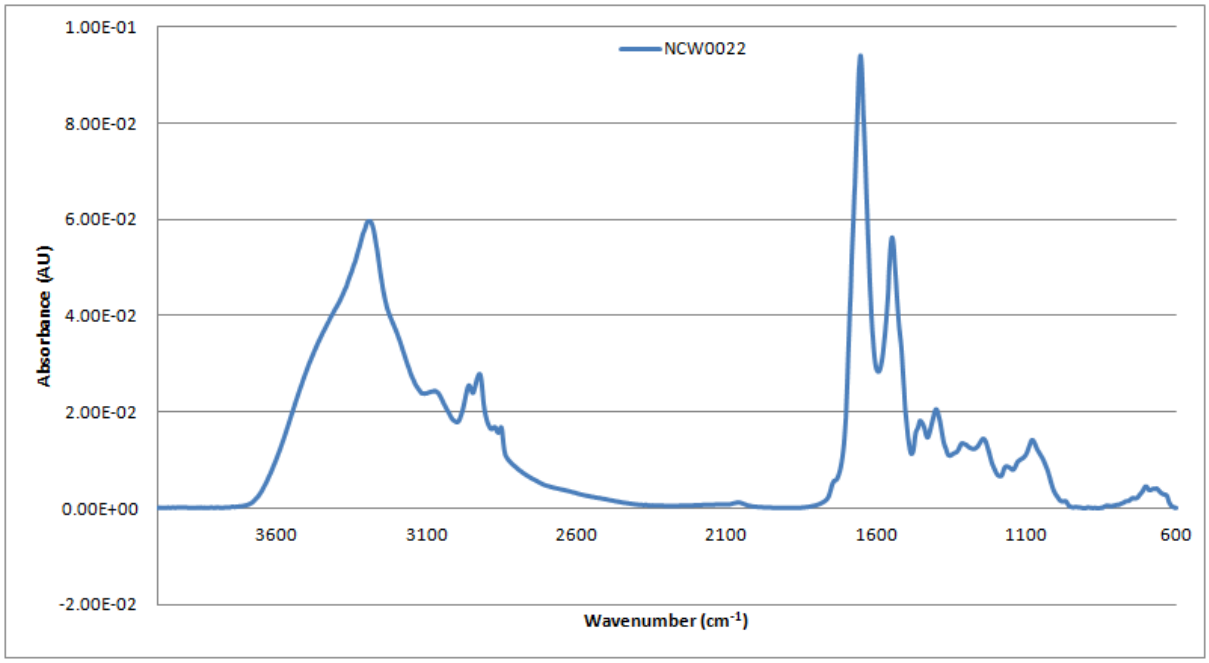


**Figure A2- 208: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0022 from 4000-600cm$^{-1}$**

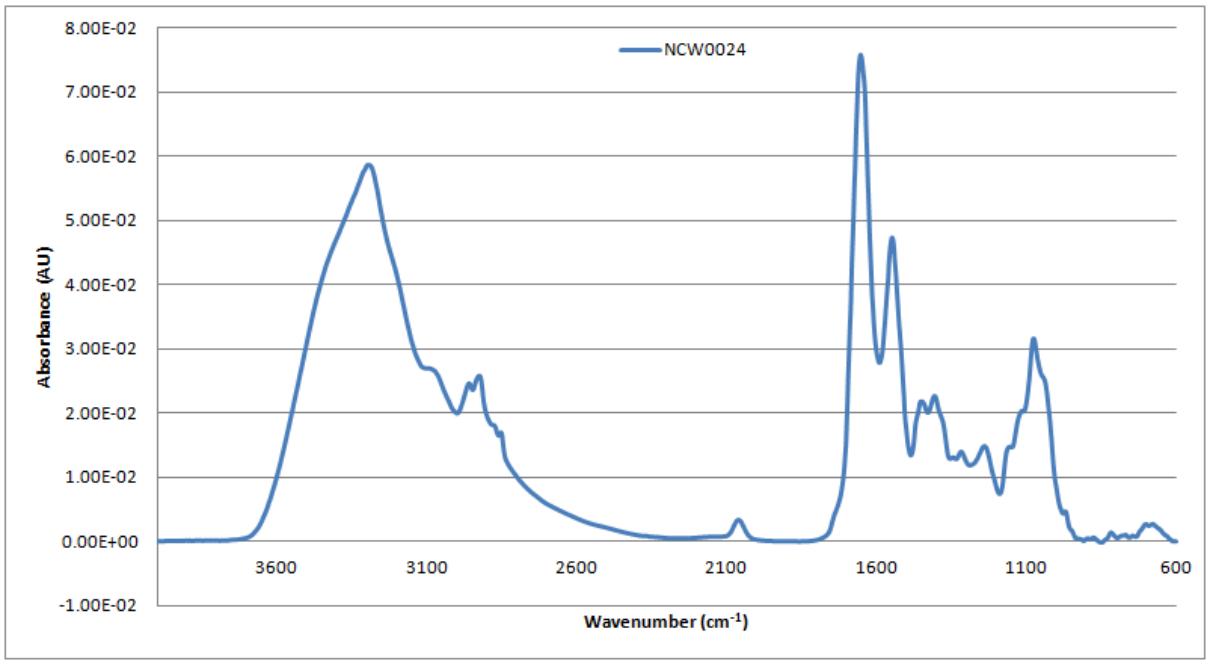


**Figure A2- 209: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0024 from 4000-600cm$^{-1}$**

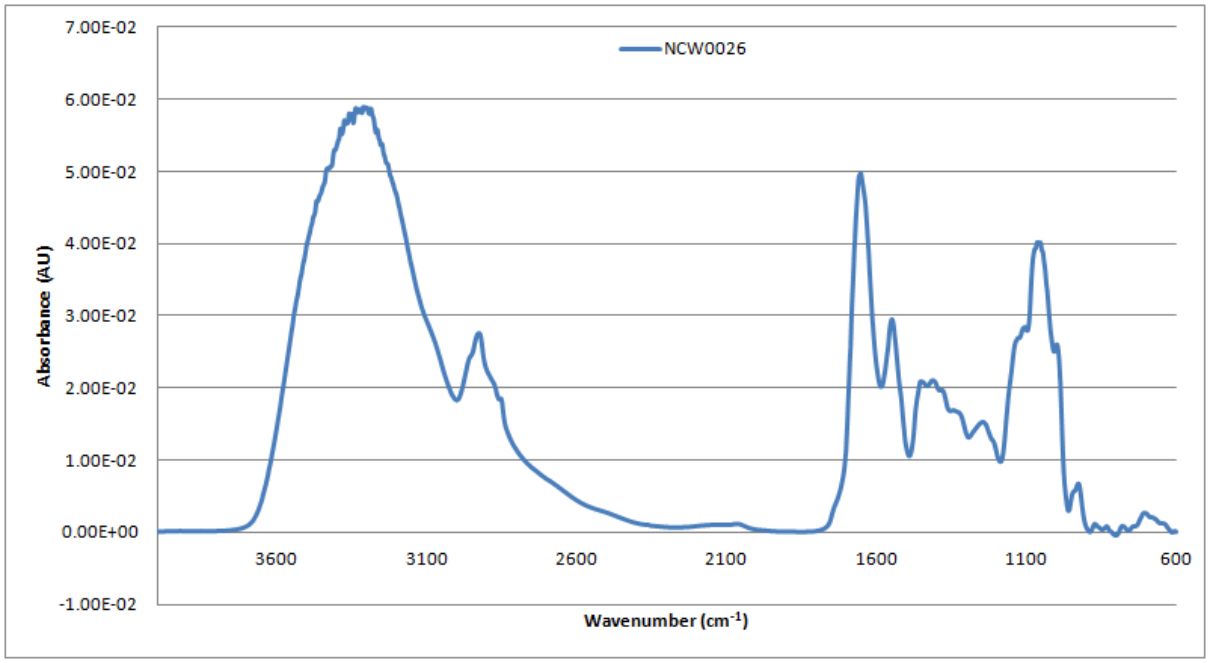


**Figure A2- 210: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0026 from 4000-600cm$^{-1}$**

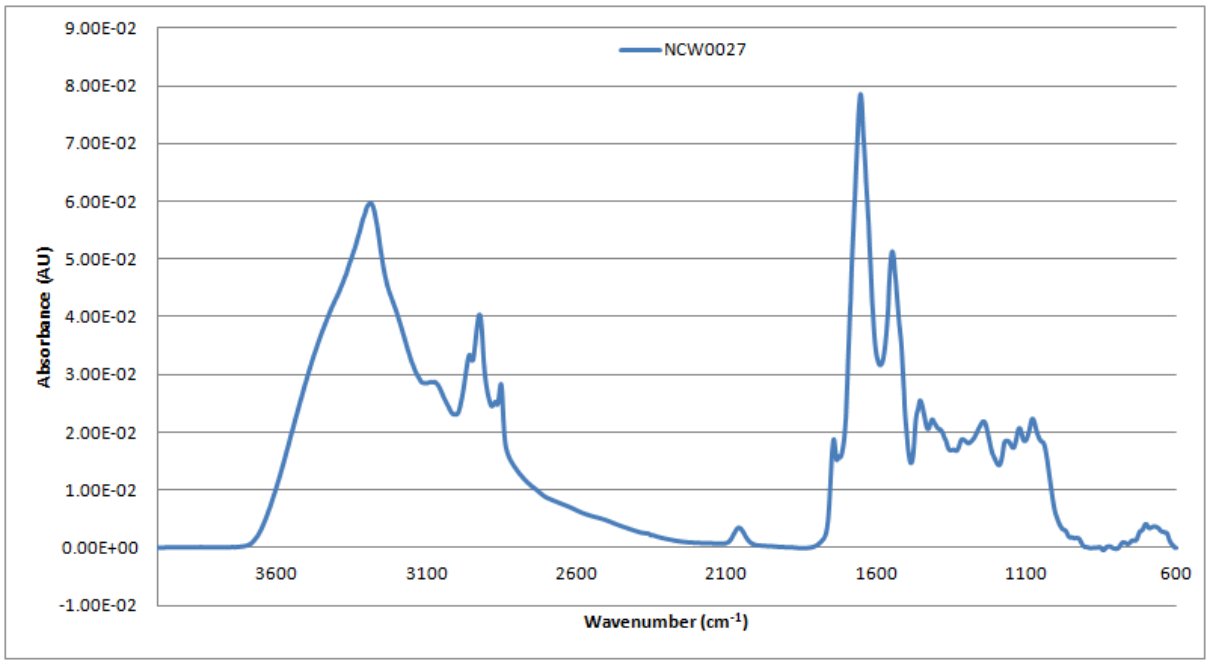


**Figure A2- 211: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0027 from 4000-600cm$^{-1}$**

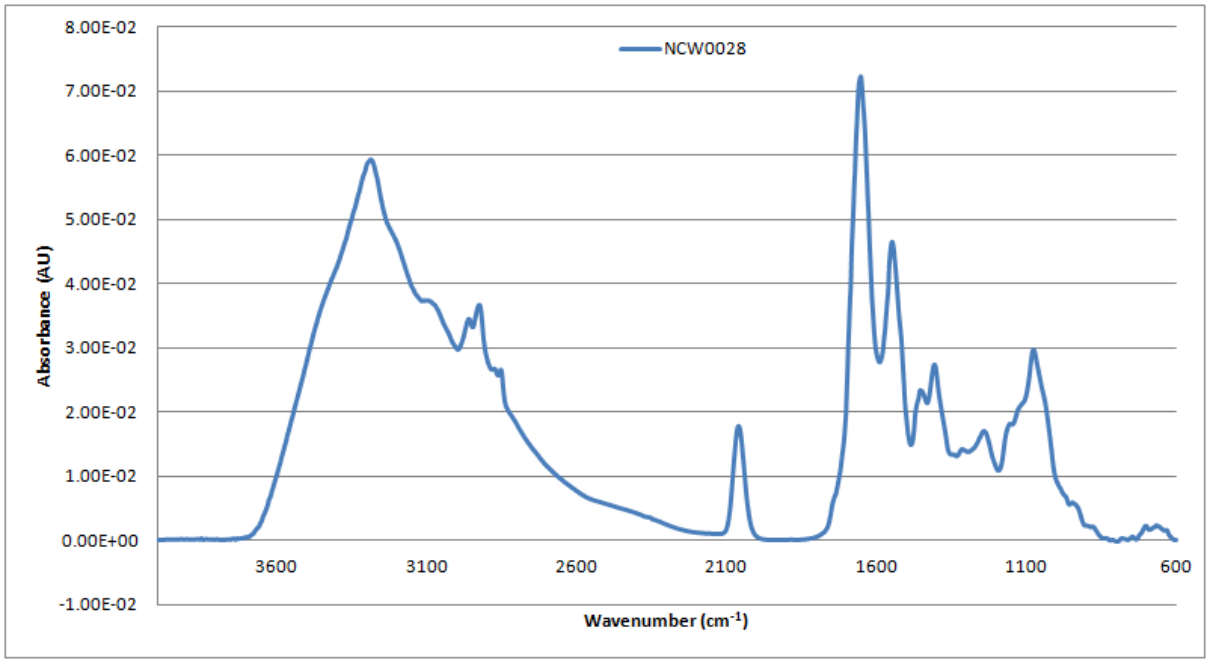


**Figure A2- 212: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0028 from 4000-600cm$^{-1}$**

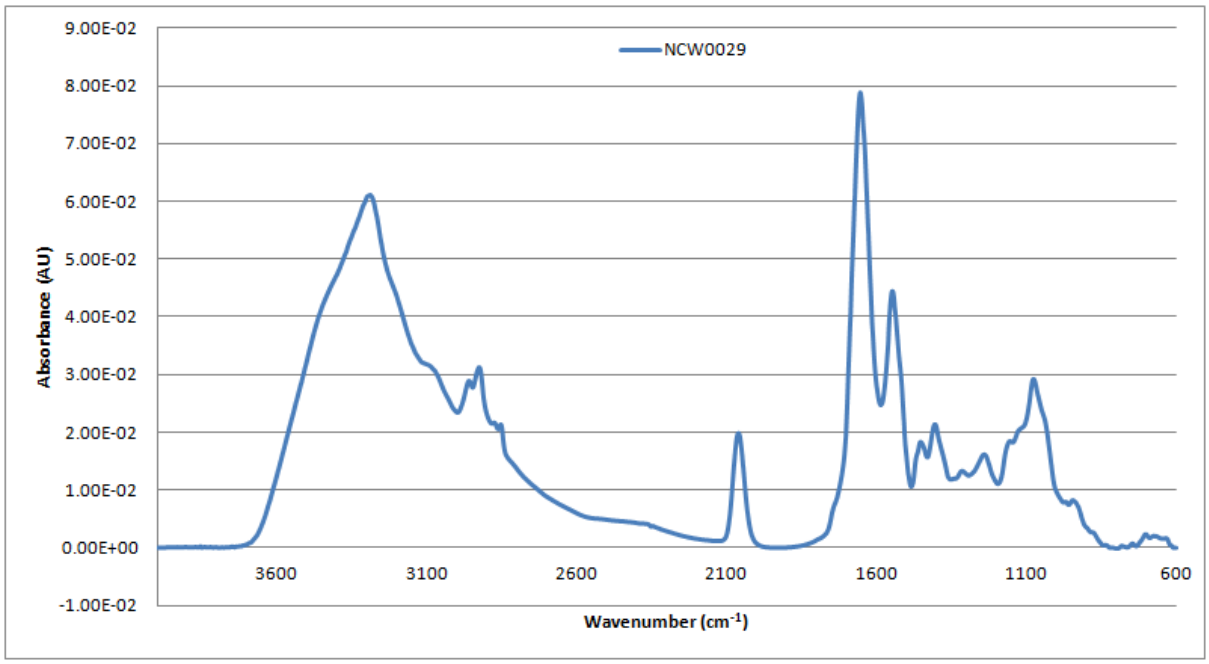


**Figure A2- 213: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0029 from 4000-600cm$^{-1}$**

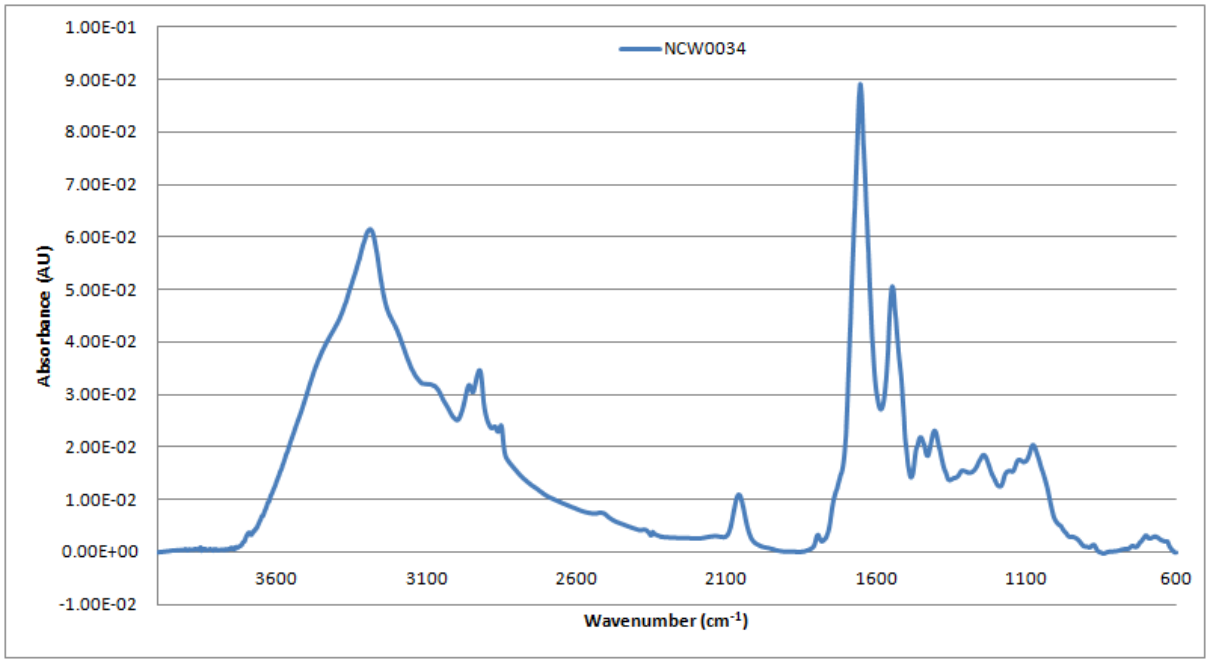


**Figure A2- 214: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0034 from 4000-600cm$^{-1}$**

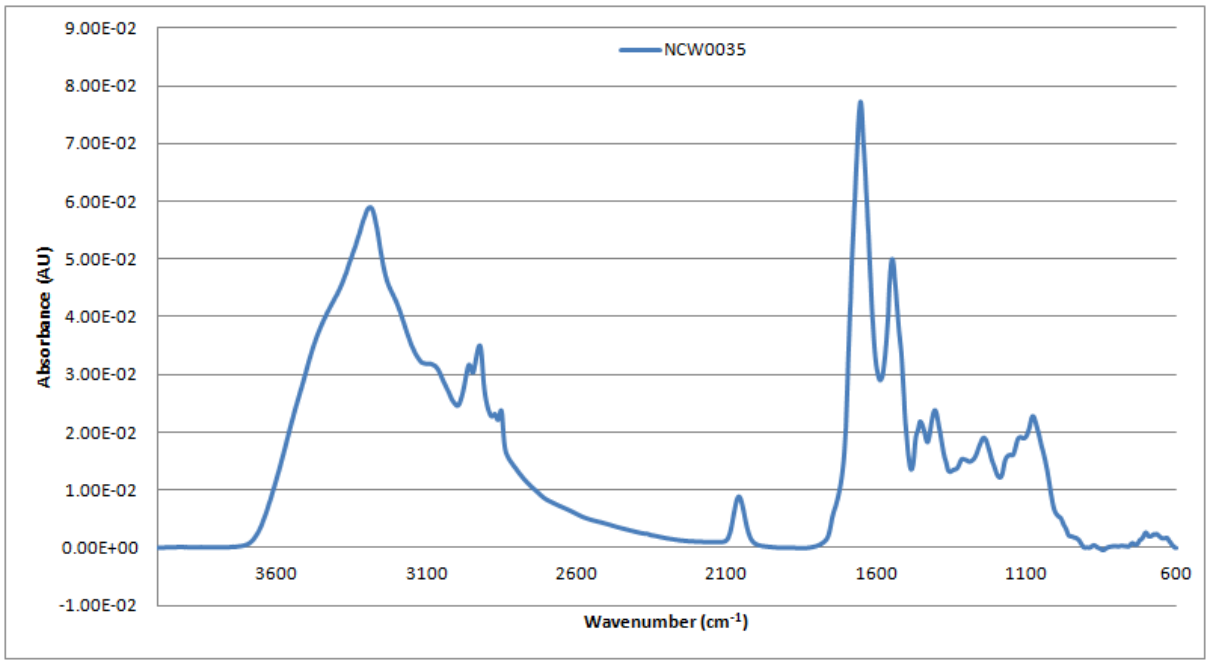


**Figure A2- 215: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0035 from 4000-600cm$^{-1}$**

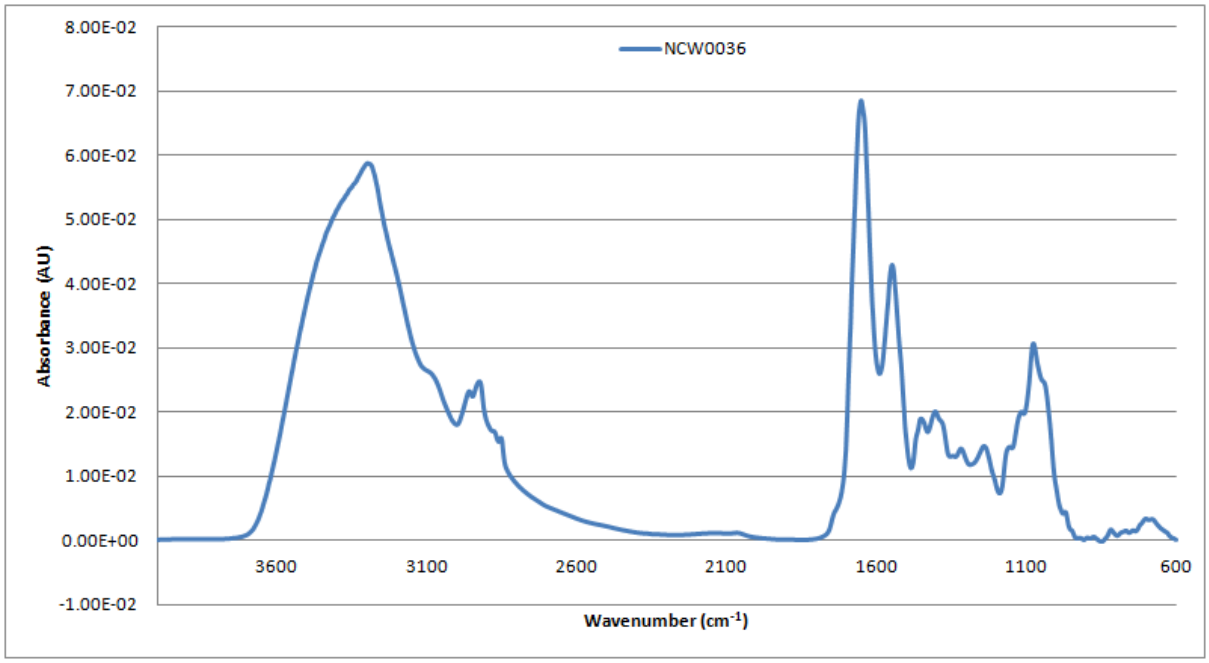


**Figure A2- 216: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0036 from 4000-600cm$^{-1}$**

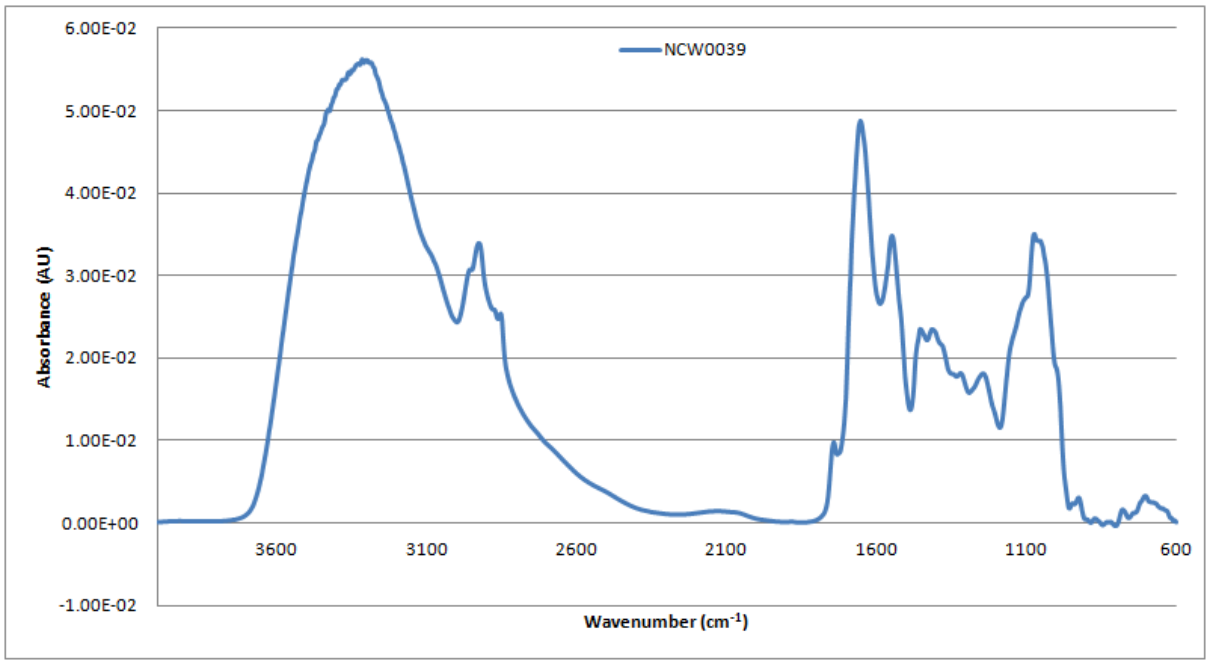


**Figure A2- 217: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0039 from 4000-600cm$^{-1}$**

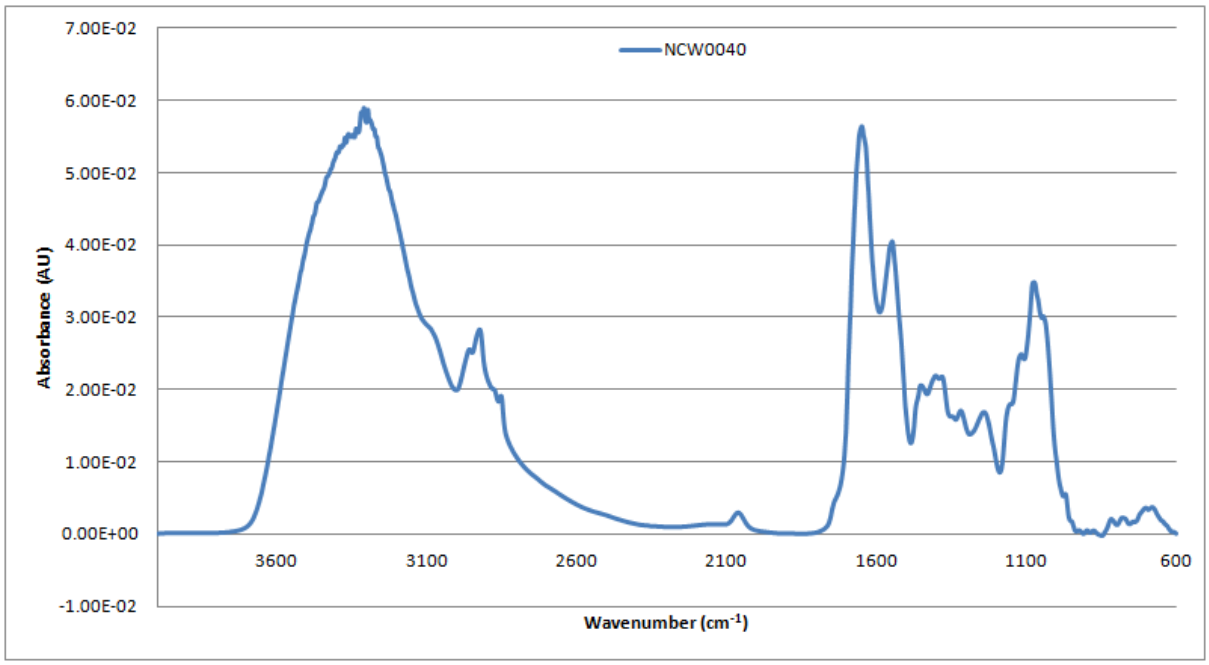


**Figure A2- 218: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0040 from 4000-600cm$^{-1}$**

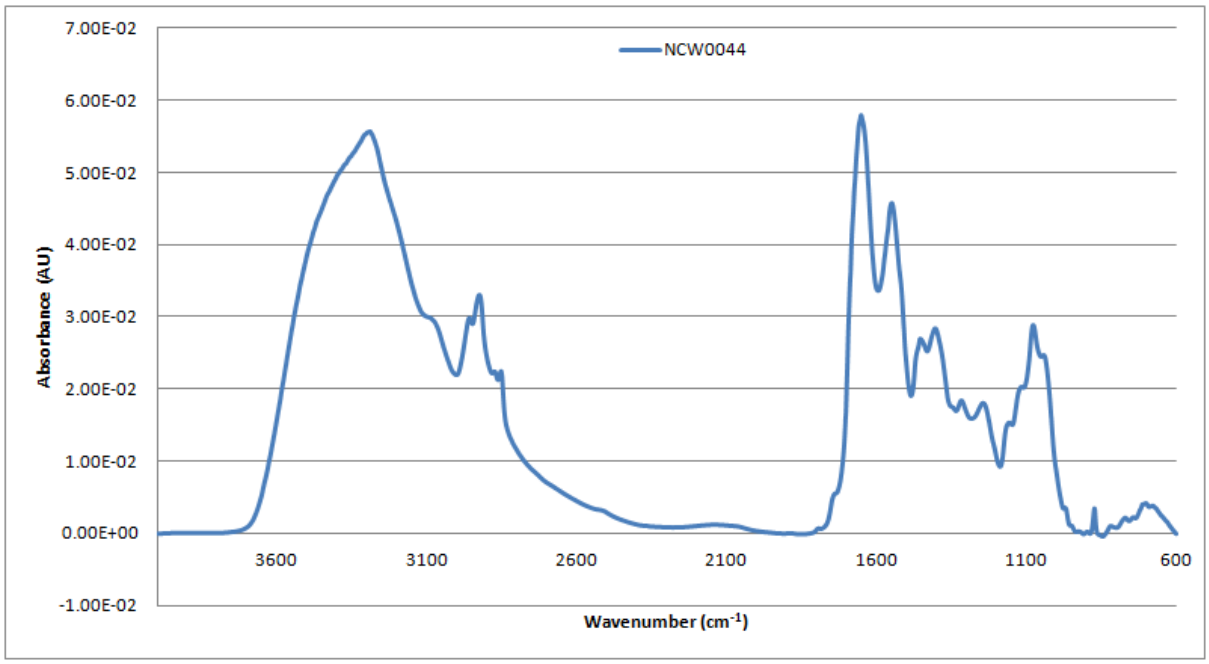


**Figure A2- 219: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0044 from 4000-600cm$^{-1}$**

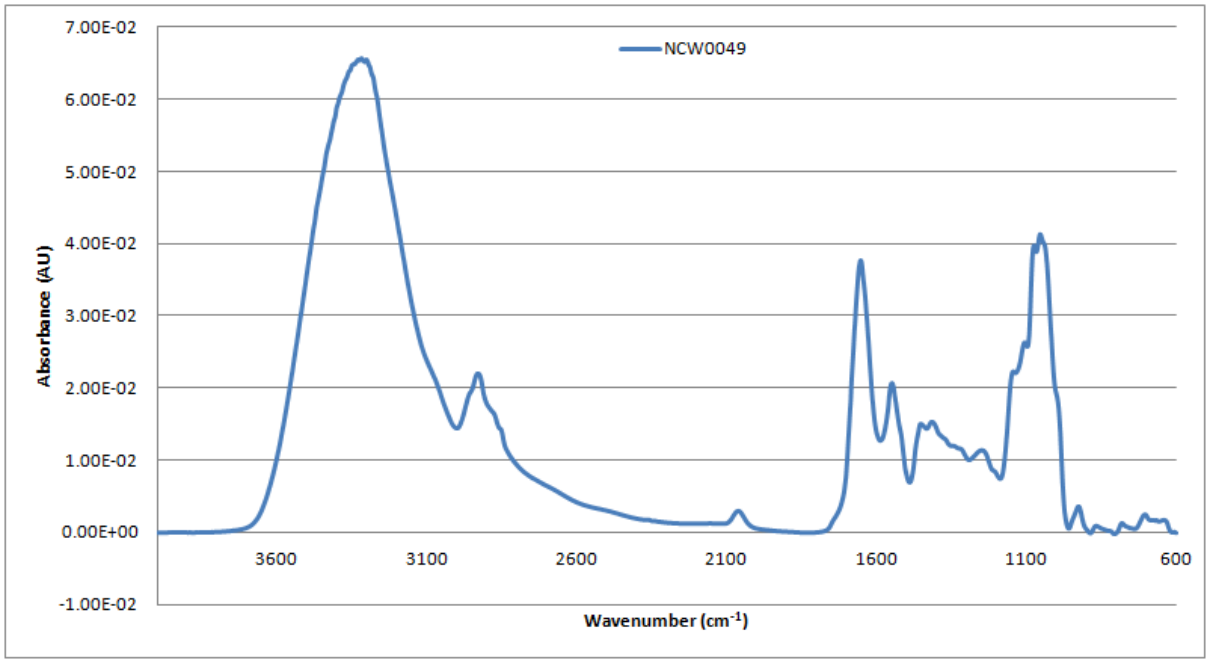


**Figure A2- 220: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0049 from 4000-600cm$^{-1}$**

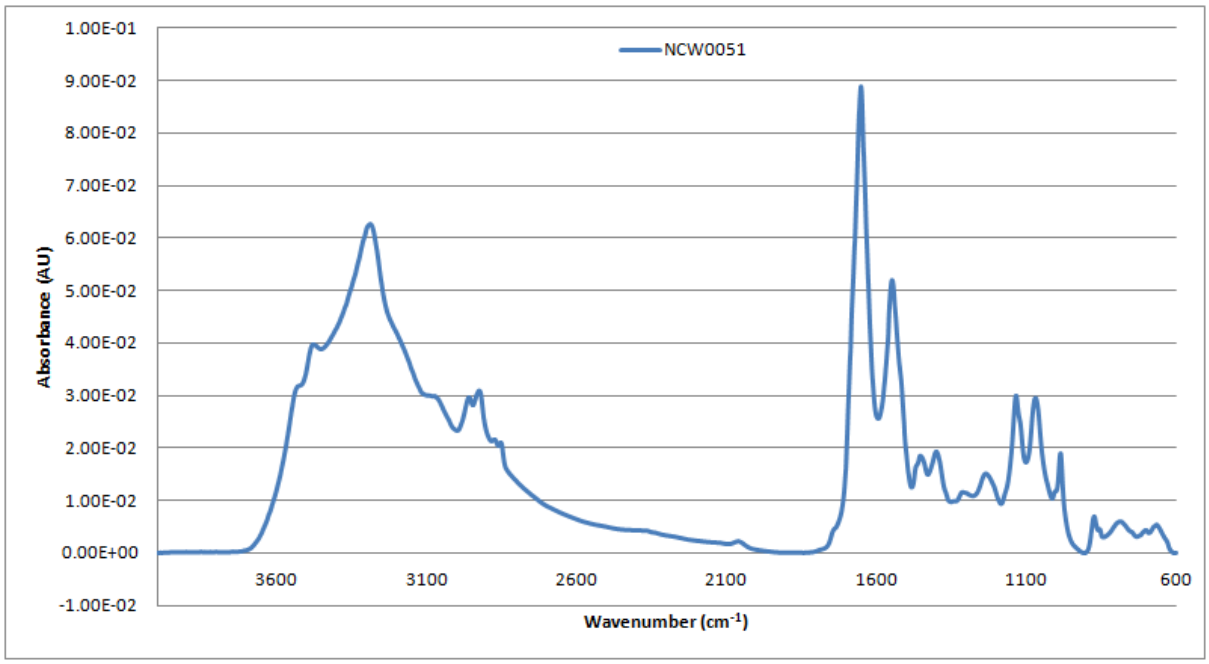


**Figure A2- 221: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0051 from 4000-600cm$^{-1}$**

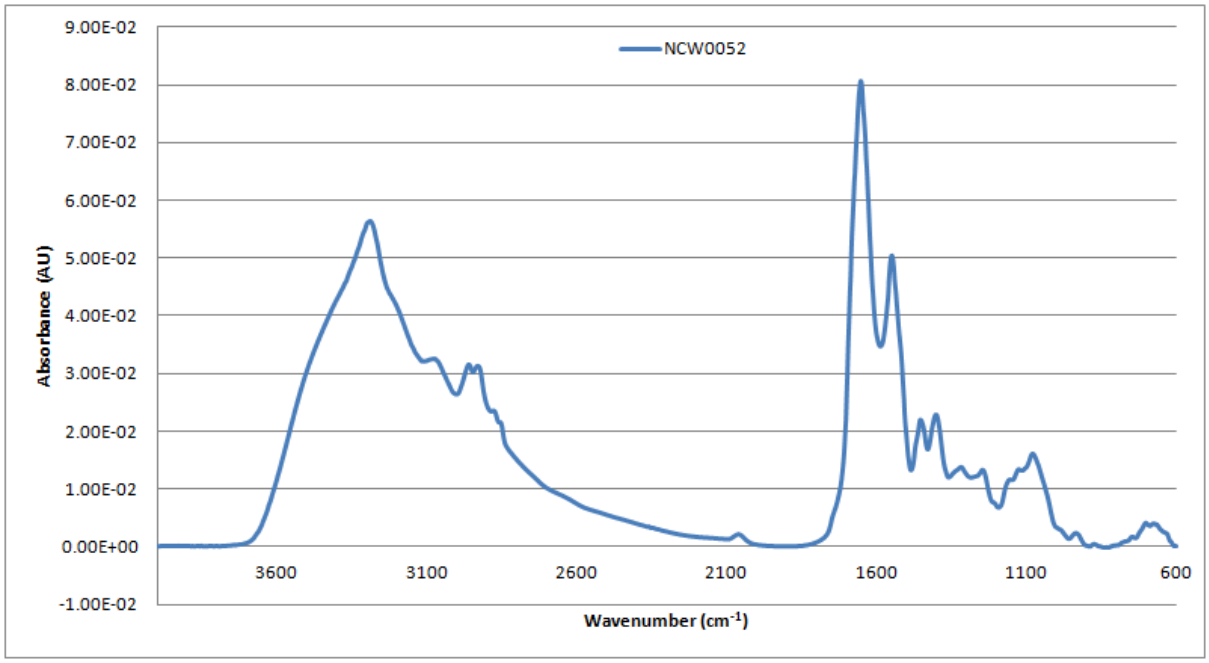


**Figure A2- 222: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0052 from 4000-600cm$^{-1}$**

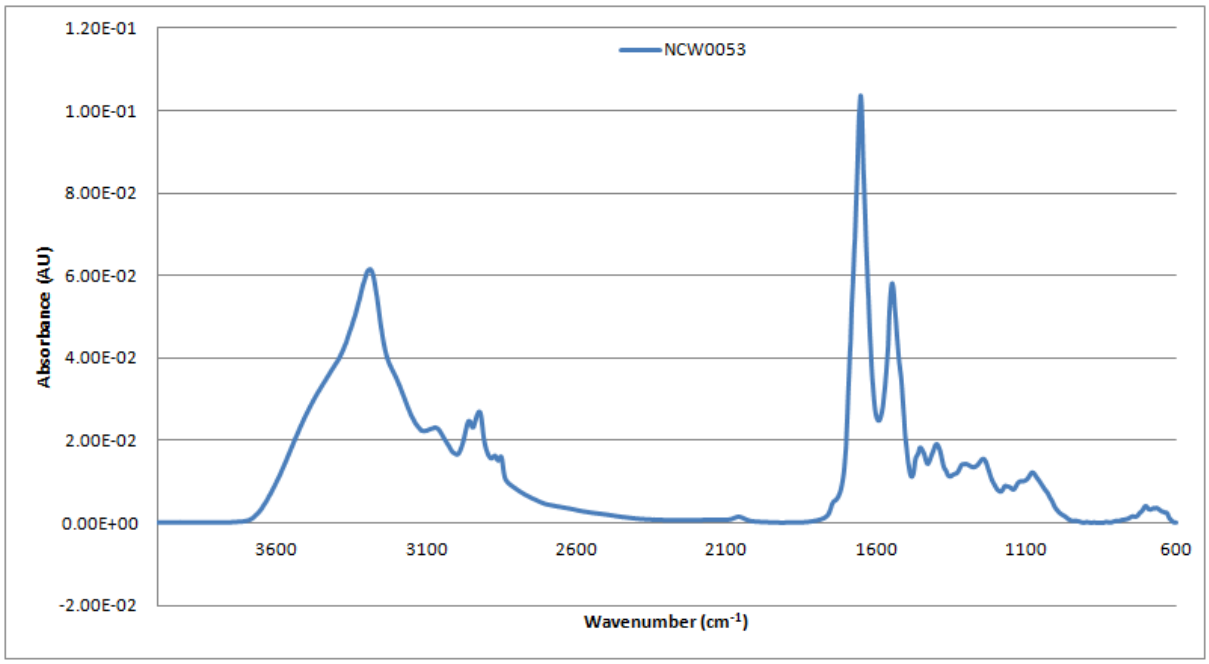


**Figure A2- 223: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0053 from 4000-600cm$^{-1}$**

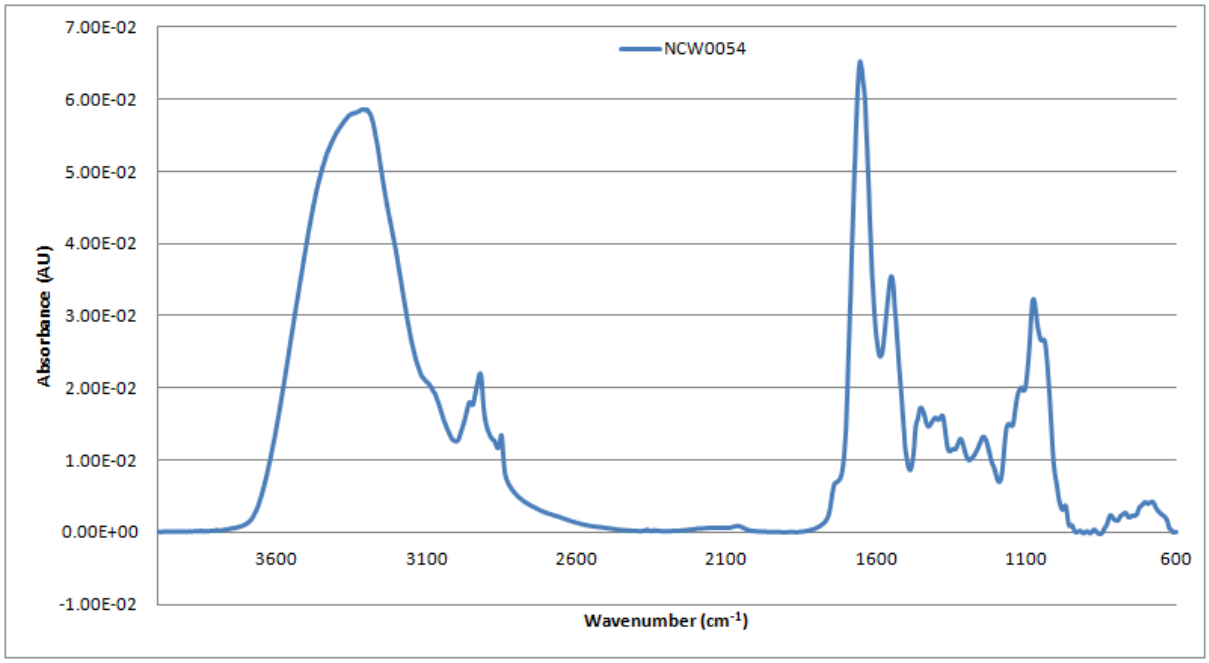


**Figure A2- 224: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0054 from 4000-600cm$^{-1}$**

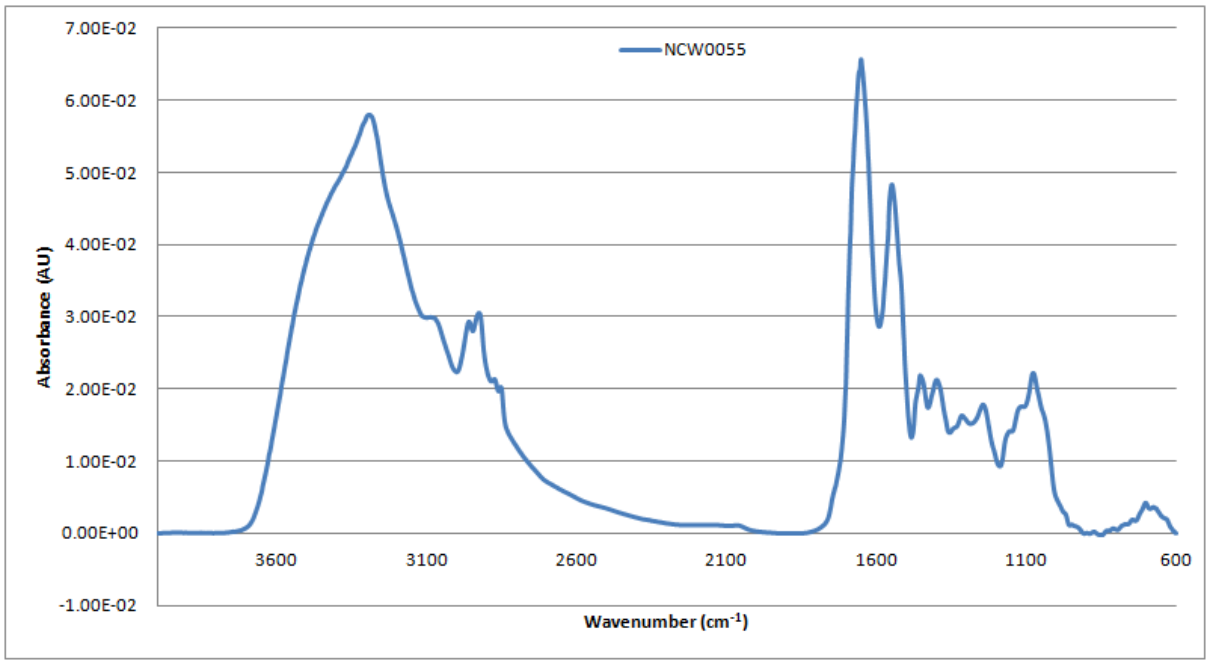


**Figure A2- 225: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0055 from 4000-600cm$^{-1}$**

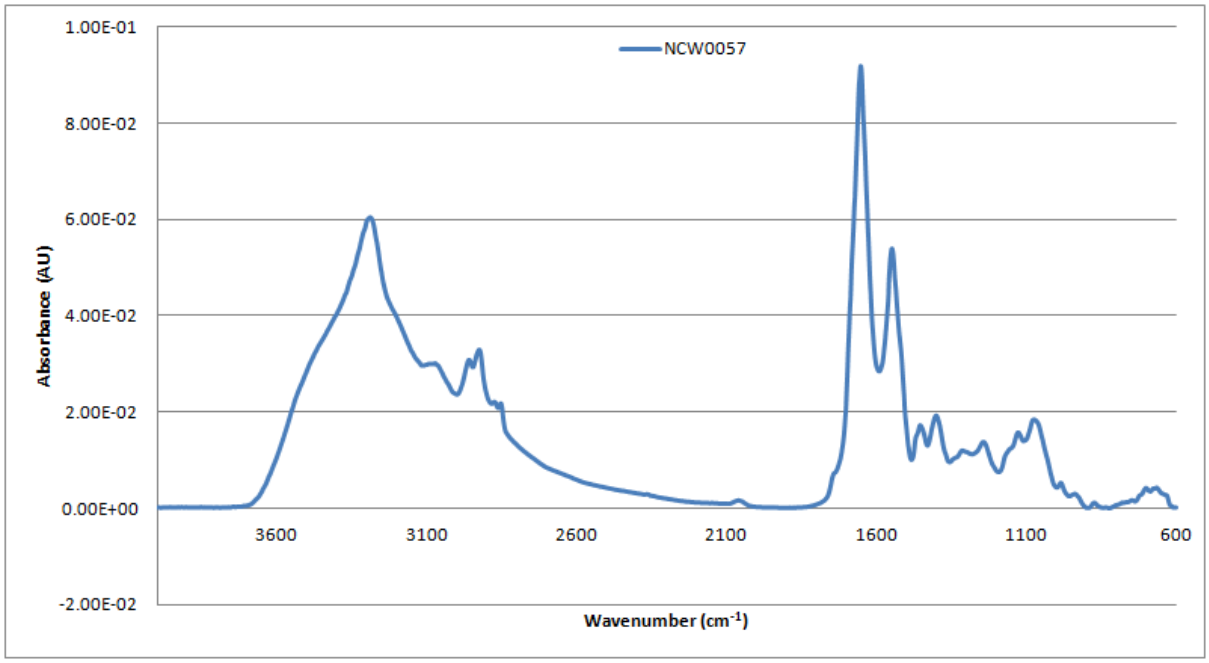


**Figure A2- 226: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0057 from 4000-600cm$^{-1}$**

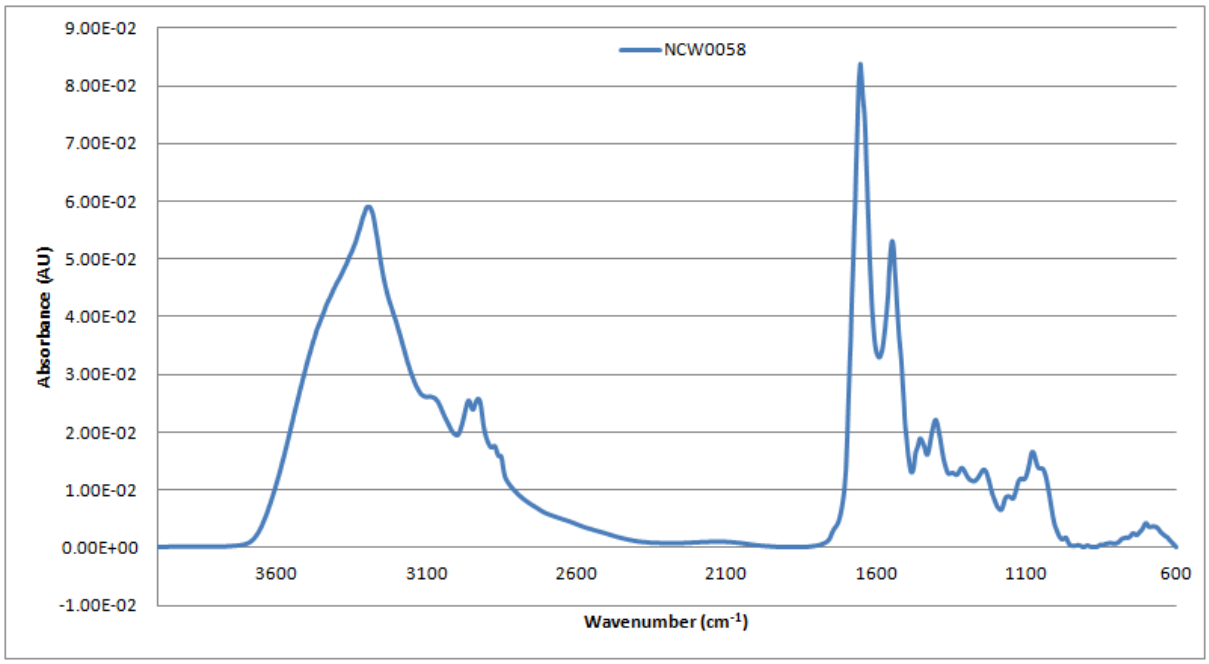


**Figure A2- 227: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0058 from 4000-600cm$^{-1}$**

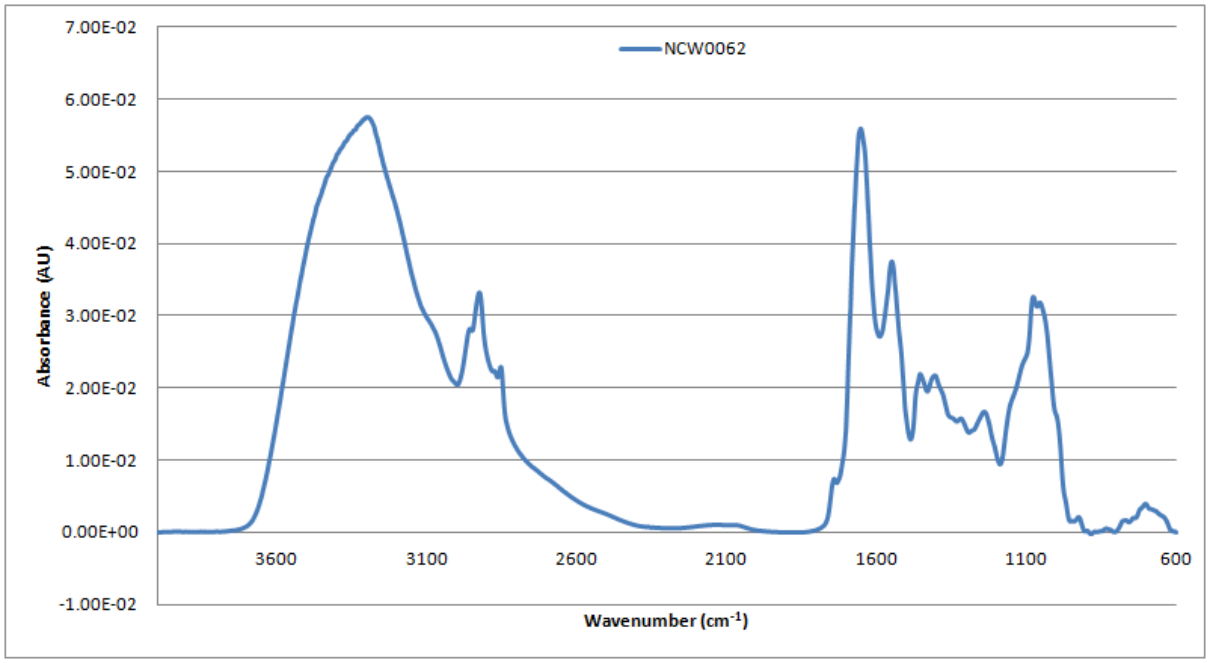


**Figure A2- 228: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0062 from 4000-600cm$^{-1}$**

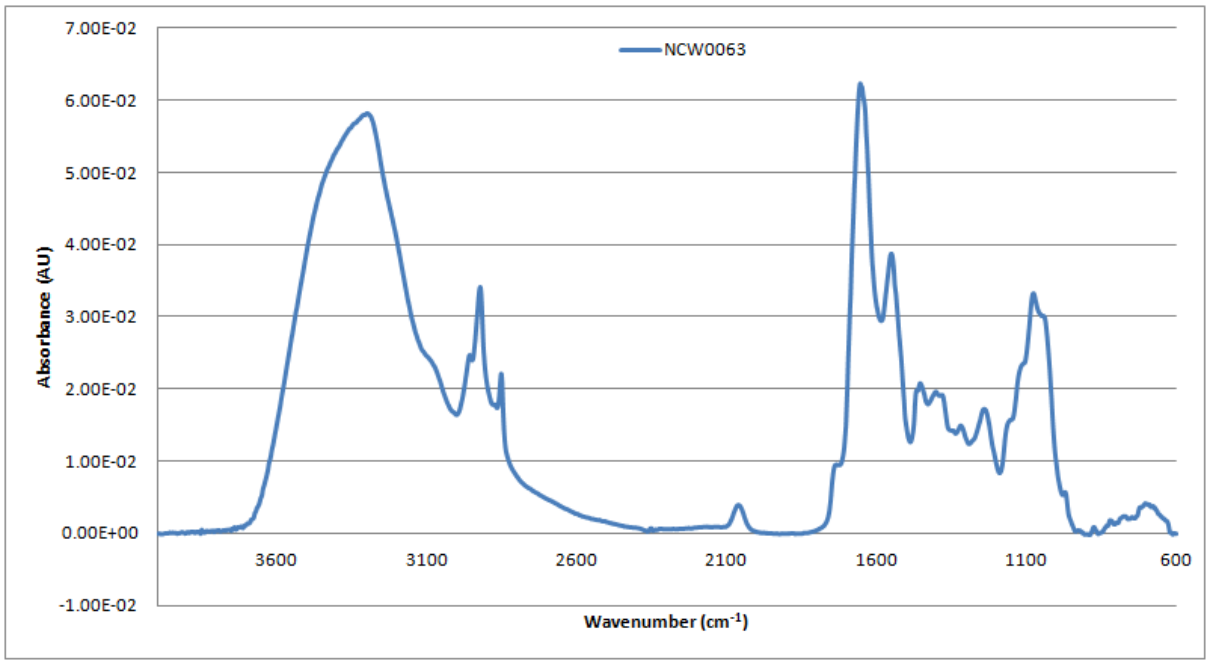


**Figure A2- 229: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0063 from 4000-600cm$^{-1}$**

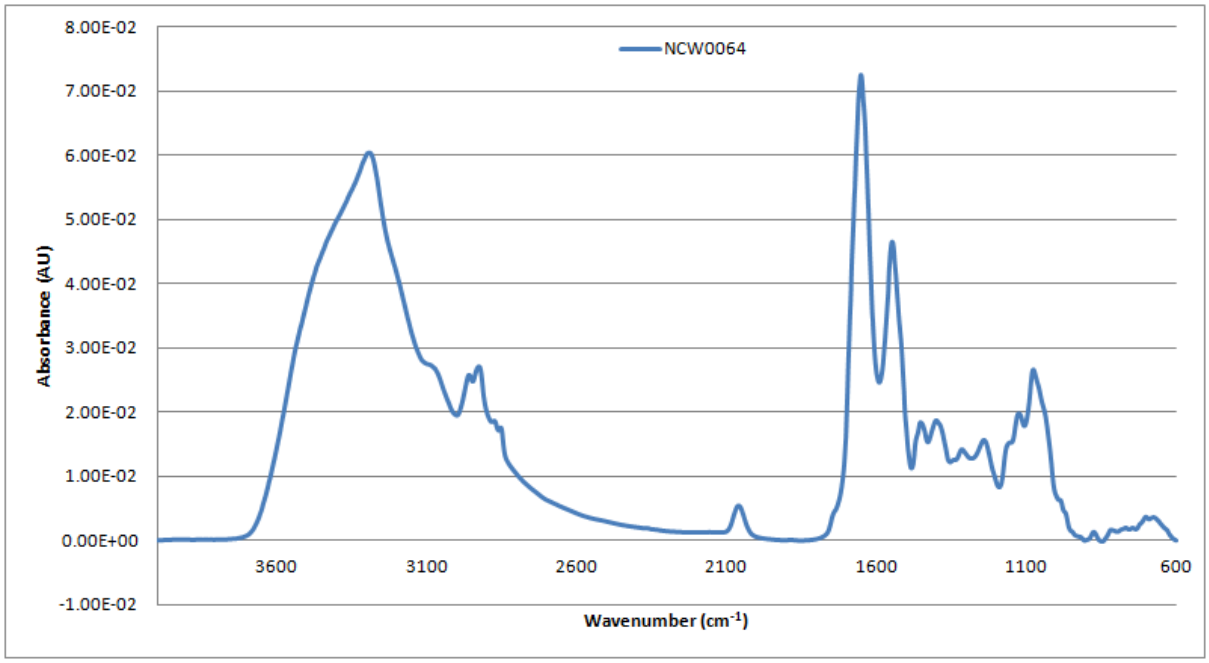


**Figure A2- 230: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0064 from 4000-600cm$^{-1}$**

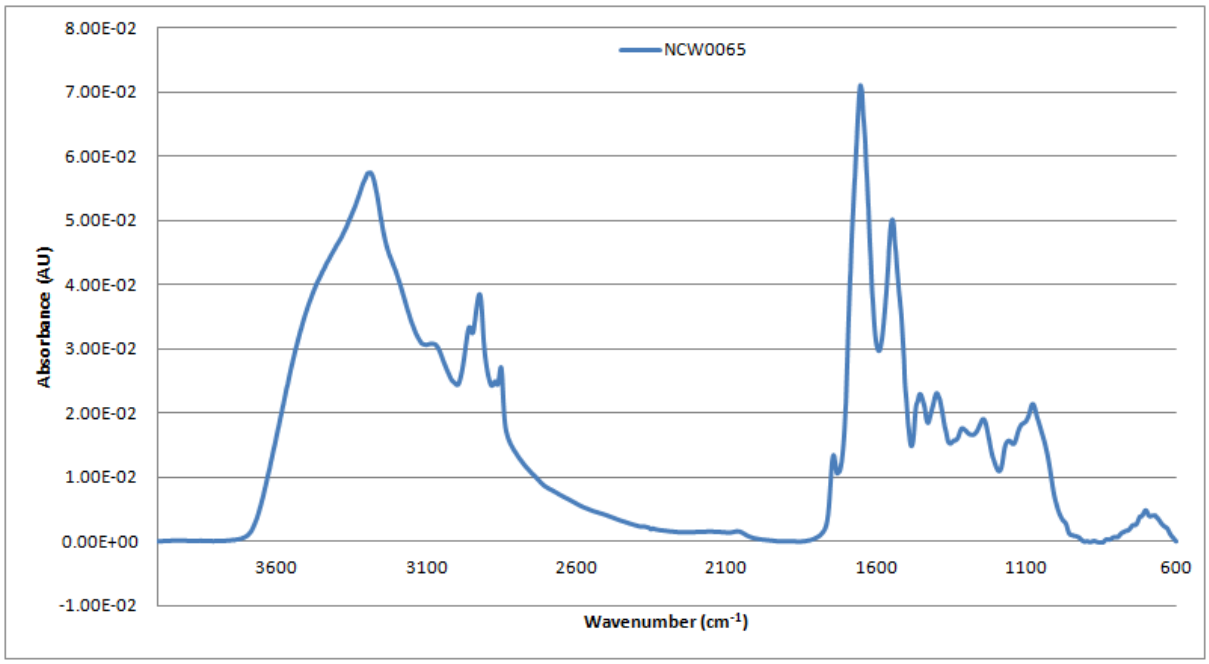


**Figure A2- 231: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0065 from 4000-600cm$^{-1}$**

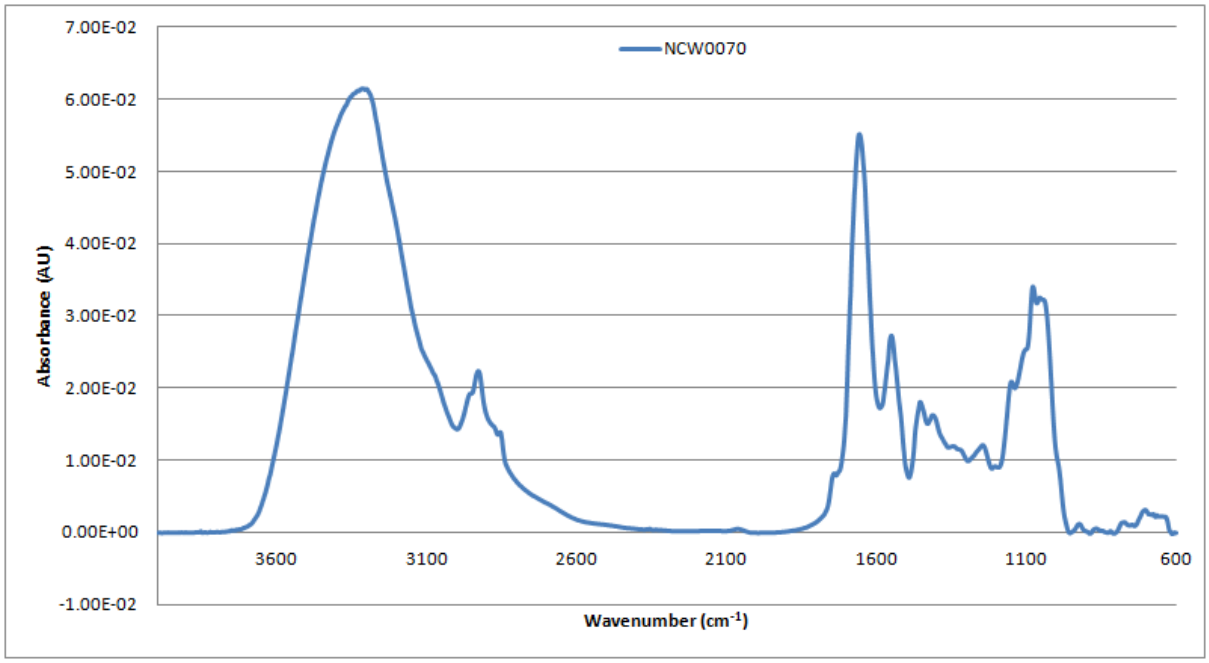


**Figure A2- 232: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0070 from 4000-600cm$^{-1}$**

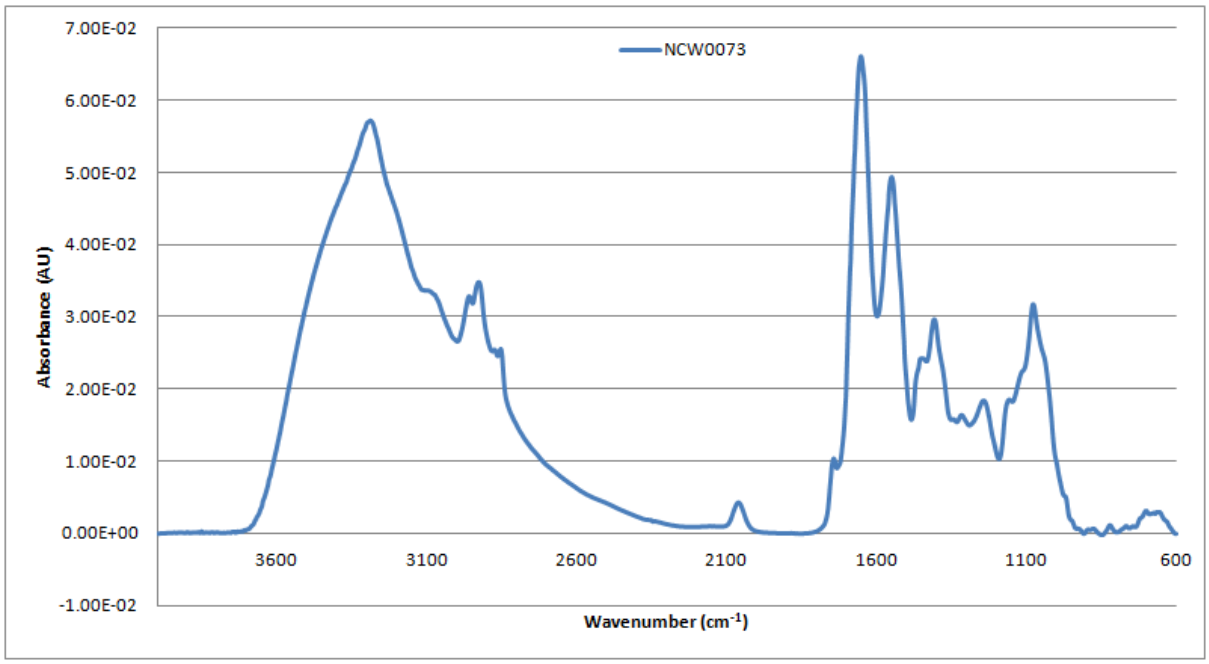


**Figure A2- 233: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0073 from 4000-600cm$^{-1}$**

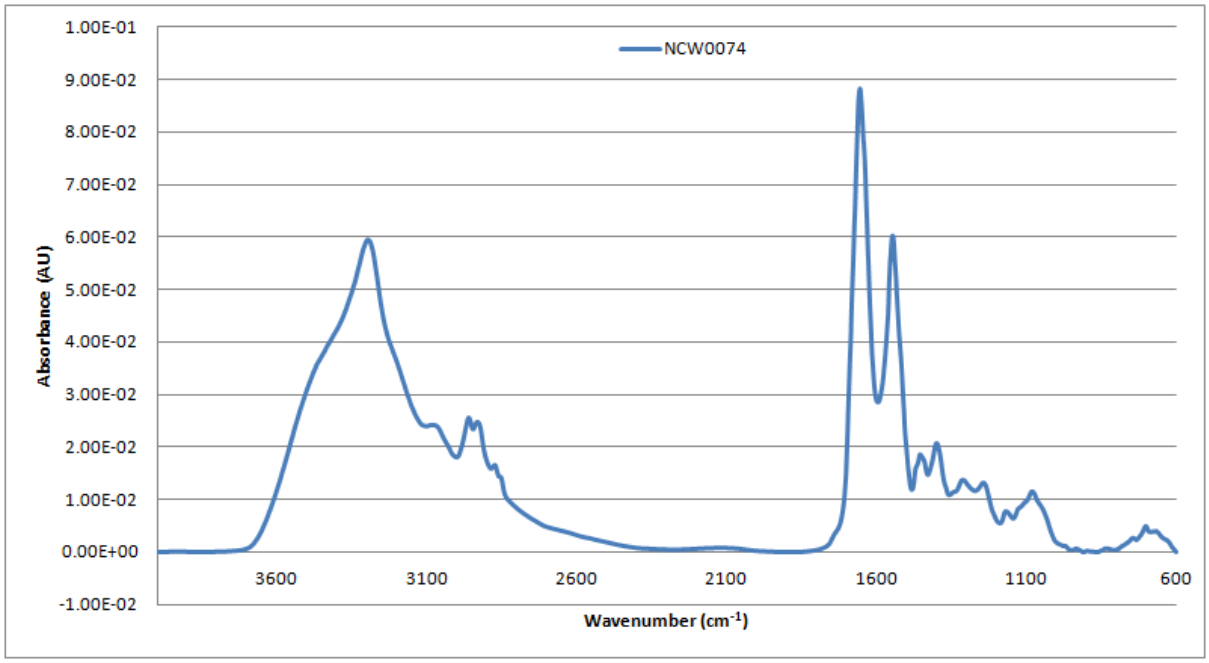


**Figure A2- 234: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0074 from 4000-600cm$^{-1}$**

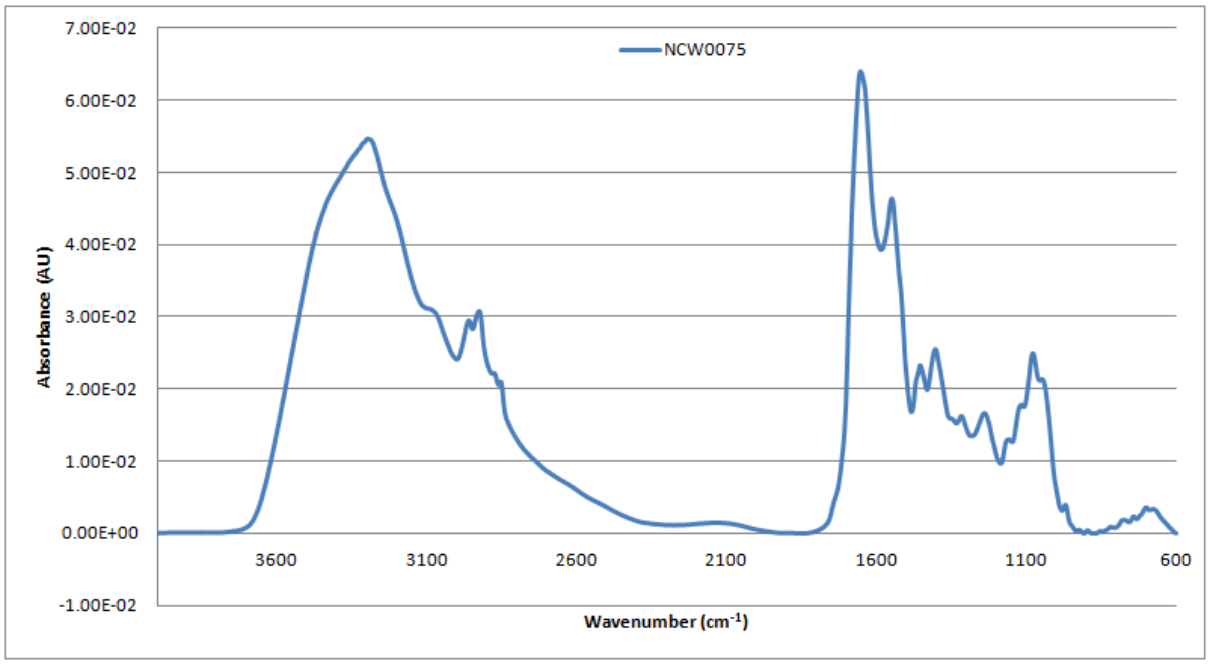


**Figure A2- 235: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0075 from 4000-600cm$^{-1}$**

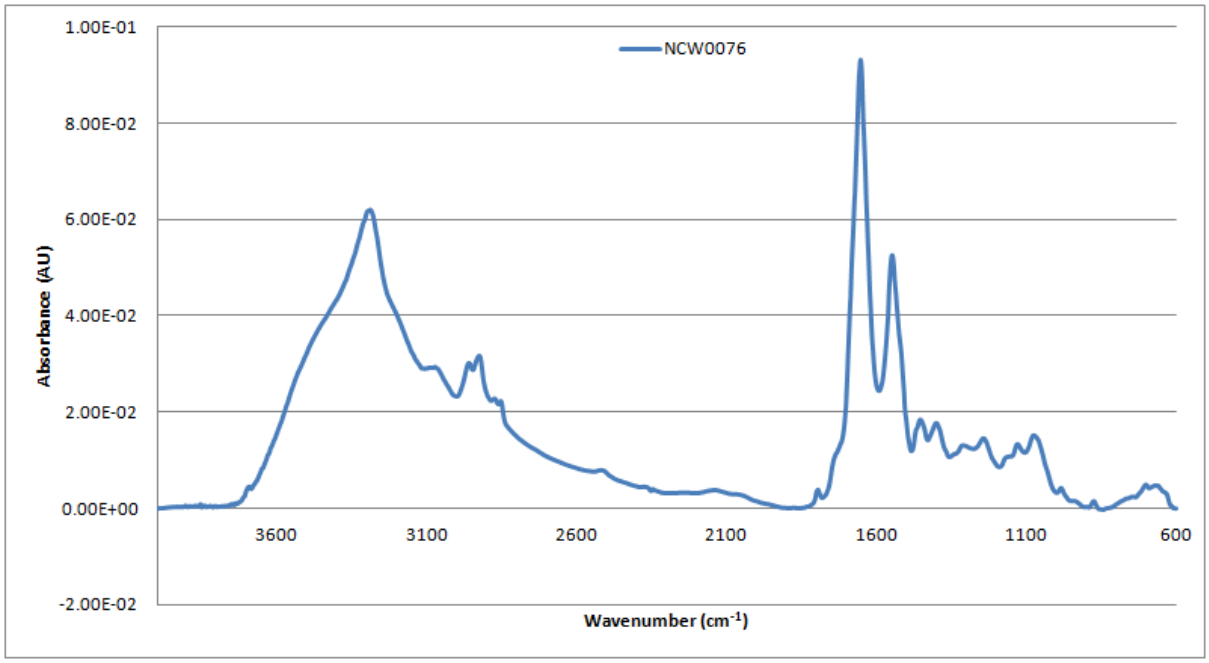


**Figure A2- 236: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0076 from 4000-600cm$^{-1}$**

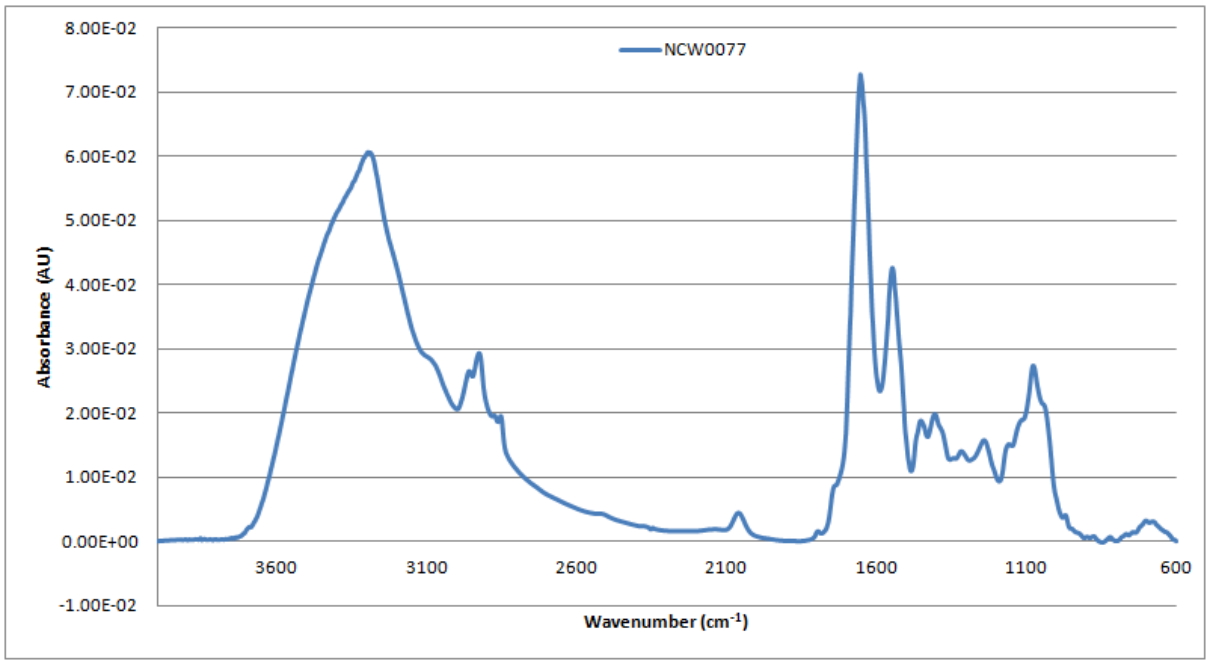


**Figure A2- 237: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0077 from 4000-600cm$^{-1}$**

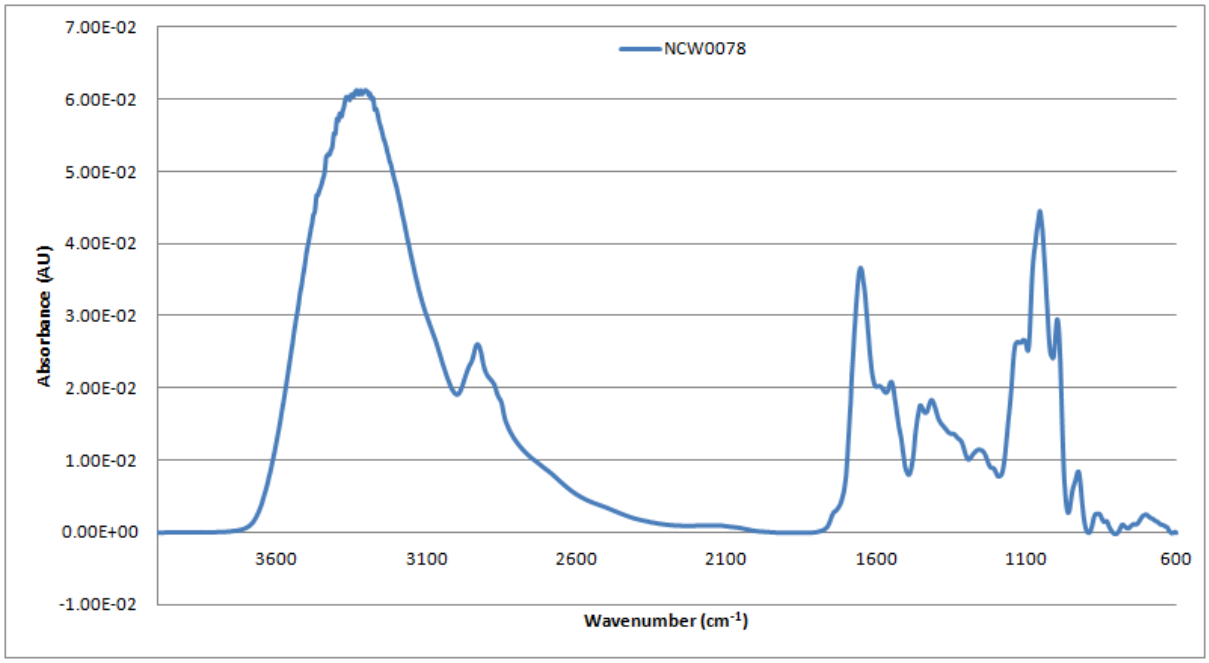


**Figure A2- 238: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0078 from 4000-600cm$^{-1}$**

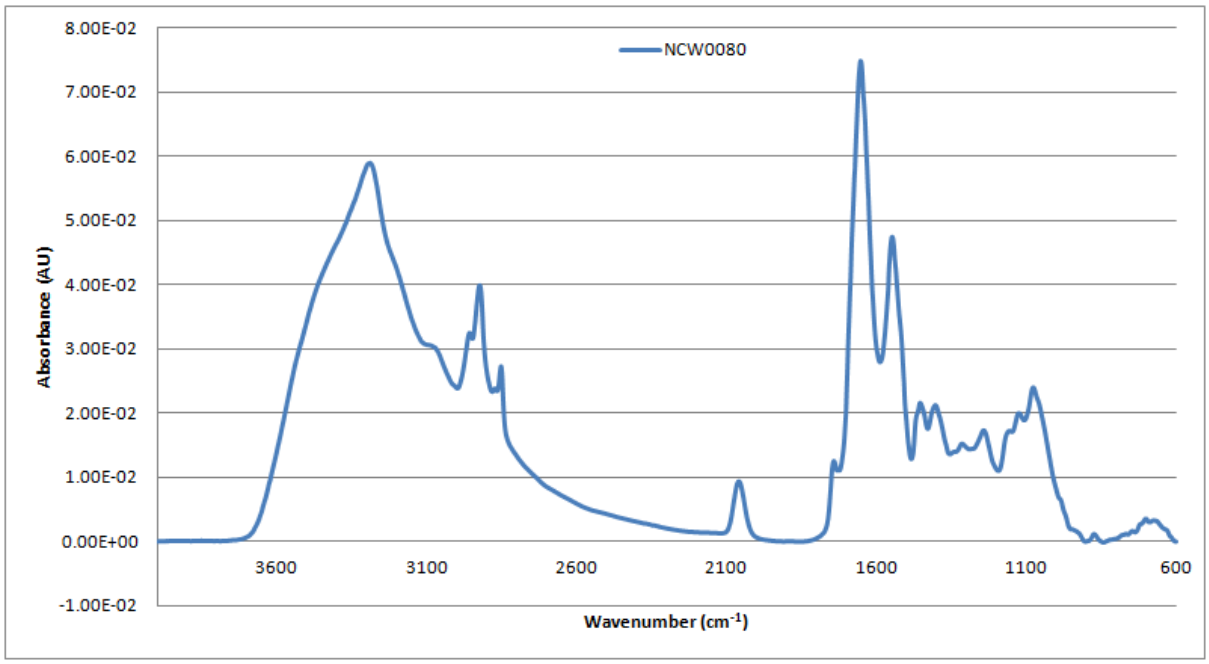


**Figure A2- 239: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0080 from 4000-600cm$^{-1}$**

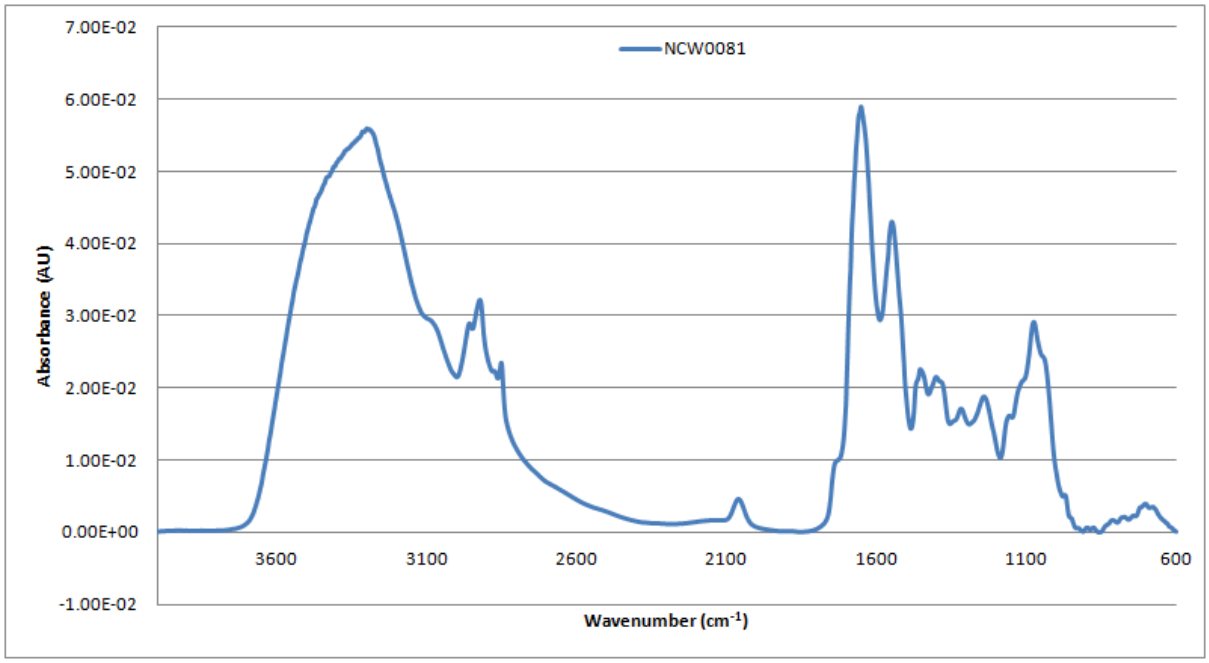


**Figure A2- 240: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0081 from 4000-600cm$^{-1}$**

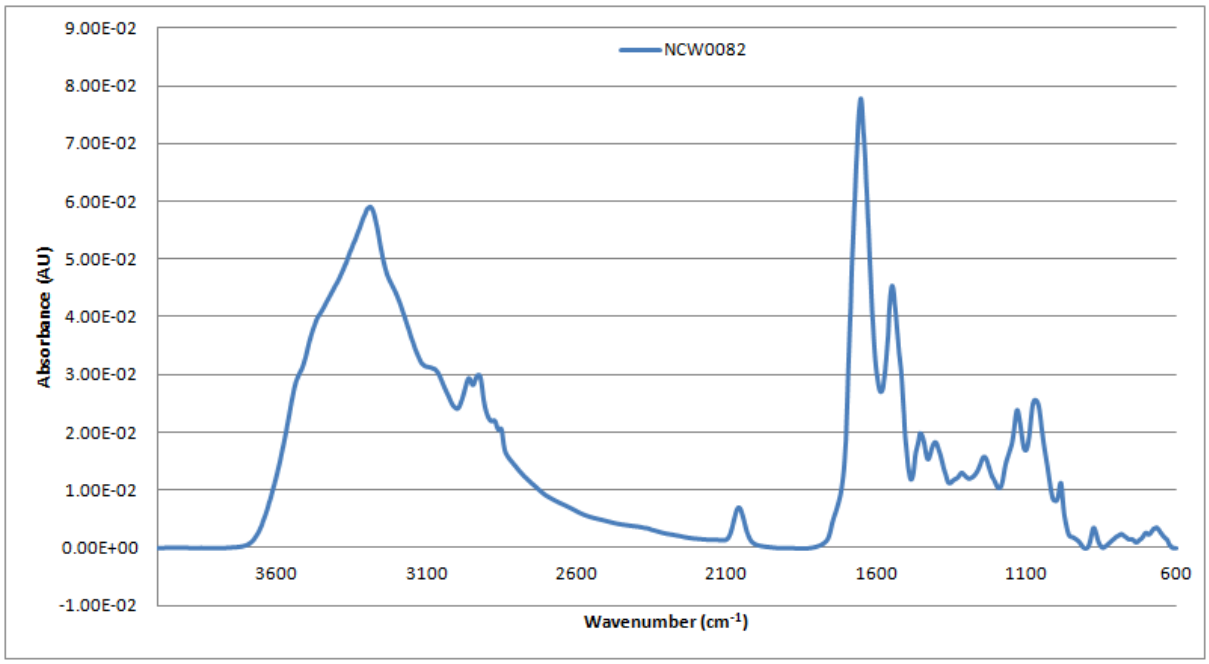


**Figure A2- 241: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0082 from 4000-600cm$^{-1}$**

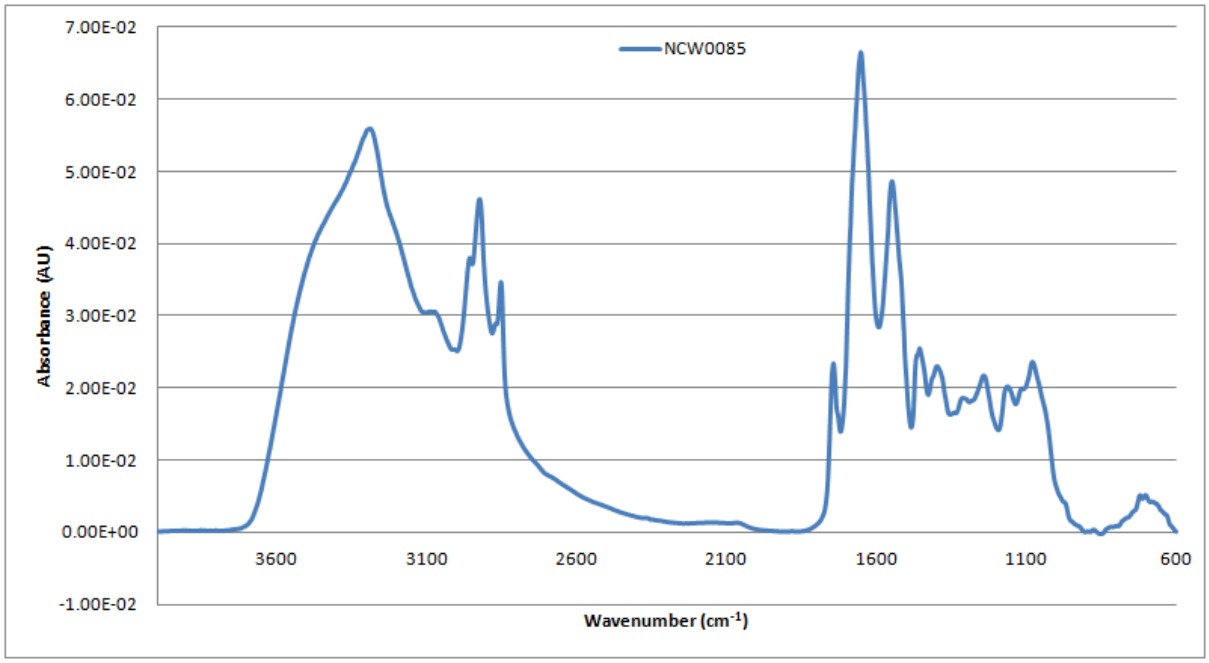


**Figure A2- 242: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0085 from 4000-600cm$^{-1}$**

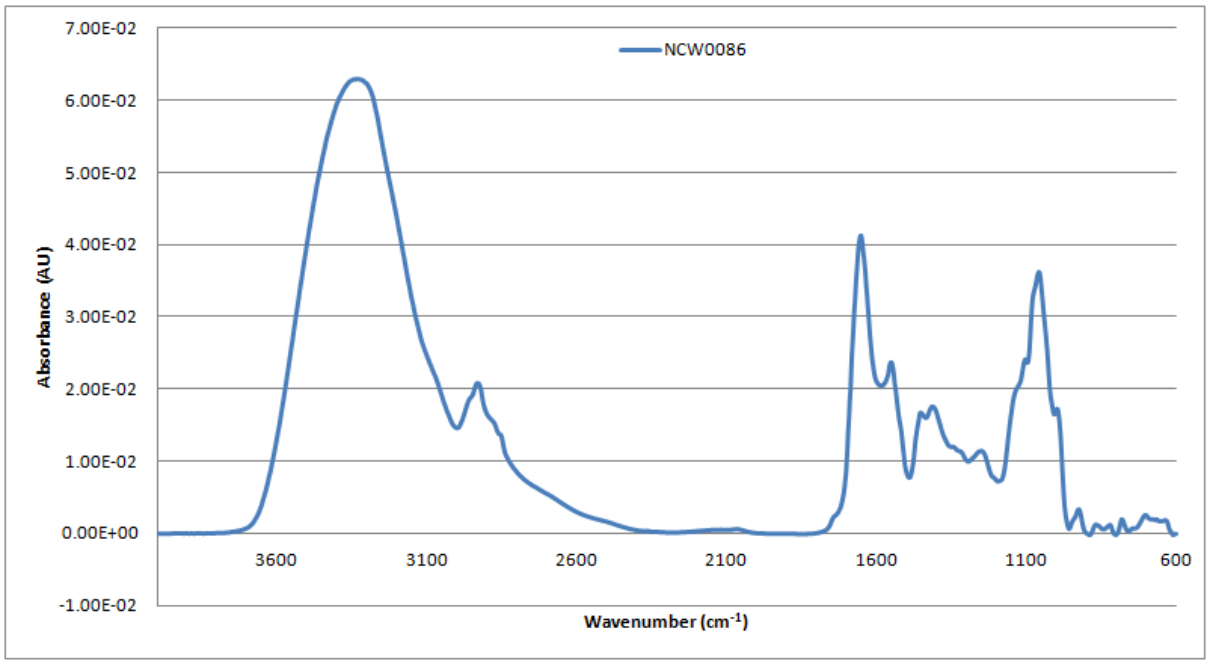


**Figure A2- 243: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0086 from 4000-600cm$^{-1}$**

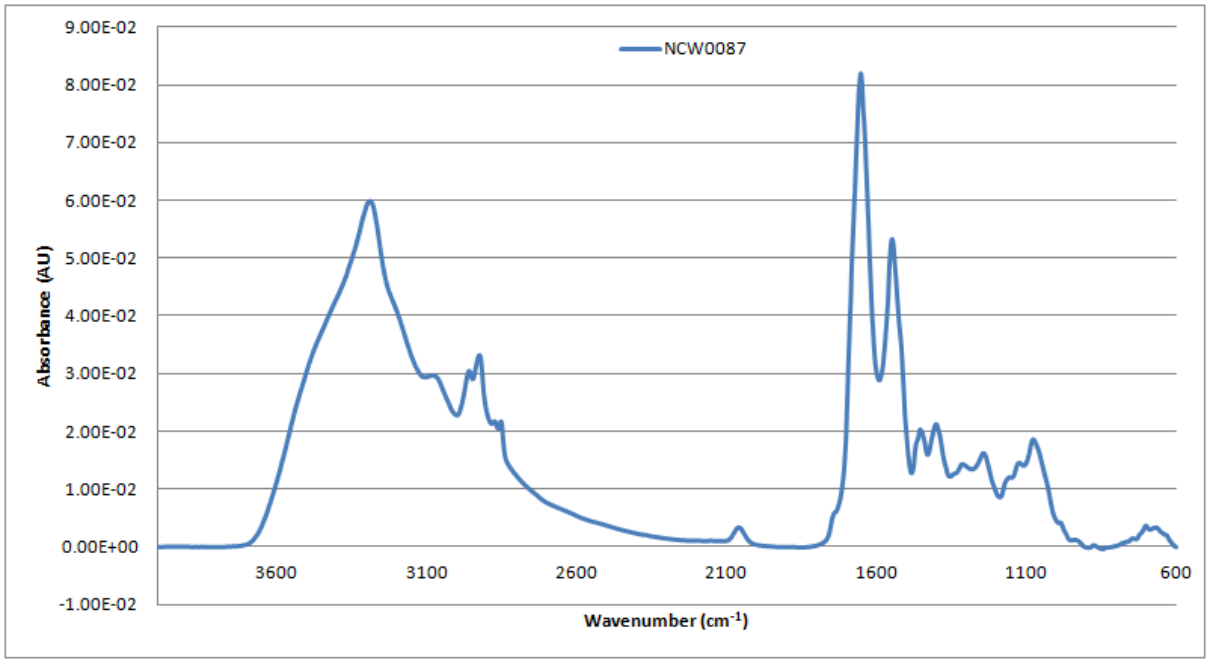


**Figure A2- 244: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0087 from 4000-600cm$^{-1}$**

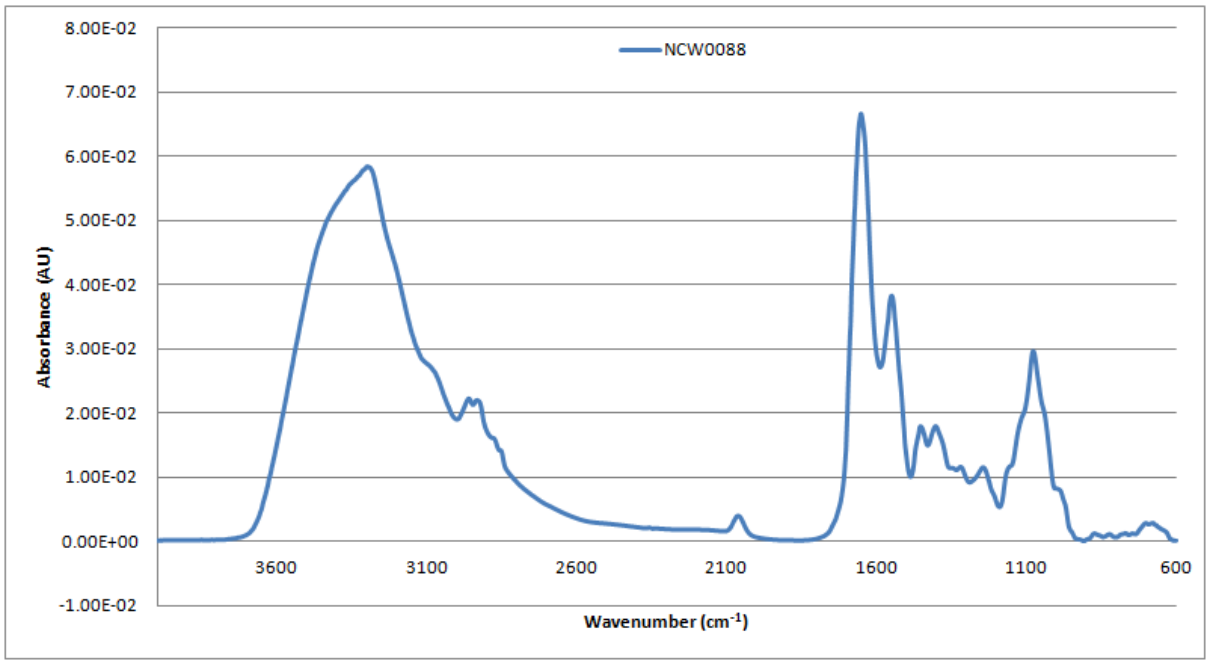


