

Georgia State University

ScholarWorks @ Georgia State University

Physics and Astronomy Dissertations

Department of Physics and Astronomy

Spring 5-4-2021

Infrared Spectroscopy of Serum Samples for Disease Diagnostics

Hemendra Mani Ghimire*

Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/phy_astr_diss

Recommended Citation

Ghimire*, Hemendra Mani, "Infrared Spectroscopy of Serum Samples for Disease Diagnostics." Dissertation, Georgia State University, 2021.
doi: <https://doi.org/10.57709/22547351>

This Dissertation is brought to you for free and open access by the Department of Physics and Astronomy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Physics and Astronomy Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

INFRARED SPECTROSCOPY OF SERUM SAMPLES FOR DISEASE DIAGNOSTICS

by

HEMENDRA MANI GHIMIRE

Under the Direction of A. G. Unil Perera, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

The fundamental vibrational modes of biological constituents in the tissues and the complex body fluids coincide with optical frequencies in the infrared region. Therefore, spatially resolved molecular compositions and interaction information within the biological materials can be extracted non-destructively using IR radiation without the use of labels or probes. However, the feasibility of this technique to elucidate constituent molecular compositions and interactions within the diagnostic mediums is not well explored. This study demonstrates an application of infrared (IR) spectroscopy of sera for monitoring inflammatory bowel diseases (IBD) and various cancers. Using samples from experimental mice and human patients, the power of IR spectroscopy in structural studies of proteins and other complex band contours are explored to find spectral signatures.

Two experimental models of IBD; interleukin 10 knockouts (IL10^{-/-}) and Dextran Sodium Sulfate (DSS) induced mouse shows diagnostic accuracy with 80-100% sensitivity and specificity values. Importantly, the findings of human IBD patients' serum also show promising results resembling our proofs-of-concept investigations of mouse models. Maximum values of sensitivity and specificity are 100% and 86%, respectively, in human samples. Similarly, in cancer studies, the EL4 mouse model of non-Hodgkin lymphoma (NHL) and a B16 mouse model of the subcutaneous melanoma are used to extract a snapshot of tumor-associated alteration in the serum. The study of both cancer-bearing mouse models in wild types (WT) and their corresponding control types emphasizes the diagnostic potential of this approach as a screening technique for the NHL and melanoma skin cancer. The breast cancer (BC) -associated protein conformational alteration in the serum samples shows the sensitivity and the specificity of identifying spectral signatures were both 90%. All in all, IR spectroscopy of serum samples accompanied by spectral

analysis technique shows some promising results for disease diagnostics. The brief outlook of the fundamentals of the infrared detection technique and their applicability for the development of portable spectroscopy is also provided.

INDEX WORDS: Breast cancer, Inflammatory bowel diseases, Infrared spectral signatures, Lymphoma, Melanoma, Protein secondary structures

Copyright by
Hemendra Mani Ghimire
2021

INFRARED SPECTROSCOPY OF SERUM SAMPLES FOR DISEASE DIAGNOSTICS

by

HEMENDRA MANI GHIMIRE

Committee Chair: A. G. Unil Perera

Committee: Ritu Aneja

Vadim Apalkov

Mukesh Dhamala

Yuan Liu

Didier Merlin

Gengsheng Qin

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2021

DEDICATION

I would like to dedicate this dissertation to my family in Nepal. My parents Kunjar Mani Ghimire and Shobha Ghimire for their love and support. My wife Susmita Gautam Ghimire, for being next to me all the time. Without her support, this dissertation would not have been possible. My dear friends in Nepal and America, please forgive me for not being able to mention my love individually in this small piece of paper. Thanks to God for all the blessings.

ACKNOWLEDGMENTS

Foremost, I would like to express my deepest appreciation to my supervisor Dr. A. G. Unil Perera for introducing me to biophysics and optoelectronics researches. I highly appreciate his guidance and constant support throughout my research period at Georgia State University. He has always remained a source of inspiration to focus and to be a better researcher.

Besides my advisor, my immeasurable appreciation and deepest gratitude go to the members of my dissertation committee: Dr. Didier Merlin, Dr. Gengsheng Qin, Dr. Mukesh Dhamala, Dr. Ritu Aneja, Dr. Vadim Apalkov and Dr. Yuan Liu for finding the time to be on my dissertation committee and giving me insightful comments. Their effort in reading, providing comments and managing time for fruitful discussion is priceless, not only for this dissertation, but also for the motivations in my career.

I would like to thank our collaborators: Dr. Didier Merlin, Dr. Ritu Aneja and Dr. Yuan Liu at the Institute of Biomedical Science; from the Biology Department; Dr. Gengsheng Qin from the Mathematics and Statistics department of our University. My sincere acknowledgment to collaborators from other institutions: Dr. Christopher S. Hourigan at Myeloid Malignancies Section, Hematology Branch, National Heart, Lung, and Blood Institute, Bethesda, MD, USA; Emiel M Janssen from Department of Pathology, Stavanger University Hospital, Stavanger, Norway; Dr. Marlina Catherine Kruger at the Department of Nutritional Science, School of Food and Advanced Technology, College of Sciences, Massey University, Tennent Drive, Palmerston North 4442, New Zealand; and Uma Krishnamurti from Department of Pathology, Emory University School of Medicine, Atlanta, GA, USA. They have provided us precious samples and/or necessary guidelines for sample handling, measurements and data analysis. I would like to express my appreciation to Dr. Mathes A. Kanishka Dayananda from the Georgia State University

perimeter college for the fruitful discussion on measurement and data analysis techniques. I am grateful to all the physicists, chemists, biologists, veterinarians, and technicians who contributed their expertise in my research work reported in this thesis.

I would like to convey special acknowledgment to former and present lab members Dr. Jito Titus, Dr. Seyoum Wolde, Dr. Dilip Chauhan, Dr. Yanfeng Lao, Kelum Perera, and Chandrama Mukharjee for valuable discussions on the subject matter. I also want to acknowledge P. V. V. Jayaweera at SPD Laboratory, Inc., Hamamatsu, Japan, who was also a former lab member, for his idea to develop the possibility of μ -controllers applicable in portable device development. My sincere thanks go to all the professors, colleagues, and staff of the Department of Physics and Astronomy. The work presented in this thesis would not have been possible without my close association with them.

Finally, I would like to thank for funding agencies as: this material is based upon work supported in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under contract/grant numbers W911NF-15-1-0018 and W911NF-12-2-0035. The molecular basis of the disease (MBD) program at GSU. Also, the dissertation grant award program at GSU, which is sponsored by the Provost and the Vice President for Research & Economic Development.

TABLE OF CONTENTS

ACKNOWLEDGMENTS		V
LIST OF TABLES		XI
LIST OF FIGURES		XIV
LIST OF ABBREVIATIONS		XXV
1 INTRODUCTION		1
1.1 Fourier Transforms Infrared (FTIR) Spectroscopy		2
<i>1.1.1 Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR)</i> <i>Spectroscopy</i>		5
1.2 ATR-FTIR Spectroscopy of Serum Samples for IBD Studies		7
<i>1.2.1 Spectroscopy of Serum Samples from Experimental Mouse Models</i>		9
<i>1.2.2 Spectroscopy of Serum Samples from Ulcerative Colitis (UC) Patients</i>		11
1.3 ATR-FTIR Spectroscopy of Serum Samples for Cancer Studies		11
<i>1.3.1 Spectroscopy of Serum Samples from the Mouse Models of Lymphoma and</i> <i>Melanoma</i>		12
<i>1.3.2 Spectroscopy of Serum Samples from Breast Cancer (BC) Patients</i>		13
1.4 Protein Conformational Studies using FTIR Spectroscopy		14
2 EXPERIMENT		17
2.1 FTIR Spectrometer		17
2.2 Animal Growth and Sample Collection		18

2.2.1	<i>Dextran Sodium Sulfate (DSS) Mouse</i>	18
2.2.2	<i>Interleukin 10 Knockout (IL10^{-/-}) Mouse</i>	18
2.2.3	<i>Collagen Antibody-Induced Arthritis Mouse</i>	20
2.2.4	<i>Toll-Like Receptor Knockout (TLR5^{-/-}) Mouse Model of Metabolic Syndrome</i> . 20	
2.2.5	<i>Human Sera of IBD Patients</i>	20
2.2.6	<i>B16 Melanoma or EL4 Lymphoblast Tumor Mouse Models</i>	20
2.2.7	<i>Human Sera of Breast Cancer (BC) Patients</i>	21
2.3	Spectral Response Measurements	21
2.4	Data Analysis	22
2.4.1	<i>Hierarchical Cluster Analysis (HCA) and Dendrogram Plot</i>	22
2.4.2	<i>Principal Component Analysis (PCA)</i>	23
2.4.3	<i>Spectral Fitting (Deconvolution) for Protein Conformational Studies</i>	23
2.4.4	<i>Statistical Analysis</i>	25
3	RESULTS	27
3.1	Identifying Infrared Spectral Signatures of IBD in Serum	27
3.1.1	<i>Spectral Signatures in Serum of DSS Mouse</i>	27
3.1.2	<i>Spectral Signatures in Serum of IL10^{-/-} Mouse</i>	30
3.1.3	<i>Spectral Signatures in Serum of Human Colitis Patients</i>	39
3.2	Spectral Signatures of Lymphoma and Melanoma in Serum Samples of Mouse Models	47

3.3	Spectral Signatures of BC in Serum of Human Patients.....	55
4	DISCUSSION	64
4.1	Discussion Based on DSS Study	64
4.2	Discussion Based on IL10-/- Study	65
4.3	Discussion Based on Colitis Human Patients Study	66
4.4	Potential Prototype for the Portable Device Development.....	69
4.5	Discussion Based on Lymphoma and Melanoma Study	74
4.6	Discussion Based on BC Study.....	75
5	CONCLUSION	80
	REFERENCES.....	82
	APPENDICES.....	94
	Appendix A	94
	<i>Appendix A.1 Understanding the Clinical Feasibility for IBD Screening.....</i>	<i>94</i>
	<i>Appendix A.2 Penetration depth of infrared light in the sera</i>	<i>98</i>
	<i>Appendix A.3 Sample Size Estimation for Diagnostic Accuracy</i>	<i>100</i>
	<i>Appendix A.4 Quantified Values at Spectral Signatures of UC Patients.....</i>	<i>100</i>
	Appendix B	1033
	<i>Appendix B.1 Information on BC Patients</i>	<i>103</i>
	Appendix C	1055
	<i>Appendix C.1 Research Protocol.....</i>	<i>1055</i>

Appendix D	1211
<i>Appendix D.1 Second Derivatives Spectra and Their Importance</i>	<i>1211</i>
Appendix E	123
<i>Appendix E.1 Number of GFEBs (Oscillators)</i>	<i>943</i>
<i>Appendix E.2 Estimation of Band's Position Using Second Derivatives Spectra Fitting</i>	<i>127</i>

LIST OF TABLES

Table 3.1 Discriminatory Infrared spectral bands of UC, with biomolecular assignments and their bond vibrations[6].	32
Table 3.2 Quantified values of identifying spectral signatures: I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , integral ratio (α/β), and the sum of the integral area of GFEBs used to fit the experimental complex band 1000-1140 cm^{-1} [6]. Tabulated p-values represent a comparison between control and colitis.	35
Table 3.3 Sensitivity, specificity and AUC value calculations of identifying spectral signatures of UC in IL10 ^{-/-} mouse[6].	37
Table 3.4 Quantified information about average integral values and average position of α -helix and β -sheet secondary structure, energy bands in human serum samples.	44
Table 3.5 Quantified information about identifying discriminatory bands. These include absorbance values at 1033 cm^{-1} and 1076 cm^{-1} , the ratios of absorbance at 1121 cm^{-1} to 1020 cm^{-1} , 1629 cm^{-1} to 1737 cm^{-1} . The integral ratios of Gaussian function energy bands representing α -helix and β -sheet protein secondary structures are also discussed. Statistical measures of discriminatory infrared spectral signatures for UC using sera shown in the shadowed region of the table. The optimal cut-off and the corresponding sensitivity and specificity are calculated based on the Youden index. P-values are calculated based on the student's t-test with a two-tailed unequal variance.	47
Table 3.6 Discriminatory infrared spectral bands of lymphoma and melanoma serum with biomolecular assignments[12].	50
Table 3.7 Quantified values of discriminatory features. Clear separation can be seen between control and cancerous cases (both B16 and EL4) while comparing integral values of α -	

helix, β -sheet structure components, and their ratios. Similarly, the altered position of amide I peak, amide I/amide II ratio and absorbance values at 1212 and 1335 cm^{-1} show significant difference only between EL4 and control[12].	55
Table 3.8 Discriminatory IR bands for BC serum samples from controls, and primary biomolecular assignments, giving rise to the main contributions for the absorbance. Amide regions and the complex region of carbohydrates and nucleic acids show the discriminating potential[13]......	59
Table 5.1 Summary of identifying spectral signatures.....	81
Table A1 Estimated human sample sizes for continuous test using ROC (FPF0) as the basis for inference with $\alpha = 0.05$ and $\beta = 0.10$. Simulation studies were based on the binormal ROC curve with slope parameter 1 and the initial values based on asymptotic theory.	100
Table A2 Quantified values of spectral signatures of UC patients. Shaded region of the table highlights the data extracted from male patients and unshaded are from female. Variability of these signatures (I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} and α/β) does not show age, gender, height, weight, smoking and alcohol consumptions dependence in the uniform pattern, so a concluding remark cannot be made.....	100
Table A3 Quantified values of control samples at spectral marker signatures. Shaded region of the table highlights the data extracted from male and unshaded are from female controls. Variability of these signatures(I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} and α/β) does not show age and gender dependence.	1012
Table A4 Information about the patients: Biopsy results such as Invasive lobular carcinoma (ILC), Invasive ductal carcinoma (IDC), Ductal Carcinoma In-Situ (DCIS) confirmed the BC diagnoses. The tumor grading assessment was performed based on the national cancer	

institute (NCI) guidelines. Grade-2 denotes the intermediate grade or moderately differentiated tumor whereas grade-3 denotes a high grade or poorly differentiated tumor.

..... 103

LIST OF FIGURES

- Figure 1.1 (A)** The schematic of FTIR spectroscopy. The infrared (IR) detector receives an interferogram and our computer does the Fourier transformation to generate the absorbance spectrum. It simultaneously collects data over a wide spectral range, reflecting how much light is absorbed by a sample at each wavelength. **(B)** Different sampling modes of FTIR spectroscopy, including transmission, transflection, and the attenuated total reflection (ATR)..... 3
- Figure 1.2** Interaction of light with sample in ATR sampling mode of FTIR spectroscopy[11]. **(A)** Schematic diagram of the evanescent wave formed on the sample at the internal reflection element surface. **(B)** Interaction of infrared signal in the material of the medium..... 6
- Figure 2.1** The process of spectral deconvolution[6]. **(A)** Second derivative spectra of amide I absorbance curve 1600-1700 cm^{-1} . **(B)** Gaussian function energy bands used to obtain a curve that fits the experimental absorption curve. **(C)** Second derivative spectra of nucleic acids and carbohydrates, complex band 1000-1140 cm^{-1} . **(D)** Deconvolution of the complex band by estimating the number and position using the minima of the second derivative curve..... 24
- Figure 3.1 (A)** Averaged (n=6) ATR-FTIR spectra of serum samples derived from DSS mice on day 0, day 3, and day 7 of the chemical feeding. The variation in infrared absorbance on day 7 is higher in the complex region of carbohydrates and nucleic acids compared to on day 3. **(B)** Second derivative curves of the absorbance spectra. Second derivative spectra shows the variation on day 7 in the amide I region. Herein, shaded regions show two major regions of interest[11]. 28

Figure 3.2 Longitudinal analysis of serum samples of DSS induced colitis mice[11]. **(A)** Normalized absorbance values in the range 1000-1140 cm^{-1} showing changes in spectral signatures representing carbohydrates and nucleic acids are significant at day 7, but not at day 3. **(B)** The integral ratio of Gaussian function energy bands representing α -helix and β -sheet secondary structures of proteins in serum samples extracted from six mice. The secondary structure analysis also shows the significant changes only at Day 7..... 29

Figure 3.3 Normalized absorbance curve and p-value calculation results[6]. **(A)** Averaged ATR-FTIR spectra of serum samples, derived from control (Non-Colitis, n=16), treated with anti-TNF α antibody (Treated, n=9) and colitis untreated (Colitis, n=9) mice. These spectra show proper anti-TNF α therapy ameliorate the colitis condition. Inset shows magnification within 1200-1000 cm^{-1} and 1660-1625 cm^{-1} , which clearly shows how the absorbance curve of anti-TNF α treatment comes close to the control levels. **(B)** P-value calculation of colitis and treated conditions with control types. The p-value is less than 0.05 at various spectral bands while comparing non-colitis and colitis, but it is always greater than 0.05 while comparing non-colitis and treated. The region highlighted by the pink box is the region with $p \leq 0.05$ 31