**Figure A2- 245: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0088 from 4000-600cm$^{-1}$**

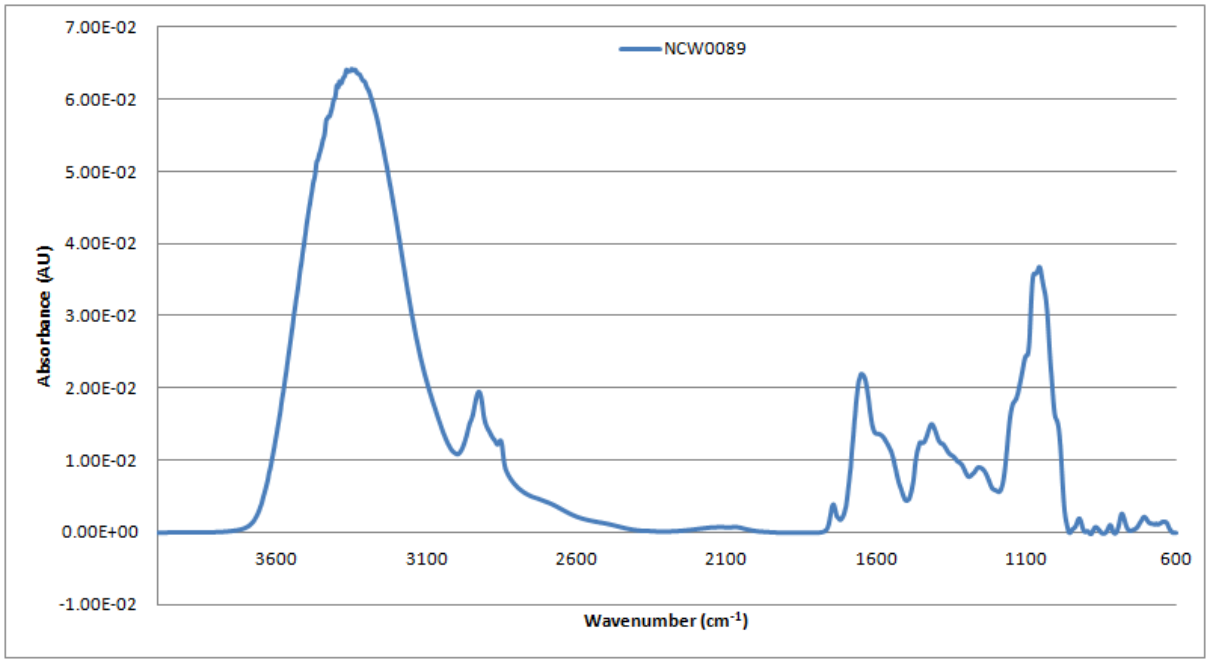


**Figure A2- 246: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0089 from 4000-600cm$^{-1}$**

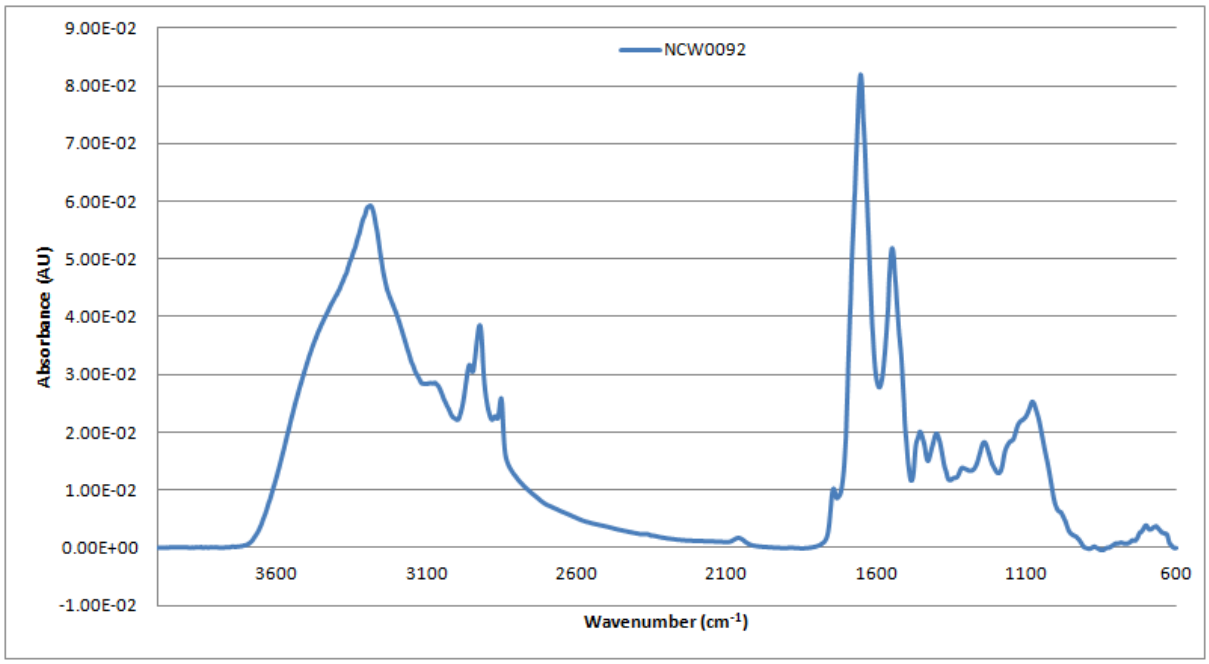


**Figure A2- 247: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0092 from 4000-600cm$^{-1}$**

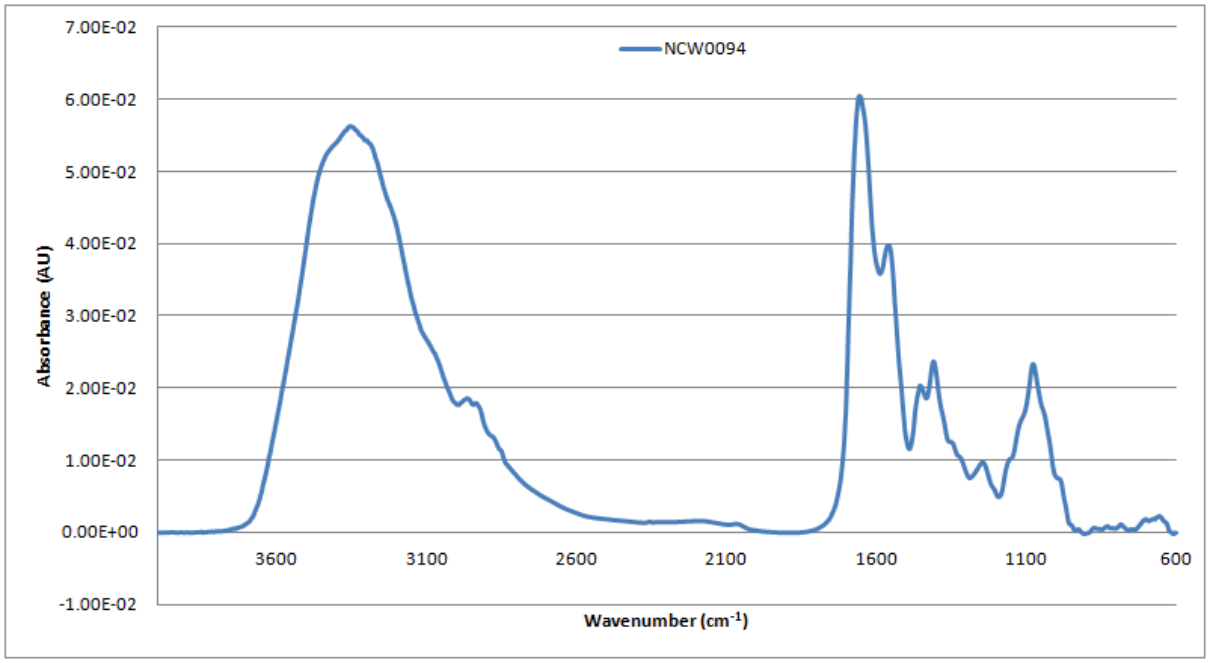


**Figure A2- 248: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0094 from 4000-600cm$^{-1}$**

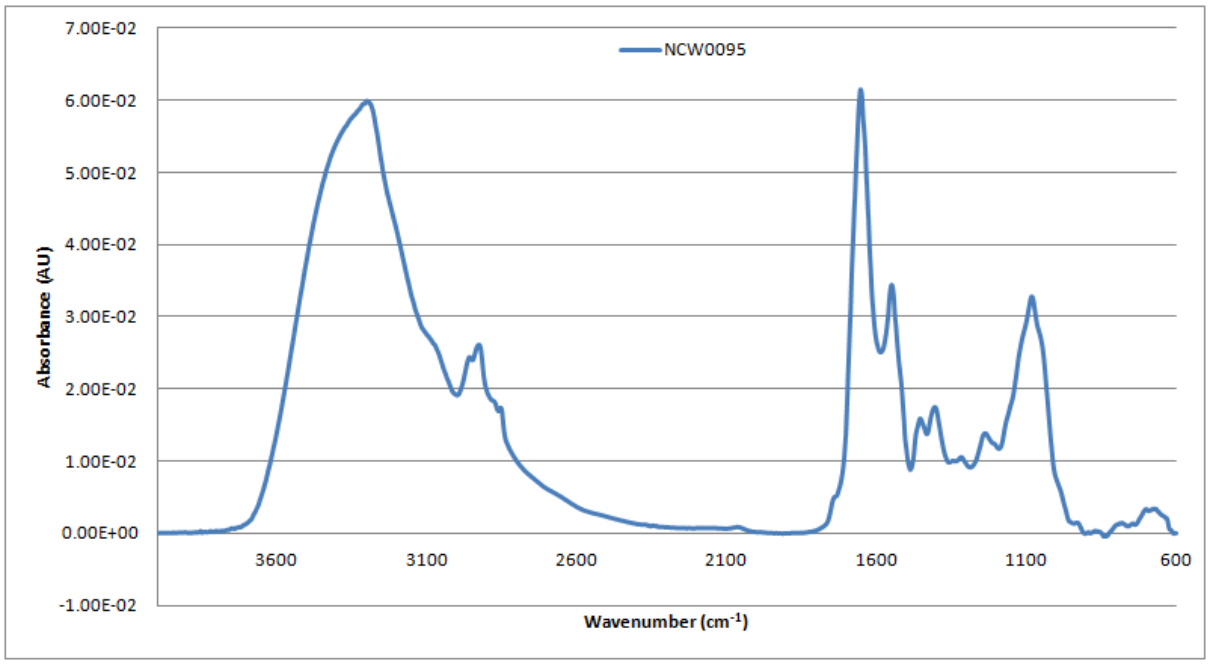


**Figure A2- 249: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0095 from 4000-600cm$^{-1}$**

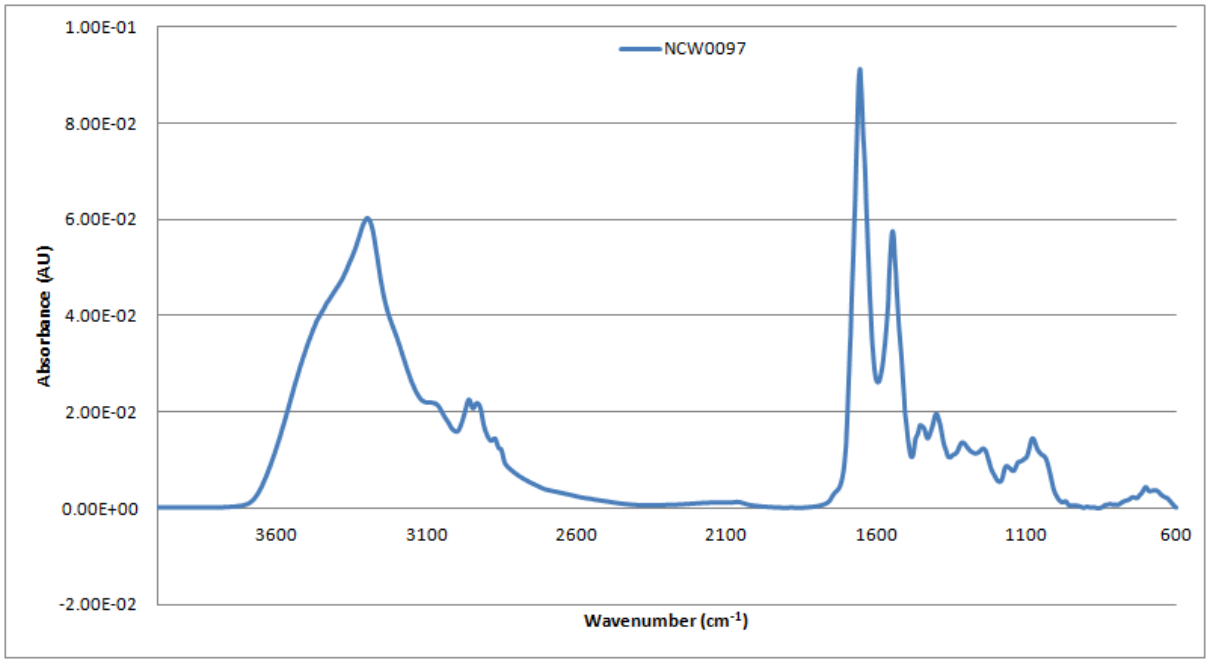


**Figure A2- 250: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0097 from 4000-600cm$^{-1}$**

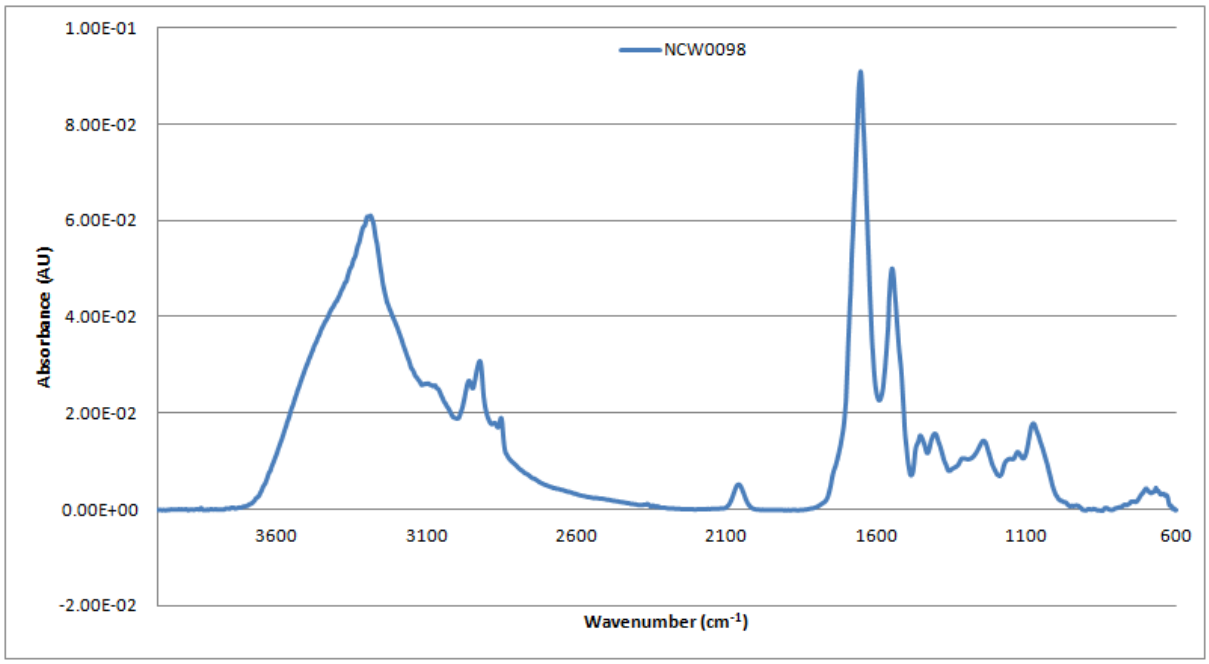


**Figure A2- 251: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0098 from 4000-600cm$^{-1}$**

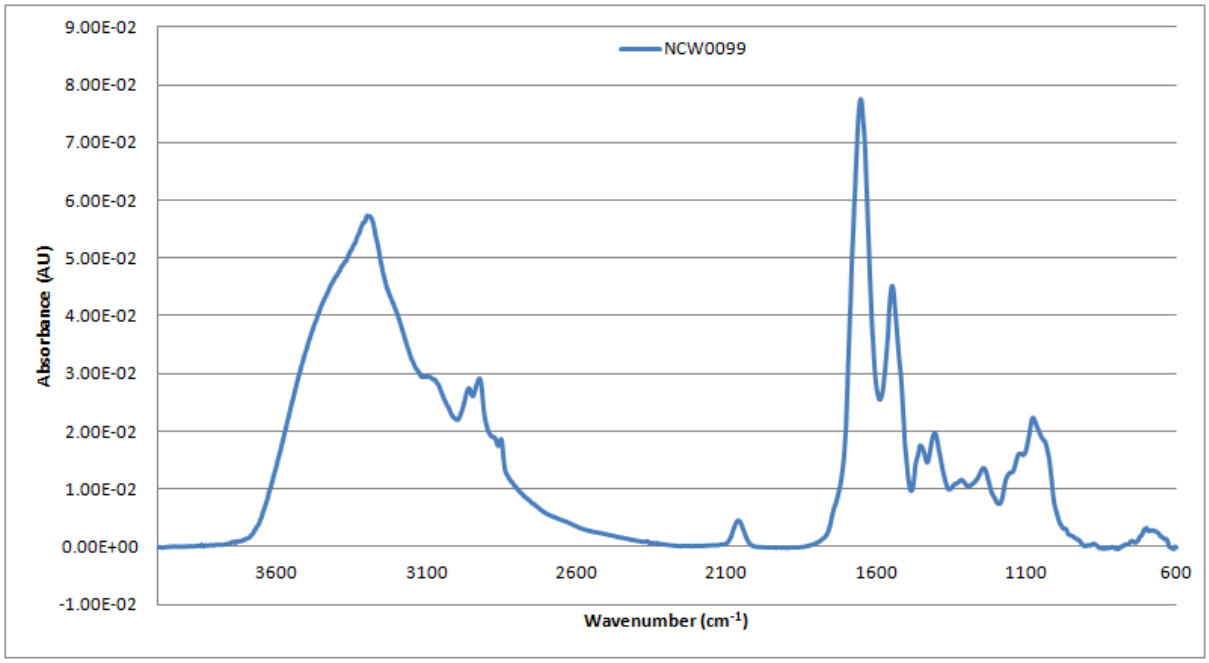


**Figure A2- 252: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0099 from 4000-600cm$^{-1}$**

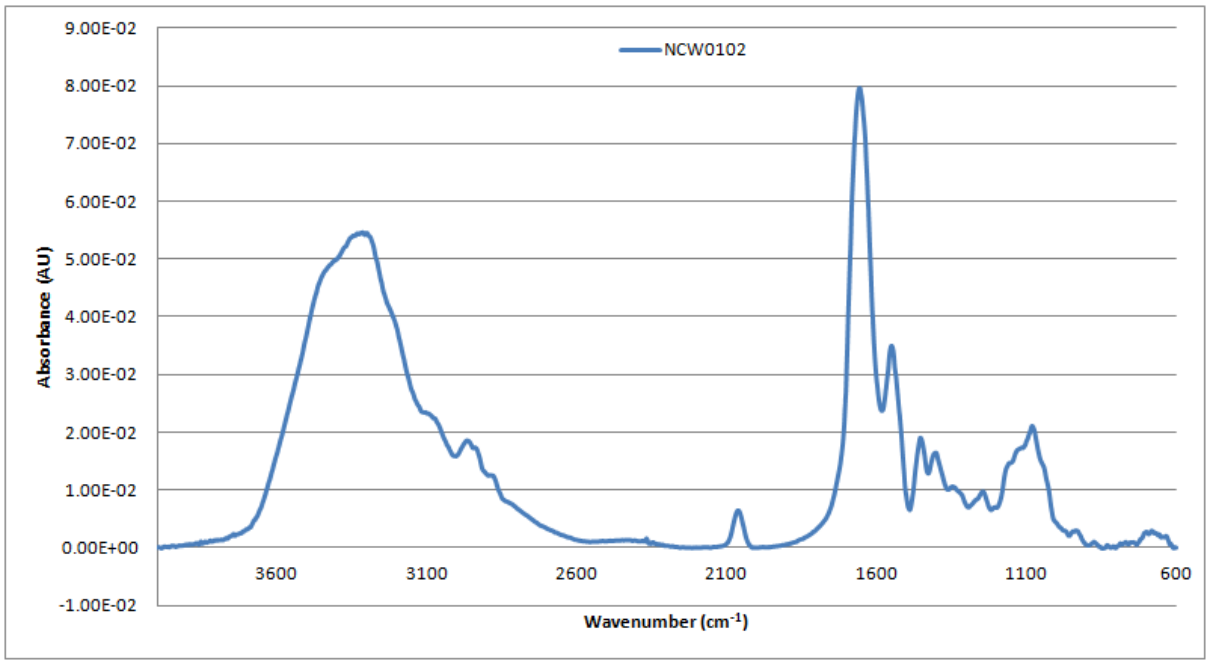


**Figure A2- 253: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0102 from 4000-600cm$^{-1}$**

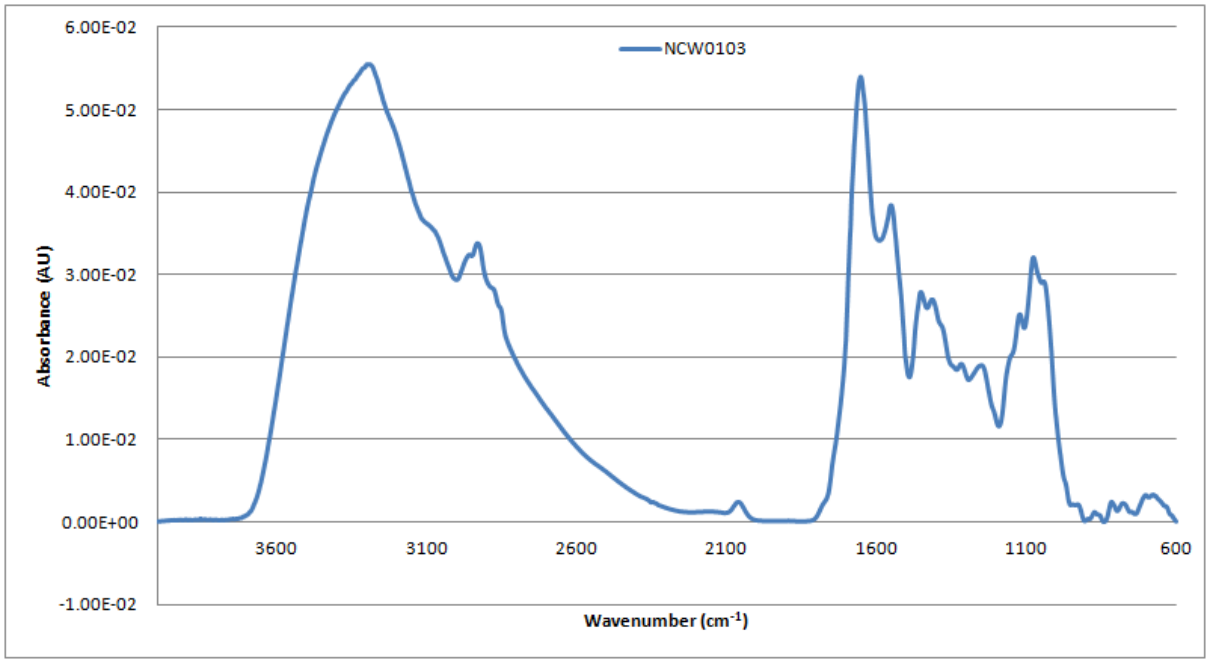


**Figure A2- 254: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0103 from 4000-600cm$^{-1}$**

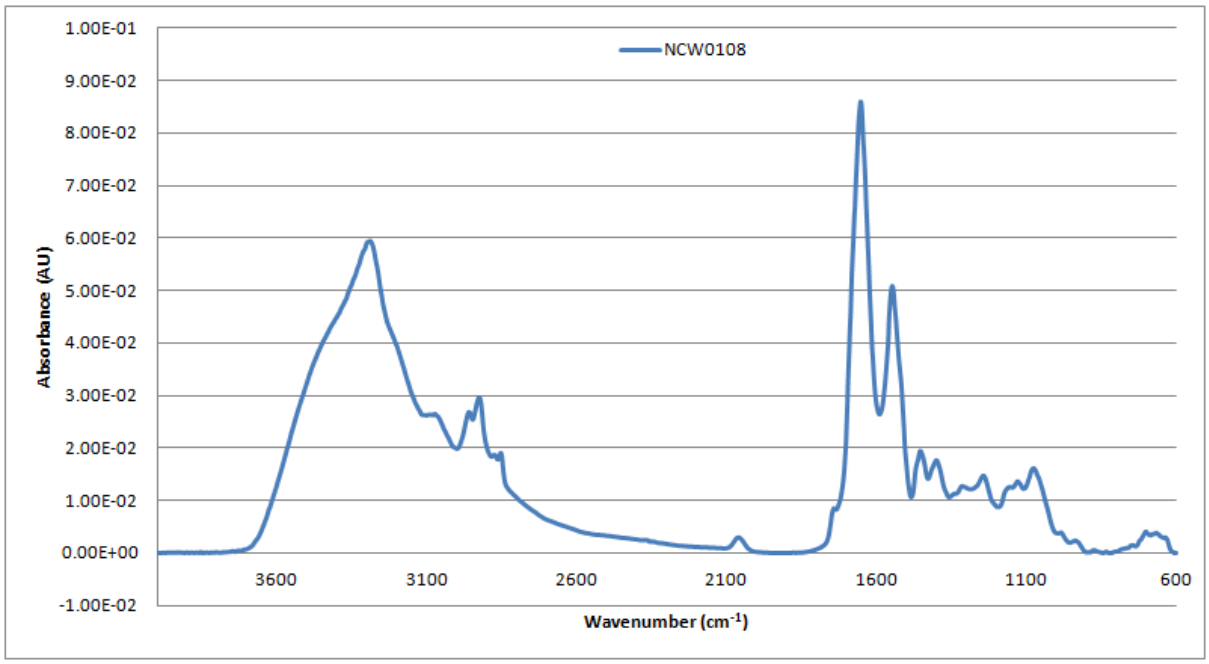


**Figure A2- 255: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0108 from 4000-600cm$^{-1}$**

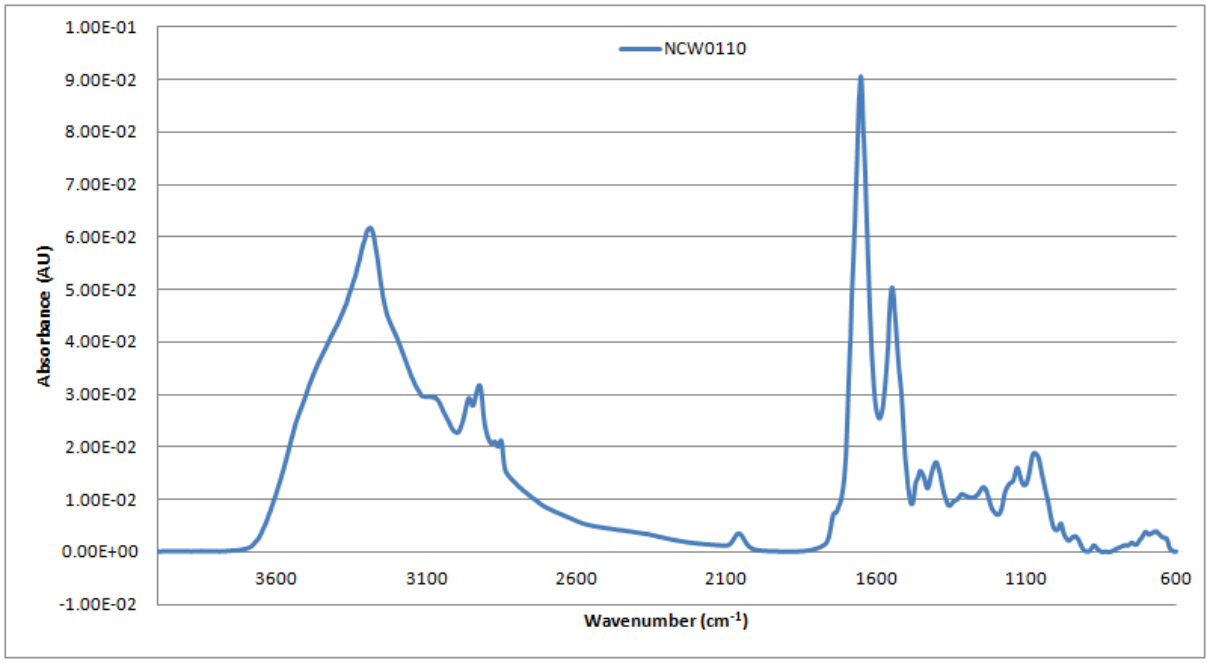


**Figure A2- 256: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0110 from 4000-600cm$^{-1}$**

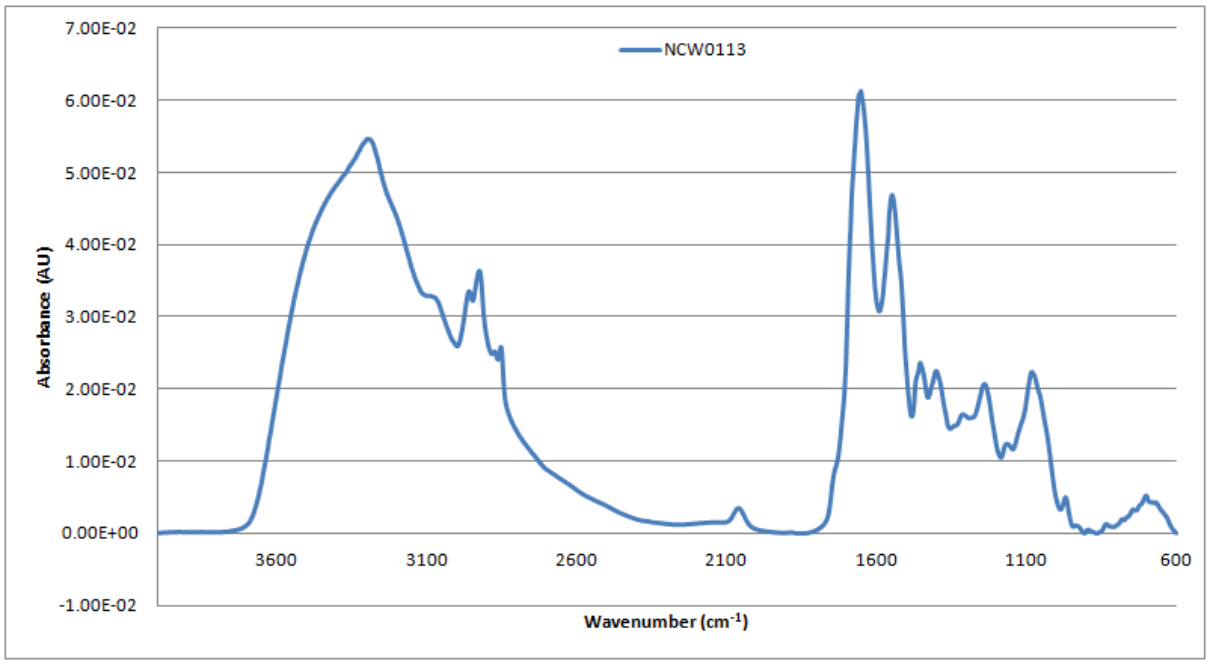


**Figure A2- 257: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0113 from 4000-600cm$^{-1}$**

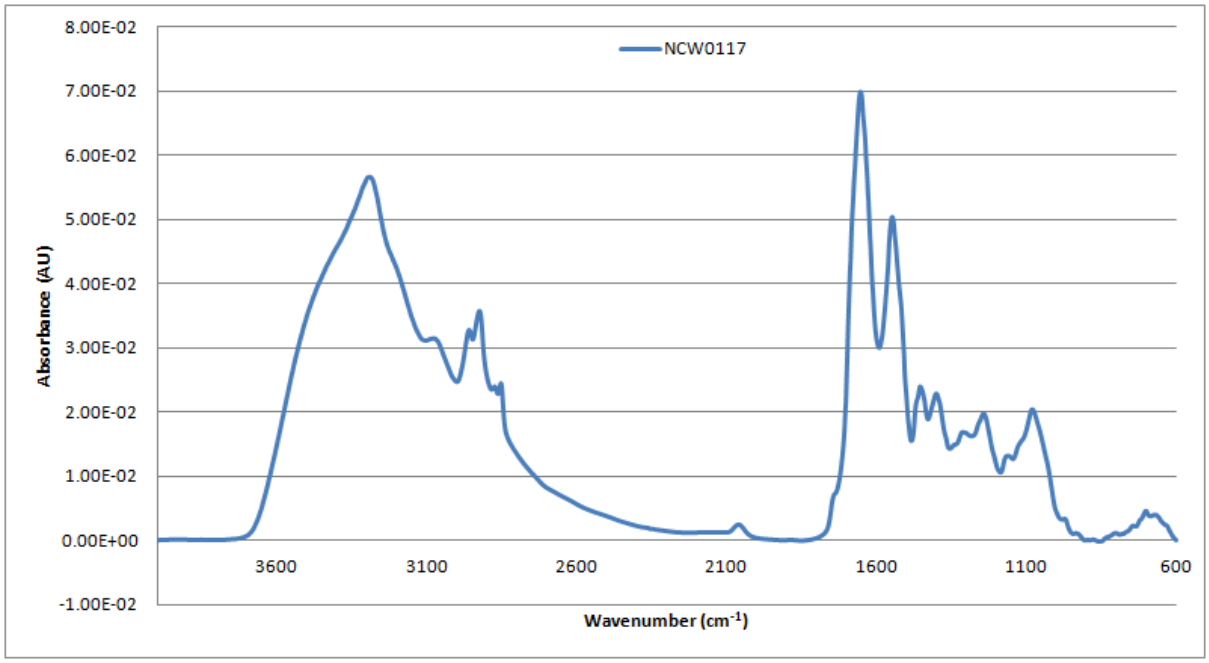


**Figure A2- 258: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0117 from 4000-600cm$^{-1}$**

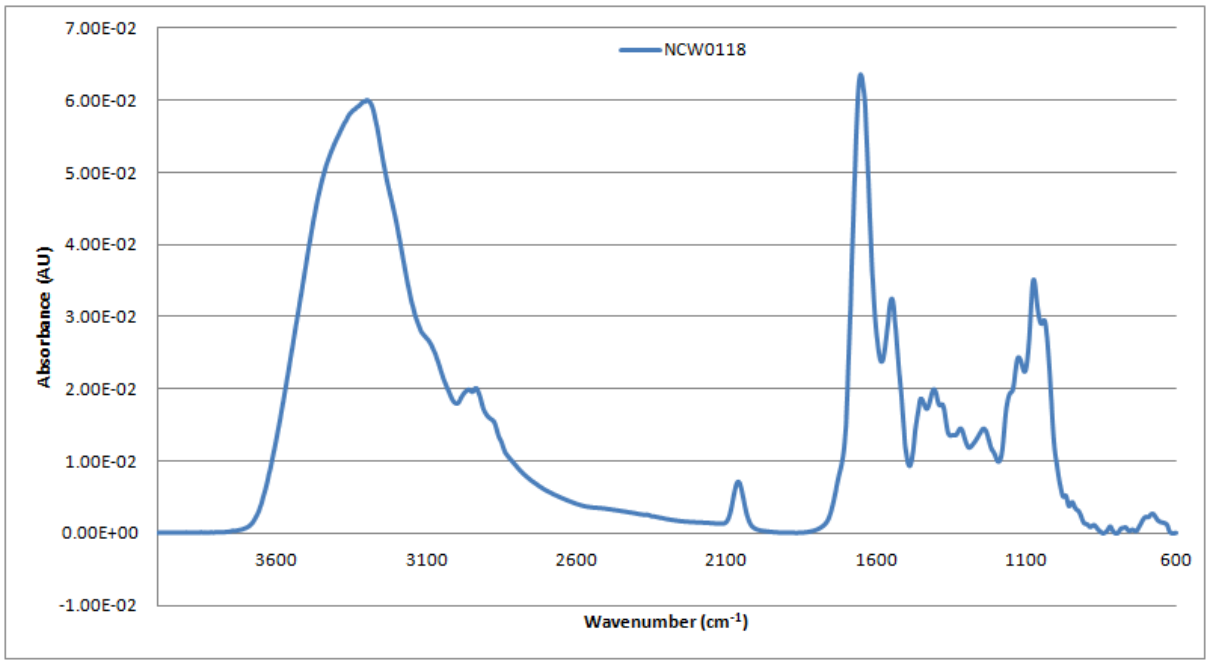


**Figure A2- 259: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0118 from 4000-600cm$^{-1}$**

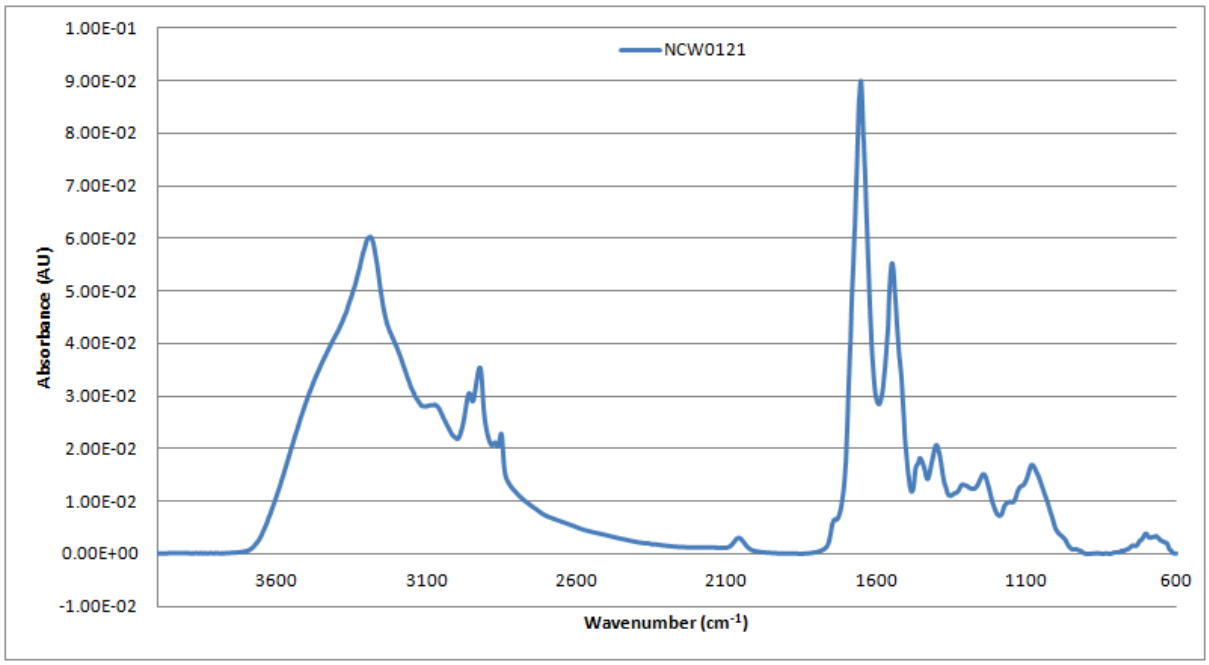


**Figure A2- 260: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0121 from 4000-600cm$^{-1}$**

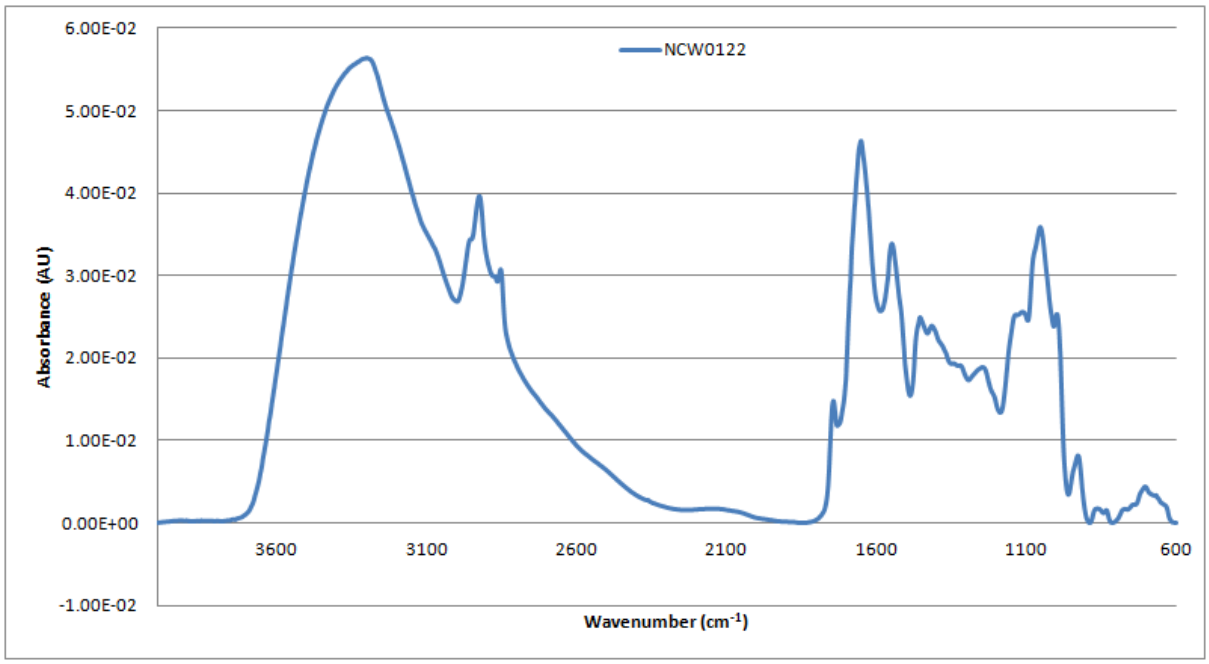


**Figure A2- 261: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0122 from 4000-600cm$^{-1}$**

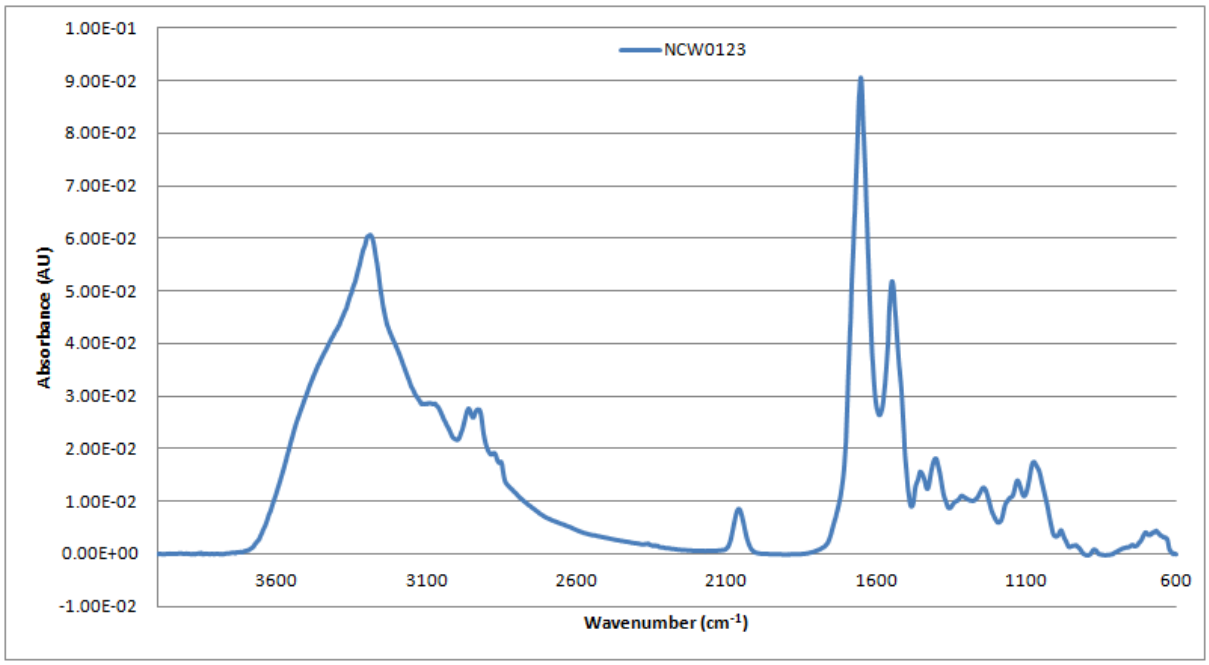


**Figure A2- 262: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0123 from 4000-600cm$^{-1}$**

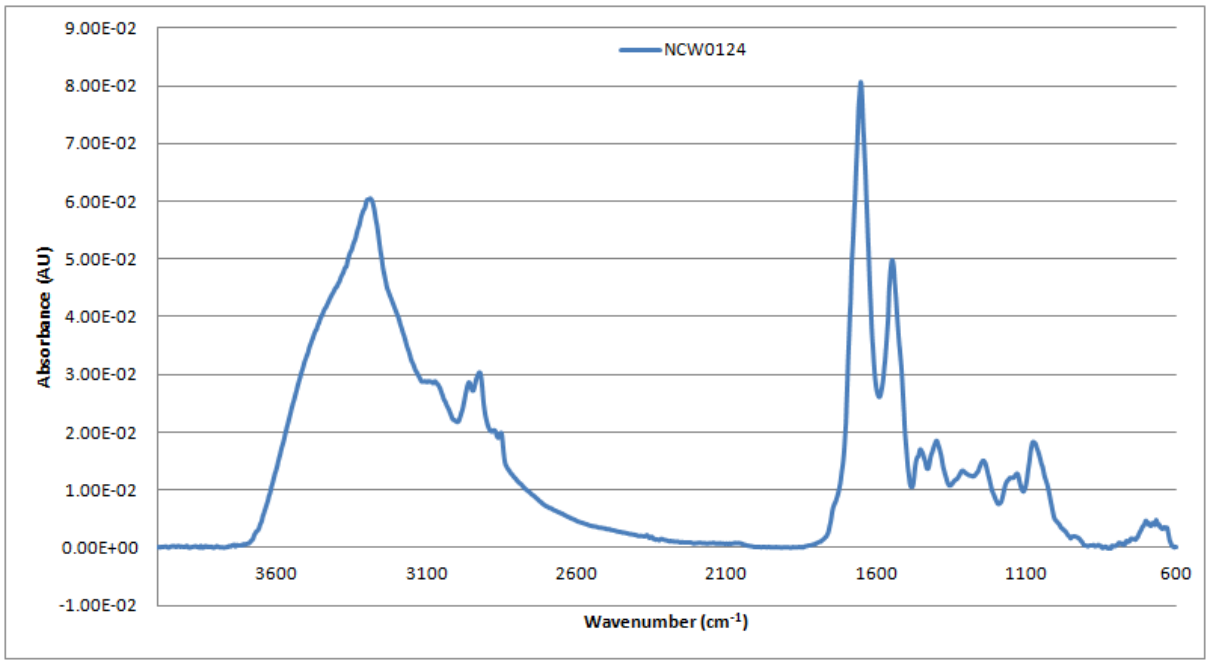


**Figure A2- 263: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0124 from 4000-600cm$^{-1}$**

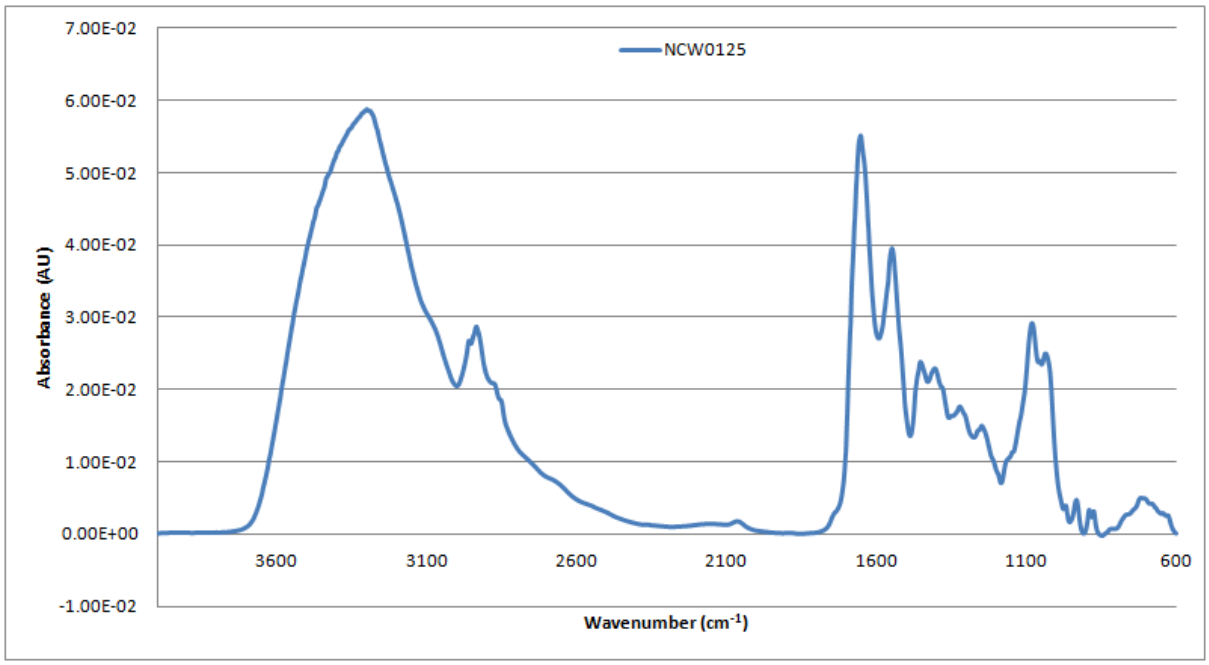


**Figure A2- 264: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0125 from 4000-600cm$^{-1}$**

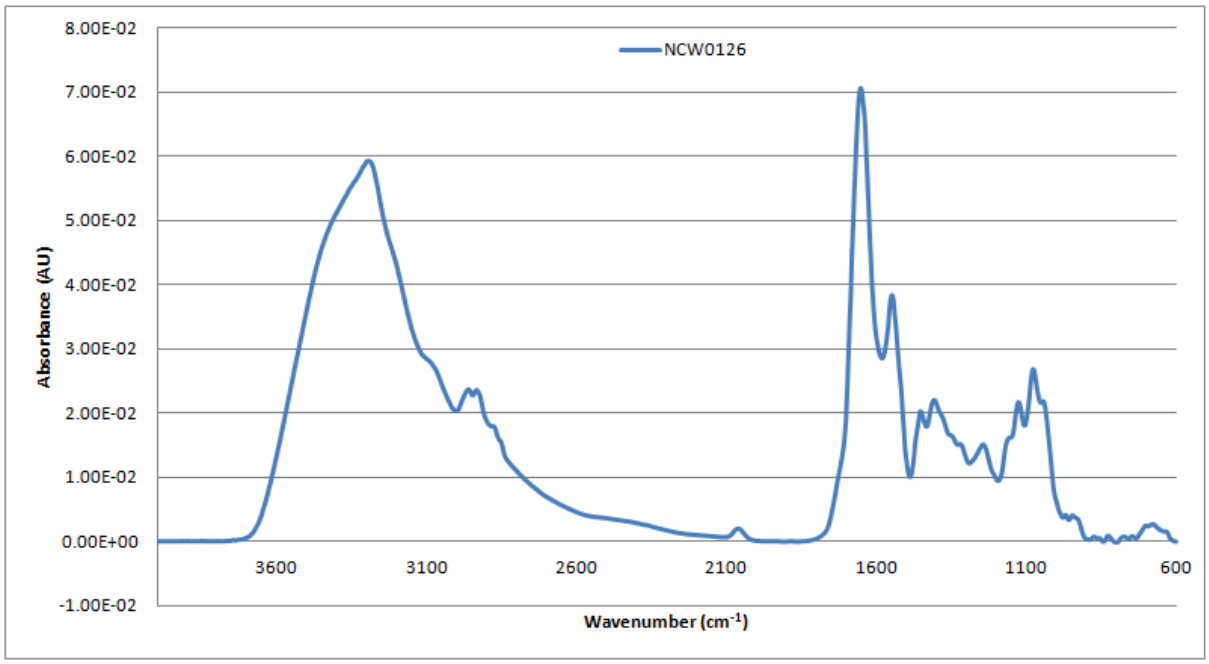


**Figure A2- 265: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0126 from 4000-600cm$^{-1}$**

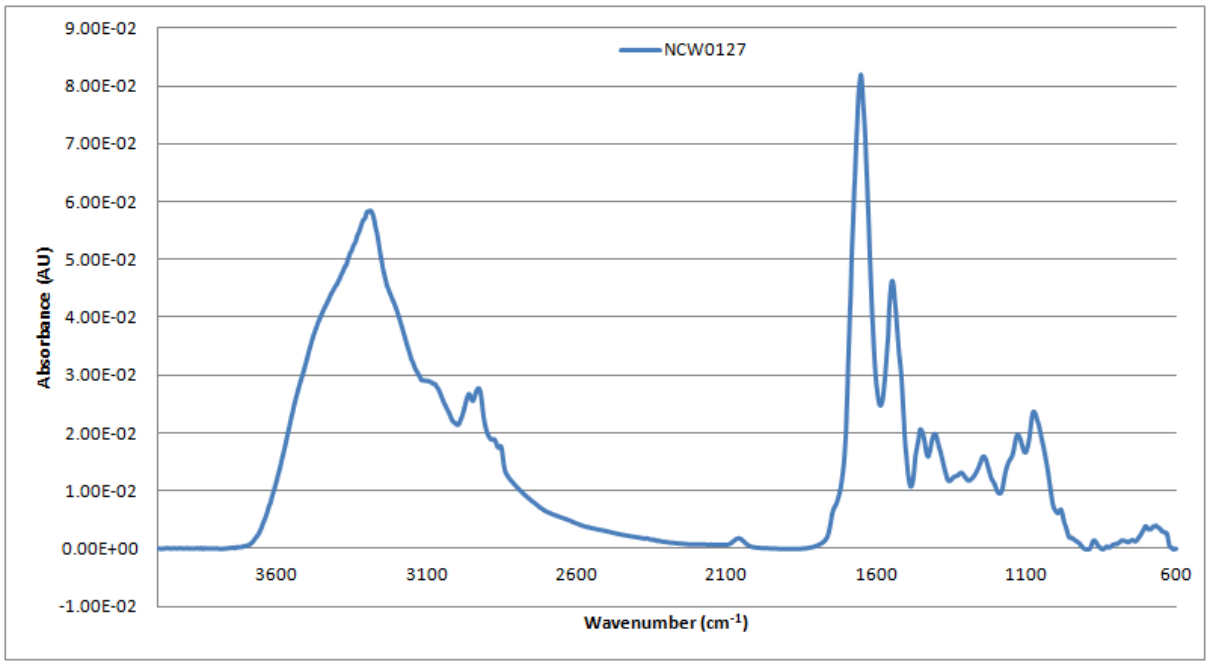


**Figure A2- 266: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0127 from 4000-600cm$^{-1}$**

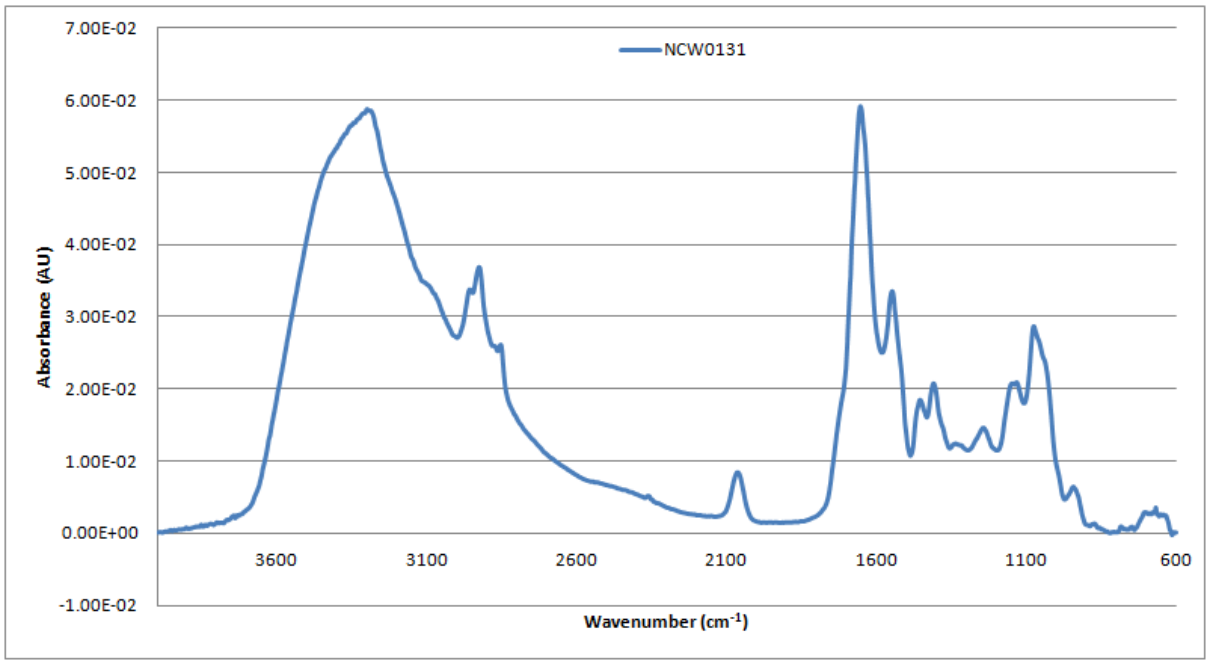


**Figure A2- 267: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0131 from 4000-600cm$^{-1}$**

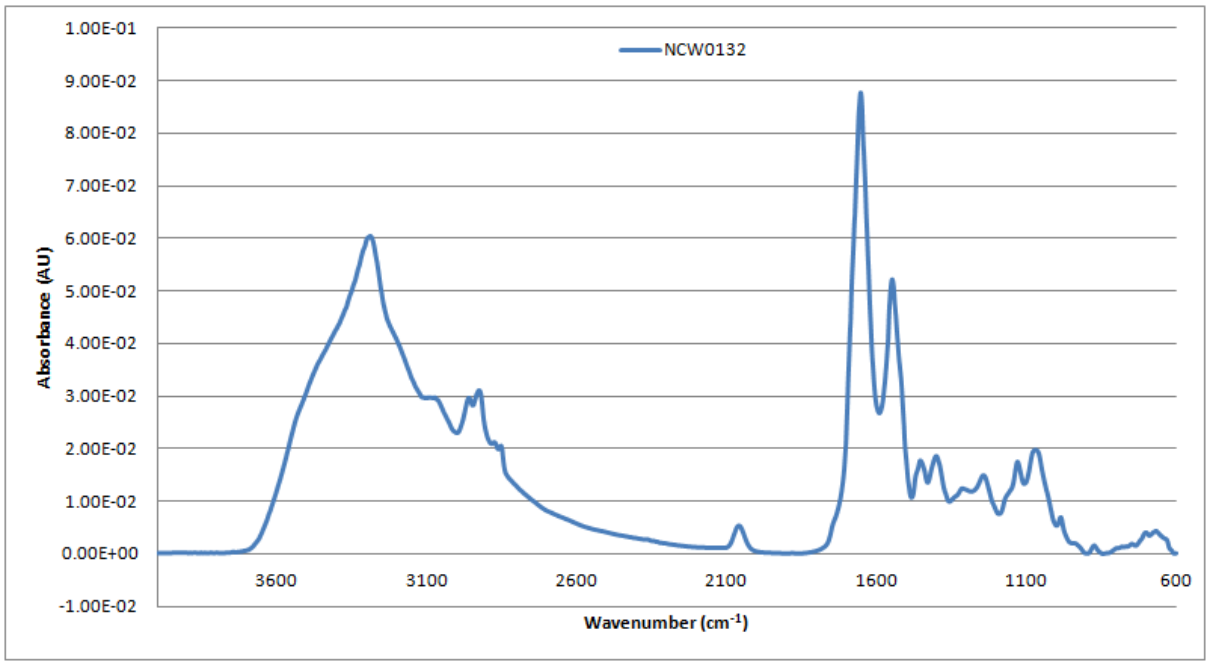


**Figure A2- 268: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0132 from 4000-600cm$^{-1}$**

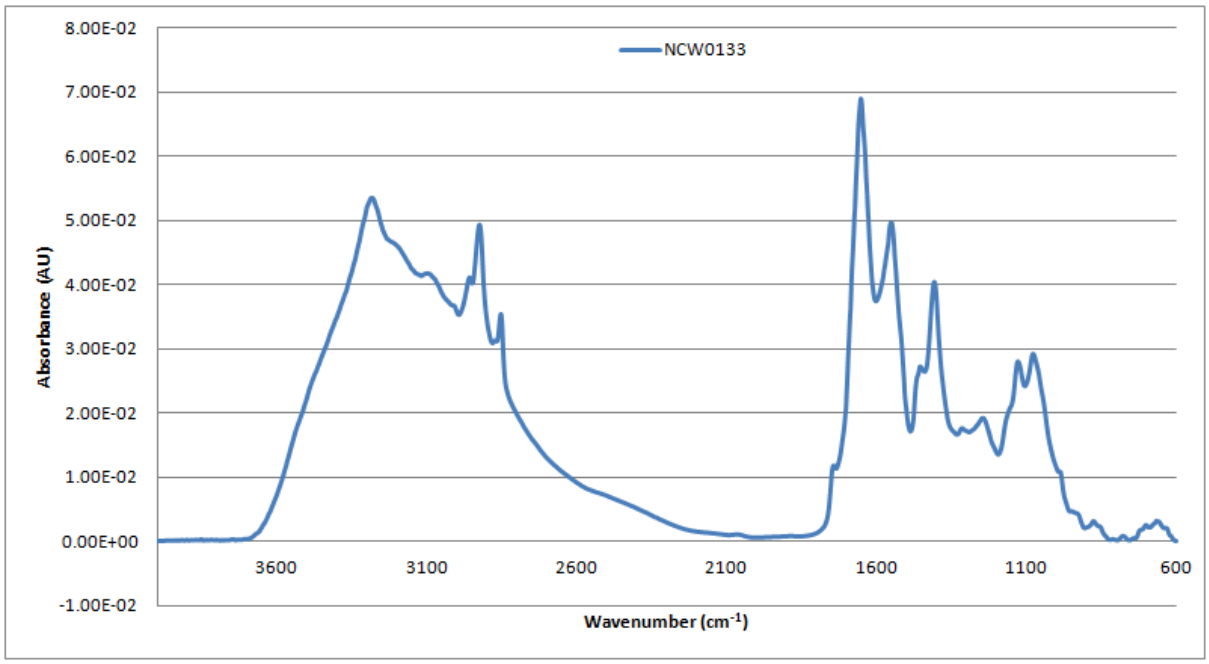


**Figure A2- 269: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient NCW0133 from 4000-600cm$^{-1}$**

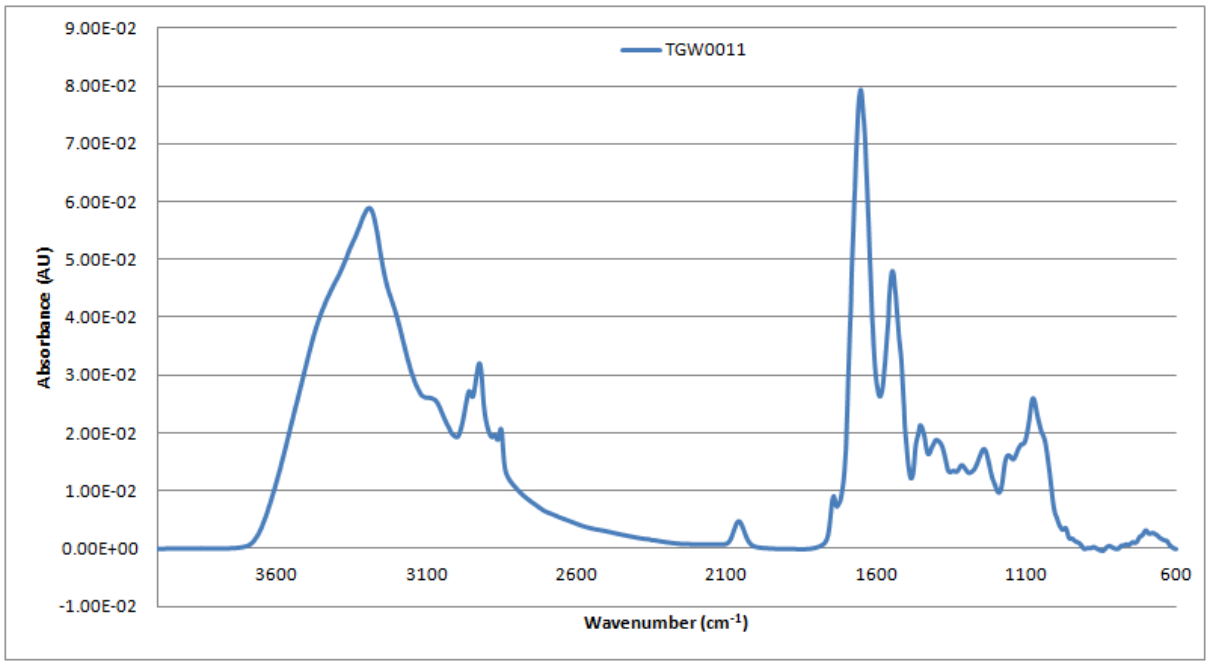


**Figure A2- 270: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0011 from 4000-600cm$^{-1}$**

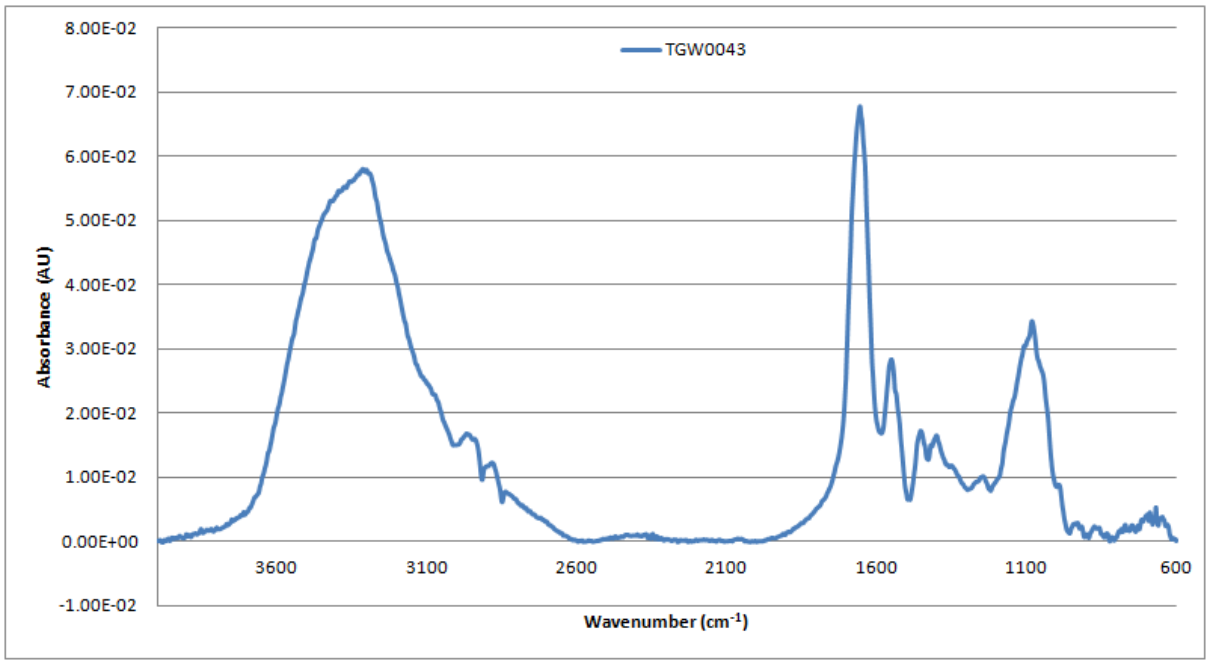


**Figure A2- 271: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0043 from 4000-600cm$^{-1}$**

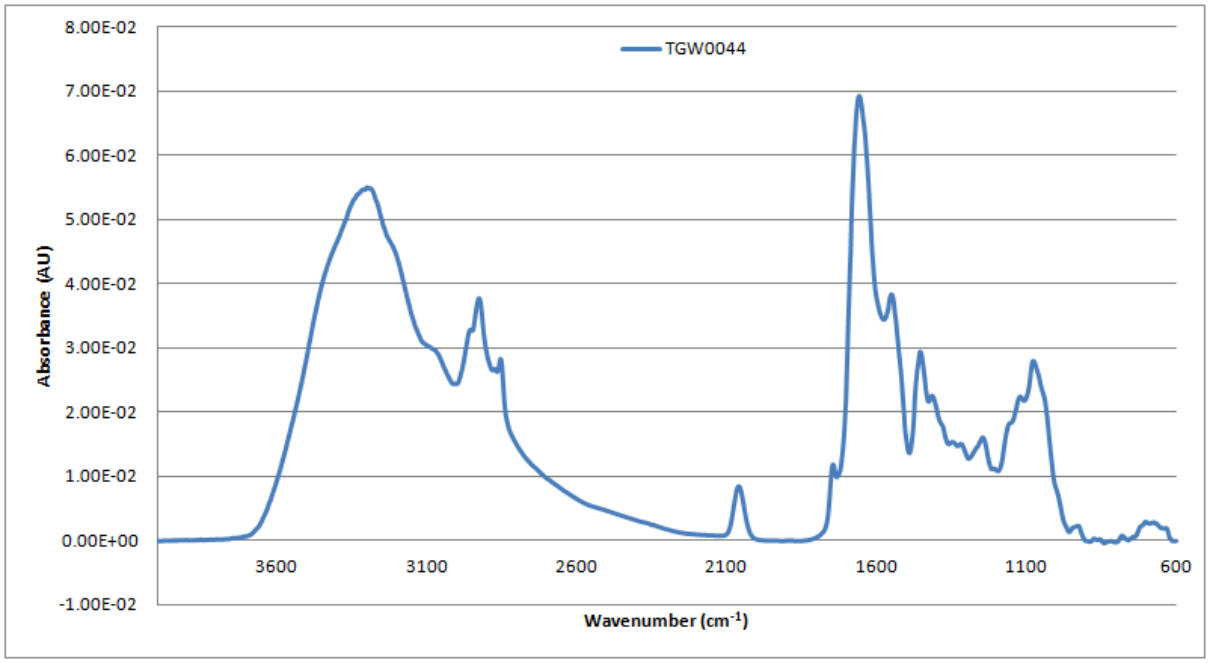


**Figure A2- 272: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0044 from 4000-600cm$^{-1}$**

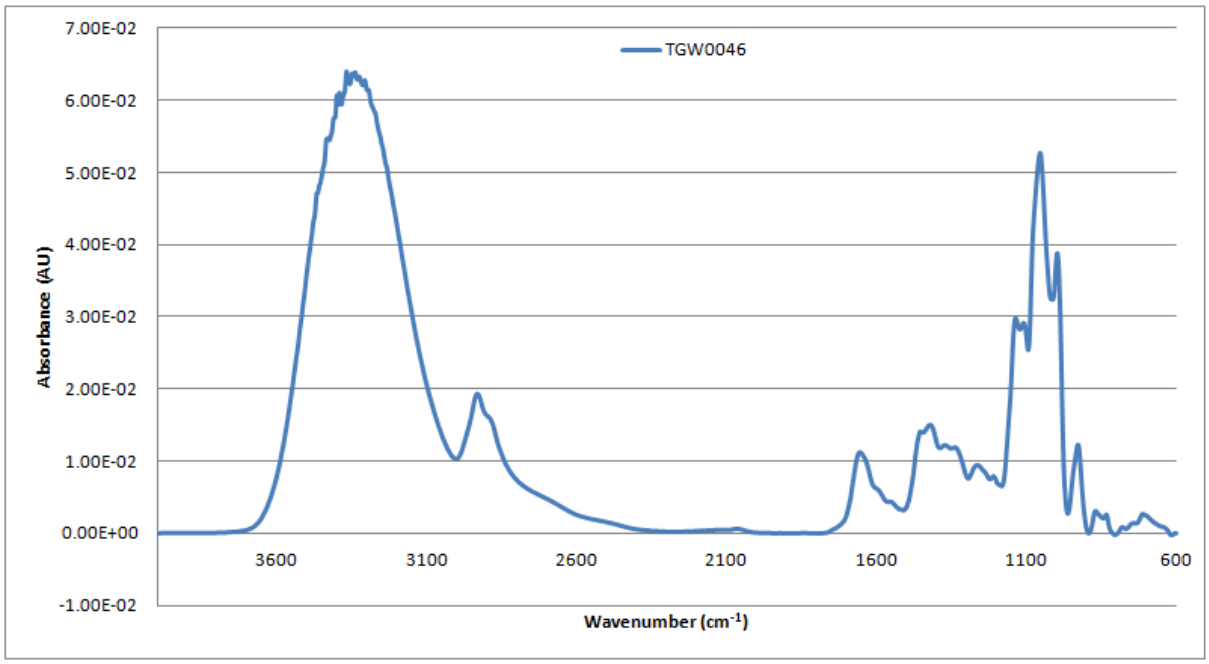


**Figure A2- 273: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0046 from 4000-600cm$^{-1}$**

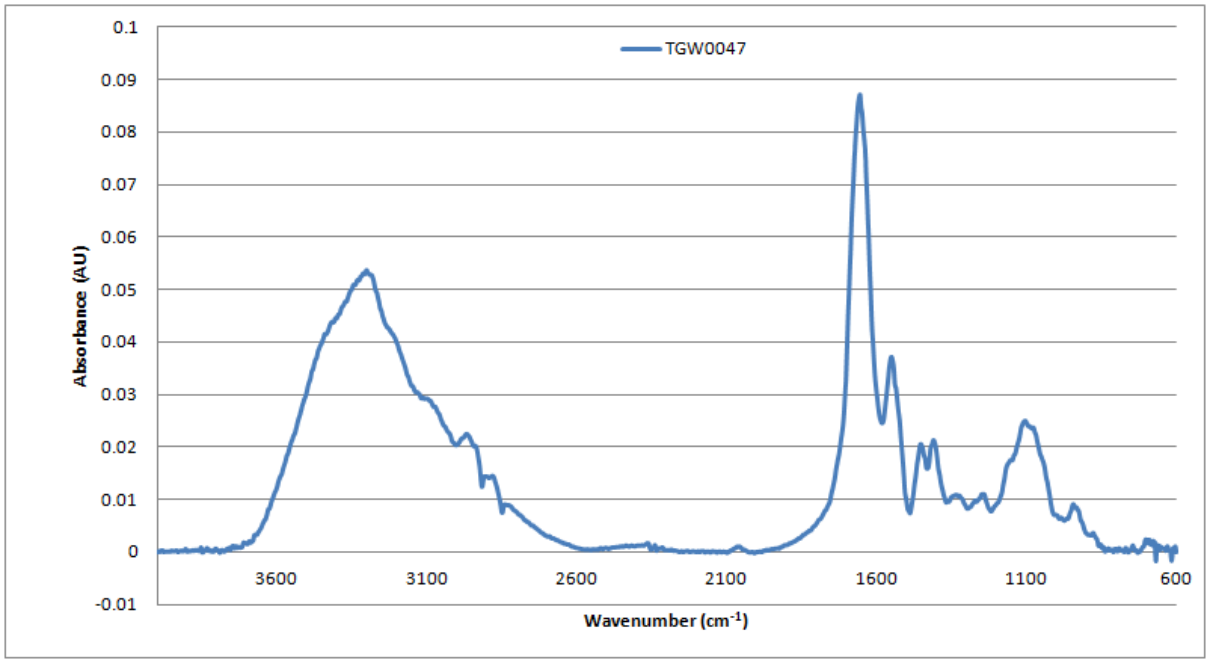


**Figure A2- 274: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0047 from 4000-600cm$^{-1}$**

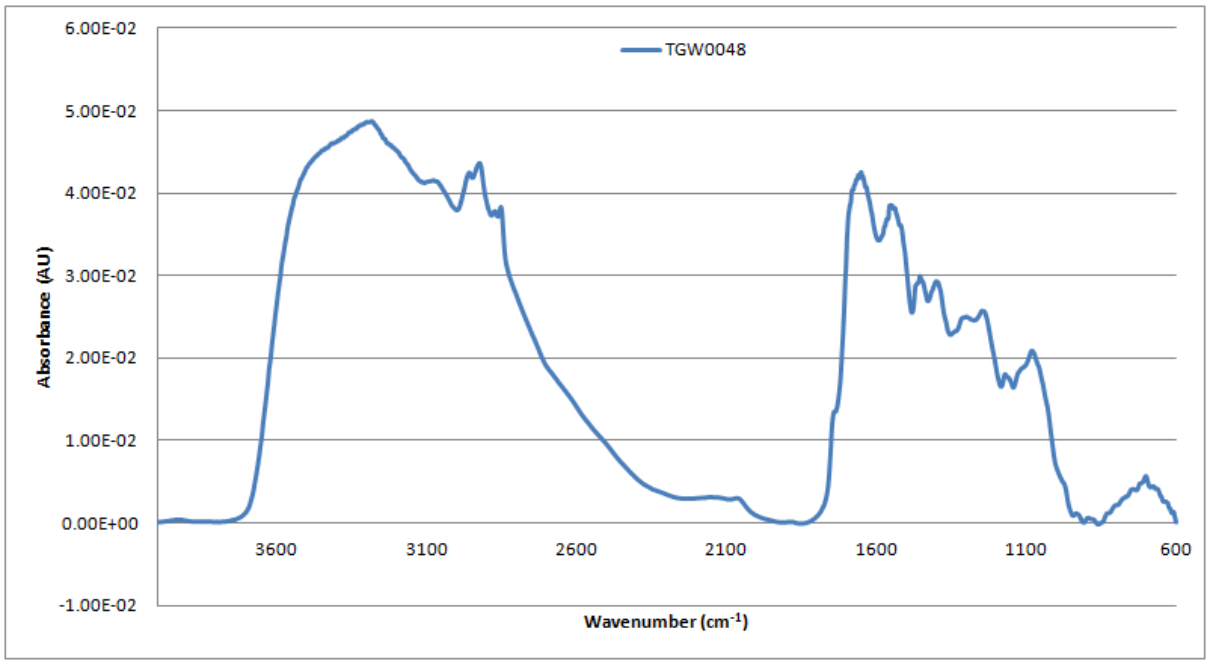


**Figure A2- 275: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0048 from 4000-600cm$^{-1}$**

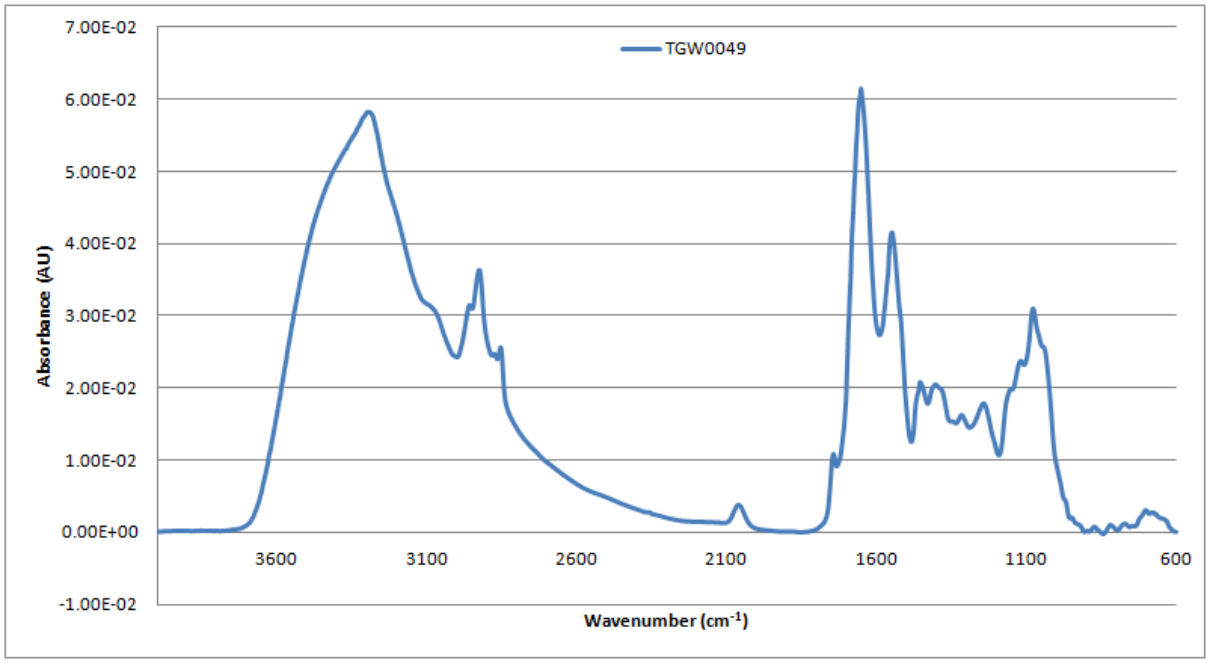


**Figure A2- 276: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0049 from 4000-600cm$^{-1}$**

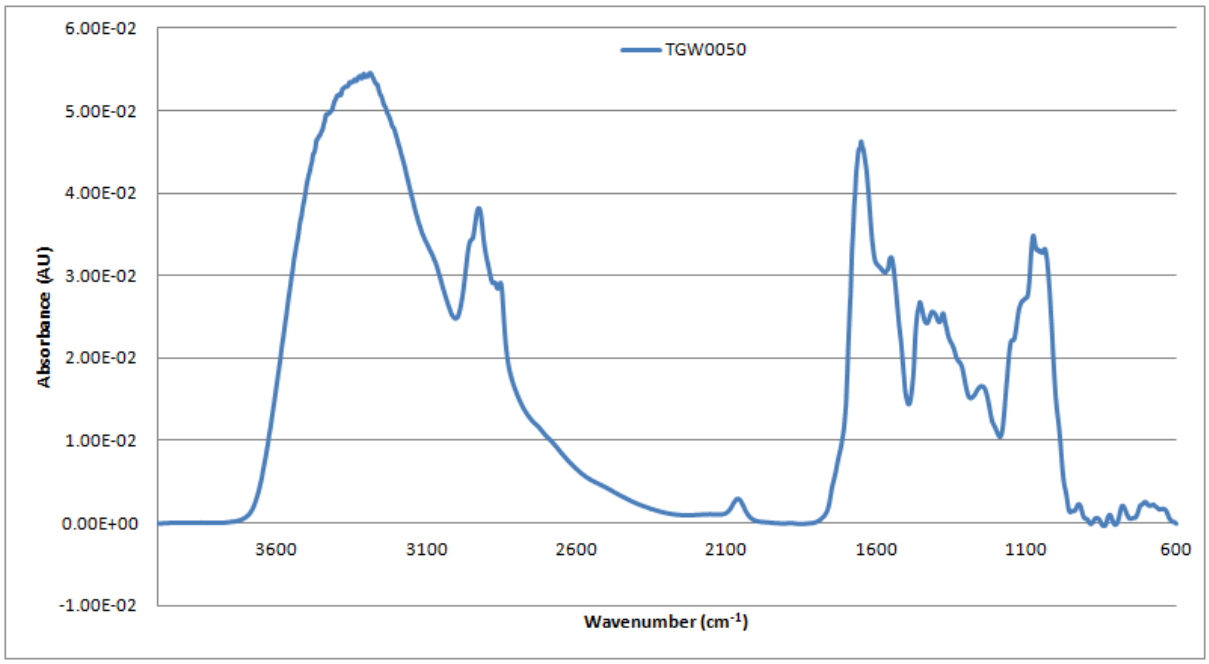


**Figure A2- 277: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0050 from 4000-600cm$^{-1}$**

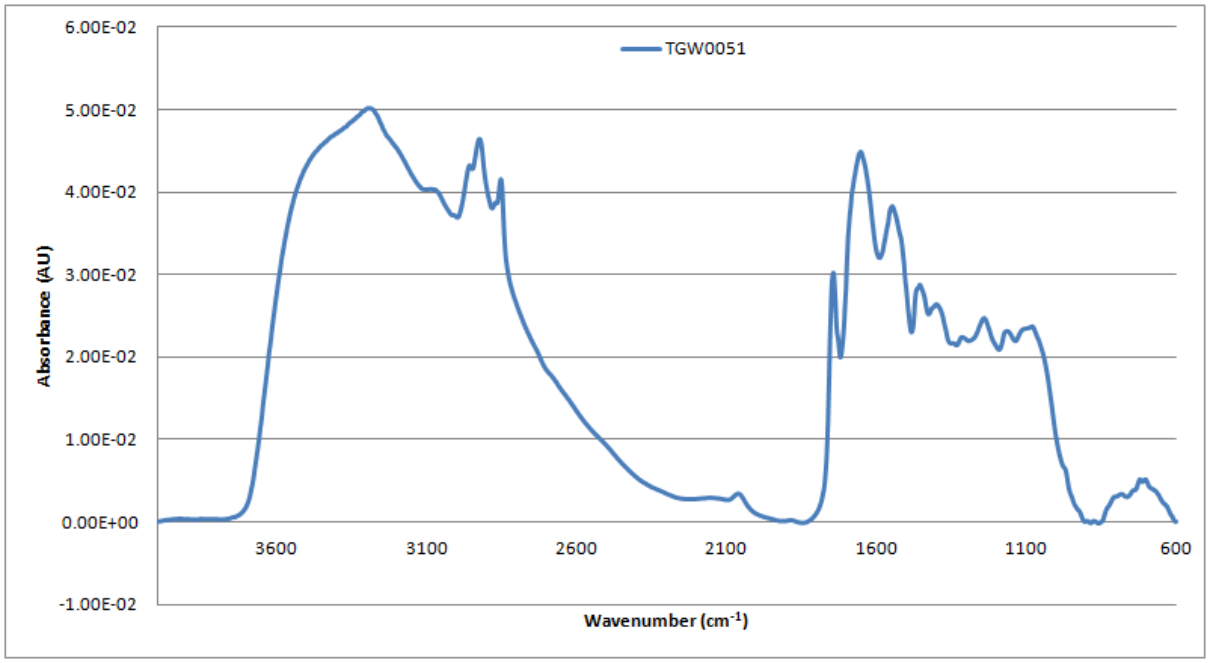


**Figure A2- 278: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0051 from 4000-600cm$^{-1}$**

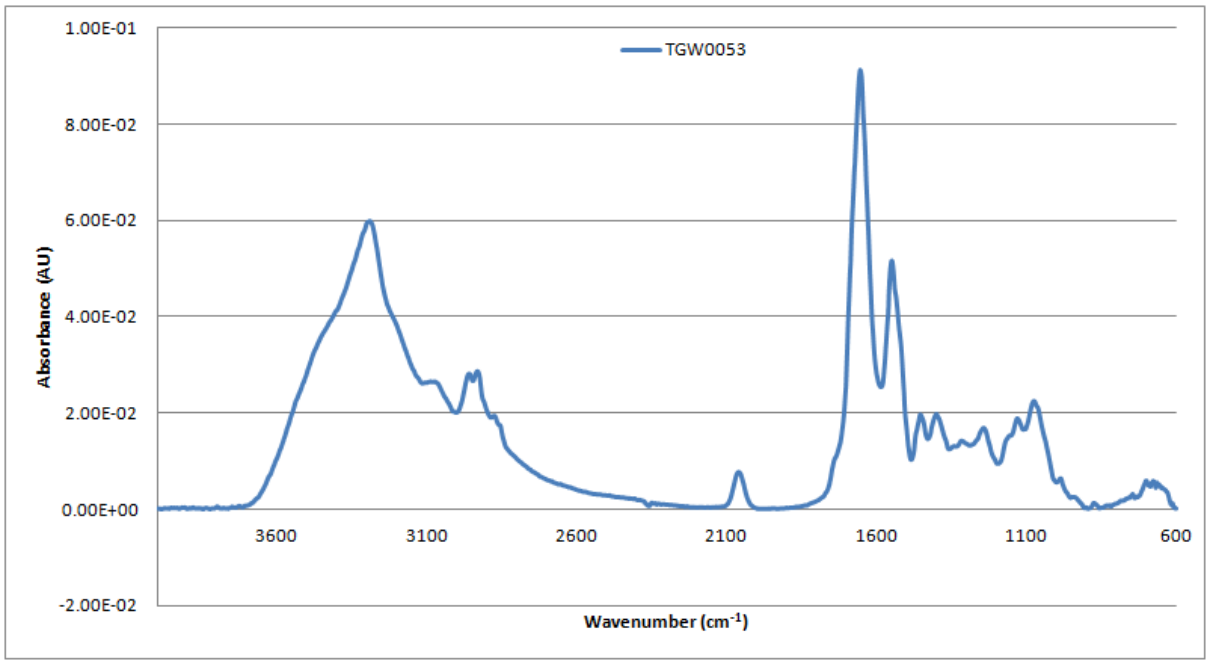


**Figure A2- 279: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0053 from 4000-600cm$^{-1}$**

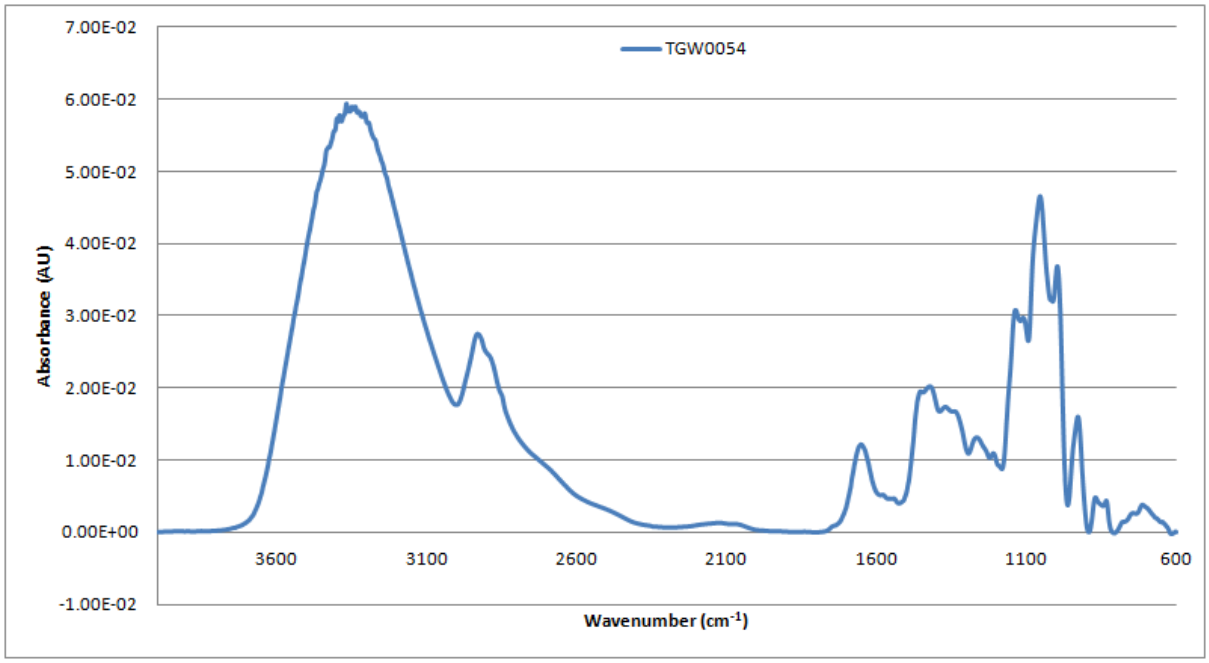


**Figure A2- 280: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0054 from 4000-600cm$^{-1}$**

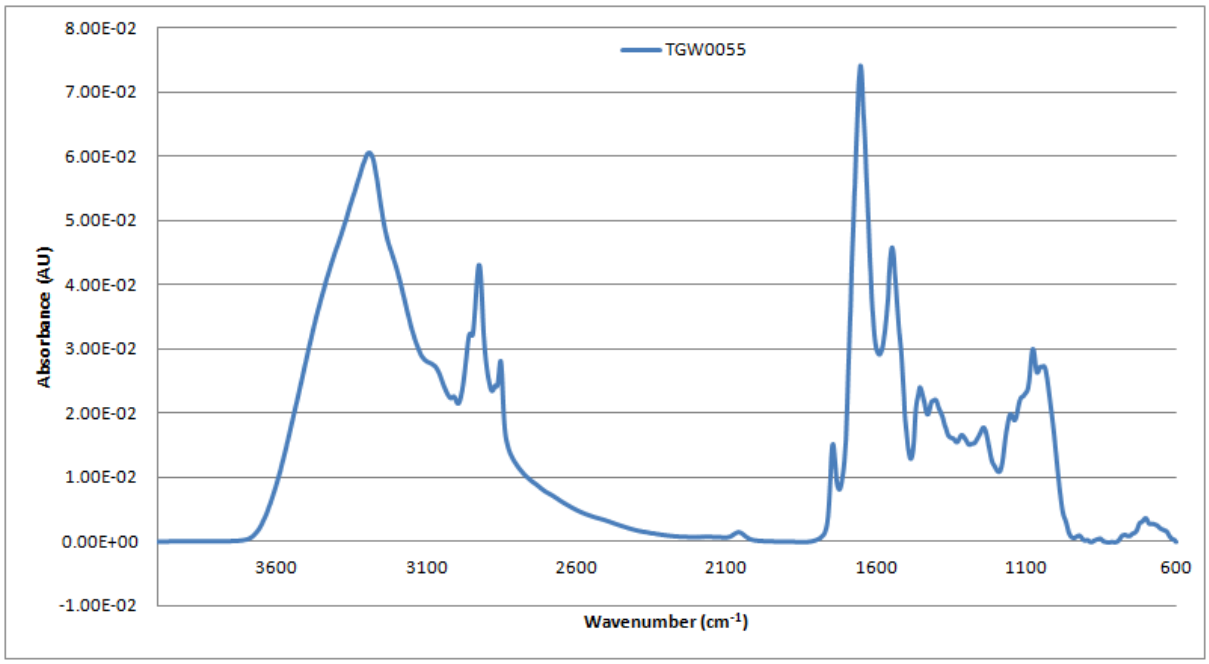


**Figure A2- 281: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0055 from 4000-600cm$^{-1}$**

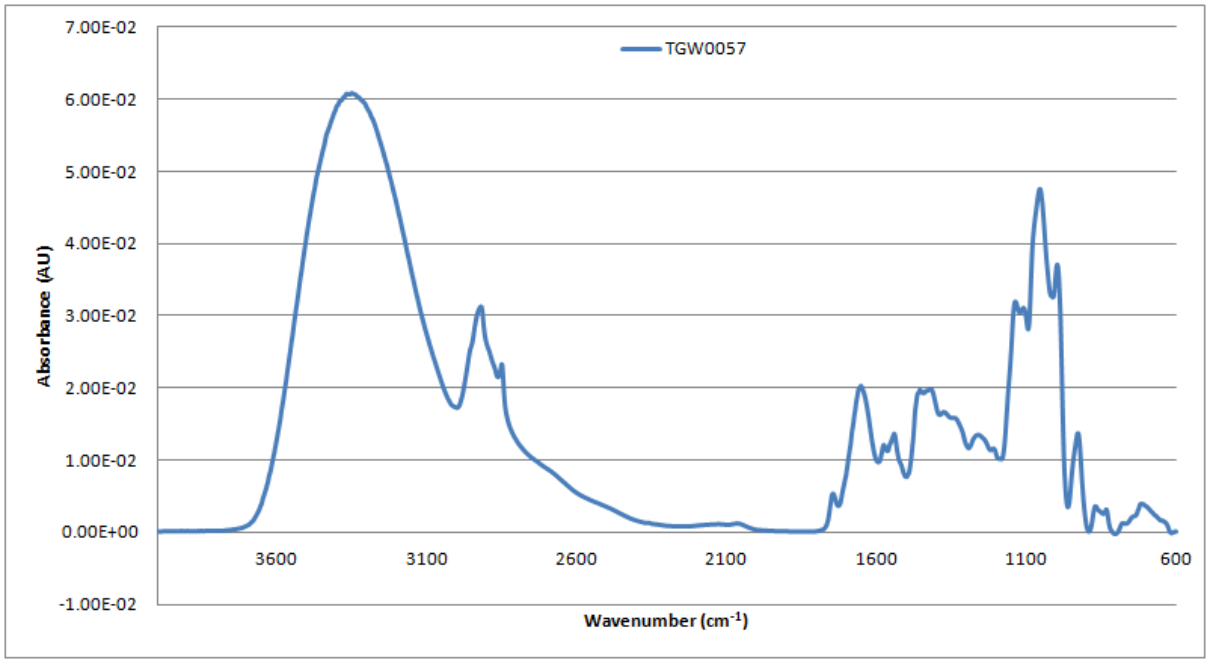


**Figure A2- 282: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0057 from 4000-600cm$^{-1}$**

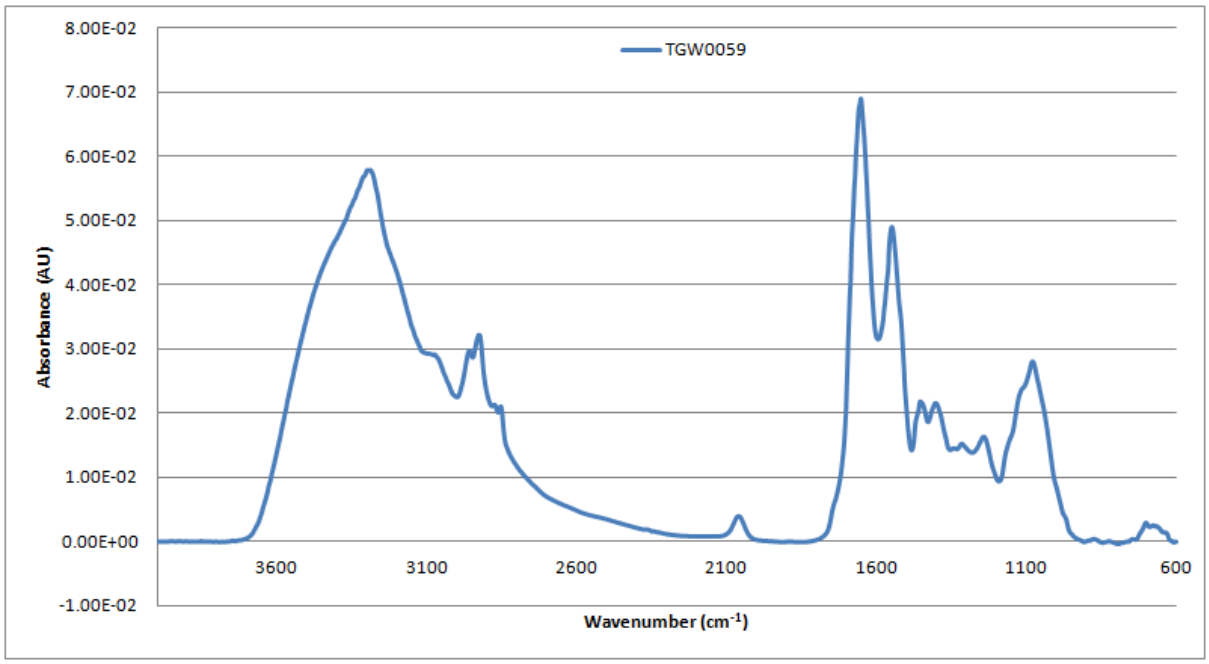


**Figure A2- 283: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0059 from 4000-600cm$^{-1}$**

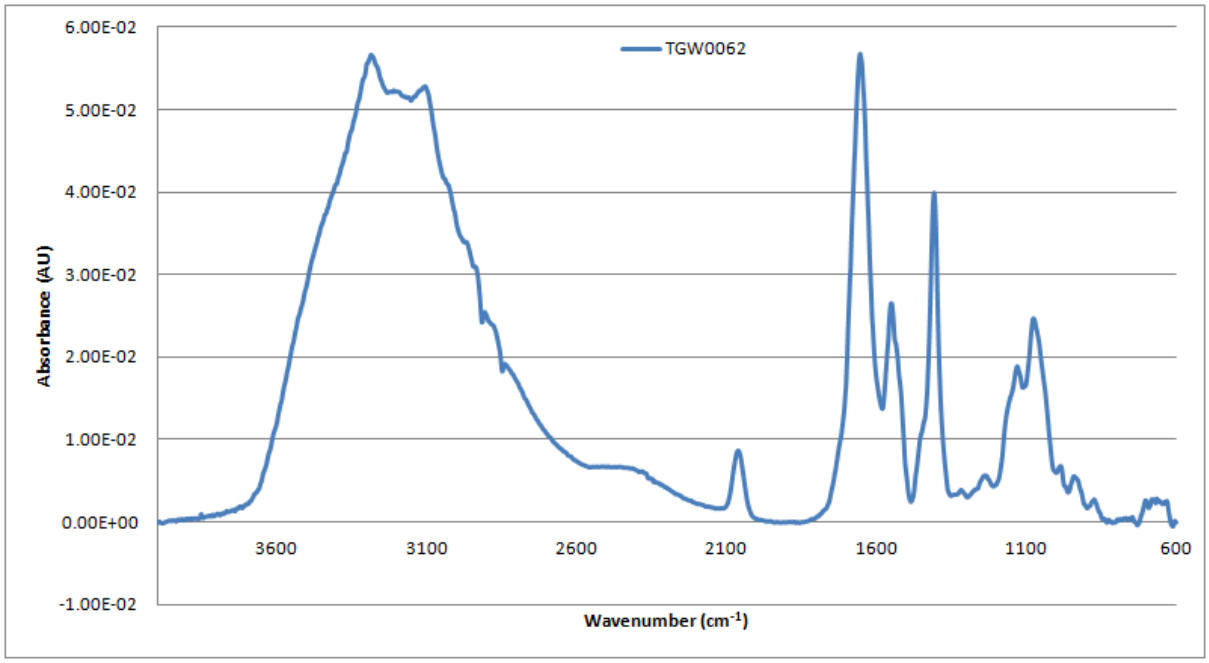


**Figure A2- 284: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0062 from 4000-600cm$^{-1}$**

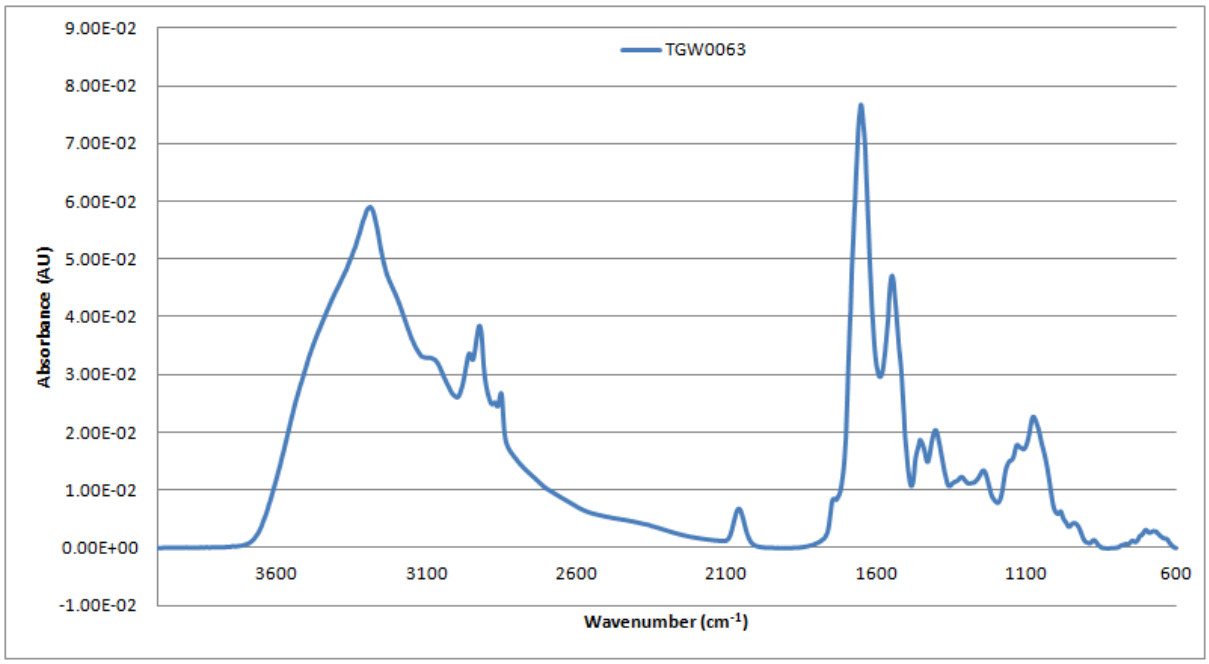


**Figure A2- 285: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0063 from 4000-600cm$^{-1}$**

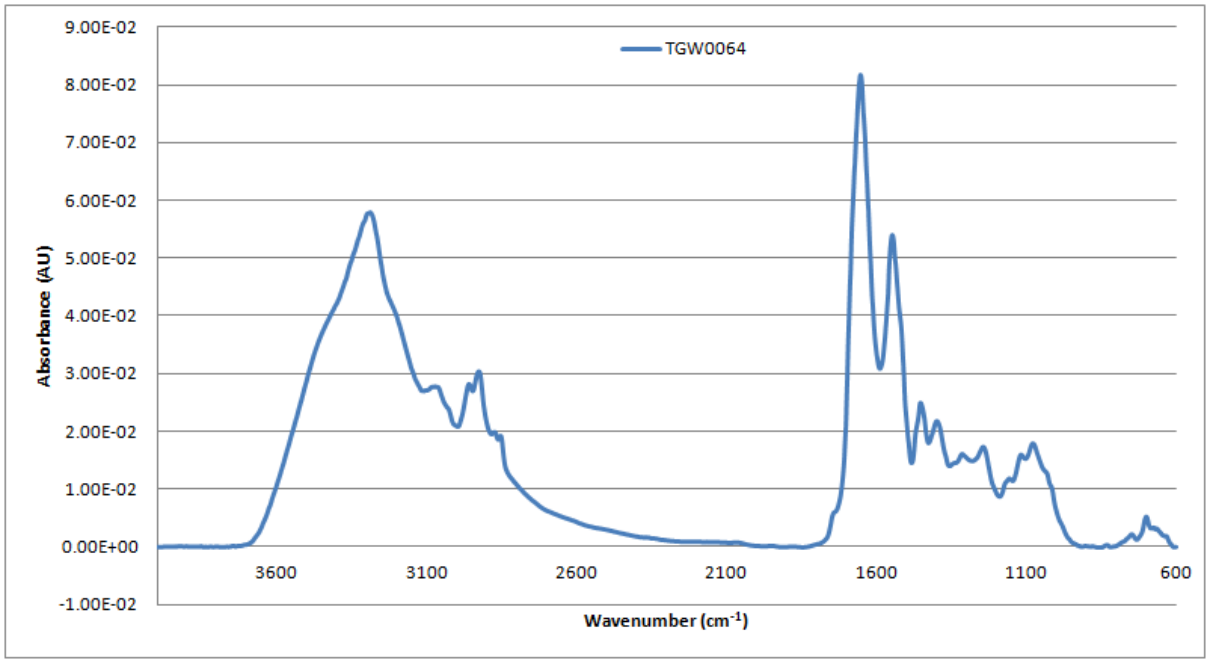


**Figure A2- 286: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0064 from 4000-600cm$^{-1}$**

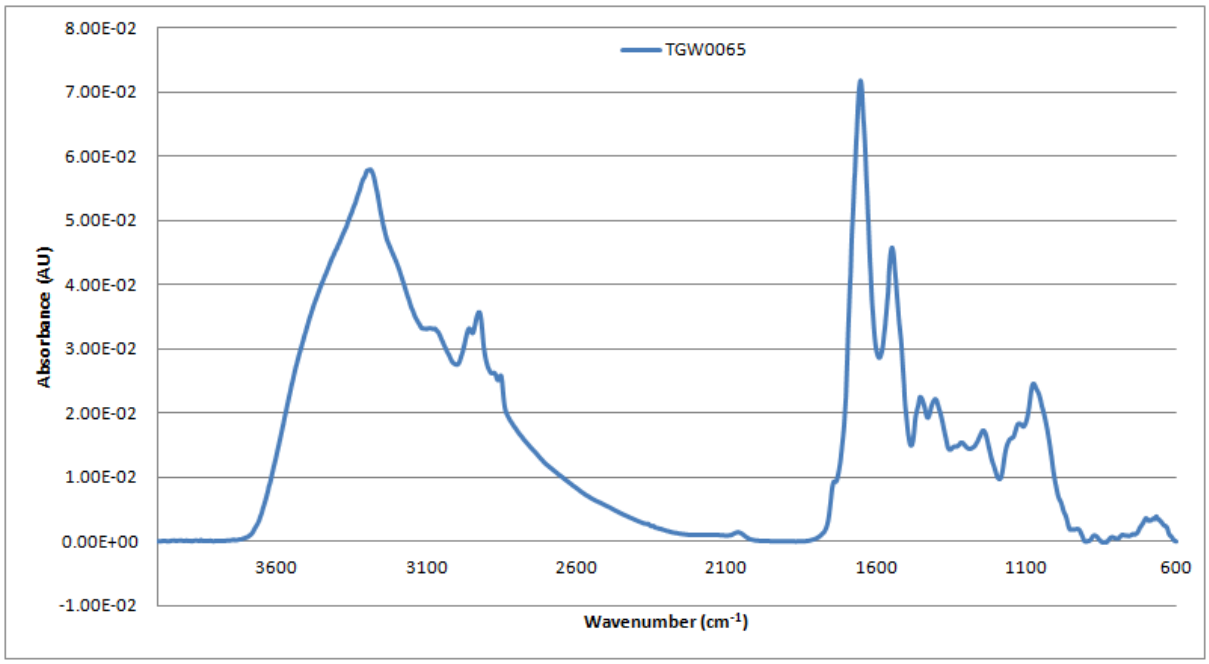


**Figure A2- 287: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0065 from 4000-600cm$^{-1}$**

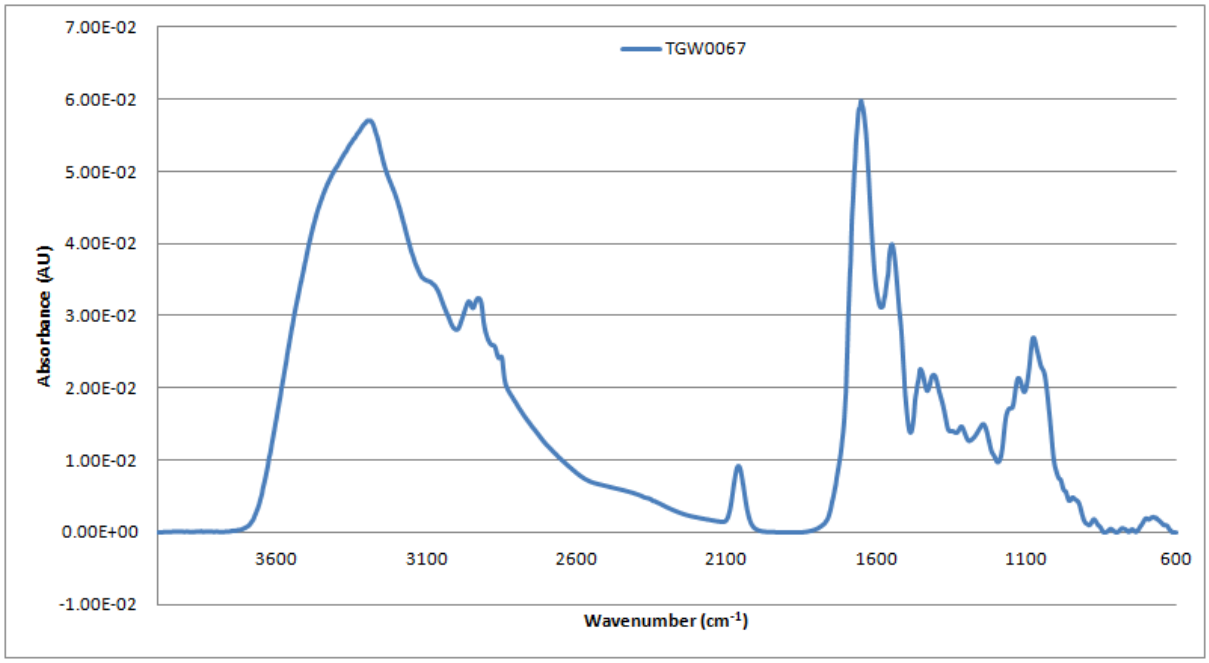


**Figure A2- 288: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0067 from 4000-600cm$^{-1}$**

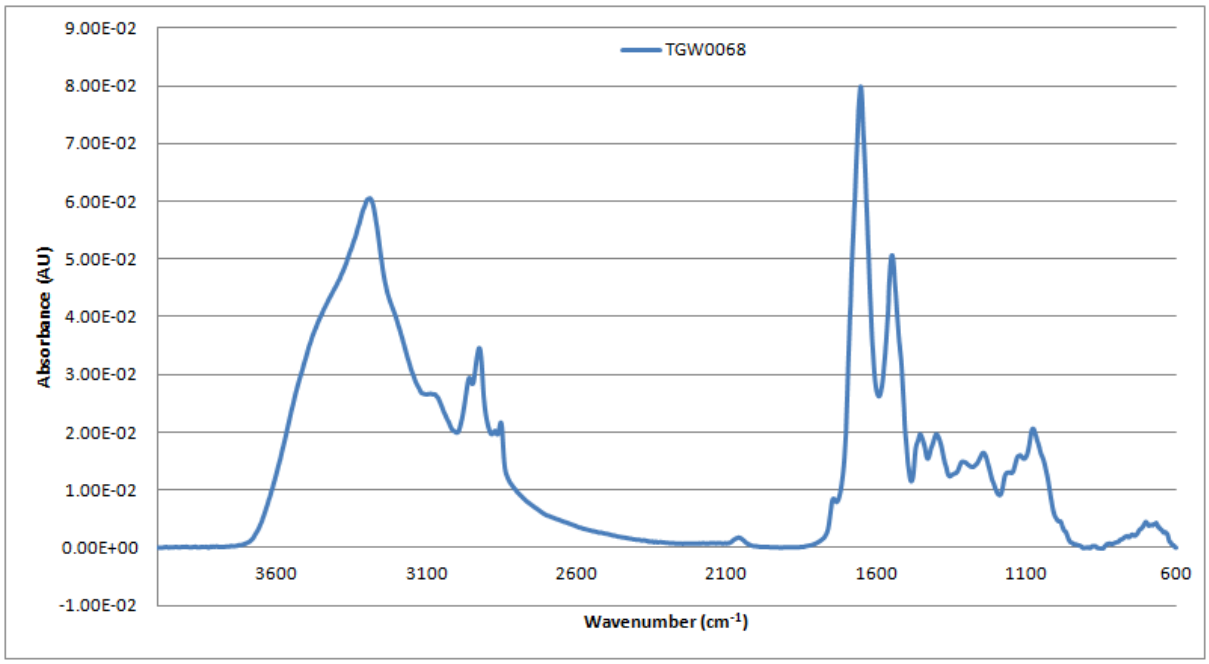


**Figure A2- 289: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0068 from 4000-600cm$^{-1}$**

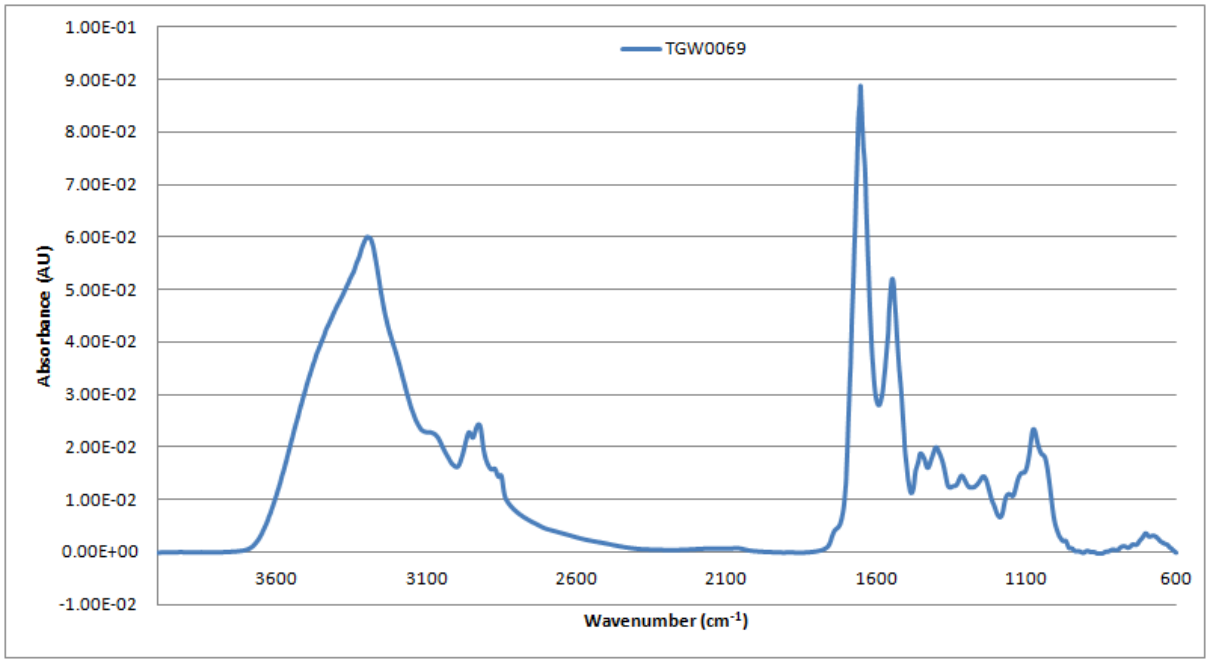


**Figure A2- 290: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0069 from 4000-600cm$^{-1}$**

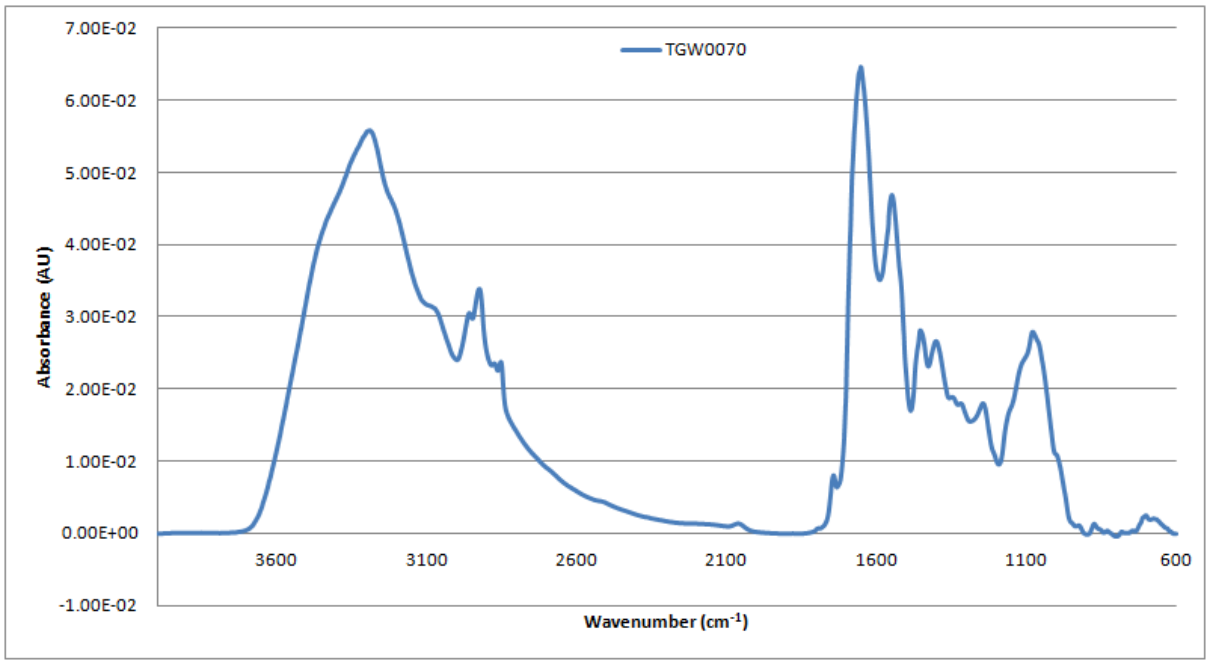


**Figure A2- 291: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0070 from 4000-600cm$^{-1}$**

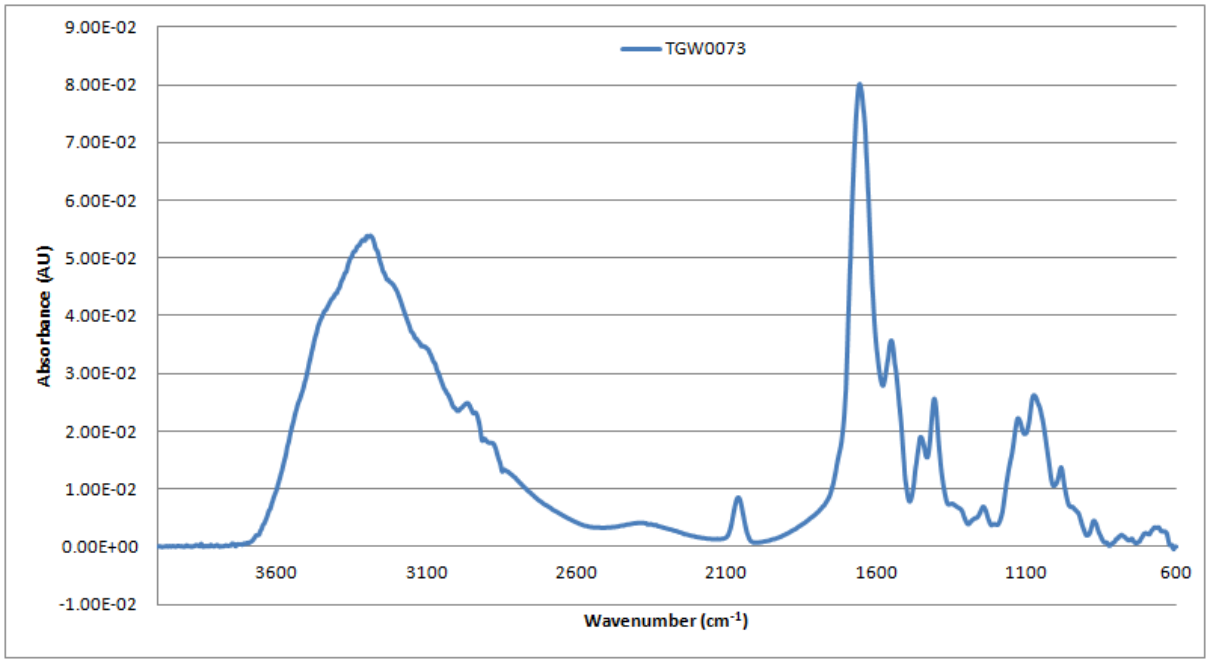


**Figure A2- 292: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0073 from 4000-600cm$^{-1}$**

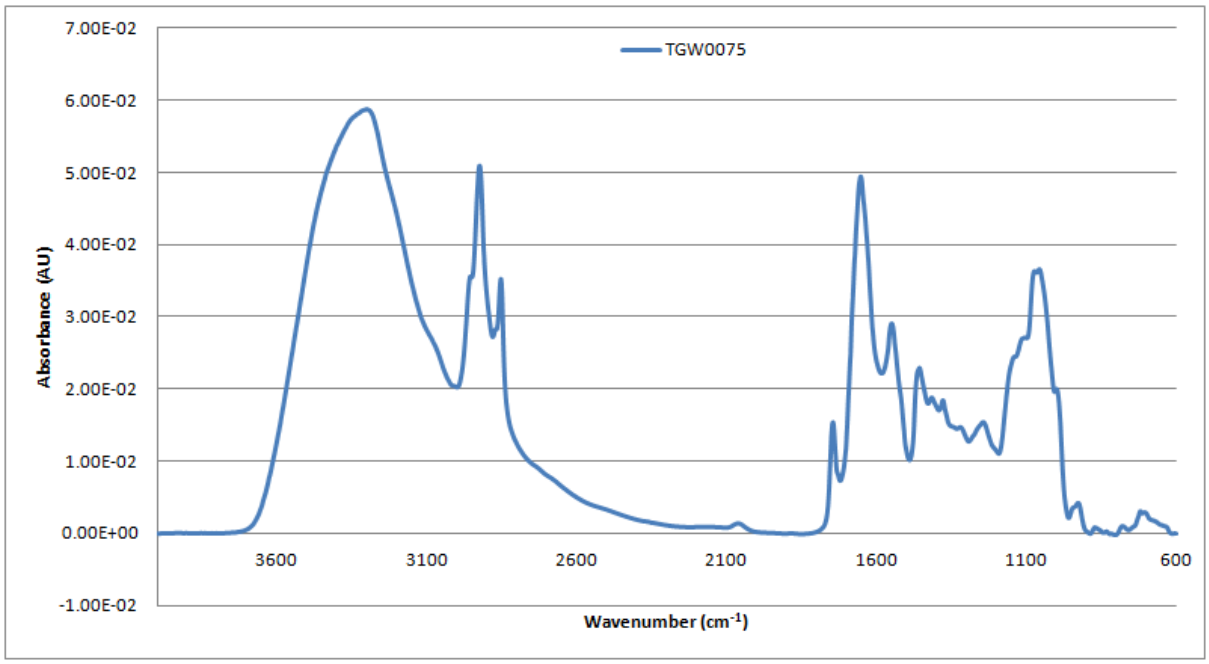


**Figure A2- 293: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0075 from 4000-600cm$^{-1}$**

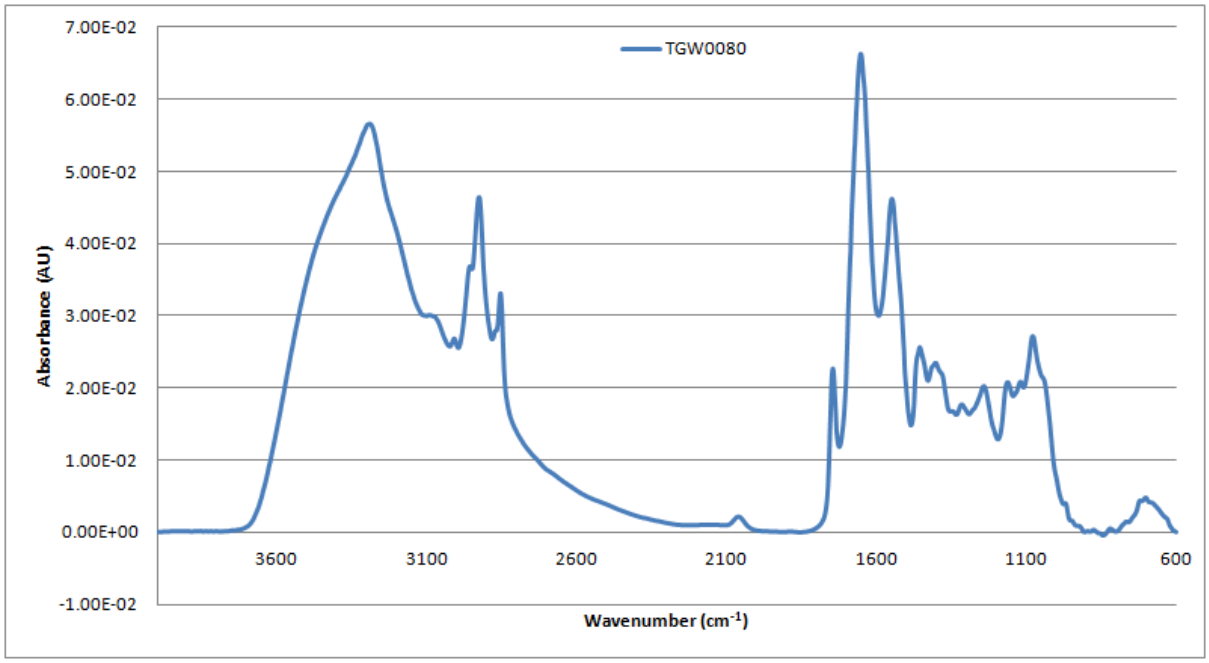


**Figure A2- 294: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0080 from 4000-600cm$^{-1}$**

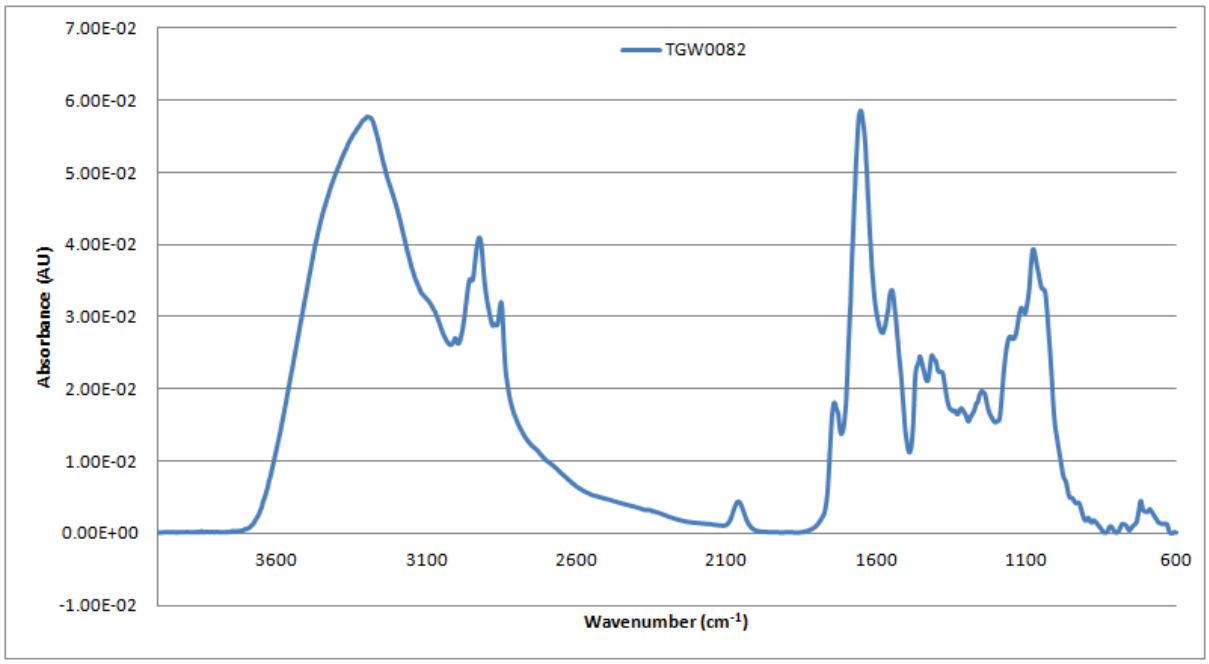


**Figure A2- 295: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0082 from 4000-600cm$^{-1}$**

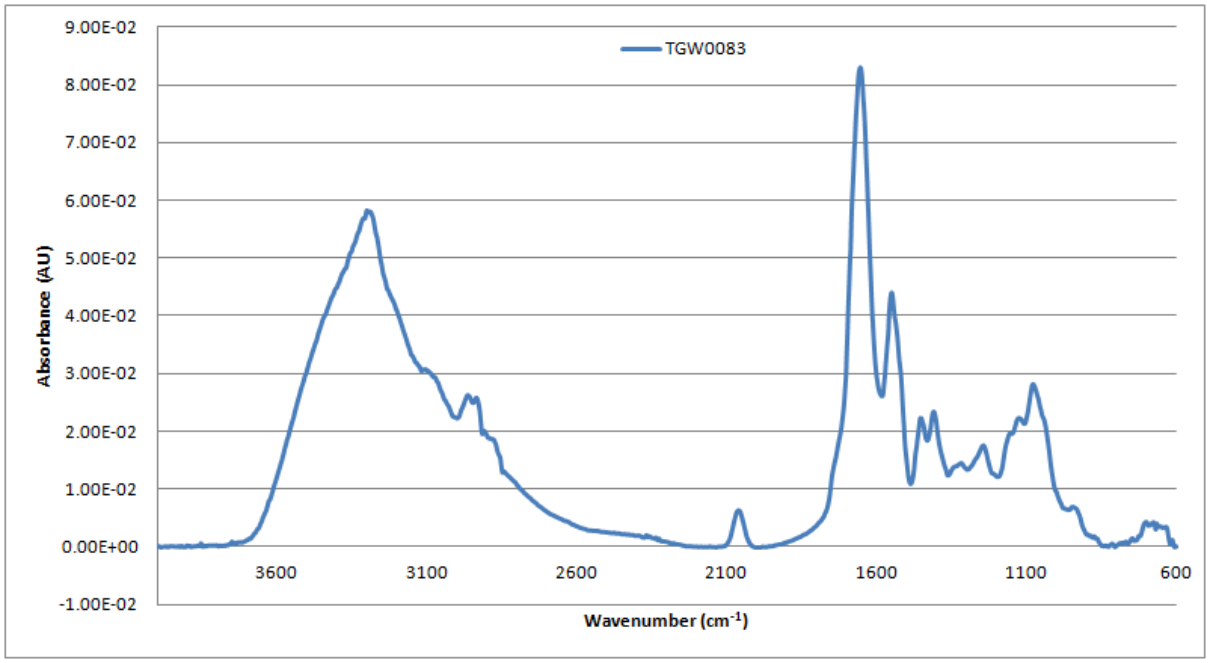


**Figure A2- 296: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0083 from 4000-600cm$^{-1}$**

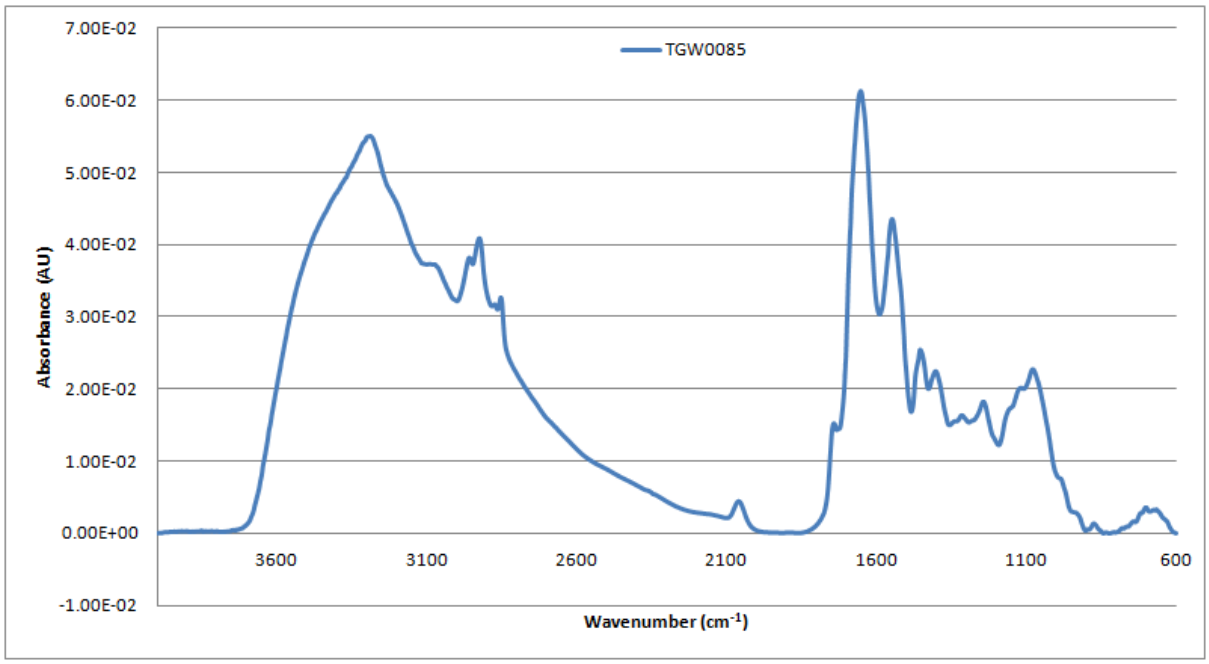


**Figure A2- 297: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0085 from 4000-600cm$^{-1}$**

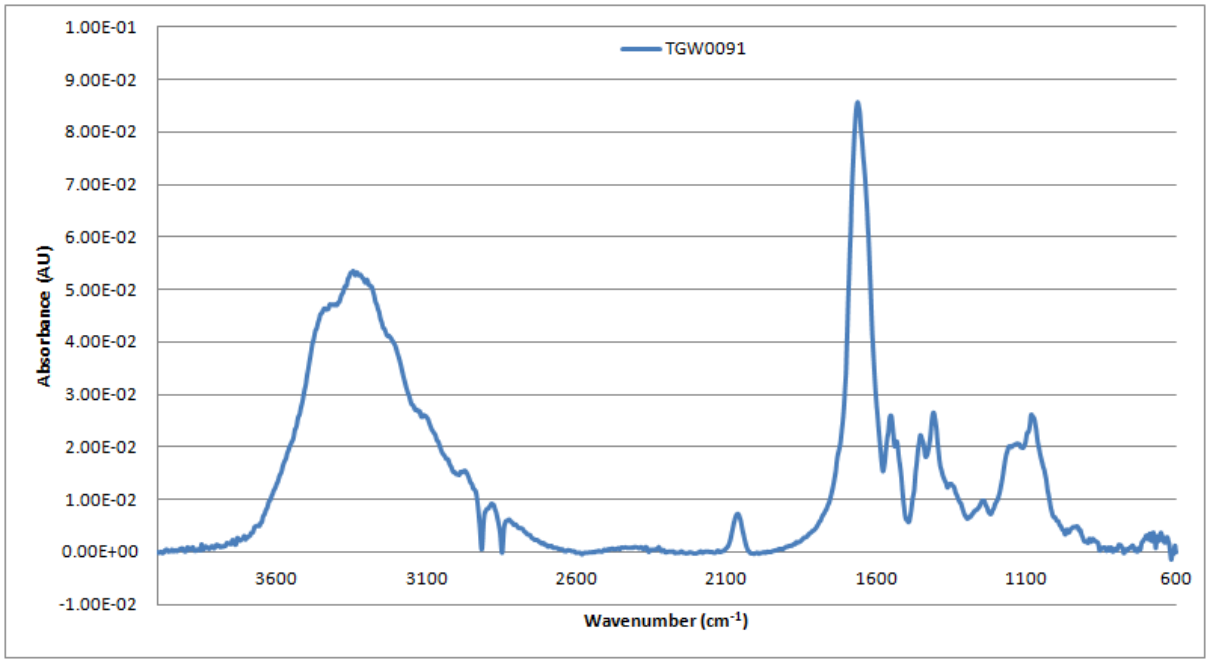


**Figure A2- 298: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0091 from 4000-600cm$^{-1}$**

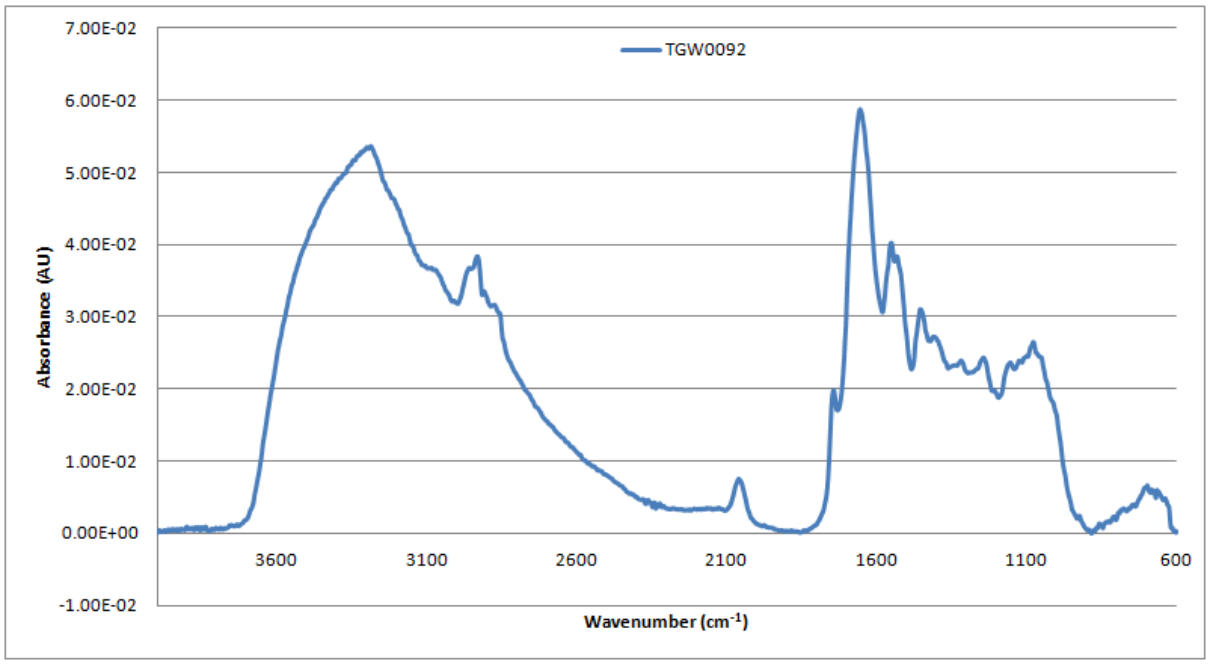


**Figure A2- 299: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0092 from 4000-600cm$^{-1}$**

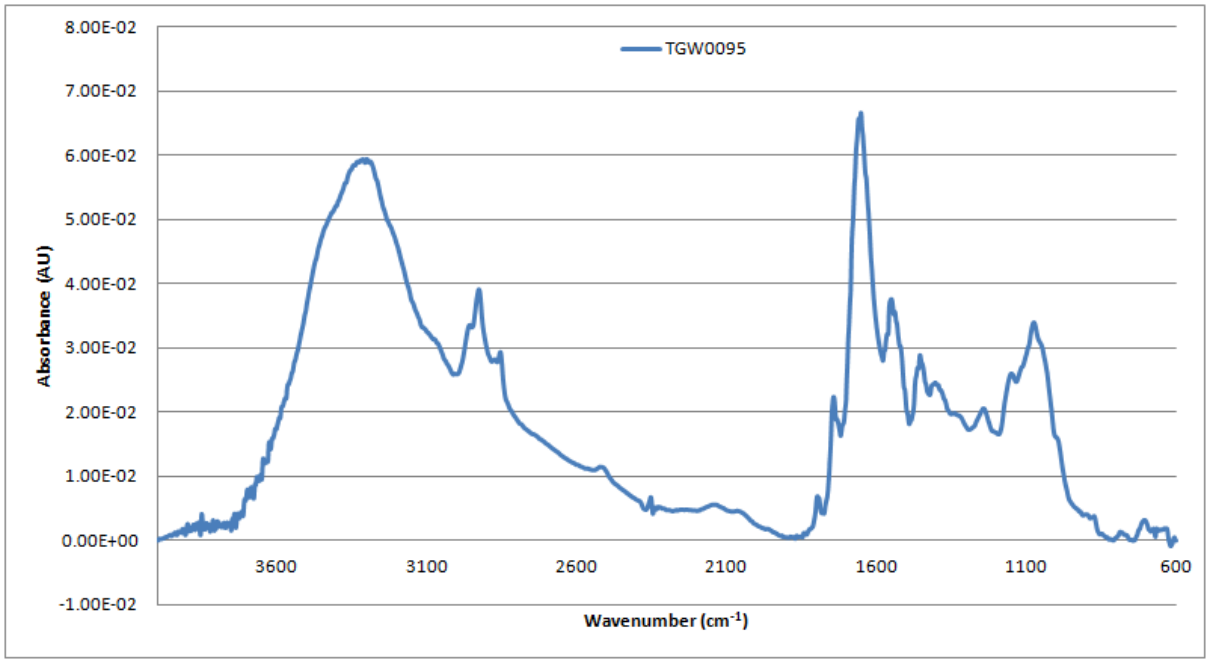


**Figure A2- 300: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0095 from 4000-600cm$^{-1}$**

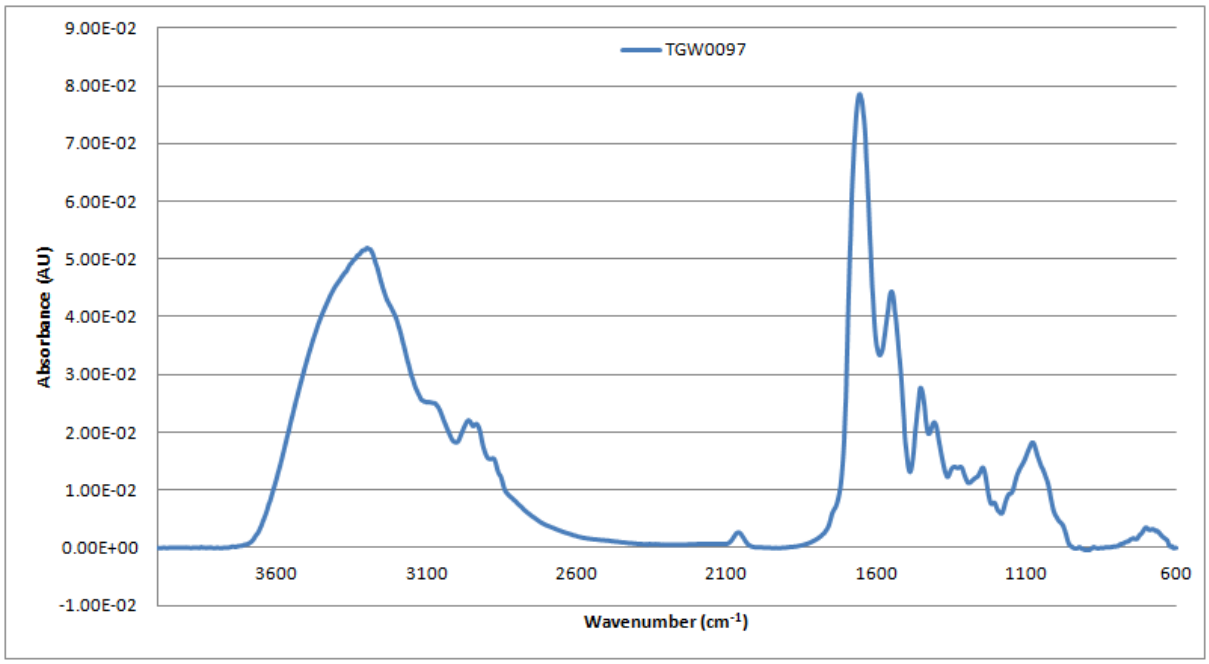


**Figure A2- 301: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0097 from 4000-600cm$^{-1}$**

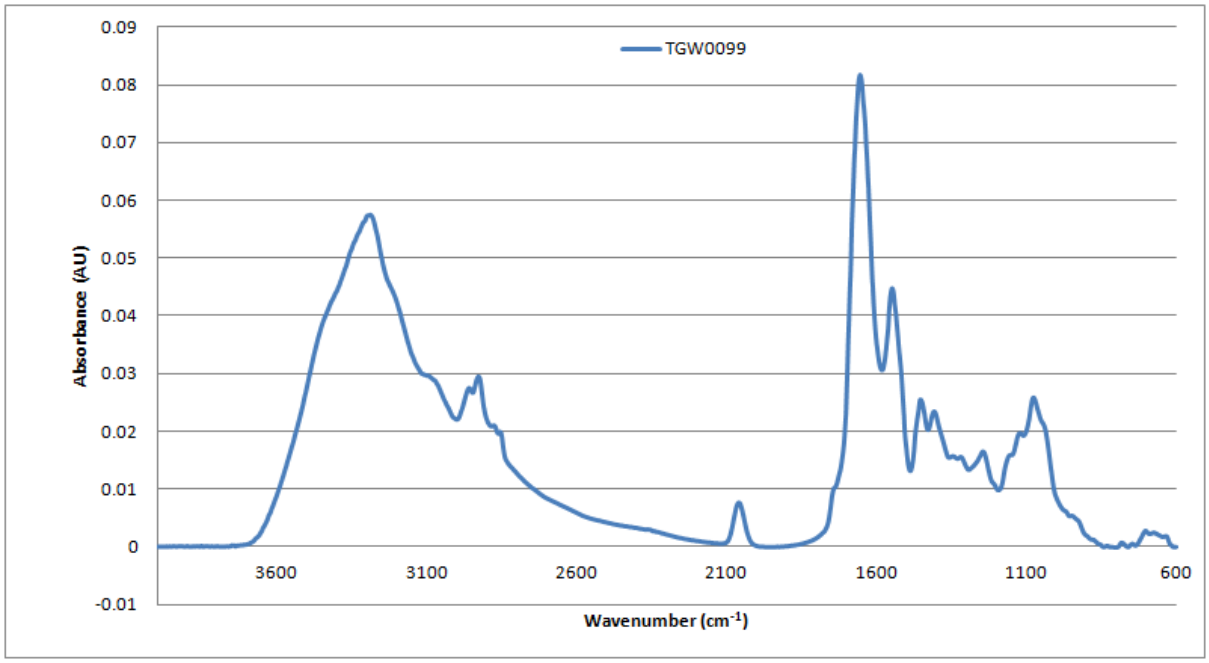


**Figure A2- 302: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0099 from 4000-600cm$^{-1}$**

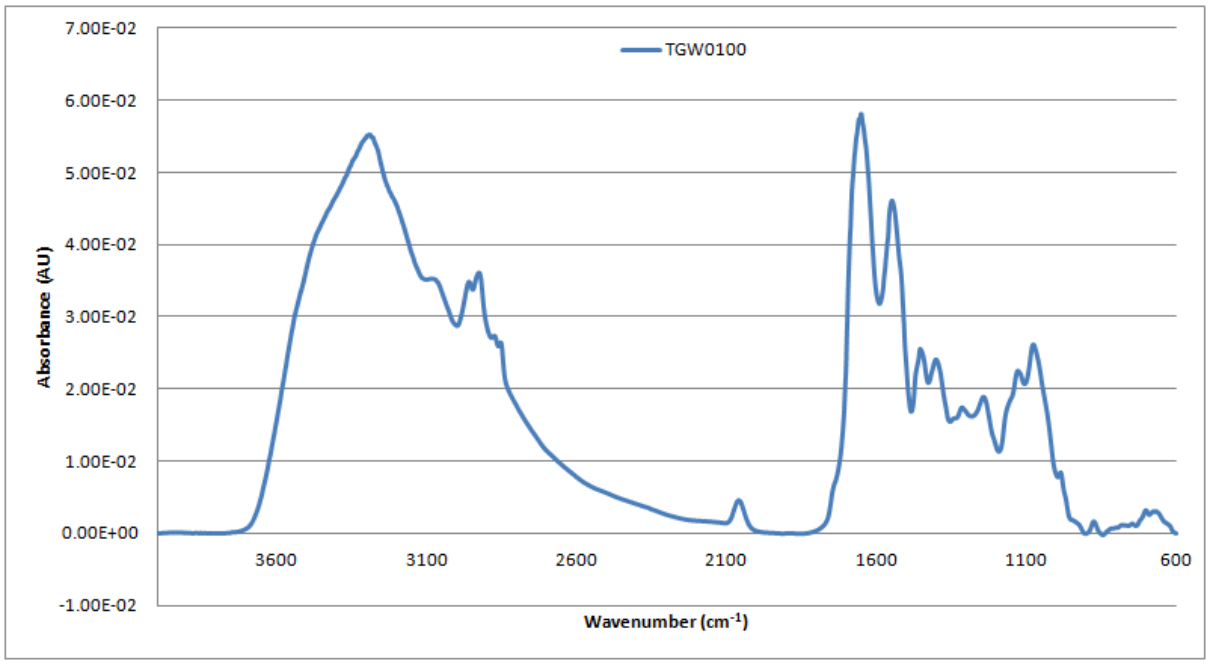


**Figure A2- 303: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0100 from 4000-600cm$^{-1}$**

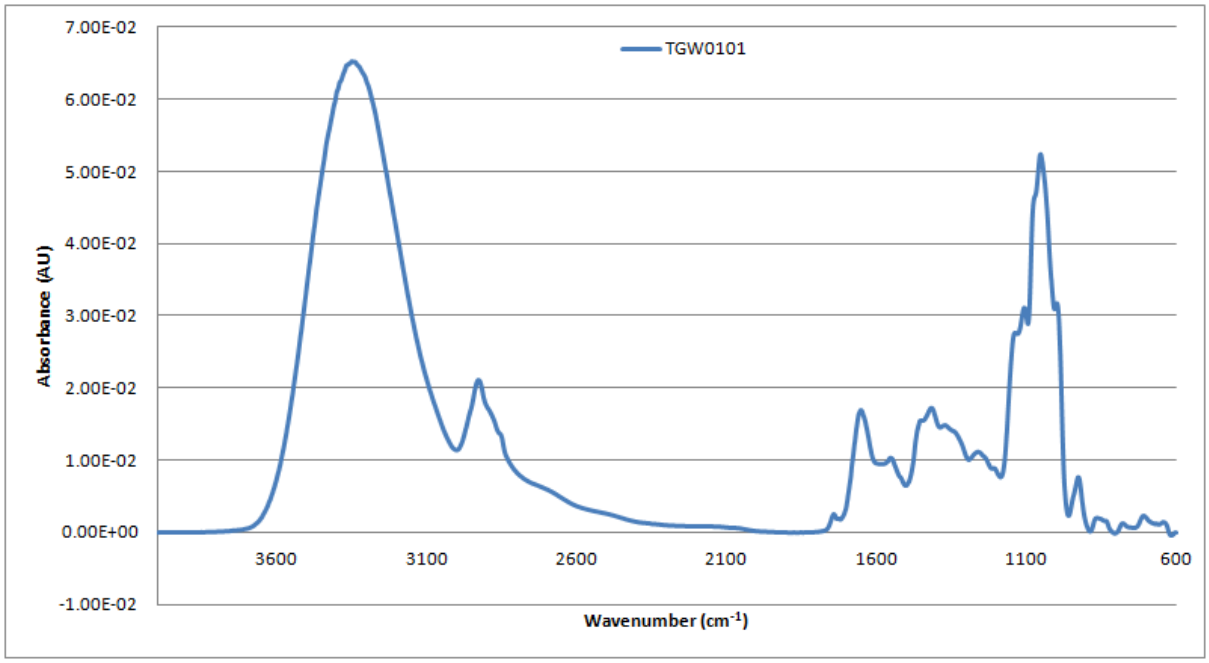


**Figure A2- 304: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0101 from 4000-600cm$^{-1}$**

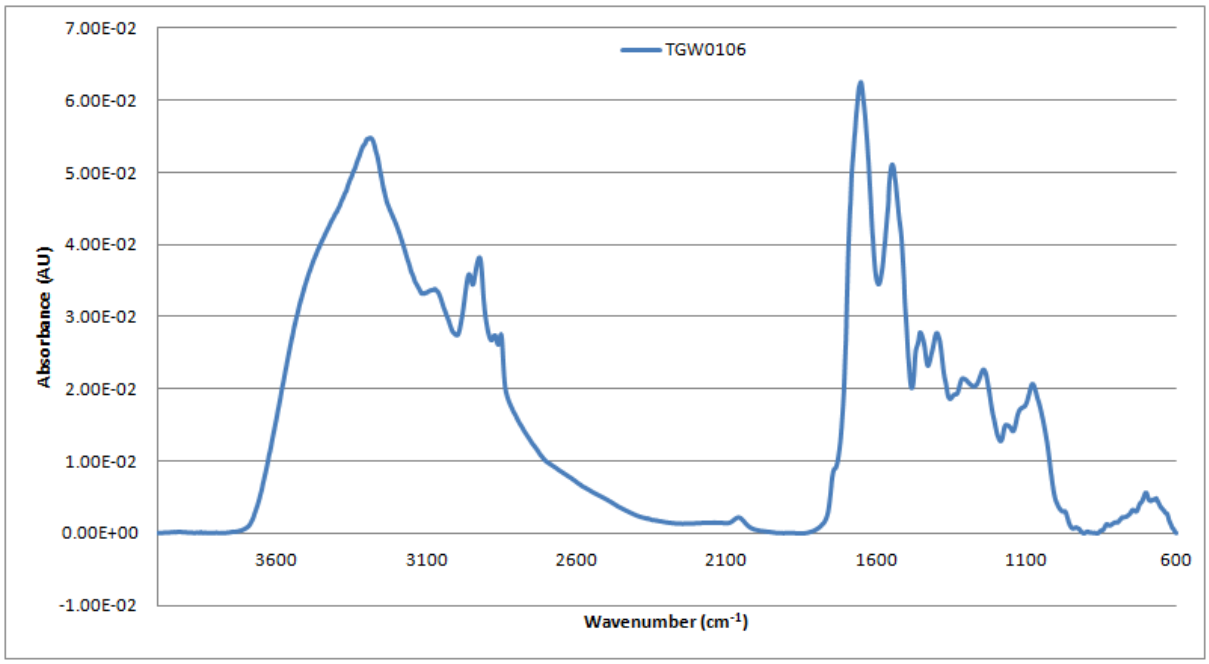


**Figure A2- 305: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0106 from 4000-600cm$^{-1}$**

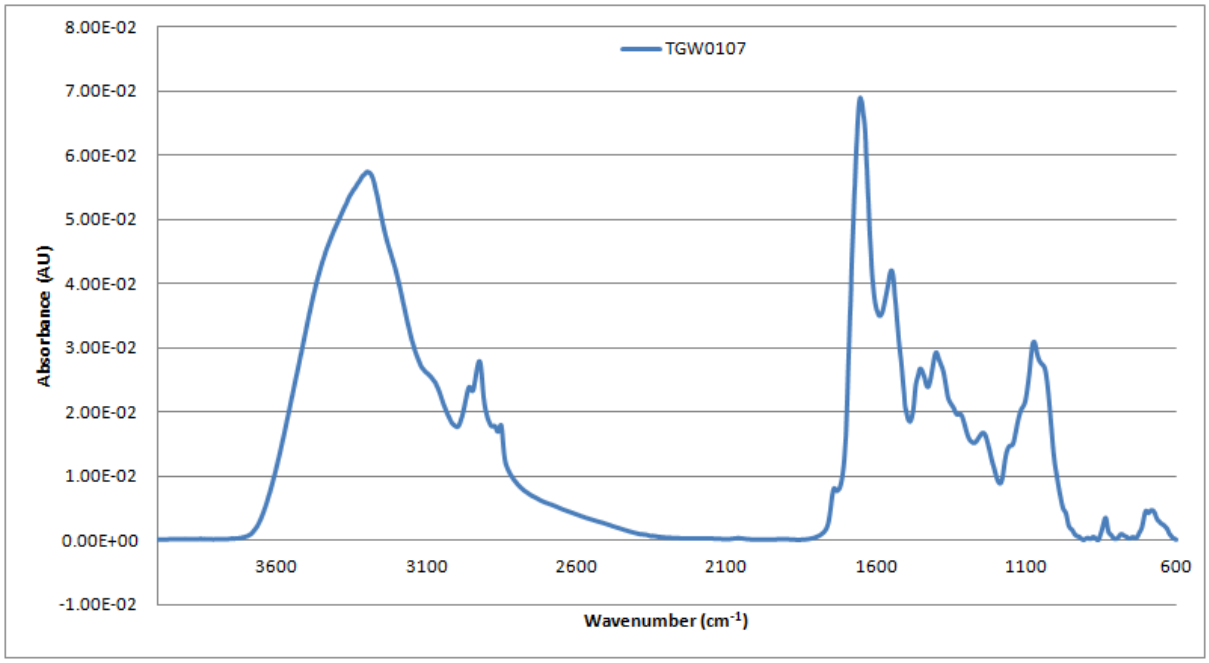


**Figure A2- 306: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0107 from 4000-600cm$^{-1}$**

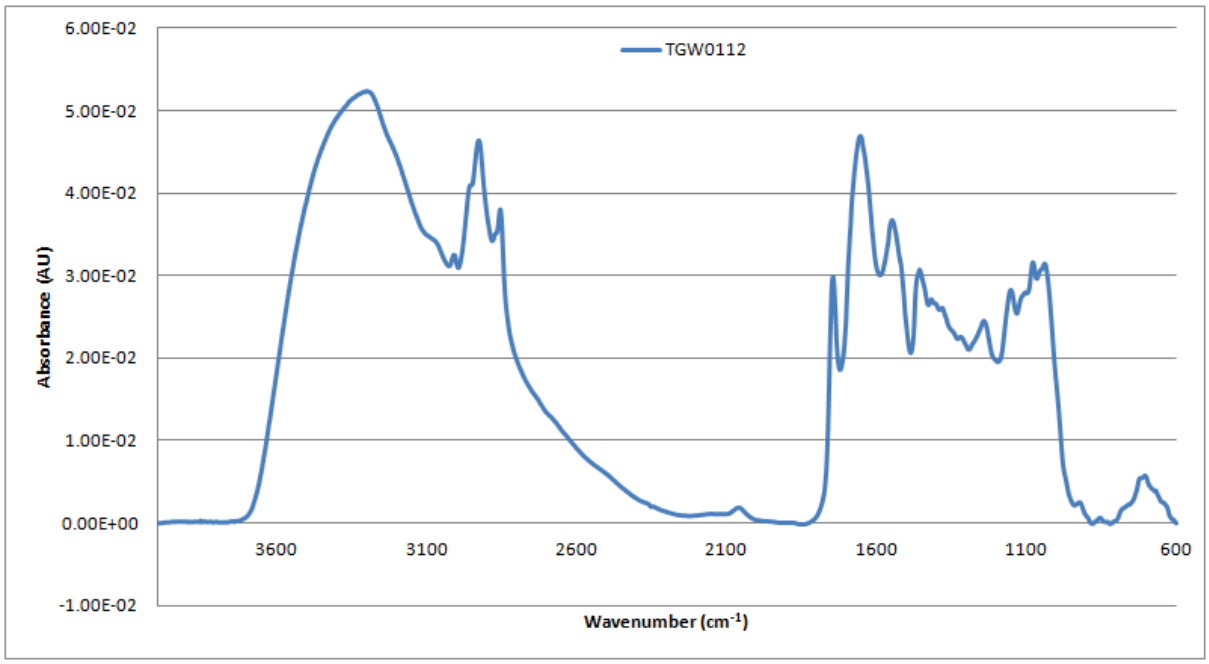


**Figure A2- 307: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0112 from 4000-600cm$^{-1}$**

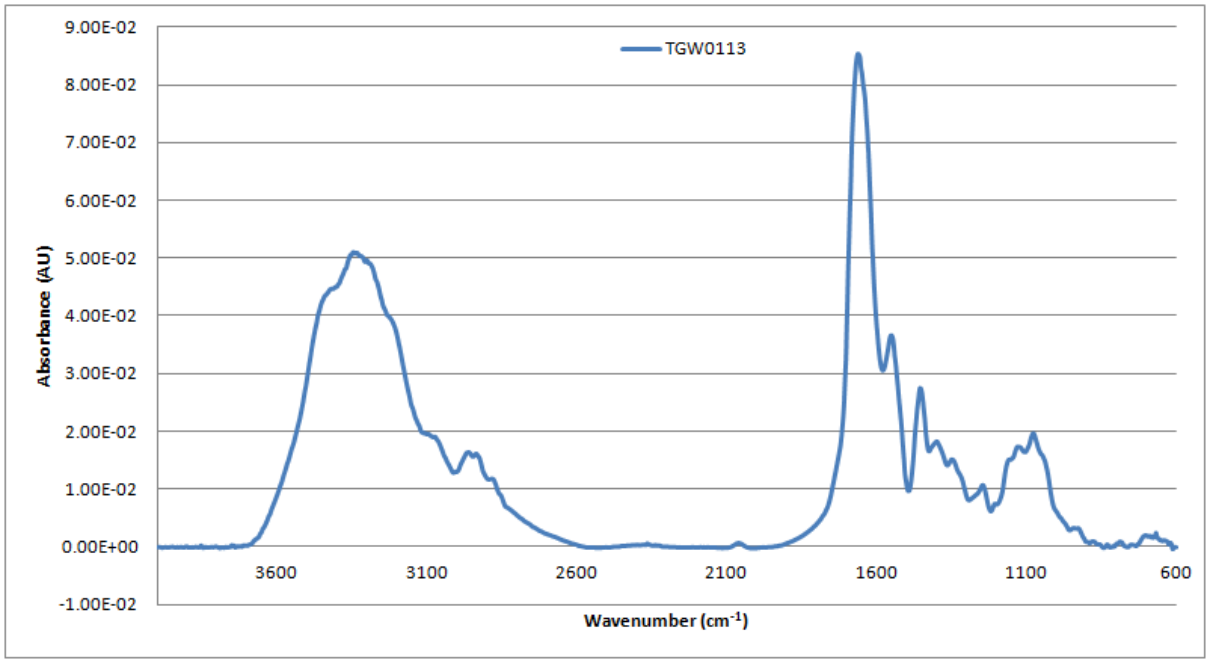


**Figure A2- 308: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0113 from 4000-600cm$^{-1}$**

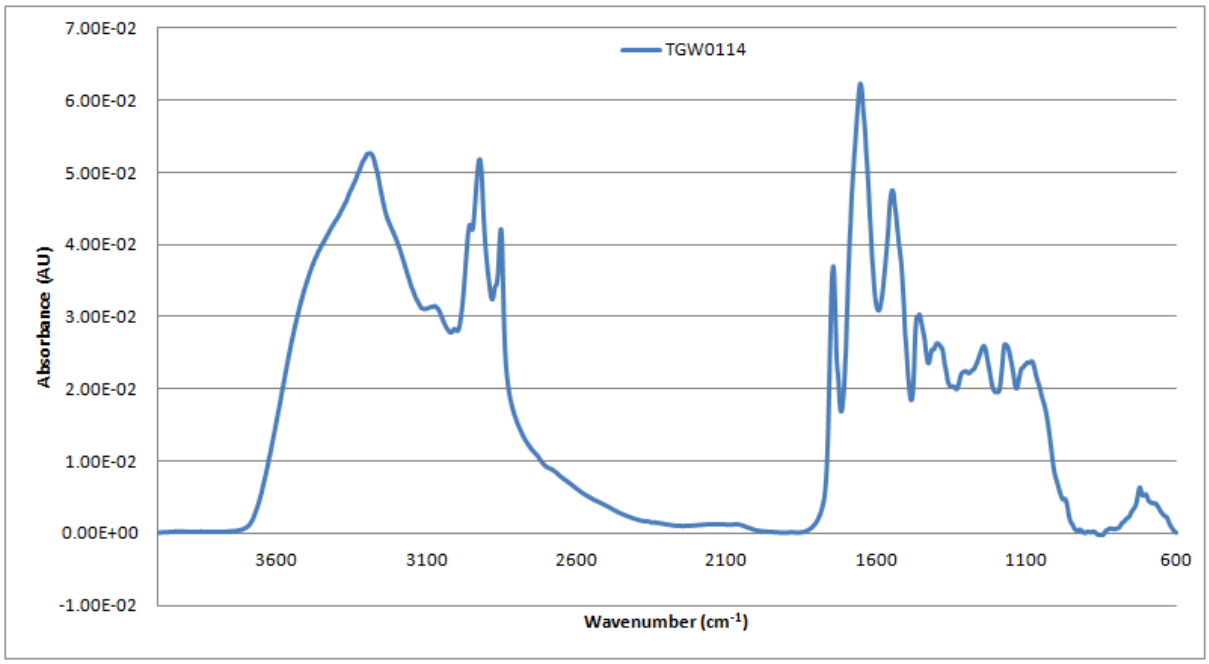


**Figure A2- 309: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0114 from 4000-600cm$^{-1}$**

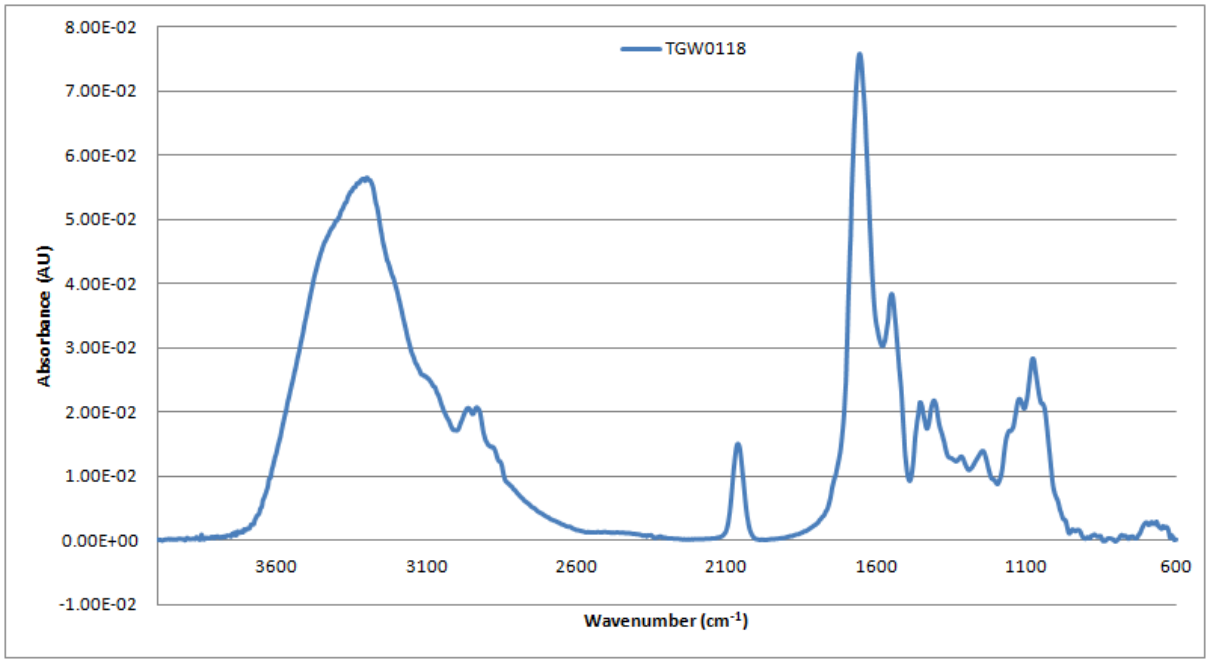


**Figure A2- 310: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0118 from 4000-600cm$^{-1}$**

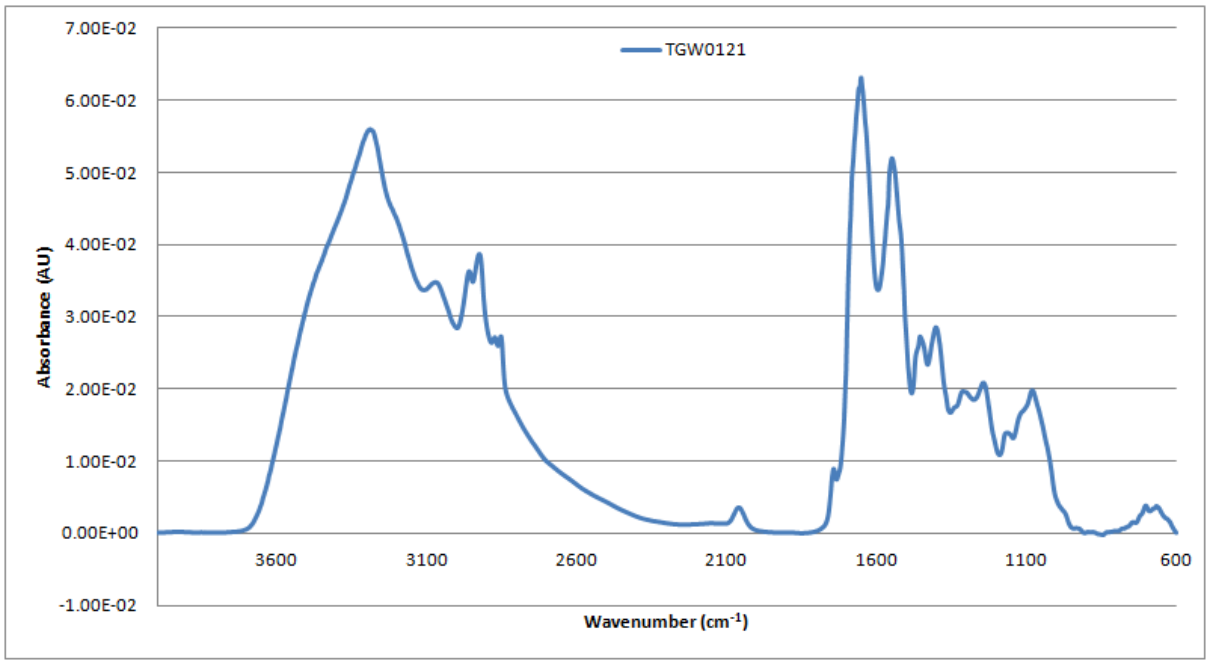


**Figure A2- 311: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0121 from 4000-600cm$^{-1}$**

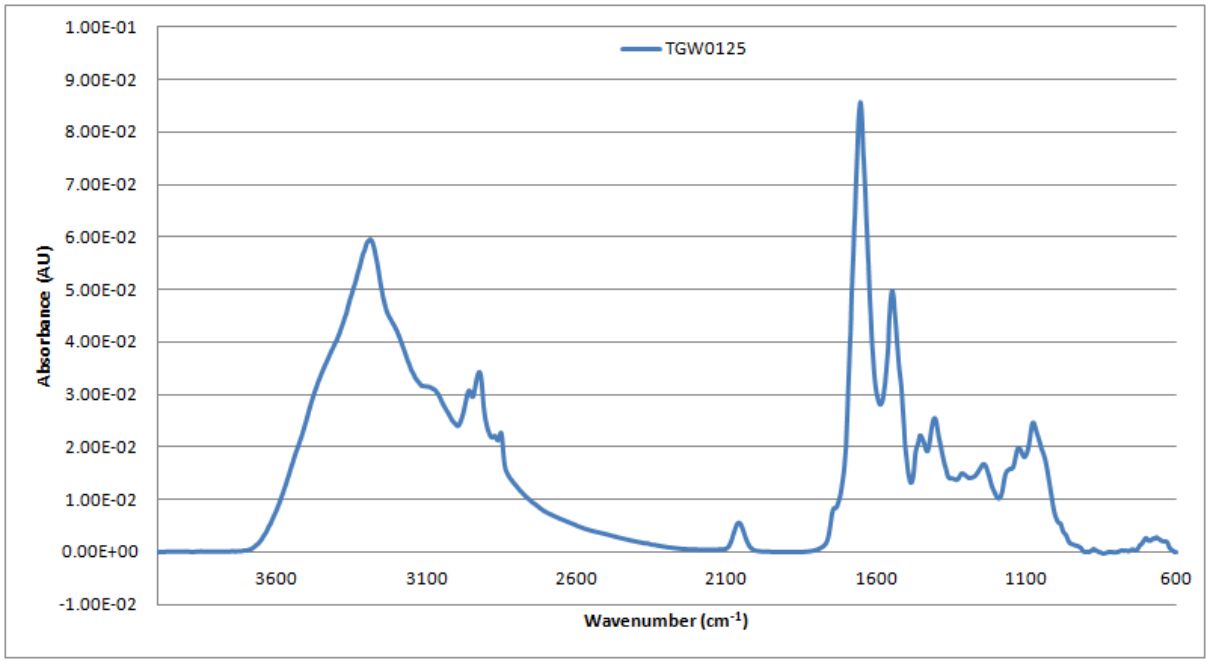


**Figure A2- 312: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0125 from 4000-600cm$^{-1}$**

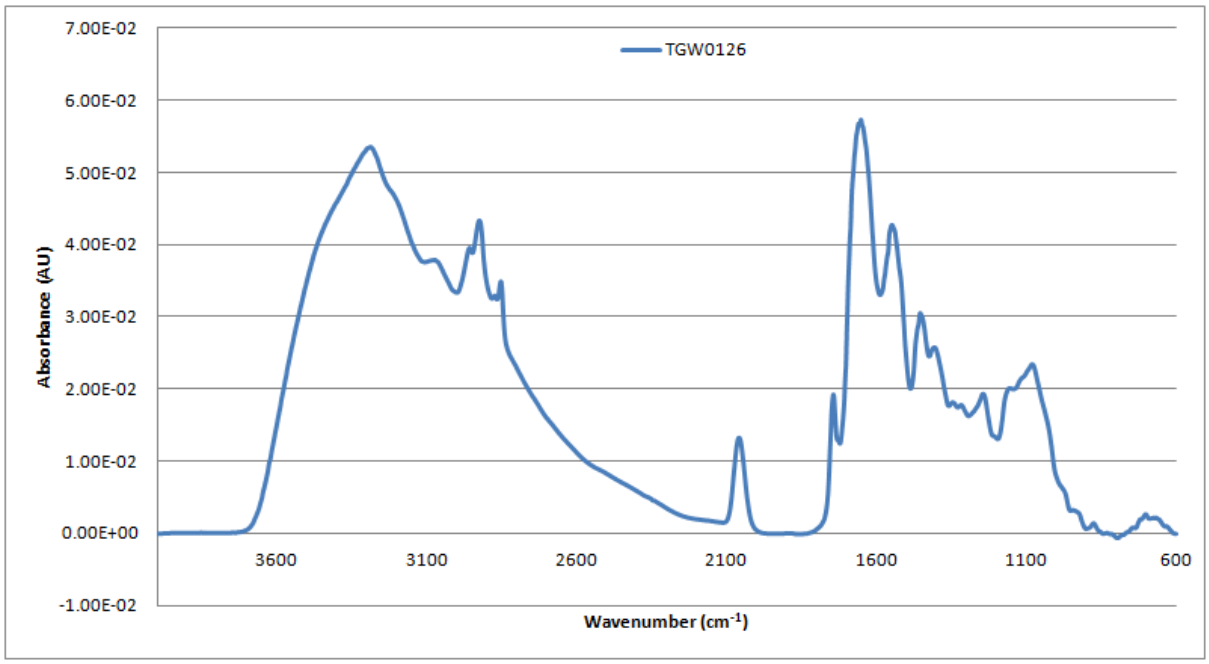


**Figure A2- 313: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0126 from 4000-600cm$^{-1}$**

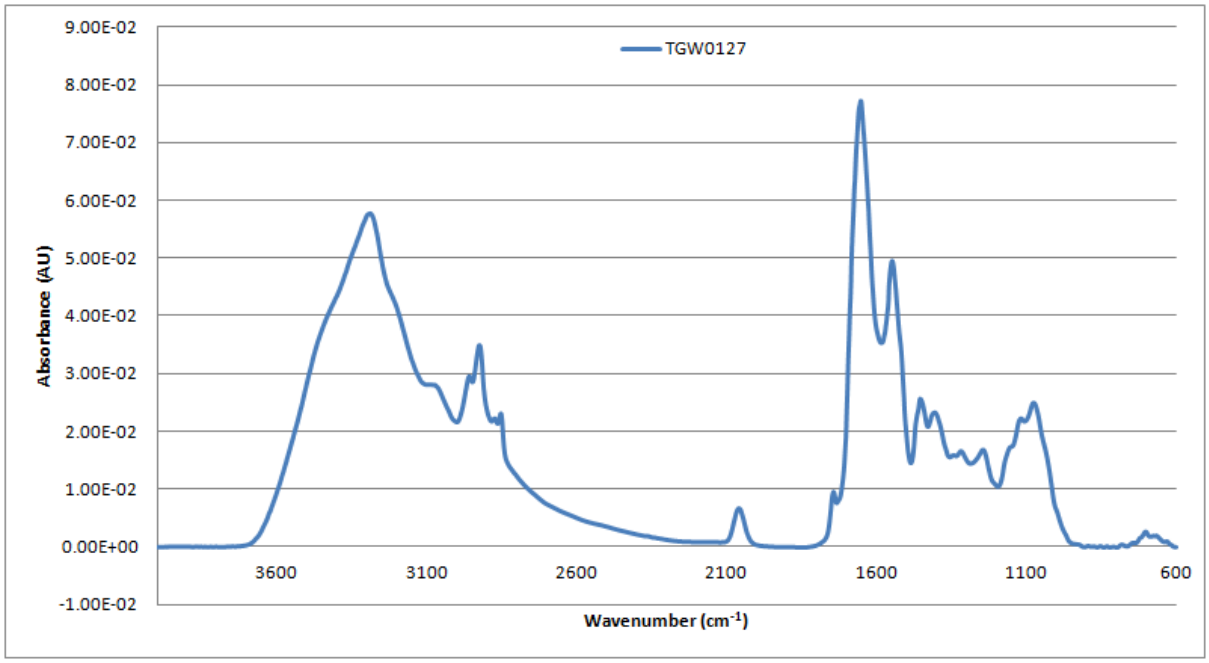


**Figure A2- 314: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0127 from 4000-600cm$^{-1}$**

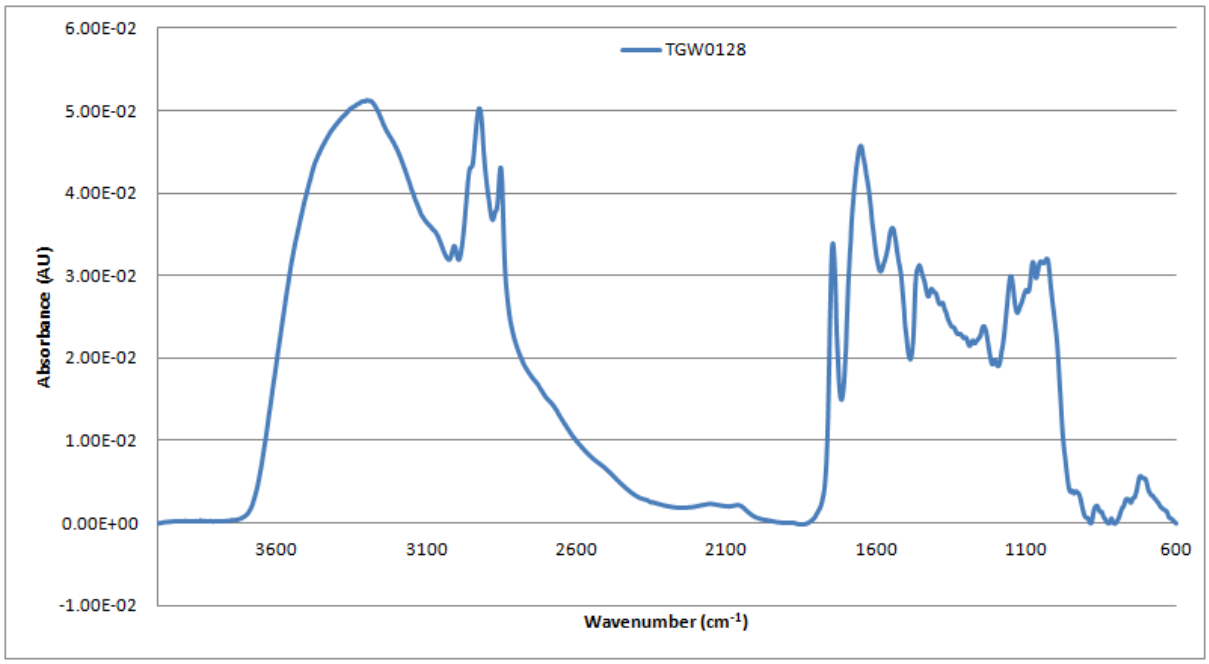


**Figure A2- 315: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0128 from 4000-600cm$^{-1}$**

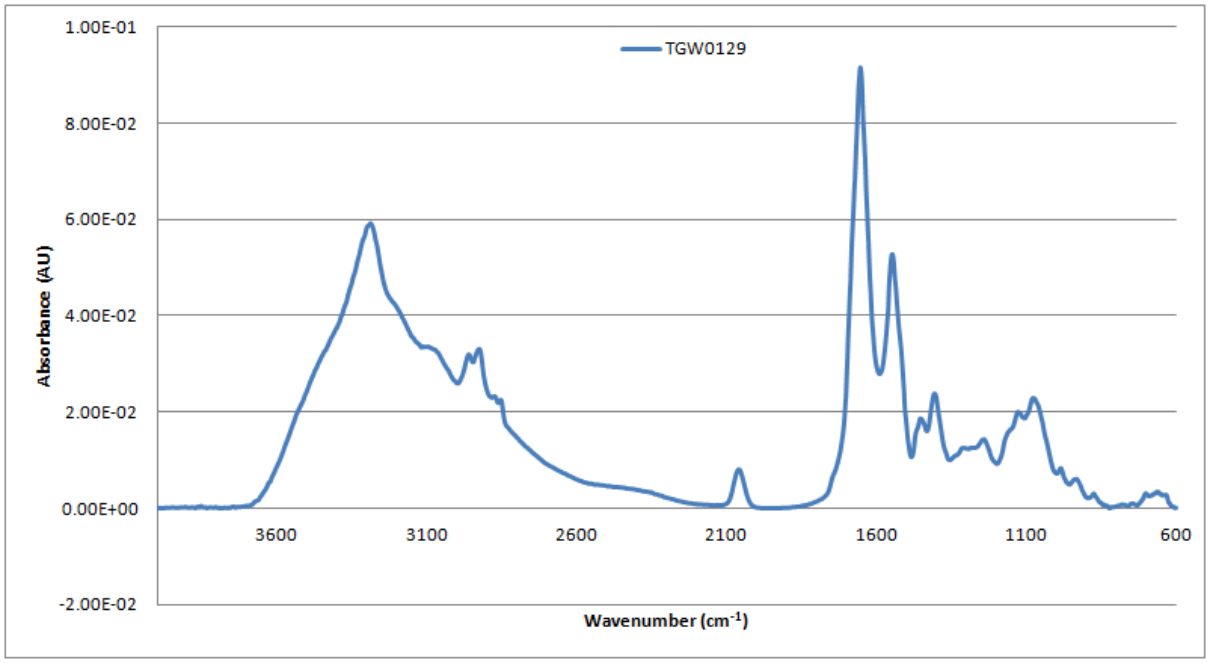


**Figure A2- 316: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0129 from 4000-600cm$^{-1}$**

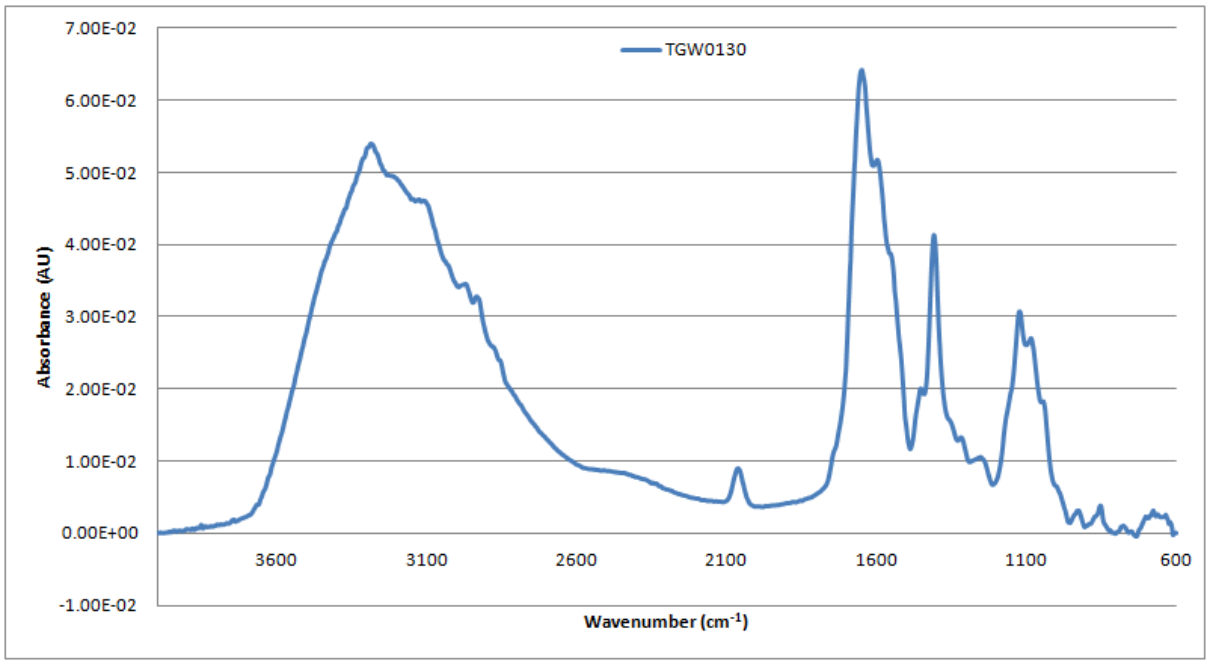


**Figure A2- 317: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0130 from 4000-600cm$^{-1}$**

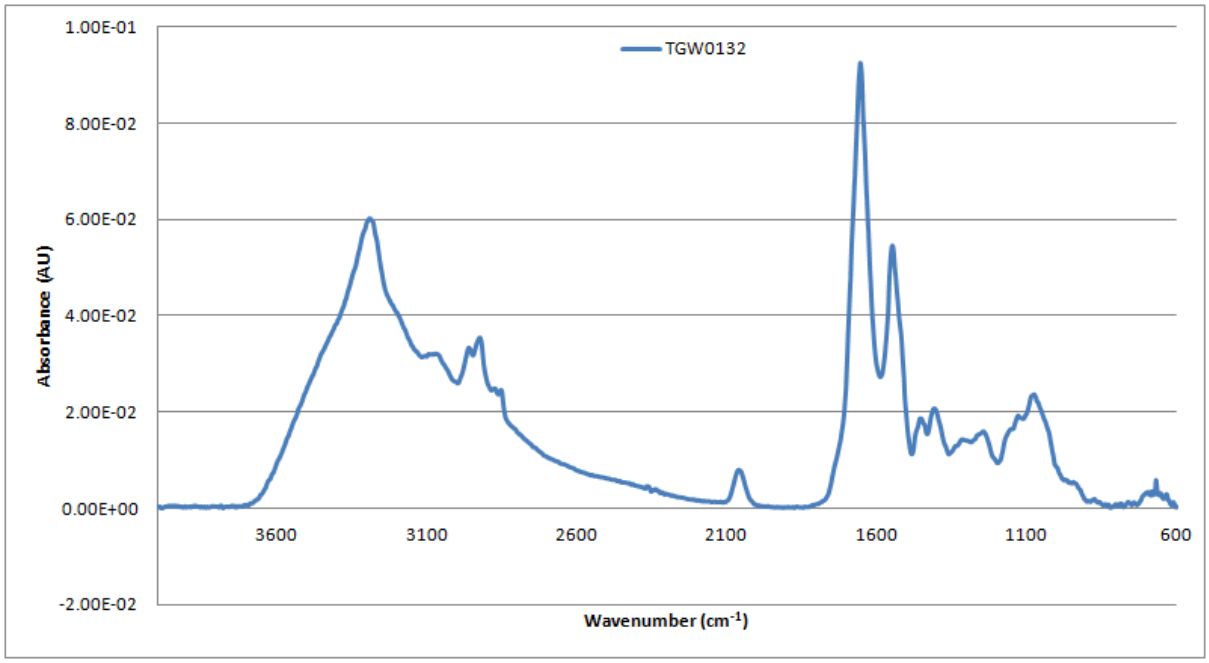


**Figure A2- 318: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0132 from 4000-600cm$^{-1}$**

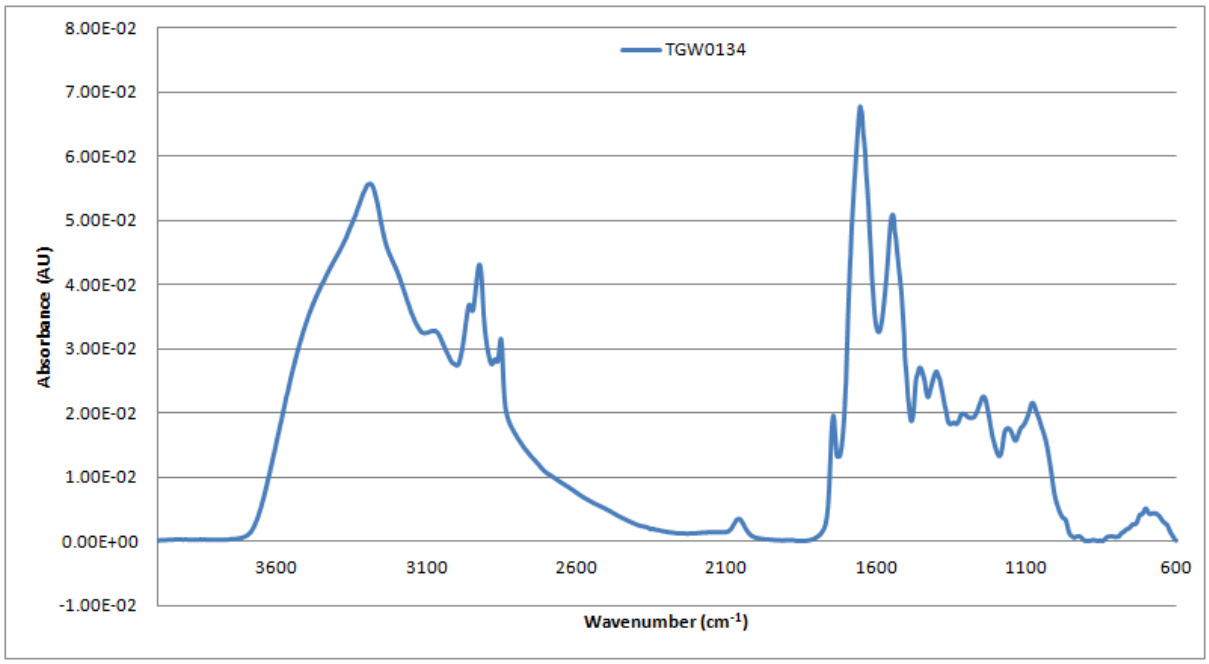


**Figure A2- 319: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0134 from 4000-600cm$^{-1}$**

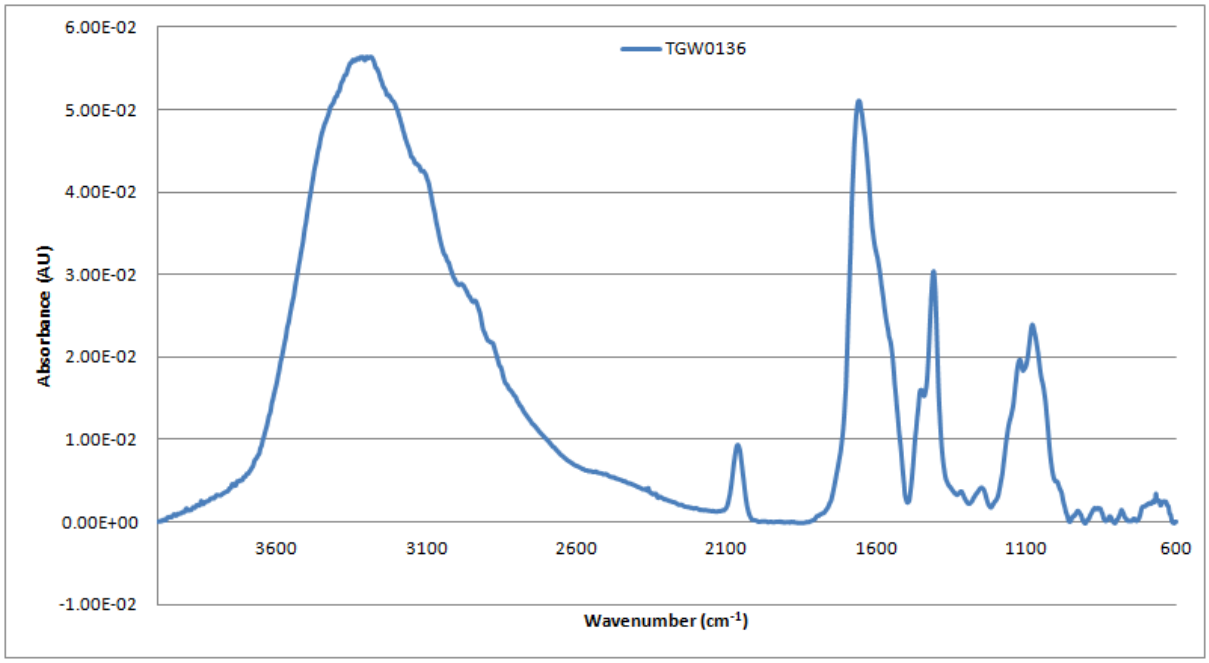