Figure 3.4 Representation of identifying spectral signatures[6]. **(A)** Ensemble average representative of absorbance at wavenumber position 1033 cm^{-1} . **(B)** Ensemble average representative of absorbance at wavenumber position 1076 cm^{-1} . **(C)** The ratio of absorption values at wavenumber 1121 cm^{-1} to 1020 cm^{-1} . **(D)** The ratio of absorbance at wavenumber 1629 cm^{-1} to 1737 cm^{-1} . **(E)** The integral values of Gaussian function energy bands representing the α helical structure of protein secondary structure. **(F)** The integral values of Gaussian function energy bands representing the β pleated sheet structure of

protein secondary structure. **(G)** The integral ratio of α -helix and β -sheet protein secondary structures. It shows the increased level of β sheet structures and a decrease in α helix structures due to colitis and their resettlement in anti- TNF α therapy mice. **(H)** The sum of the integral area of GFEBs used to fit the experimental curve in the complex band 1000-1140 cm^{-1} 34

Figure 3.5 (A) The expression levels of TNF α in the Total RNAs extracted from colons. The level of TNF α is quantified by qPCR and it confirms the effect of anti-TNF α therapy. **(B)** ROC curve for the data obtained from the ratio of the integral area of energy bands representing α -helix and β -sheet protein secondary structures. AUC is 0.95 while comparing colitis and non-colitis, but its value is 0.46 while comparing non-colitis and the treated. High diagnostic accuracy (with AUC = 0.95) can be seen for colitis, but poor diagnostic accuracy (with AUC =0.46) for the treated[6]. 36

Figure 3.6 Infrared absorption spectra of sera to discriminate control and UC conditions. **(A)** Ensemble average representative of min-max normalized absorbance spectra of 28 control serum samples (healthy volunteers) and 18 UC patients indicating biomolecular assignments. **(B)** The spectral response difference between UC and control. The average absorbance values of UC conditions are elevated at carbohydrate and nucleic acids regions and fluctuating up and down in the Amides region. The difference between repeat measurements of the same sample (Multiple Controls and Multiple UCs) shown by the (— the blue dashed and ••• the red dotted lines respectively) lines are negligible compared to the difference spectra between the average UC and average control (— green; UC-Control). **(C)** p-values of the Student's t-test (with two-tailed unequal variance) for the absorbance representing healthy volunteers and UC patients. p-values less than 0.05 are

used to highlight the strongest discriminatory features of the spectrum. p-values less than 0.025 (horizontal pink-dotted line) show the validity of all the features except lipids.

..... 41

Figure 3.7 Plots of integral values of Gaussian function energy bands representing α -helix, β -sheet protein secondary structures in human sera. **(A)** Quantified integral values of α -helix components. **(B)** The integral values of β -sheet components. **(C)** Ensemble average representation of integral values of α -helix. P-values for the integral area of α -helix is 0.05 and for the mean position is 0.03. **(D)** Ensemble average representation of integral values of the β -sheet. P-values for the integral area of β -sheet is $9E-07$ and mean position is 0.3. **(E)** Representation of integral ratios between corresponding α -helix and β -sheet structures. Average values of control and UC cases for the integral ratio of α -helix to β -sheet (Integral (α/β)) shows a statistically significant difference with a p-value of $1E-04$. **(F)** The ROC curve of the Integral (α/β) and the corresponding AUC. 43

Figure 3.8 Histogram representation of identifying spectral signatures. **(A)** Ensemble average representative of absorbance of wavenumber position 1033 cm^{-1} . **(B)** Ensemble average representative of absorbance of wavenumber position 1076 cm^{-1} . **(C)** The ratio of absorbance values (I_{1121}/I_{1020}) at wavenumber 1121 cm^{-1} to 1020 cm^{-1} . **(D)** The ratio of absorbances (I_{1629}/I_{1737}) at wavenumber 1629 cm^{-1} to 1737 cm^{-1} 45

Figure 3.9 The discriminatory region of infrared absorbance spectra. **(a)** Average normalized ATR-FTIR spectra of serum samples extracted from EL4-lymphoma (n=8), B16-melanoma (n=8) mouse models in wild types and corresponding control types (n=15). The inset (i) shows B16-melanoma mouse with tumor size approximately 1000 mm^3 (day 18). An increase in volume of the tumor from day 9 to day 18 of tumor inoculation in B16-

melanoma mouse is as in inset (ii). Like the B16 mouse, the elevation of tumor size is also monitored in EL4-lymphoma. The serum sample is extracted for both types of mice when tumor size becomes bigger than 1000 mm³. **(b)** Student's t-test (two-tailed unequal variance) p-values of absorbance. Discriminatory region for lymphoma with higher significance ($p < 0.05$) are amide I of protein, amide II of protein, C-H bends of CH₃/CH₂ groups in α - and β - anomers, asymmetric phosphate I, and carbohydrates with predominant contributions nucleic acids (DNA/RNA via PO₂⁻ stretches). Discriminatory regions of melanoma are amide I and carbohydrates with predominant contributions of nucleic acids[12]..... 49

Figure 3.10 Plots of the protein secondary structures (α -helix, β -sheet) and their ratio[12]. **(a)** Quantified integral (area covered) values of α -helix components are less for tumorous cases compared to control. **(b)** Integral values of β -sheet components are higher for tumorous cases compared to control **(c)** Bar graph representation of average integral ratios between α -helix and β -sheet for control, B16 and EL4. Significant alteration in integral ratio (α -helix/ β -sheet) is found between control and tumorigenic case. 51

Figure 3.11 Dendrogram of hierarchical cluster analysis[12]. The dendrogram tree diagram performed within spectral range 1140-1000 cm⁻¹, by using Ward's algorithm and squared Euclidian distance measurements. The spectra are correctly classified. Control spectra appear grouped together, which describes a high degree of similarity within the groups. Similarly, there is a distinct clustering in the cancer spectra in two subgroups showing the higher degree of heterogeneity between cancerous spectra. 53

Figure 3.12 The integral sum of Gaussian energy profiles used to fit experimental curves within 1000-1140 cm⁻¹. **(a)** The sum of integral values of control groups cluster within the

approximate range 12-14, B16 covers the range 15-17 and EL4 covers 15-18. **(b)** Bar graph representation of the average value of the integral sum which shows significant differences between control and tumorigenic case[12]. 54

Figure 3.13 PCA analysis of second derivatives curves of FTIR absorbance spectra[13]. **(A)** PCA scores plots (PC1 x PC2) with 95% confidence ellipse. The data related to control groups (• • black dots enclosed by a black shaded ellipse) and BC (▪ ▪ red dots surrounded by a red-shaded ellipse) are clustered together with different magnitudes and directions. **(B)** Scree plot of eigenvalues showing the percentage variance of components one and two is significant compared to others. 56

Figure 3.14 Identification of discriminatory bands. **(A)** Ensemble averages of normalized serum spectra derived from control, $n = 10$ and BC, $n = 10$. This wider range of spectra is presented to show the quality of the spectra, which overcomes the noise and atmospheric contamination while measuring them at resolution 4 cm^{-1} . **(B)** Corresponding student t-test p-values for the control and BC. **(C)** The second derivative absorbance spectra confined to the amide-I region, covering $1600\text{-}1700 \text{ cm}^{-1}$. **(D)** Difference between the absorbance spectra of control and BC, indicating up- and down-regulation of proteins, carbohydrates, and nucleic acids in the serum of BC patients[13]. 58

Figure 3.15 Protein secondary structure analysis[13]. **(A)** Representative second derivatives of absorbance spectra at the amide-I absorbance region. **(B)** Deconvolution of the amide-I region: the baseline-corrected spectra fitted with 6 GBEF by approximating number and position of minima of second derivatives, which simulated fits (•••) to the experimental curve (—). **(C)** Integral area of GBEF representing α helix and β sheet. **(D)** The ratio of α helix and β sheet energy bands, which proves an elevation of β sheet and drop off α helix

structures due to malignancies. **(E)** The ratio of IR absorbance at amide II (I_{1556}) to its value at amide III (I_{1295}). **(F)** ROC curves for the ratio of the integral area of energy bands representing α -helix and β -sheet protein secondary structures and the respective absorbance at amide II and amide III. The maximum values of sensitivity and specificity are 90% and 90% for signature α/β , while these values are 100% and 80% for signature I_{1556}/I_{1295} , respectively..... 60

Figure 3.16 Understanding protein secondary structures and physics of IR interaction. **(A)** Parallel β -pleated sheet structure of proteins. N-H groups in the backbone of one strand form hydrogen bonds with the C=O groups in the backbone of the adjacent strand to form a β -sheet. **(B)** Right-handed α -helix structures of proteins. The backbone N-H group donates a hydrogen bond to the backbone C=O group, contributing to the helical structure of the α -helix[13]..... 62

Figure 4.1 Structure of the software program showing 9 subroutines to perform each specific task[11]. Each of these specific programming tasks was individually carried out in our preliminary work in order to get the conclusion. In the portable device development, all those will be automated into one program which will be done at a single touch command. 69

Figure 4.2 Flow chart for the programming to study protein secondary structures[11]. Only the α -helix and β -sheet ratio analysis is shown in the flow chart. Similarly, other biomarker analysis will also be in the program as needed to achieve the needed confidence limit. . 70