**Figure A2- 320: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0136 from 4000-600cm$^{-1}$**

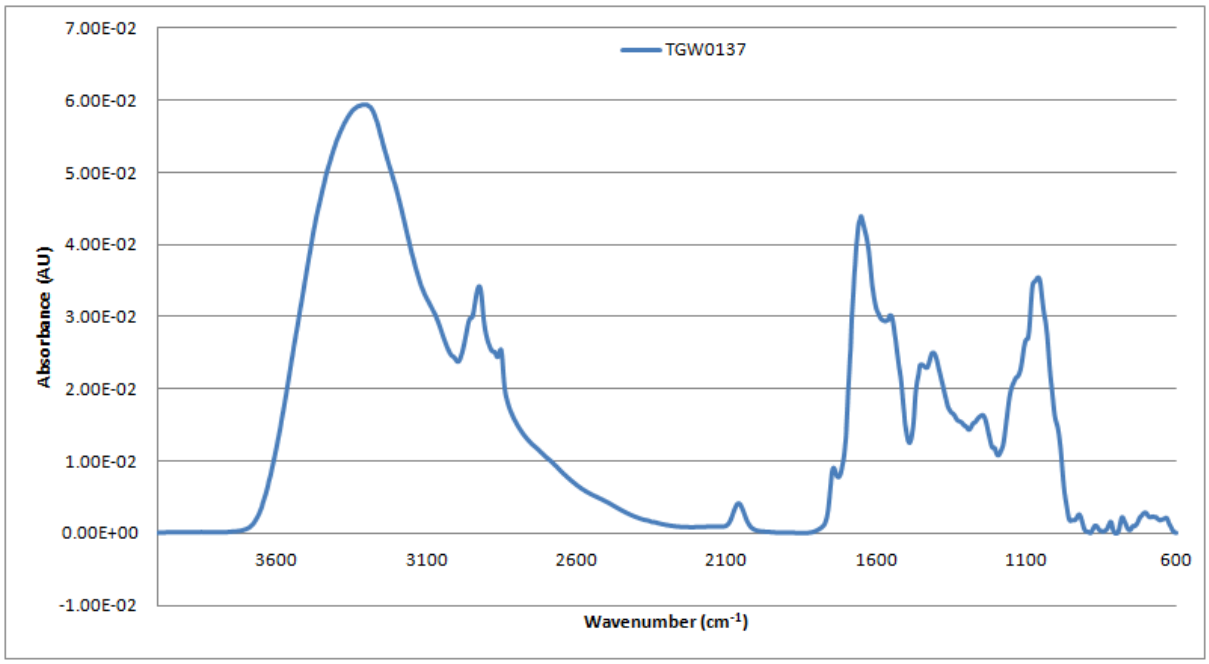


**Figure A2- 321: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0137 from 4000-600cm$^{-1}$**

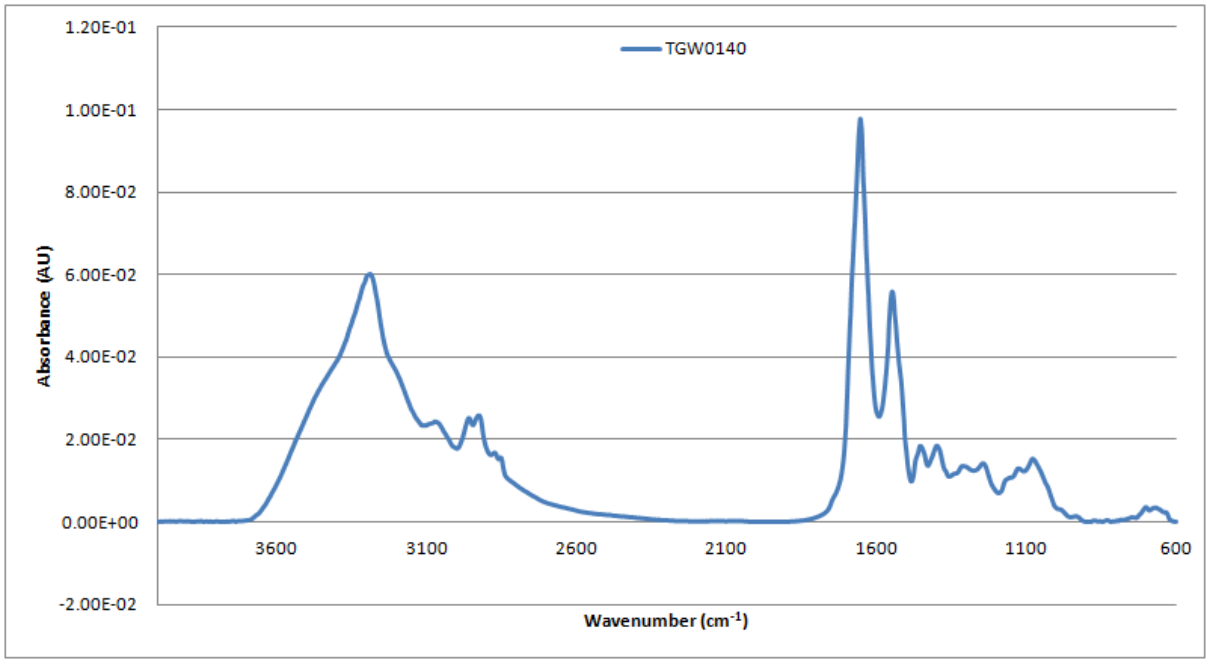


**Figure A2- 322: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0140 from 4000-600cm$^{-1}$**

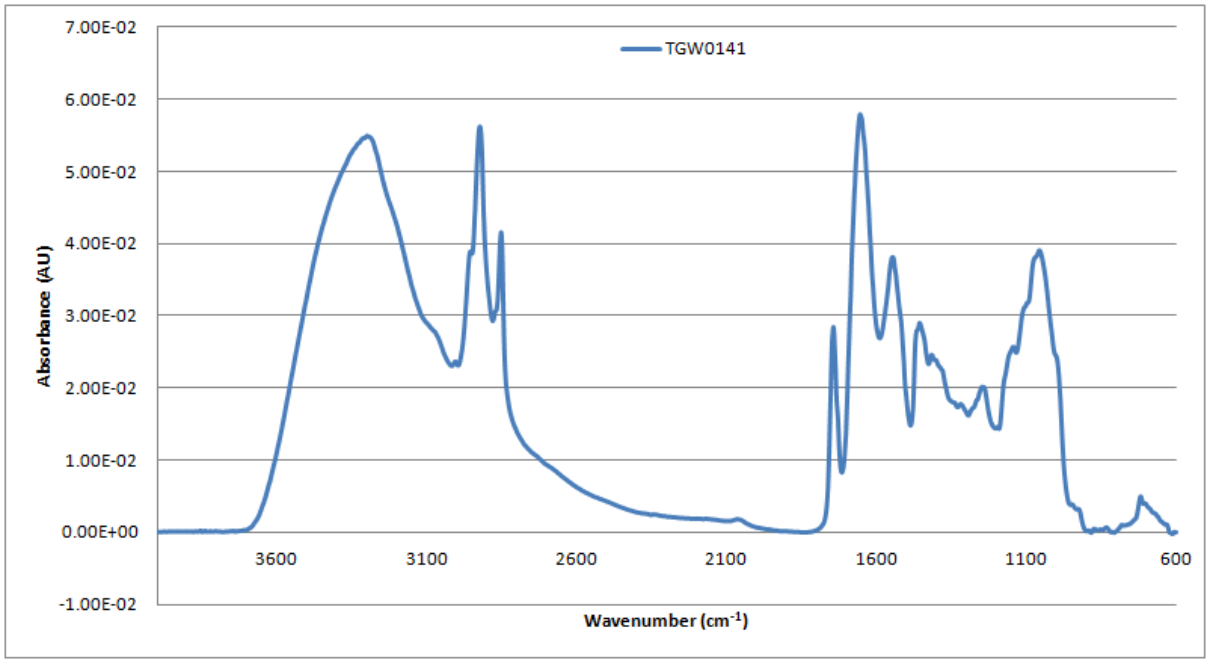


**Figure A2- 323: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0141 from 4000-600cm$^{-1}$**

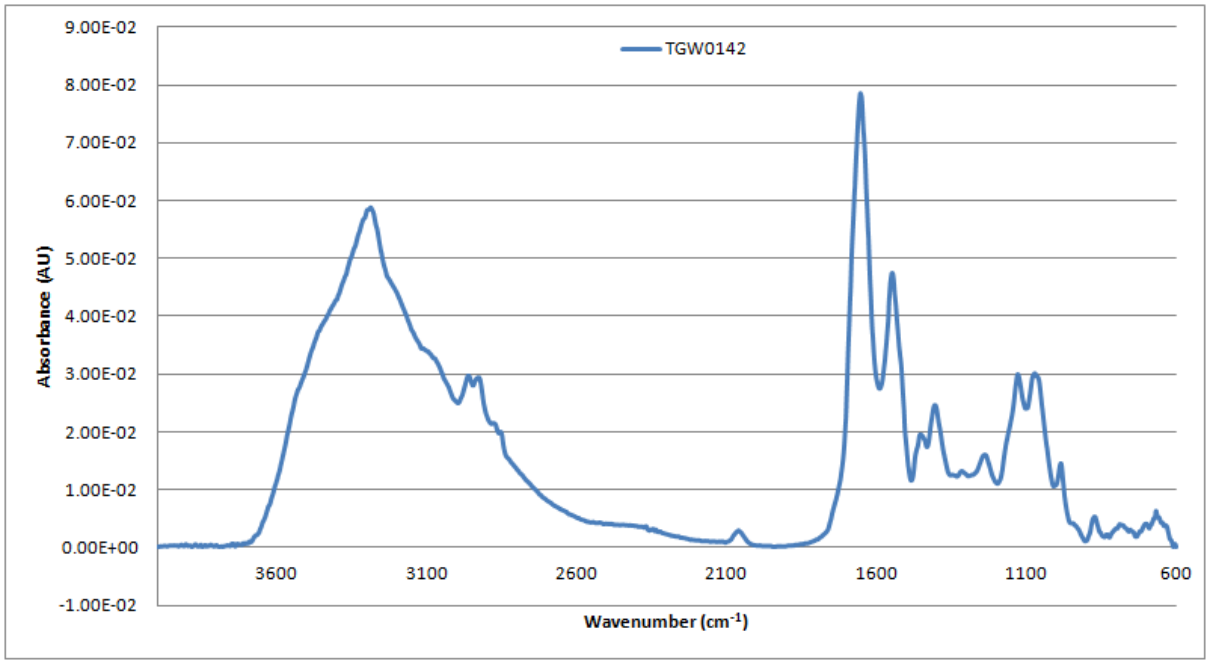


**Figure A2- 324: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0142 from 4000-600cm$^{-1}$**

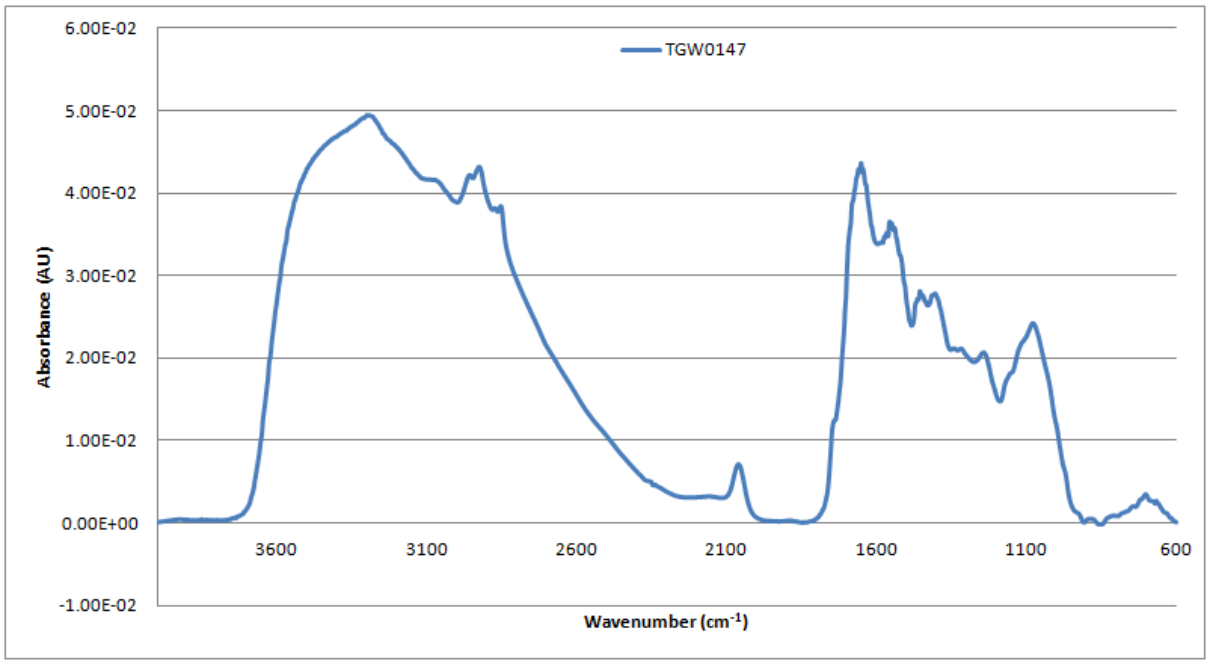


**Figure A2- 325: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0147 from 4000-600cm$^{-1}$**

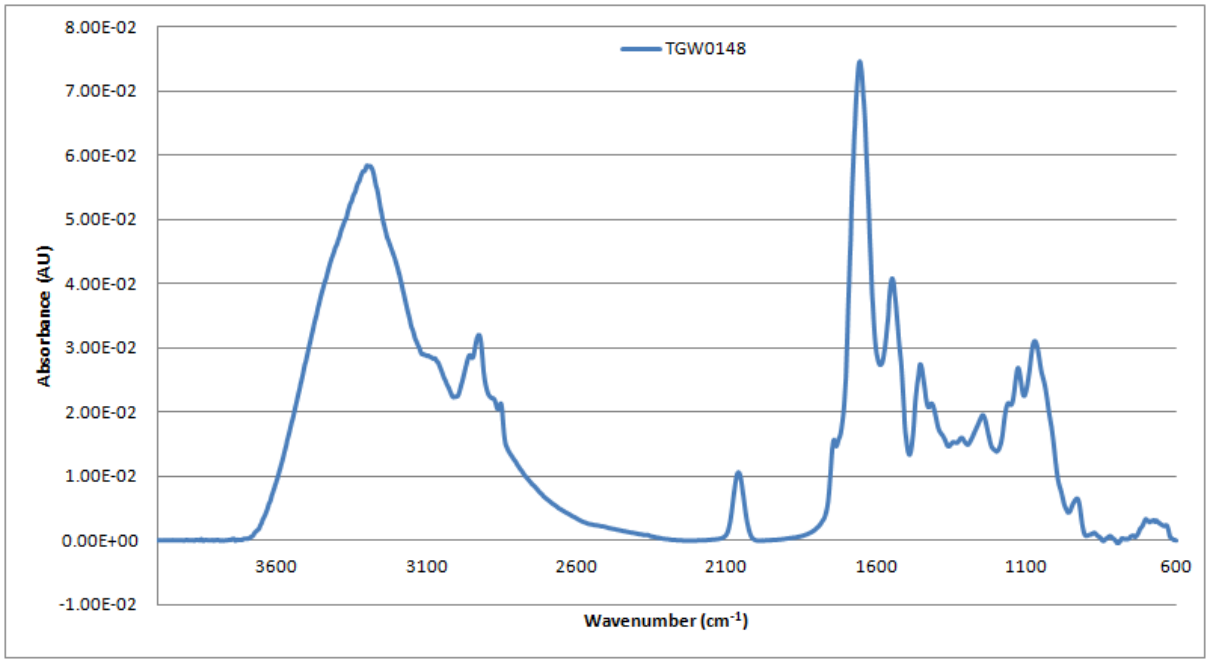


**Figure A2- 326: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0148 from 4000-600cm$^{-1}$**

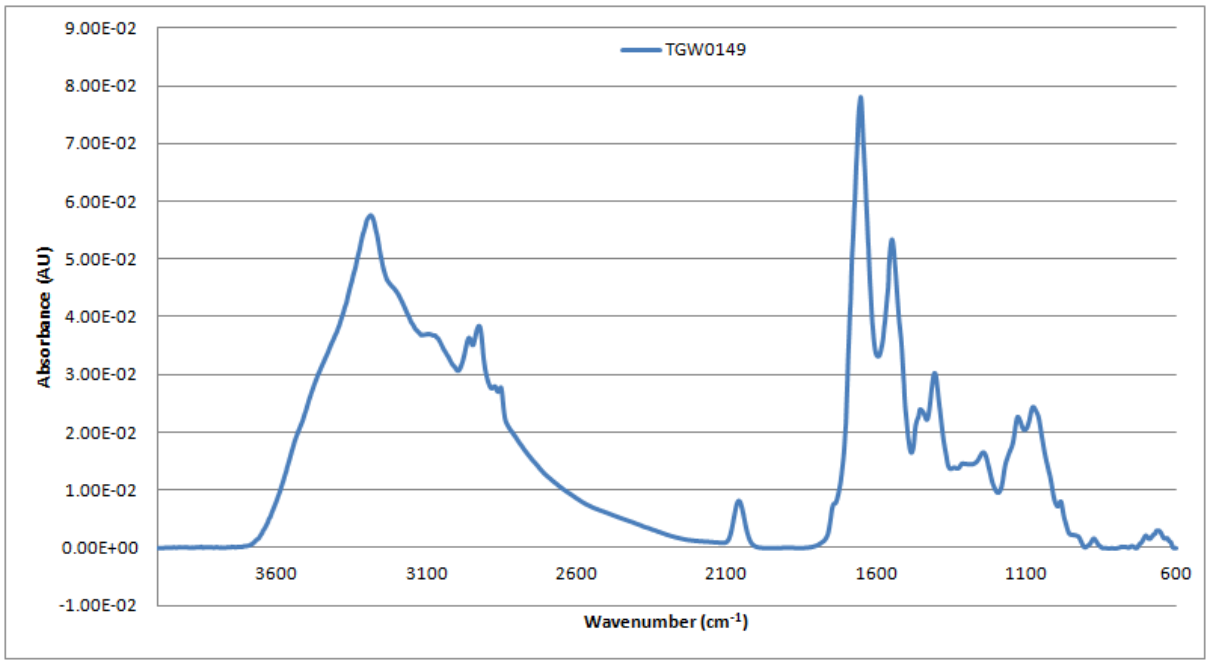


**Figure A2- 327: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0149 from 4000-600cm$^{-1}$**

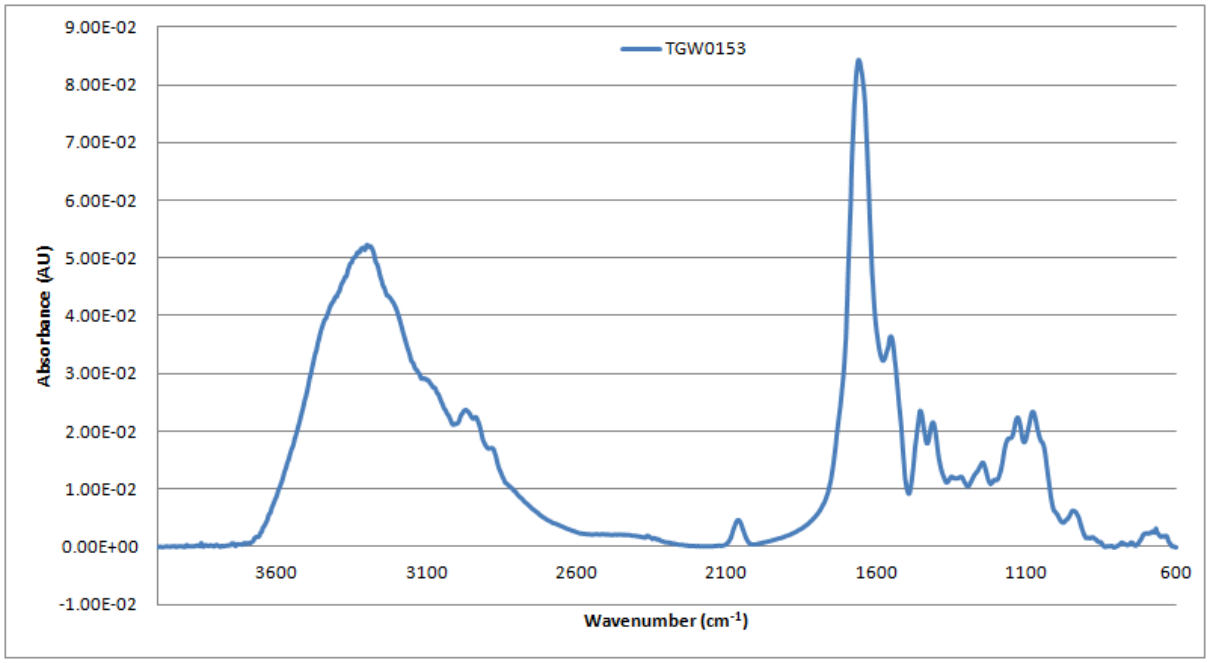


**Figure A2- 328: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0153 from 4000-600cm$^{-1}$**

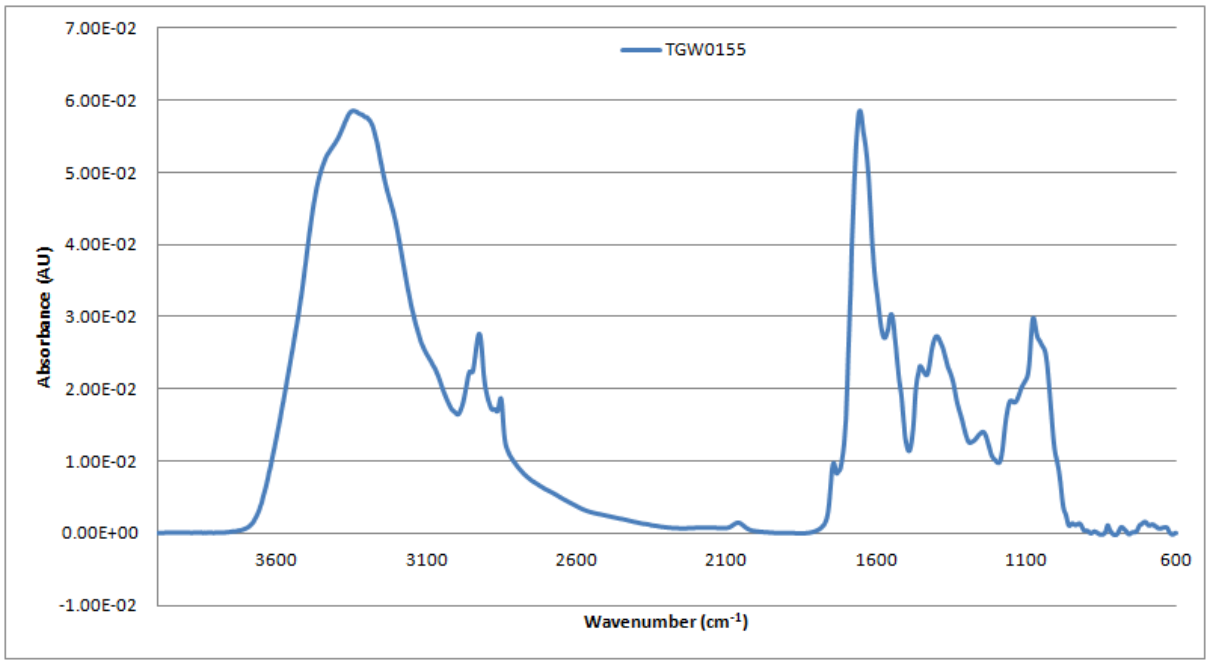


**Figure A2- 329: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0155 from 4000-600cm$^{-1}$**

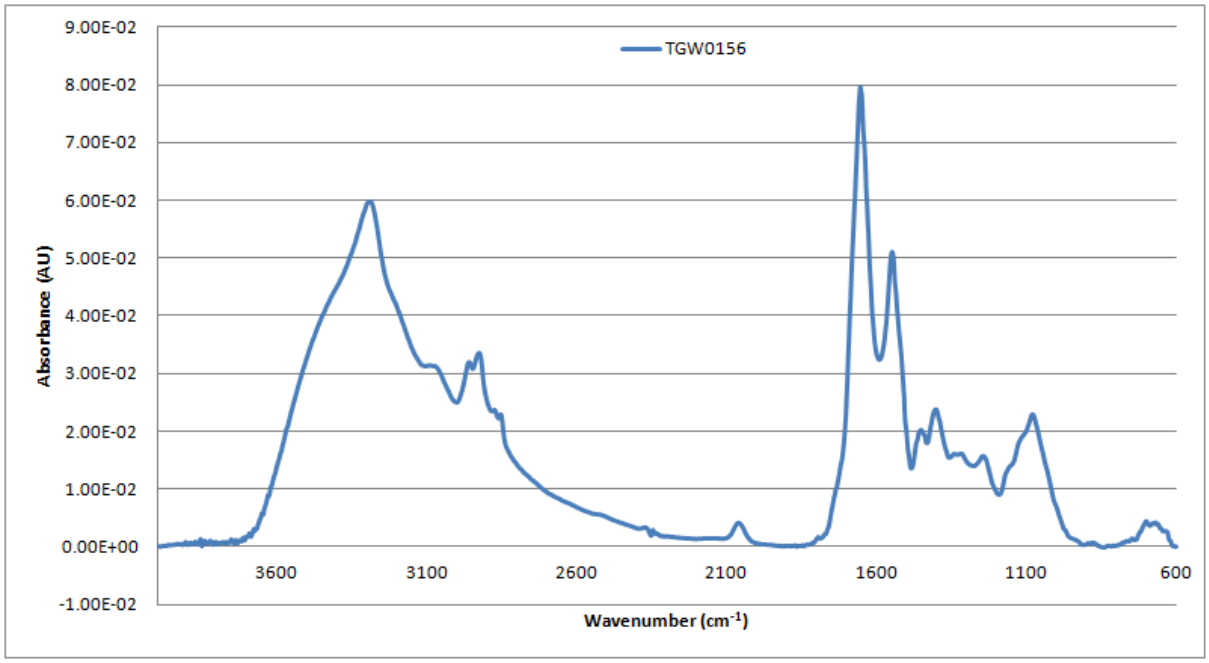


**Figure A2- 330: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0156 from 4000-600cm$^{-1}$**

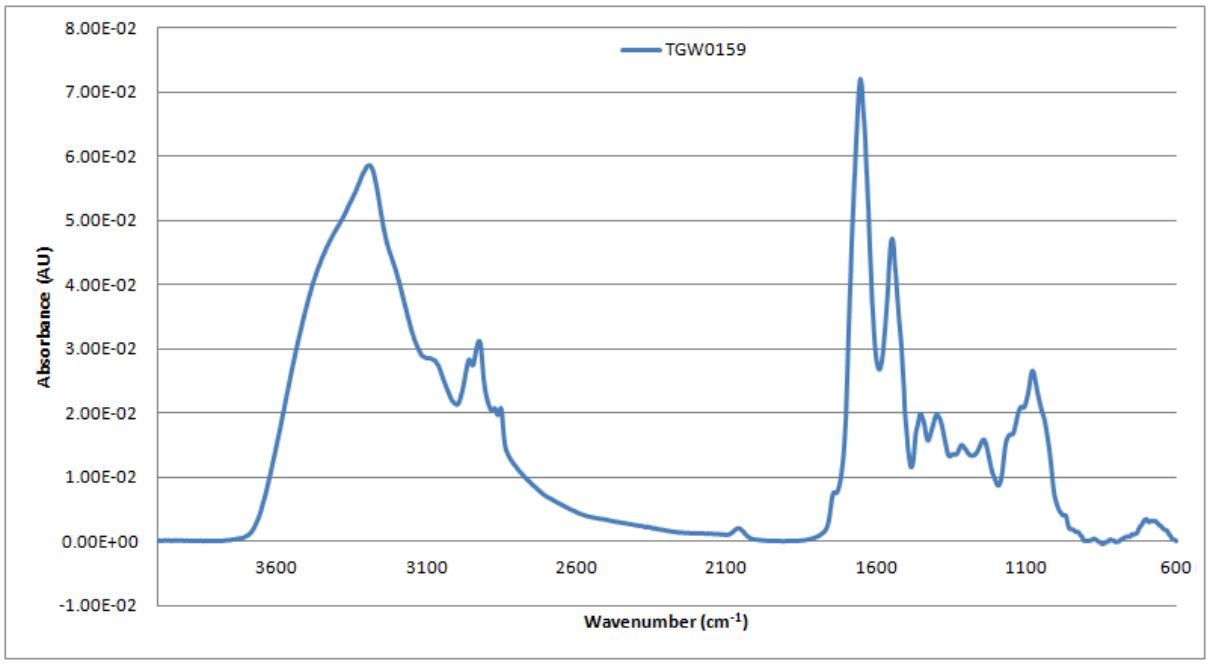


**Figure A2- 331: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0159 from 4000-600cm$^{-1}$**

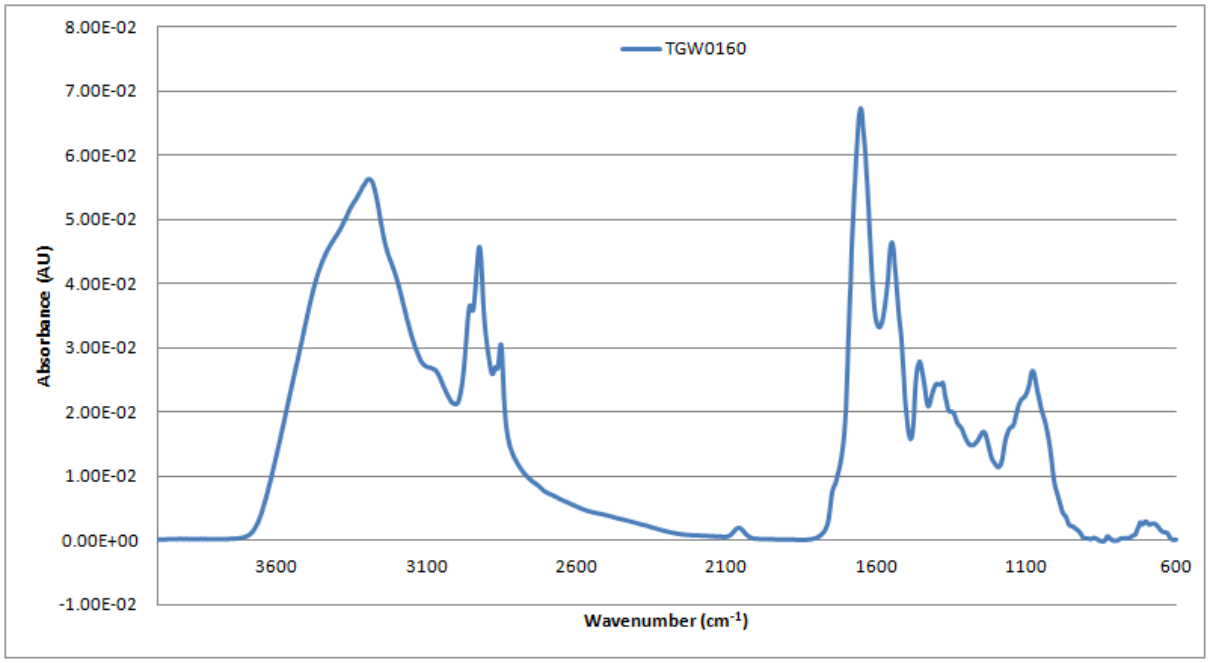


**Figure A2- 332: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0160 from 4000-600cm$^{-1}$**

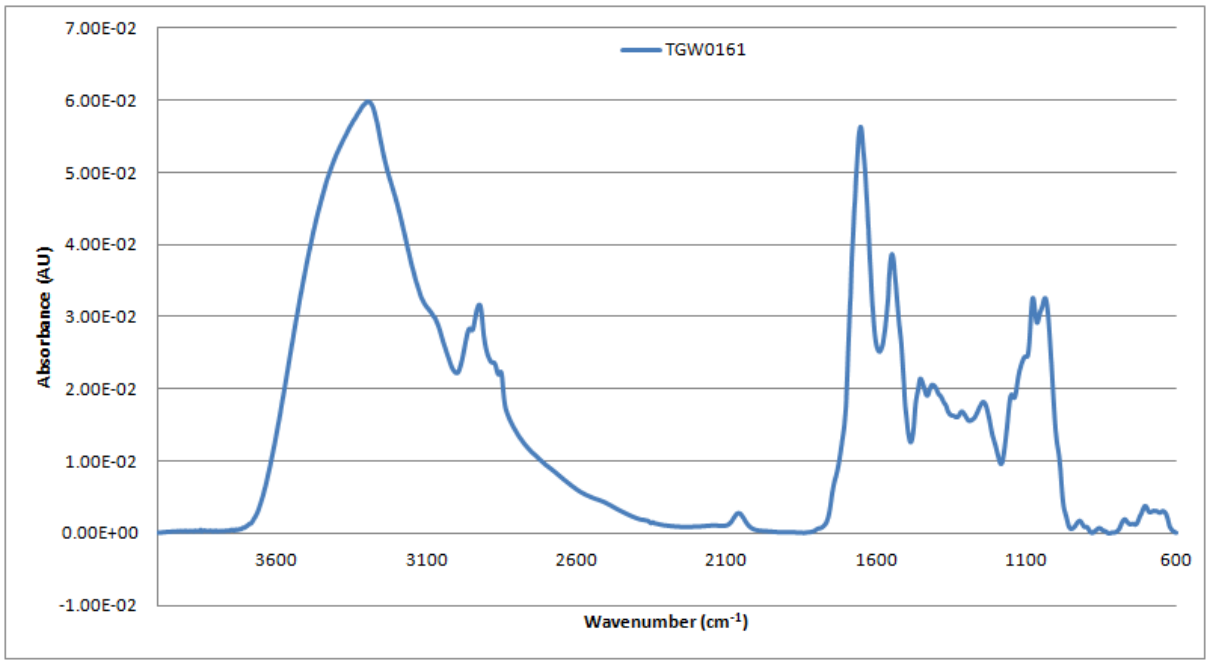


**Figure A2- 333: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0161 from 4000-600cm$^{-1}$**

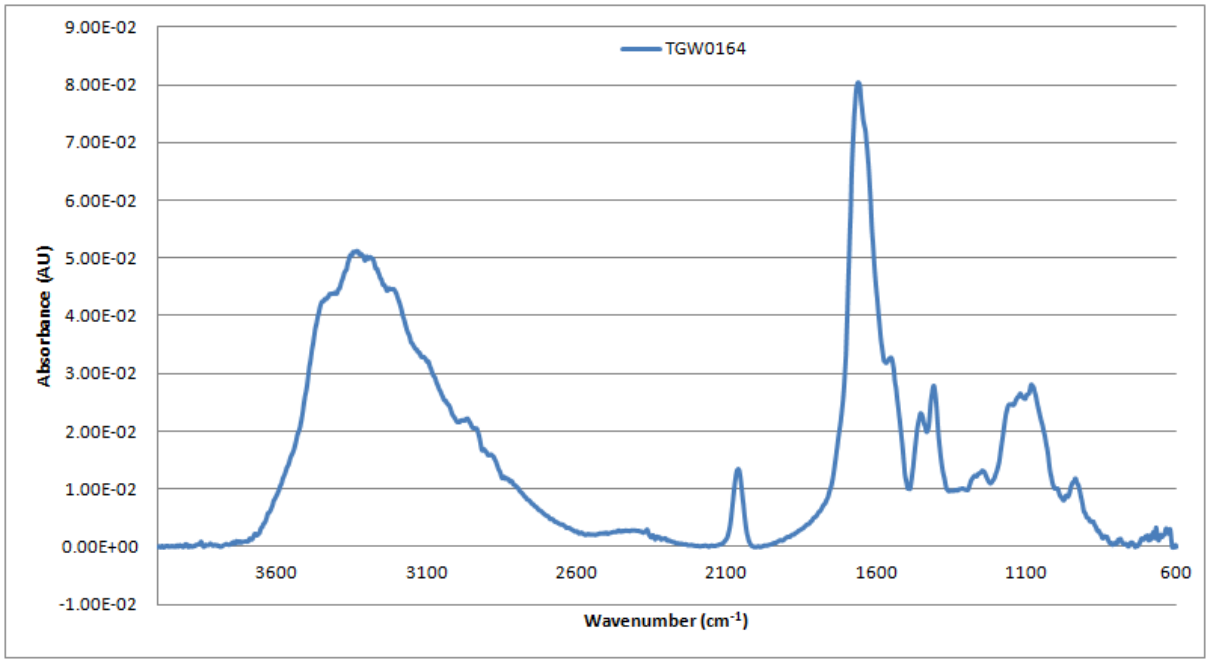


**Figure A2- 334: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0164 from 4000-600cm$^{-1}$**

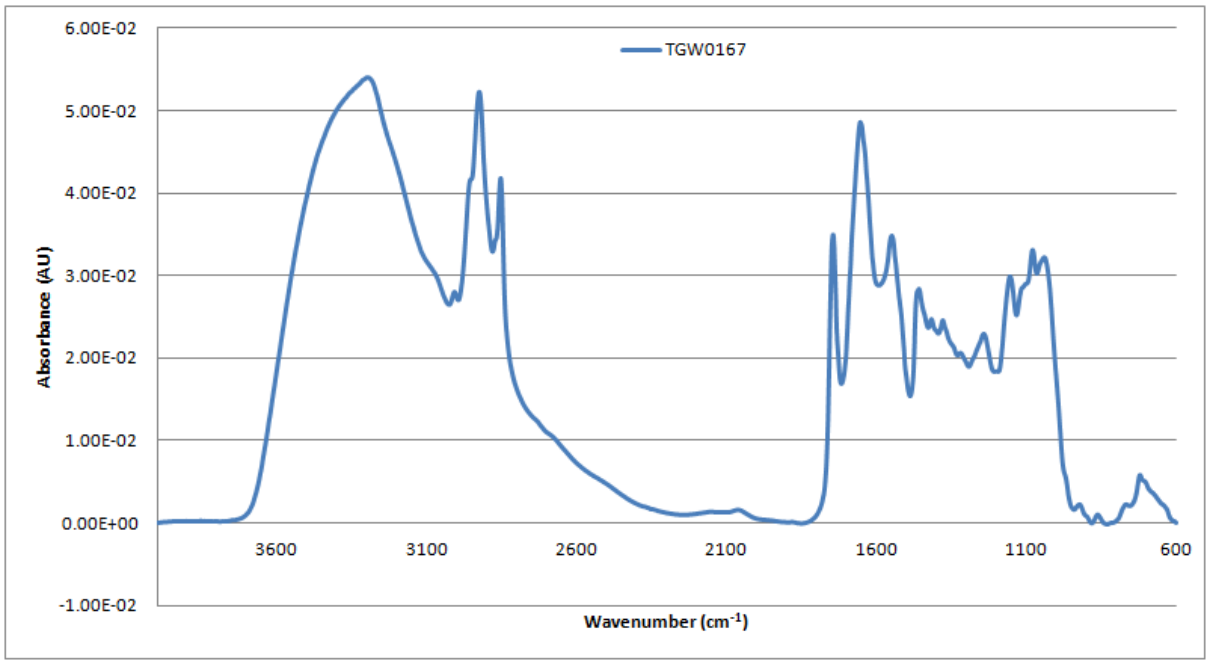


**Figure A2- 335: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0167 from 4000-600cm$^{-1}$**

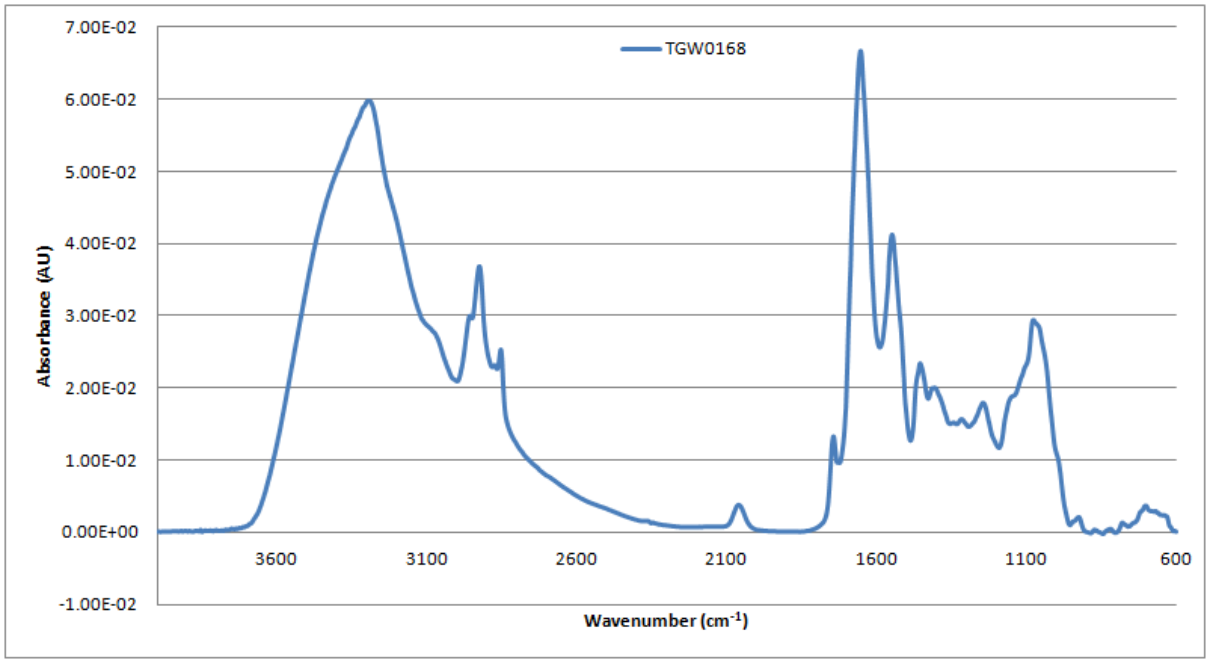


**Figure A2- 336: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0168 from 4000-600cm$^{-1}$**

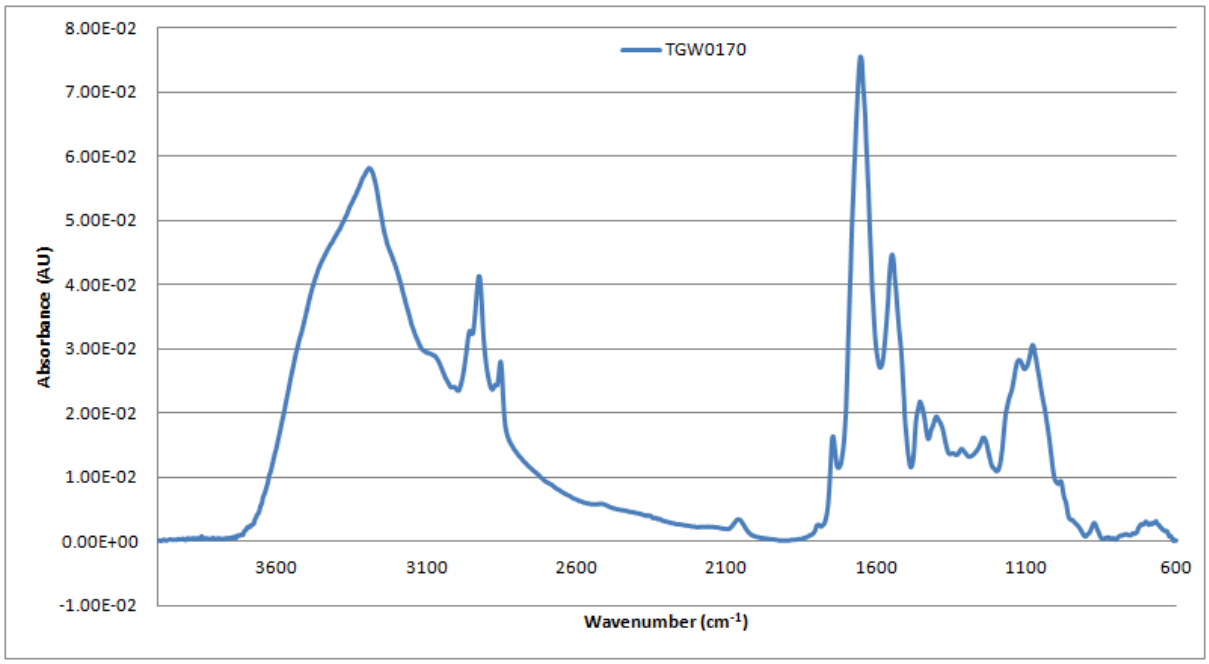


**Figure A2- 337: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0170 from 4000-600cm$^{-1}$**

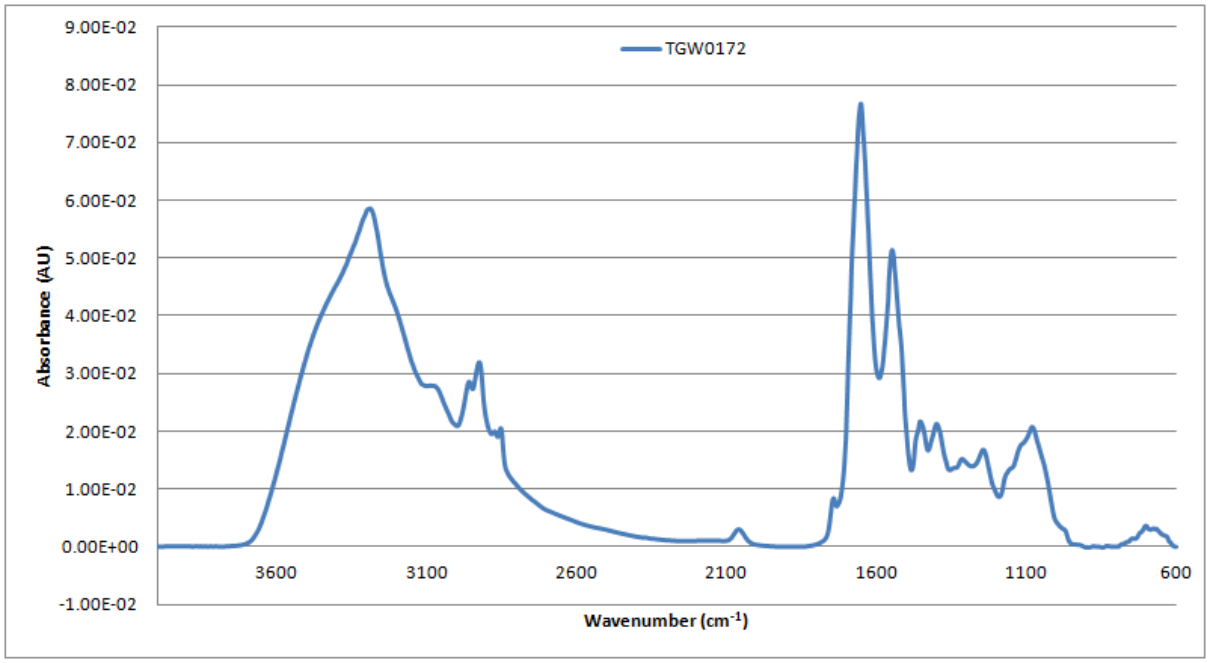


**Figure A2- 338: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0172 from 4000-600cm$^{-1}$**

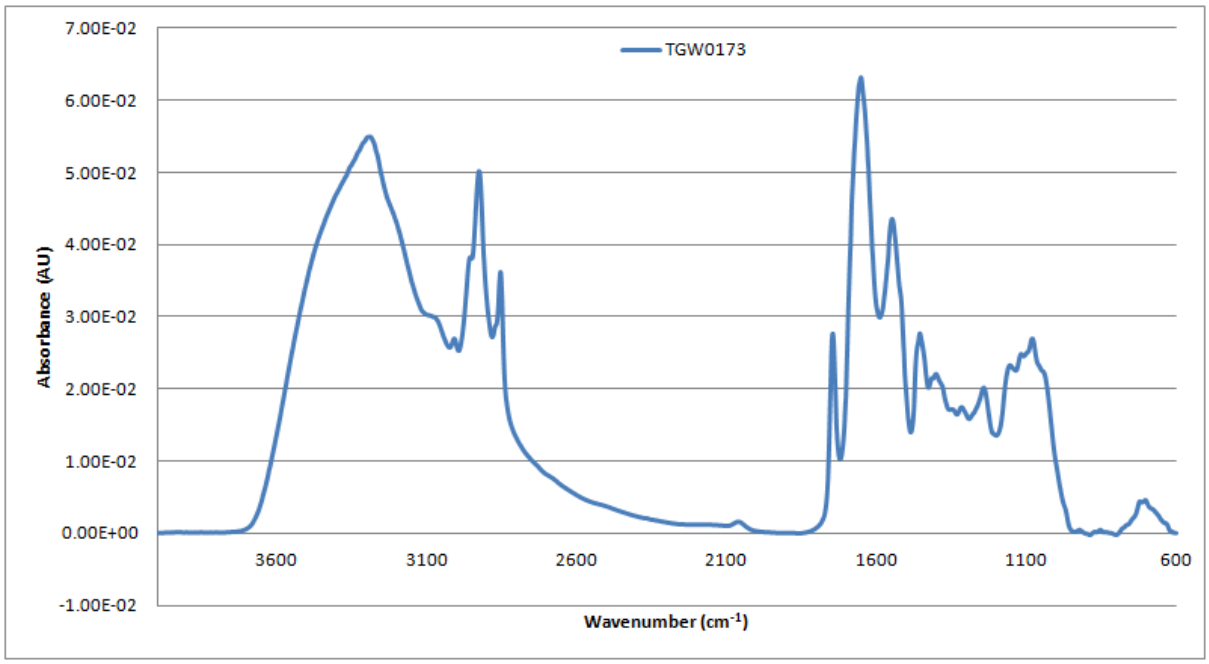


**Figure A2- 339: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0173 from 4000-600cm$^{-1}$**

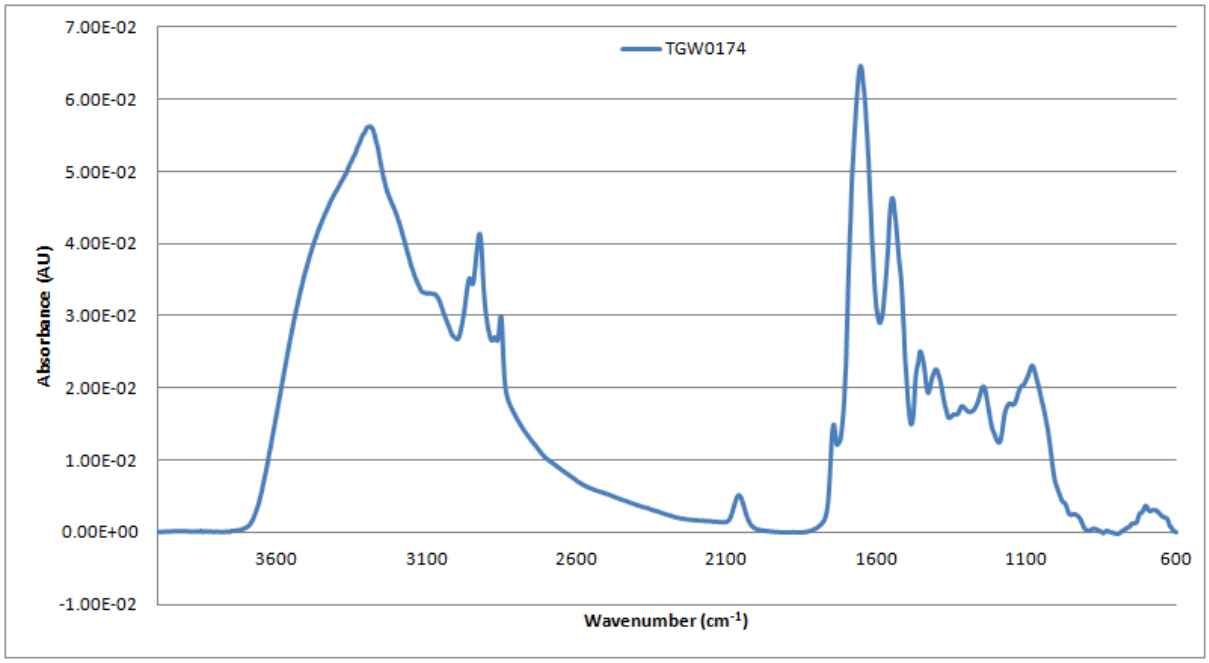


**Figure A2- 340: Vector-normalised, baseline-corrected absorbance spectrum of MEDLUNG lung cancer patient TGW0174 from 4000-600cm$^{-1}$**

## Appendix 3: SPEDIC Patients' Longitudinal Q-values plots

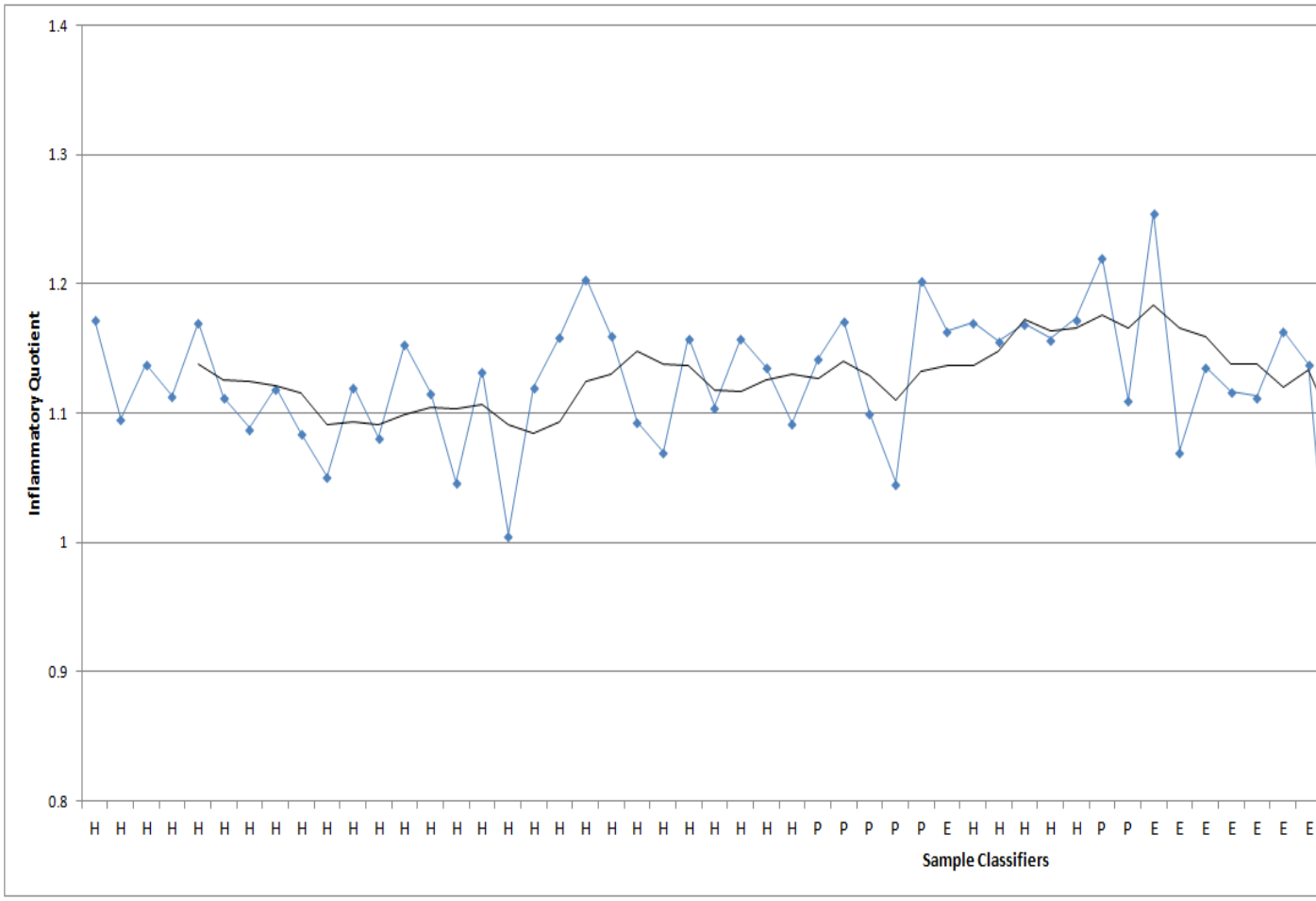


**Figure A3-1: Inflammatory quotient values (blue) for patient SP01, plotted over time, with sample classifications exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

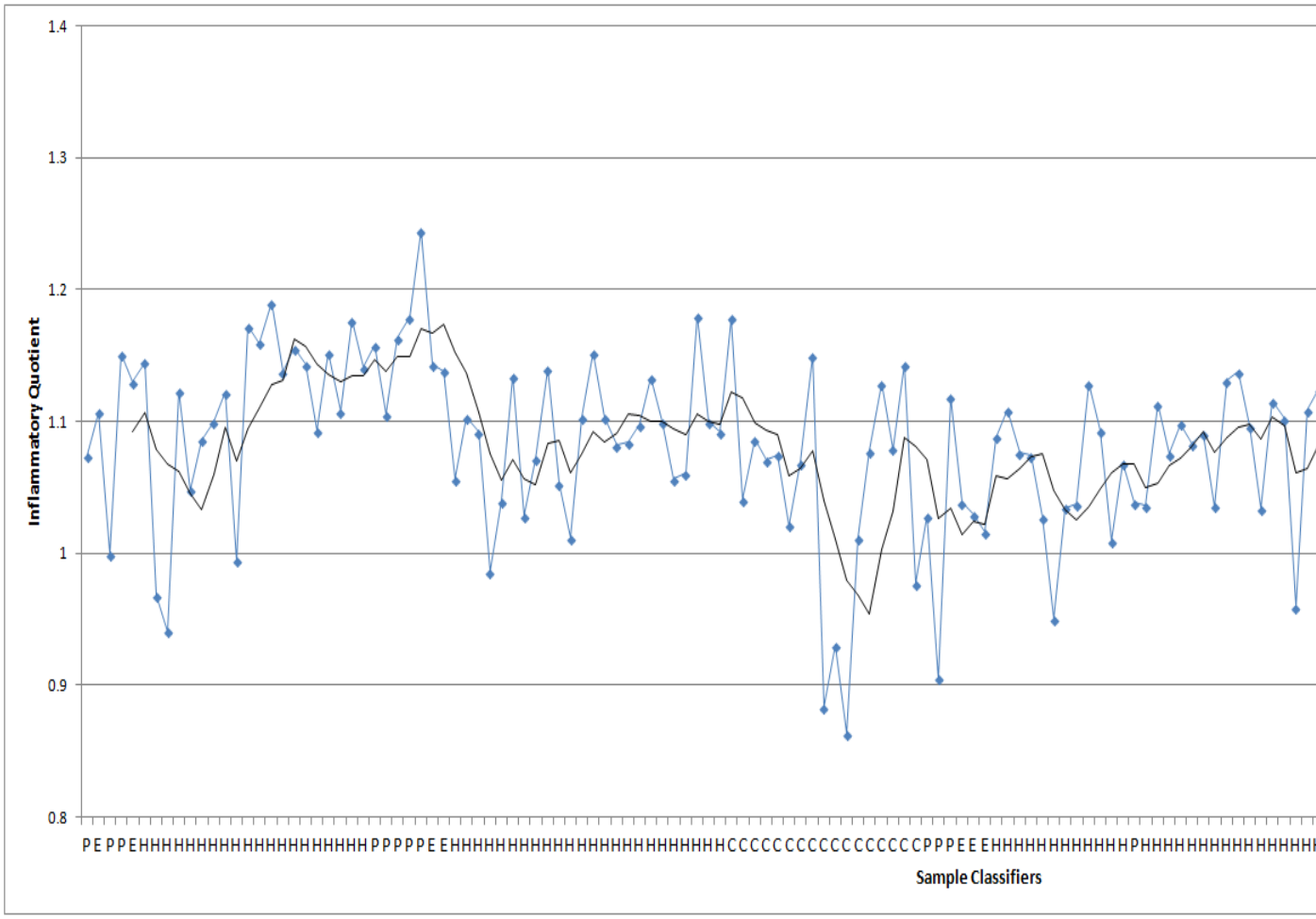


**Figure A3-2: Inflammatory quotient values (blue) for patient SP02, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

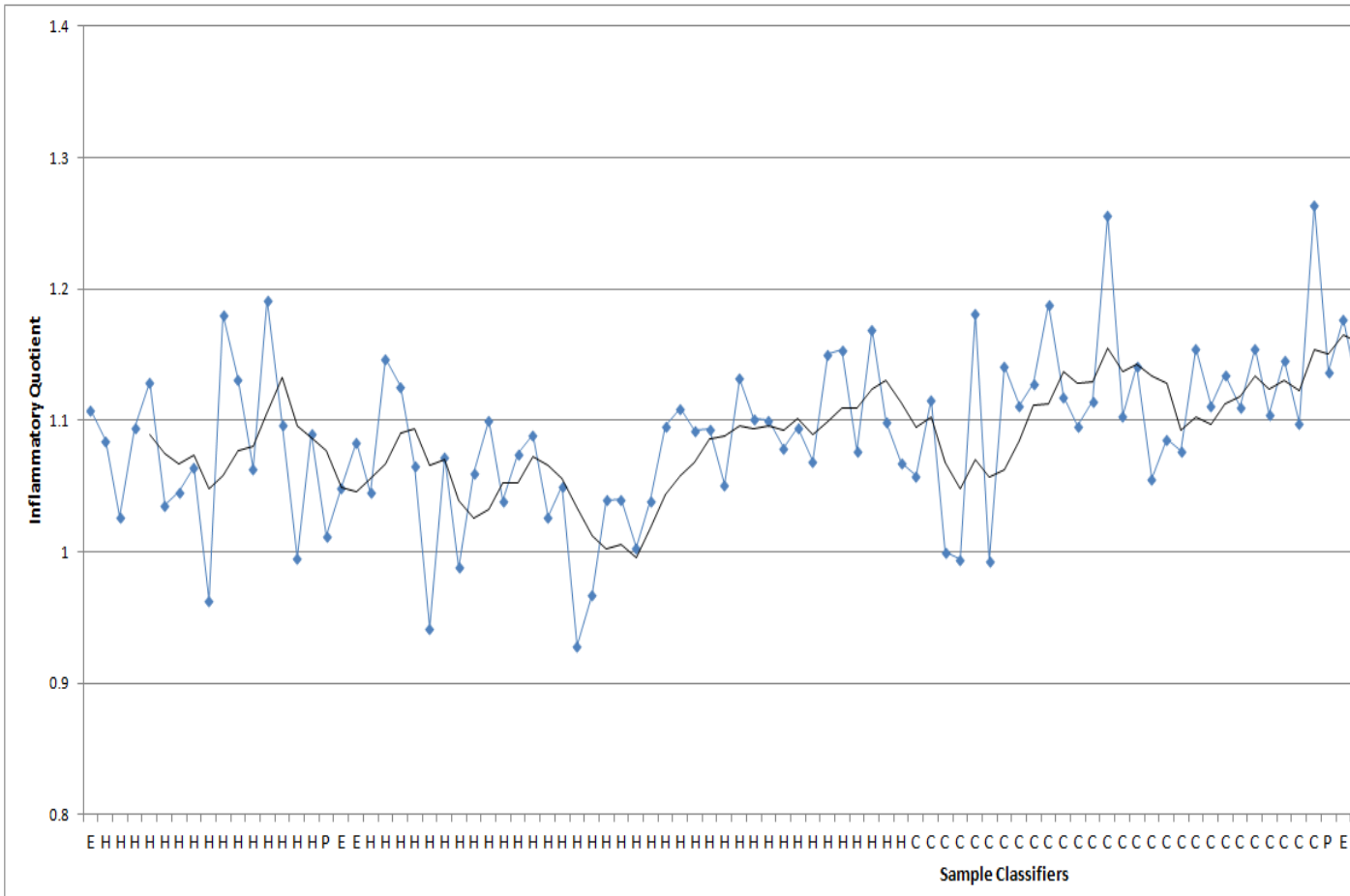

**Figure A3-3: Inflammatory quotient values (blue) for patient SP05, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

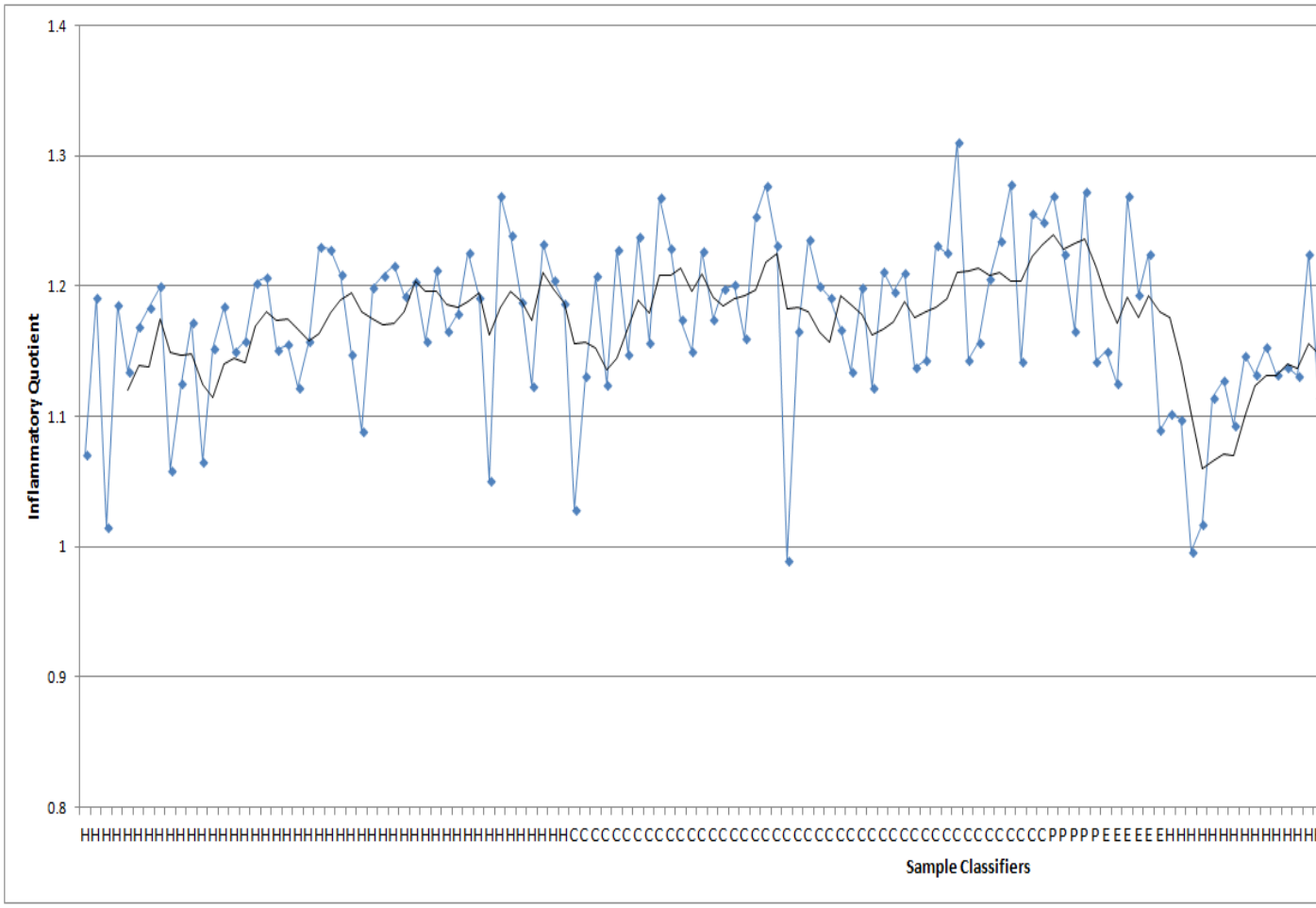


**Figure A3-4: Inflammatory quotient values (blue) for patient SP06, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

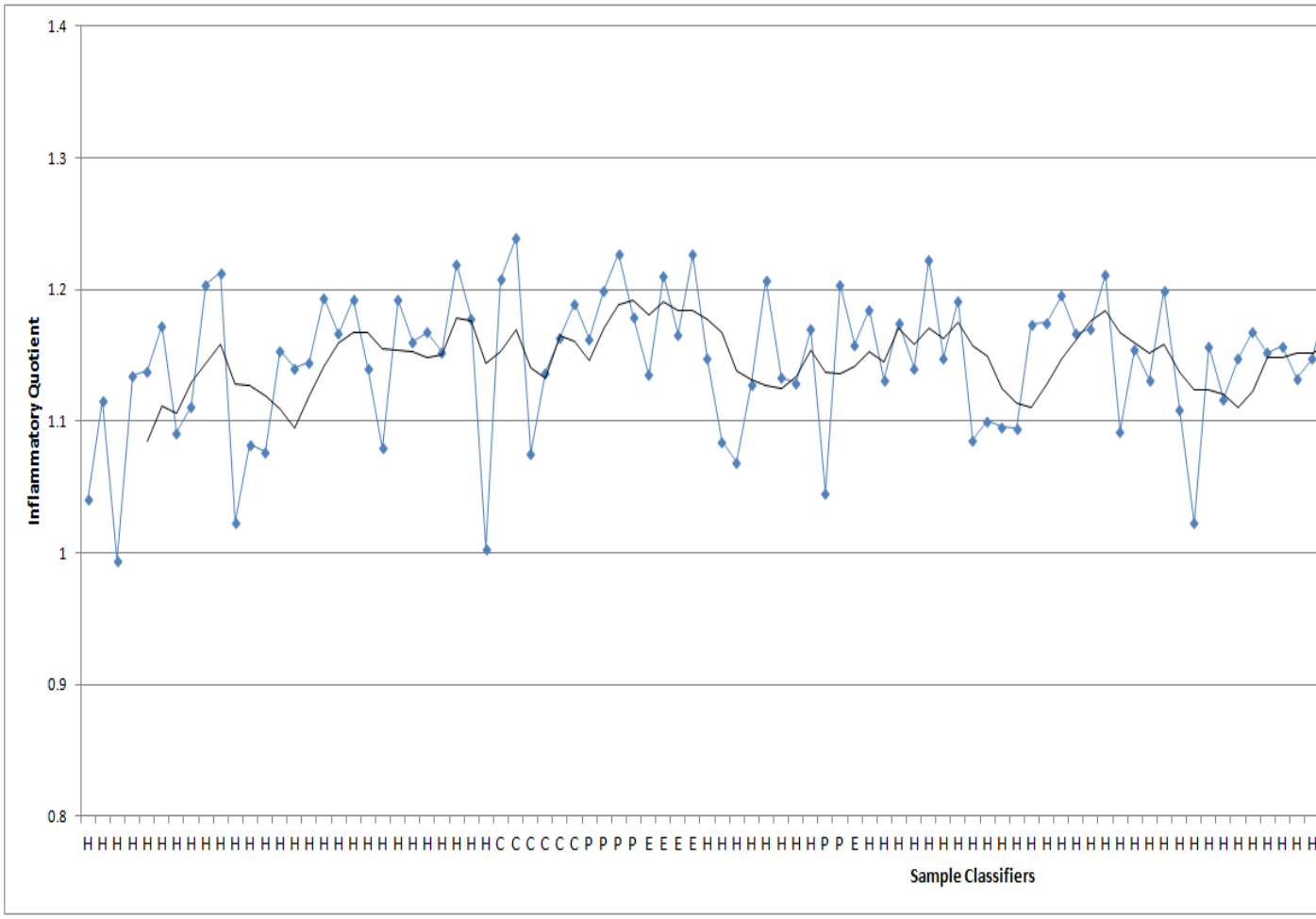

**Figure A3-5: Inflammatory quotient values (blue) for patient SP07, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

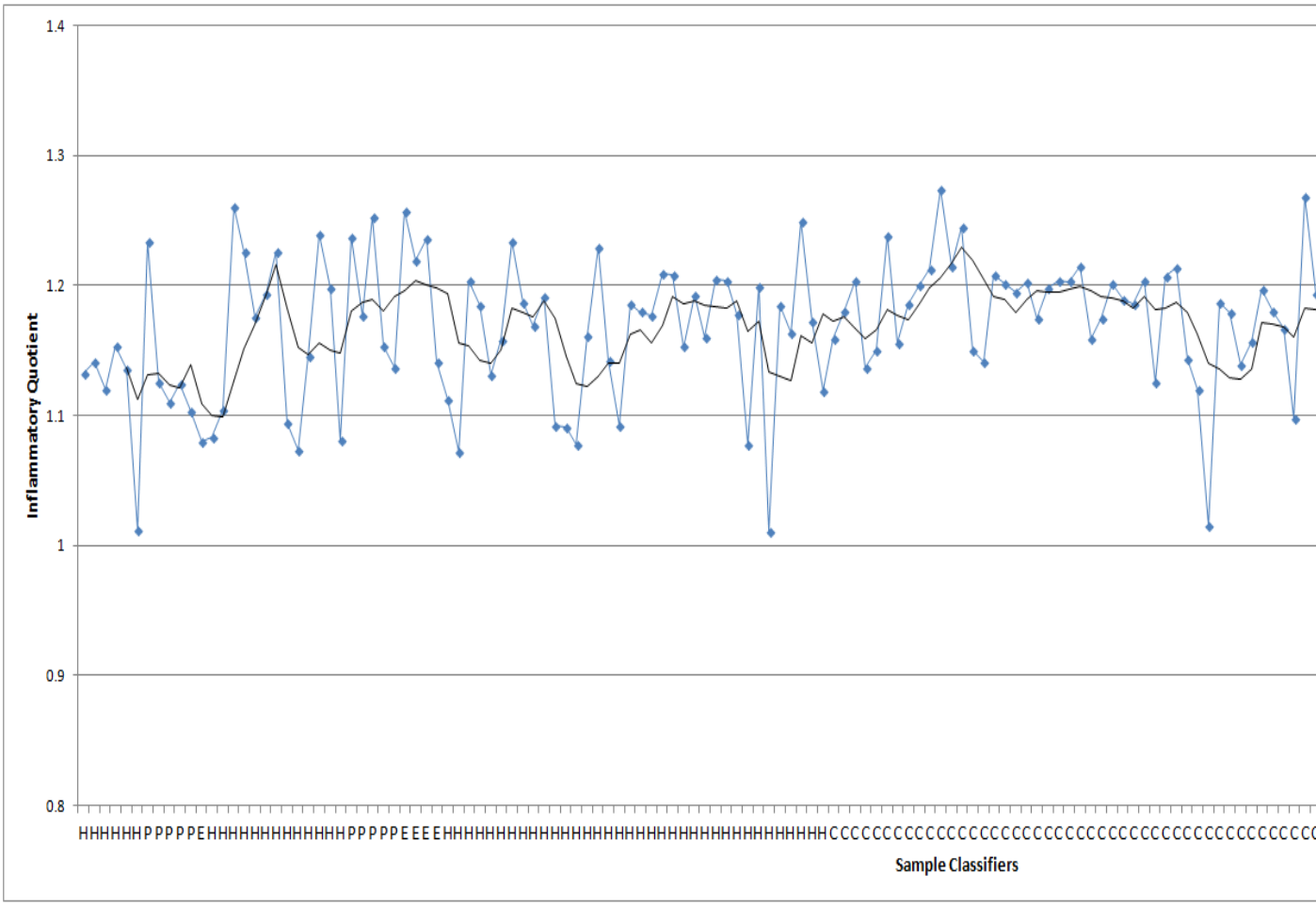

**Figure A3-6: Inflammatory quotient values (blue) for patient SP09, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

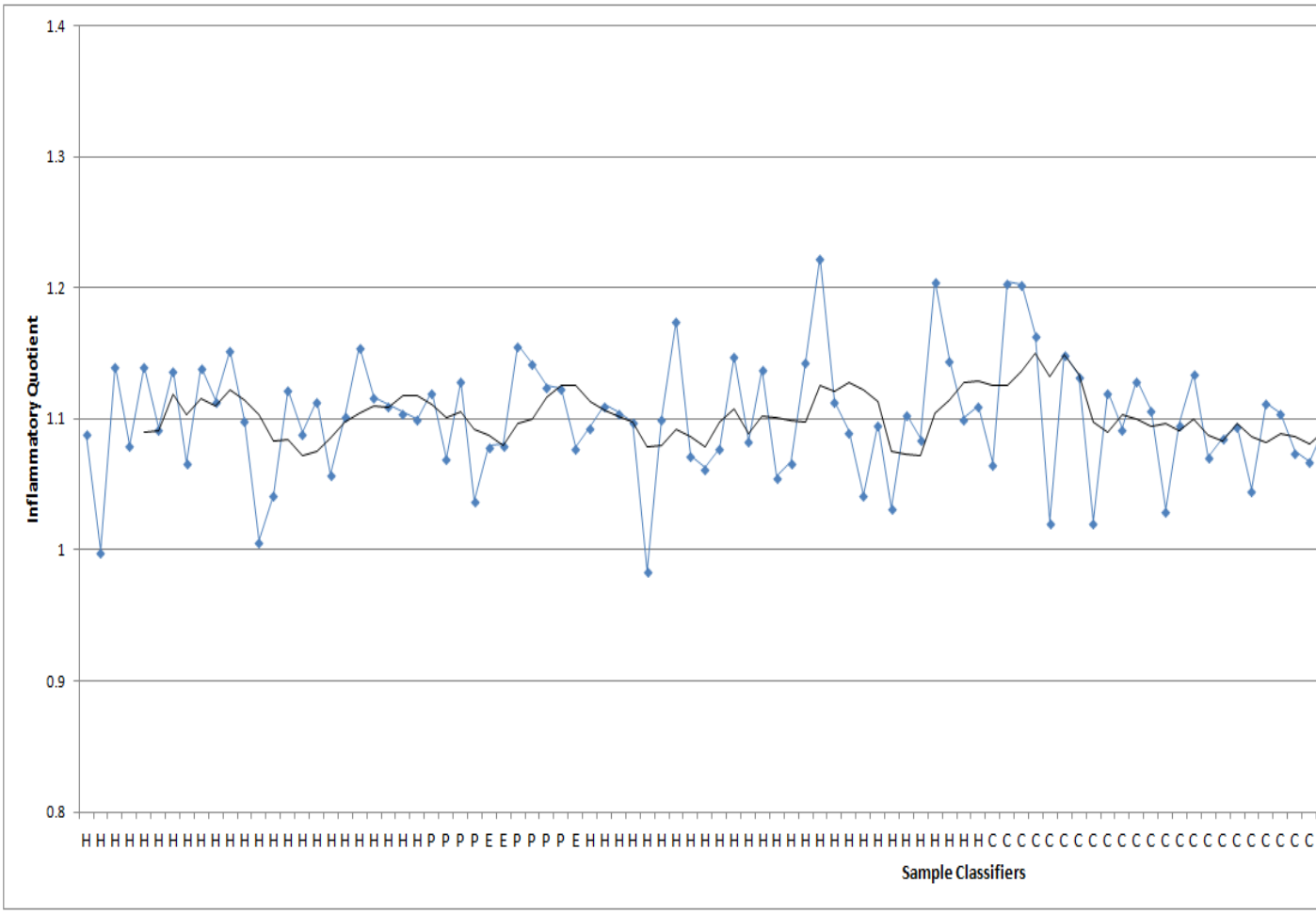

**Figure A3-7: Inflammatory quotient values (blue) for patient SP10, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

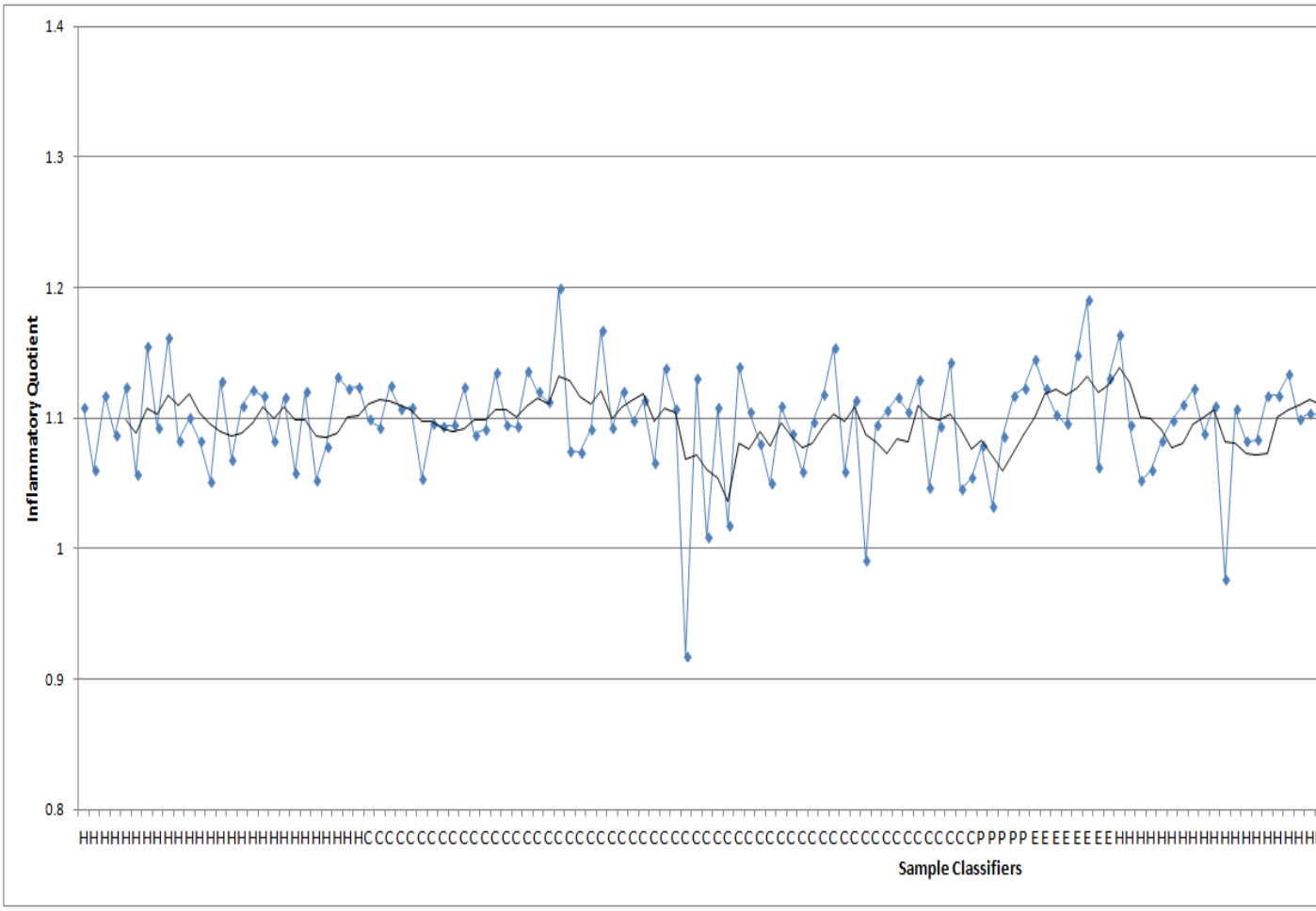


**Figure A3-8: Inflammatory quotient values (blue) for patient SP12, plotted over time, with sample classifications exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

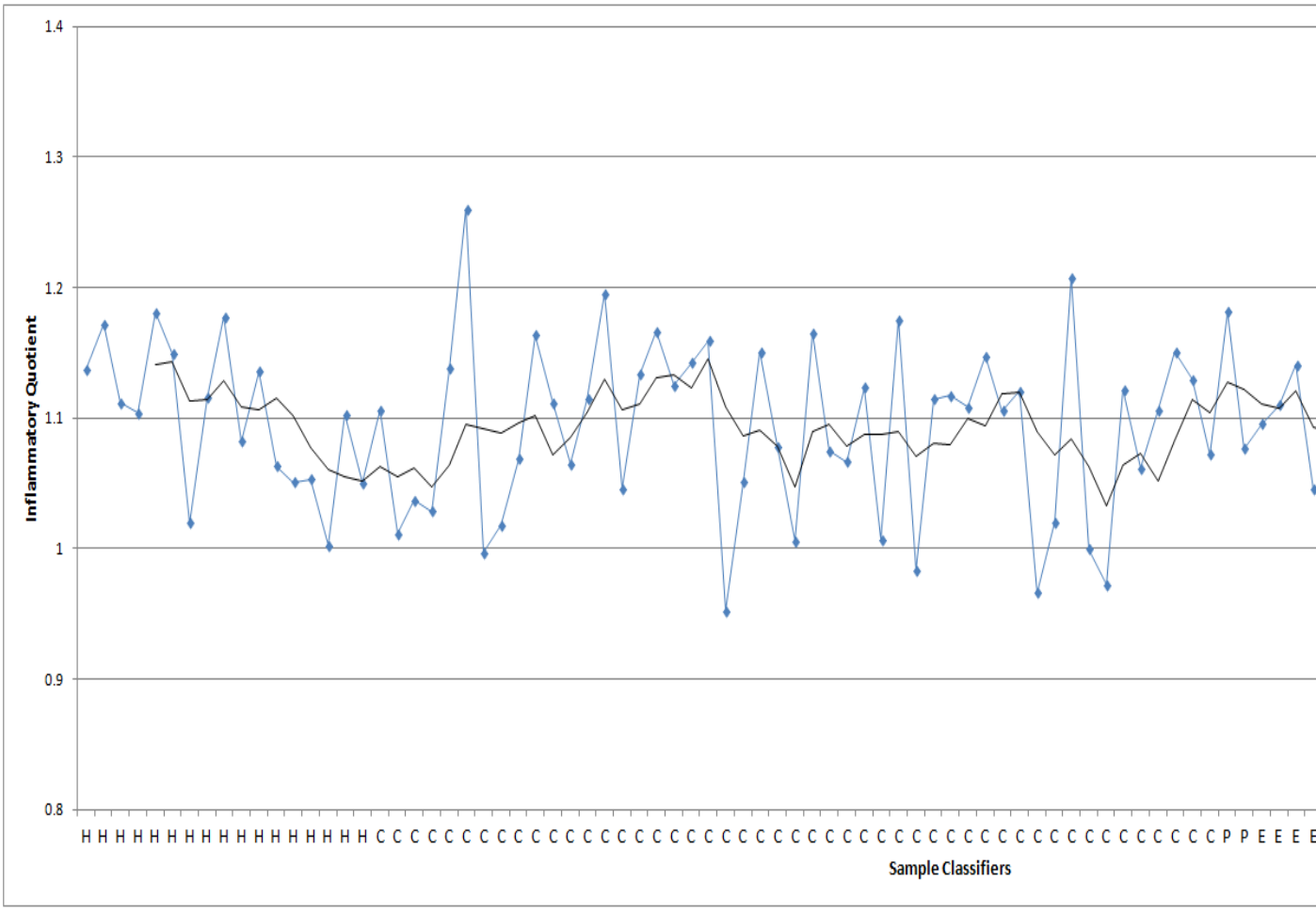

**Figure A3-9: Inflammatory quotient values (blue) for patient SP14, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

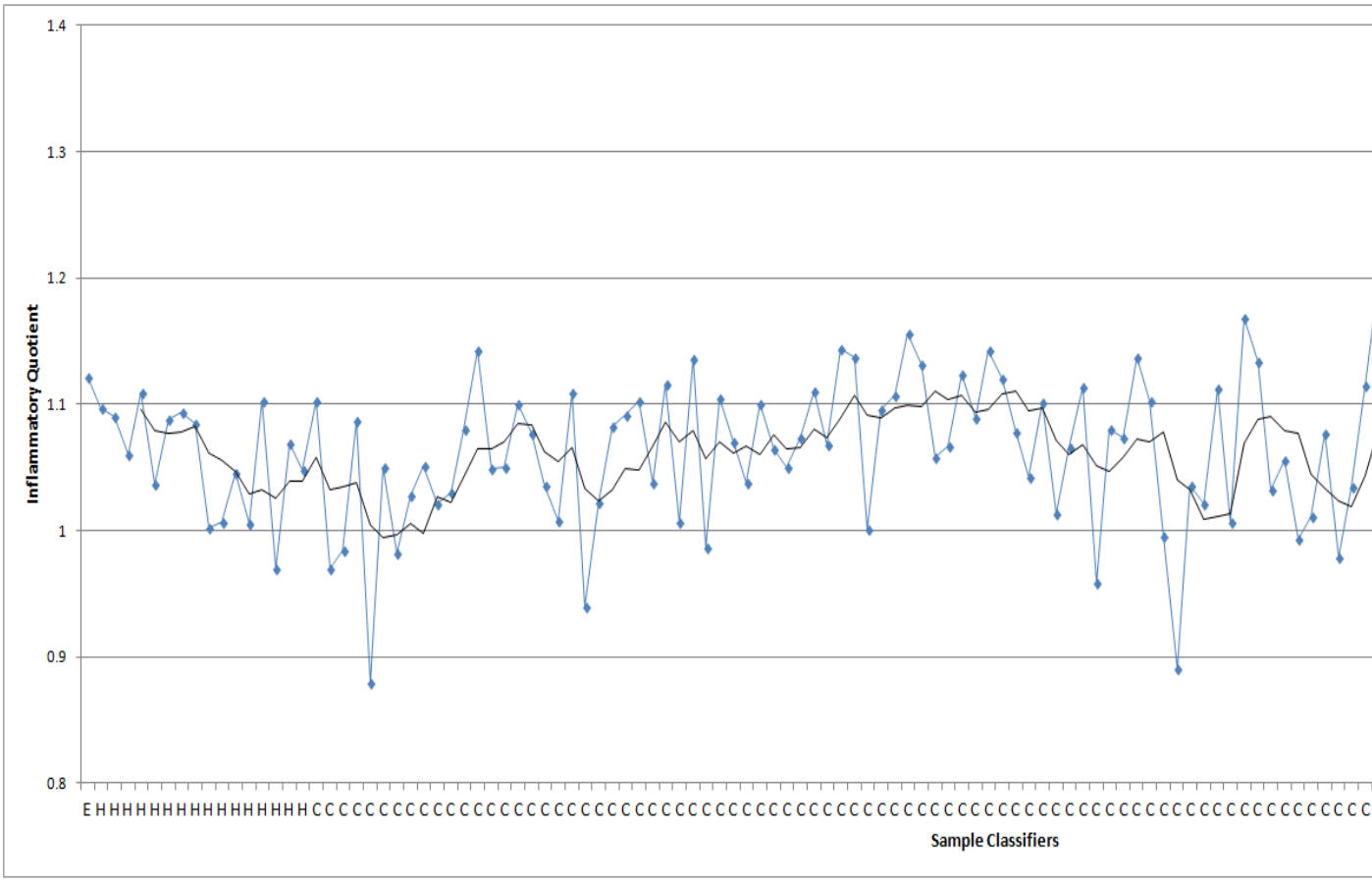

**Figure A3-10: Inflammatory quotient values (blue) for patient SP16, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

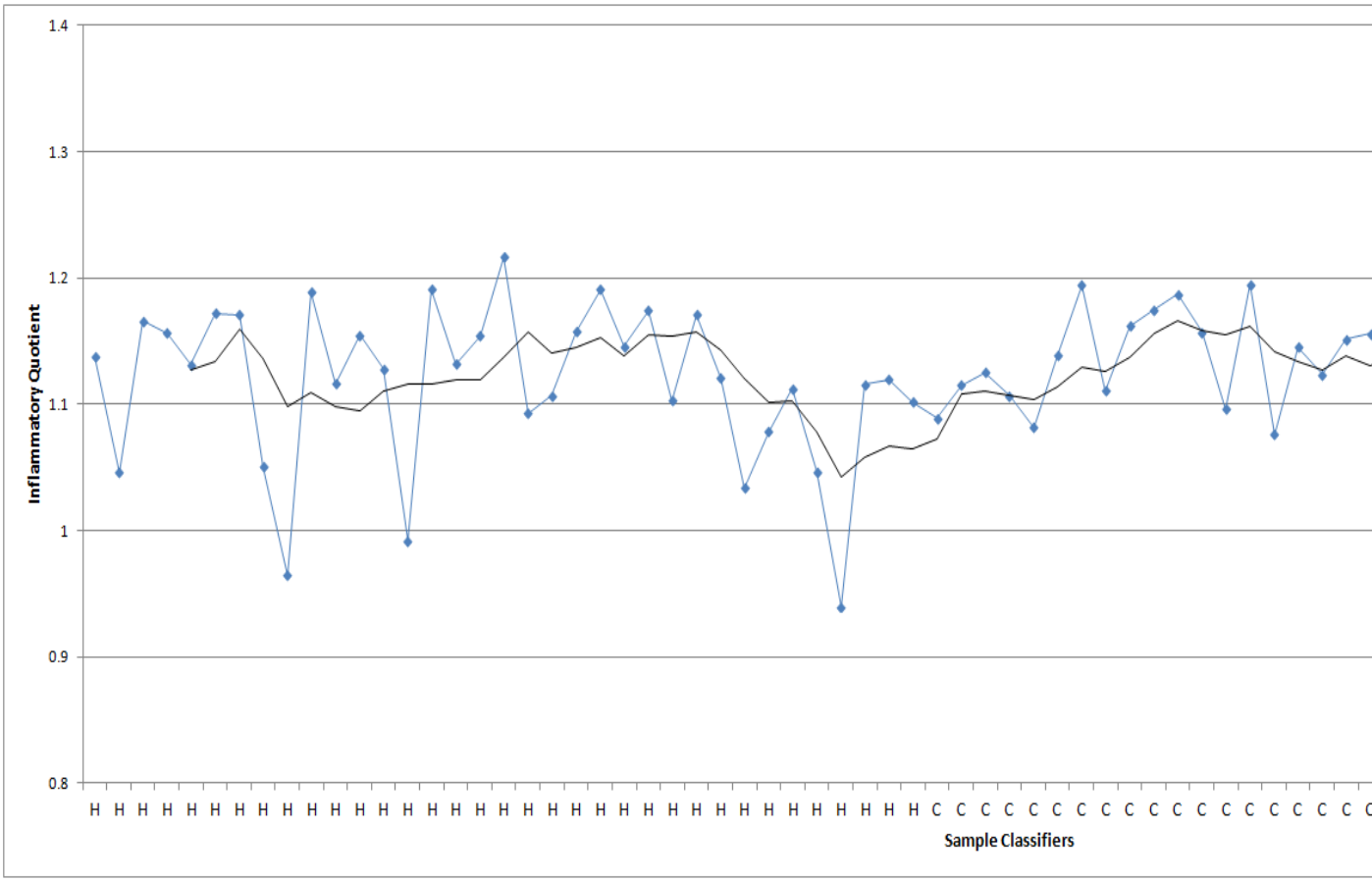

**Figure A3-11: Inflammatory quotient values (blue) for patient SP17, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

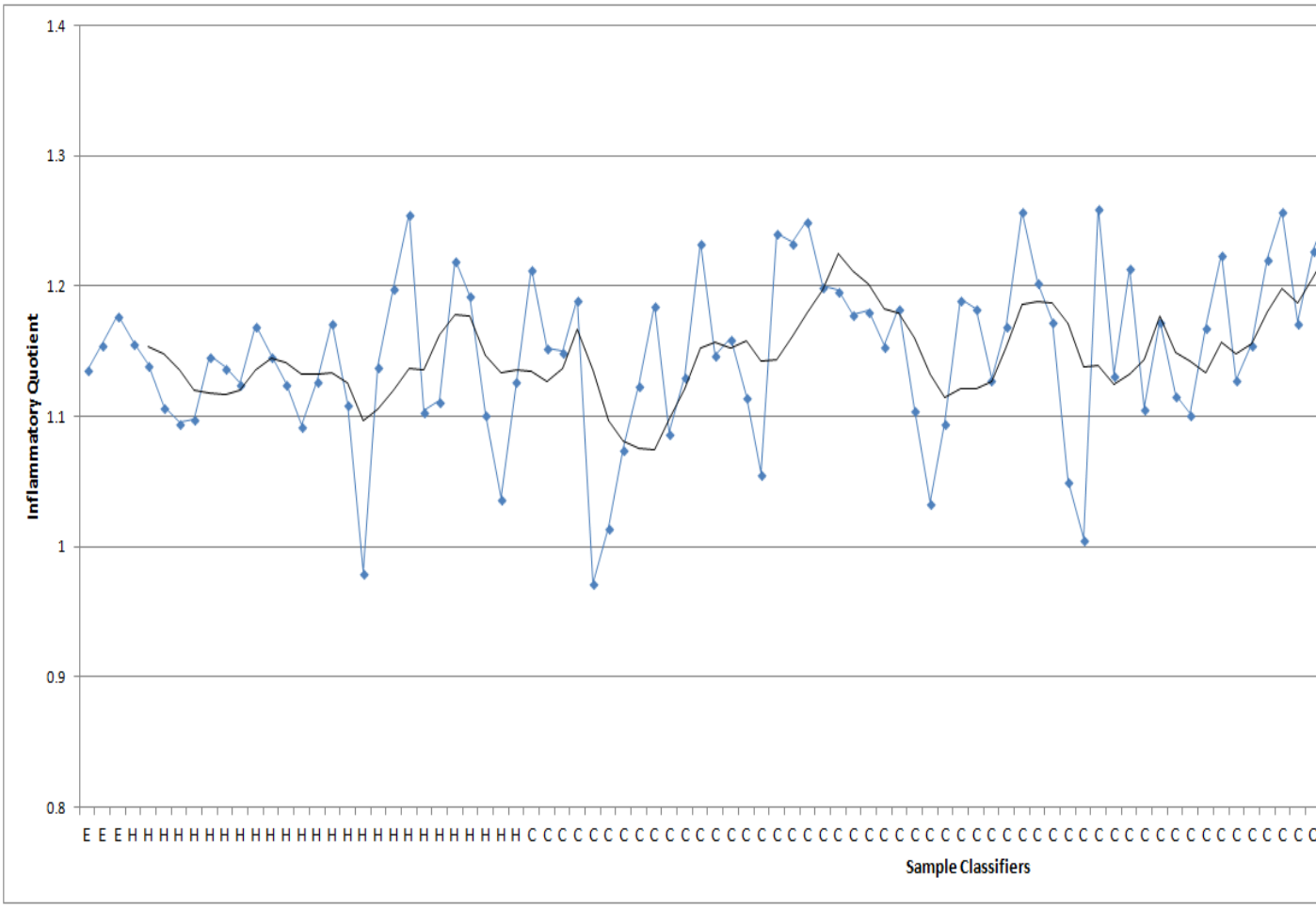

**Figure A3-12: Inflammatory quotient values (blue) for patient SP18, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

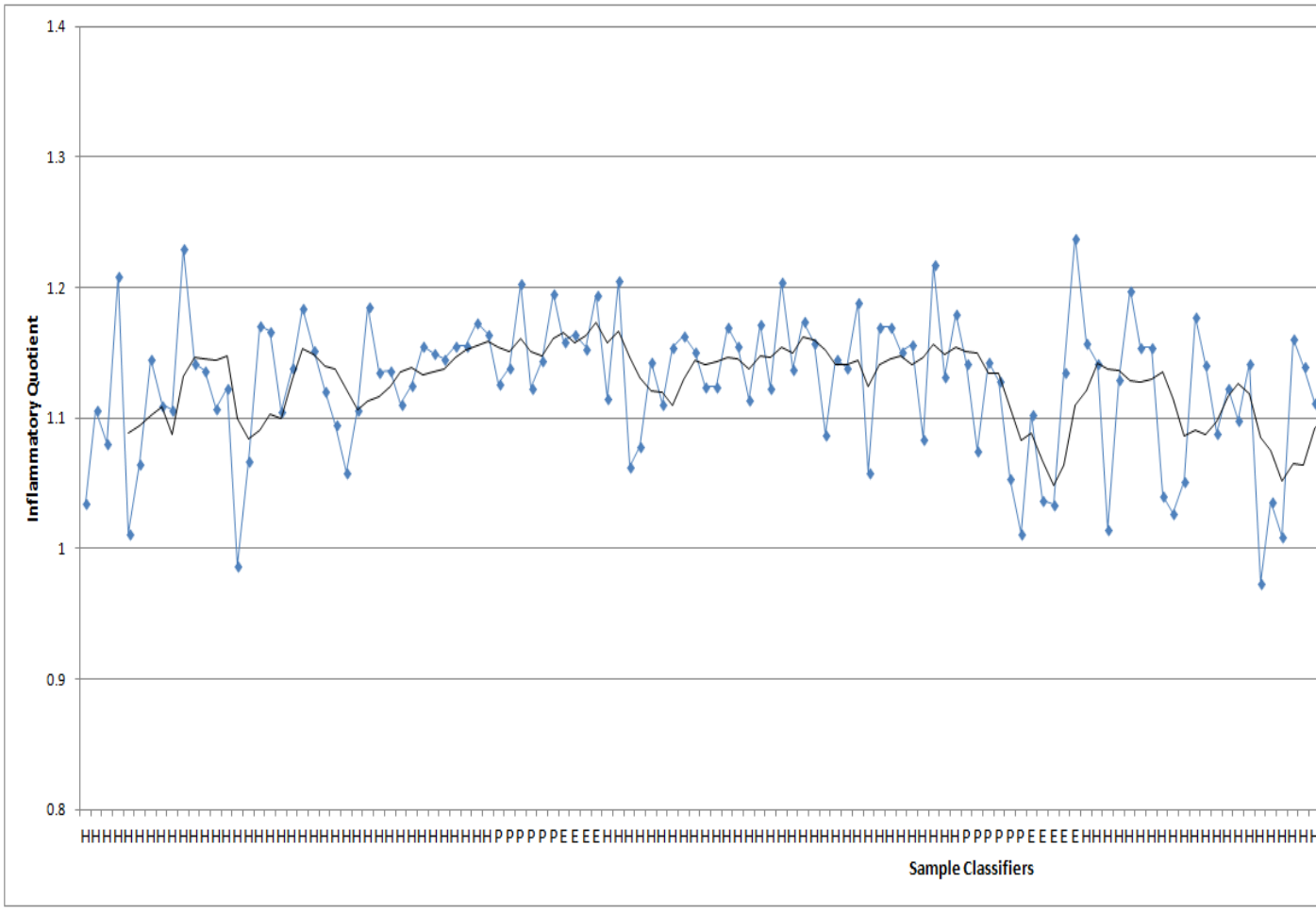


**Figure A3-13: Inflammatory quotient values (blue) for patient SP20, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

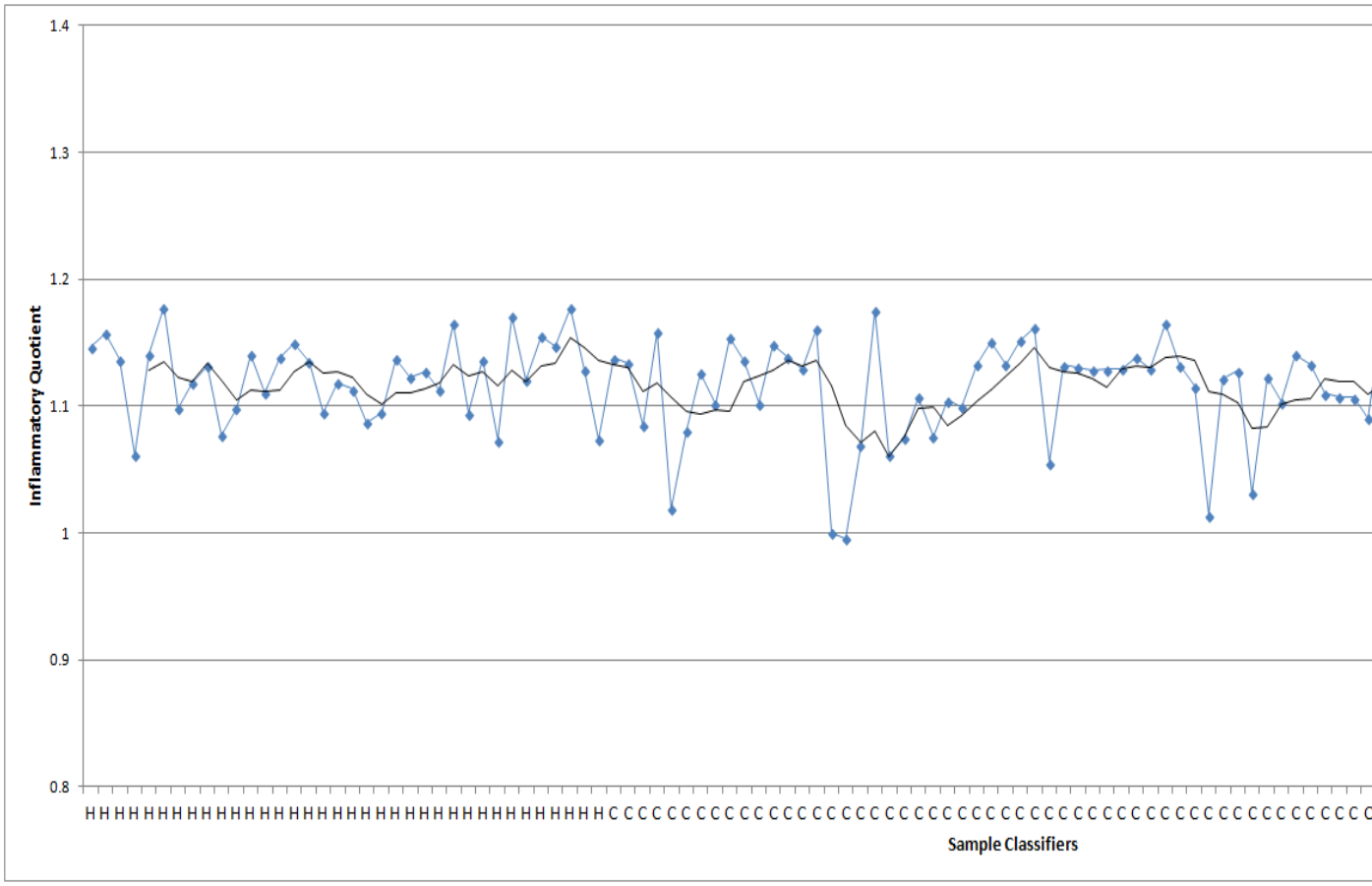


**Figure A3-14: Inflammatory quotient values (blue) for patient SP21, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

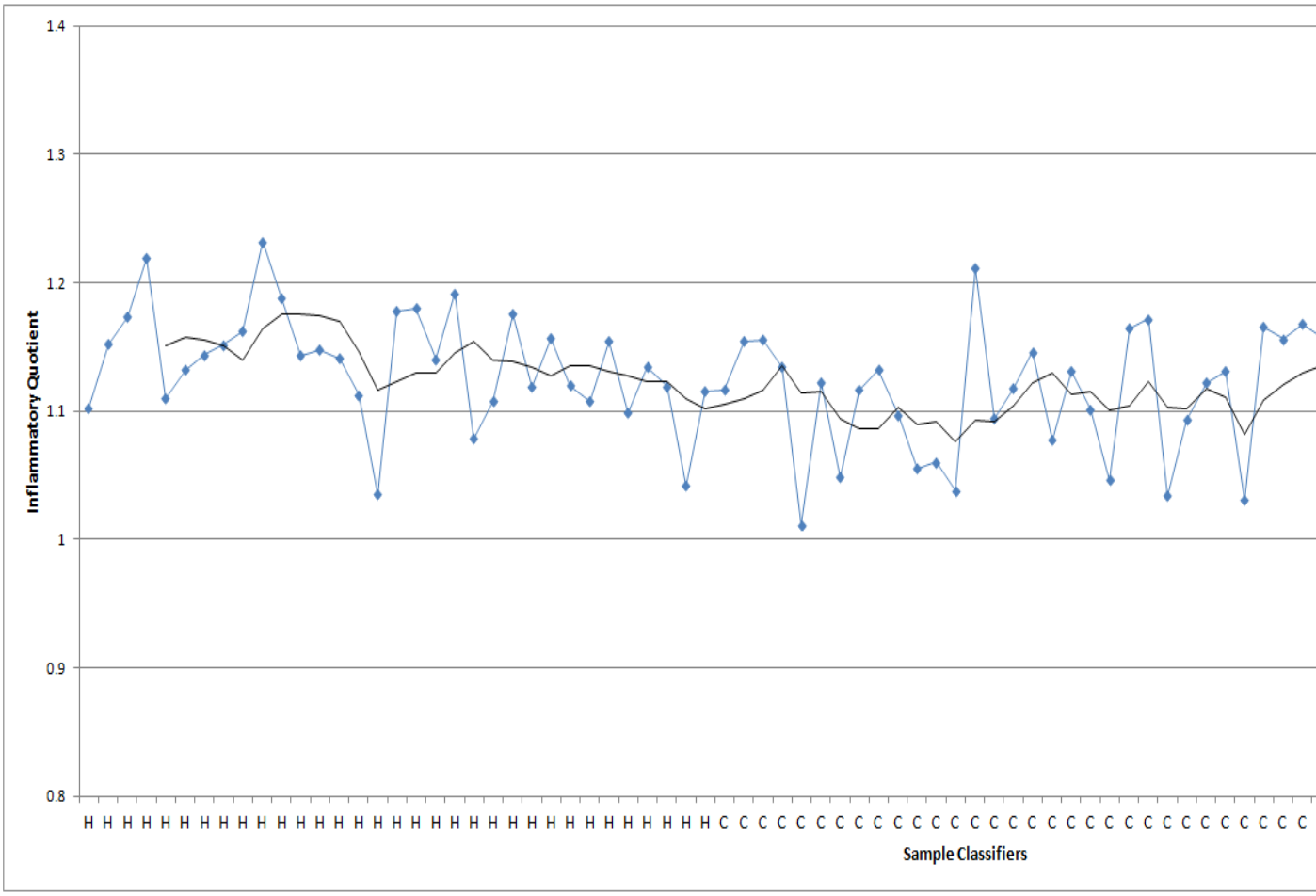

**Figure A3-15: Inflammatory quotient values (blue) for patient SP23, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

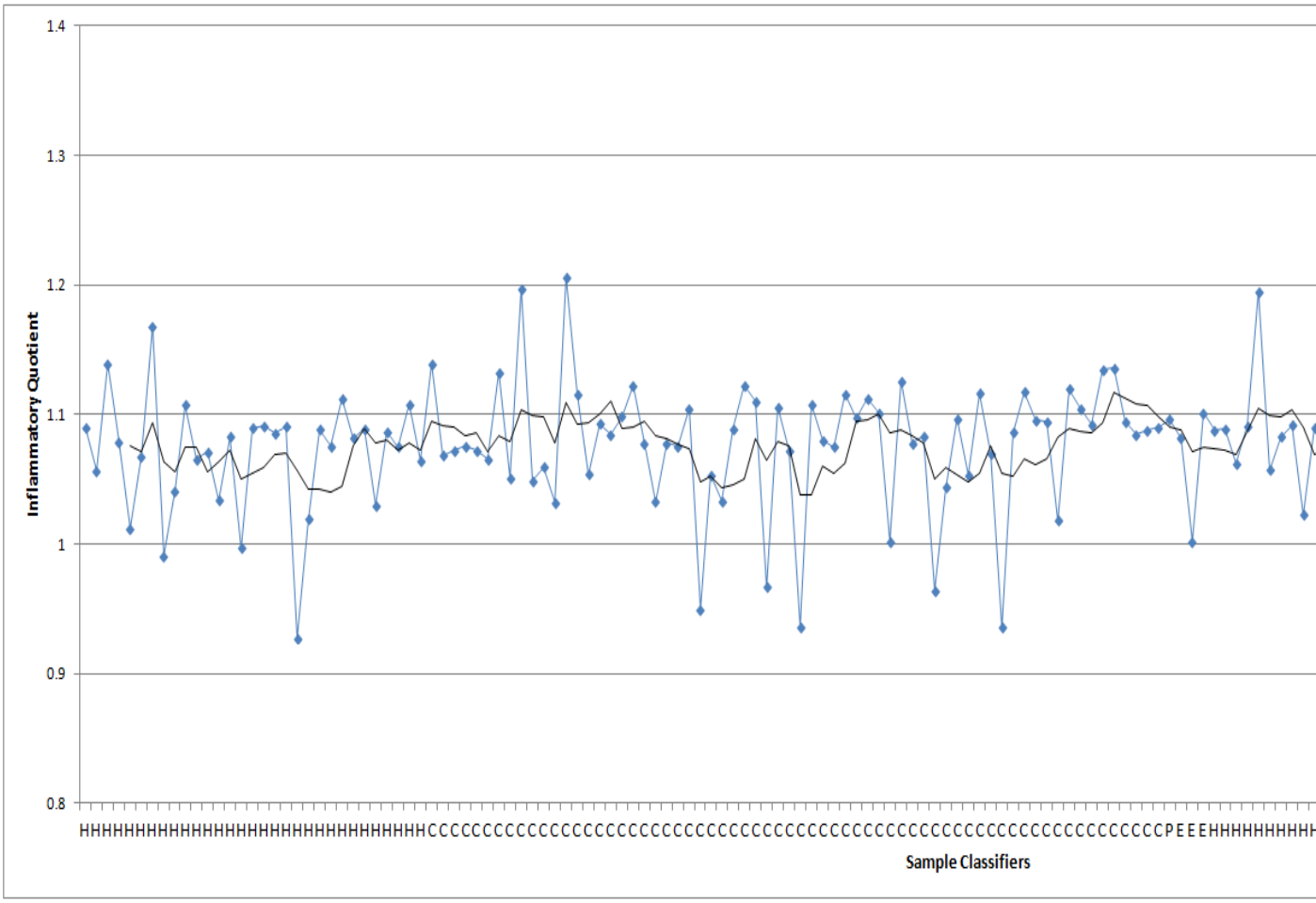

**Figure A3-16: Inflammatory quotient values (blue) for patient SP24, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

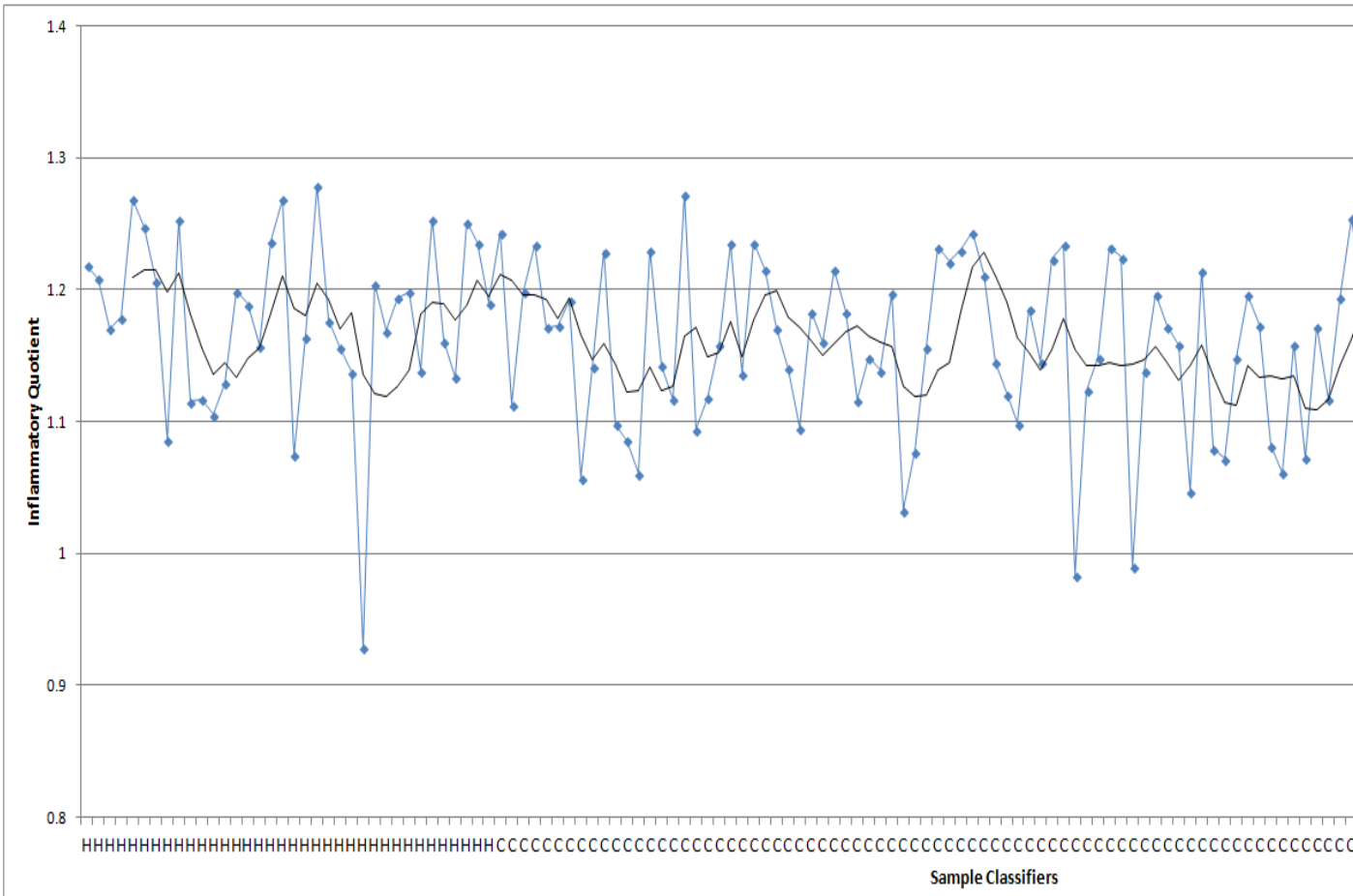


**Figure A3-17: Inflammatory quotient values (blue) for patient SP25, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

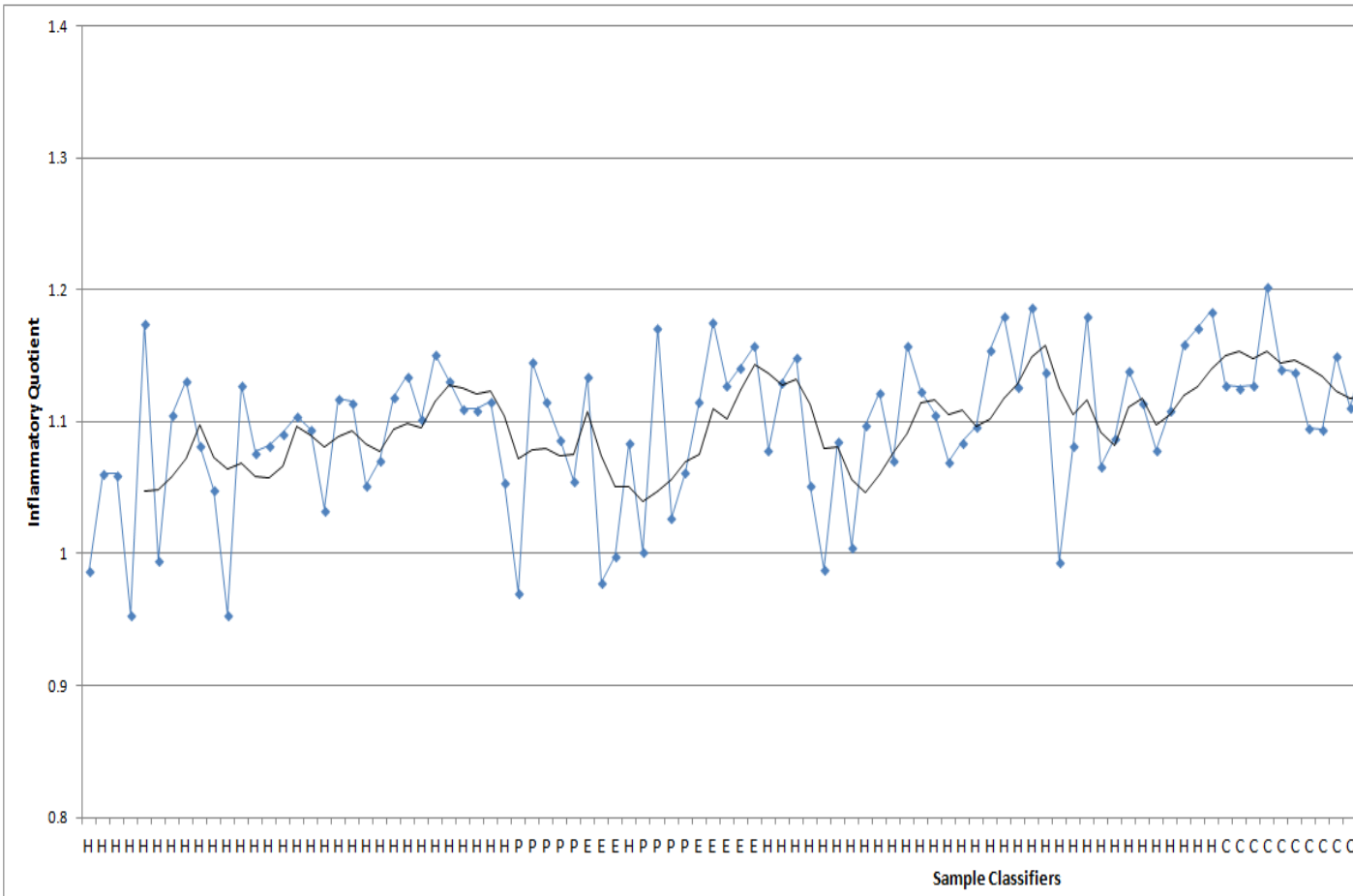


**Figure A3-18: Inflammatory quotient values (blue) for patient SP27, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

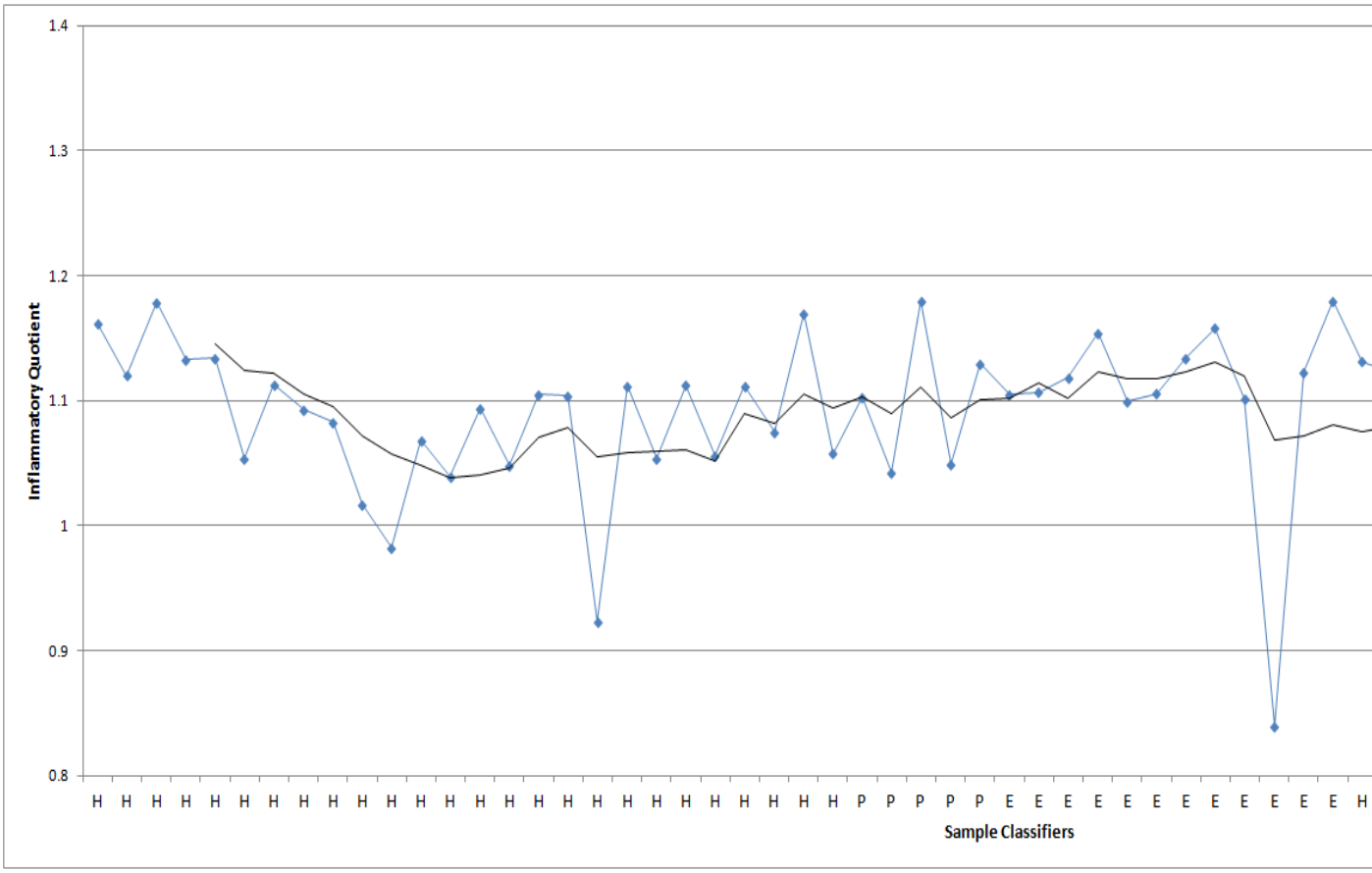


**Figure A3-19: Inflammatory quotient values (blue) for patient SP28, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

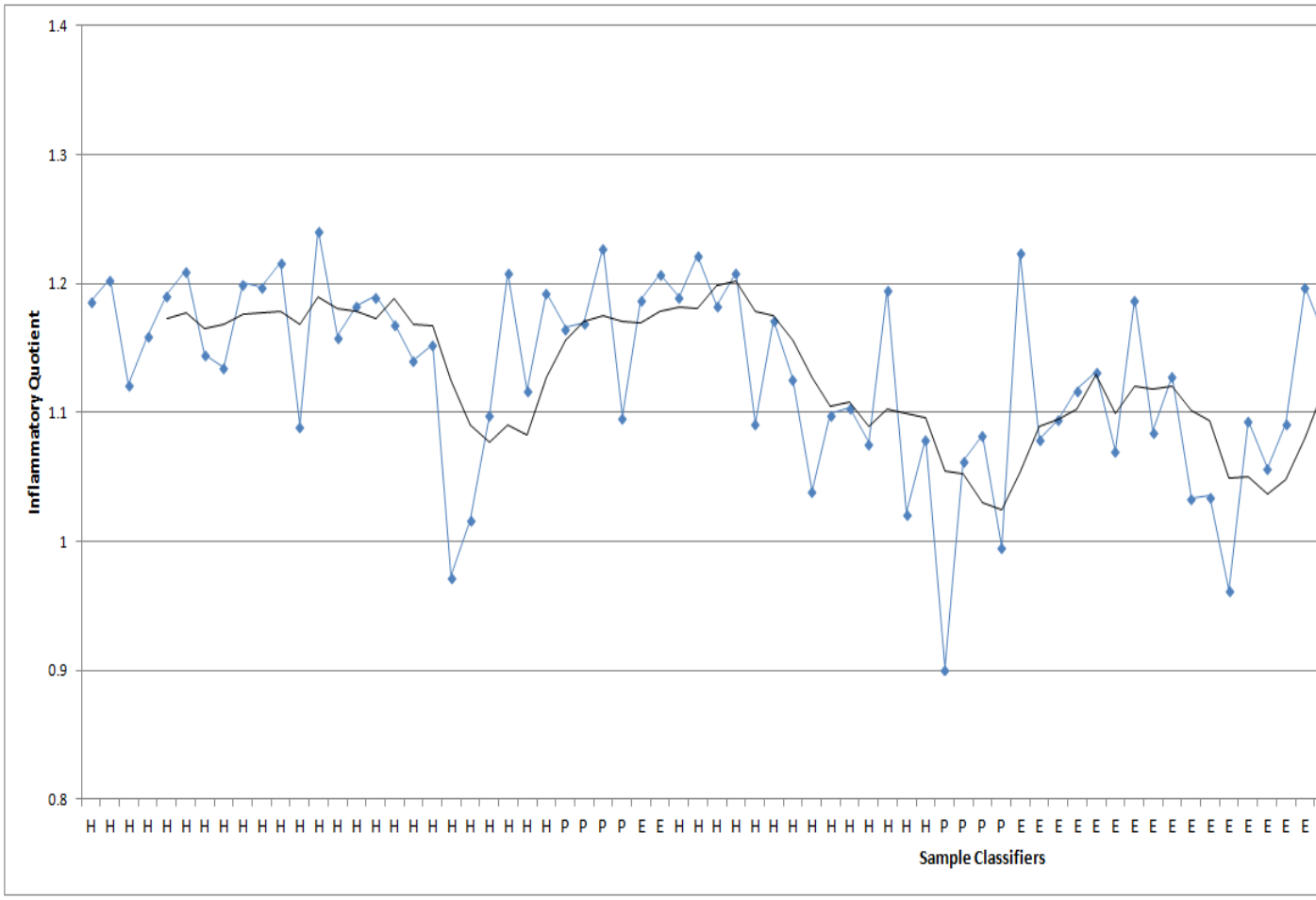


**Figure A3-20: Inflammatory quotient values (blue) for patient SP30, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

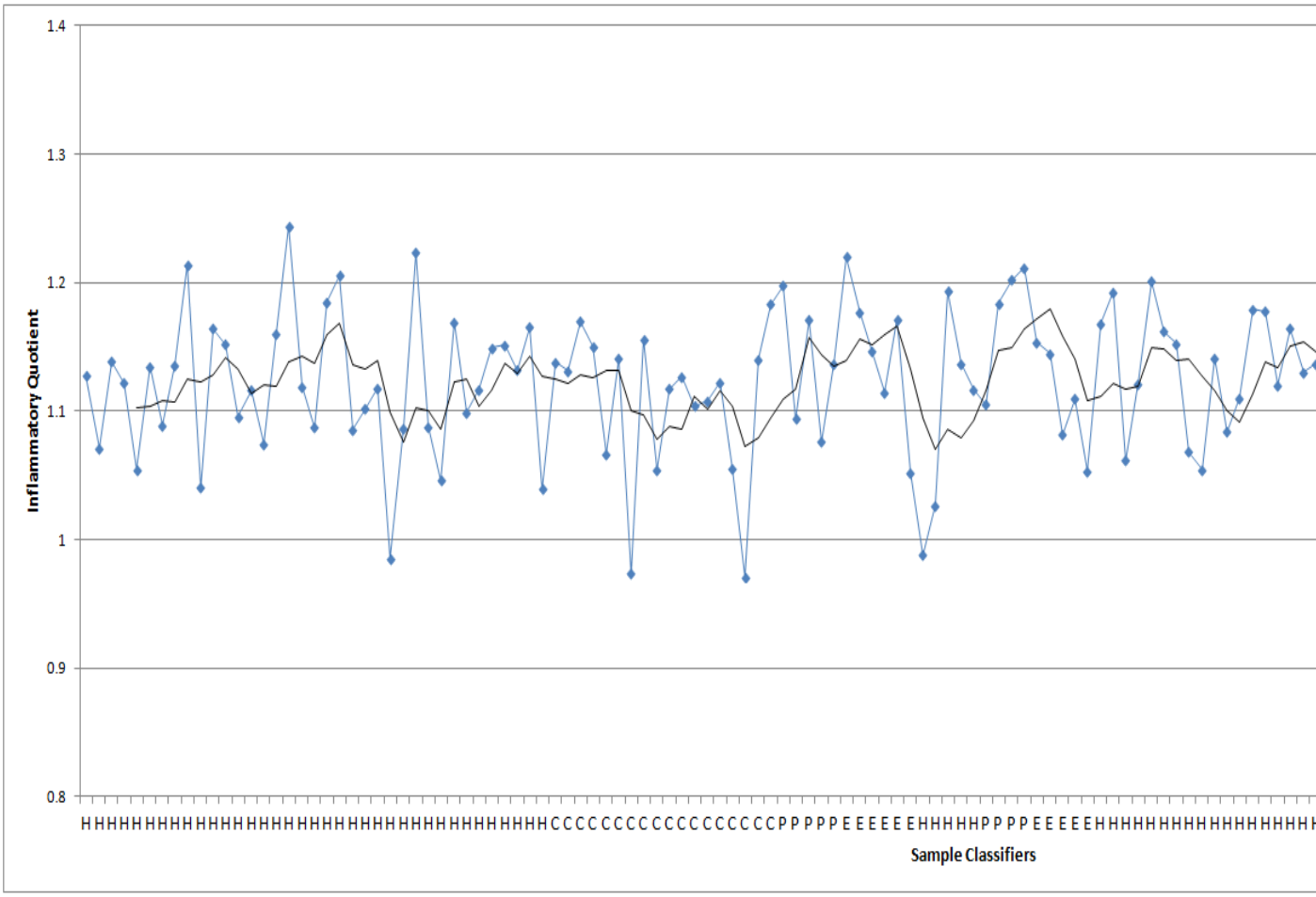


**Figure A3-21: Inflammatory quotient values (blue) for patient SP31, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

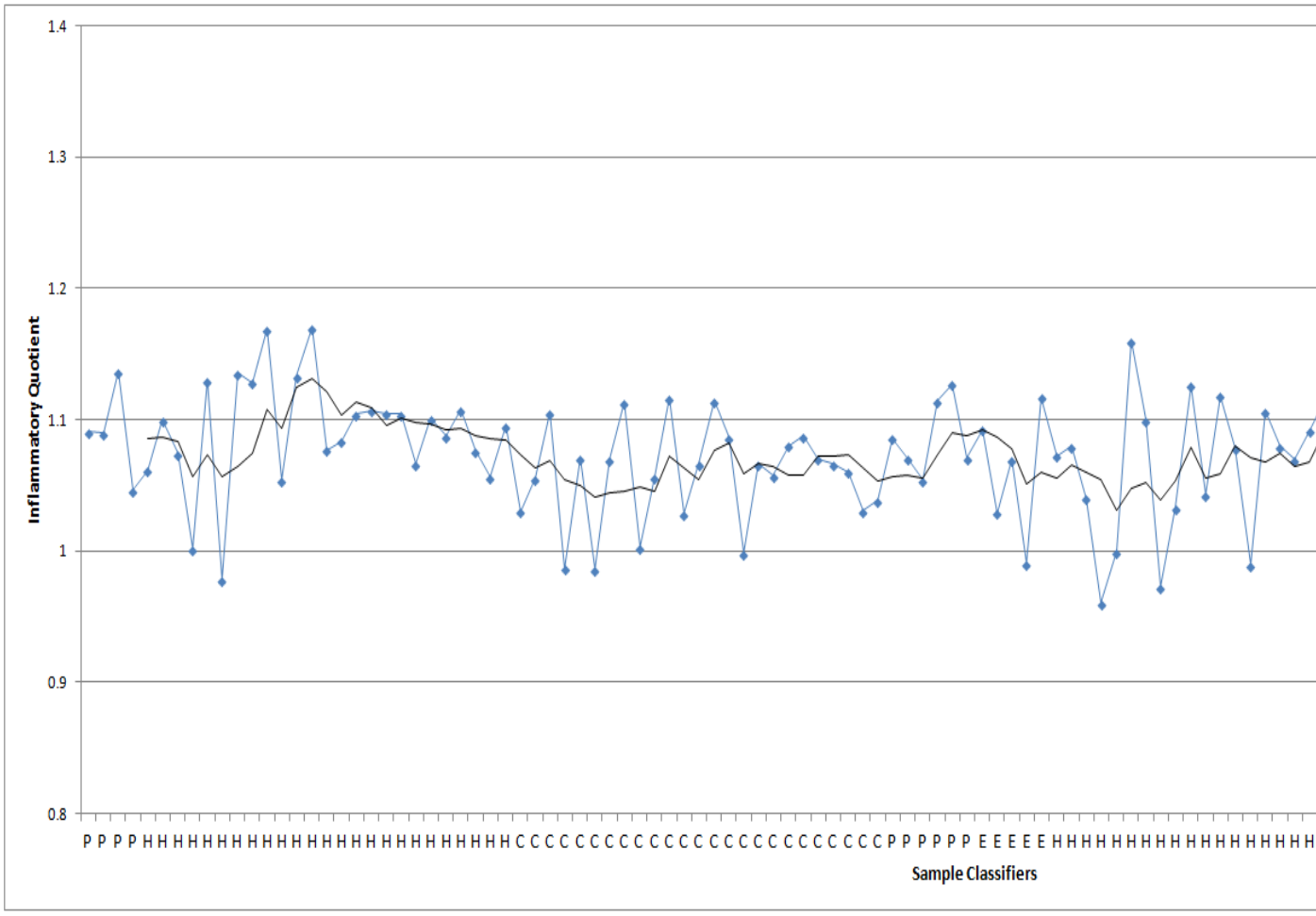

**Figure A3-22: Inflammatory quotient values (blue) for patient SP34, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

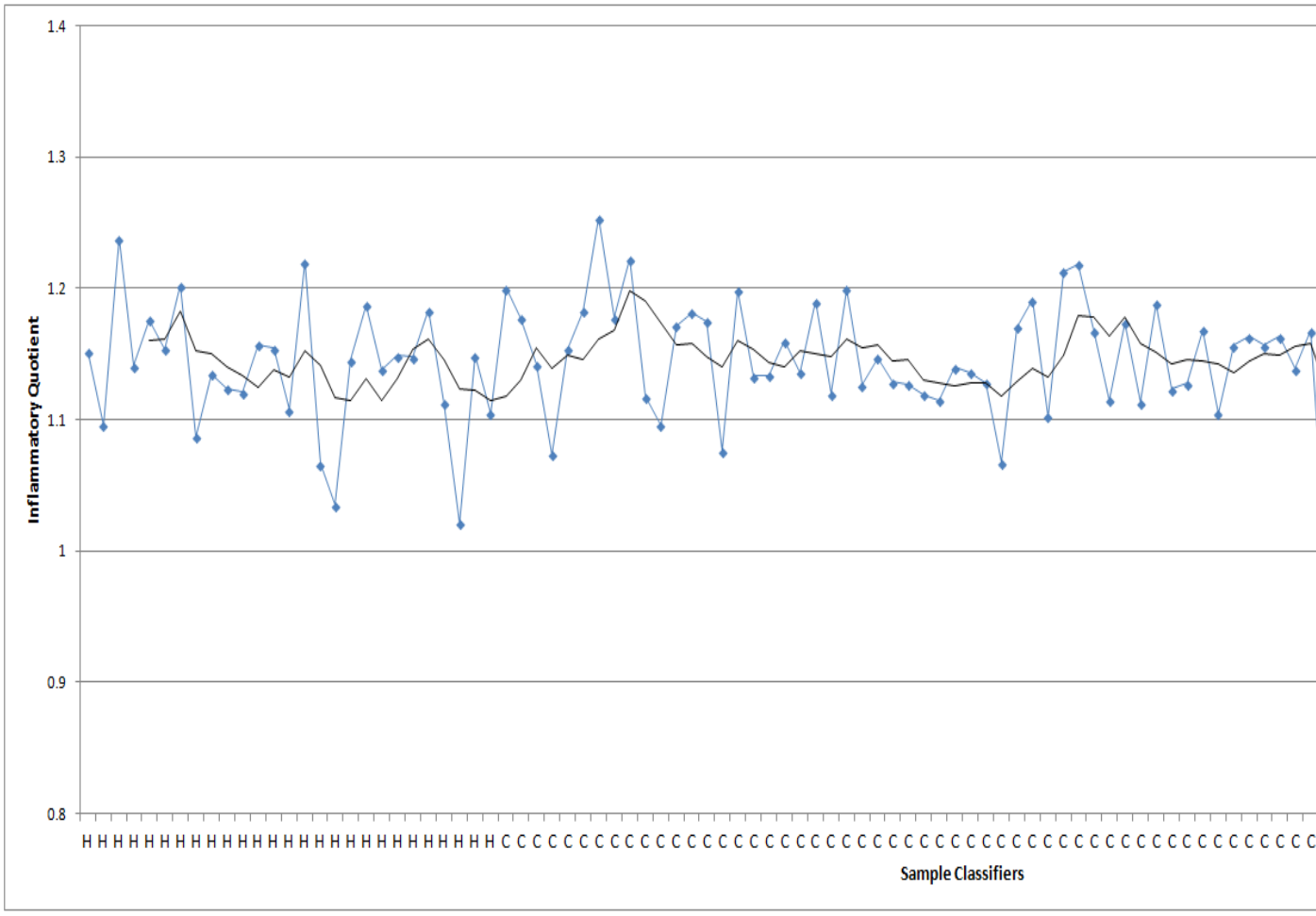

**Figure A3-23: Inflammatory quotient values (blue) for patient SP35, plotted over time, with sample classification exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

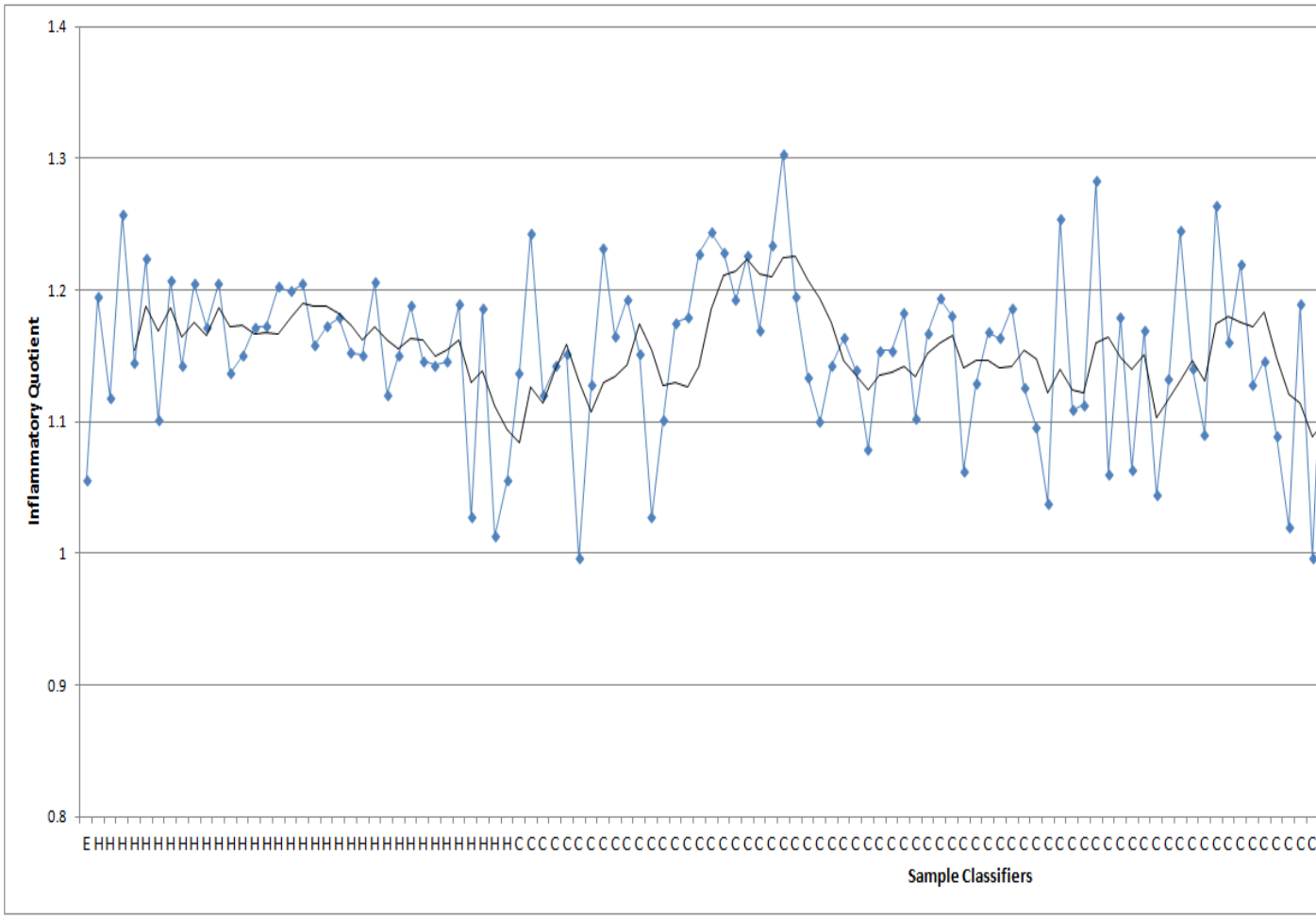

**Figure A3-24: Inflammatory quotient values (blue) for patient SP36, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

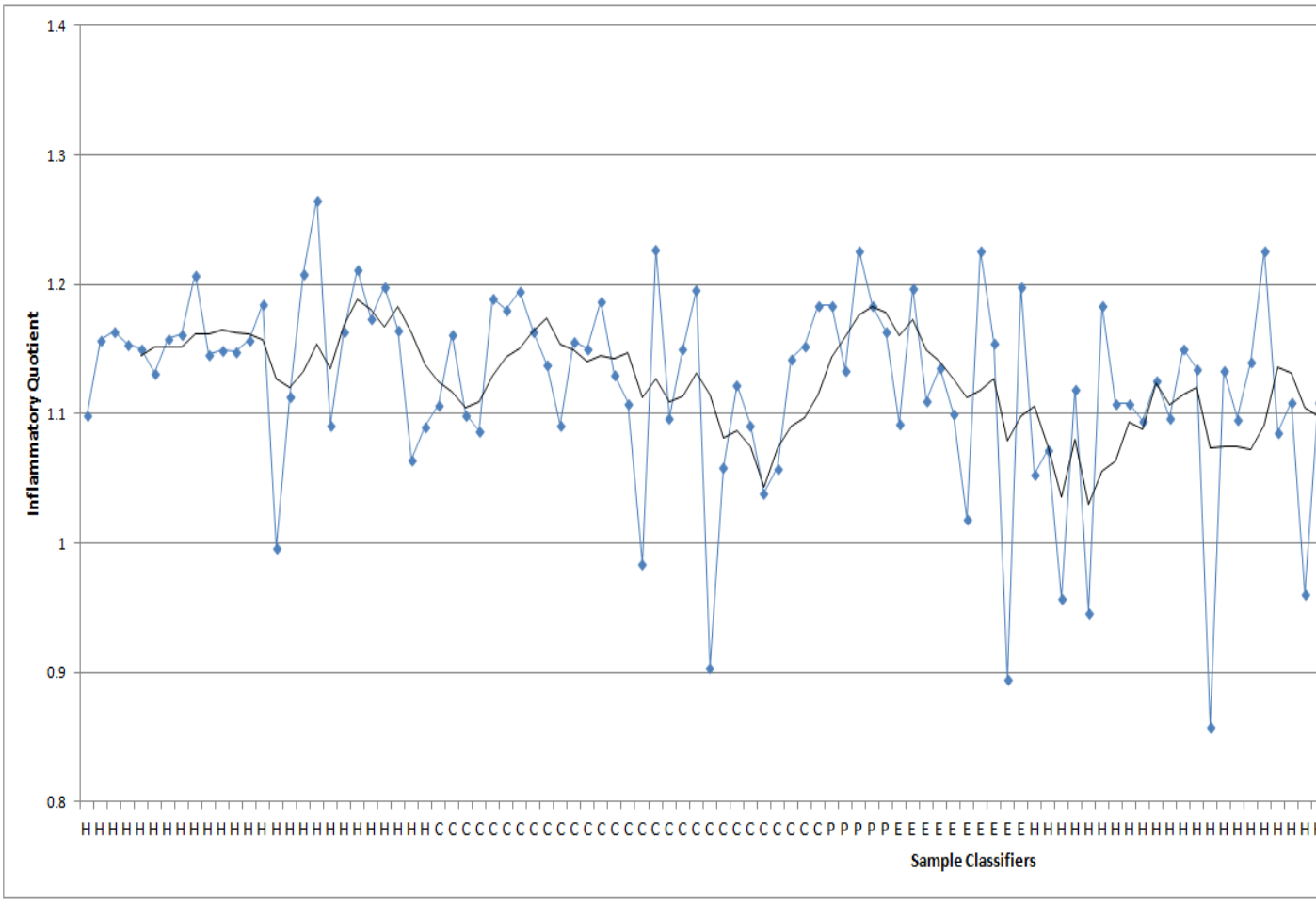


**Figure A3-25: Inflammatory quotient values (blue) for patient SP38, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

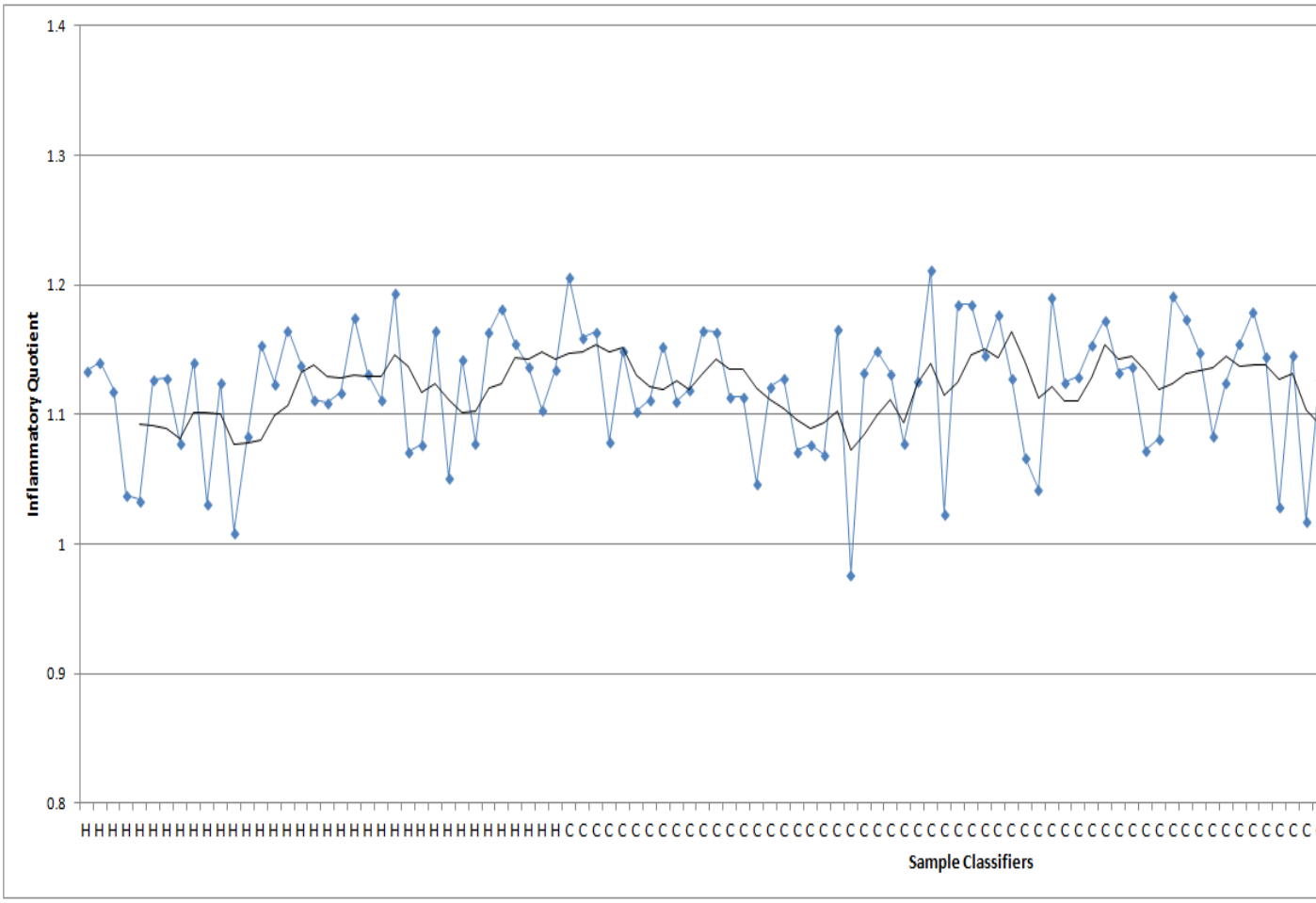

**Figure A3-26: Inflammatory quotient values (blue) for patient SP39, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

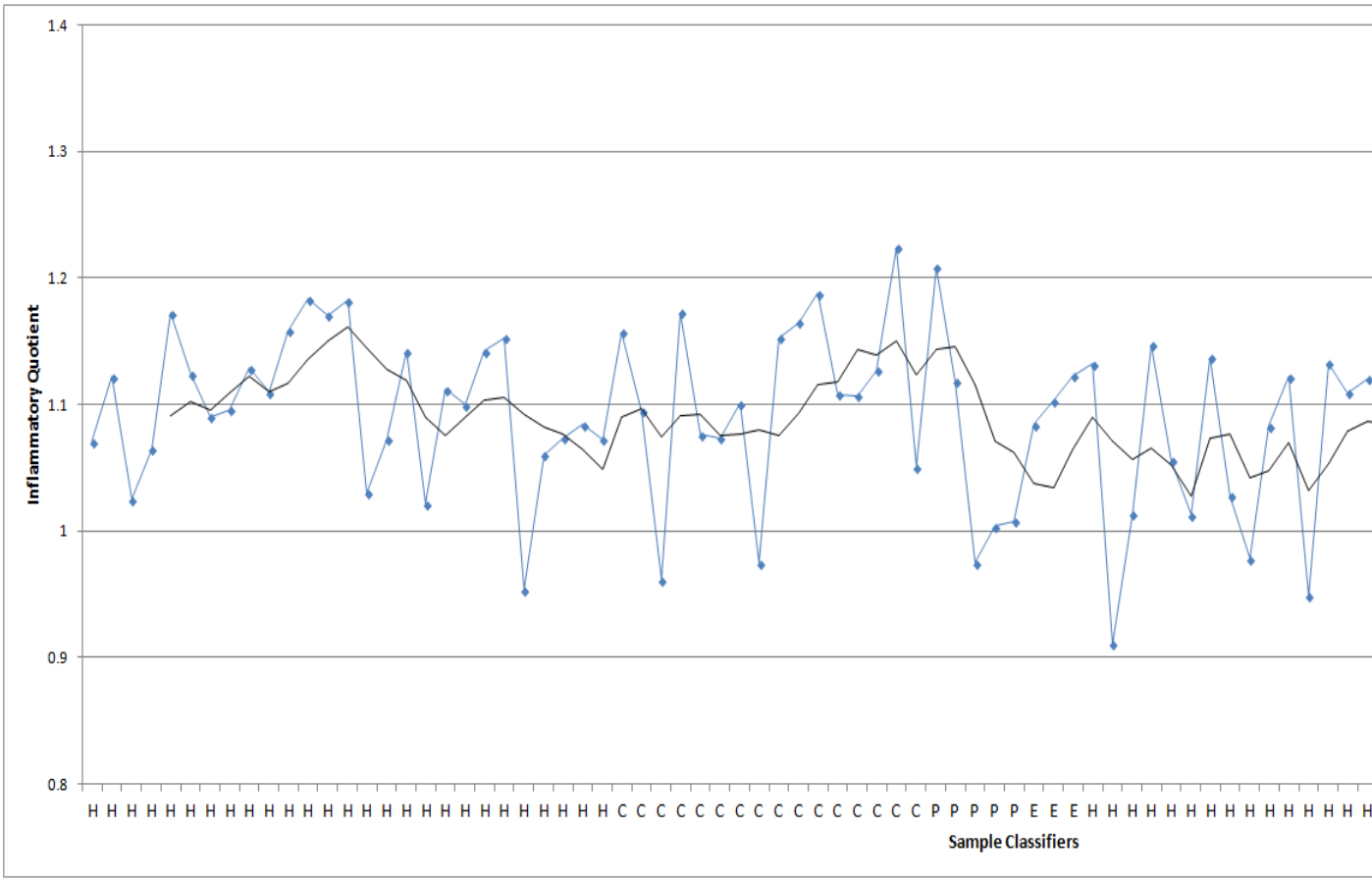

**Figure A3-27: Inflammatory quotient values (blue) for patient SP40, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

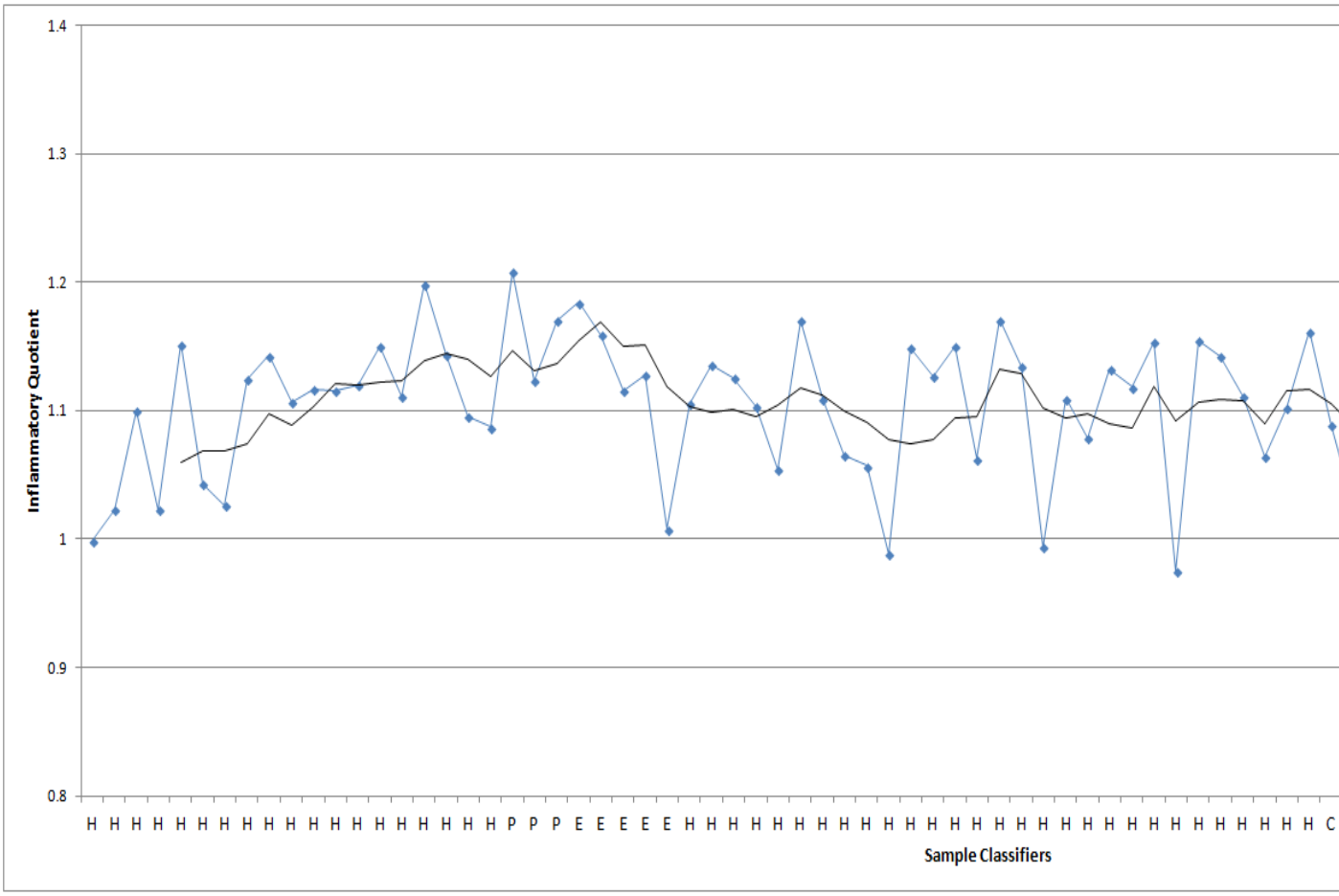

**Figure A3-28: Inflammatory quotient values (blue) for patient SP43, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

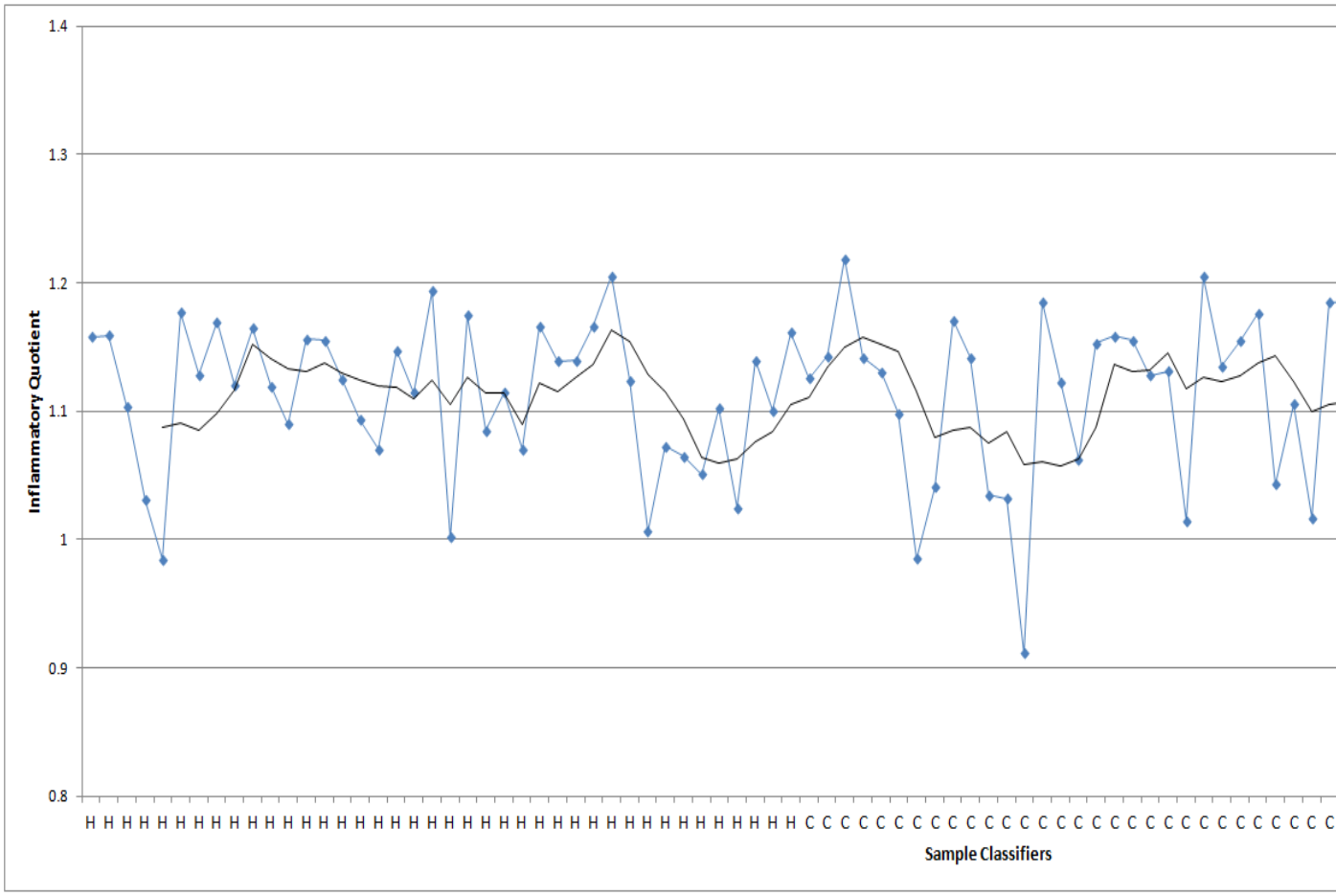


**Figure A3-29: Inflammatory quotient values (blue) for patient SP44, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving average identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

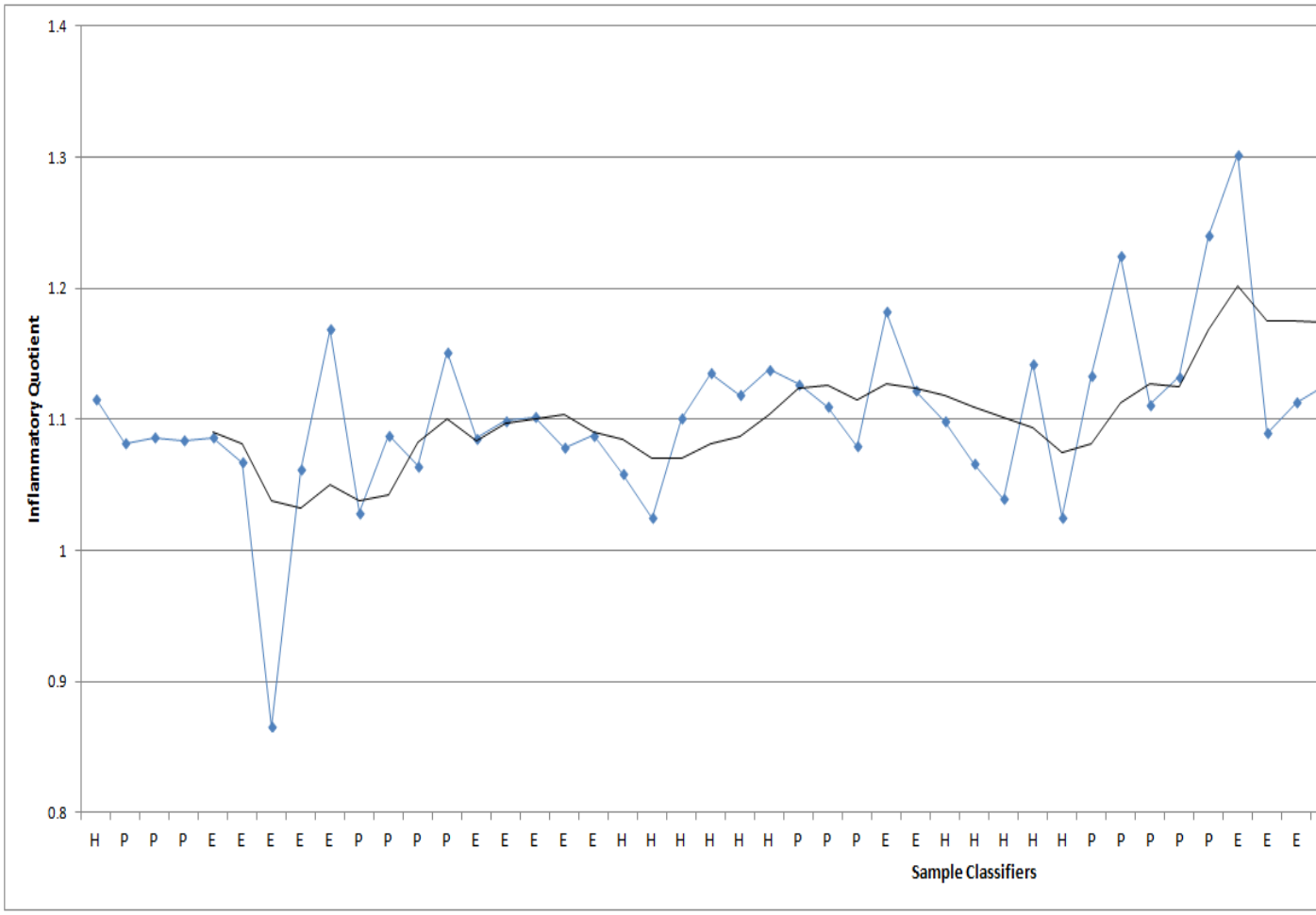


**Figure A3-30: Inflammatory quotient values (blue) for patient SP47, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

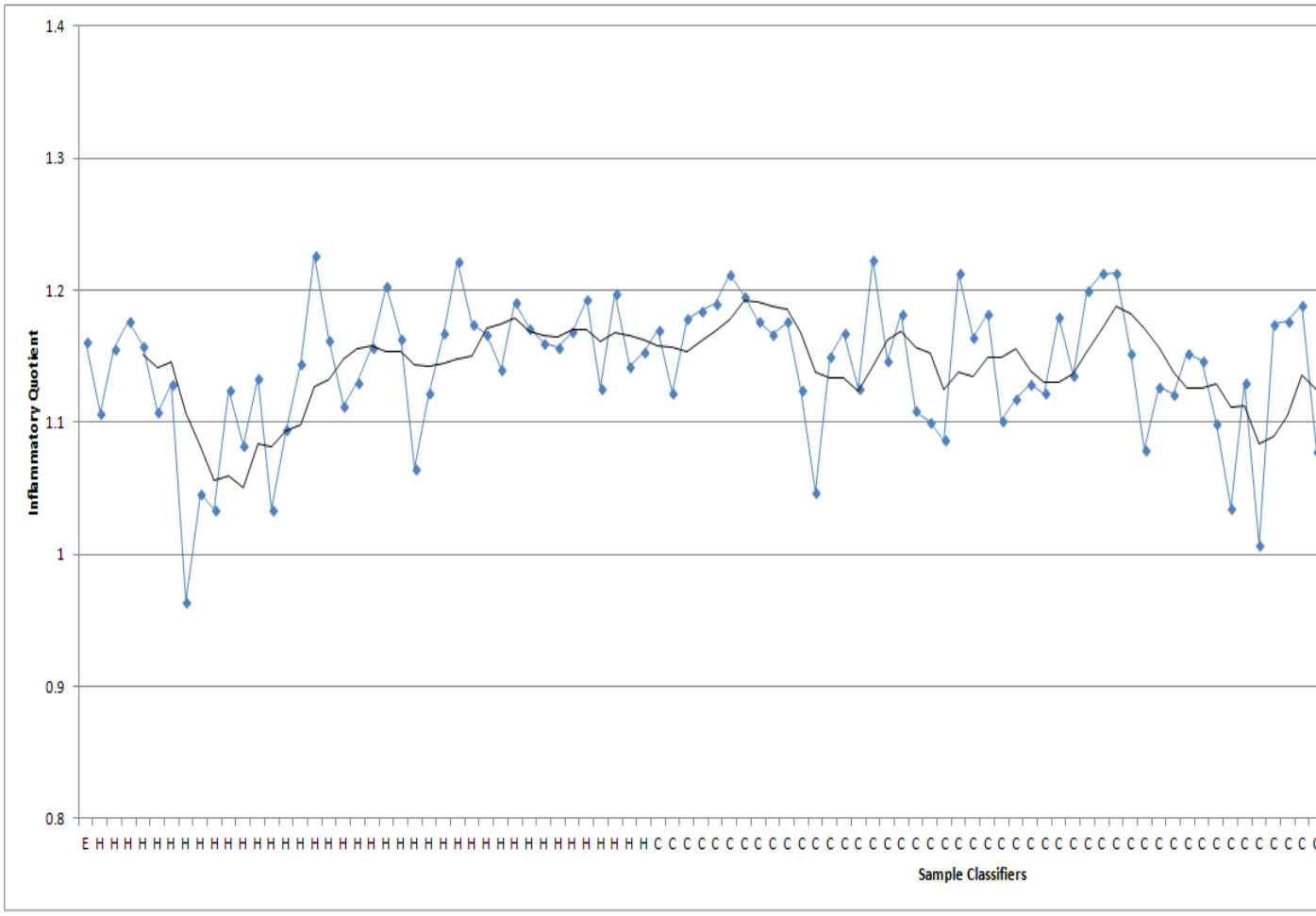

**Figure A3-31: Inflammatory quotient values (blue) for patient SP48, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

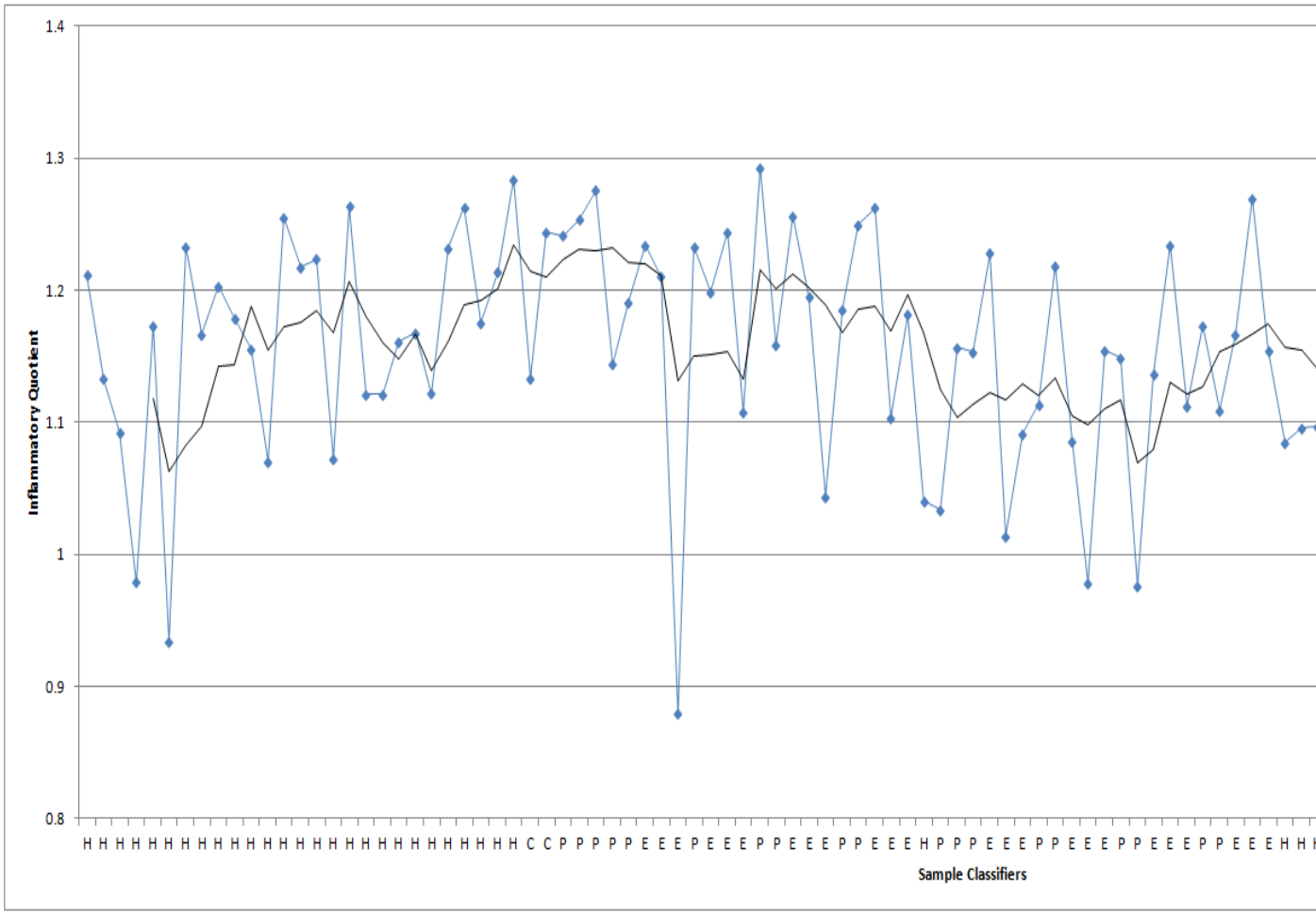


**Figure A3-32: Inflammatory quotient values (blue) for patient SP49, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

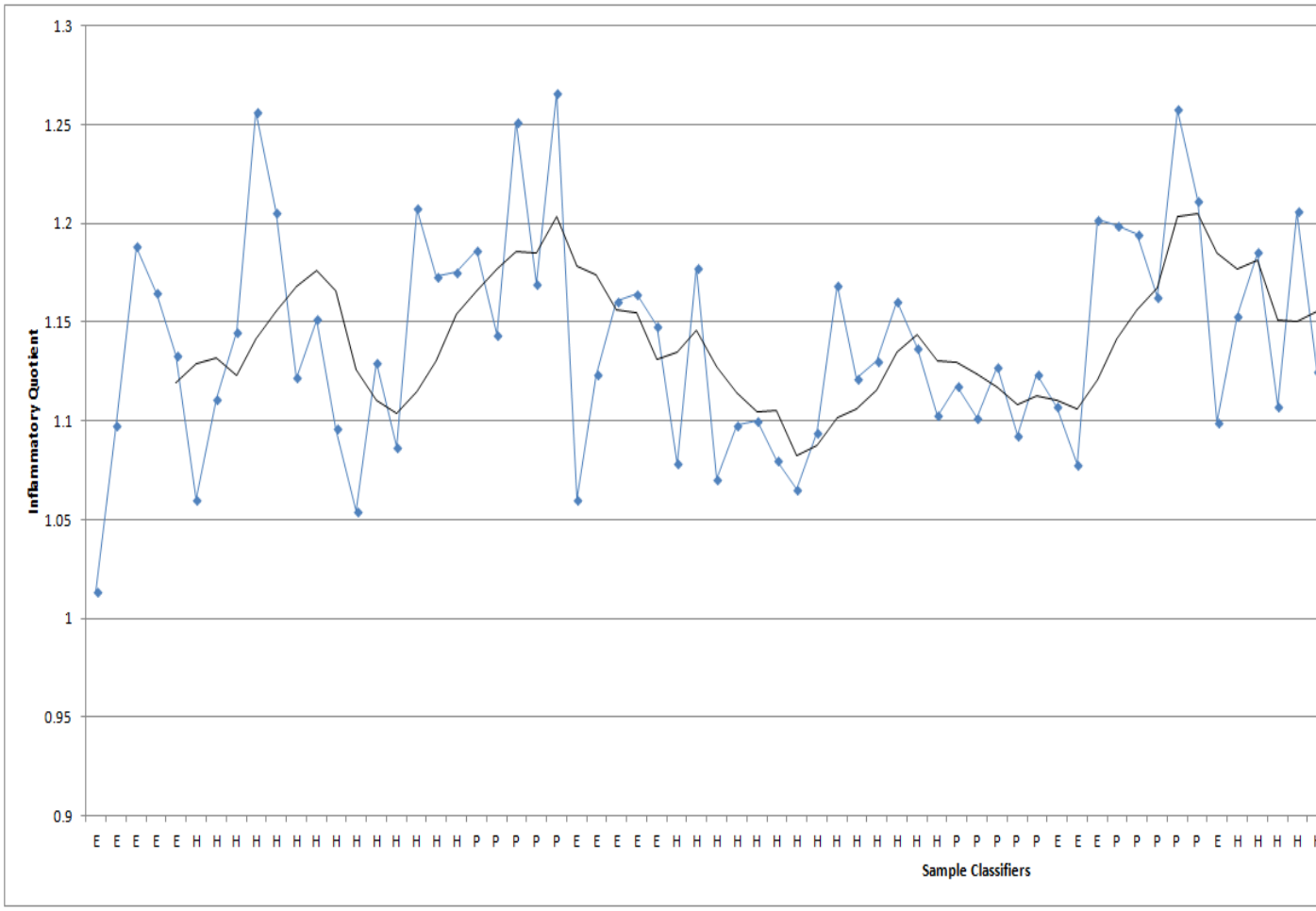


**Figure A3-33: Inflammatory quotient values (blue) for patient SP50, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

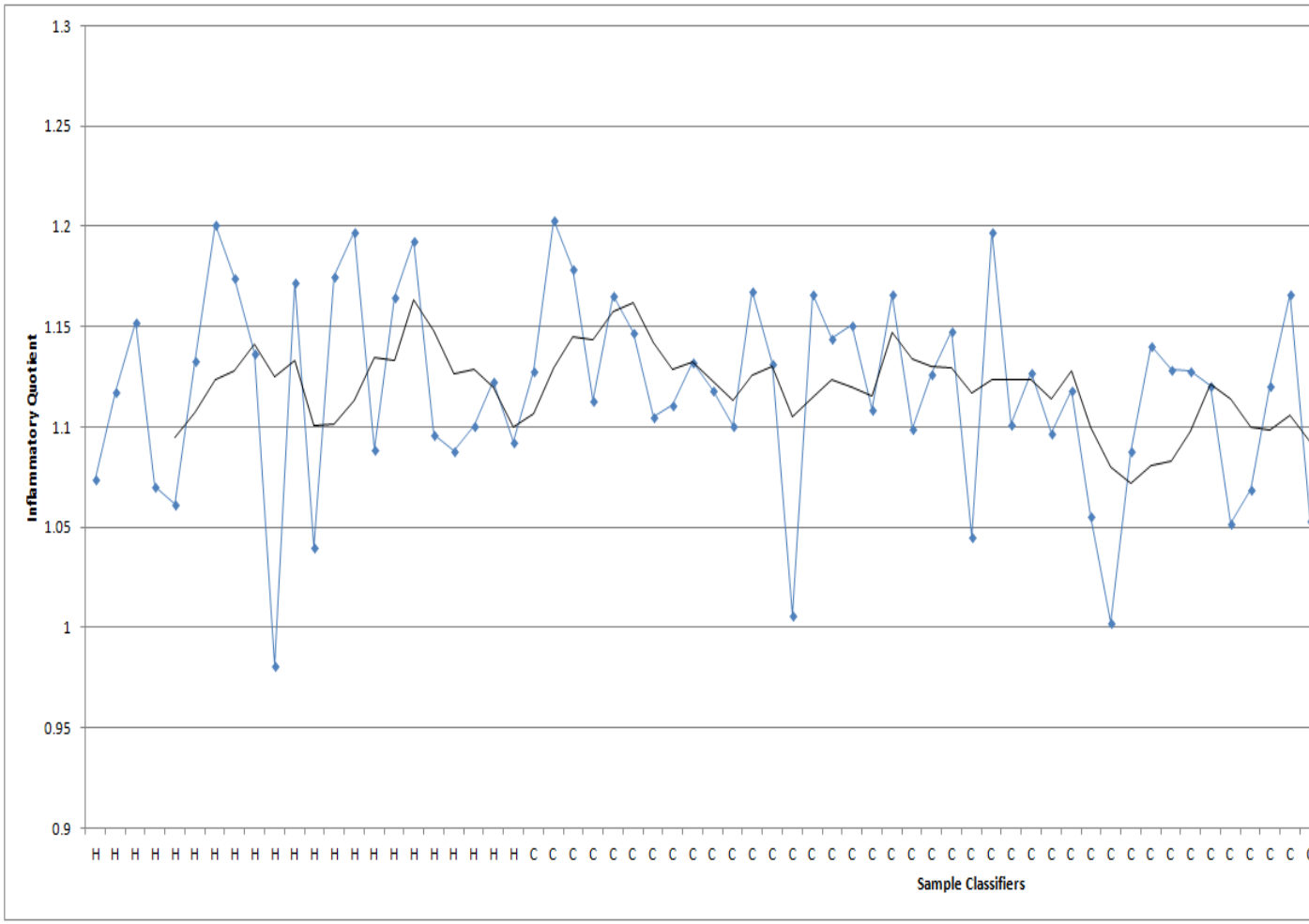

**Figure A3-34: Inflammatory quotient values (blue) for patient SP54, plotted over time, with sample classificatio exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving averag identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

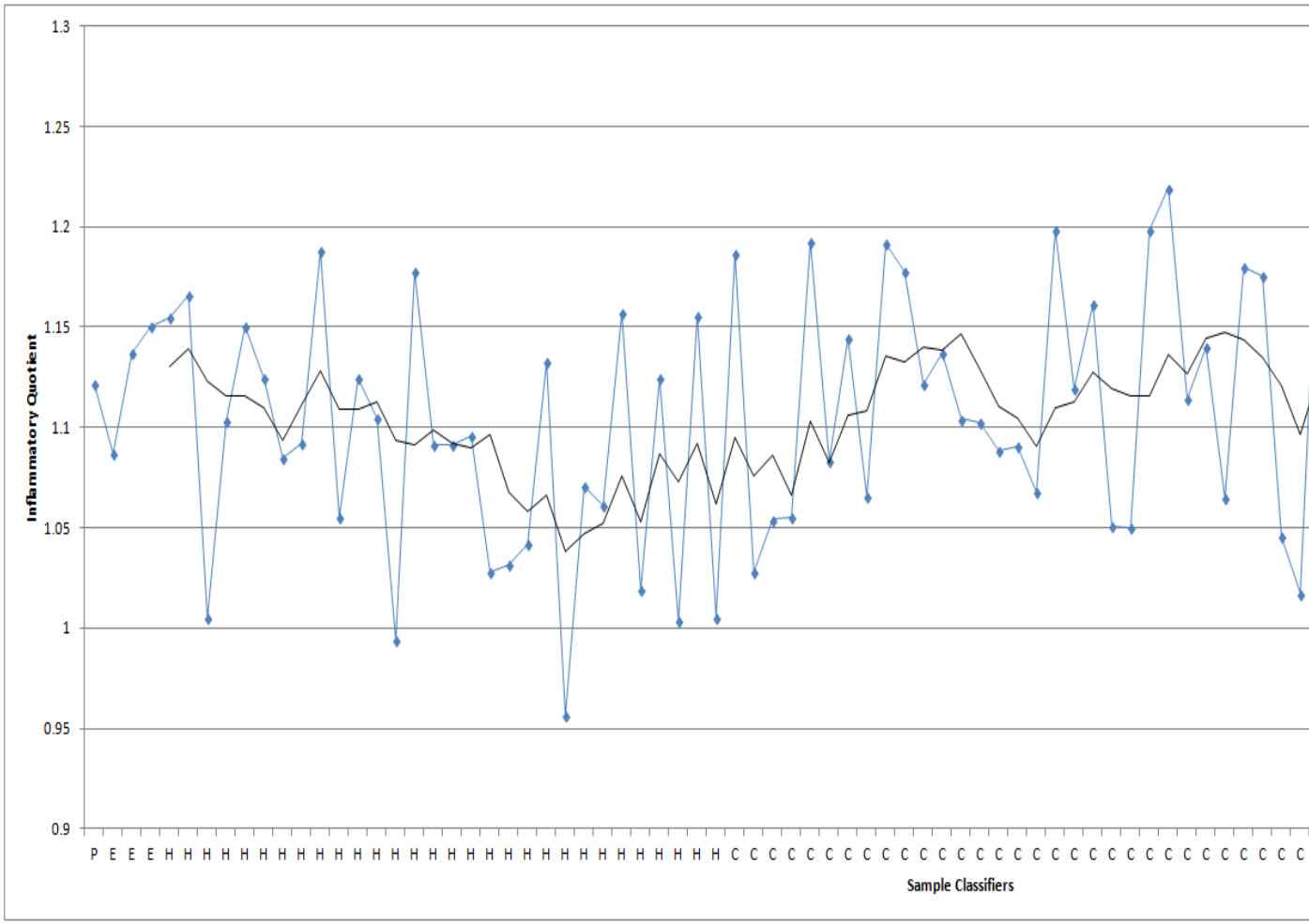

**Figure A3-35: Inflammatory quotient values (blue) for patient SP55, plotted over time, with sample classificat exacerbation (E), baseline control (C), pre-exacerbation (P), or post-exacerbation (H). A 5-point moving av identify overall short-term, weekly and long-term patterns in sputum molecular structural change.**

## Appendix 4: Spectra Generated with and without the Sampling Strip on the Bruker Alpha

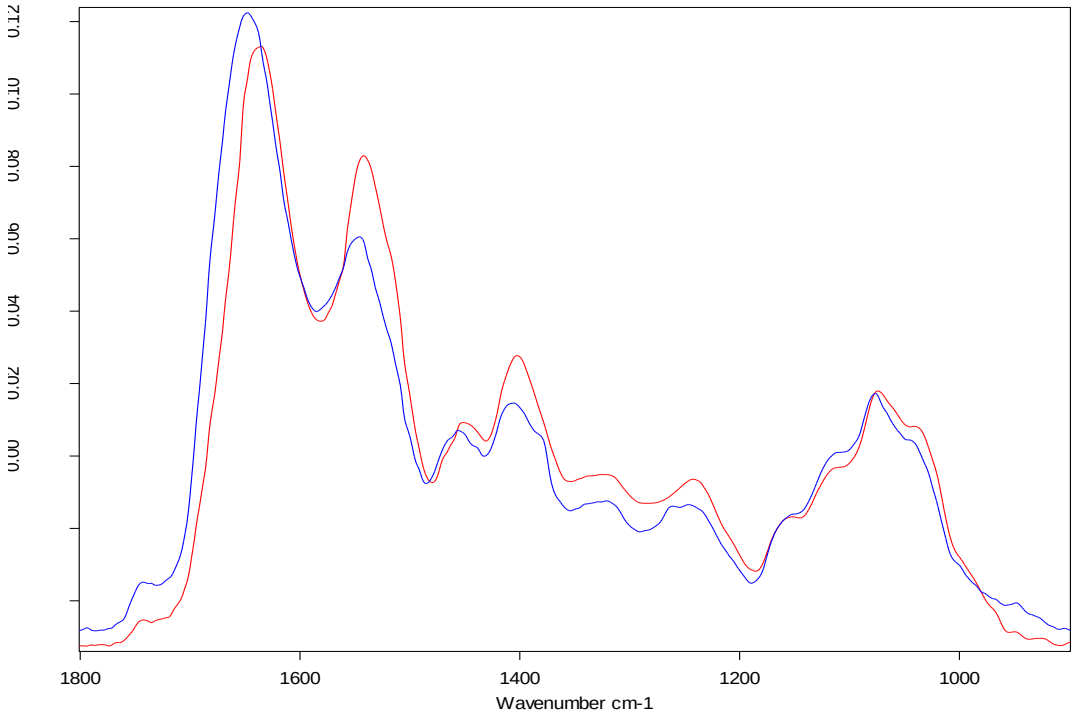


**Figure A4-1: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0002 dried onto the sampling strip (red), and the ATR crystal (blue)**

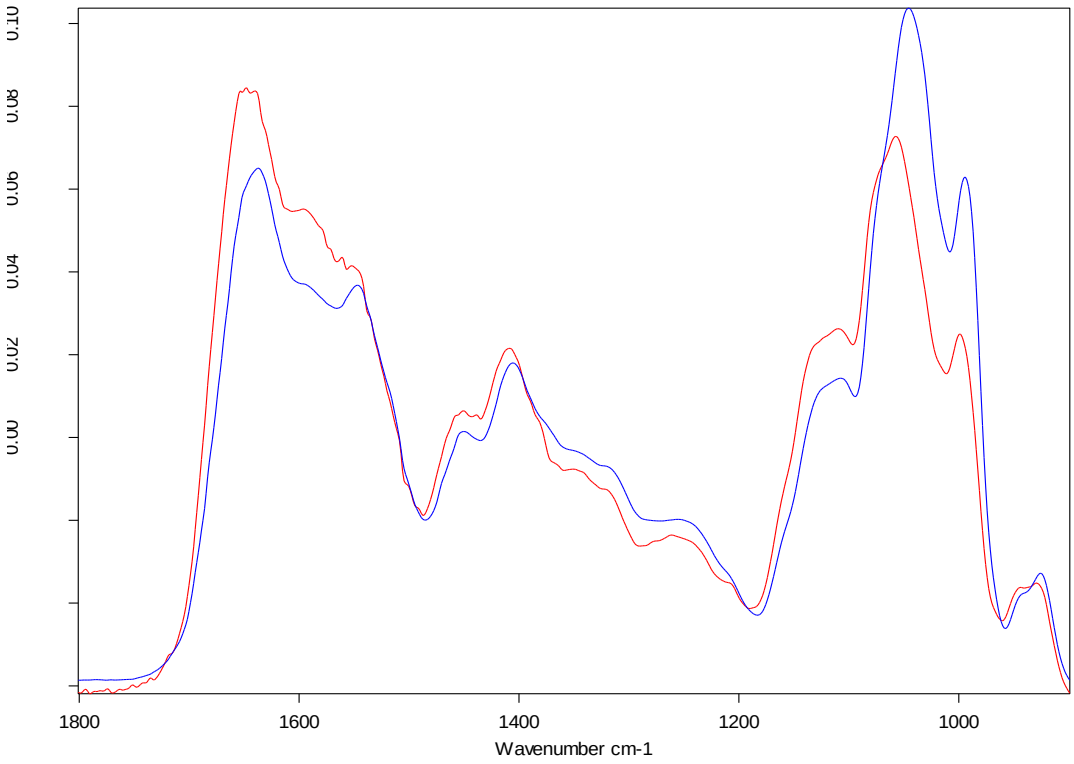


**Figure A4-2: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0005 dried onto the sampling strip (red), and the ATR crystal (blue)**

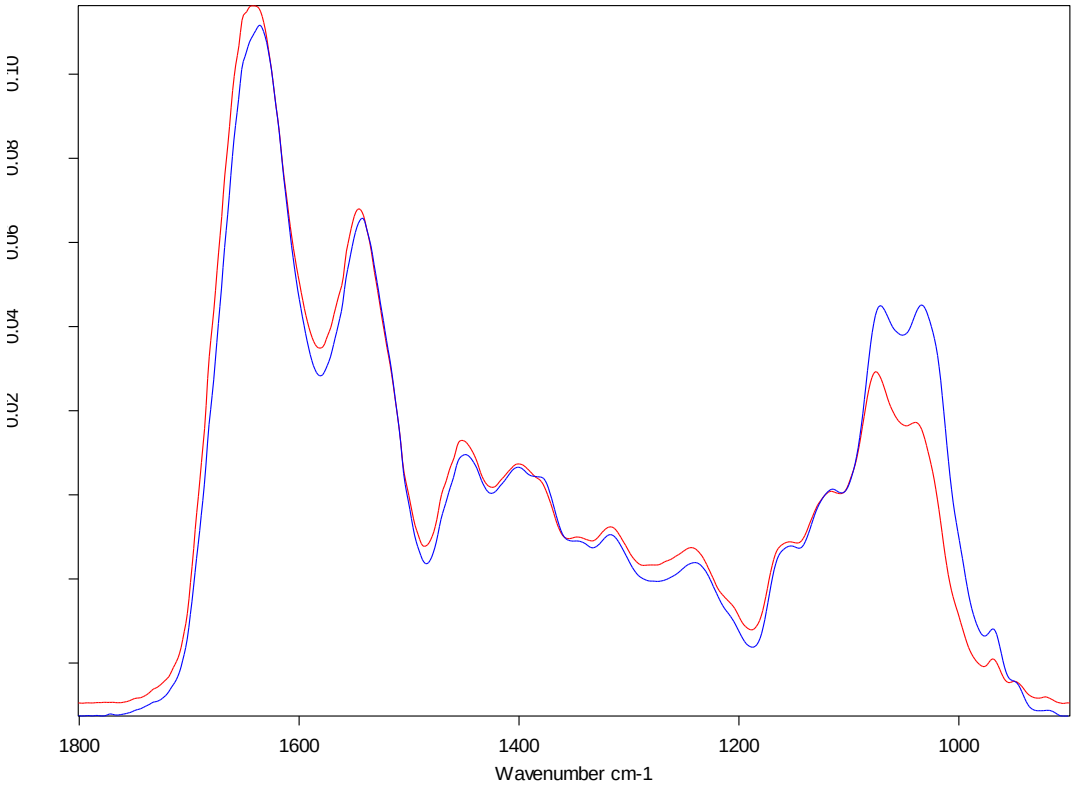


**Figure A4-3: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0006 dried onto the sampling strip (red), and the ATR crystal (blue)**

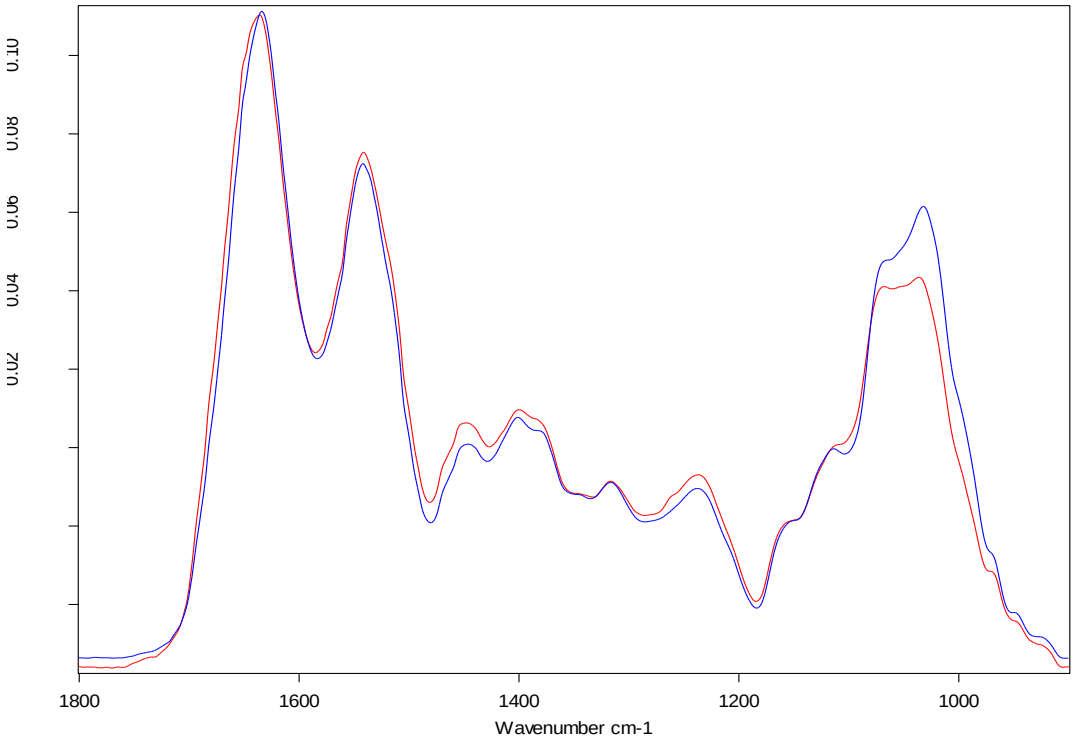


**Figure 4: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0010 dried onto the sampling strip (red), and the ATR crystal (blue)**

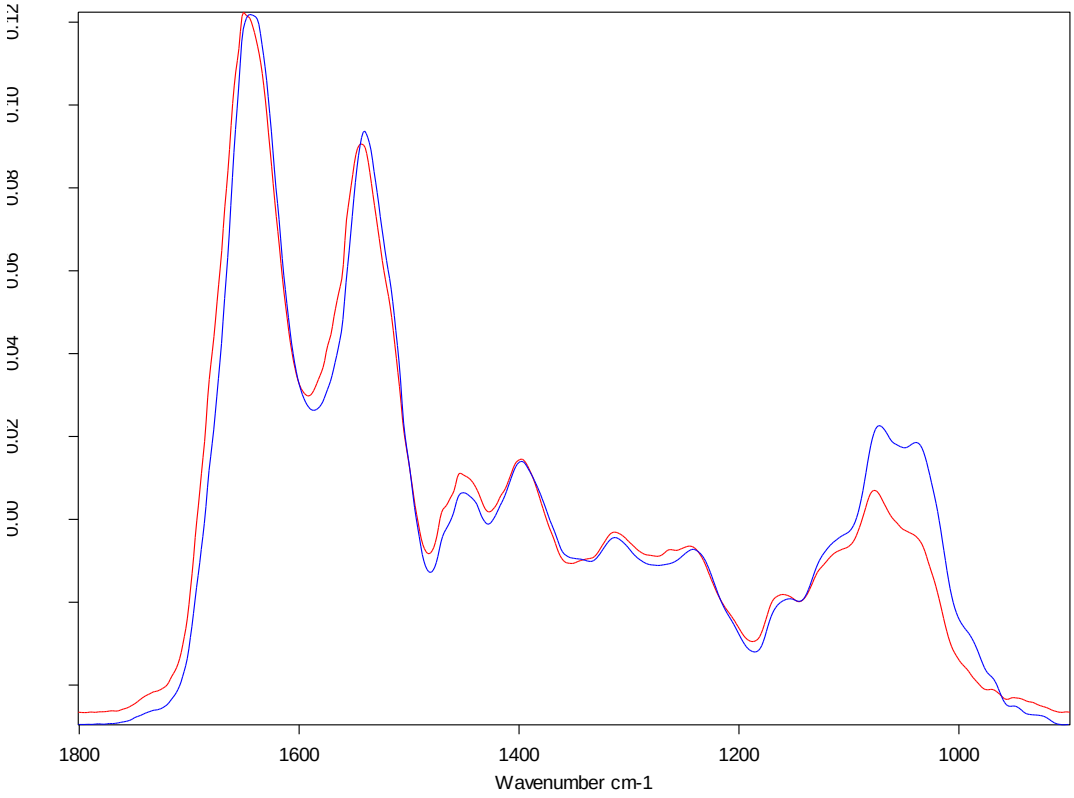


**Figure A4-5: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0013 dried onto the sampling strip (red), and the ATR crystal (blue)**

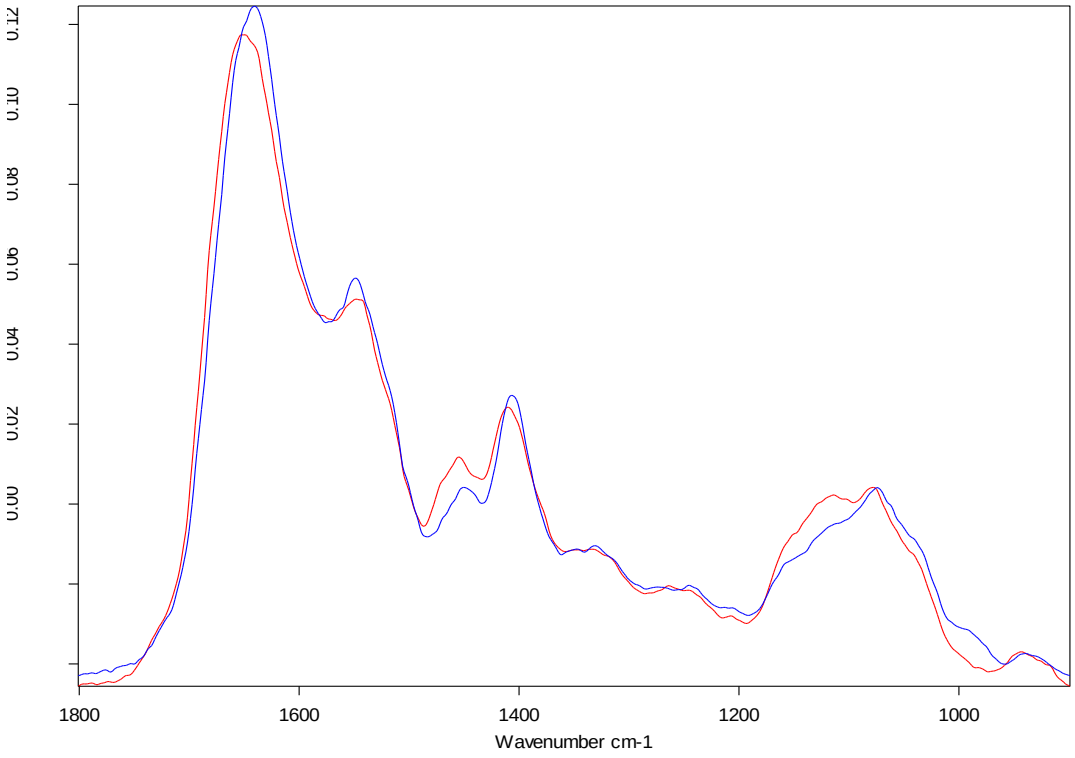


**Figure A4-6: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0014 dried onto the sampling strip (red), and the ATR crystal (blue)**