Figure 4.3 Schematic of measurements and data analysis tool permanently integrated into the anticipated modified spectroscope. "Sample Scan" tab allows users to enter the patient ID, select the disease followed by pressing "Start Scan" to initiate the program[11]..... 71

Figure 4.4 Anticipated portable device with the touch screen showing "Test Report". The test report can be printed or save the patient record in the hospital database[11]. 73

Figure 4.5 (A) Deconvolution of the complex band of carbohydrates and nucleic acids 1000-1140 cm^{-1} . The number and position of six bands used to fit the experimental curve were determined by using the minima of secondary curves as in amide I case. **(B)** Bar graph representation of the average value of the integral sum which shows a significant difference between control and BC case. **(C)** Bar graph of average absorbances at wavenumber position 1020 cm^{-1} , which is mainly due to the presence of DNA. It also shows a significant difference between control and BC cases[13]. 77

Figure 4.6 Schematic of ATR-FTIR spectrometer integrated with two micro-controllers (micro-processors) A and B. The controller A extracts the information about the signal-sample interaction, while controller B stores the spectral analyzing software application in the clinical domain[13]. 78

Figure A1 (A-H) Microscopic images of one microliter serum sample during different stages of drying at room temperature. Pictures (A) to (H) indicate the image of the serum in 2-minute intervals starting with (A) 4 minutes after its deposition and (H) 18 minutes after deposition. (I) FTIR spectra showing variation in absorbance during the air-drying of the serum samples. Spectra representing 1 is the first spectra taken after 4 minutes of sample deposition, 2 is second spectra (~ 5.2 minutes), 3 is third spectra (~ 6.4 minutes). Overlapping spectra (solid blue \rightarrow) indicate the stabilization of the sample. Inset I(i) and I(ii) show the magnified view of spectral range 3225-3350 cm^{-1} and 1610-1670 cm^{-1} to highlight spectral reproducibility nearly after 10 minutes of sample deposition. 966

Figure A2 (A) Spectral discrimination by student's t-test (with two-tailed unequal variance) p-values of normalized absorbance using 12 measurements of the same sample at two different times. P-values $\gg 0.05$ clearly show the similarity between these spectra verifying better measurement-to-measurement reproducibility of the technique. (B) P-values calculated for the absorbance data measured by three different researchers. Significantly, similar p-values ($p \geq 0.18 \gg 0.05$) are found while comparing data from individual researchers, verifying the user-to-user consistency. 988

Figure A3 (A) Light-sample interaction and formation of evanescent wave in ATR sampling mode. (B) Output spectra of $1\mu\text{l}$ of serum samples using diamond and Silicon crystals. 9999

Figure A4 Verifications of mathematical formula for normalization. (A) Min-max normalization of absorption spectra. (B) Vector normalization of second derivative curves of the spectra. Spectra obtained from the formula overlaps with the corresponding spectra, as obtained from the OPUS software. It verifies, these formulas can be used while writing code in another programming language. 19914

Figure A5 Spectral data before and after smoothing. 9916

Figure A6 Dendrogram tree diagram performed, by using Ward's algorithm and Euclidian distance measurements of vector normalized second derivatives curves within spectral range $1600\text{-}1700\text{ cm}^{-1}$. (A) A wide difference can be seen between spectra belonging to control and colitis mouse, compared to the degree of the dissimilarity between controls and treated while considering three groups constraints (1 for control: c1 to c9, 2 for treated: t1 to t9, and 3 colitis: d1 to d9). (B) Stratigraphic constraints of three groups showing colitis

group has larger heterogeneity from control and treated groups. While there is a similarity between control and treated groups in this classification. 1997

Figure A7 The process of rubber band baseline correction. The sectioned region 1600-1700 cm^{-1} is baseline corrected using baseline correction, such that the absorbance at the two extreme ends is 0..... 119

Figure A8 Derivatives spectra and their importance for spectral analysis. (A) Gaussian function distribution. (B) First derivatives spectra and (C) Second derivatives spectra. The minima position of second derivatives spectra exactly represent the maxima of Gaussian distribution..... 121

Figure A9 (A) Second derivative of ATR-FTIR absorbance of control sera at resolution 8 cm^{-1} , showing the deconvoluted amide I region. (B) The individual secondary structure components were modeled using GFEB whose positions were determined from the second derivative of the absorbance to obtain simulated fits to the experimental curves. (C) Secondary structures were modeled using GFEB with the inclusion of random coil components approximately at wavenumber position 1641 cm^{-1} [4]..... 124

Figure A10 (A) Second derivative of ATR-FTIR absorbance of control sera at resolution 4 cm^{-1} , showing amide I, region fitting. (B) The individual secondary structure components were modeled using GFEB whose positions were determined from the second derivative of the absorbance to obtain simulated fits to the experimental curves. (C) Secondary structures were modeled using GFEB with the inclusion of random coil components..... 125

Figure A11 The integral ratio of α -helix and β -sheet protein secondary structures using five and six bands on experimental curves at resolution 4 cm^{-1} . (A) Ratio of integral values while

fitting experimental curve without random coils structures. **(B)** Ratio of integral values while fitting experimental curves with including random coils structures. 126

Figure A12 Second derivative curve fitting technique. (A) Second derivative spectra. (B) Values of second derivative spectra subtracted from one at each wavenumber position. (C) Rubber band correction of the curve obtained from one minus second derivative. (D) GFEBs fitting to estimate the position of the bands. 128

LIST OF ABBREVIATIONS

- 2D-DIGE: Two-dimensional Differential Gel Electrophoresis
- α -helix: Alpha helix
- β -sheet: Beta sheet
- ASSURED: Affordable, Sensitive, Specific, User-friendly, Robust and rapid, Equipment-free, and Deliverable
- ATCC: American Type Culture Collection
- ATR: Attenuated Total Reflectance
- ATR-FTIR: Attenuated Total Reflectance Fourier Transforms Infrared
- AUC: Area Under Curve
- BALB/C: Bagg and Albino Laboratory-bred strain of house mouse with C sub-strain
- BC: Breast Cancer
- BMI: Body Mass Index
- CAIA: Collagen Antibody Induced Arthritis
- CD: Crohn's Disease
- CEA: Cancer Embryonic Antigen
- DMEM: Dulbecco's Modified Eagle Medium
- DNA: Deoxyribonucleic Acids
- DSS: Dextran Sodium Sulfate
- ER: Estrane Receptor
- FBS: Fetal Bovine Serum
- FPF: False Positive Fraction
- FTIR: Fourier Transforms Infrared

FWHM: Full Width Half Maximum

GI: Gastrointestinal

GFEB: Gaussian Function Energy Bands

HCA: Hierarchical Cluster Analysis

HER2: Human Epidermal Growth Factor Receptor-2

IACUC: Institutional Animal Care and Use Committee

IARC: International Agency for Research on Cancer

IBD: Inflammatory Bowel Diseases

IDC: Invasive Ductal Carcinoma

IL10-/-: Interleukin 10 knockouts

ILC: Invasive Lobular Carcinoma

IRE: Internal Reflection Element

Lcn-2: Lipocalin 2

miRNAs: microRNAs

MPO: Myeloperoxidase

NCI: National Cancer Institute

NGAL: Neutrophil Gelatinase-associated Lipocalin

NHL: Non-Hodgkin's Lymphoma

NMR: Nuclear Magnetic Resonance

PAST-4: PAleontological Statistics-4

PC: Principal Components

TLR5: Toll-like Receptor 5

PCA: Principal Component Analysis

PR: Progesterone Receptor

qPCR: Quantitative Polymerase Chain Reaction

RMS: Root Mean Square

RNA: Ribonucleic Acids

ROC: Receiver Operating Characteristic

TLR5: Toll-Like Receptor 5 knockouts

TNF: Tumor Necrosis Factor

TPF: True Positive Fraction

WT: Wild Type

1 INTRODUCTION

Spectroscopic techniques are becoming popular exploratory tools to investigate the interaction between matter and electromagnetic radiations[1]. The popularity is due to the extraordinary sensitivity of spectroscopy to understand the molecular structures depending on bonding between the constituent atoms and their spatial arrangements. Based on the nature of the measurement process, radiative energy involved in the interaction and their applications, these techniques are classified into several types[2]. These include classification based on atomic or molecular probing, electronic or magnetic features, Gamma to terahertz radiation, absorption or emission of radiations, and many more. The selection of an individual spectroscopic technique depends on the purpose of the study. Among these several spectroscopic techniques, the infrared spectroscopy is applicable to investigate vibrational and the rotational motions of molecular bonds and the spatial arrangement of the constituent atoms[3]. The usefulness of infrared spectroscopy arises due to the unique spectral fingerprint of each molecule type.