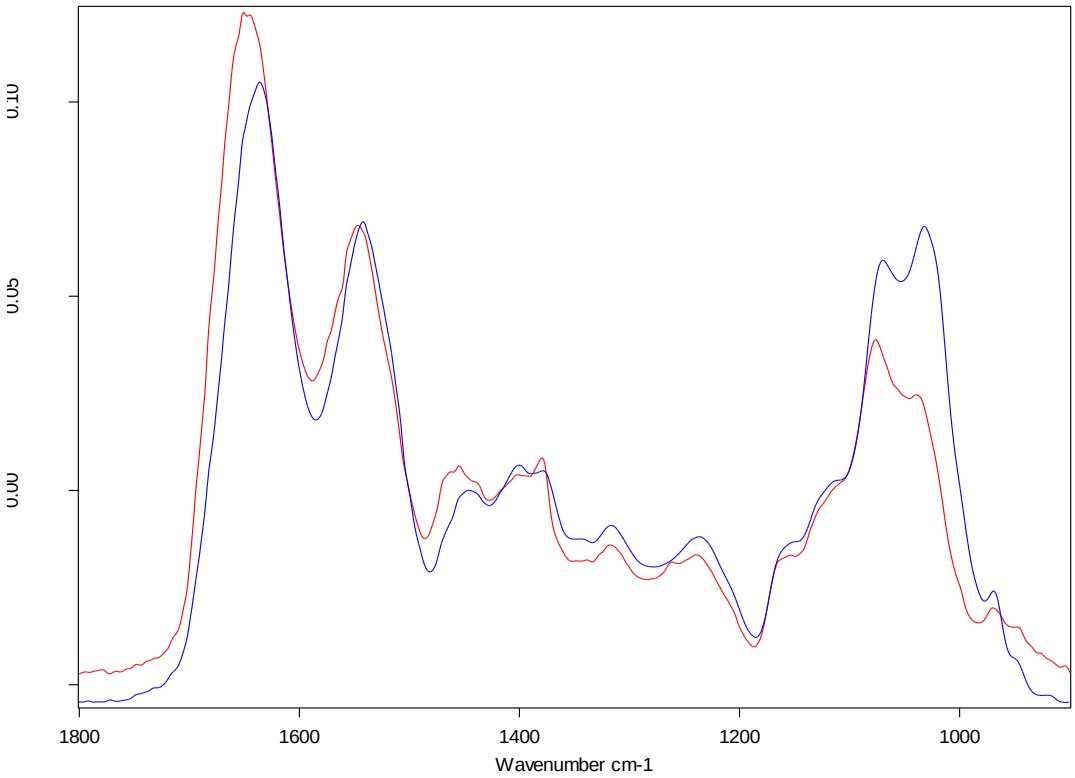


**Figure A4-7: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0015 dried onto the sampling strip (red), and the ATR crystal (blue)**

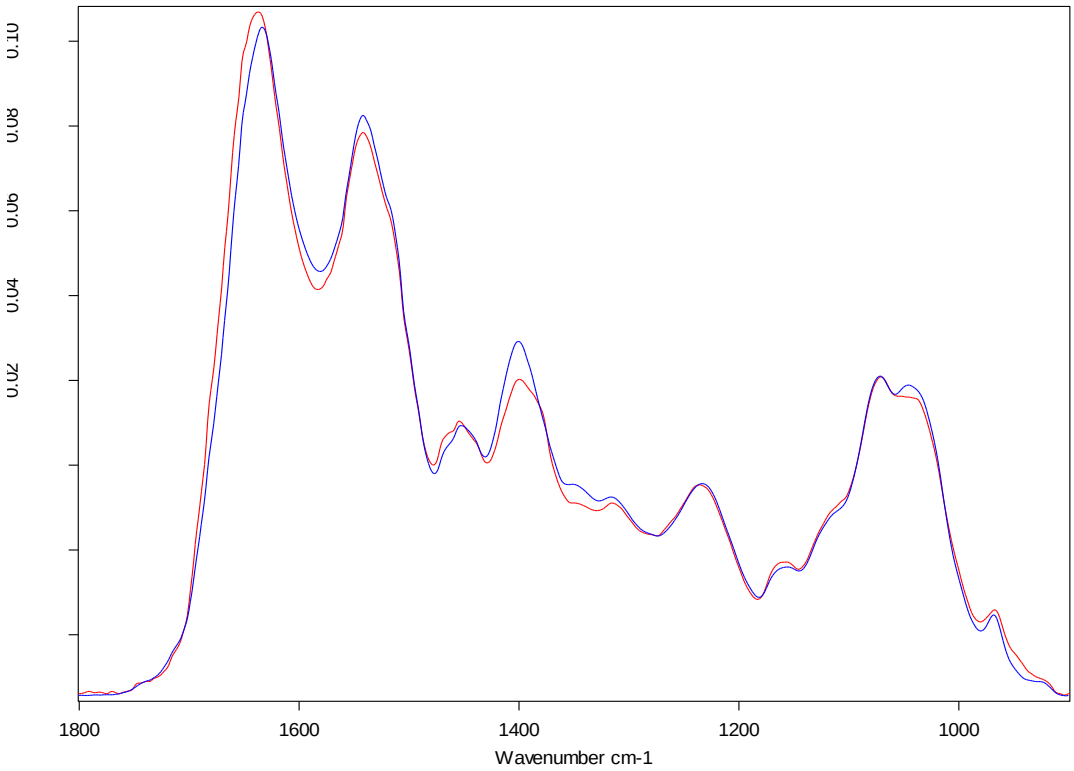


**Figure A4-8: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0016 dried onto the sampling strip (red), and the ATR crystal (blue)**

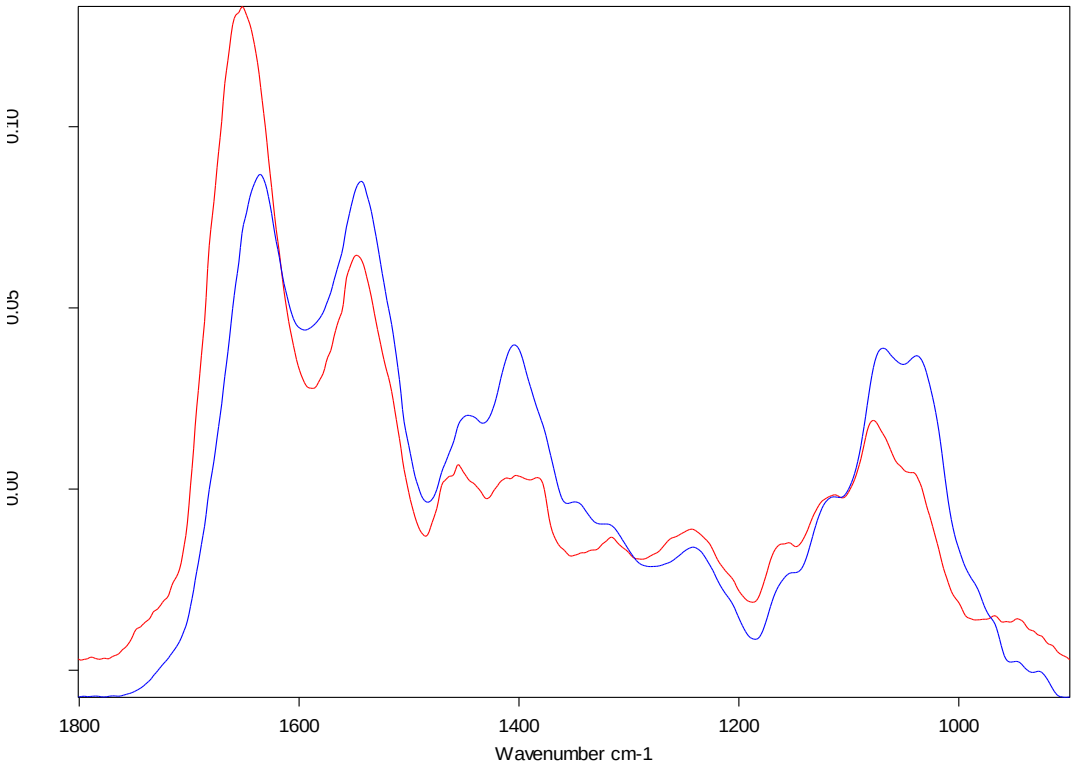


**Figure A4-9: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0017 dried onto the sampling strip (red), and the ATR crystal (blue)**

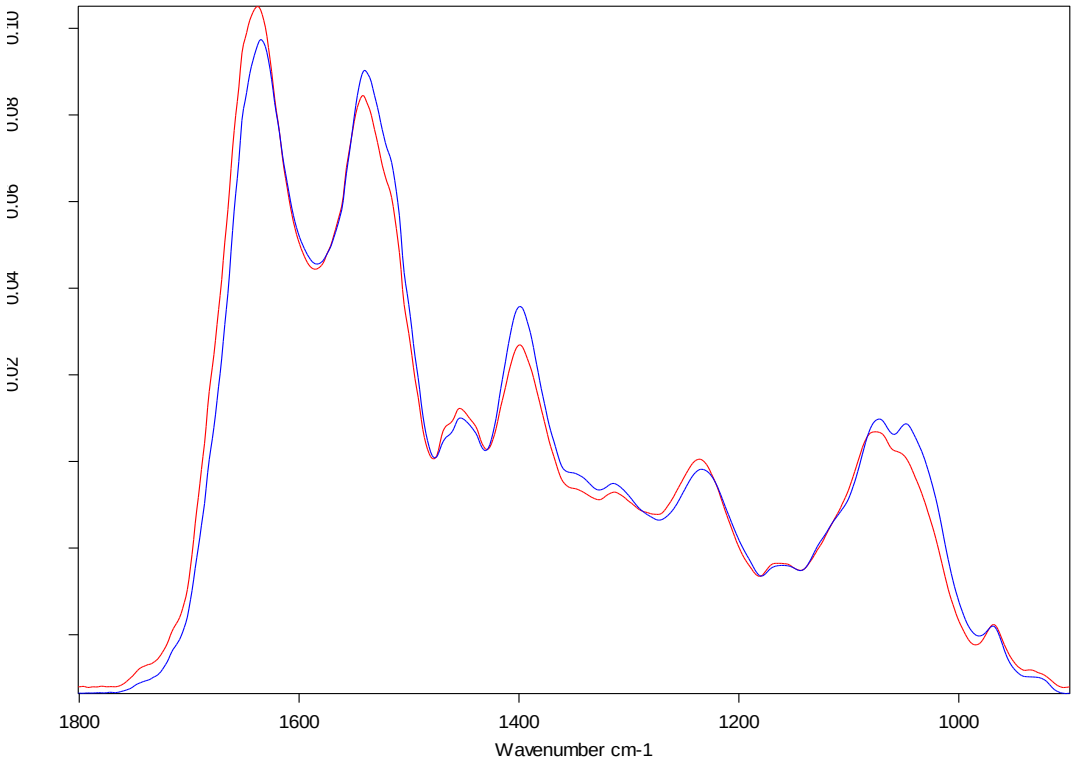


**Figure A4-10: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0018 dried onto the sampling strip (red), and the ATR crystal (blue)**

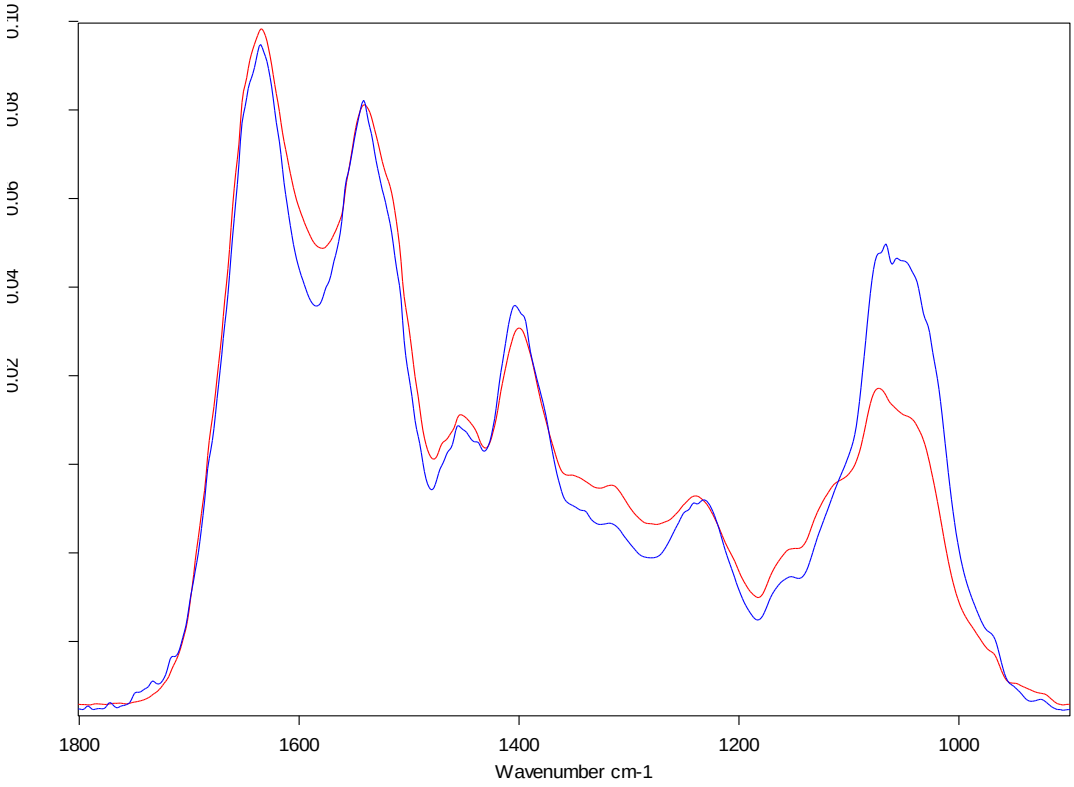


**Figure A4-11: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0020 dried onto the sampling strip (red), and the ATR crystal (blue)**

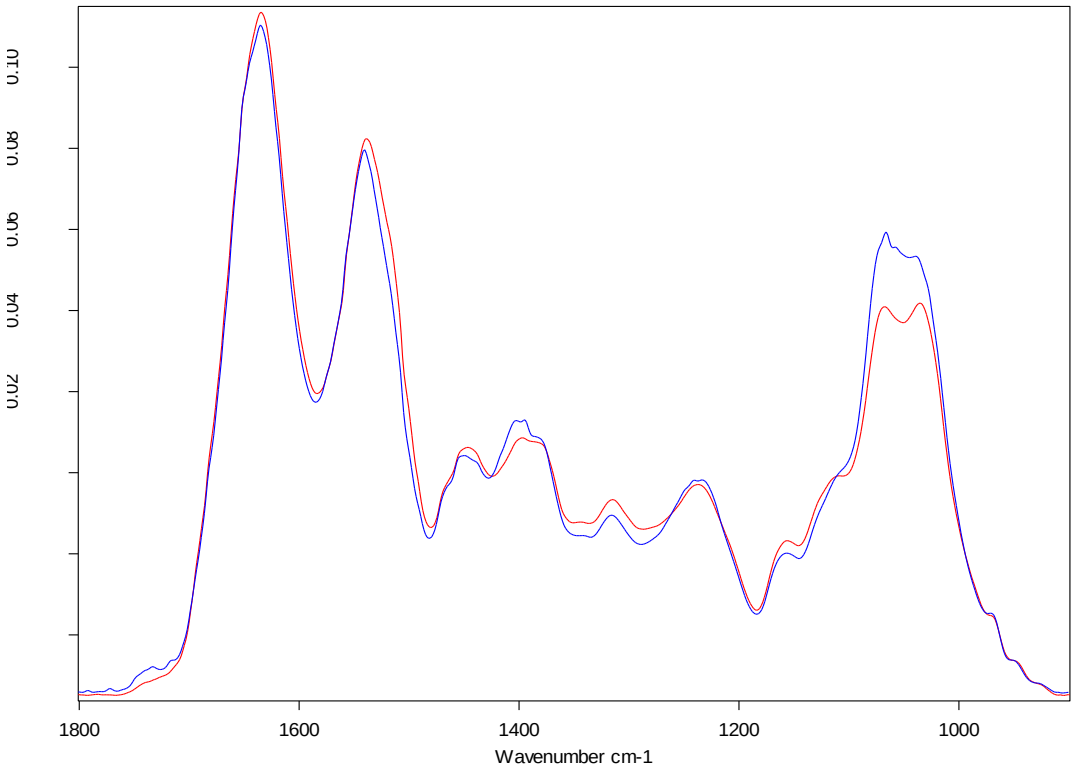


**Figure A4-12: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0021 dried onto the sampling strip (red), and the ATR crystal (blue)**

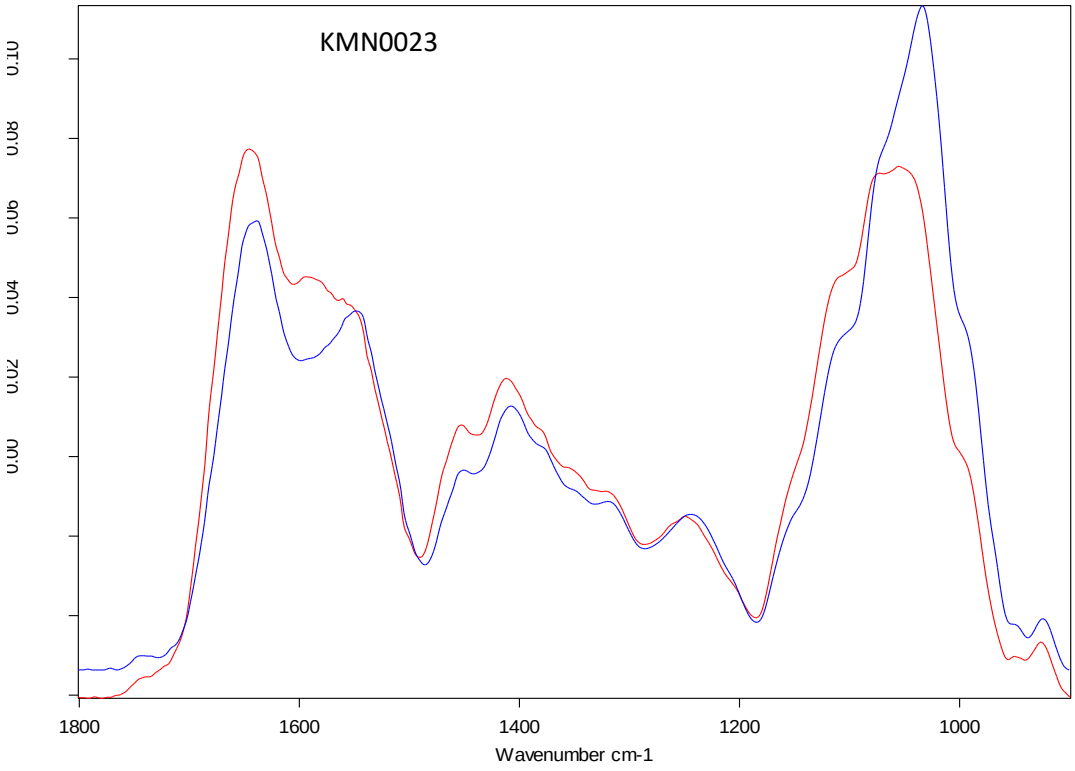


**Figure A4-13: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0023 dried onto the sampling strip (red), and the ATR crystal (blue)**

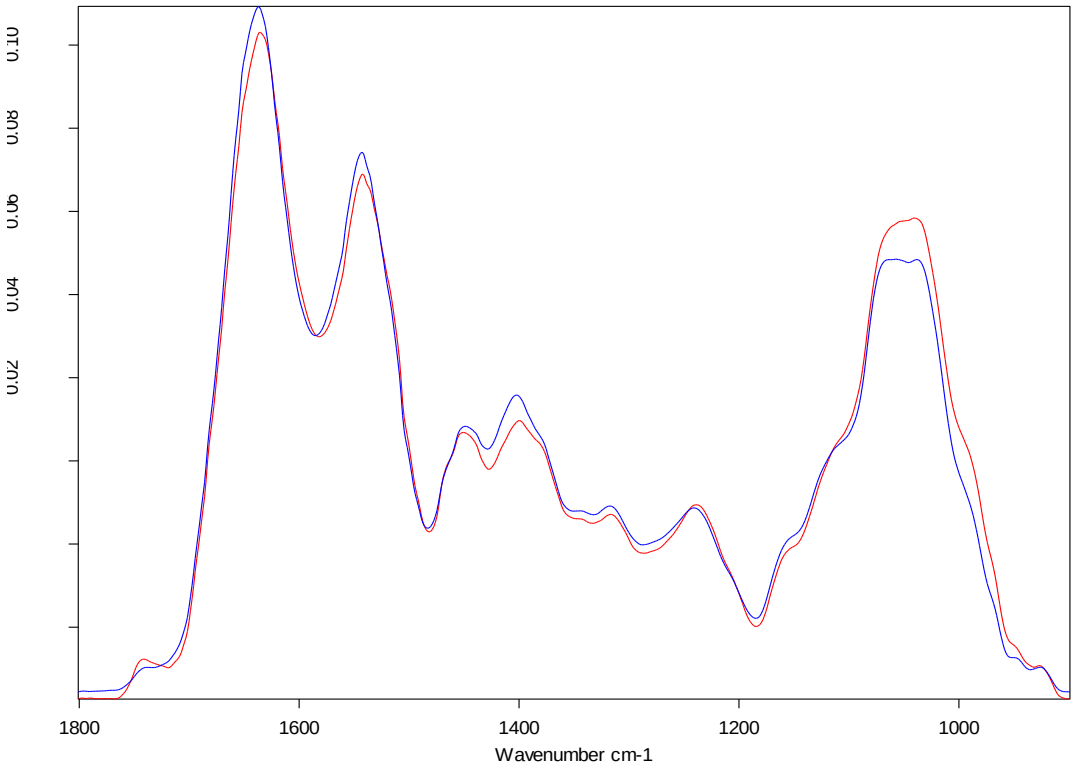


**Figure A4-14: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm$^{-1}$ of COPD sputum sample KMN0024 dried onto the sampling strip (red), and the ATR crystal (blue)**

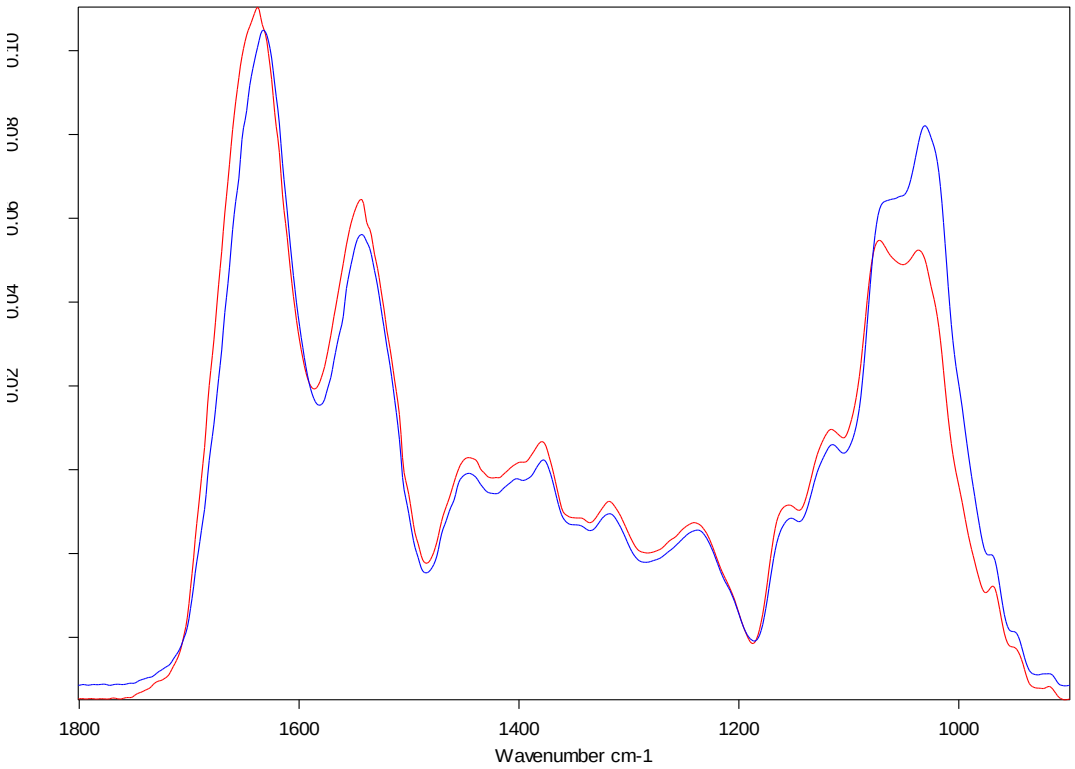


**Figure A4-15: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0028 dried onto the sampling strip (red), and the ATR crystal (blue)**

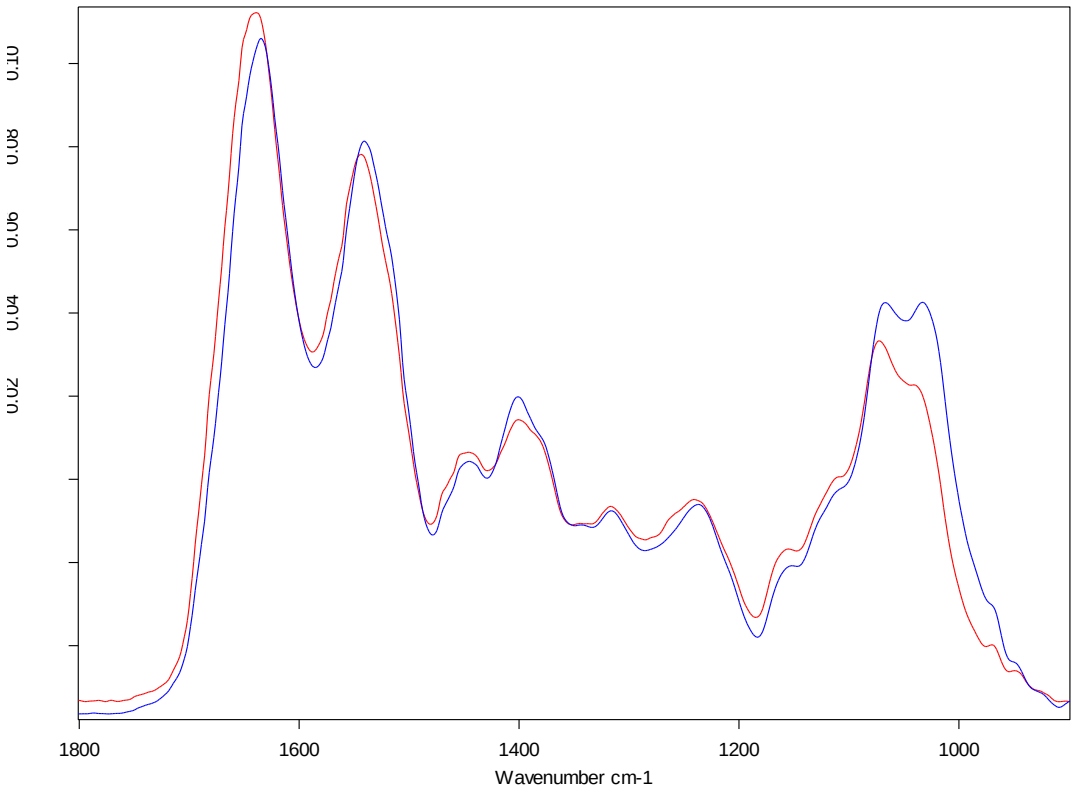


**Figure A4-16: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0031 dried onto the sampling strip (red), and the ATR crystal (blue)**

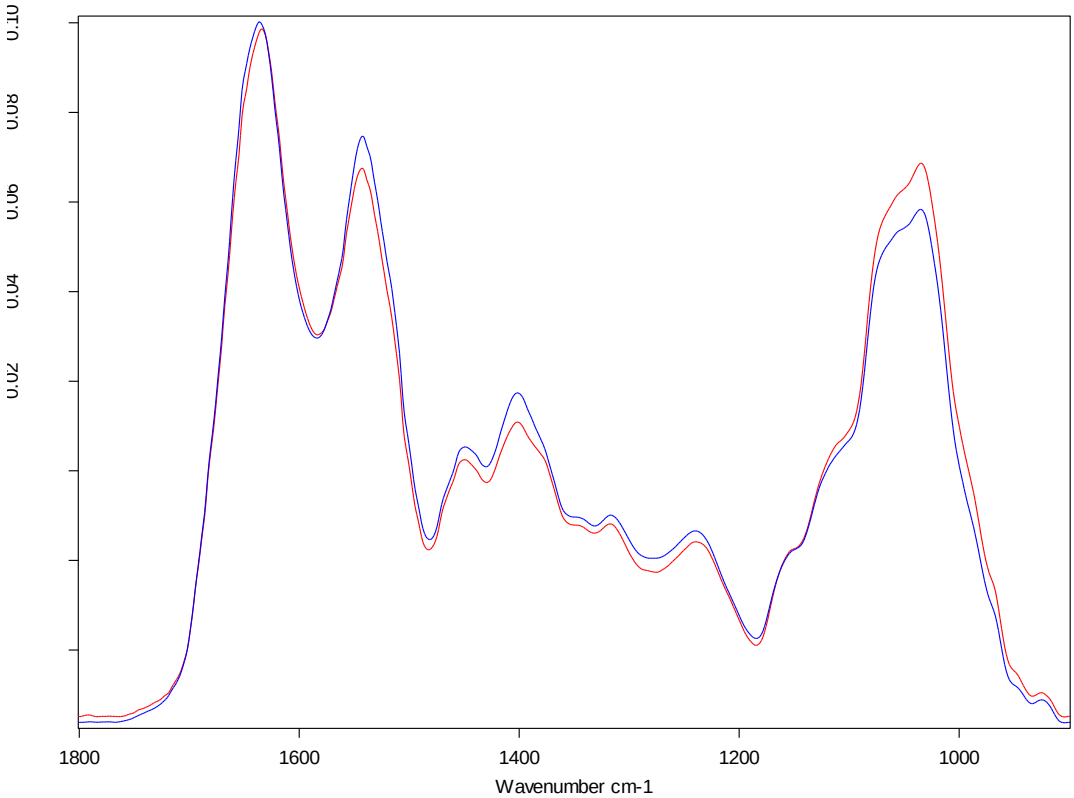


**Figure A4-17: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0031 dried onto the sampling strip (red), and the ATR crystal (blue)**

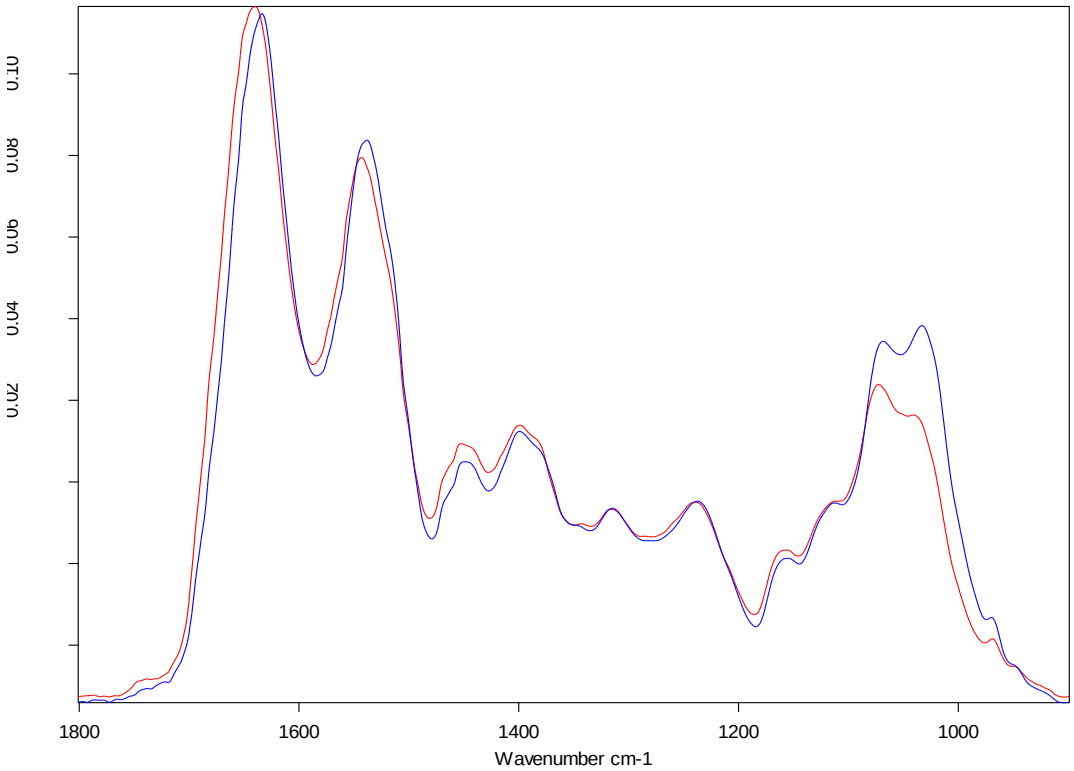


**Figure A4-18: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0039 dried onto the sampling strip (red), and the ATR crystal (blue)**

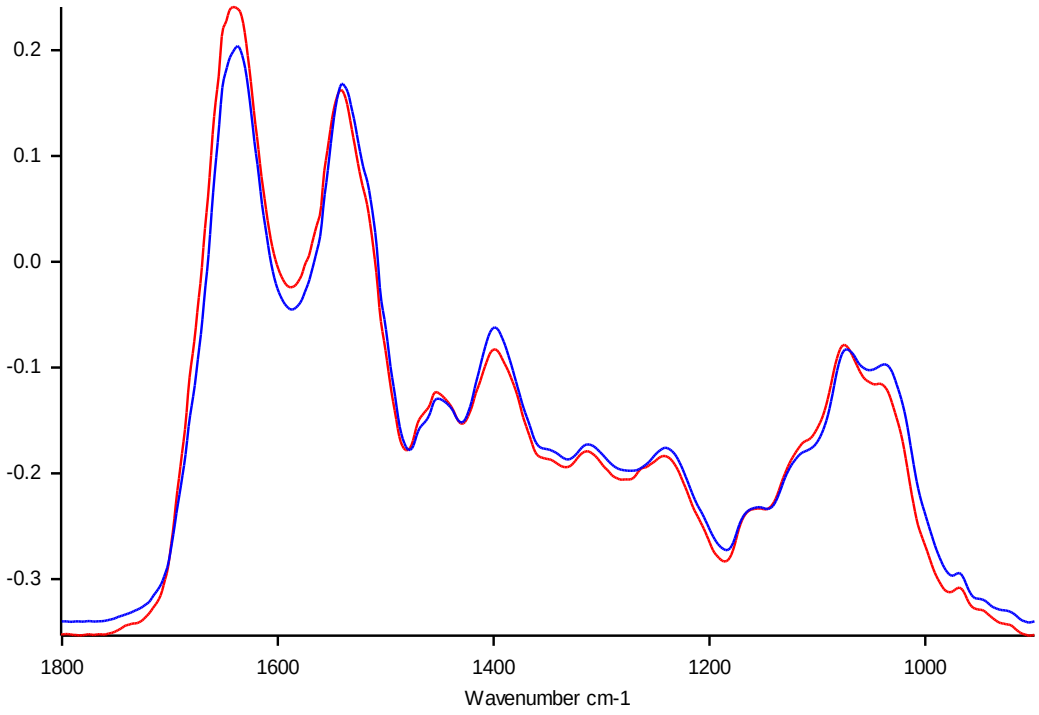


**Figure A4-19: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0040 dried onto the sampling strip (red), and the ATR crystal (blue)**

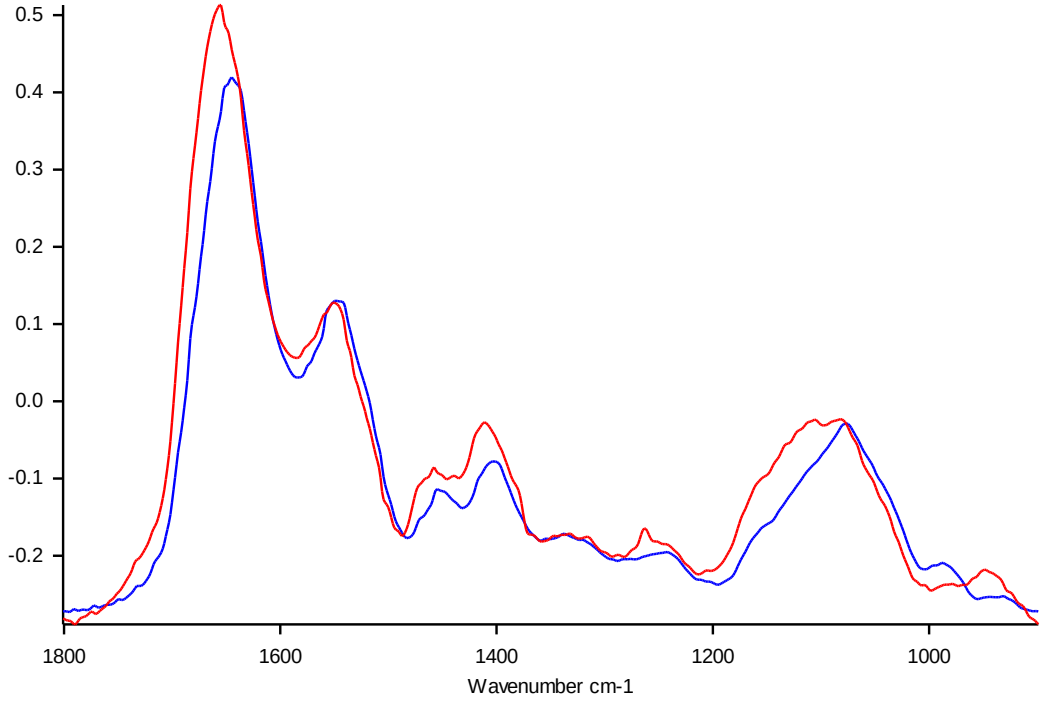


**Figure A4-20: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0042 dried onto the sampling strip (red), and the ATR crystal (blue)**

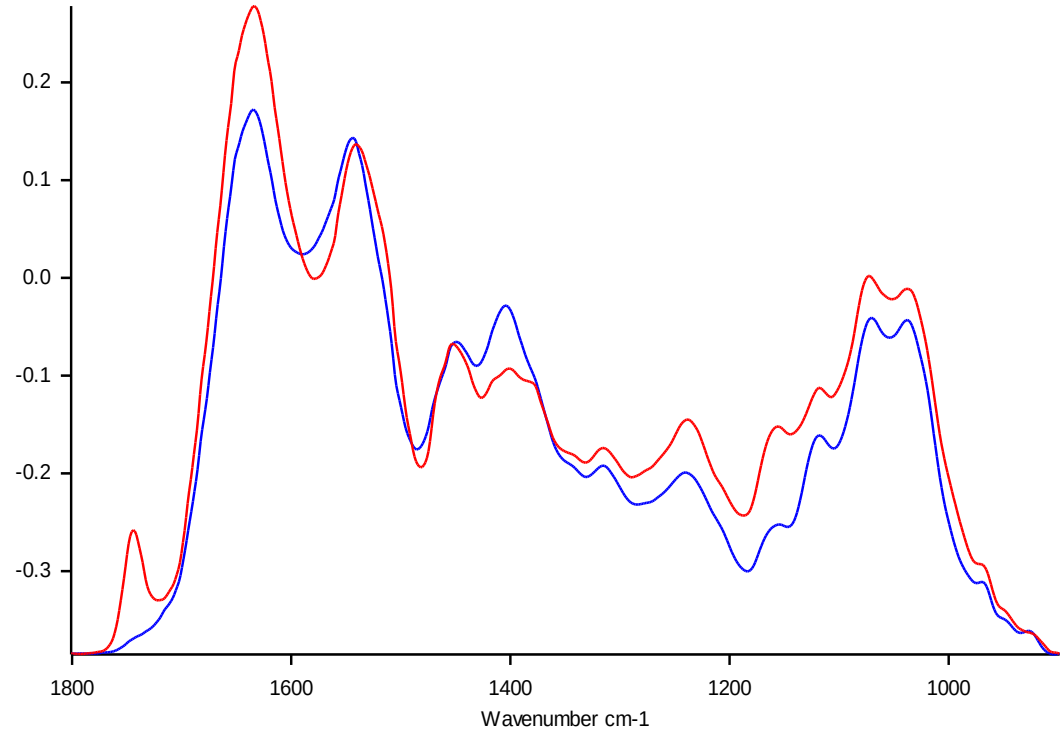


**Figure A4-21: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0043 dried onto the sampling strip (red), and the ATR crystal (blue)**

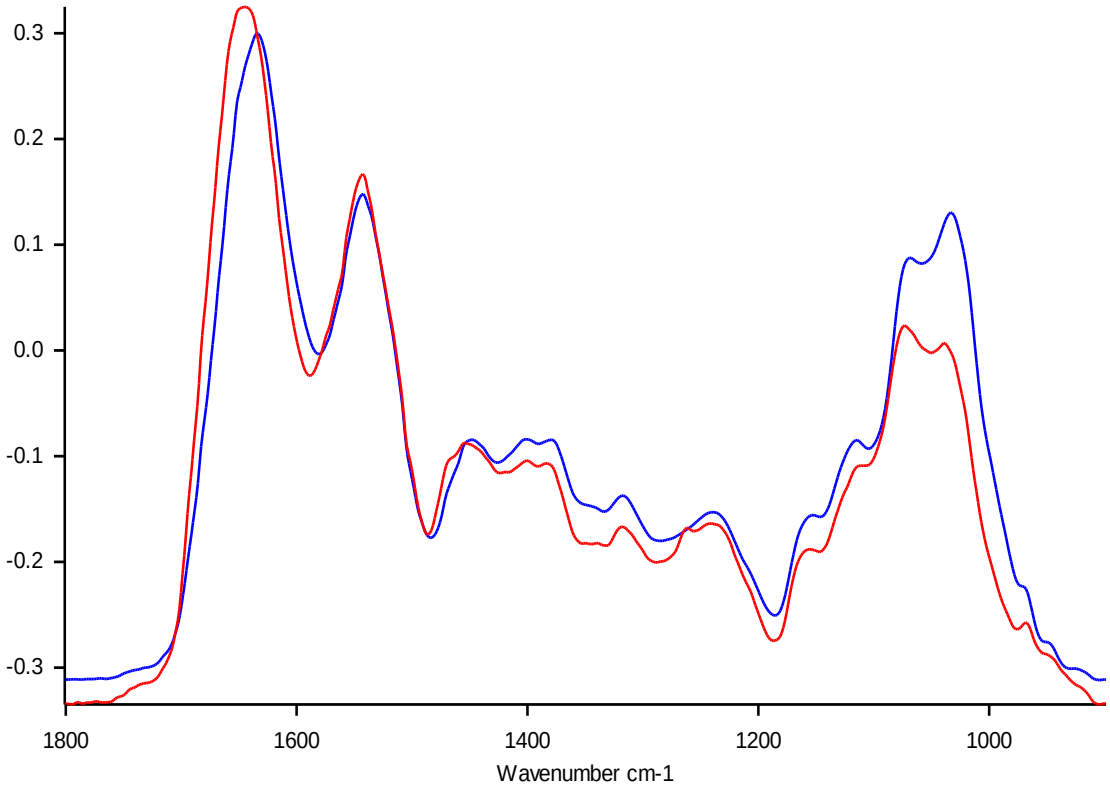


**Figure A4-22: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0046 dried onto the sampling strip (red), and the ATR crystal (blue)**

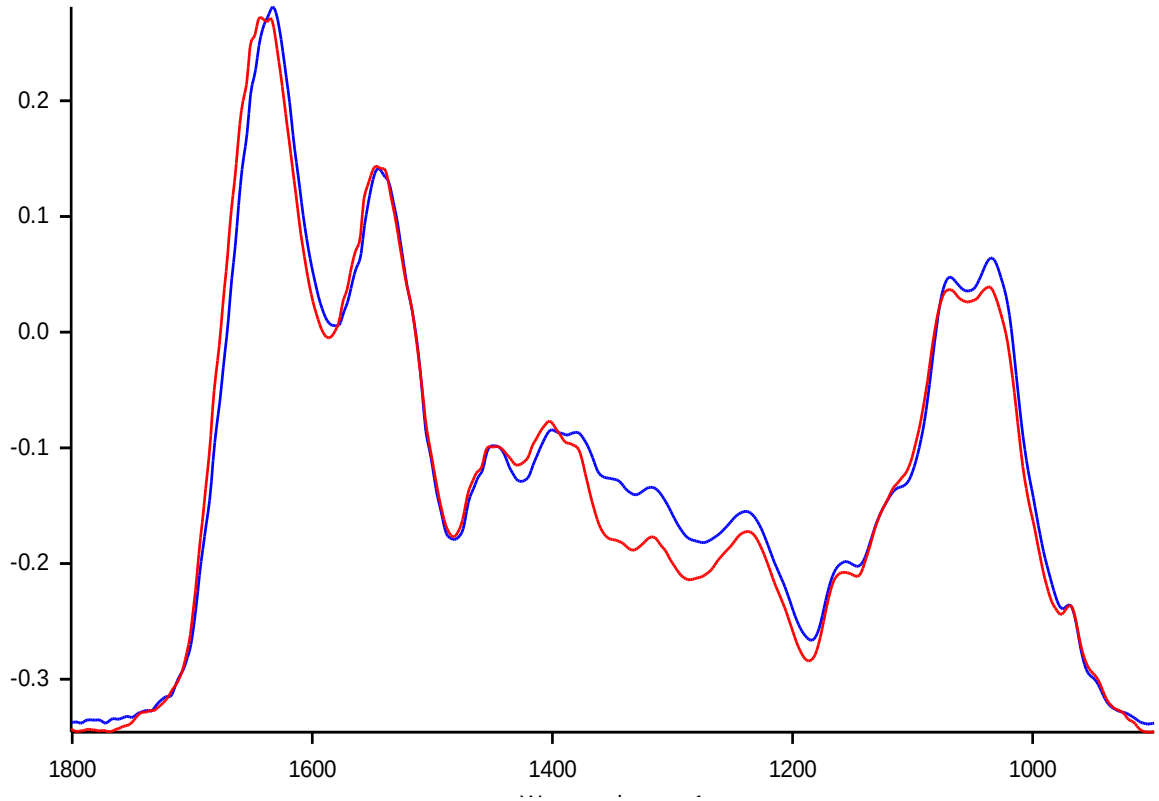


**Figure A4-23: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0049 dried onto the sampling strip (red), and the ATR crystal (blue)**

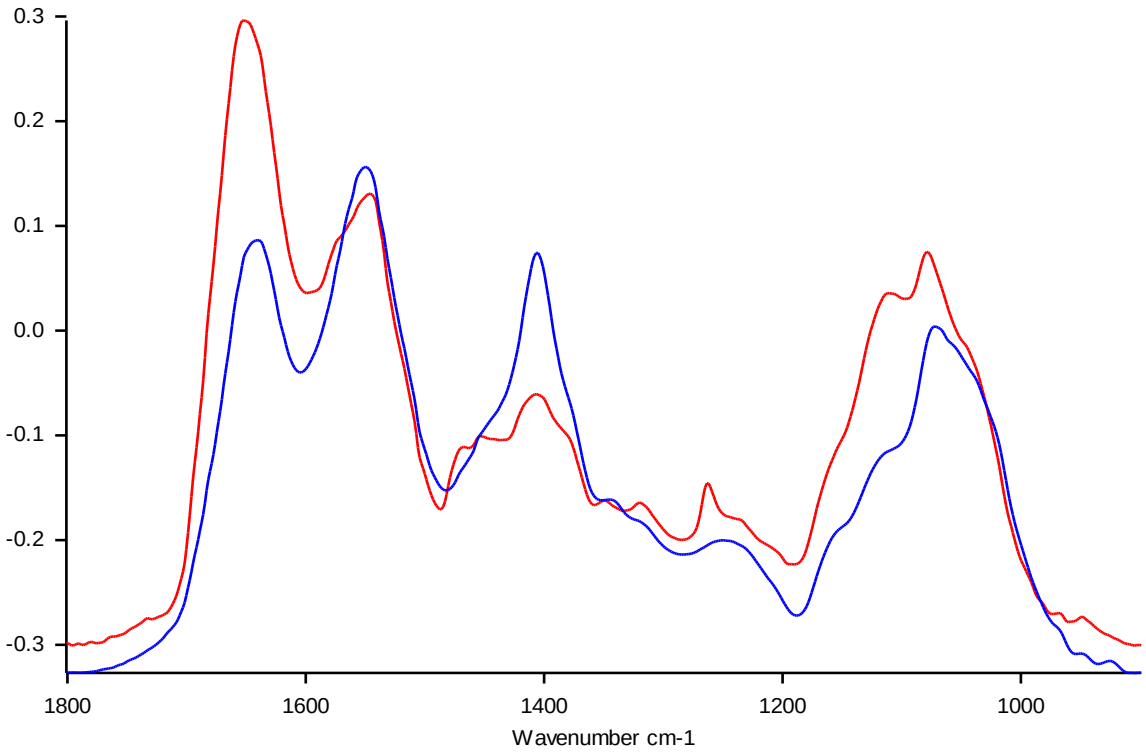


**Figure A4-24: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0050 dried onto the sampling strip (red), and the ATR crystal (blue)**

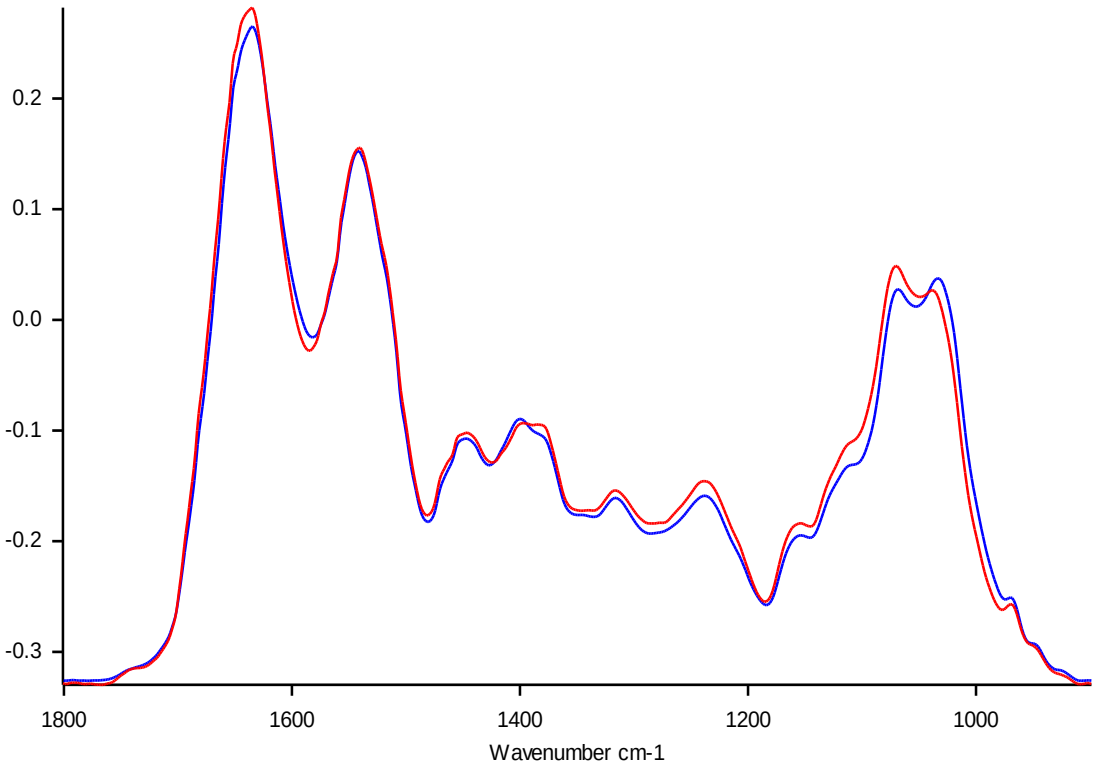


**Figure A4-25: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0052 dried onto the sampling strip (red), and the ATR crystal (blue)**

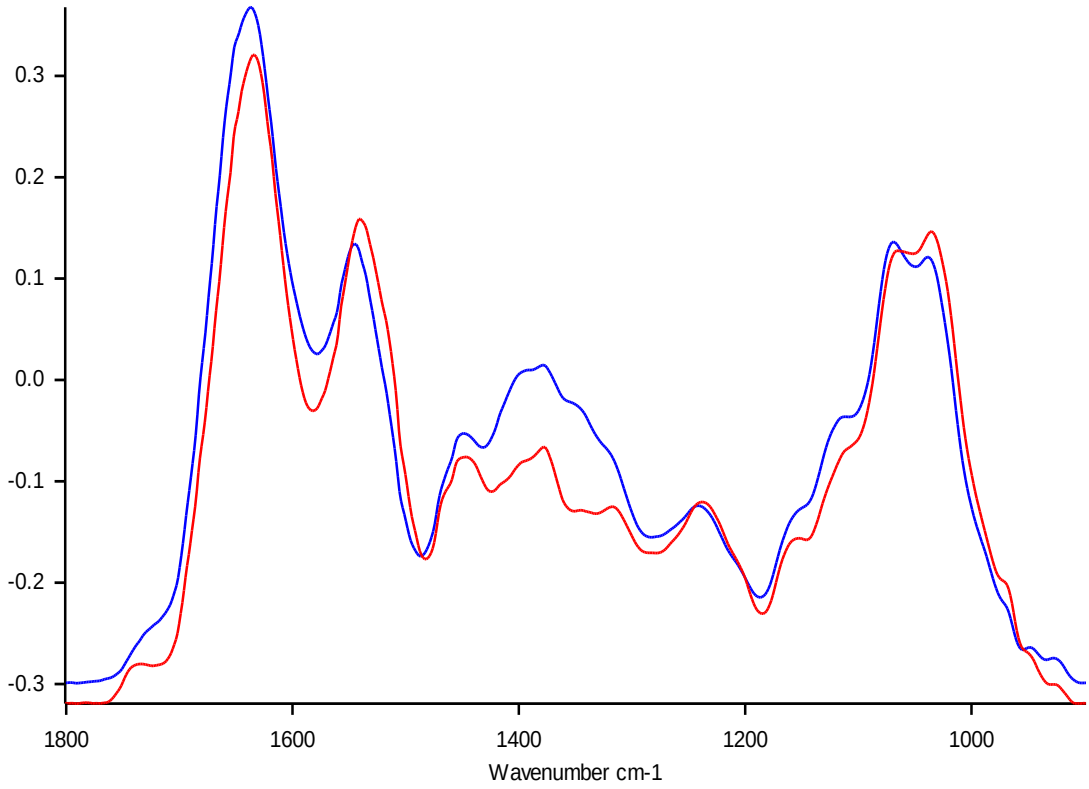


**Figure A4-26: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0053 dried onto the sampling strip (red), and the ATR crystal (blue)**

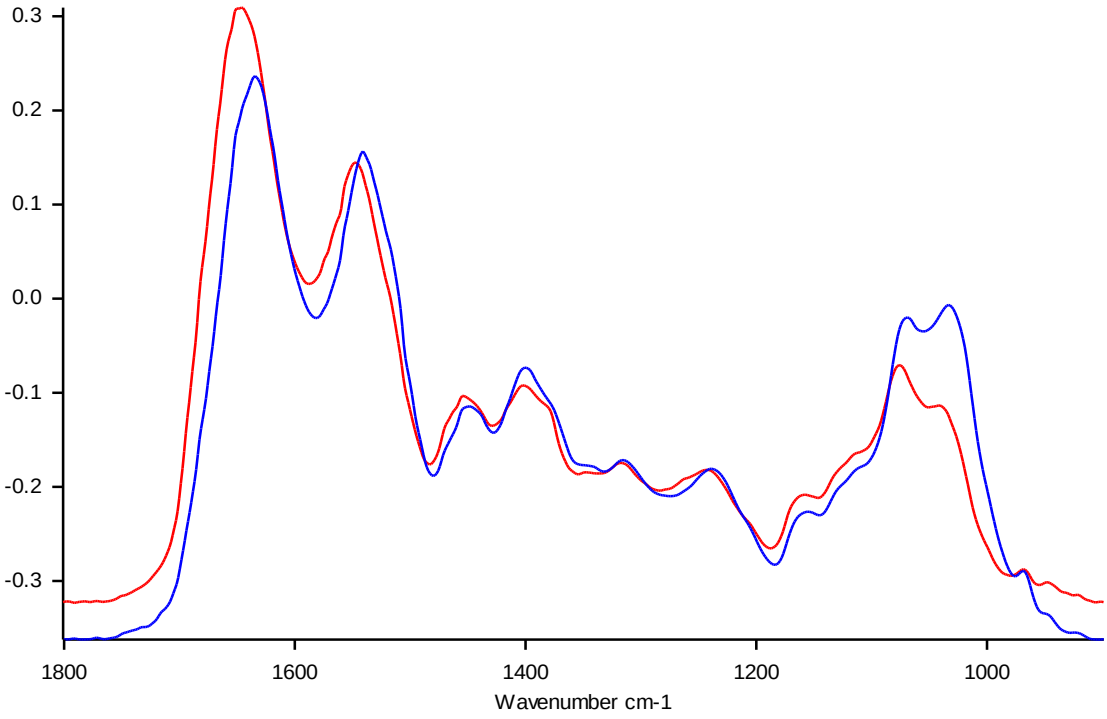


**Figure A4-27: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0056 dried onto the sampling strip (red), and the ATR crystal (blue)**

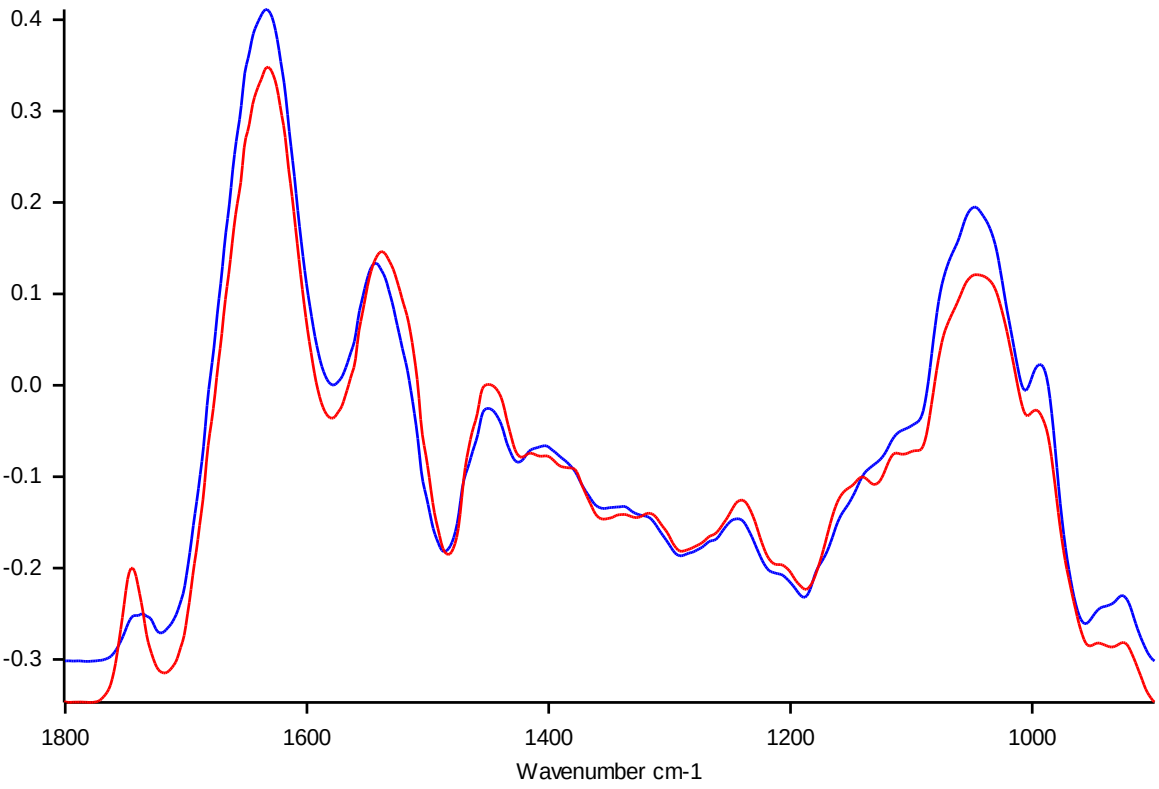


**Figure A4-28: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0057 dried onto the sampling strip (red), and the ATR crystal (blue)**

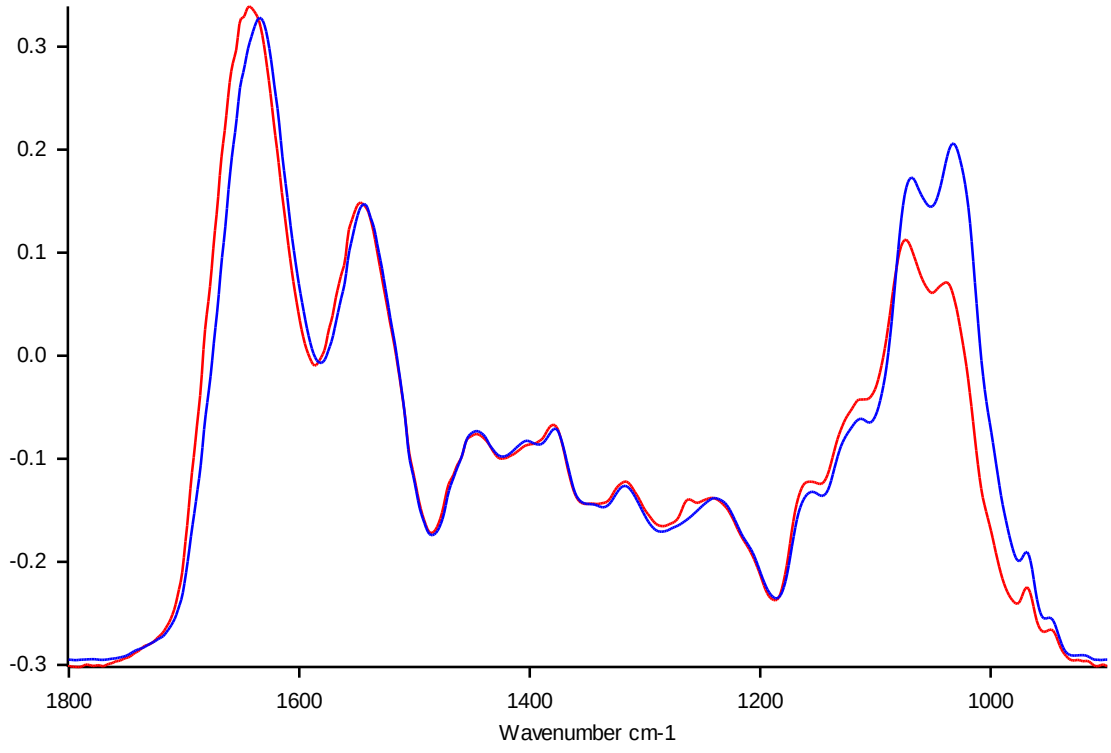


**Figure A4-29: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0058 dried onto the sampling strip (red), and the ATR crystal (blue)**

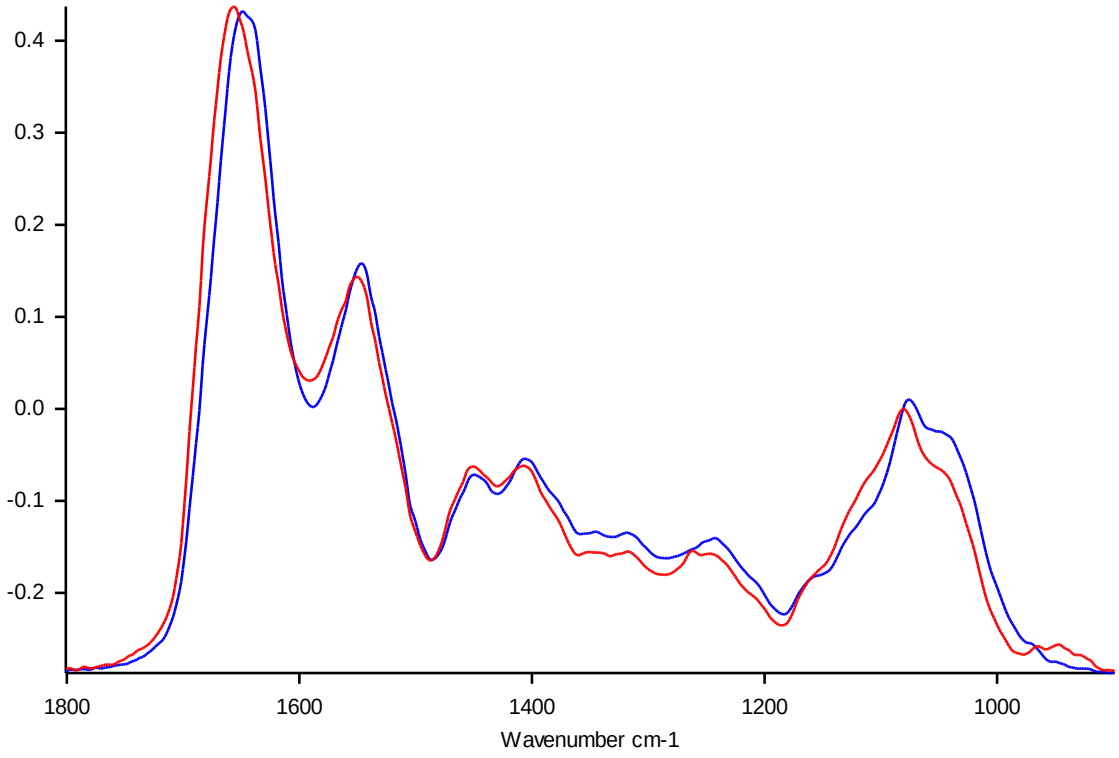


**Figure A4-30: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0059 dried onto the sampling strip (red), and the ATR crystal (blue)**

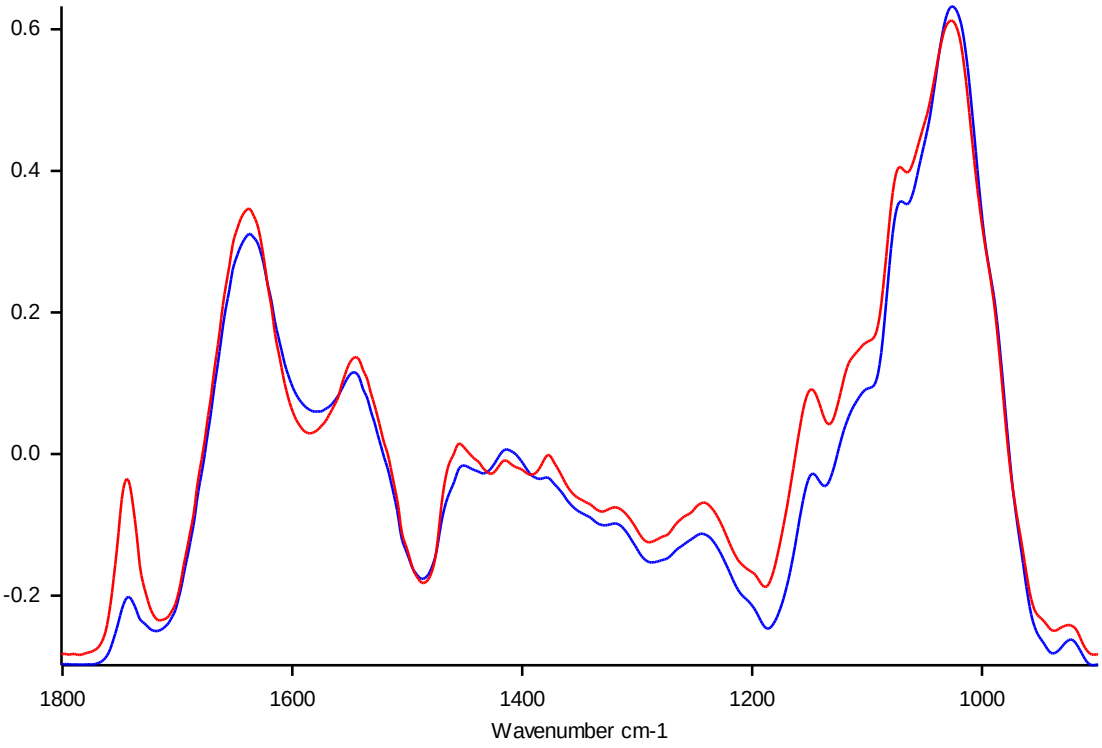


**Figure A4-31: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0061 dried onto the sampling strip (red), and the ATR crystal (blue)**

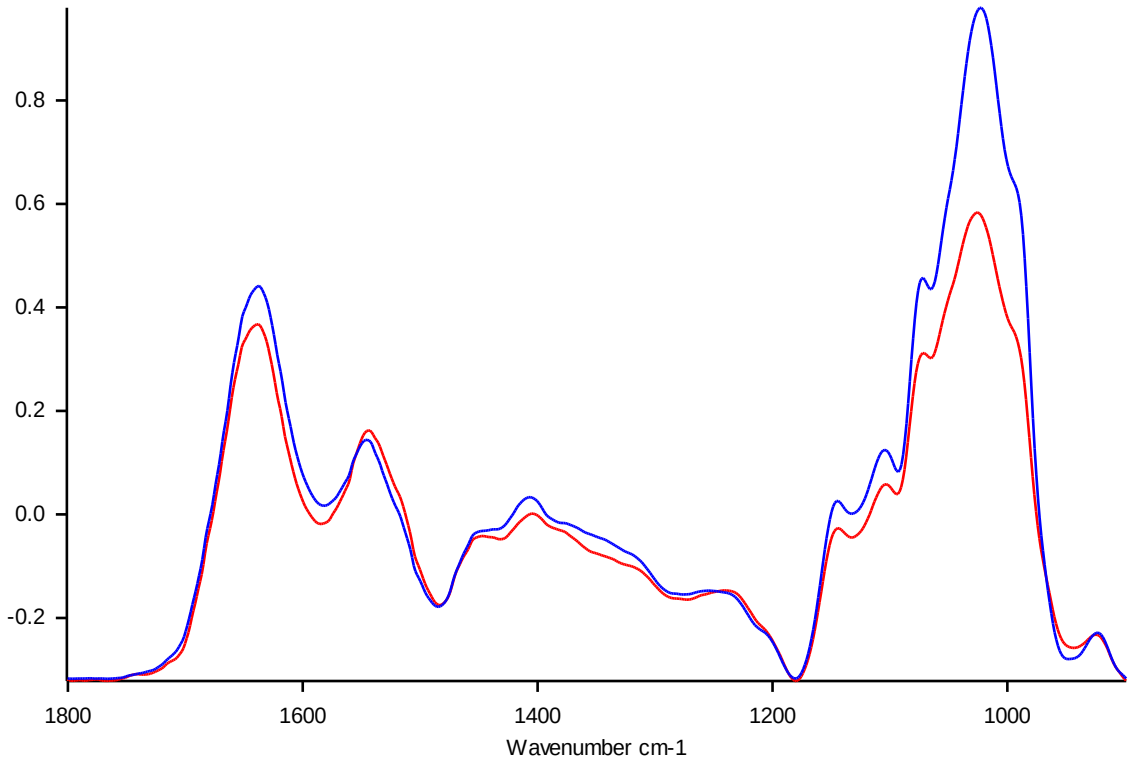


**Figure A4-32: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0062 dried onto the sampling strip (red), and the ATR crystal (blue)**

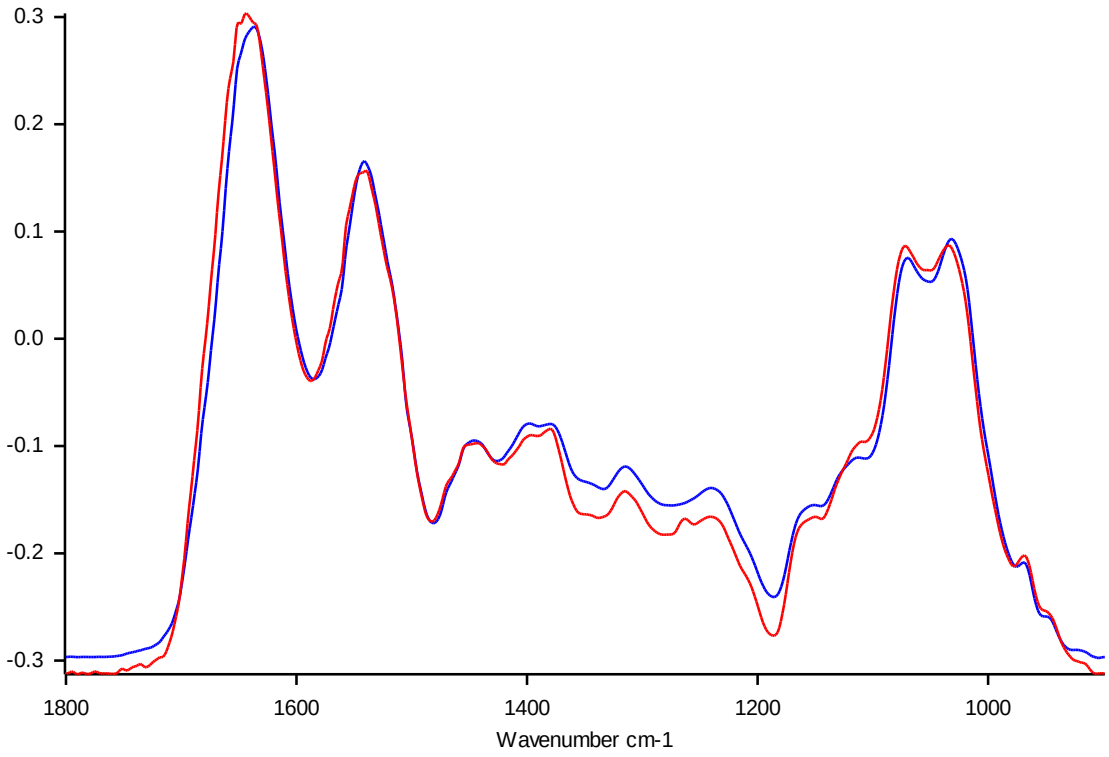


**Figure A4-33: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0063 dried onto the sampling strip (red), and the ATR crystal (blue)**

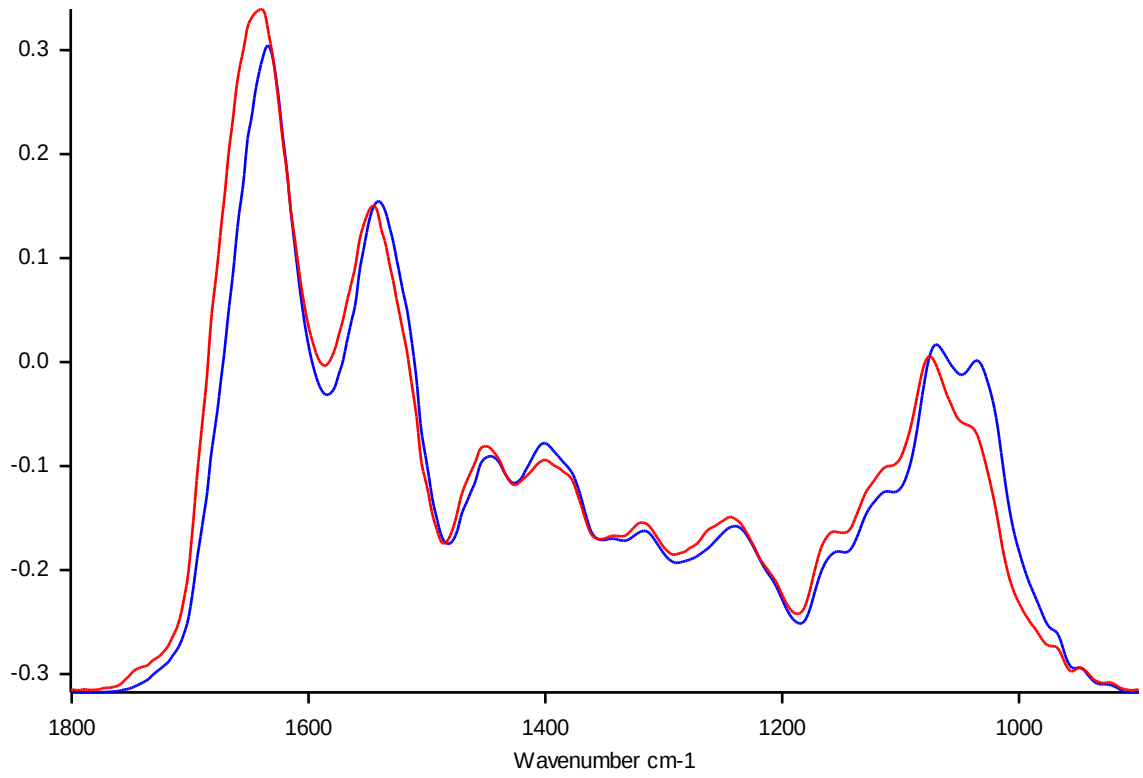


**Figure A4-34: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0064 dried onto the sampling strip (red), and the ATR crystal (blue)**

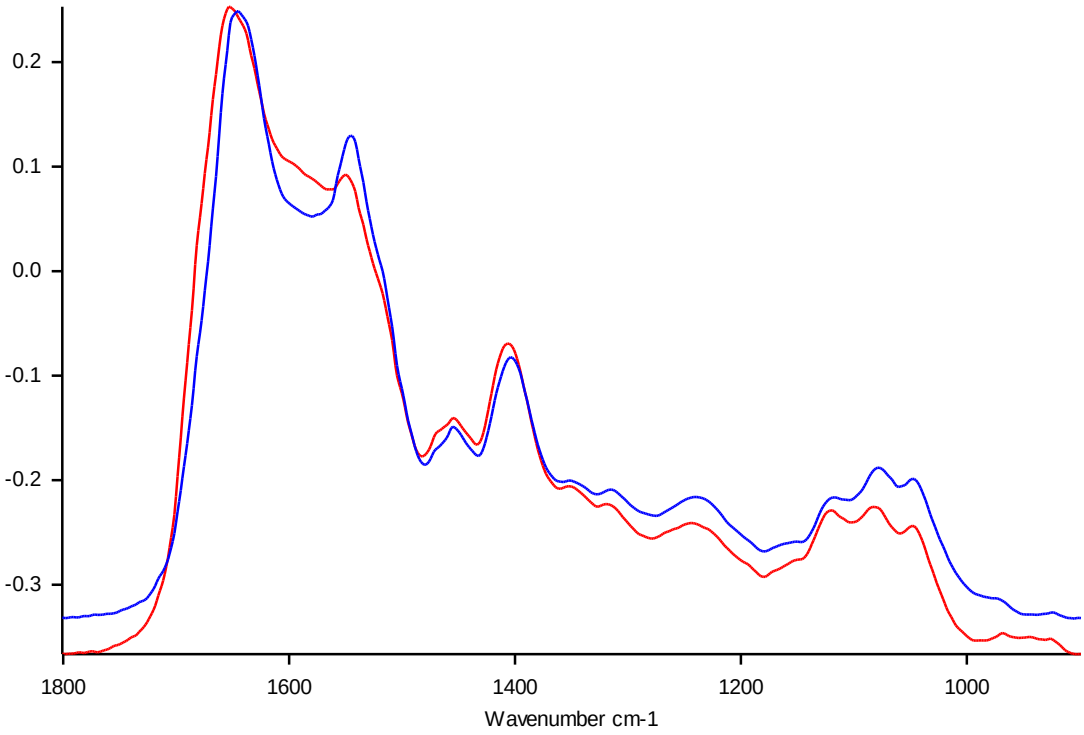


**Figure A4-35: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0066 dried onto the sampling strip (red), and the ATR crystal (blue)**

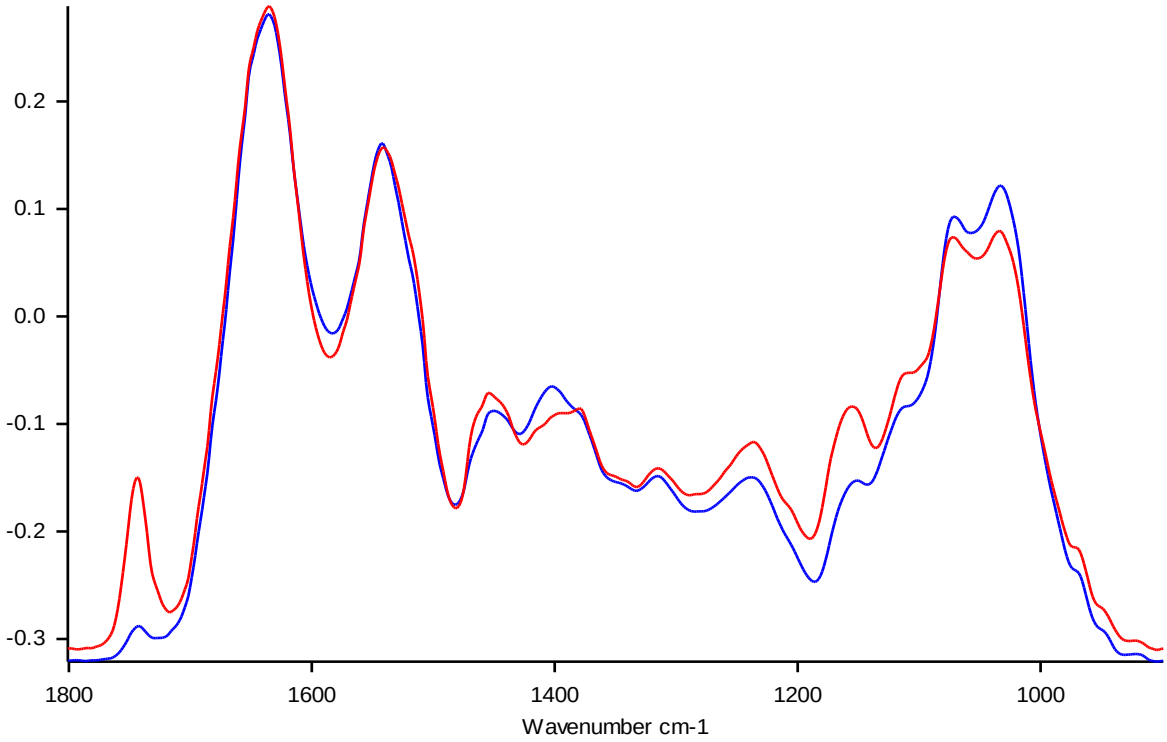


**Figure A4-36: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0069 dried onto the sampling strip (red), and the ATR crystal (blue)**

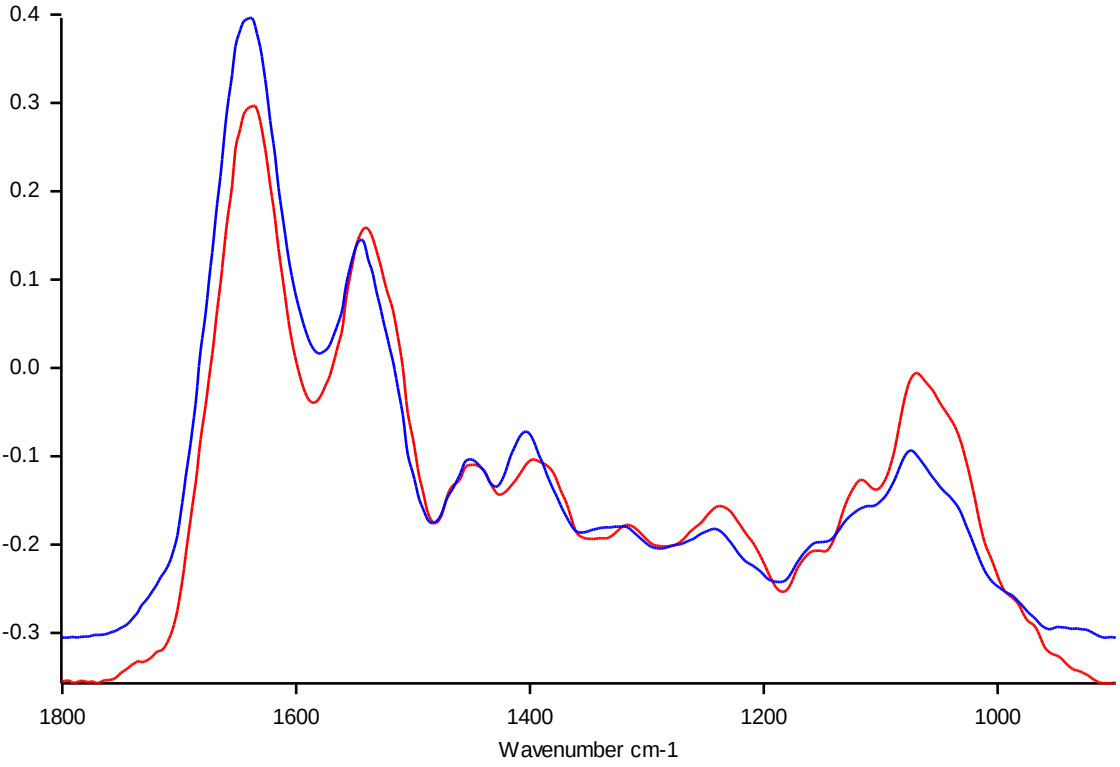


**Figure A4-37: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0071 dried onto the sampling strip (red), and the ATR crystal (blue)**

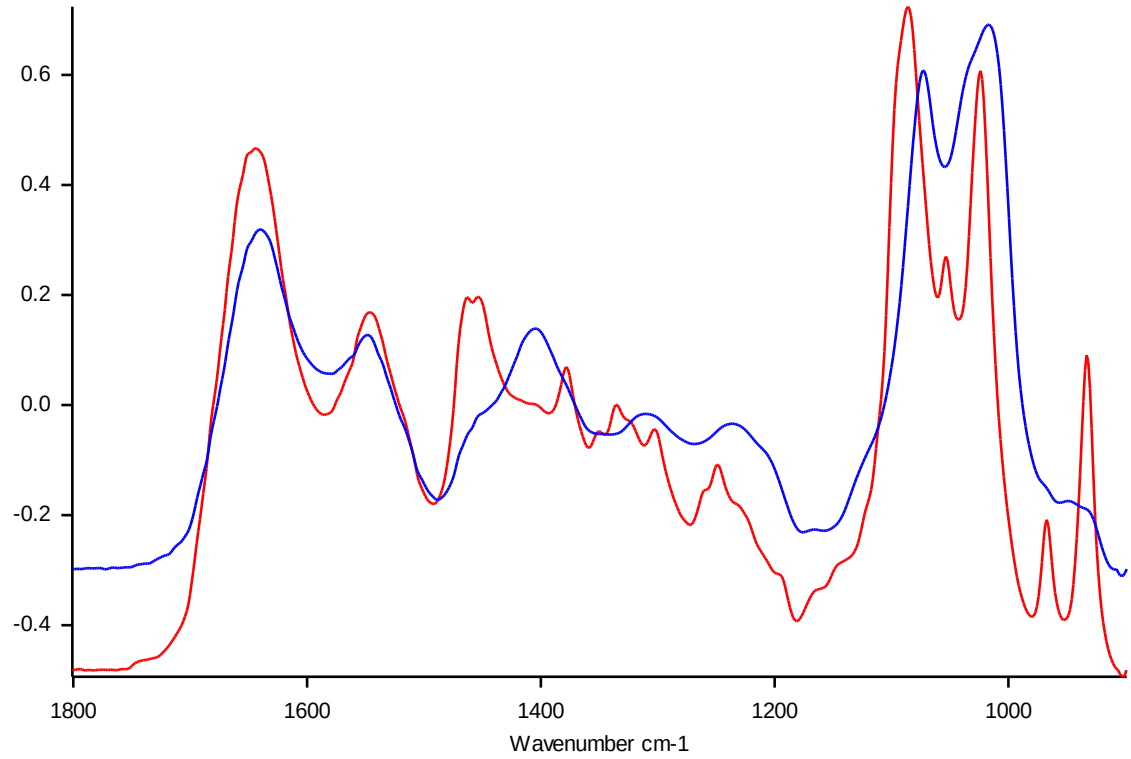


**Figure A4-38: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0072 dried onto the sampling strip (red), and the ATR crystal (blue)**

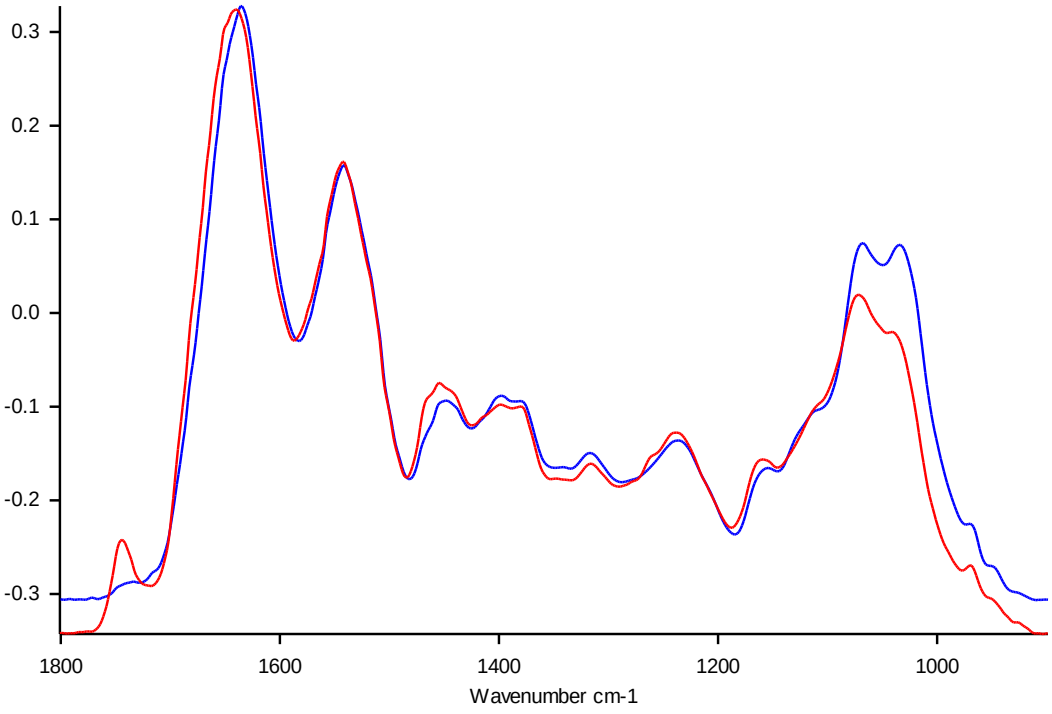


**Figure A4-39: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0074 dried onto the sampling strip (red), and the ATR crystal (blue)**

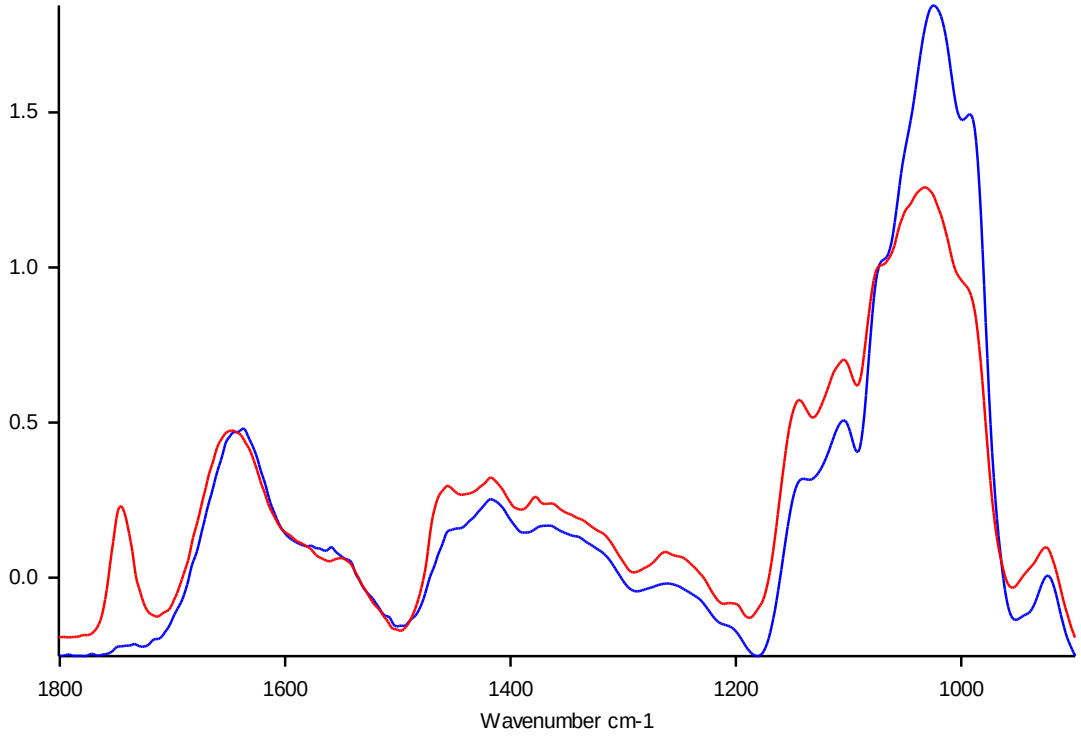


**Figure A4-40: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0075 dried onto the sampling strip (red), and the ATR crystal (blue)**

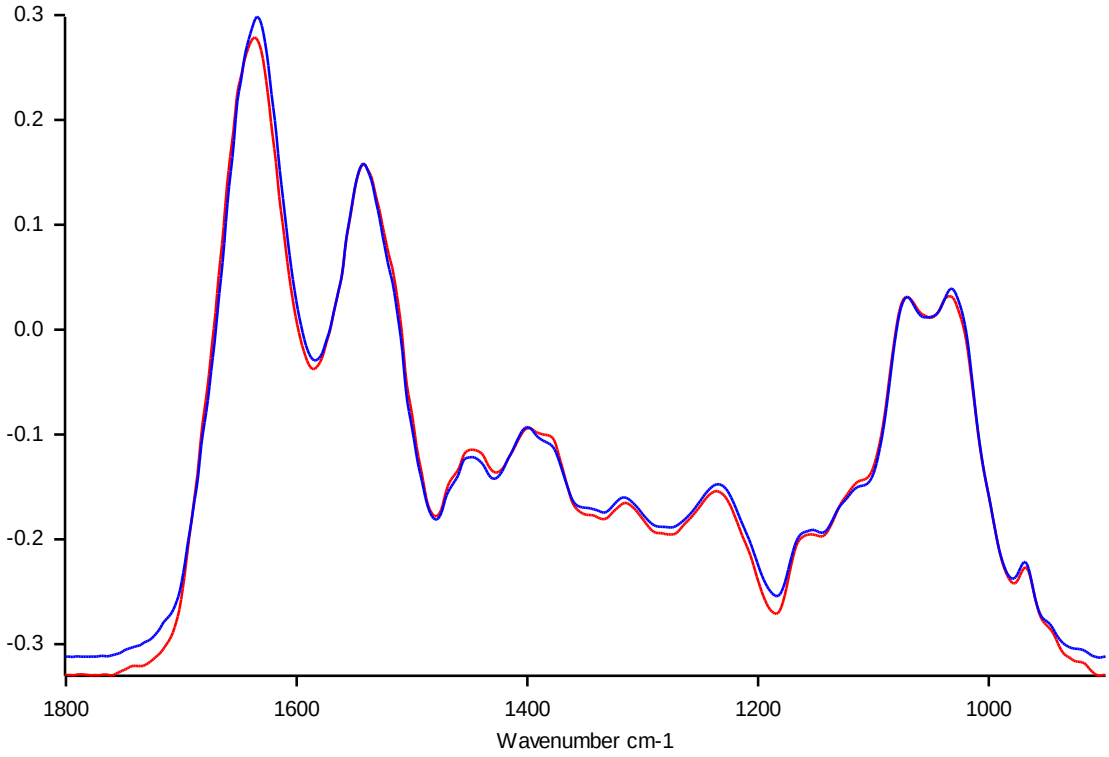


**Figure A4-41: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0077 dried onto the sampling strip (red), and the ATR crystal (blue)**

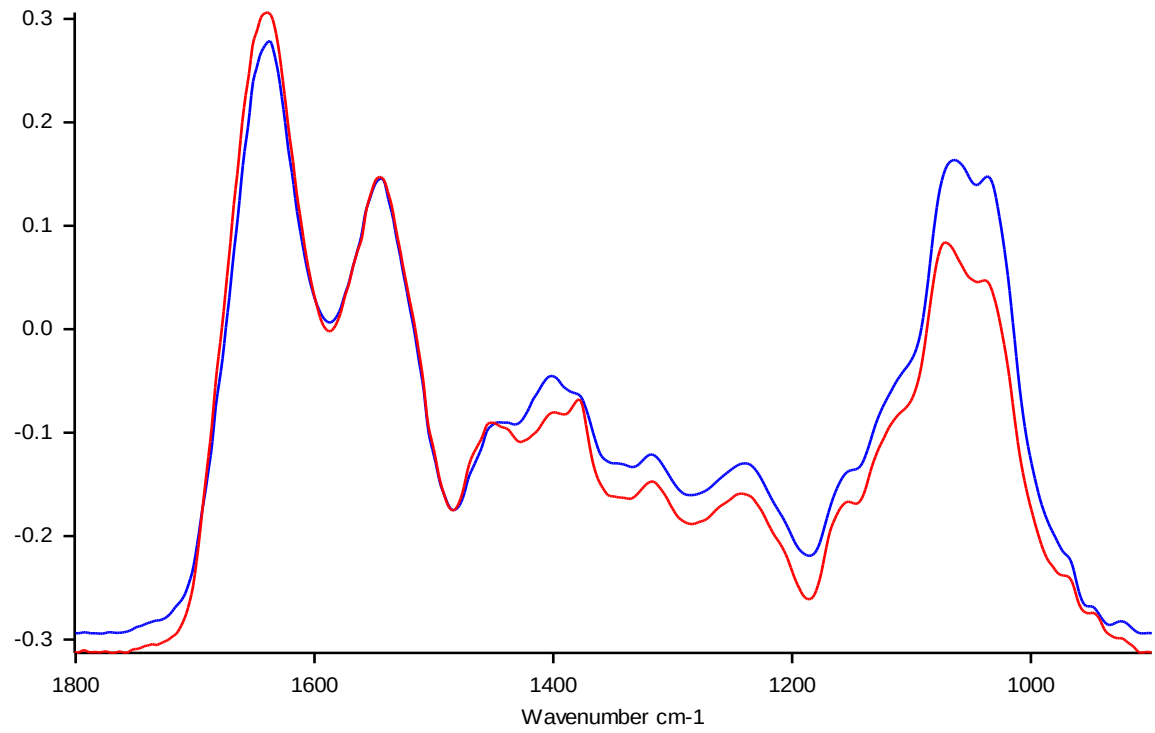


**Figure A4-42: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0078 dried onto the sampling strip (red), and the ATR crystal (blue)**