In this dissertation, infrared spectroscopy is utilized to investigate the structural and compositional features of constituents in serum samples and disease-associated alterations within them. Accompanied by data analysis techniques, Fourier transforms infrared (FTIR) spectroscopy of sera has been explored to find disease-associated alterations in the constituent's biomolecules[3]. The FTIR uses interferometry to record information about the material and these interferograms are being decoded into recognizable spectra via Fourier transformations. The first part of the dissertation investigates the feasibility of the technique to understand inflammatory bowel diseases (IBD) associated alteration in the serum samples of experimental mice[4-11] and human patients. In experimental models, acute (Dextran Sodium Sulfate induced: DSS) and chronic (Interleukin 10 knockouts; IL10^{-/-}) mice are tested. Arthritis (Collagen Antibody Induced

Arthritis: CAIA) and metabolic syndrome (Toll-Like Receptor 5 knockouts: TLR5) models are also tested as controls[4, 5]. The second part of the dissertation investigates its potential to discriminate cancer-associated changes in the serum samples[12, 13]. In the cancer study, an EL4 mouse model of non-Hodgkin lymphoma (NHL) and a B16 mouse model of the subcutaneous melanoma are used[12]. The samples of Breast cancer (BC) patients were also used to investigate the feasibility of the technique[13]. An outlook on the ongoing study of the standardization of technique for the acquisition of spectral signatures applicable in the clinical domain and preparing unique database is also discussed.

1.1 Fourier Transforms Infrared (FTIR) Spectroscopy

The interaction of light with matter depends on the nature of the material, the wavelength of the incident light, and the intensity[14]. Within the infrared region of the electromagnetic spectrum, as light interacts with matter and the chemical bonds will stretch, contract and bend. It is because different bonds and functional groups of constituents have a range of quantized vibrations and bending modes. When these constituent molecules are irradiated with infrared photons matching the energy of those vibrational energy levels, molecules absorb the radiation and start to vibrate. As a result, the amplitude of the vibration (which depends on intensity) is increased, but it does not affect the frequency of those vibrations and remains the same[15]. Therefore, a chemical functional group tends to adsorb infrared photons/radiation in a specific wavelength region, regardless of the structure of the rest of the molecule. The analytical technique, which measures the vibration characteristics (i.e. Infrared intensity versus the wavelength or wavenumber of light) of chemical functional groups in the sample is called infrared spectroscopy[3].

FTIR spectrometer obtains infrared spectra by first collecting an interferogram of the sample that measures IR frequencies simultaneously, and digitizes it followed by performing the Fourier transformation[16]. It takes advantage of how IR light changes the dipole moments of molecules at the specific vibrational energy ($E = hv = hc/\lambda = hc\bar{\nu}$, where c is the velocity of light and h is Planck's constant) corresponding to reduced mass (μ) and the bond spring constant (k)[17];

$$\bar{\nu} = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad \dots\dots\dots (1)$$

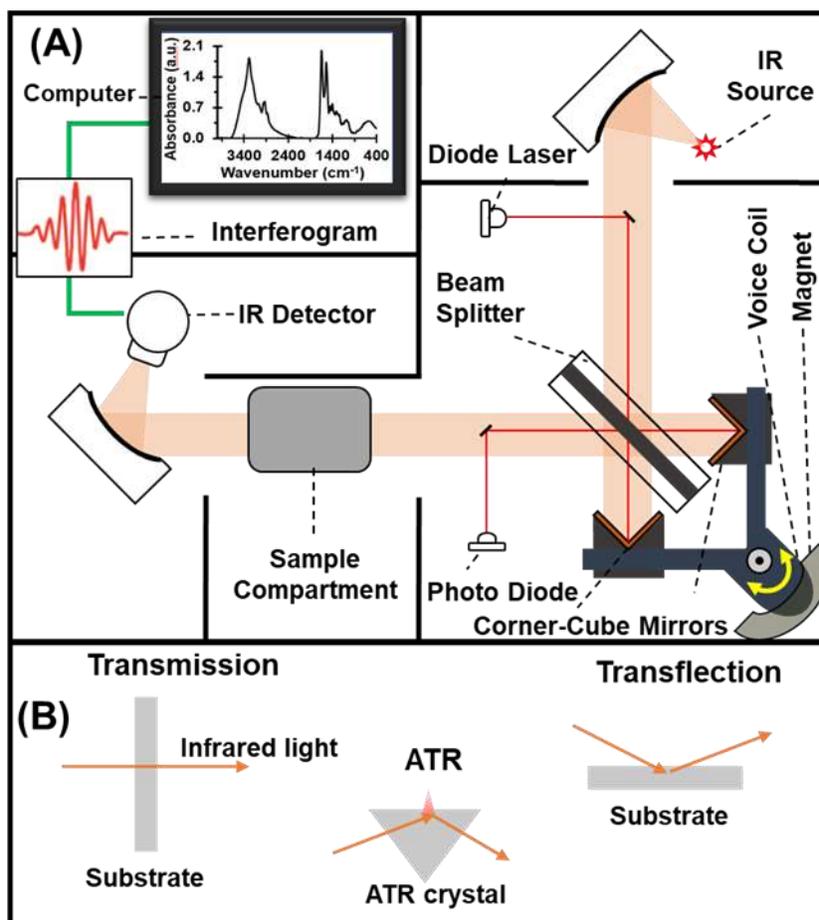


Figure 1.1 (A) The schematic of FTIR spectroscopy. The infrared (IR) detector receives an interferogram, and our computer does the Fourier transformation to generate the absorbance

spectrum. It simultaneously collects data over a wide spectral range, reflecting how much light is absorbed by a sample at each wavelength. (B) Different sampling modes of FTIR spectroscopy, including transmission, transflection, and the attenuated total reflection (ATR).

Figure 1.1 (A) shows the schematic of an FTIR spectrometer. The Michelson interferometer modulates the incoming optical radiation by changing the optical path difference between the two possible paths. Thus, formed interferograms are the coded representation of the target spectrum. Fourier transformation decodes the interferogram and provides the spectrum of the target radiation. In biochemical research, FTIR spectroscopy is a powerful analytical tool that can be used to provide insight into the composition, structure, and interaction of constituent molecules within biological solutions[18, 19]. Changes in the characteristics of biological fluids that often occur in disease can be detected in spectral data and have emerged as a robust tool used in clinical[20] and forensic[21] studies.

An increasing number of studies have demonstrated the effectiveness and promising application of this technique in several applications in the broad field of biological sciences[22]. The frequencies and their biochemical assignments of prominent spectral bands are also identified[23]. Standard protocols for the measurement of diagnostic mediums using FTIR spectroscopy[24] and spectral analysis techniques are well established[25]. However, examining the differences between the FTIR spectra has proven to be challenging due to the complexity of the biological constituents, which have different vibrational modes[26, 27]. Therefore, in FTIR studies, sophisticated spectral analysis techniques are employed to overcome these inherent spectral interpretation challenges caused by highly overlapping absorbance peaks[26].

A variety of choices are available for the sampling modes. The three major sampling modes of FTIR spectroscopy include transmission, transflection and the attenuated total reflectance (ATR). The spectra measured by FTIR are similar in each sampling mode, but subtle differences exist in the interaction between light-matter interaction. The Figure 1.1 (B) shows the schematic of light-matter interaction in each sampling mode; transmission, transflection, and ATR. In transmission and transflection sampling mode samples can be sectioned into a thin layer, allowing for accurate spectral data acquisition. For illustration, in transflection mode, the sample is placed on IR-reflecting surface and measurement are performed. In this mode, the IR/source light passes through the sample and will reflect back from the reflecting surface (substrate) through the sample[24]. In transmission mode, incident infrared beam passes through the sample (prepared in the form of pallet, mull films etc.) and the transmitted energy is measured to generate a spectrum. Similarly, in the ATR sampling mode IR beam is directed through an internal reflection element (IRE) of high refractive index[24].

1.1.1 Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR)

Spectroscopy

ATR-FTIR spectroscopy is an excellent analytical technique to understand the molecular composition of soft biological mediums and complex body fluids. It provides a complementary technique compared to other IR approaches [28, 29] for clinical diagnostics. Thus, it is an attractive technique for the rapid and reliable extraction of the molecular composition within the diagnostic medium[24]. Combined with appropriate data analysis, this technique has been applied in a few biological studies as reported in a review article[23]. In ATR-FTIR spectroscopy, light is totally reflected inside the prism of the high refractive index[28], and the reflected light from the sample (Goos-Hanchen effect)[30] creates an evanescent wave, penetrating the sample placed in contact

with crystal[31]. The energy of the evanescent wave first interacts with the sample medium, and the incident wave is entirely reflected in the system as shown in figure 1.2 (A). The output spectra represent a snapshot of molecular components within the sample medium. Herein, the penetration depth (Figure 1.2 (B)) of evanescent energy waves, $d_p = \lambda/2\pi(n_1^2 \sin^2 \theta - n_2^2)^{1/2}$ depends on the wavelength of incident radiation (λ), the refractive index of the crystal (n_1), the angle of incidence (θ), and the refractive index of the sample (n_2). It is noted that the evanescent wave decays exponentially in relation to the distance from the interface[32], so the penetration depth is only a fraction of incident IR wavelength. The evanescent waves lose energy at frequencies identical to the sample's absorption because soft biological materials[33], such as the serum, make intimate contact with the ATR element.

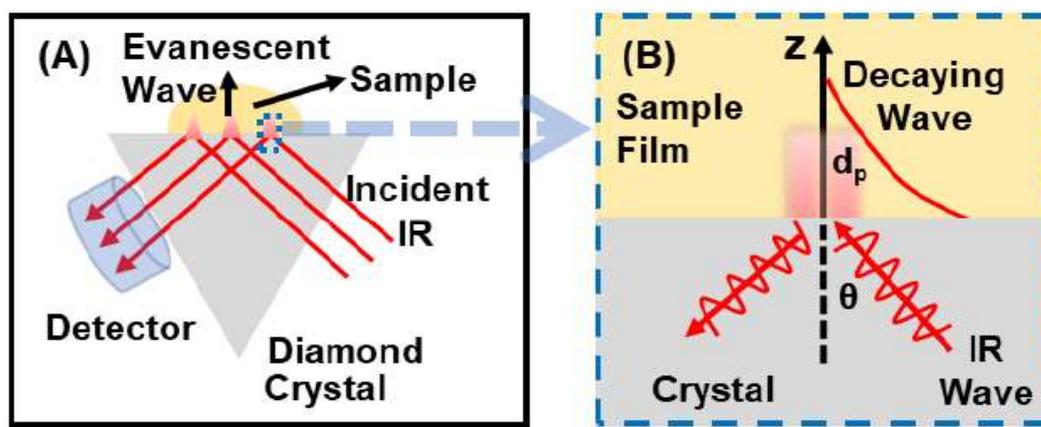


Figure 1.2 Interaction of light with sample in ATR sampling mode of FTIR spectroscopy[11]. (A) Schematic diagram of the evanescent wave formed on the sample at the internal reflection element surface. (B) Interaction of infrared signal in the material of the medium.

1.2 ATR-FTIR Spectroscopy of Serum Samples for IBD Studies

The two primary clinically identified IBDs; Ulcerative Colitis (UC) and Crohn's disease (CD), are debilitating, with a chronically inflamed gastrointestinal (GI) tract[34]. It can limit an active lifestyle and lead towards the life-threatening complications[34], including secondary infections, organ degeneration, nutritional deficiencies, and even GI cancer[35]. The incidence and prevalence rates[36] of these diseases with unknown etiology have been increasing globally. According to The Crohn's and Colitis Foundation of America, approximately 1.6 million Americans suffer from IBD, with 70,000 new cases being diagnosed every year [37].

Rapid detection and preventative interventions of these diseases enable the early administration of therapeutic strategies when the treatment is most effective. Simultaneous monitoring of mucosal healing during the treatment is also critical for IBD. Furthermore, IBDs are chronic inflammatory disorders that affect the GI tract and are characterized by episodes of relapse and remission requiring continuous monitoring. Certainly, ground-breaking advances in the interruption of these diseases[38] have been achieved using human and animal models of intestinal inflammation. However, a reliable screening test to diagnose these illnesses by primary care physicians, and hence the early administration of the therapeutic strategies, is not yet available[39].

At present, the gold standard routine tests such as colonoscopy, sigmoidoscopy and small bowel follow-through are recommended for colorectal and GI diseases[40], including IBDs. The existing guidelines recommend that individuals, fifty to eighty years of age should get a colonoscopy done every five years as part of a routine checkup[41]. Nonetheless, these gold-standard tests are expensive (requiring specialized instruments and qualified personnel and prior preparation), invasive, risky (such as colonic perforation, hemorrhaging, post-polypectomy electrocoagulation syndrome and infections), unpleasant for patients, and are not easily accessible

to the eligible population for disease screening. These factors have become significant hindrances to patient adherence to regular preventative IBD disease screening and monitoring. Studies further show that the onset of IBD typically occurs in the second and third decades of life[42], progressing into episodes of relapse and remission in the GI tract. Therefore, developing an affordable, sensitive, specific, user-friendly, robust and rapid, equipment-free and deliverable to the end-users (ASSURED) prescreening test before a colonoscopy is a critical need.

Blood and its component-based biomarkers[43] have been previously studied, utilizing gene panels to predict and discriminate IBD disease activity[44]. The blood gene expression analysis offers the significant potential to expand the understanding of the underlying genes' involvement[44], and can play a major role in the development of personalized medicine[45]. Data interpretation in gene expression analysis is not straightforward and requires several steps appropriate to the experimental protocols[43-45]; thus, gene expression analysis is not yet able to provide a low cost, quick and easier early screening test, keeping the traditional means of evaluations (such as a colonoscopy) as the gold standard. We explore the utility of FTIR spectroscopy of serum samples in ATR sampling mode[24], which could, in the future, allow physicians to diagnose and/or monitor IBDs in its early stage. This technique is simple but possesses the potential for radical change from standard practices as it will lead to the development of a preventive screening technique before the use of invasive, risky colonoscopy.

The loss of energy in the decaying process will be at frequencies identical to the sample's absorption for soft biological materials like body fluids[33]. Similarly, metabolic discharges into the body fluids (saliva, excreta, blood, and other tissue fluids) from the proximate diseased tissue change the constituent molecules, providing strong guidance for subsequent clinical assessment[33, 46]. Therefore, body fluids are considered as precious and ideal diagnostic

mediums of clinical biomarkers owing to the advantages of minimal invasiveness, low cost, and rapidity of sample collection and processing. ATR-FTIR spectroscopy of body fluids has garnered much attention in the scientific community, including from clinicians[47, 48]. Serum is the protein-rich blood component extracted after the removal of cells and clotting factors from the whole blood sample[49] and it is used in numerous diagnostic tests[50, 51]. Thus, present infrared spectral studies enriched with data analysis techniques on samples of human IBD patients and experimental mice[4-10] can pave the way for identifying diseases in the blood.

1.2.1 Spectroscopy of Serum Samples from Experimental Mouse Models

Mouse models have been a vital tool for research in GI Diseases. The understanding of intestinal inflammation in the IBD has been immeasurably advanced using several mouse models. The study presented in this dissertation also demonstrates the applicability of ATR-FTIR spectroscopy by analyzing biochemical changes in samples of IBD mice.

First, detailed insight into the molecular structural changes in serum samples of the DSS induced colitis mice[5] is presented. The DSS mouse model provides us an inexpensive, simple, and reproducible model to study human IBD. In addition to serum from DSS mice, this dissertation reports the use of serum samples of IL10^{-/-} deficient mice. Herein, mice with IL10^{-/-} deficiency treated with and without anti-TNF α antibody therapy was used to understand the efficacy of the technique for drug signature monitoring[6]. IBD patients tend to have a low IL10^{-/-} producer genotype (a regular cytokine that plays a major role in the homeostasis of the gut) more often than healthy controls[52, 53]. Therefore, genetically engineered IL10^{-/-} deficient mice were most frequently used as experimental models for chronic IBD studies. Longitudinal analysis of serum samples of IL10^{-/-} mice using two-dimensional differential gel electrophoresis (2D-DIGE) proteomic analysis has shown that the accumulation of various proteins was altered under

inflammatory conditions[54]. Importantly, a study has shown that this mouse model develops spontaneous chronic IBD and are showing a promise as a model for adequately recapitulate the full complexity of the human disease[54]. It has been found that, human IBD is a chronic intestinal inflammation disorder with an unknown etiology. The genetic and epigenetic factors also play an important role in the onset of disease and its progression[35]. Appropriateness of IL10^{-/-} mouse model to analyze the multifactorial nature of this disease including genetic factors, immune aspects, and the role of the microbiota contributing to IBD is also reported[55]. Therefore, the spectral signatures presented in this dissertation are common to both DSS and IL10^{-/-} mouse[4-10]. The quantified information of Lipocalin 2 (Lcn-2) levels in feces of IL10^{-/-} mice and Myeloperoxidase (MPO) activity of distal colon in DSS induced colitis to monitor disease progression are also used to evaluate disease level on each mouse[4, 5].