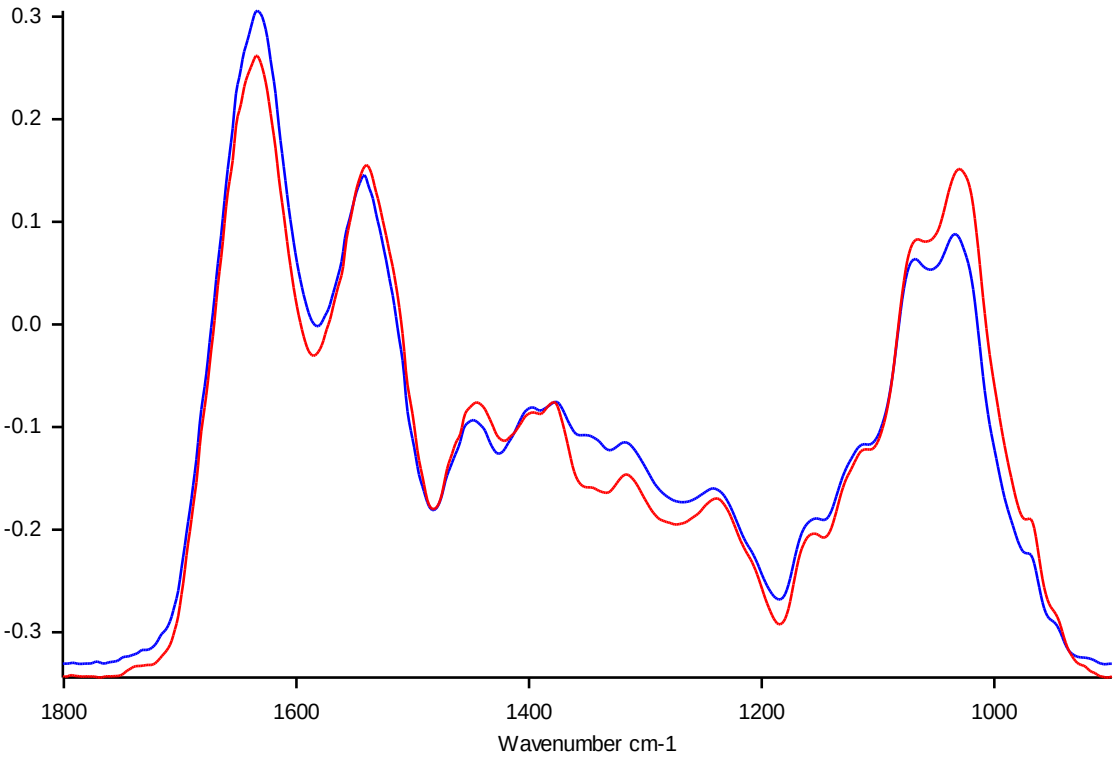


**Figure A4-43: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0079 dried onto the sampling strip (red), and the ATR crystal (blue)**

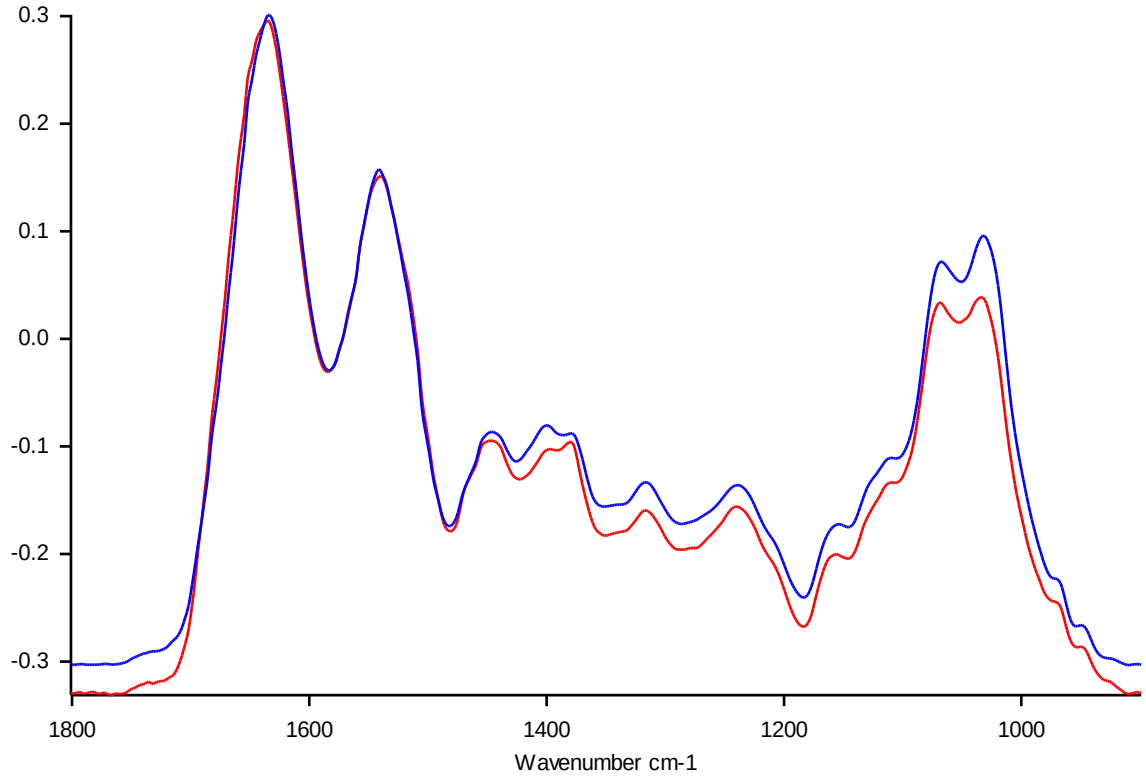


**Figure A4-44: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMNo080 dried onto the sampling strip (red), and the ATR crystal (blue)**

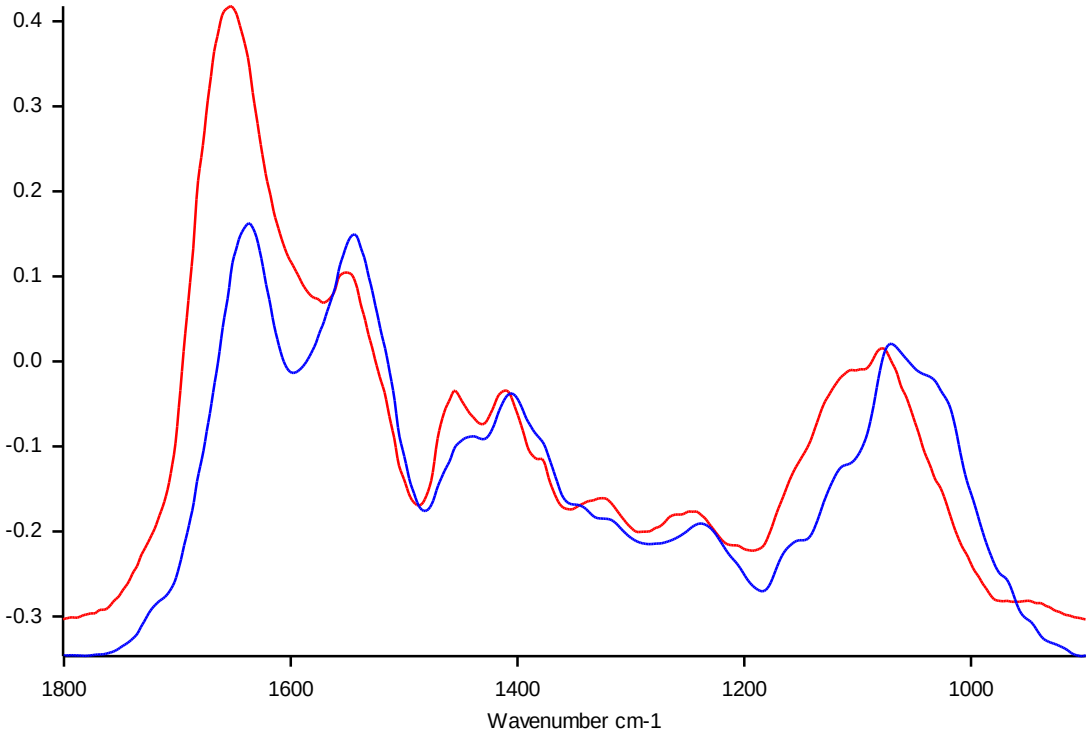


**Figure A4-45: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMNo081 dried onto the sampling strip (red), and the ATR crystal (blue)**

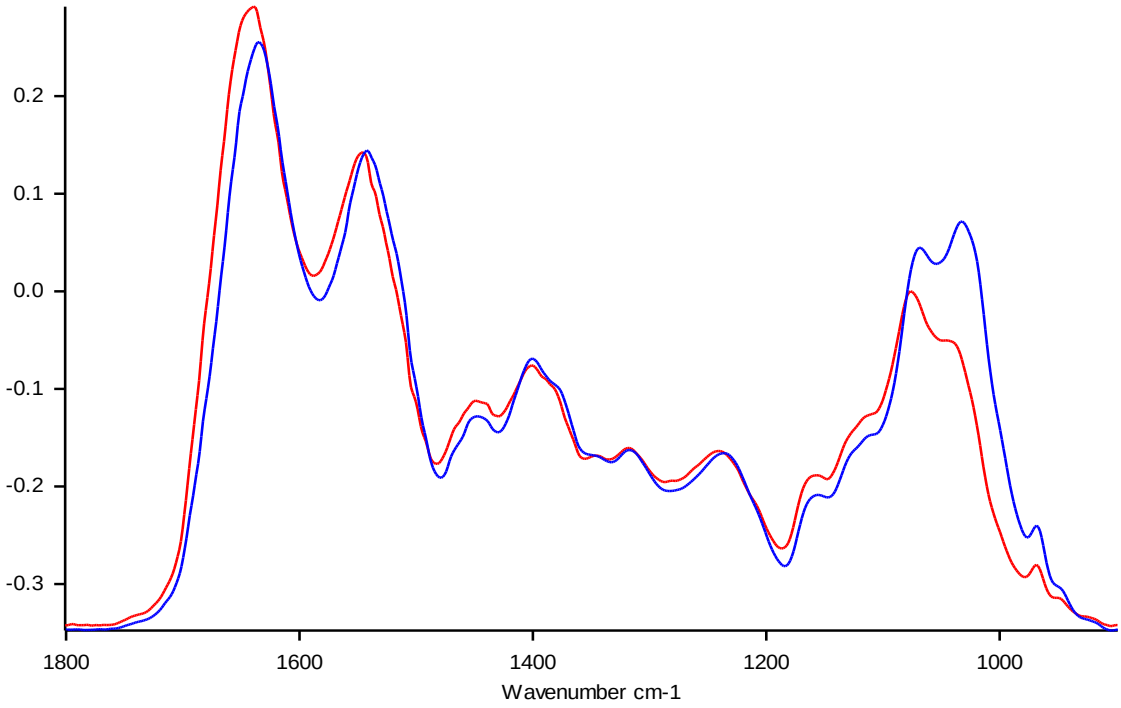


**Figure A4-46: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0084 dried onto the sampling strip (red), and the ATR crystal (blue)**

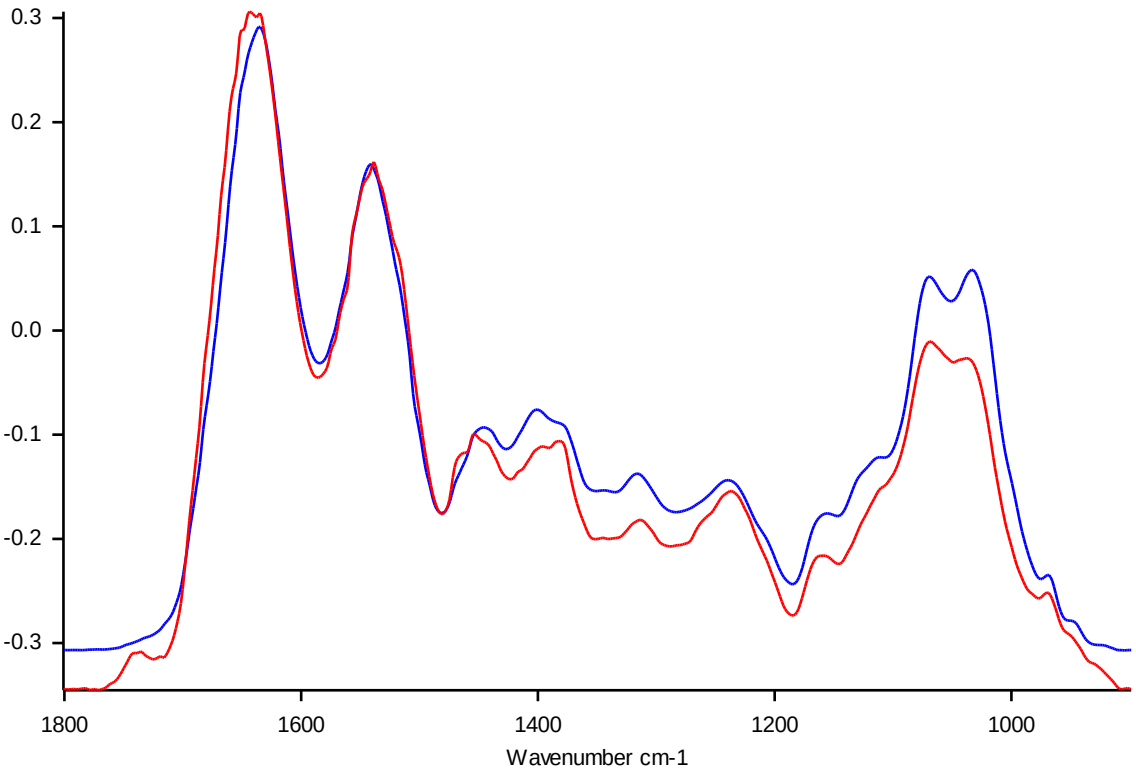


**Figure A4-47: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0085 dried onto the sampling strip (red), and the ATR crystal (blue)**

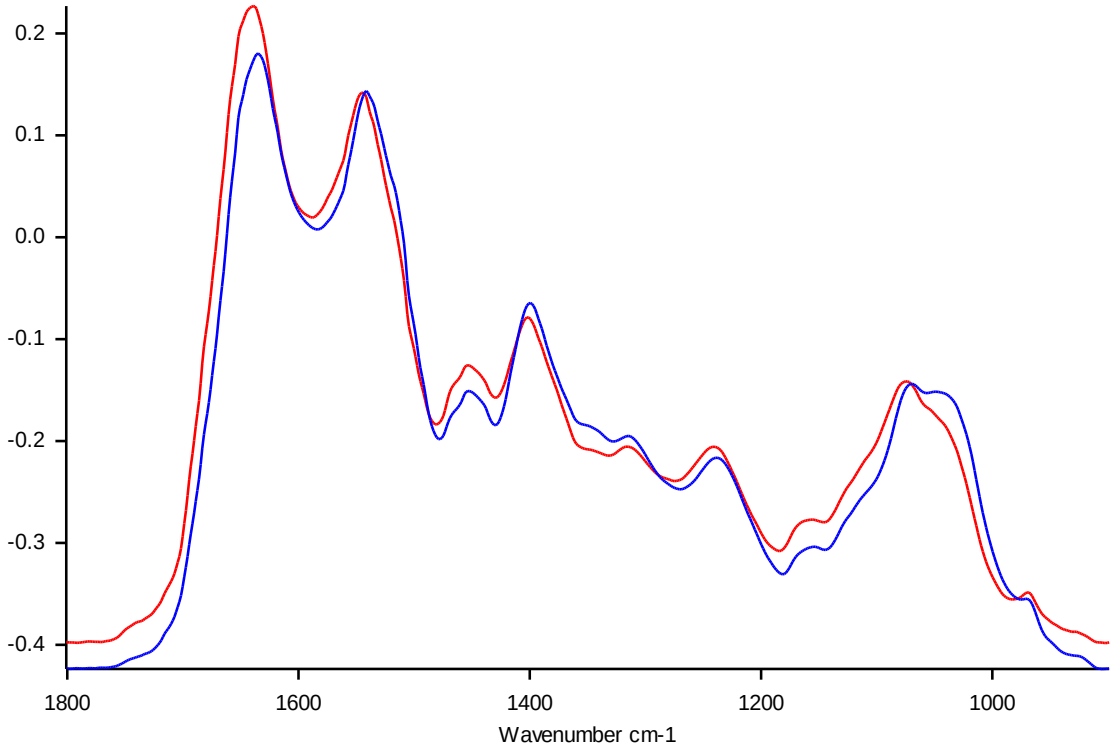


**Figure A4-48: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMNo086 dried onto the sampling strip (red), and the ATR crystal (blue)**

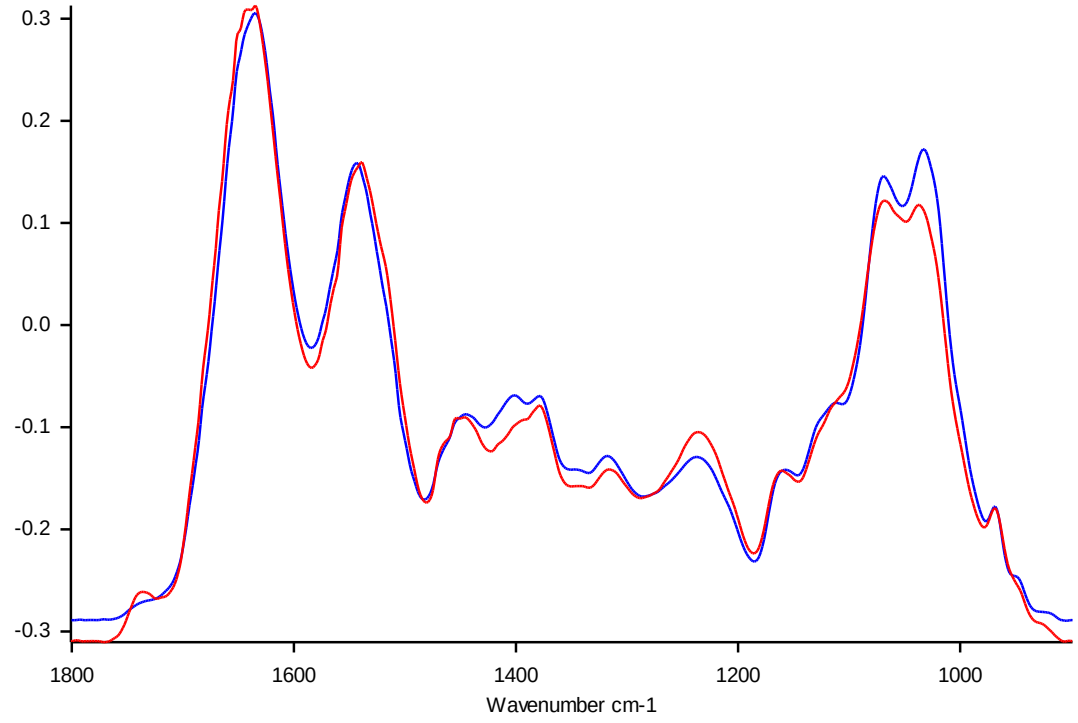


**Figure A4-49: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMNo088 dried onto the sampling strip (red), and the ATR crystal (blue)**

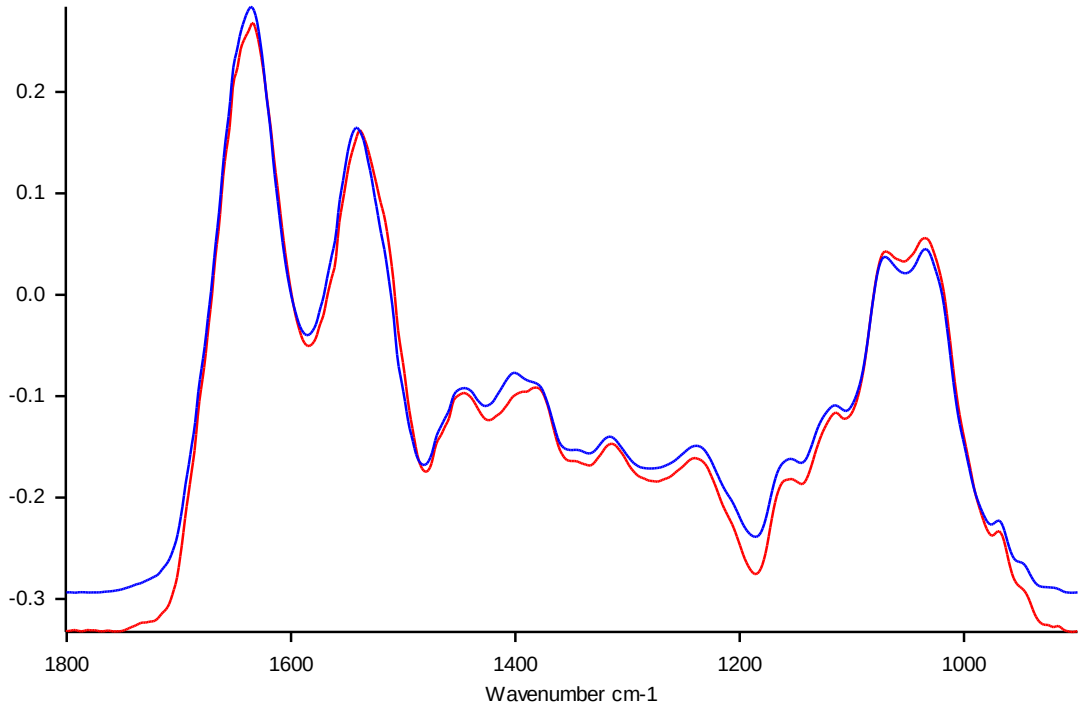


**Figure A4-50: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0091 dried onto the sampling strip (red), and the ATR crystal (blue)**

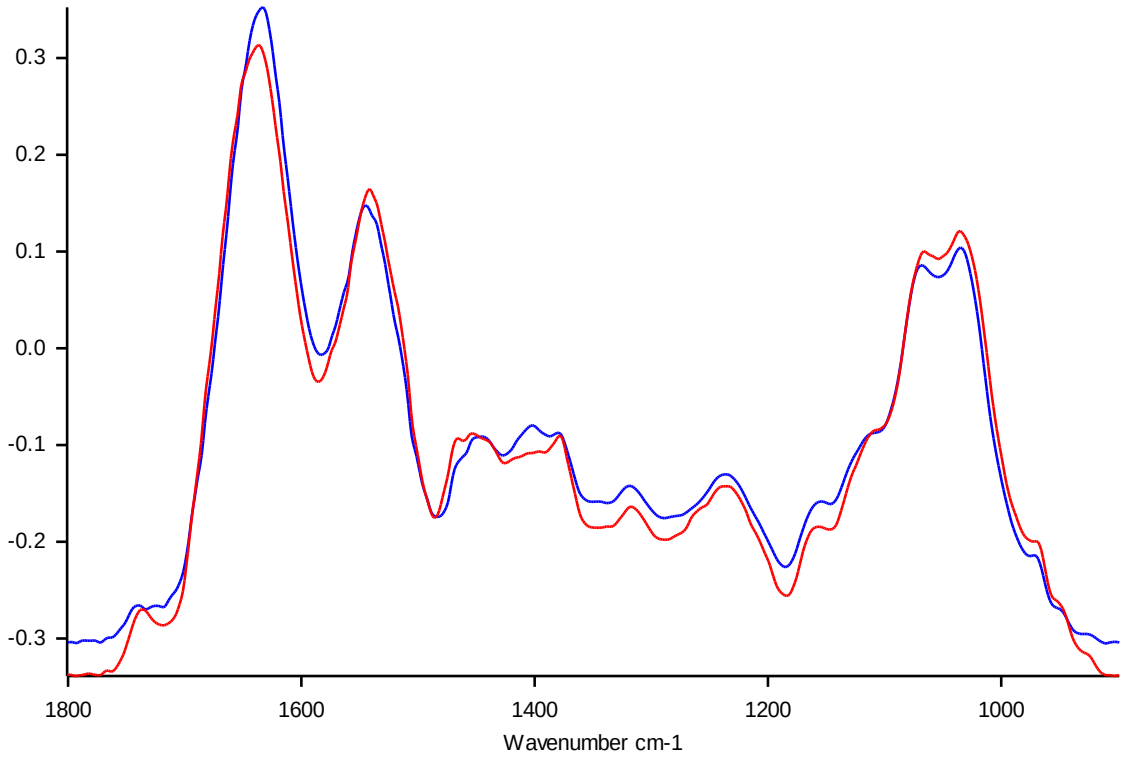


**Figure A4-51: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0093 dried onto the sampling strip (red), and the ATR crystal (blue)**


**Figure A4-52: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0094 dried onto the sampling strip (red), and the ATR crystal (blue)**

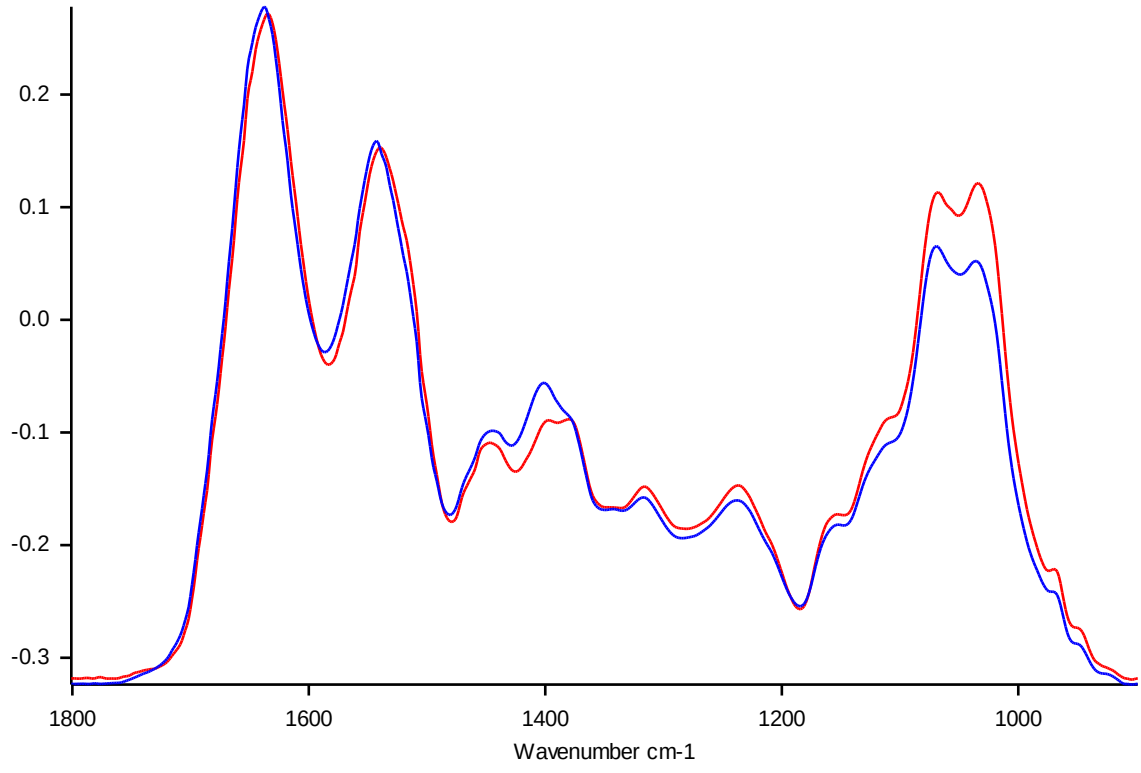


**Figure A4-53: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0095 dried onto the sampling strip (red), and the ATR crystal (blue)**

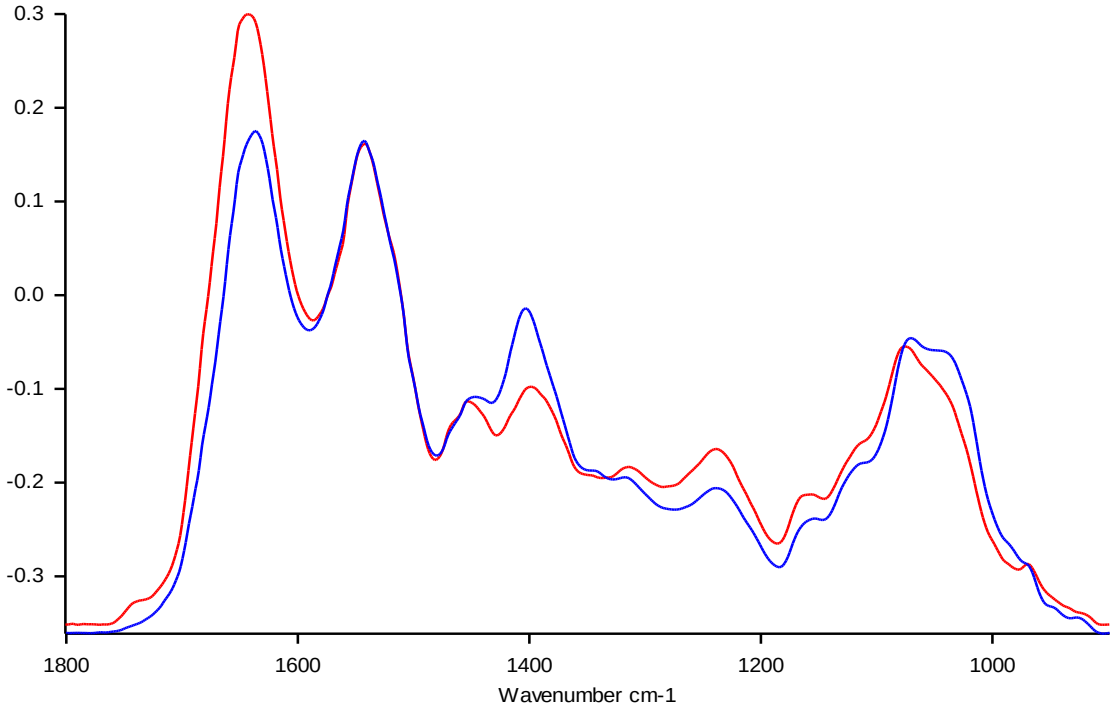


**Figure A4-54: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0097 dried onto the sampling strip (red), and the ATR crystal (blue)**

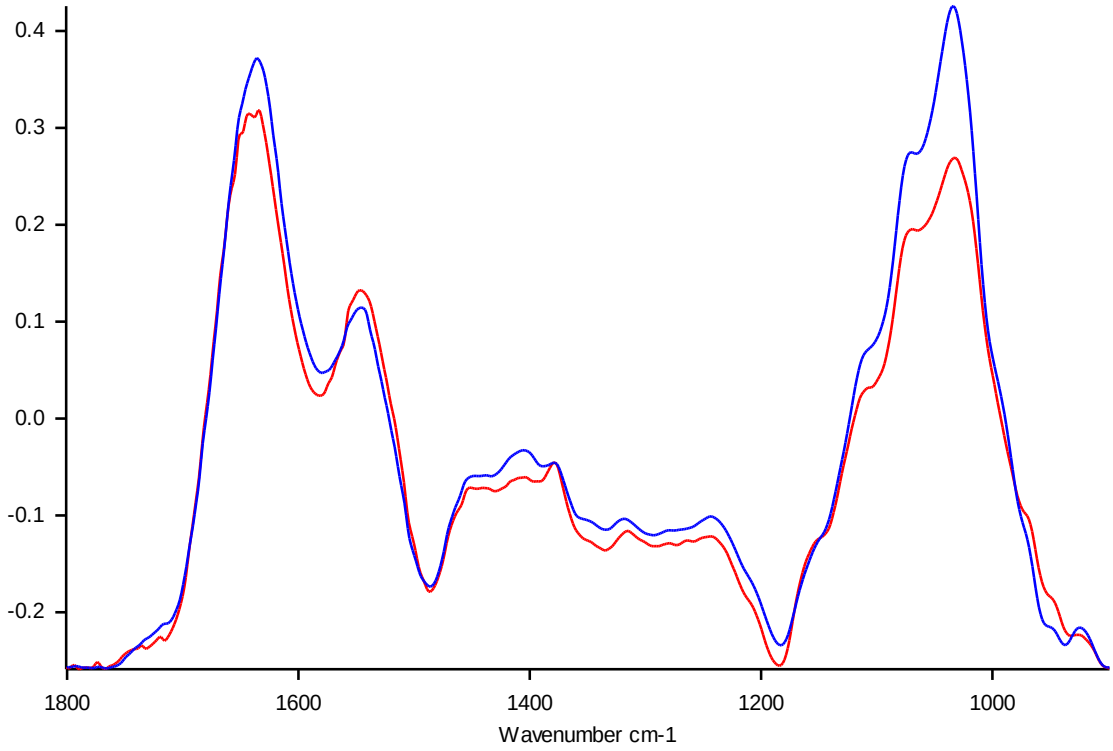


**Figure A4-55: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0099 dried onto the sampling strip (red), and the ATR crystal (blue)**


**Figure A4-56: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0102 dried onto the sampling strip (red), and the ATR crystal (blue)**

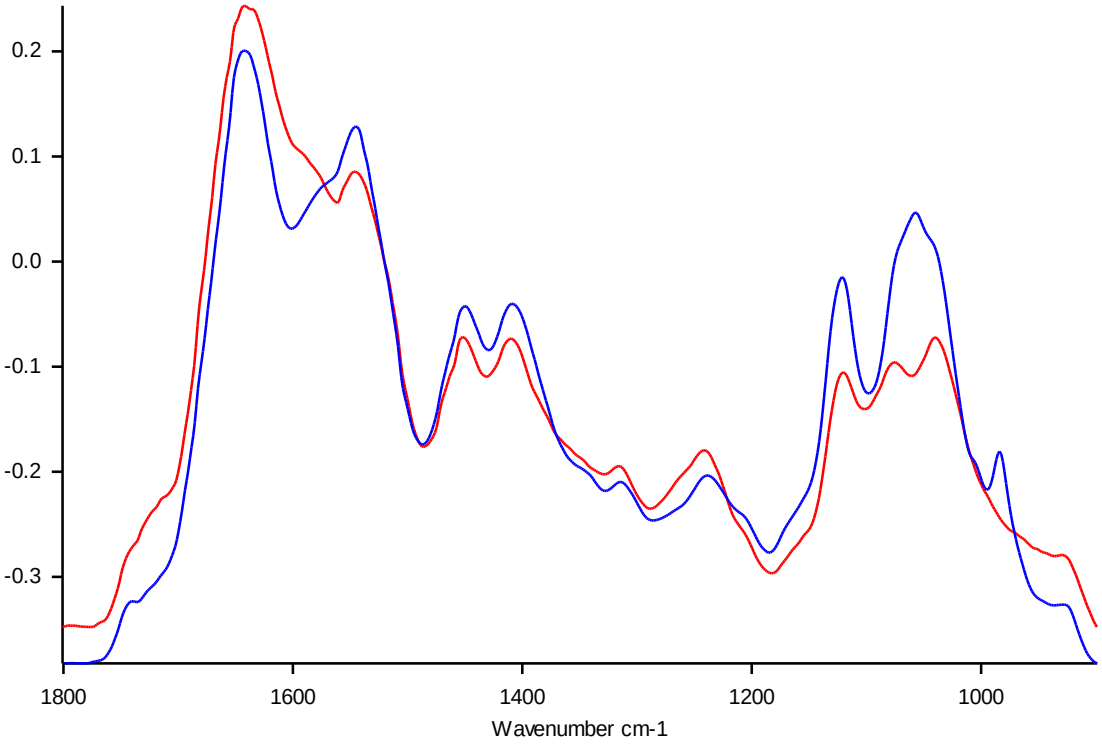


**Figure A4-57: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0103 dried onto the sampling strip (red), and the ATR crystal (blue)**

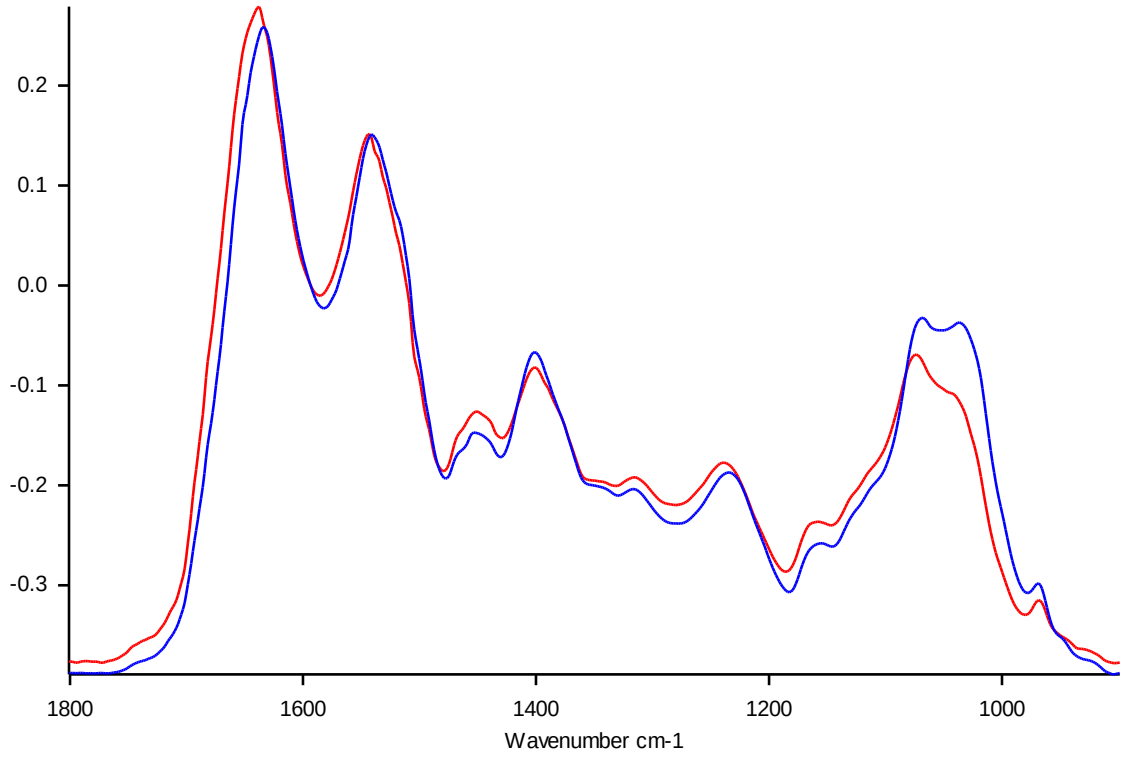


**Figure A4-58: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0104 dried onto the sampling strip (red), and the ATR crystal (blue)**

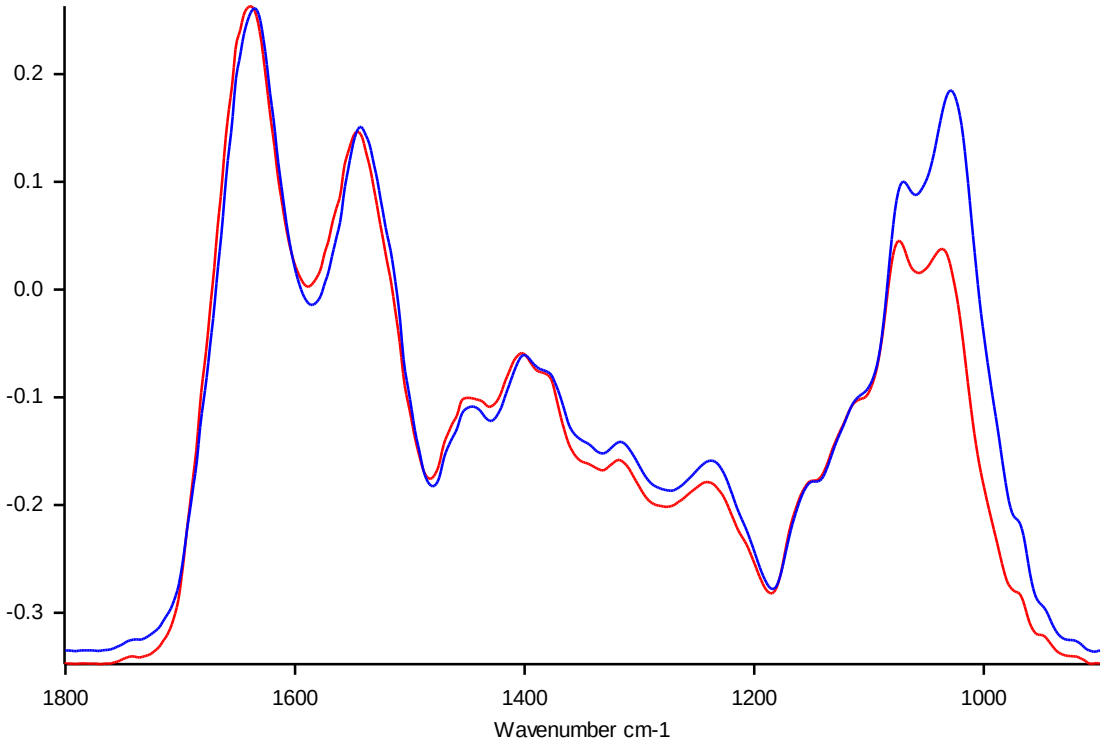


**Figure A4-59: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0105 dried onto the sampling strip (red), and the ATR crystal (blue)**

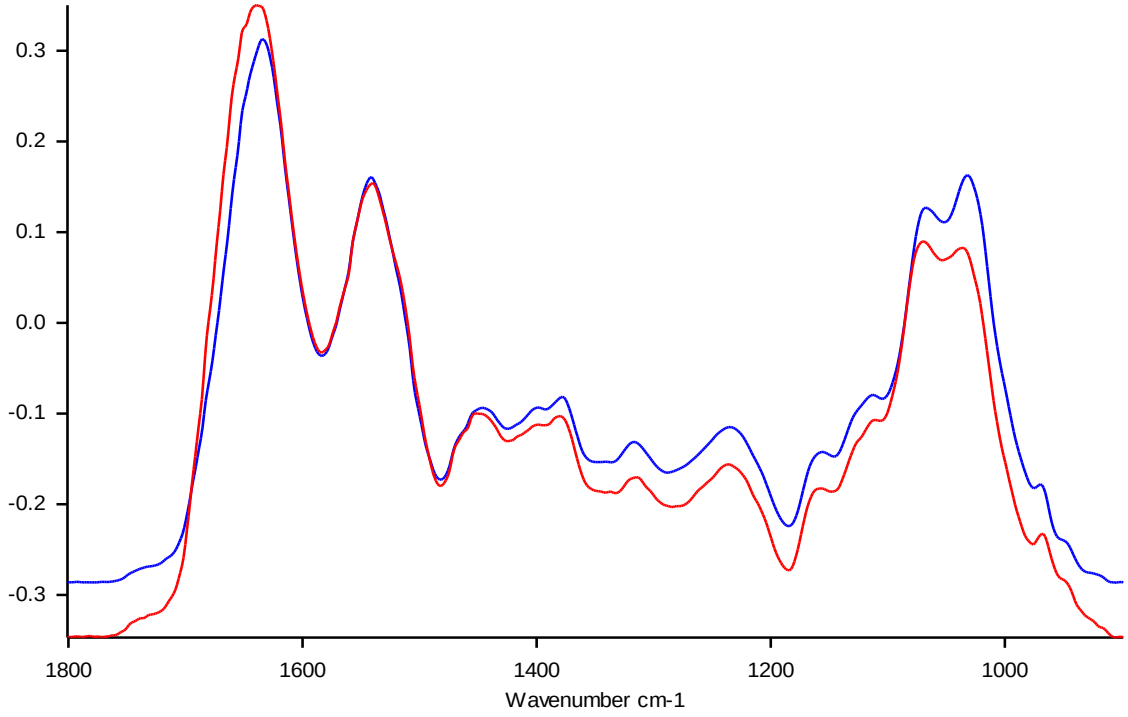


**Figure A4-60: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0106 dried onto the sampling strip (red), and the ATR crystal (blue)**

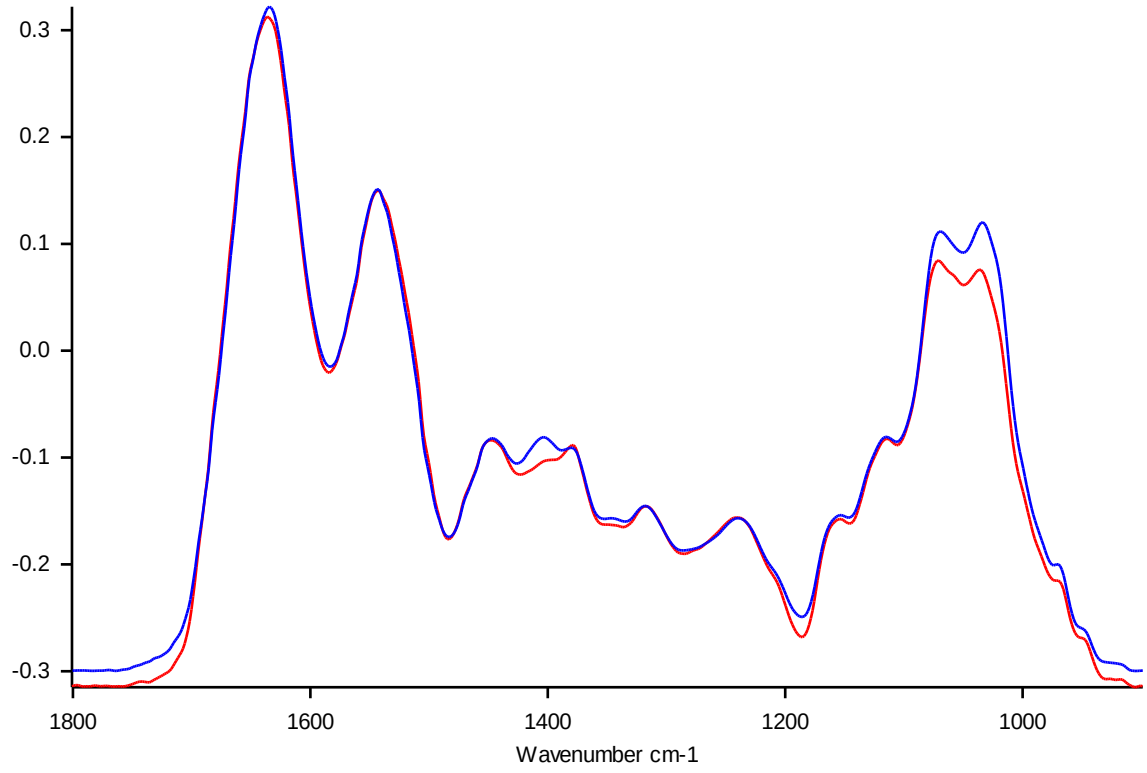


**Figure A4-61: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0110 dried onto the sampling strip (red), and the ATR crystal (blue)**

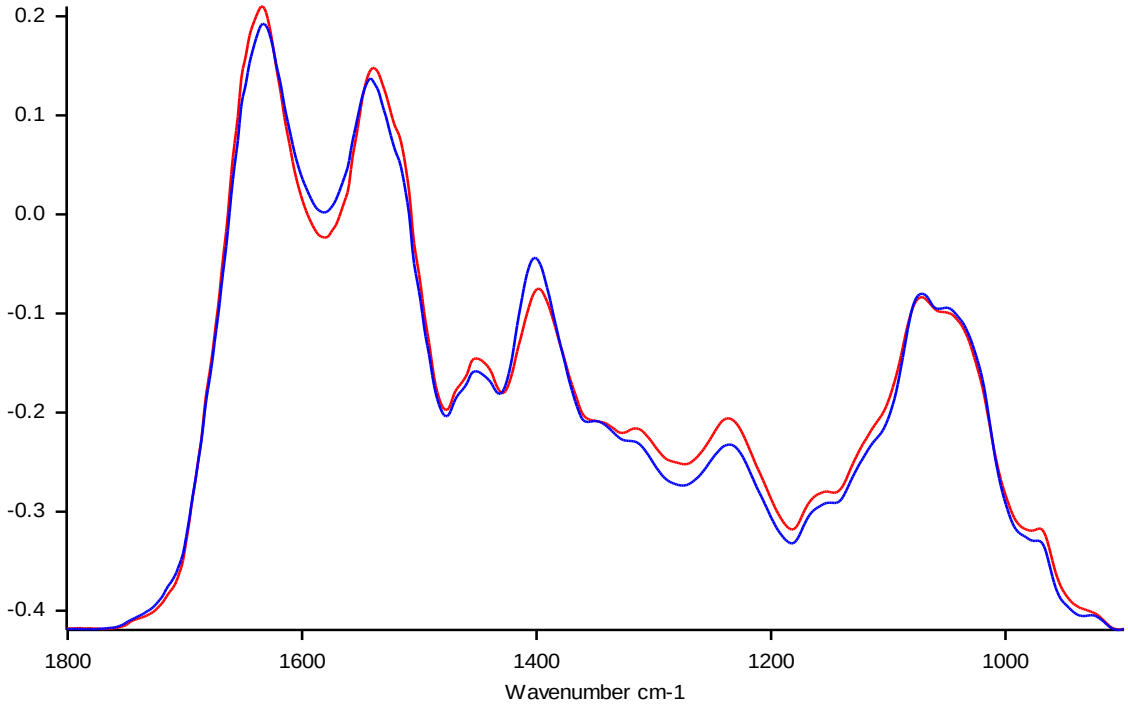


**Figure A4-62: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0111 dried onto the sampling strip (red), and the ATR crystal (blue)**

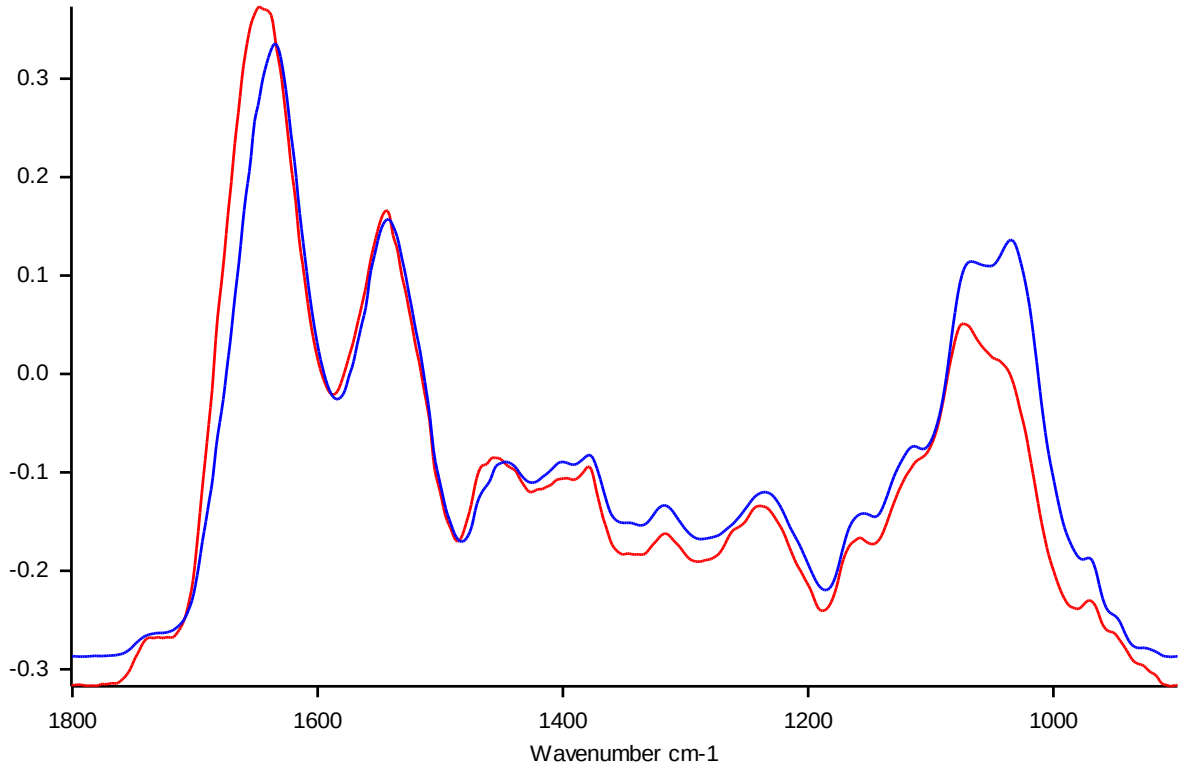


**Figure A4-63: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0112 dried onto the sampling strip (red), and the ATR crystal (blue)**

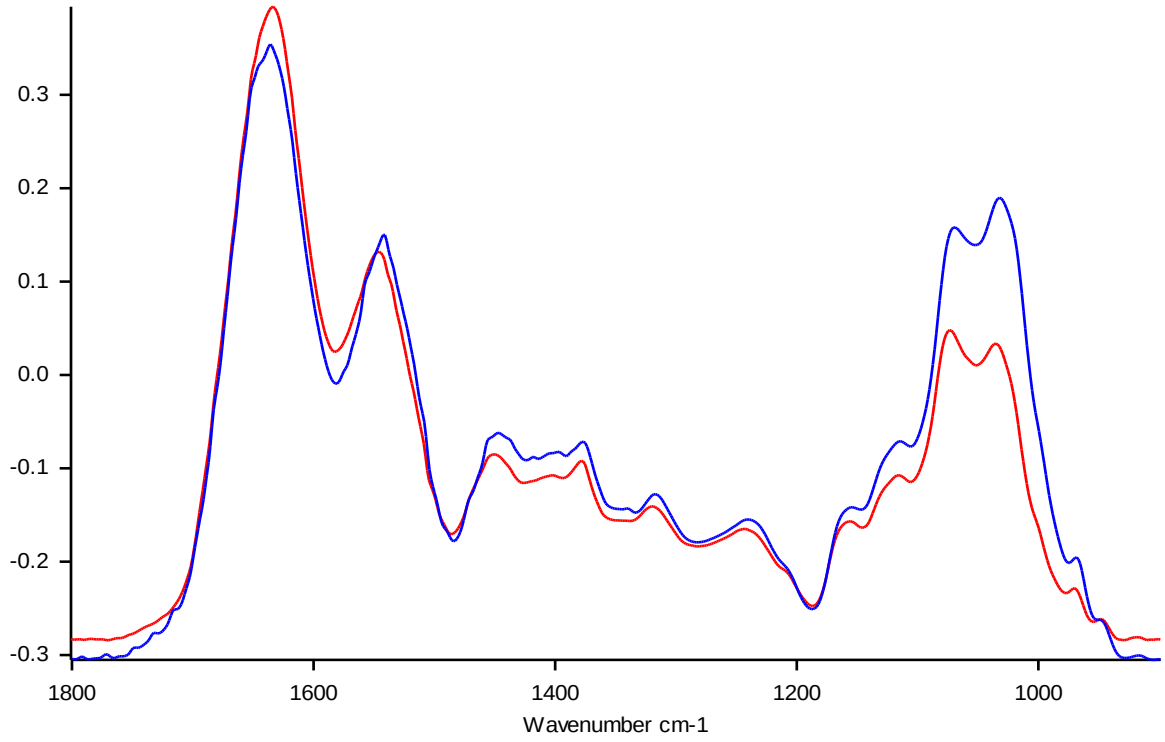


**Figure A4-64: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0116 dried onto the sampling strip (red), and the ATR crystal (blue)**

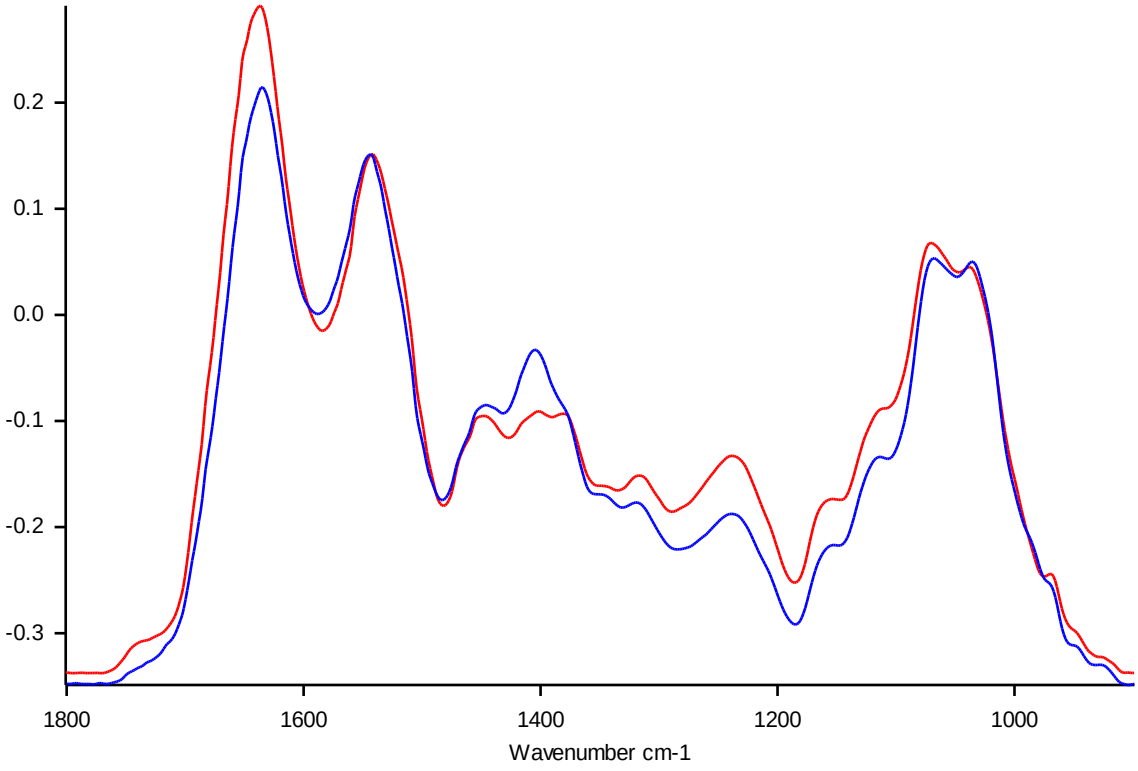


**Figure A4-65: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0117 dried onto the sampling strip (red), and the ATR crystal (blue)**

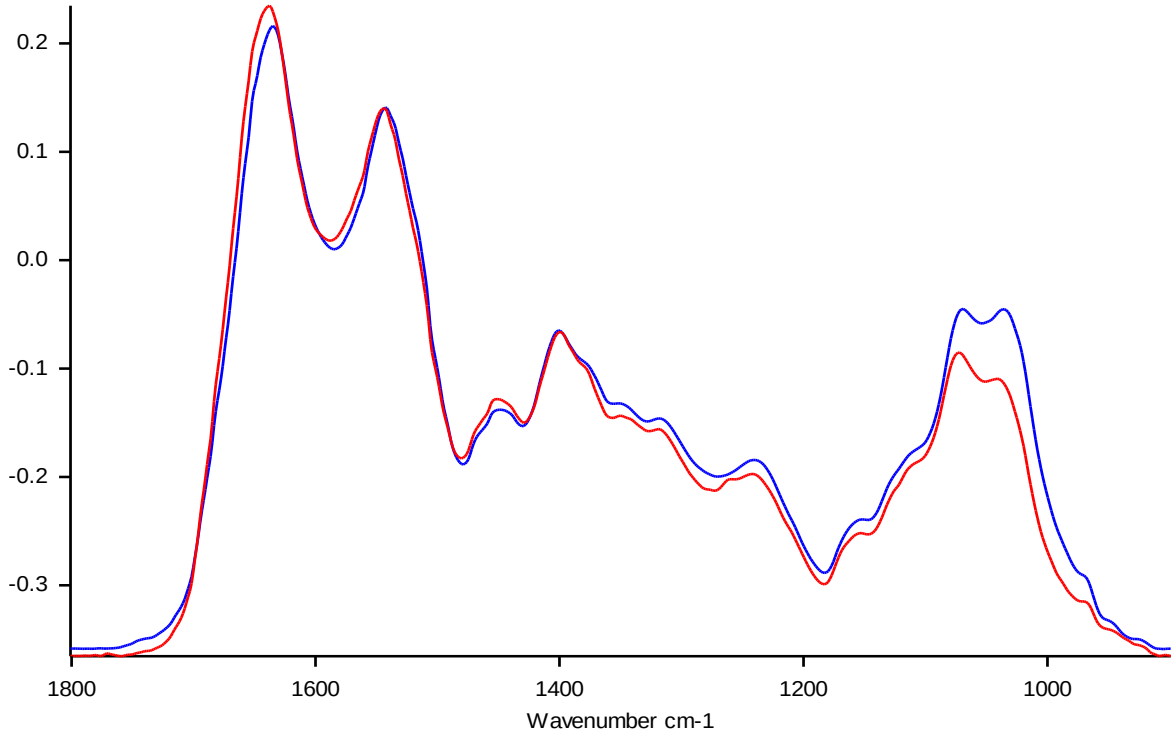


**Figure A4-66: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0118 dried onto the sampling strip (red), and the ATR crystal (blue)**

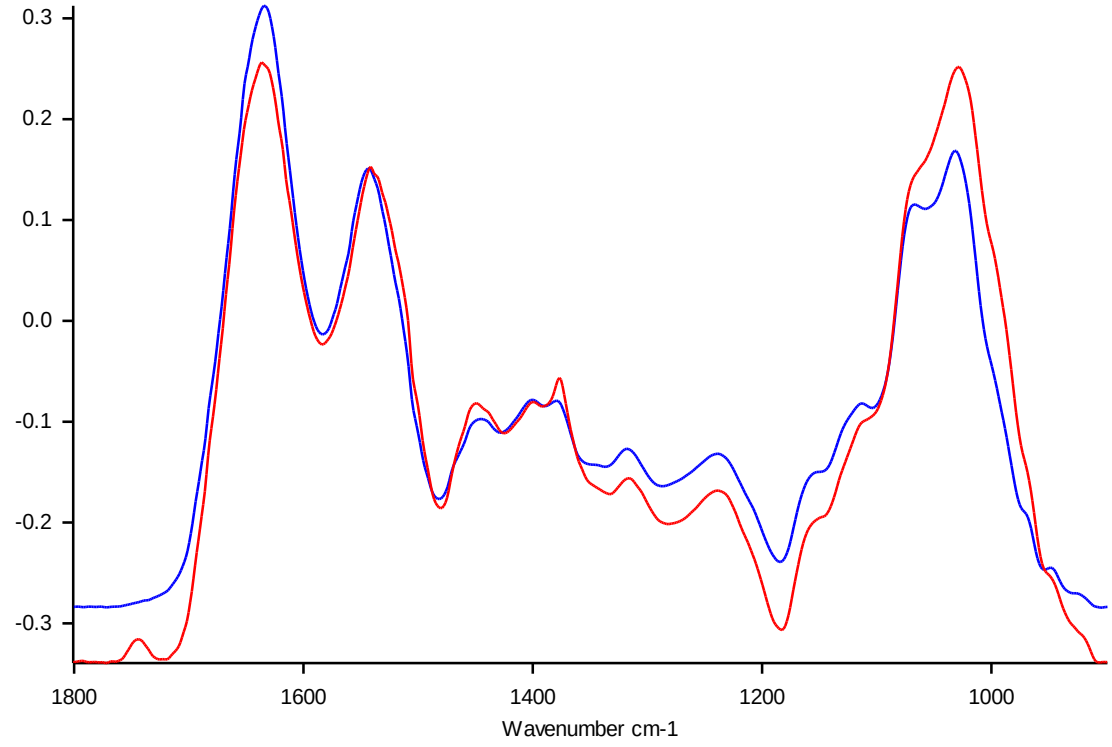


**Figure A4-67: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample KMN0119 dried onto the sampling strip (red), and the ATR crystal (blue)**

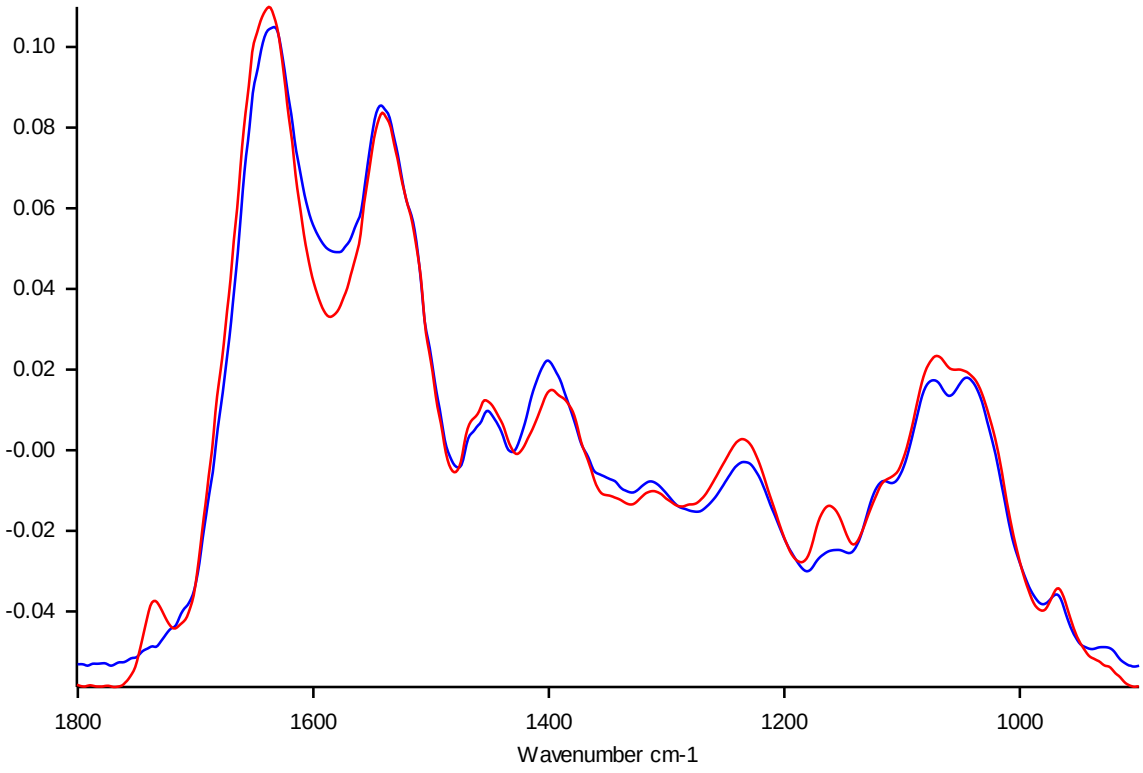


**Figure A4-68: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL01 dried onto the sampling strip (red), and the ATR crystal (blue)**

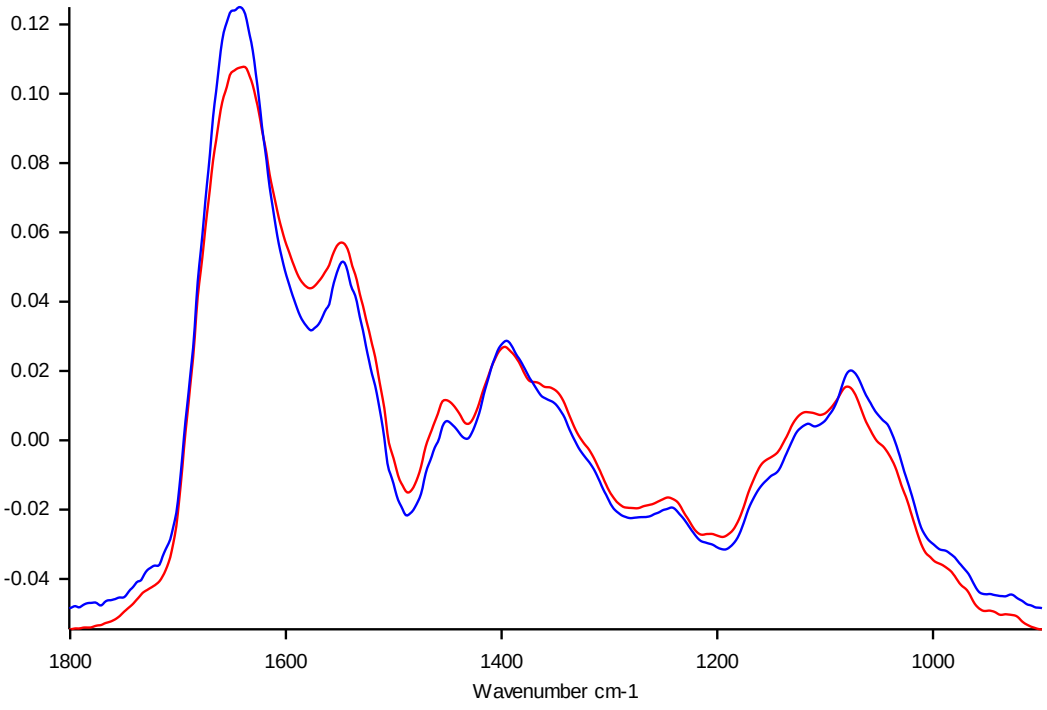


**Figure A4-69: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL02 dried onto the sampling strip (red), and the ATR crystal (blue)**

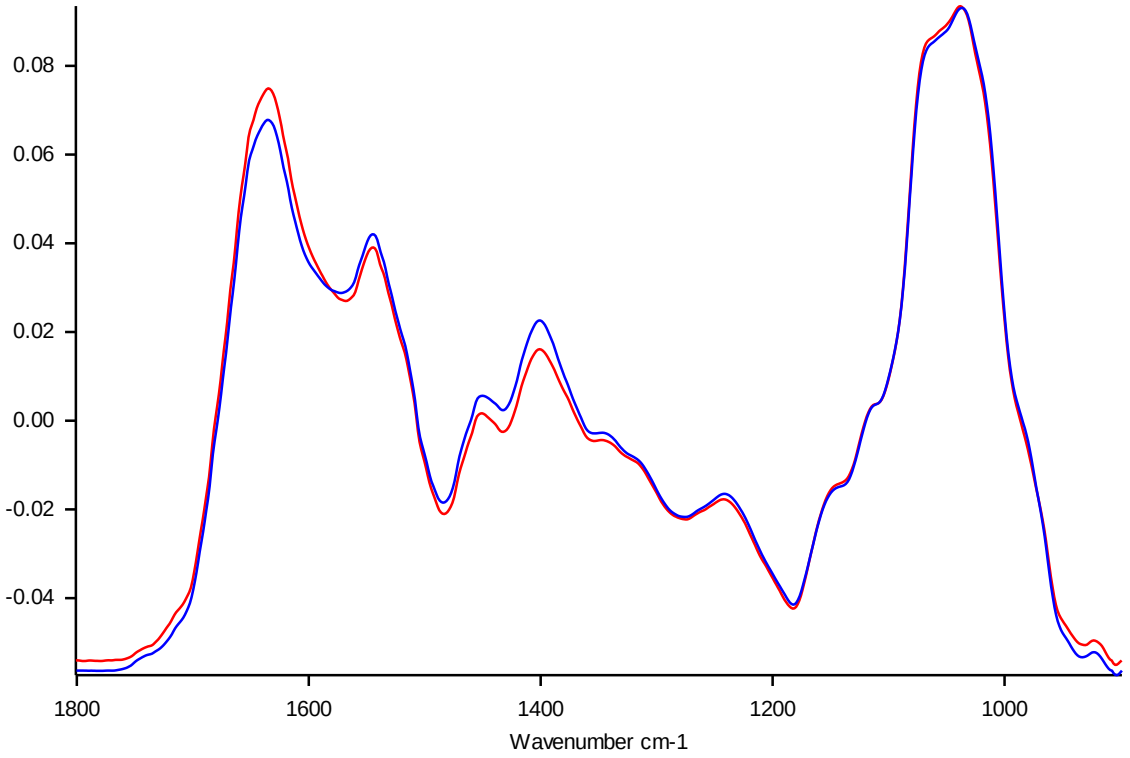


**Figure A4-70: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL03 dried onto the sampling strip (red), and the ATR crystal (blue)**

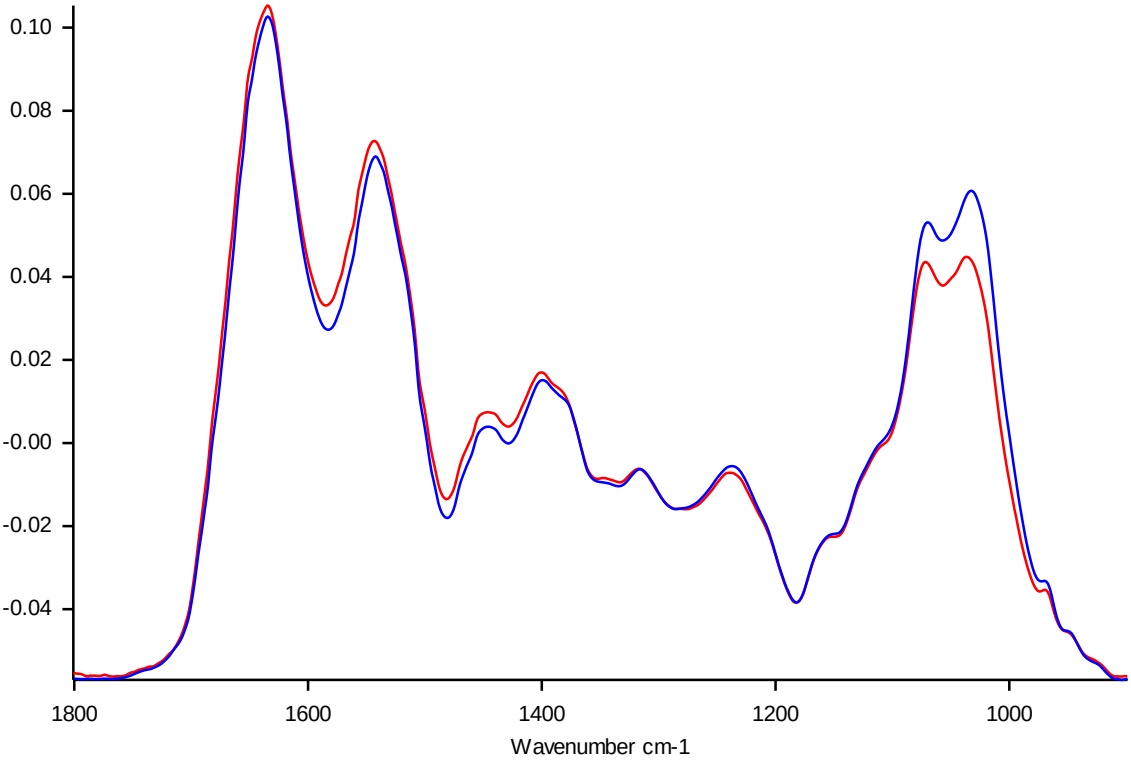


**Figure A4-71: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL04 dried onto the sampling strip (red), and the ATR crystal (blue)**

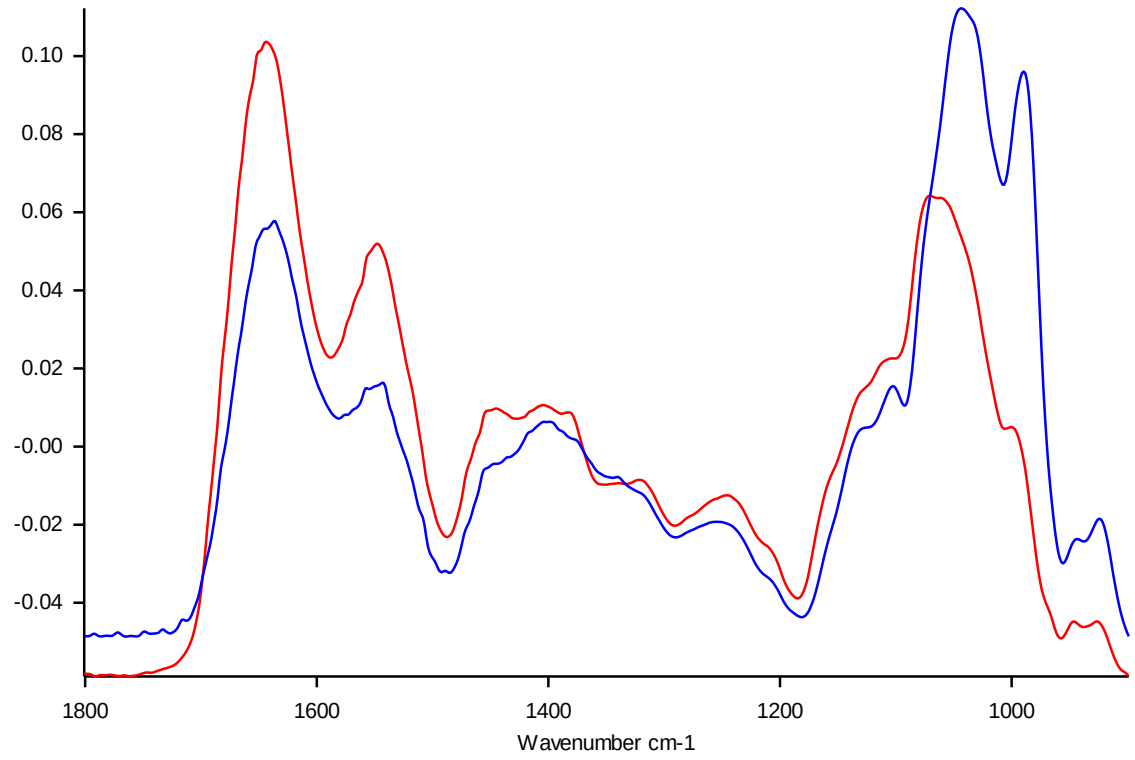


**Figure A4-72: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL05 dried onto the sampling strip (red), and the ATR crystal (blue)**

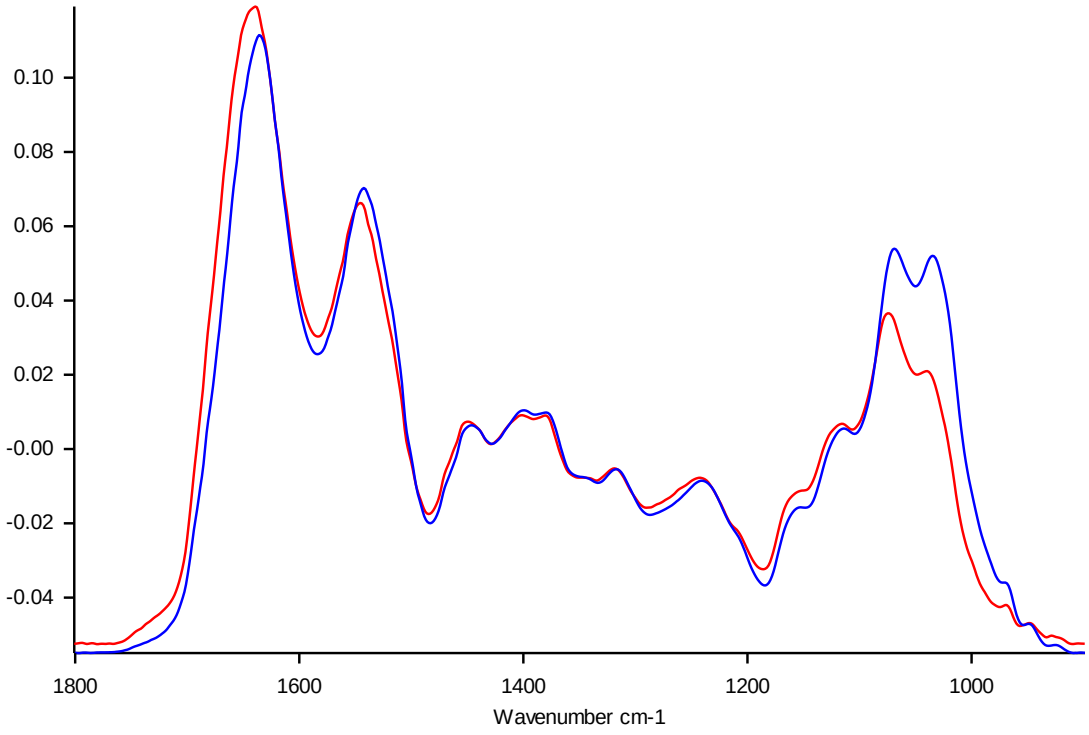


**Figure A4-73: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL06 dried onto the sampling strip (red), and the ATR crystal (blue)**

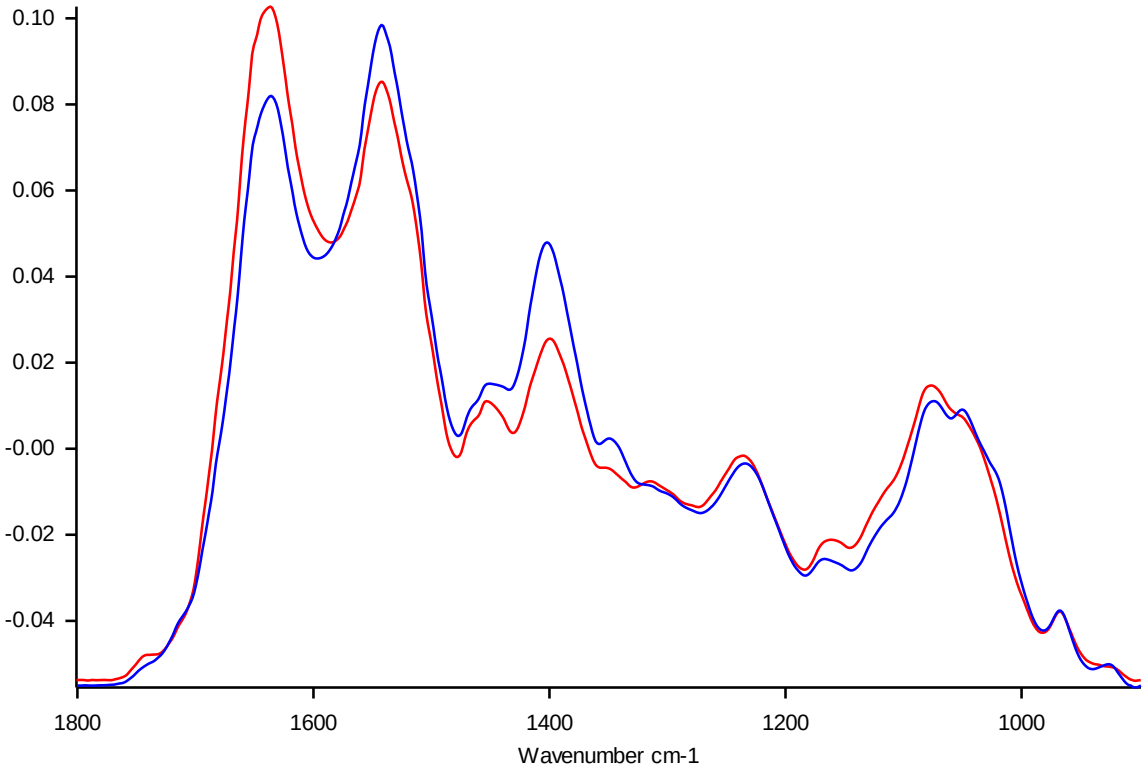


**Figure A4-74: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL08 dried onto the sampling strip (red), and the ATR crystal (blue)**

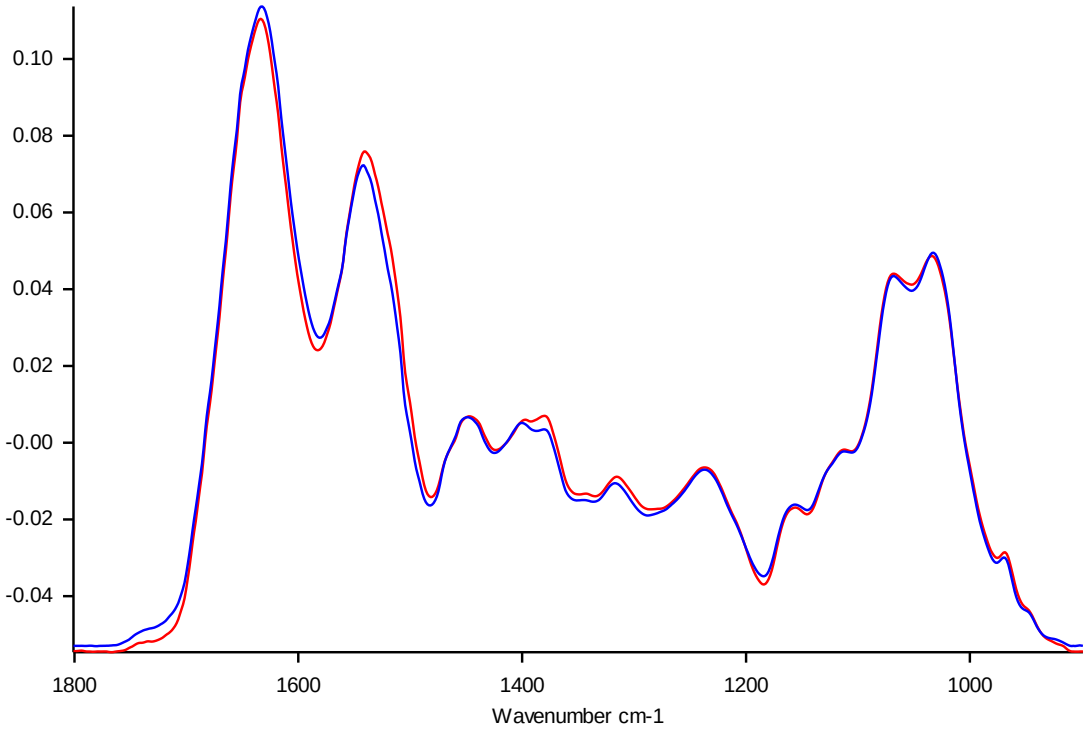


**Figure A4-75: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL09 dried onto the sampling strip (red), and the ATR crystal (blue)**

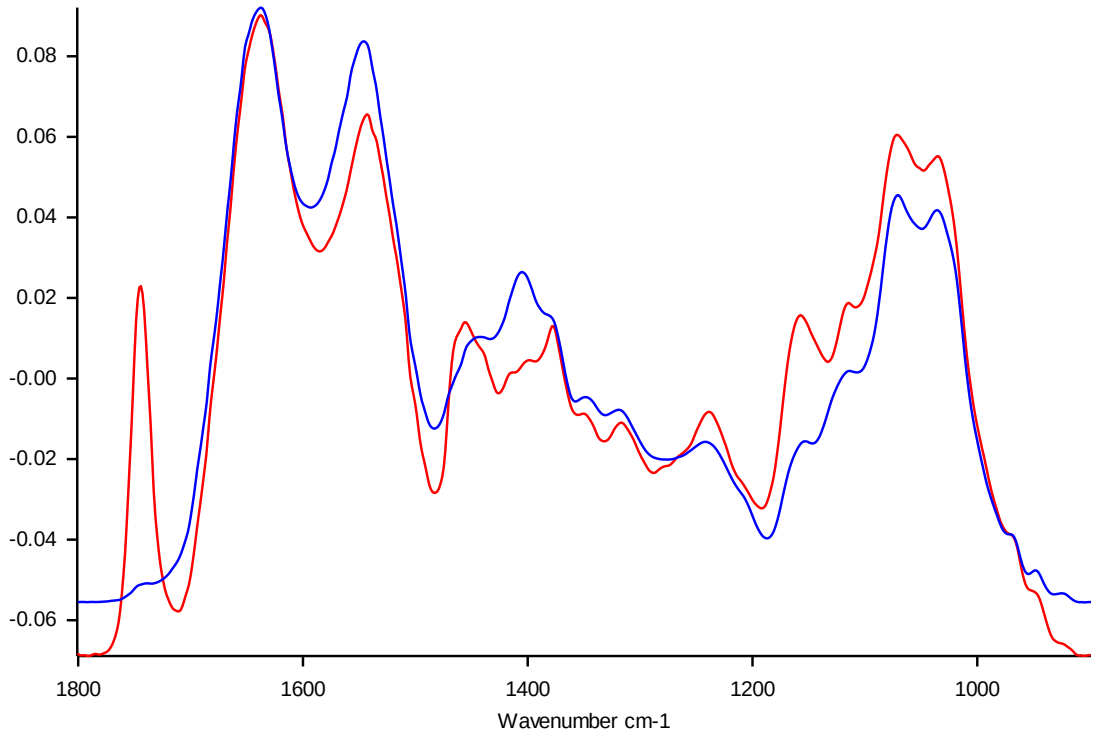


**Figure A4-76: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL10 dried onto the sampling strip (red), and the ATR crystal (blue)**

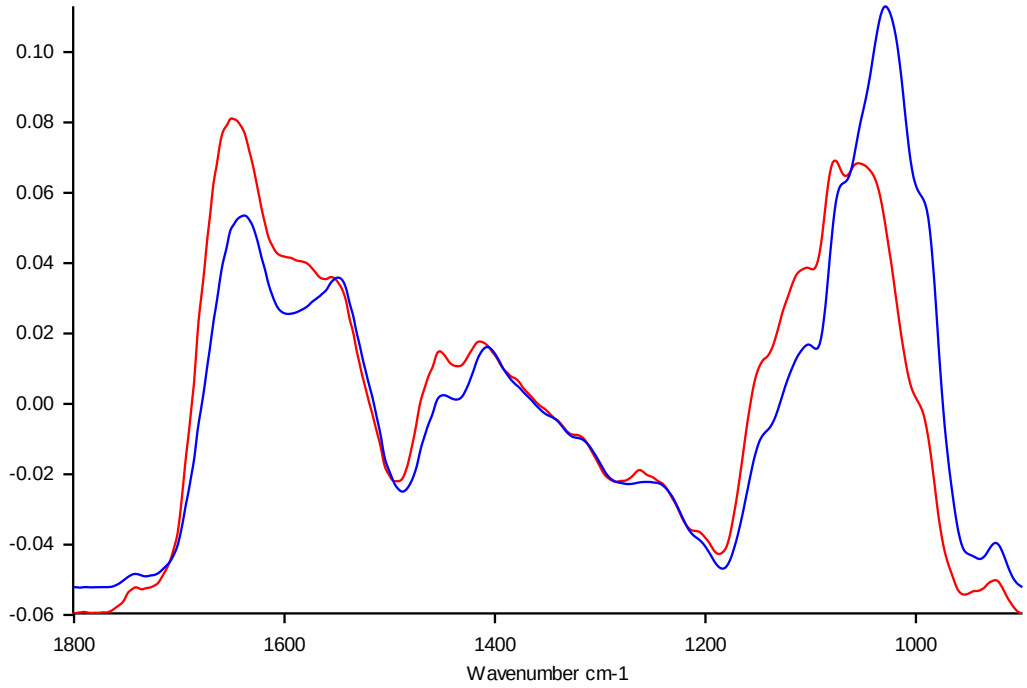


**Figure A4-77: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL11 dried onto the sampling strip (red), and the ATR crystal (blue)**

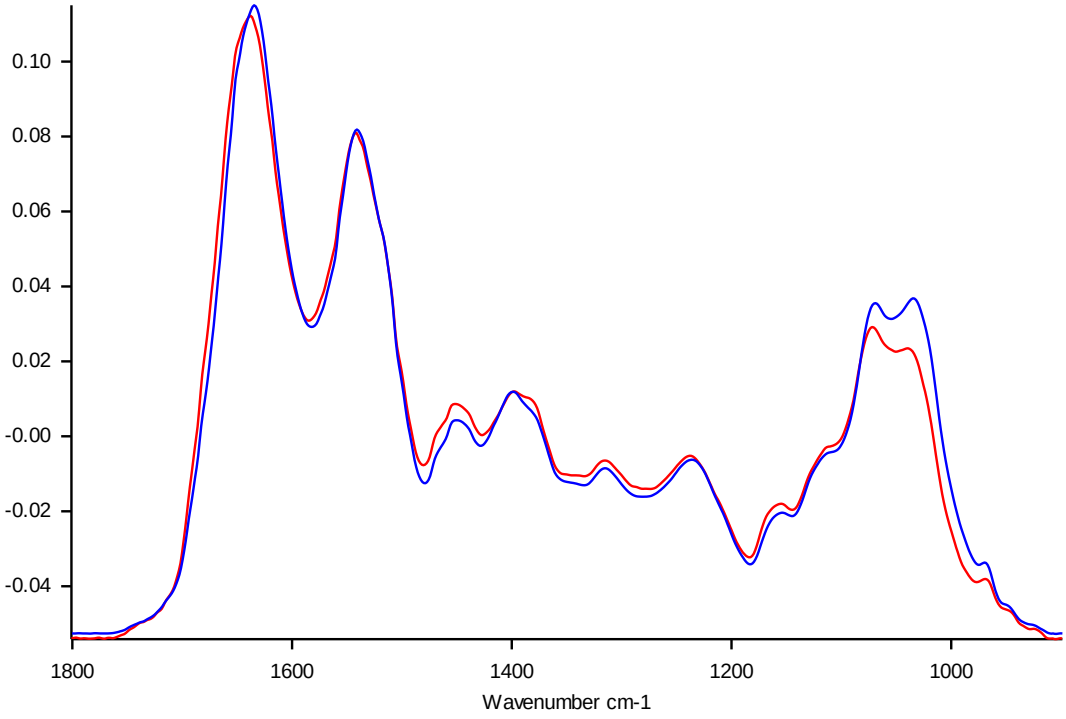


**Figure A4-78: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL12 dried onto the sampling strip (red), and the ATR crystal (blue)**

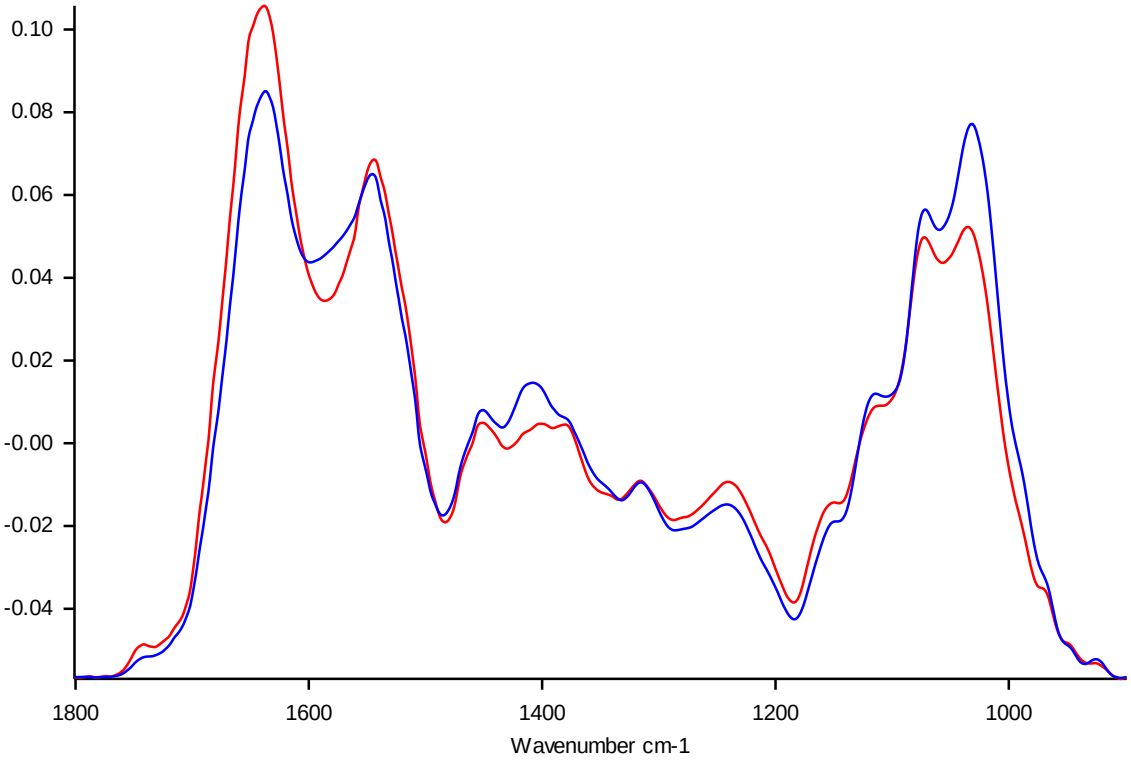


**Figure A4-79: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL13 dried onto the sampling strip (red), and the ATR crystal (blue)**

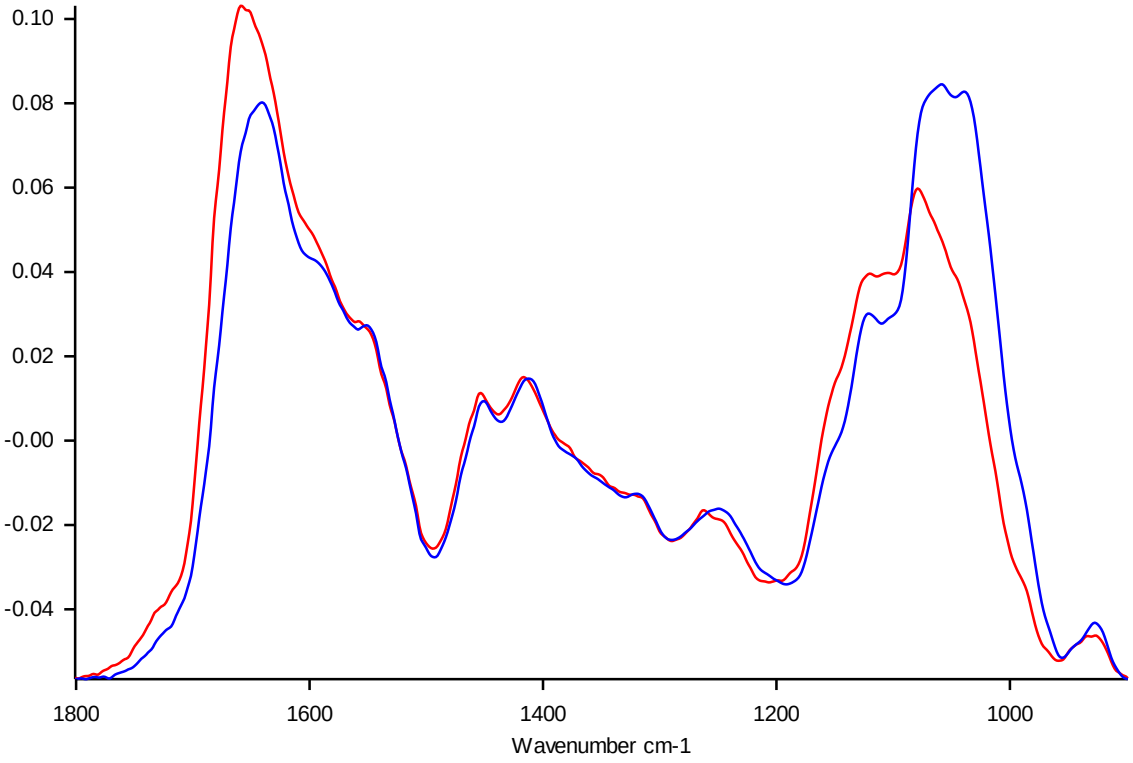


**Figure A4-80: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL14 dried onto the sampling strip (red), and the ATR crystal (blue)**

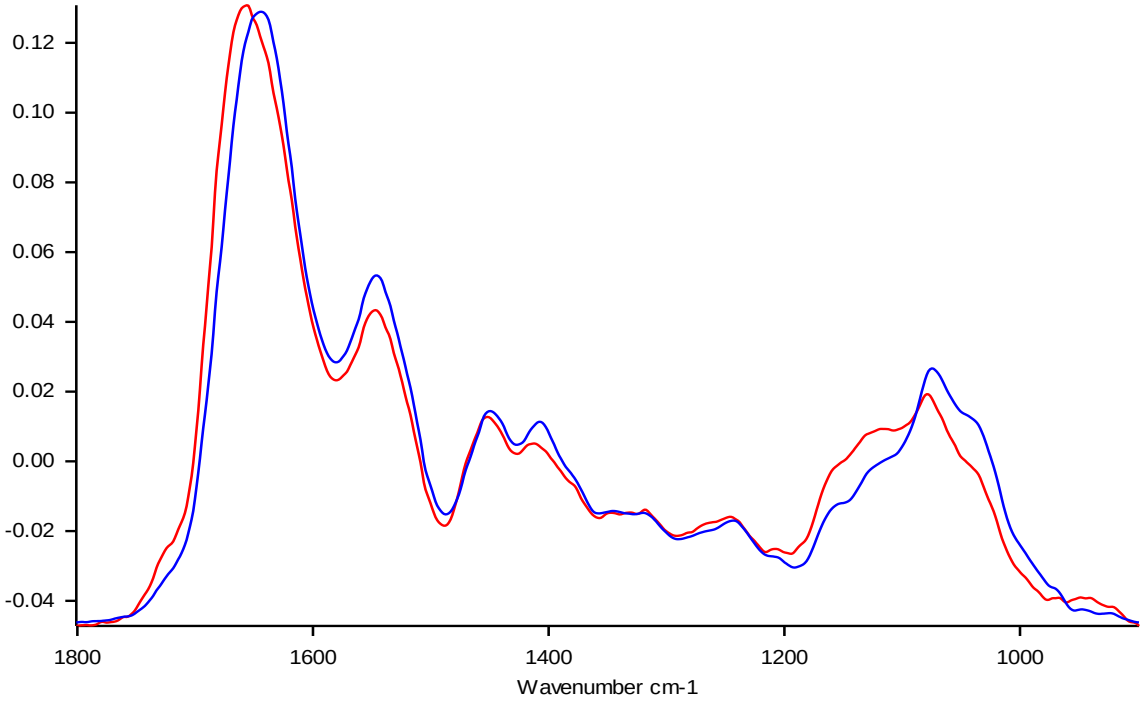


**Figure A4-81: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL16 dried onto the sampling strip (red), and the ATR crystal (blue)**