This dissertation also covers the efficacy of the FTIR technique to monitor anti-tumor necrosis factor-alpha (anti-TNF α) antibody therapy in IL10^{-/-} mouse[6]. IBD patients tend to have an increased level of tumor necrosis factor (TNF) in serum, stool, or mucosal biopsy specimens[56]. TNF α is a cell-signaling protein (cytokine) that plays a central role in inflammation. Therefore, TNF α became a key target for antibody treatment for inflammation[55]. It is found that the diversity of fecal microbiota in the IBD patients shifts towards the control individuals[57] after anti-TNF therapy. This treatment option leads to mucosal healing, reducing hospitalizations, and surgeries while improving patients' quality of life[58]. Experiments on IL10^{-/-} deficient mouse models[59] also demonstrates the same outcomes[60]. The cytokines in the stool samples show, the debilitating IBD conditions of IL10^{-/-} deficient mice can be improved by anti-TNF α antibody therapy[61].

1.2.2 Spectroscopy of Serum Samples from Ulcerative Colitis (UC) Patients

The animal models are not enough for preclinical human disease studies and for confirming the potential clinical applicability in humans. Therefore, as a follow-up step, we investigate a diagnostic application of IR spectroscopy coupled with data analysis techniques to understand ulcerative colitis-induced alterations in the molecular components of serum samples extracted from IBD human patients. Alterations reflected as fluctuations in vibrational modes and are used to identify suitable spectral signatures. The findings of this study resemble the proof-of-concept investigation with experimental (both DSS and IL10^{-/-}) mouse models, providing preliminary evidence for a reliable diagnostic test for human ulcerative colitis.

This study enriched with data analysis techniques on human sera that resemble mouse study can pave the way for identifying diseases in the blood while translating findings from an animal study to a human study. The perspective of the technique with its appropriateness and necessary improvements for the development of a new diagnostic regimen for IBDs using a serum is also provided.

1.3 ATR-FTIR Spectroscopy of Serum Samples for Cancer Studies

The FTIR spectroscopy combined with appropriate data handling frameworks has been widely applied in many oncological studies[23] such as studies using cancerous tissues from the cervix[62], the lung[63], the breast[64], the skin[65], the gastro-intestine[66], the prostate[67], the colon[68], the ovary[69], the urinary bladder[46]. These studies have reported that the molecular structural rearrangement associated with cancer development alters the vibrational mode of the molecular functional groups of the affected tissues as manifested in spectral signatures or signatures. Furthermore, the ATR-FTIR[70] spectroscopy represents a complementary approach for the clinical applications, compared to other IR approaches[29]. In this mode, high-quality

results with better spectral reproducibility compared to other modes can be obtained by the use of fluid samples[32]. It has been noted that metabolic discharges into the body fluids from the cancerous tissue change the constituent molecules, providing strong guidance for subsequent clinical assessment[33, 46].

1.3.1 Spectroscopy of Serum Samples from Mouse models of Lymphoma and Melanoma

The incidence rates of cutaneous melanoma[71], a deadly form of skin cancer, has been increasing in many regions and populations over the last few decades[72]. The increase has been of the order of 3-7% per year among fair-skinned populations[73]. At the same time, non-Hodgkin's lymphoma (NHL)[74], a solid tumorous condition of the immune system with a wide range of histological appearance and clinical features, accounts 4.3% of all new cancer cases in the United States[75]. Although significant improvement has been made to stabilize the number of NHL cases and to increase its five-year survival rate, the existing diagnostic techniques, which include the histological examination using a biopsy, are time-consuming, invasive, costly, and are not accessible to the entire at-risk population. Developing a rapid and reliable prescreening strategy for melanoma and lymphoma is thus critical because of early diagnosis and treatment of these malignancies better improve[76, 77] the patient's chances of survival.

This study presents, ATR-FTIR spectroscopy of serum samples in an effort to assess biochemical changes induced by non-Hodgkin's lymphoma and subcutaneous melanoma[12]. An EL4 mouse model of non-Hodgkin lymphoma and a B16 mouse model of subcutaneous melanoma are used to extract a snapshot of tumor-associated alteration in the serum. The study of both cancer-bearing mouse models in wild types and their corresponding control types, emphasizes the

diagnostic potential of this approach as a screening technique for NHL and melanoma skin cancer[12].

1.3.2 Spectroscopy of Serum Samples from Breast Cancer (BC) Patients

BC is the most common invasive cancer among women worldwide[78]. The international agency for research on cancer (IARC) reports that BC comprises 22.9% of invasive cancers in women[78, 79]. At present, personal inspection and imaging remain the preferred methods for screening the asymptomatic women for BC. Nonetheless, the gold standard mammography entails high costs, is not available in all medical centers, and has a low sensitivity in young women and in the dense breast. Furthermore, BC typically produces less to no symptoms when the tumor is small and is easily treatable[80]. Established mammography screening, may miss up to 20% of underlying BCs [81]. It may also lead to a 30% rate of overdiagnosis and may increase unnecessary surgical procedures and patient anxiety[82]. These limitations have led to the investigation of blood associated protein markers[83] that can be used for BC screening before mammography. The feasibility of markers like CA15-3, HSP90A and PAI-1 for the early prognosis[84] is still unclear. It is thus critical to explore potential new markers that can help for early detection of BC. Our study focuses on evaluating the feasibility of FTIR spectral discrimination of serum from healthy controls and BC patients using protein conformational alteration in serum samples via spectral deconvolution[13].

Moreover, the protein regulation[85], expression[86], and profiling[87] of tissues are commonly used as indicators for the diagnosis, treatment, and prognosis of various stages of BC[88]. FTIR spectrometry has also been successfully applied to commonly used diagnostic material, such as blood components[88-91], breast tissue[92, 93], hair[94, 95], and other biological samples[96], to discriminate BC samples[89-95]. Contrastingly, the applicability of the curve

fitting technique while discriminating IR spectra of control and BC sera samples have not been well understood[97]. Spectral fitting using characteristic Gaussian Function Energy Bands (GFEB), improves the resolution and ease the inherent IR spectral analysis difficulties involving highly overlapping absorbance peaks[25]. In fact, a key element appears to be missing; the details of GFEB that have attributed to the specific functional groups present in sera and their BC induced changes.

Additionally, body fluids, including blood-components, are considered as precious and ideal diagnostic mediums of clinical biomarkers[98] owing to advantages of minimal invasiveness, low cost, and rapidity for sample collection and their processing. The assessment of BC associated changes in protein secondary structures of body fluids will thus be an emerging interest over the existing histopathological examination of the breast biopsy materials. Alterations in the biochemical composition of the serum could reflect changes of physiological states due to BC, enabling early disease diagnosis and treatment[99].

1.4 Protein Conformational Studies using FTIR Spectroscopy

X-ray crystallography and nuclear magnetic resonance (NMR) have been widely used to examine the structure of proteins and other biological macromolecules[100]. Despite the fact they have been successfully used in biochemical studies over the years when it comes to the high-resolution analysis of protein structure and function, the use of these complementary spectroscopic methods is hindered by the need for sampling protocols[101] and sophisticated data analysis tools. The X-ray diffraction technique requires a well-ordered crystal[101], while the use of NMR spectroscopy is limited to small proteins[102]. Data analysis protocols for these techniques are also complex, complicating the interpretation of the results. These limitations have led to the development of alternative methods for determining protein structures.

FTIR spectroscopy is one alternative method that can be used for protein secondary structure analysis [4, 103, 104]. In previous reports, the FTIR spectroscopic investigation of protein secondary structures[105] in BC patient serum samples was validated by several other analytical techniques[106] such as X-ray, NMR[107] and Circular Dichroism spectra (CD)[108]. FTIR technique has also been tested with various sample types and conditions, including living cells[109], aqueous media[110], Hydrogen deuteration[111] in serum[112], dehydration[113], and heat-induced[114, 115] denaturation of serum. Additionally, spectral deconvolution[108] has been employed to diagnose or monitor various ailments, including prostate cancer[116], lymphoma, melanoma, Alzheimer's disease[117], Parkinson's disease[118], colitis, and scrapie[119]. Moreover, this method has been successfully used to study protein-protein interactions[120], the structure of calcium-binding proteins[121], and the understanding of the uses and misuses of techniques[122], their optimizations,[123] and instrumental improvisations[124]. Protein structure, as well as protein conformational changes[125], structural dynamics, and stability, have also been successfully determined using the second derivative curves[126]. All in all, FTIR spectroscopy has emerged as a powerful tool to study protein secondary structures and can be clinically useful in the early diagnosis of diseases.