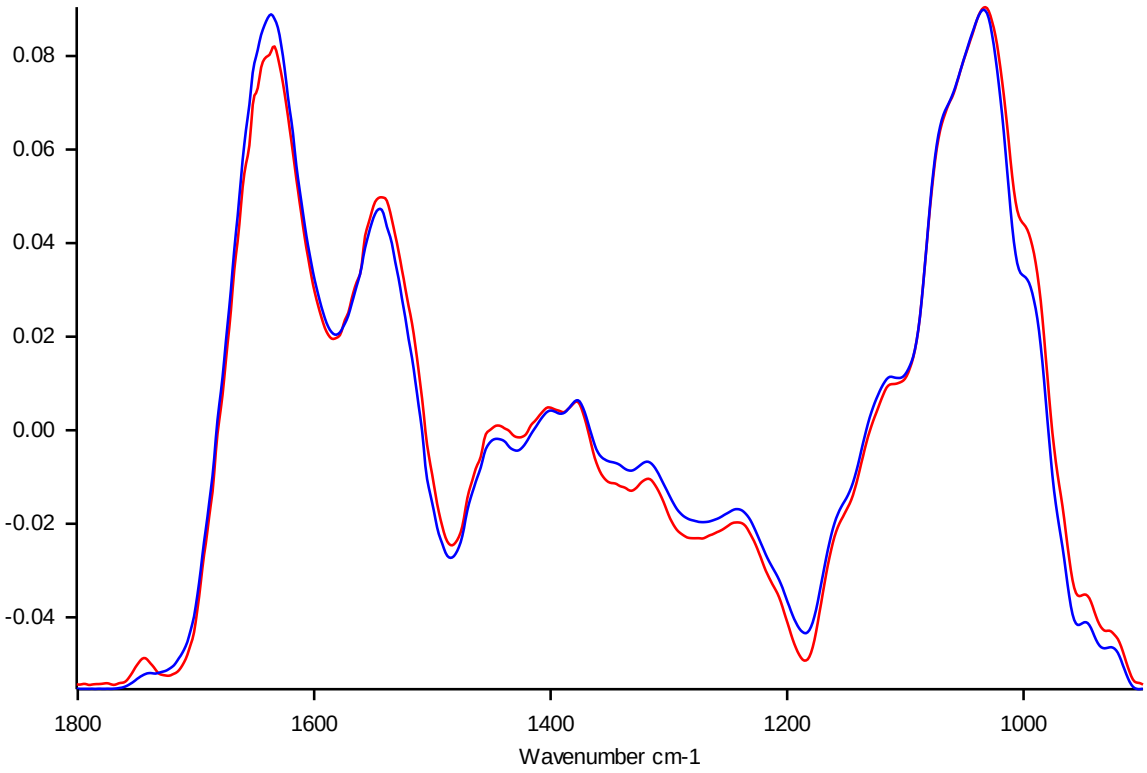


**Figure A4-82: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL17 dried onto the sampling strip (red), and the ATR crystal (blue)**

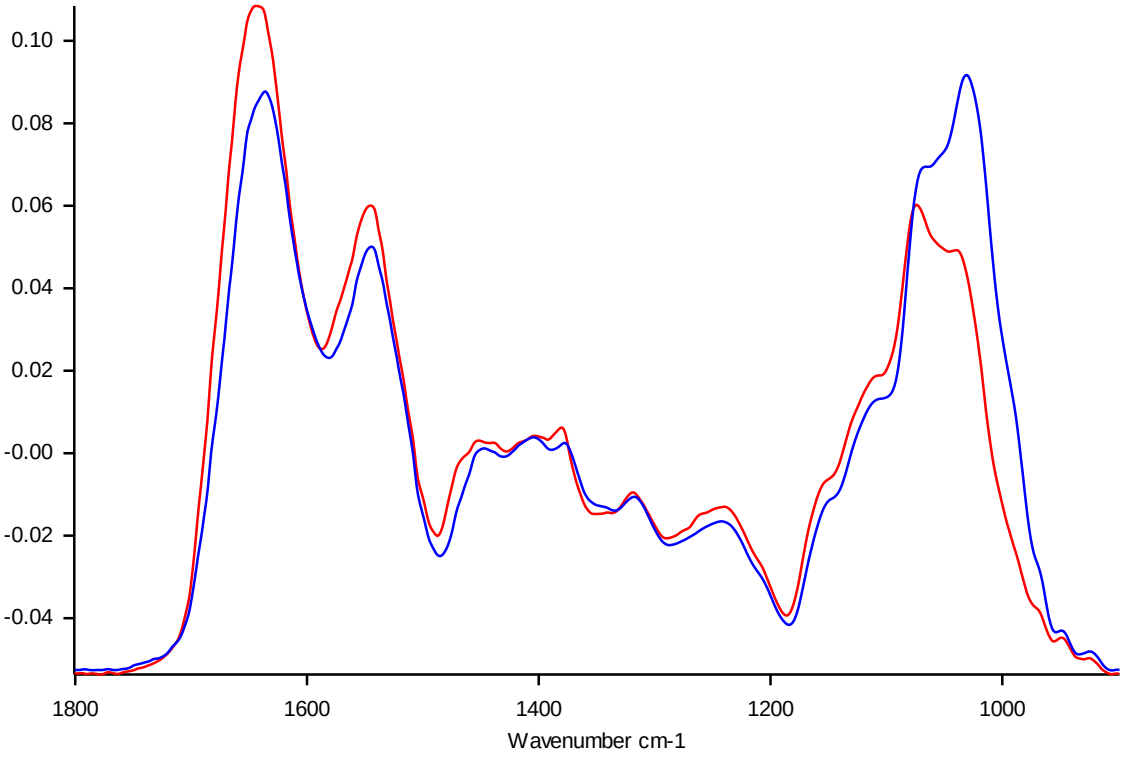


**Figure A4-83: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL19 dried onto the sampling strip (red), and the ATR crystal (blue)**

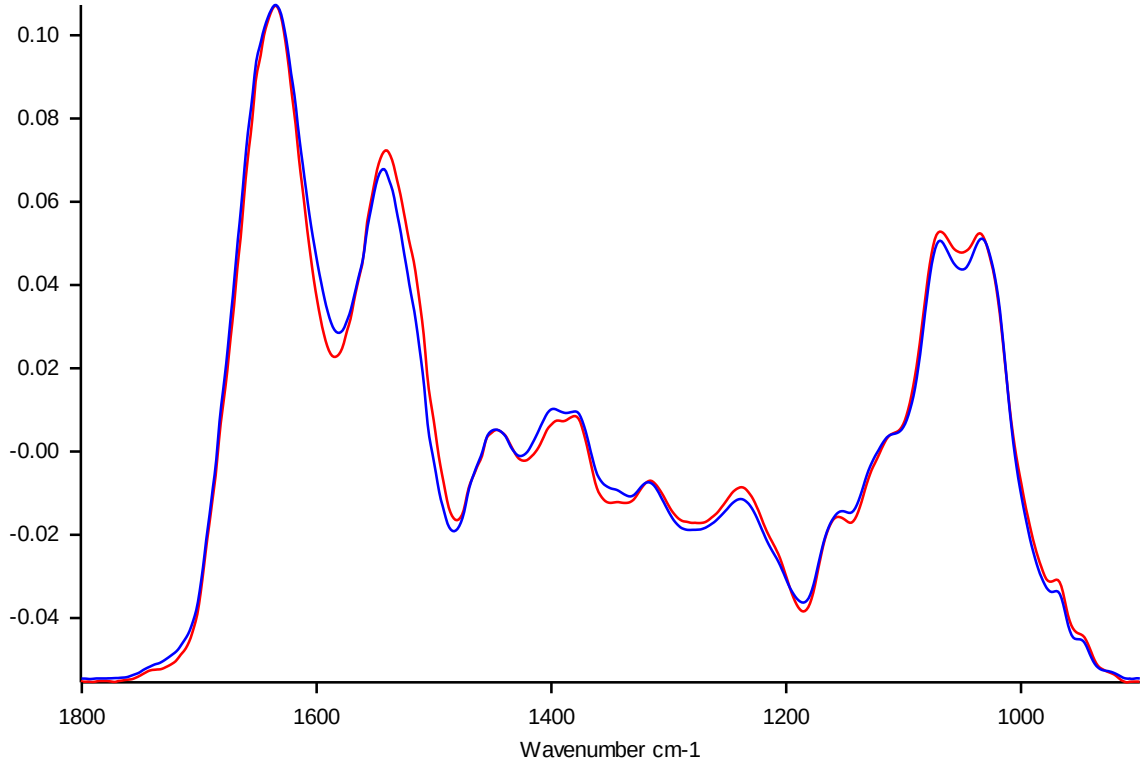


**Figure A4-84: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL20 dried onto the sampling strip (red), and the ATR crystal (blue)**

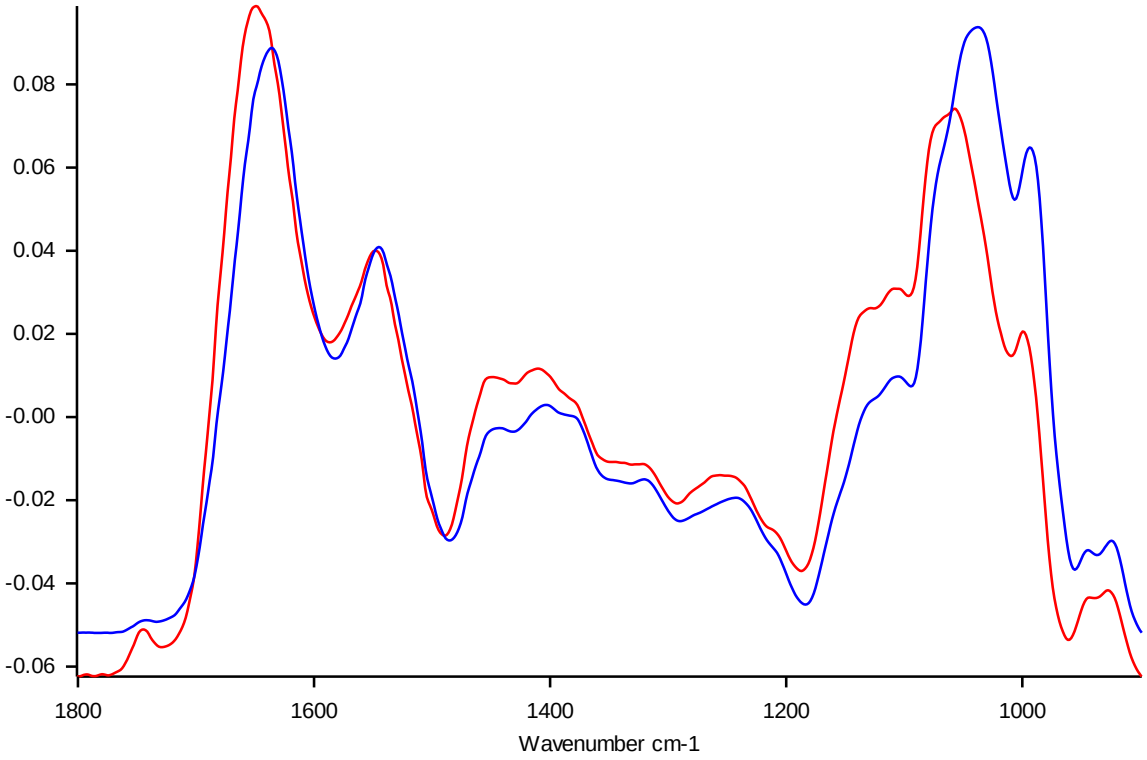


**Figure A4-85: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL21 dried onto the sampling strip (red), and the ATR crystal (blue)**

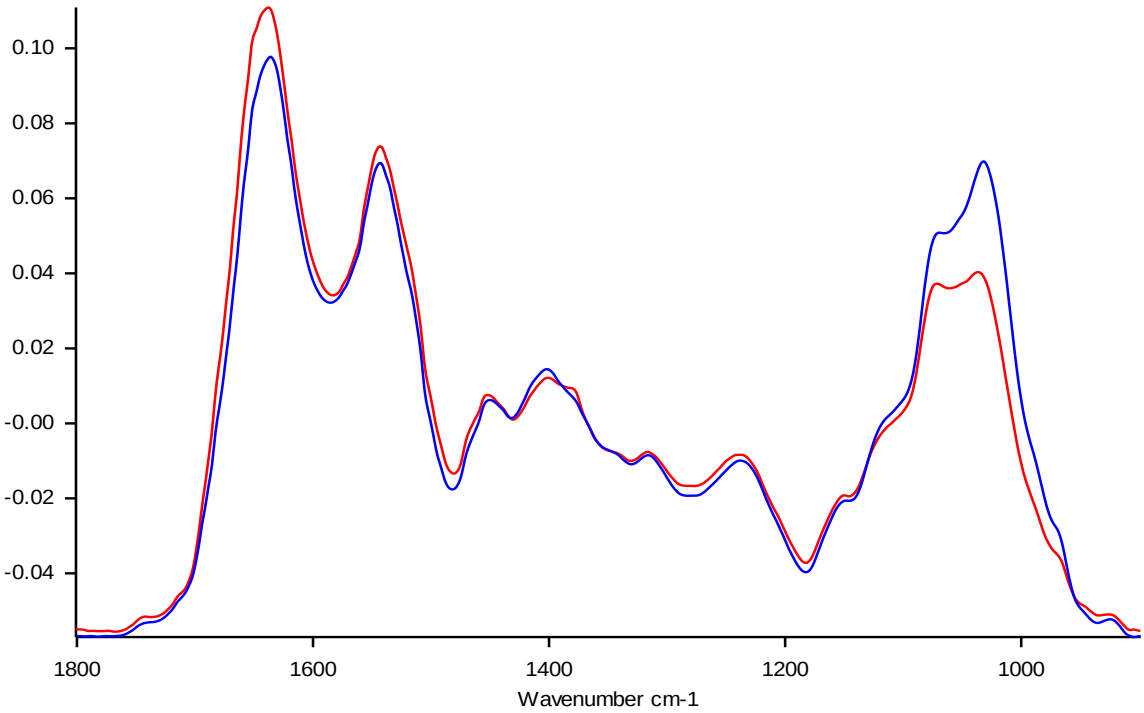


**Figure A4-86: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL22 dried onto the sampling strip (red), and the ATR crystal (blue)**

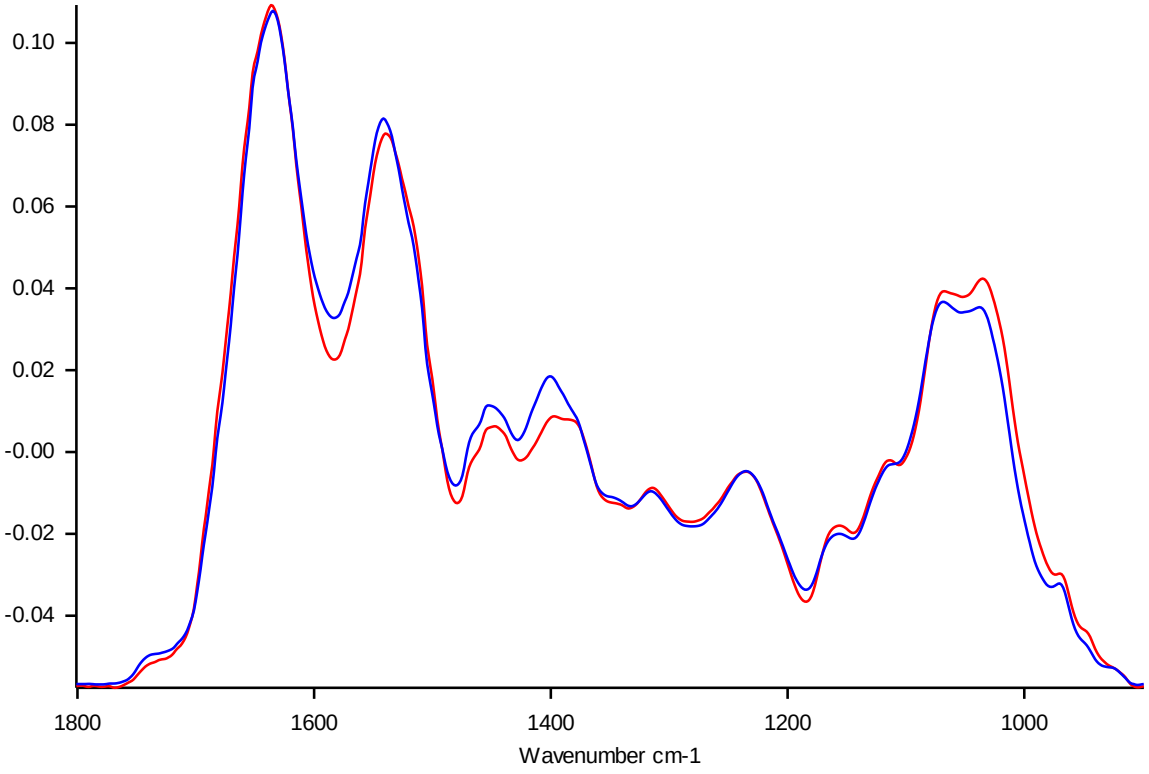


**Figure A4-87: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL23 dried onto the sampling strip (red), and the ATR crystal (blue)**

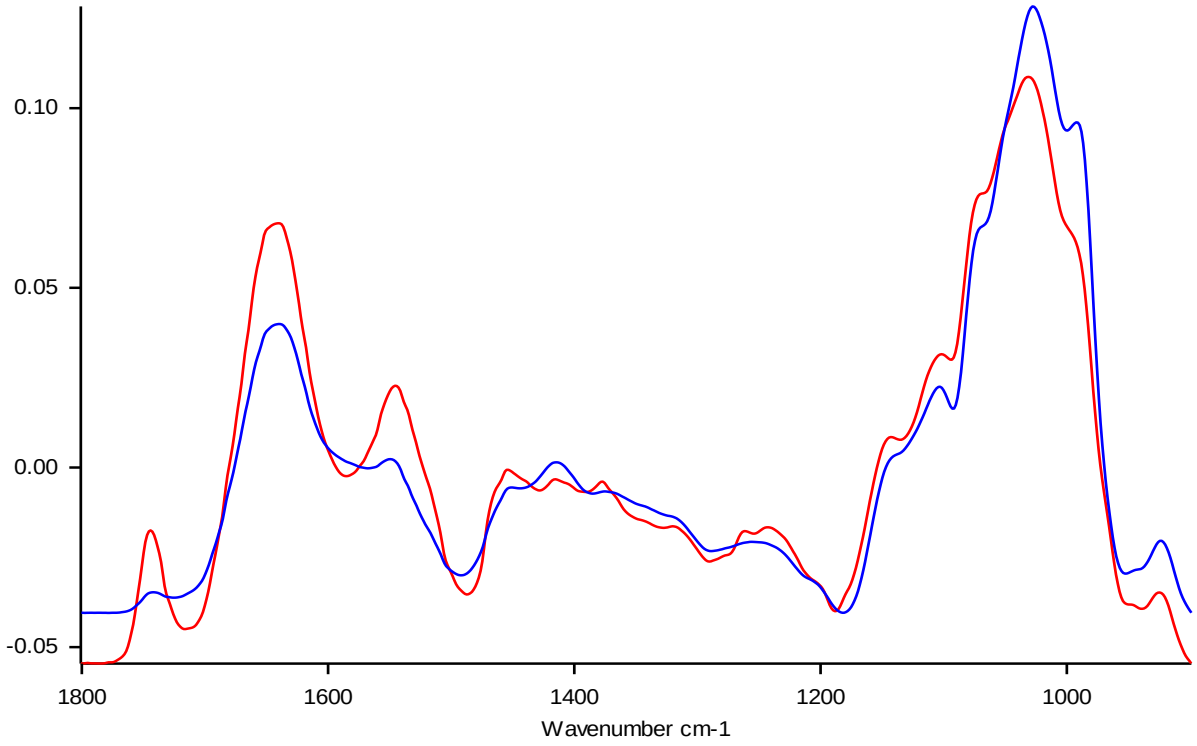


**Figure A4-88: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL24 dried onto the sampling strip (red), and the ATR crystal (blue)**

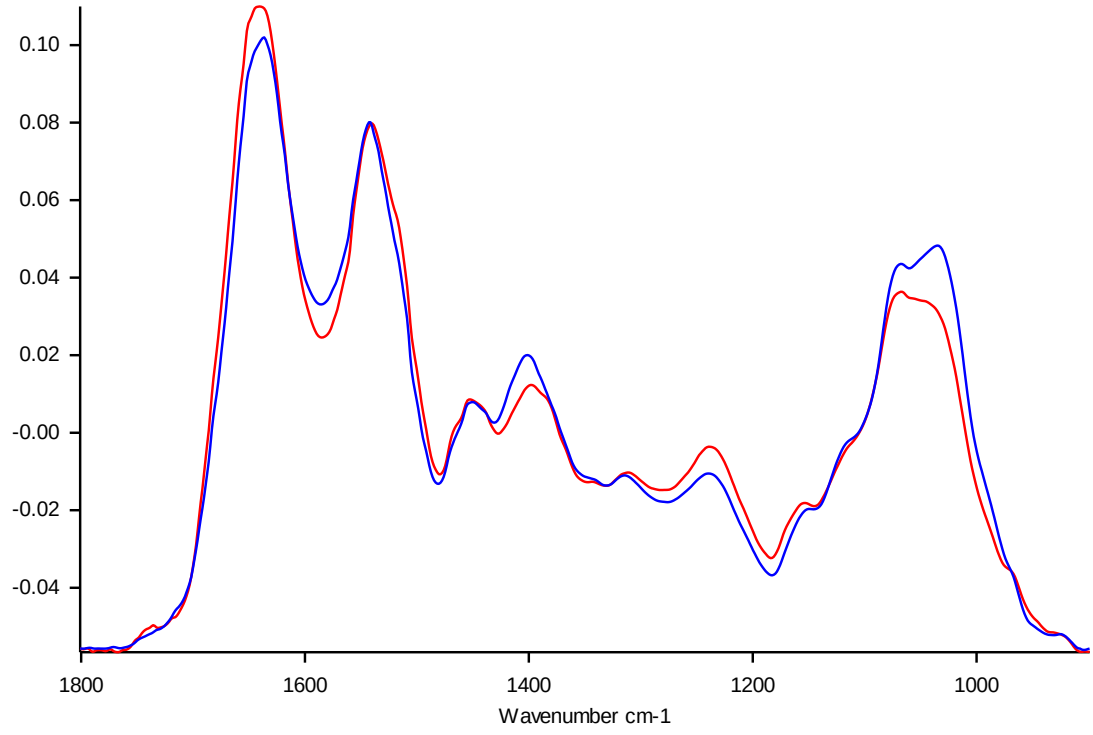


**Figure A4-89: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL25 dried onto the sampling strip (red), and the ATR crystal (blue)**

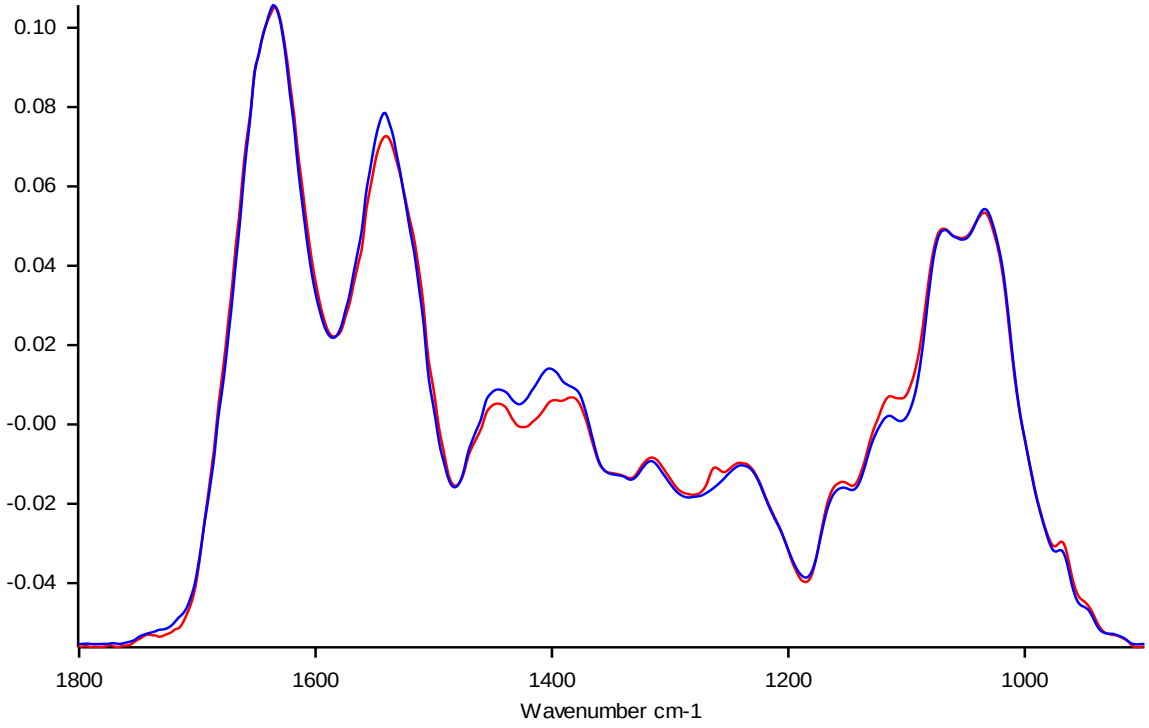


**Figure A4-90: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL26 dried onto the sampling strip (red), and the ATR crystal (blue)**

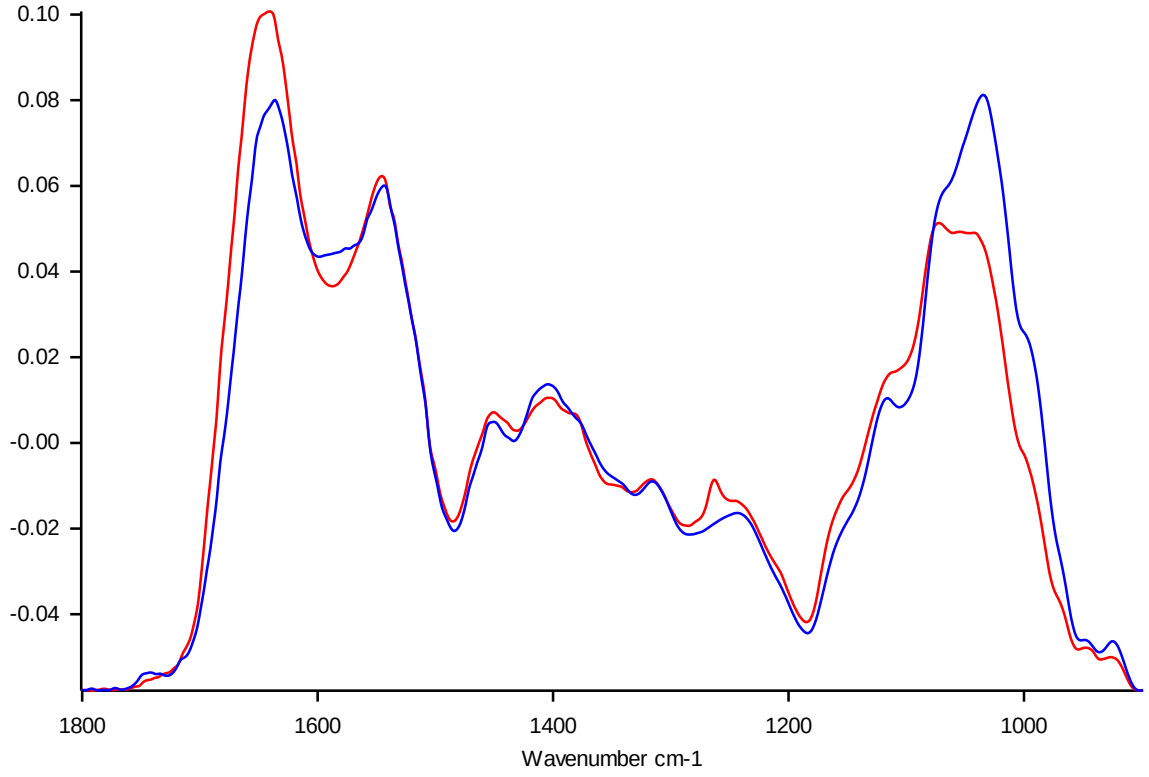


**Figure A4-91: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL27 dried onto the sampling strip (red), and the ATR crystal (blue)**

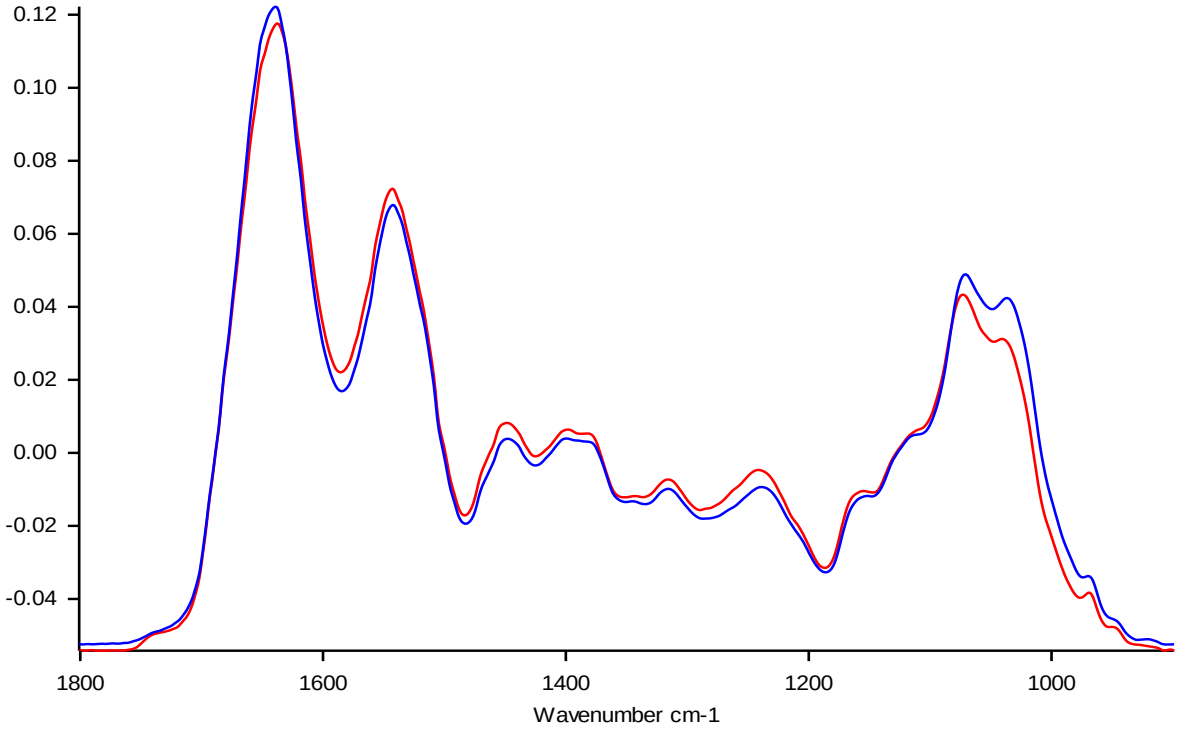


**Figure A4-92: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL28 dried onto the sampling strip (red), and the ATR crystal (blue)**

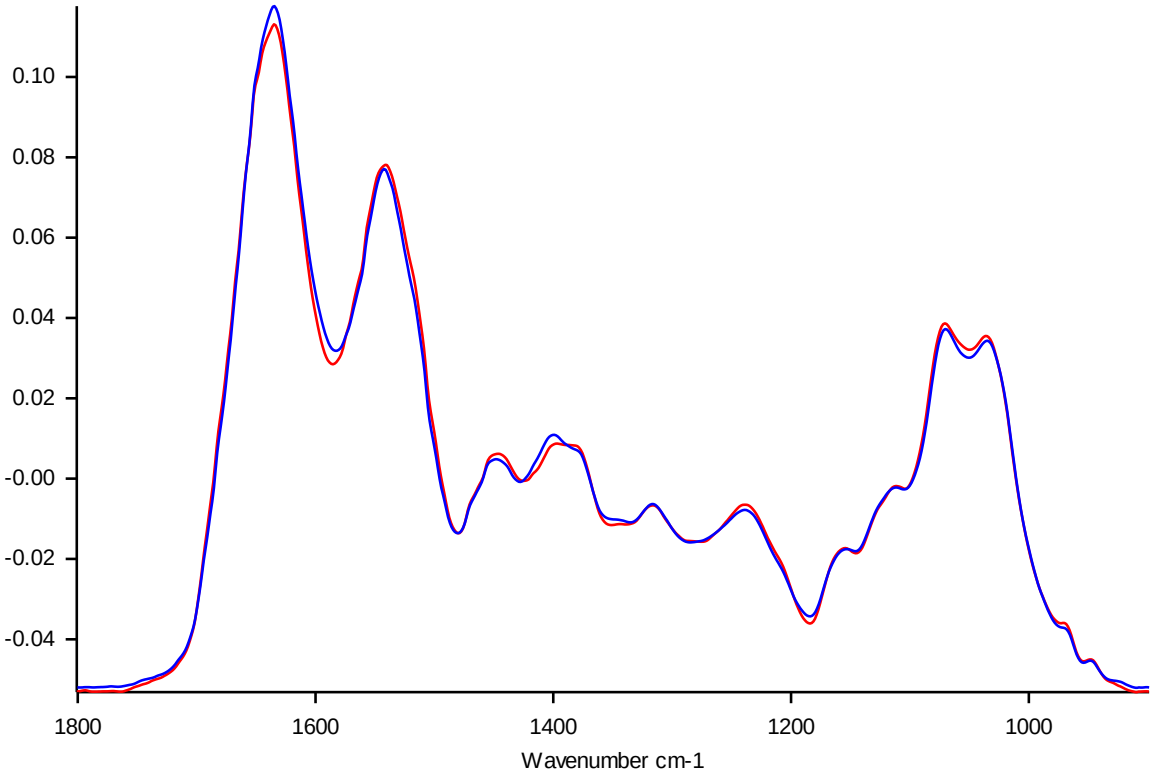


**Figure A4-93: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL29 dried onto the sampling strip (red), and the ATR crystal (blue)**

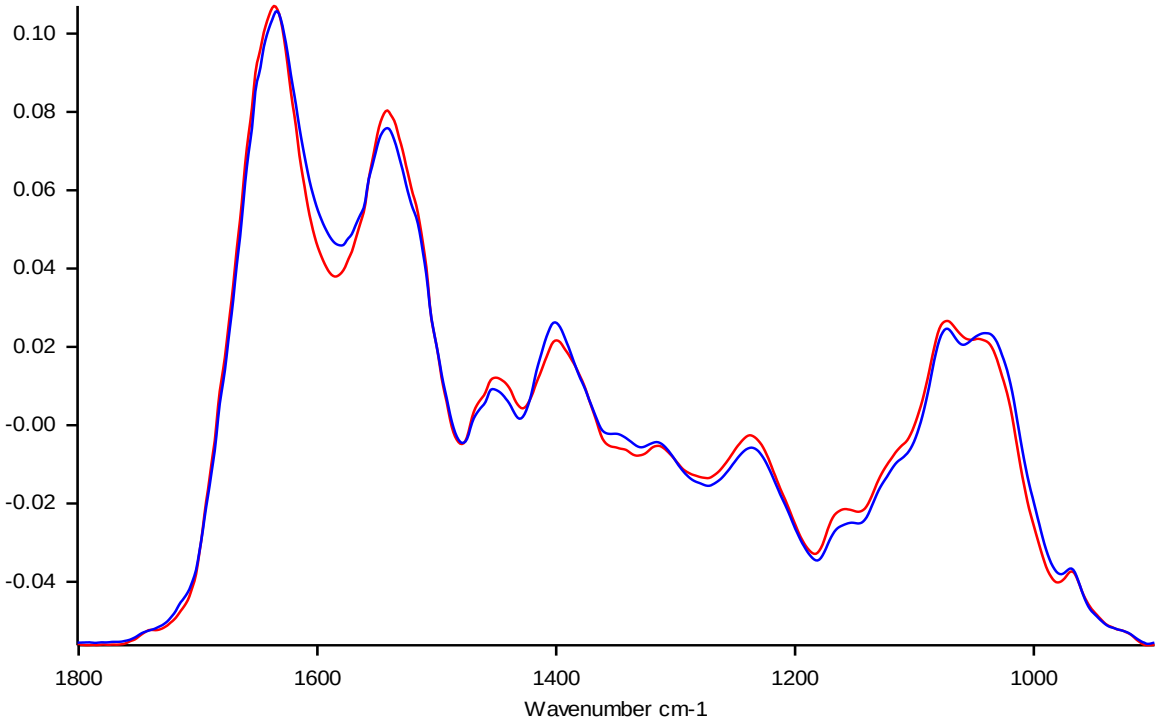


**Figure A4-94: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL30 dried onto the sampling strip (red), and the ATR crystal (blue)**

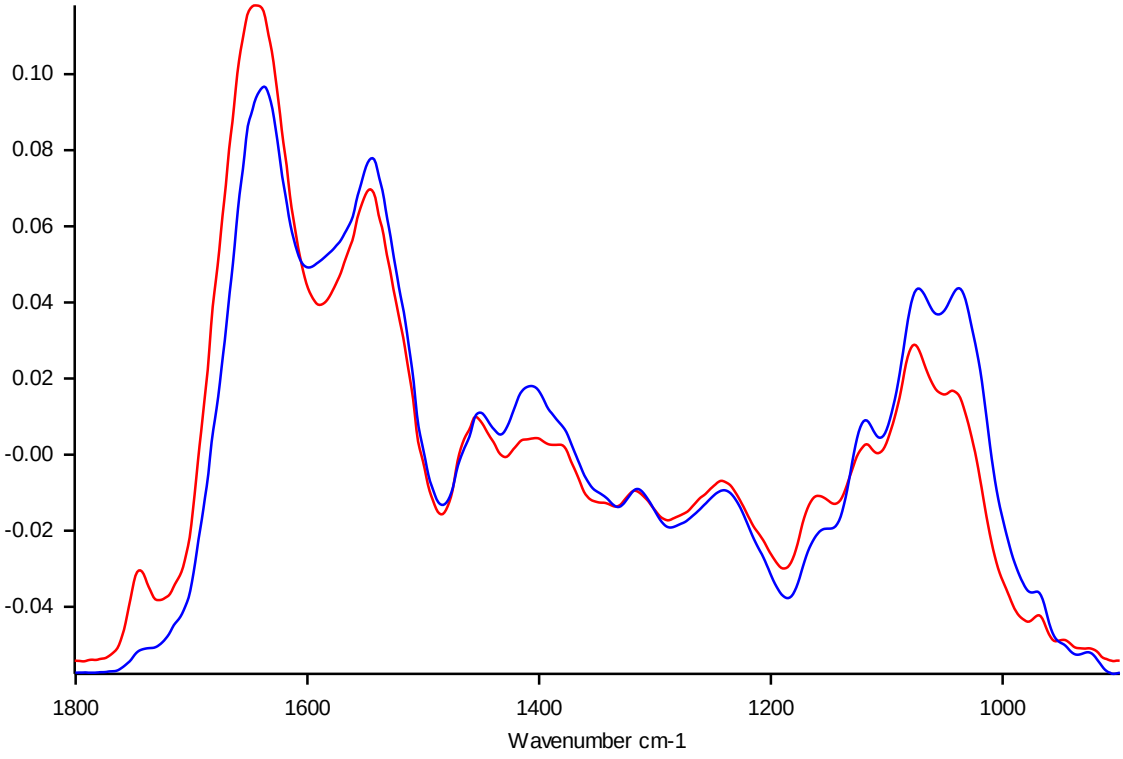


**Figure A4-95: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL32 dried onto the sampling strip (red), and the ATR crystal (blue)**

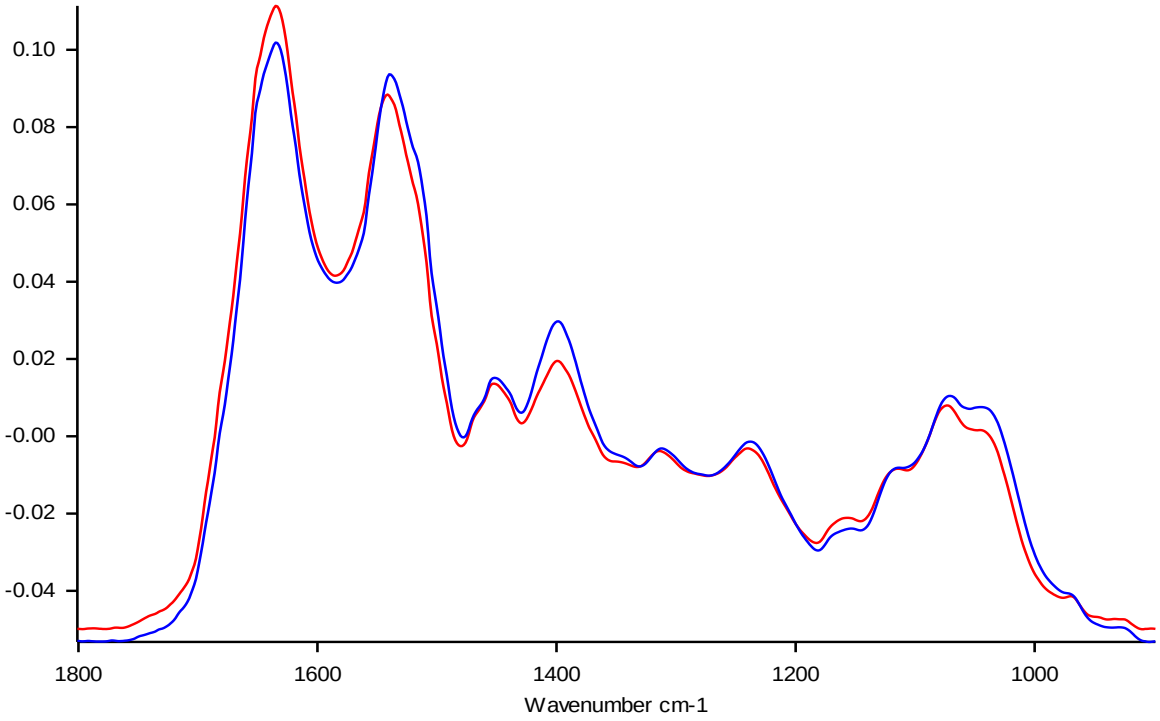


**Figure A4-96: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL33 dried onto the sampling strip (red), and the ATR crystal (blue)**

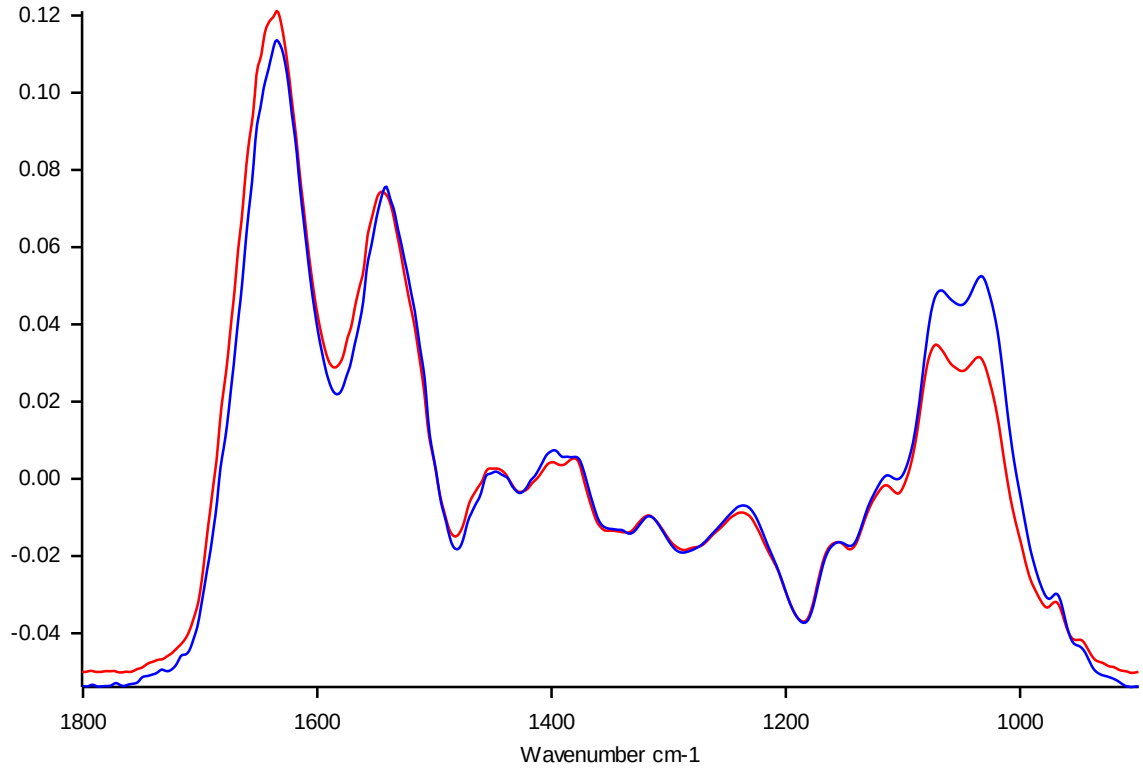


**Figure A4-97: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL34 dried onto the sampling strip (red), and the ATR crystal (blue)**

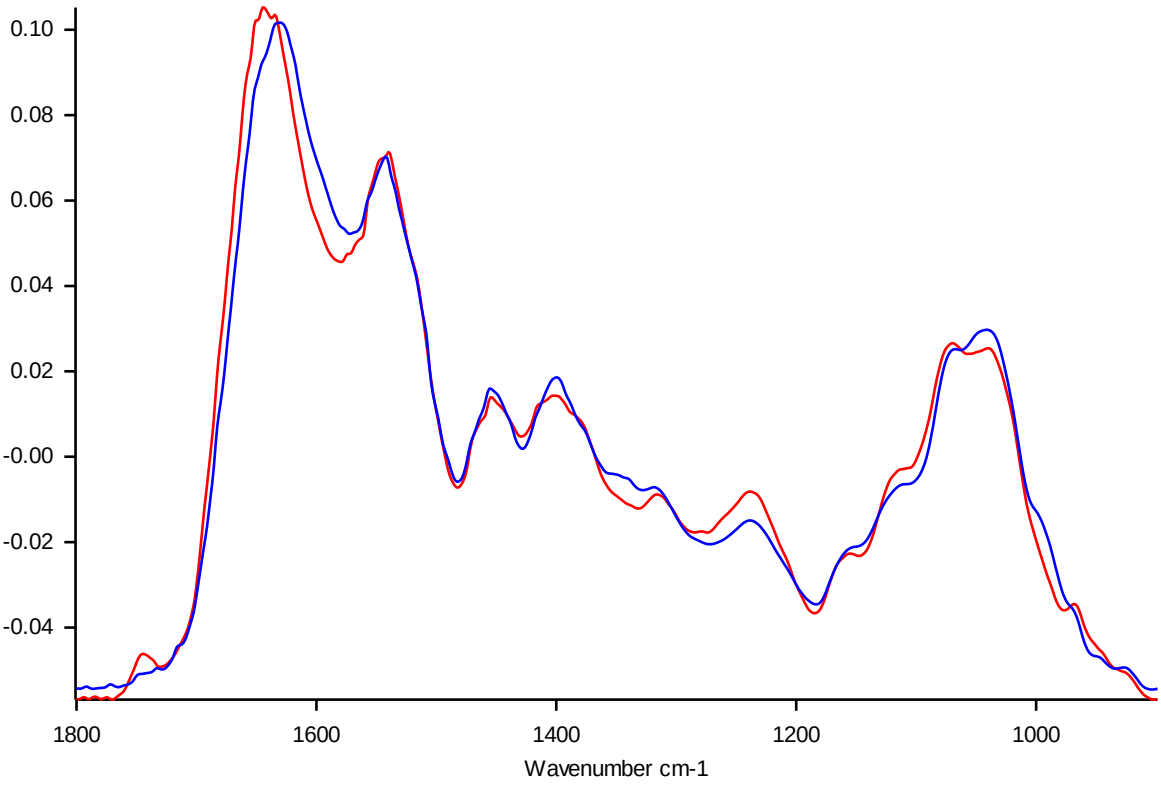


**Figure A4-98: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL35 dried onto the sampling strip (red), and the ATR crystal (blue)**

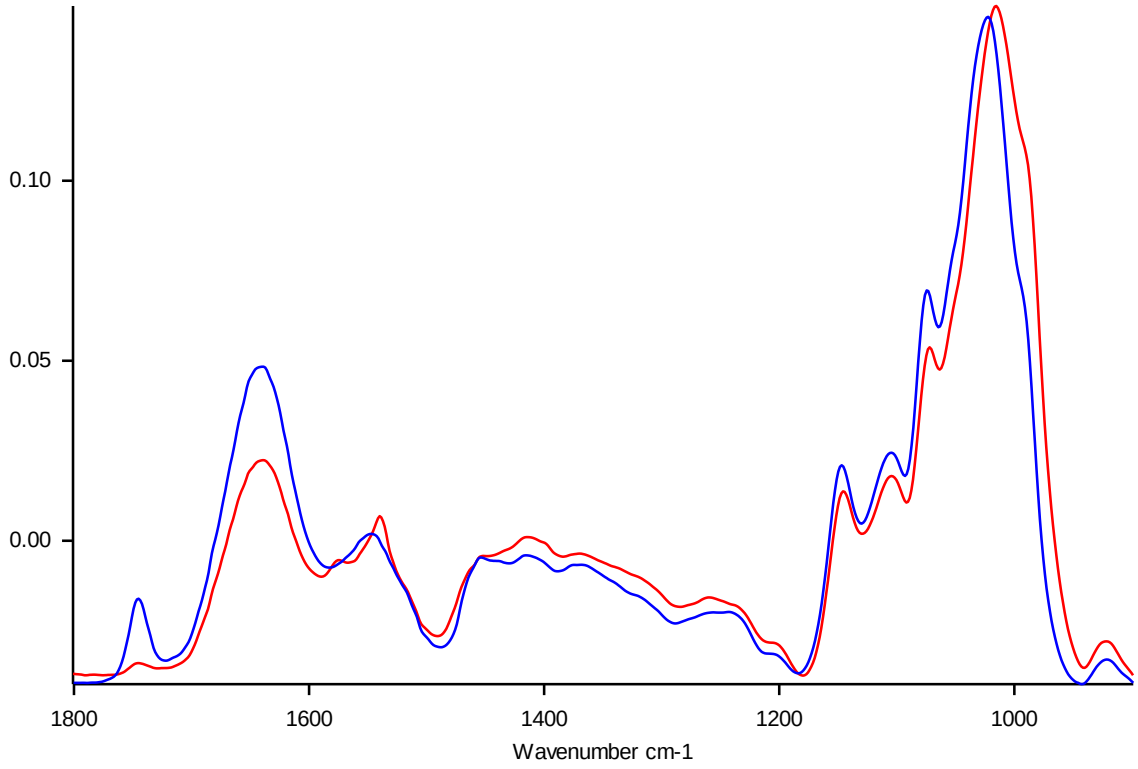


**Figure A4-99: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL36 dried onto the sampling strip (red), and the ATR crystal (blue)**

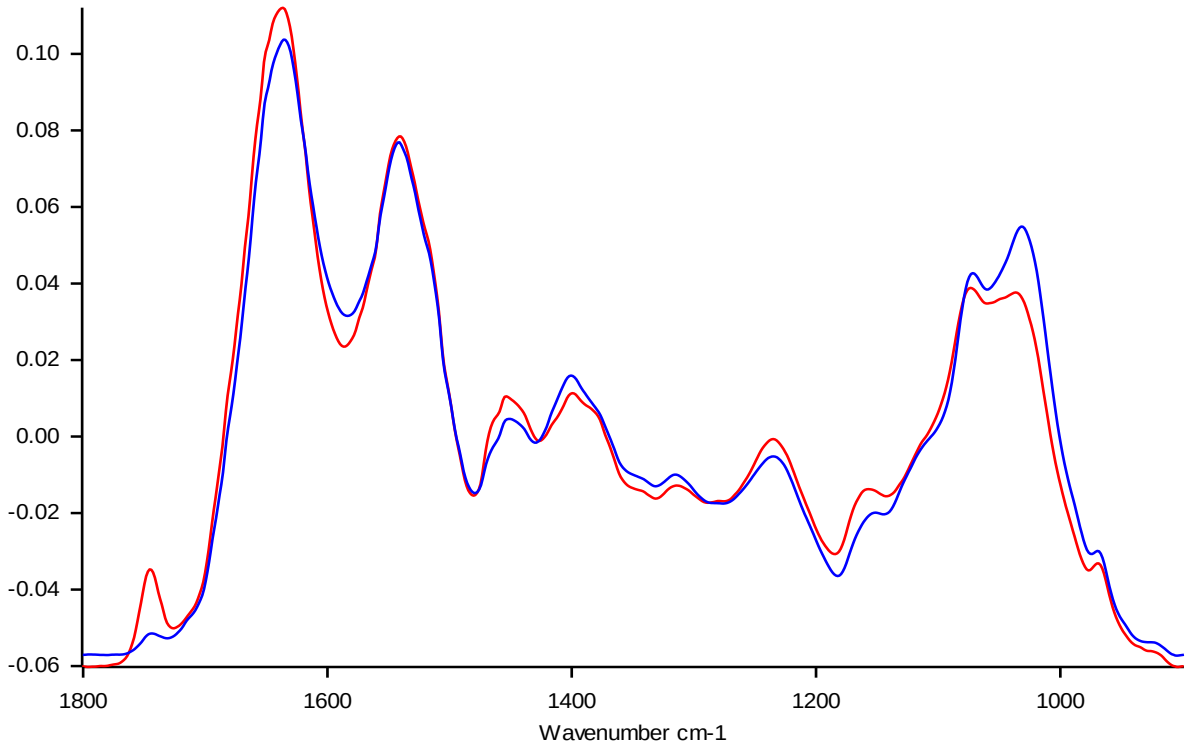


**Figure A4-100: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL37 dried onto the sampling strip (red), and the ATR crystal (blue)**

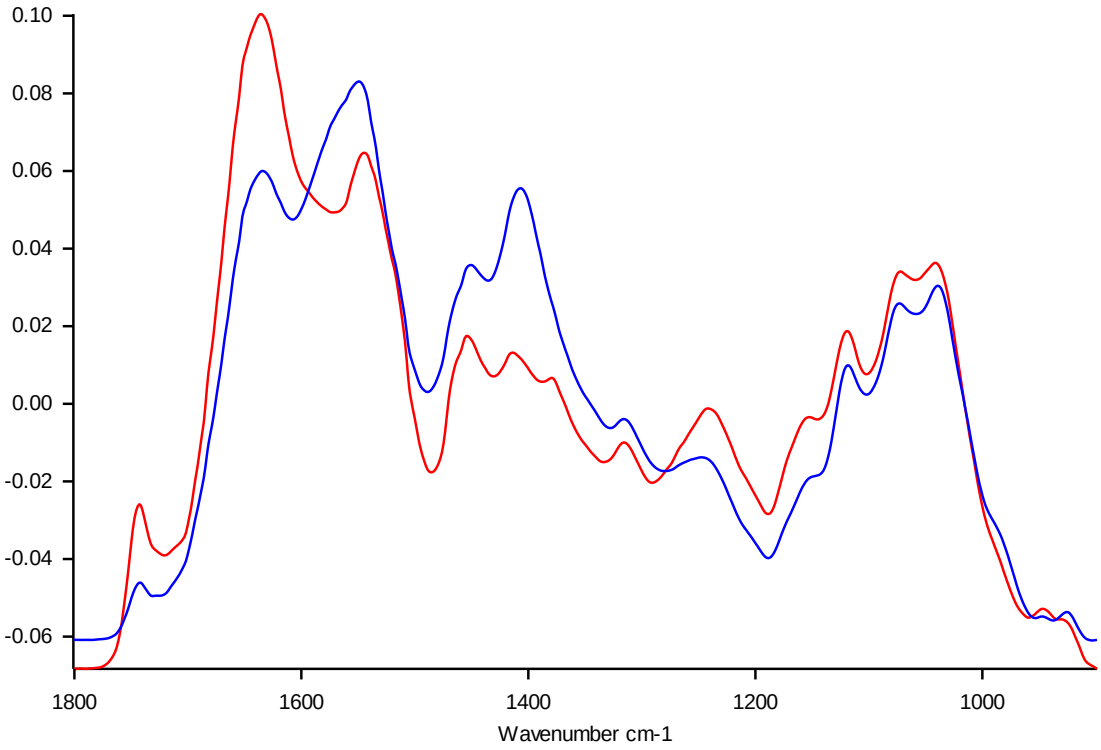


**Figure A4-101: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample SGUL38 dried onto the sampling strip (red), and the ATR crystal (blue)**

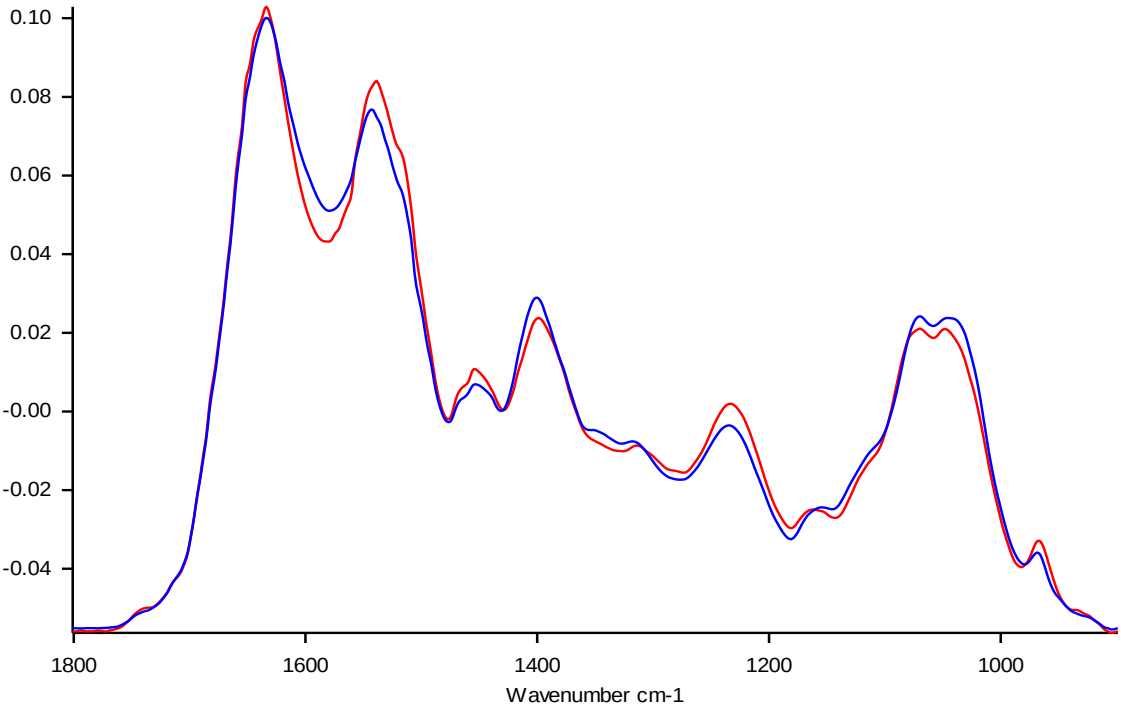


**Figure A4-102: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR02 dried onto the sampling strip (red), and the ATR crystal (blue)**

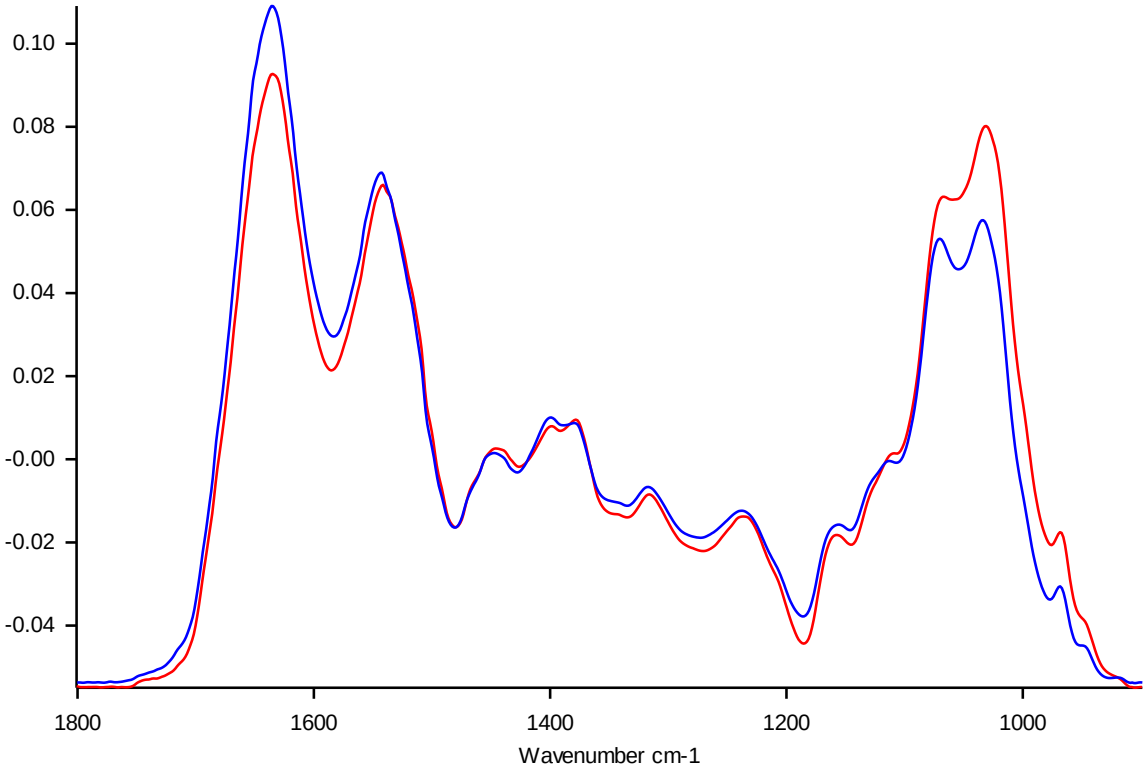


**Figure A4-103: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR04 dried onto the sampling strip (red), and the ATR crystal (blue)**

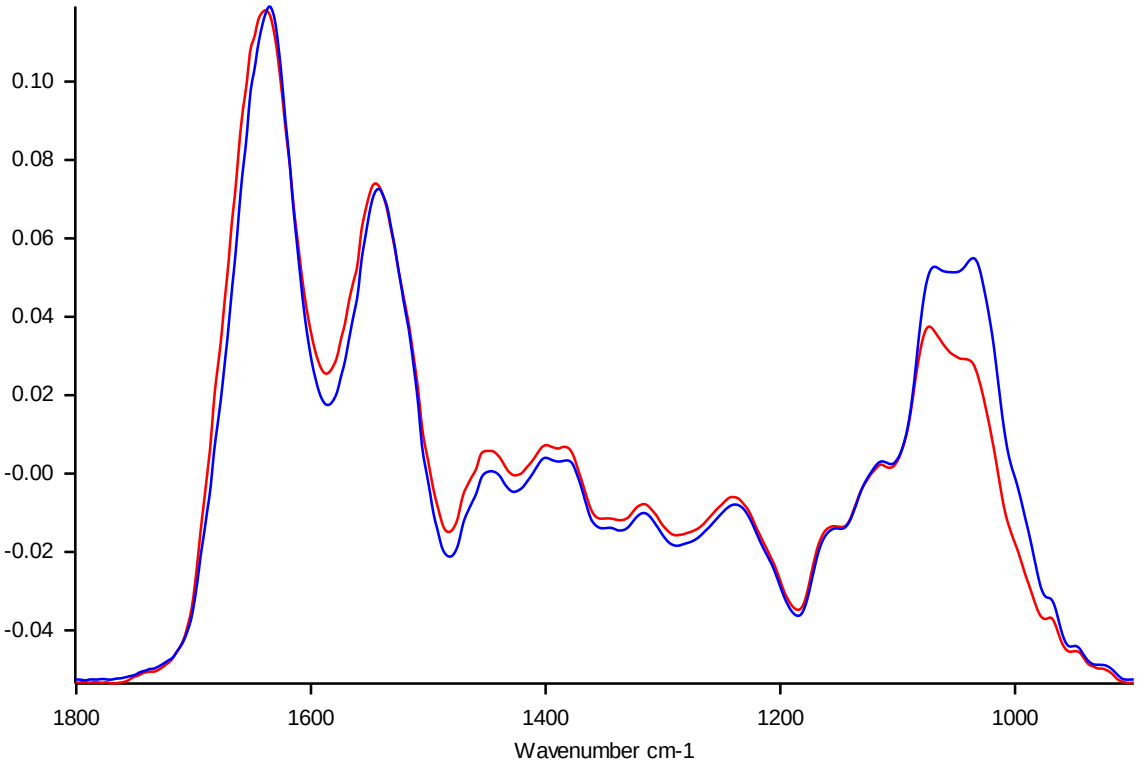


**Figure A4-104: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR05 dried onto the sampling strip (red), and the ATR crystal (blue)**

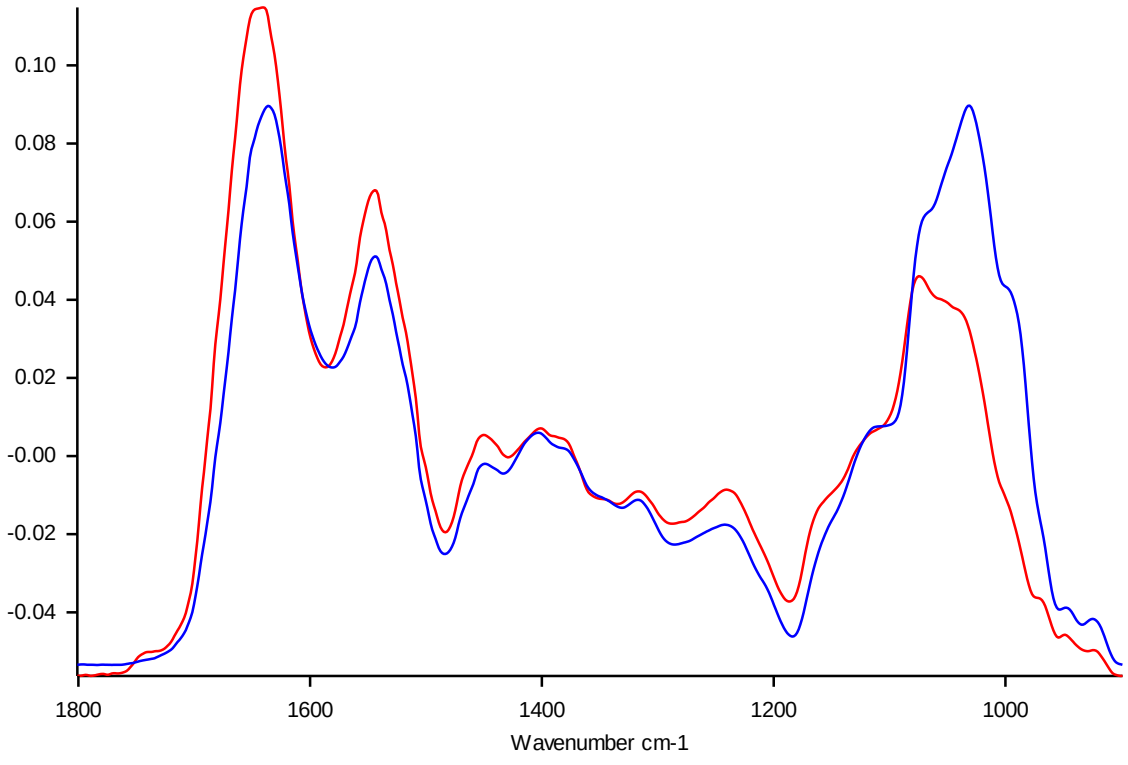


**Figure A4-105: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR06 dried onto the sampling strip (red), and the ATR crystal (blue)**

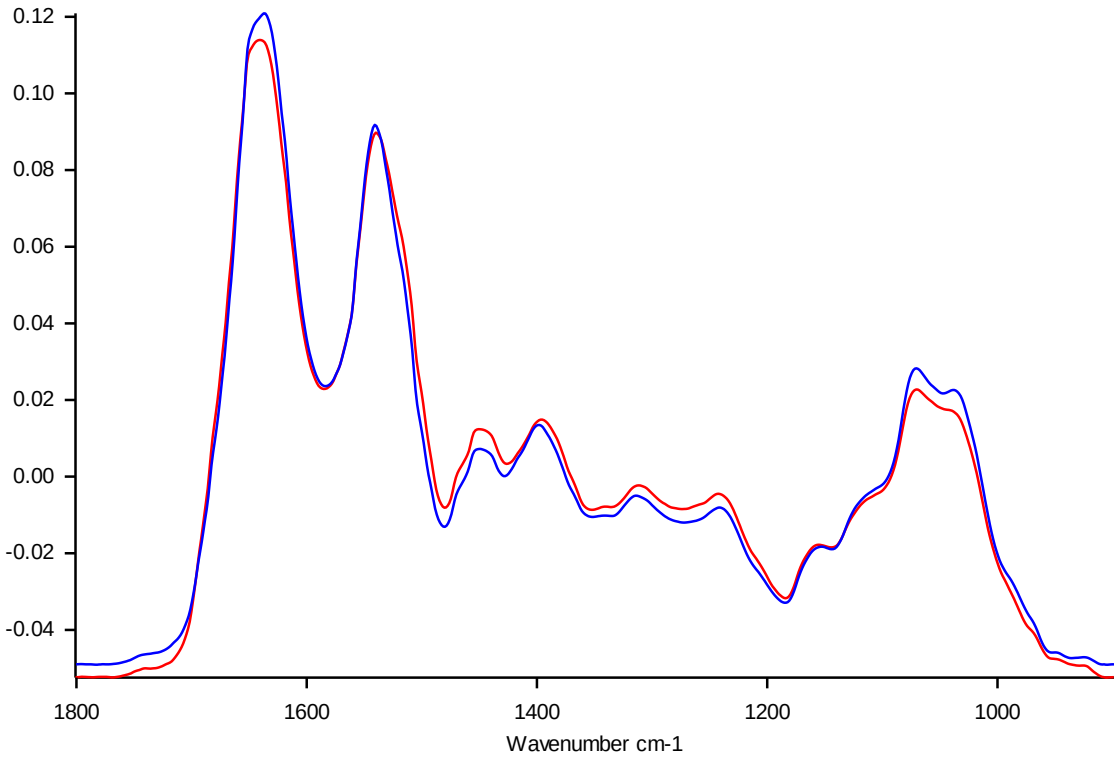


**Figure A4-106: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR08 dried onto the sampling strip (red), and the ATR crystal (blue)**

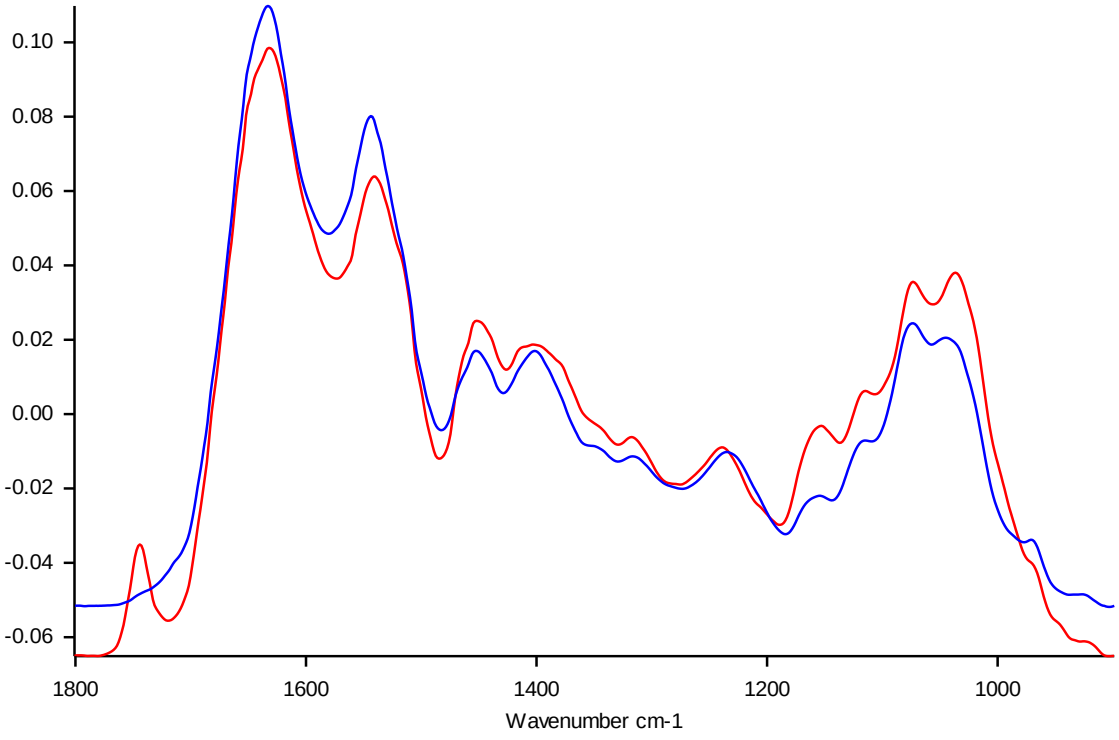


**Figure A4-107: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR09 dried onto the sampling strip (red), and the ATR crystal (blue)**

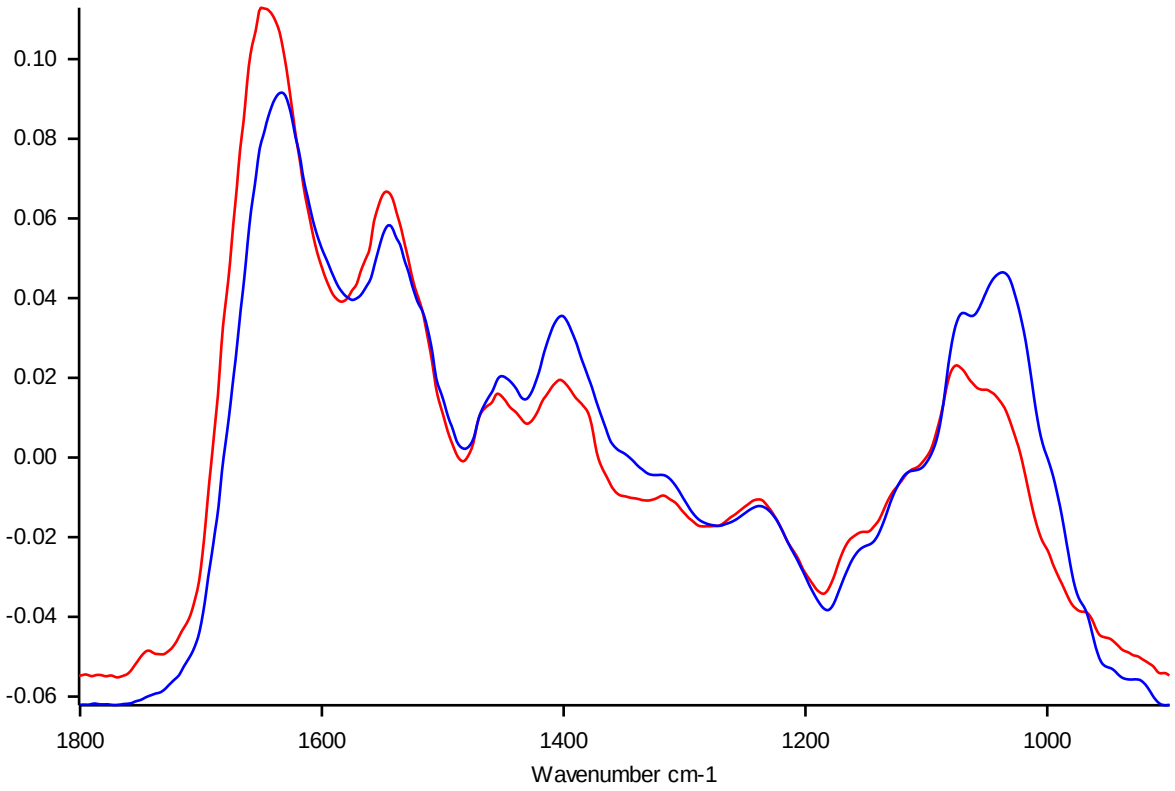


**Figure A4-108: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR11 dried onto the sampling strip (red), and the ATR crystal (blue)**

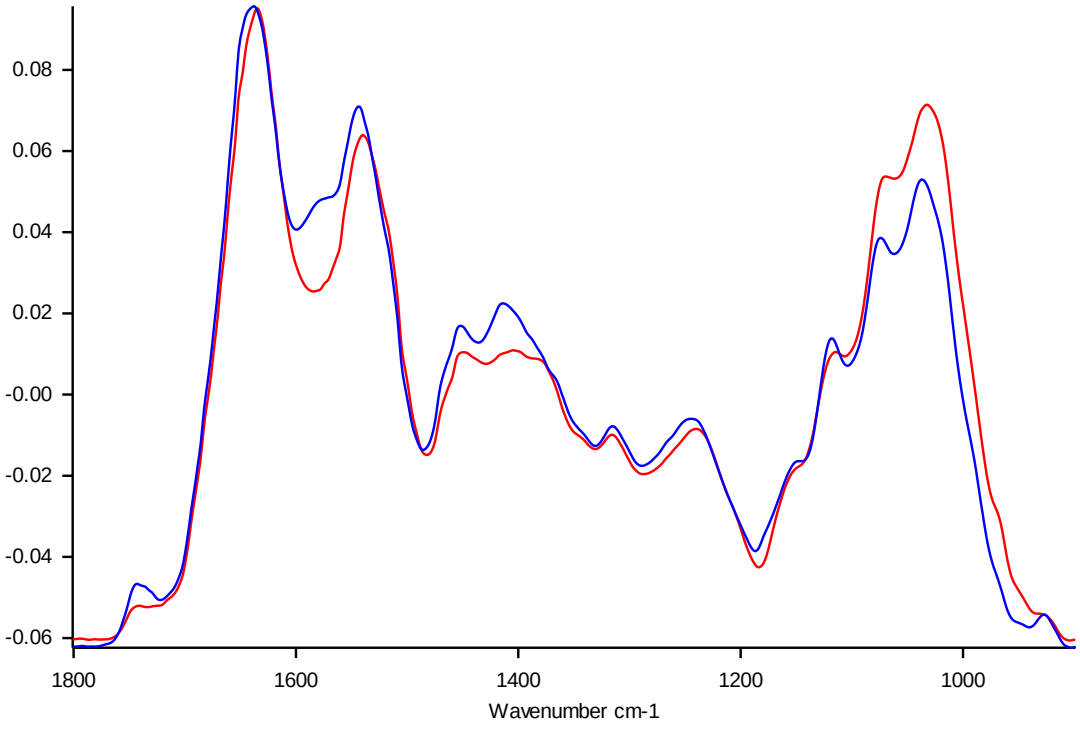


**Figure A4-109: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR12 dried onto the sampling strip (red), and the ATR crystal (blue)**

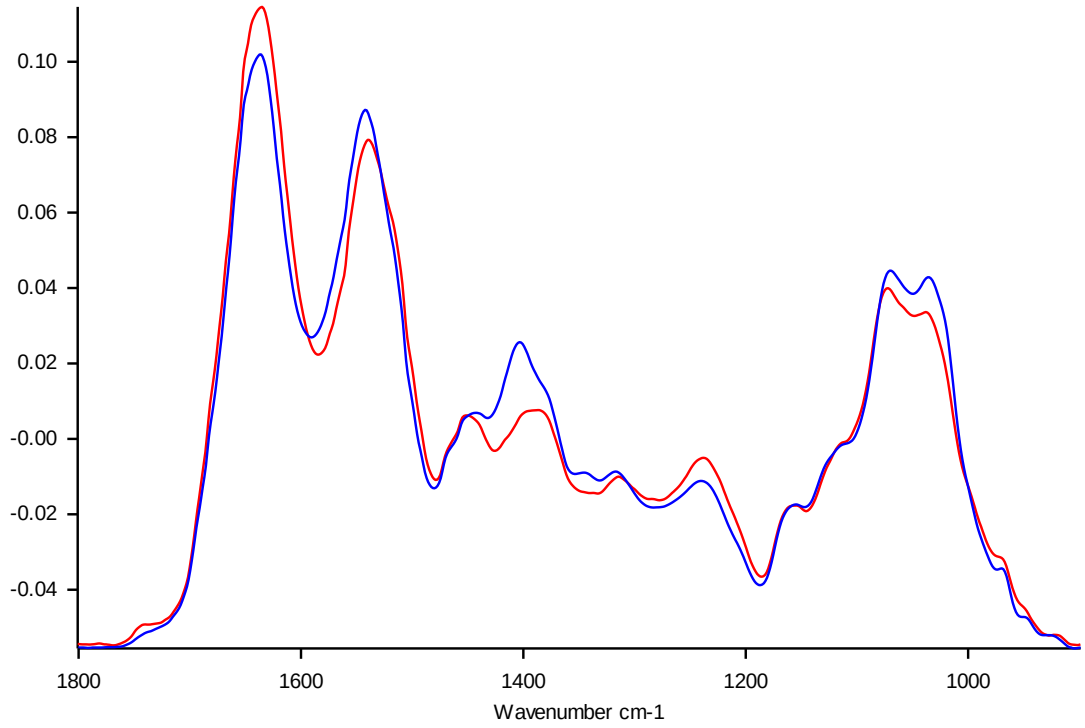


**Figure A4-110: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR13 dried onto the sampling strip (red), and the ATR crystal (blue)**

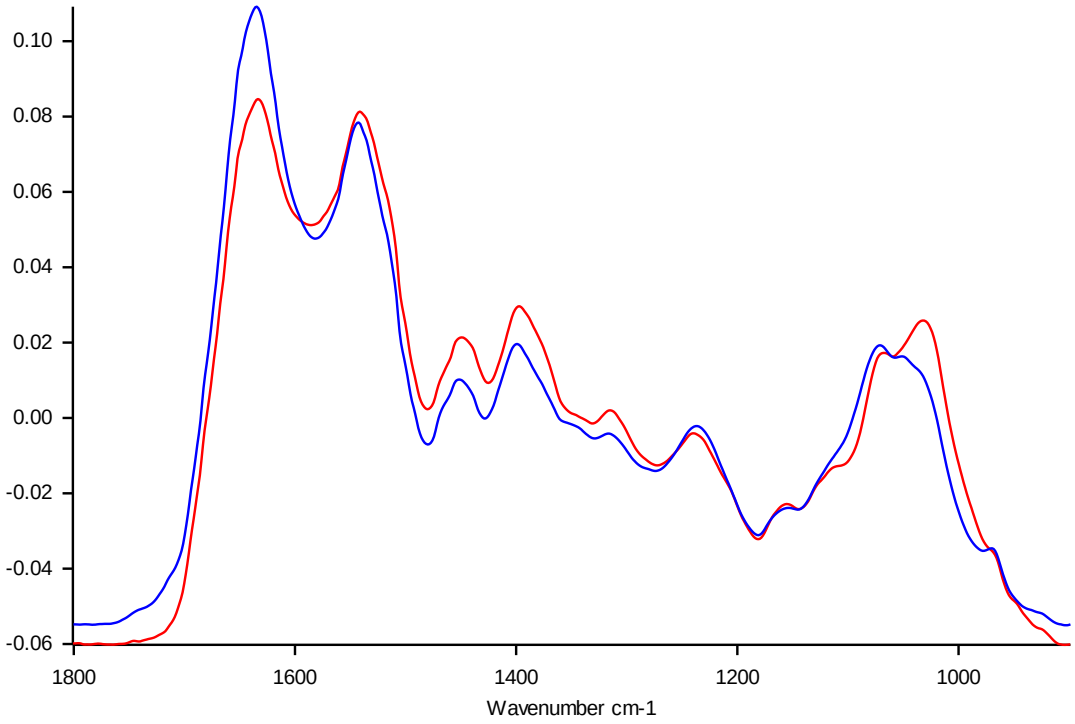


**Figure A4-111: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR16 dried onto the sampling strip (red), and the ATR crystal (blue)**

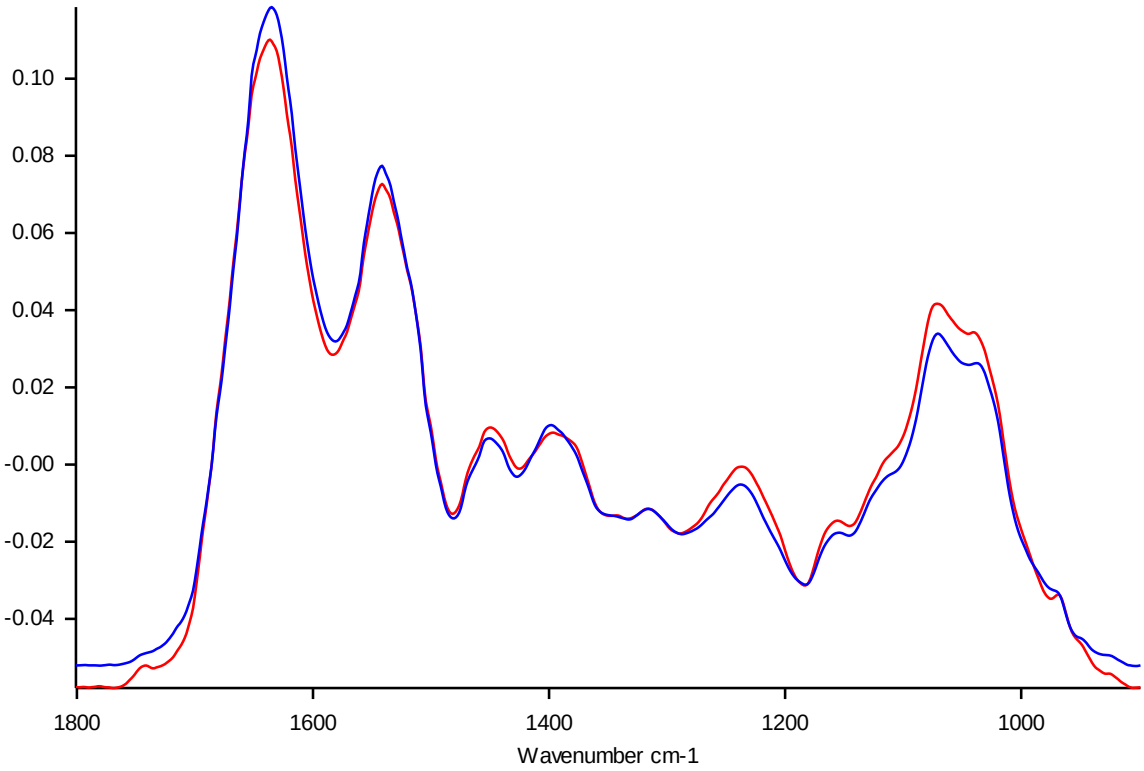


**Figure A4-112: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR17 dried onto the sampling strip (red), and the ATR crystal (blue)**

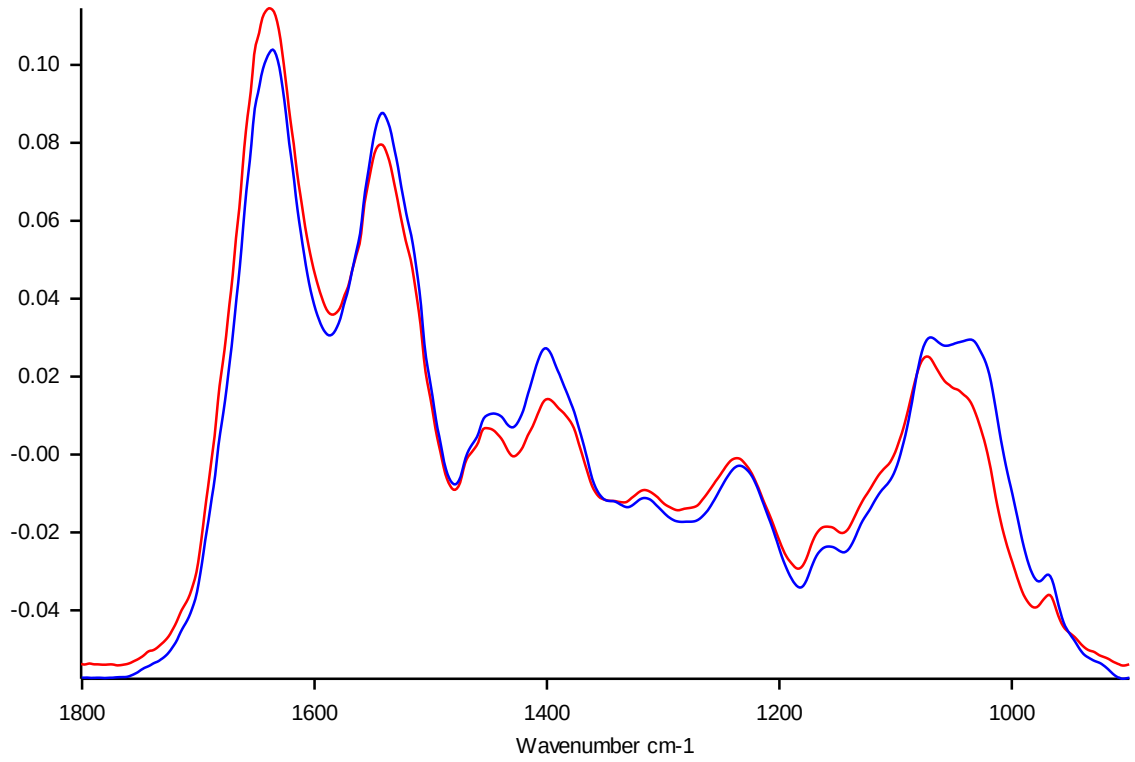


**Figure A4-113: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR18 dried onto the sampling strip (red), and the ATR crystal (blue)**

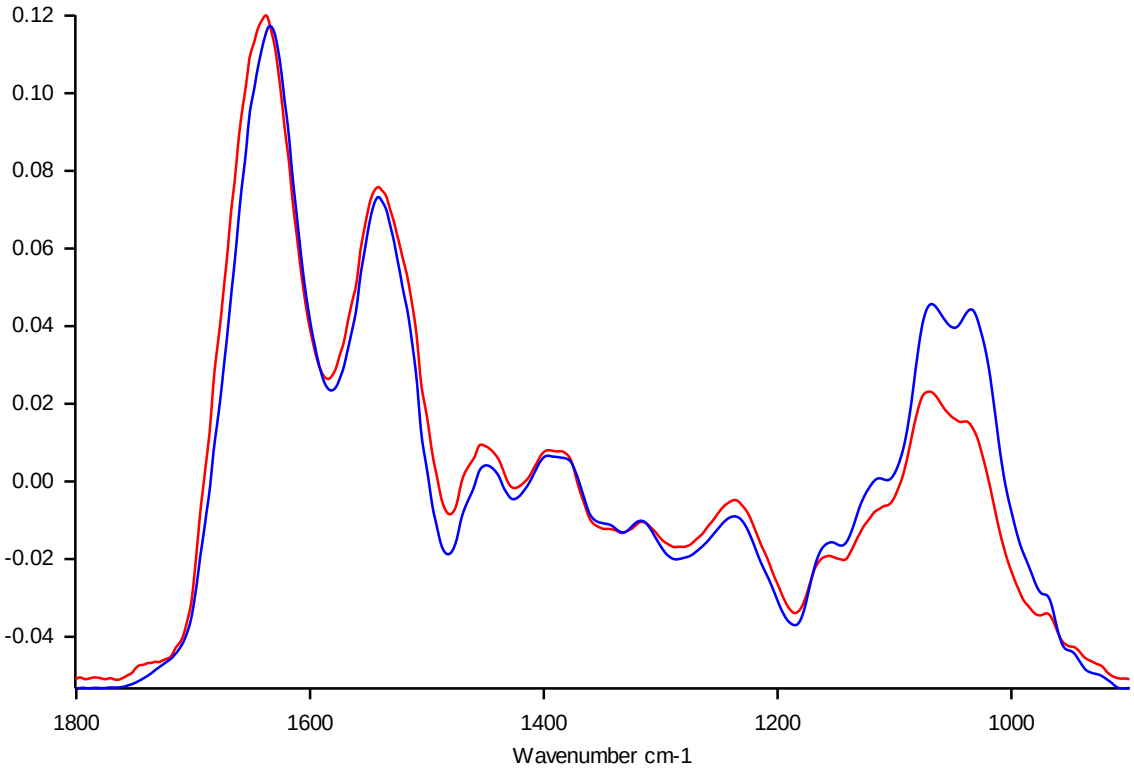


**Figure A4-114: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR20 dried onto the sampling strip (red), and the ATR crystal (blue)**

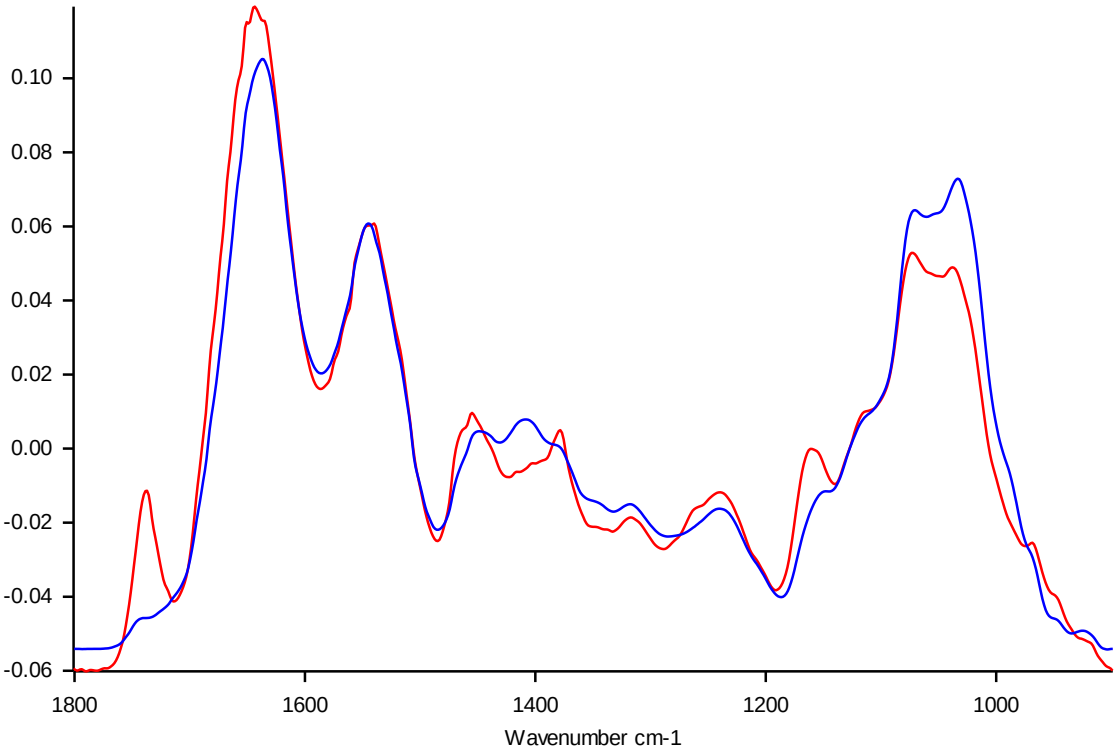


**Figure A4-115: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR21 dried onto the sampling strip (red), and the ATR crystal (blue)**

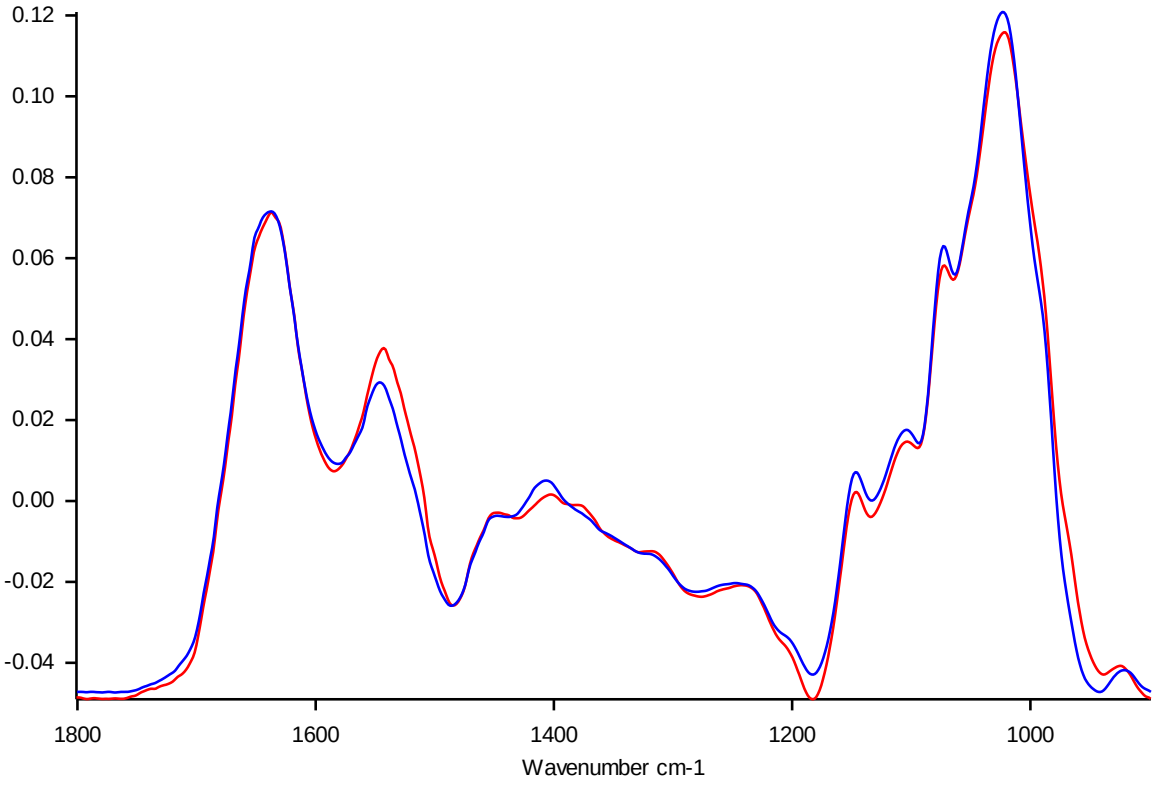


**Figure A4-116: Baseline-corrected, vector-normalised average ATR-FTIR spectra from 1800:900cm-1 of COPD sputum sample TR22 dried onto the sampling strip (red), and the ATR crystal (blue)**