In the present proof-of-concept pilot study, we have used the FTIR spectral discrimination using curve fitting, to obtain the best fit that reflects protein conformational changes in serum samples of BC patients. The curve fitting technique is also elaborated on the complex spectral region of carbohydrates and nucleic acids, $1000-1140\text{ cm}^{-1}$. By deconvoluting these regions of experimental spectra with the corresponding GFEB of various biological components, the differentiating signatures of controls and cancerous spectra were determined. Other IR spectral markers such as peak positions of the absorbance curves and spectral signatures such as the ratio

of absorbance values in amide II to amide III bands are also considered for discrimination. Statistical analysis is further performed in these identifying spectral signatures to understand discriminating potential. Herein, the accepted scientific premise is that the BC associated genetic alteration in serum is reflected in the complex region of nucleic acids including deoxyribonucleic acids (DNA) and ribonucleic acids (RNA). Therefore, our discussion also includes the possible application of genetic and proteomic molecular mapping in serum samples via FTIR spectroscopy for early detection of BC. We have incorporated statistical measures, holistically evaluated the biochemical mapping of protein structures and circulating nucleic acid components by using IR spectral deconvolution. A unified fitting protocol for all the samples and a potential prototype applicable in clinical domain is also presented. These findings go beyond the earlier study[97] providing spectral signatures with higher sensitivities and specificity. Similarly, the implementation of optimized experimental and data analysis protocols, and quantification of the spectral signatures by scrutinizing molecular entities rather than relying entirely on wider spectral ranges, are the improvements over an earlier study[89].

2 EXPERIMENT

Sample handling and measurement methods were carried out in accordance with relevant guidelines and regulations. The sera were thawed upon purchase/obtain and stored at -80°C until analysis. Instrumentation, measurements and analysis follow the standard protocols[24, 25]. Herein, the selection of instruments, their measurement parameters, spectral measurement/analyzing protocols, animal model selection, and human patients' samples are described in stepwise.

2.1 FTIR Spectrometer

All the IR spectral data were obtained with the use of a Bruker Vertex-70 FTIR spectrometer fitted a KBr beam splitter and Deuterated Tri-Glycine Sulfate (DTGS) detector. An MVP-Pro ATR accessory fitted with a diamond crystal configured to have a single reflection was used. Medium Blackman-Harris apodization function was employed with a resolution of 4 cm^{-1} and a zero-filling factor of four to provide the best resolution ability (maximum signal-to-noise ratio). The aperture size is also set to 2.5 mm for the optimization of the detector response without saturation. A Parker-Ballston gas purging system was also used to maintain purified ambient air in the spectrometer.

We have performed repeated measurements and extract information from several studies, including references [23-25, 32], while selecting these parameters. This selection allows us to reduce noise, and to analyze the spectral data without losing information during spectral deconvolution. Resolution of 2 cm^{-1} gives noisy derivative spectra and of 8 cm^{-1} resolution gives rise to missing secondary structure information (spectral signatures show a lower degree of variability), while using a diamond ATR crystal. Therefore, a resolution of 4 cm^{-1} with a zero-filling factor of 4 to give a data spacing of 1 cm^{-1} . Diamond has a smaller acceptance angle cone,

which allows having a good optical design to extract good spectra. Because of the strong robust, and chemically inert feature of the diamond, it is considered as the best ATR crystal for routine measurements of biological samples[127]. Thus, optimized values of measurement parameters are used in this experimental and equipment setup.

2.2 Animal Growth and Sample Collection

2.2.1 Dextran Sodium Sulfate (DSS) Mouse

Three-week-old female C57BL/6 wild type (WT) mice were obtained from Jackson Laboratories (Bar Harbor, ME). Mice were group-housed under controlled temperature (25°C) and photoperiod (12:12-h light–dark cycle) and fed ad libitum. Animal experiments were approved by the Institutional Animal Care and Use Committee of Georgia State University (Atlanta, GA), and performed in accordance with the guide for the Care and Use of Laboratory Animals by U.S. Public Health Service. All procedures were approved under the IACUC protocol #A14010.

C57BL/6 WT mice were administered DSS (MP Biomedicals, Solon, OH) at 3% in drinking water ad libitum for 7 days. Feces and blood were collected on day 0 (before DSS treatment) and day 7. Hemolysis-free serum was collected by centrifugation using serum separator tubes. Mice were sacrificed by CO₂ euthanasia.

2.2.2 Interleukin 10 Knockout (IL10^{-/-}) Mouse

The IL10^{-/-} mouse model develops IBD in a time-dependent manner[59]. Similarly, the anti-TNF α treatment on IL10^{-/-} mice show the therapeutic efficacy in resolving intestinal inflammation[60]. A study indicates that IBD development and their healing is independent of sex (male or female) in IL10^{-/-} mouse[61]. Similarly, Anti-TNF drugs have proven highly effective for targeting and diminishing the downstream effects of TNF α activation. However, the precise mechanism is still unclear[128]. Studies have shown that a significant proportion of patients do

not respond to the treatment, or they lose the effect over time. Therefore, 30-40 % of CD patients and 25-30 % of UC patients still need surgery at some point during their life addition to the biological treatment[129, 130].

In this study, eighteen, three weeks-old female (IL10^{-/-}) knockouts mice, obtained from Jackson Laboratories (Bar Harbor, ME), were used. Nine mice were treated with anti-TNF α therapy, twice a week by intraperitoneal injection of anti-TNF antibody, starting from the 28th day of their birth, and rest (n=9) are treated with Phosphate buffered saline (PBS). Feces collected on day 28 (4 weeks) and day 98 (14 weeks), and Lipocalin-2 (Lcn-2) was measured to assess the intestinal inflammation in those mice[131]. Total RNA was also extracted from colon tissues for further analysis.

Similarly, blood samples collected on day 28 and day 98 was centrifuged by using serum separator tubes for IR spectral measurements. Herein, samples collected from mice on day 28, before they develop colitis constitute the control or non-colitis (n=16). The samples extracted on days 98 of PBS injected mice are colitis (n=9). Similarly, the samples extracted on days 98 from anti- TNF α injected colitis- mice are treated (n=9). All studies were performed following the Institutional Animal Care and Use Committee at Georgia State University (Atlanta, GA), permit number: A14010.

Herein, we have analyzed microRNAs (miRNAs) to analyze the efficacy of anti-TNF treatment. It is noted that miRNAs are promising biomarkers for IBD[132]. We examined the expression levels of pro-inflammatory cytokines in the colon, to understand the response of therapeutics. Herein, the colonic miRNA expression analysis is performed by the quantitative polymerase chain reaction (qPCR) by following the protocol as discussed[132].

2.2.3 Collagen Antibody-Induced Arthritis Mouse

BALB/C WT mice received collagen antibody injections (200 μ L) on day 0 by an intravenous (tail vein) injection. (On day 6, mice received lipopolysaccharide (200 μ L) by intraperitoneal injection in order to boost the effect). Blood samples were collected from each mouse on pre-treatment (day 2) and on post-treatment (day 12) from the jugular vein. Hemolysis-free serum was collected by centrifugation of the collected blood using serum separator tubes.

2.2.4 Toll-Like Receptor Knockout (TLR5^{-/-}) Mouse Model of Metabolic Syndrome

TLR5^{-/-} mice were grown in our facility (Dr. Merlin's lab at the Biology department). TLR5^{-/-} spontaneously develop metabolic syndrome as previously described[133]. Age-matched wild type (WT) and TLR5^{-/-} mice were fasted for 5-h and baseline blood glucose levels measured with a blood glucose meter (Roche) using blood collected from the tail vein.

2.2.5 Human Sera of IBD Patients

Serum samples of IBD human patients were provided by Dr. Emilie Vienna's from Dr. Didier Merlin lab at the Institute for Biomedical Sciences, GSU, Atlanta, GA 30302, USA. These human sera were obtained from Equitech-Bio, Inc. (Kerrville, TX, USA). Information about IBD patients is provided below in Appendix A.4.

2.2.6 B16 Melanoma or EL4 Lymphoblast Tumor Model

C57BL/6J mice (6-8 weeks, 20-22 g, the Jackson Laboratory) were engrafted with B16 melanoma or EL4 lymphoblast via subcutaneous (s. c.) route with 2×10^5 of each cell line. B16 and EL4 cells were obtained from the American Type Culture Collection (ATCC) and maintained in DMEM with 10% FBS prior to use. Mice were euthanized after 3 weeks of tumor inoculation when tumors were larger than 1000 mm³ in size (see figure 1, inset (i)). Serum samples from tumor-bearing mice and healthy mice were isolated and stored in -80°C until analysis. All

experiments using animals described in this study were approved (protocol number: A17015) by the Institutional Animal Care and Use Committee (IACUC) of Georgia State University, Atlanta, GA, and experiments were conducted according to the guideline of Office of Laboratory Animal Welfare (OLAW), Assurance number: D16-00527(A3914-01).

2.2.7 Human Sera of Breast Cancer (BC) Patients

Human sera from BC patients were obtained from the Breast Satellite Tissue Bank, Winship Cancer Institute, Emory University, Atlanta GA, USA. The Helsinki Declaration guidelines were followed for sample collection, and informed consent was obtained from all patients. Blood was collected without additives from patients after informed consent. The blood was then centrifuged at ~3200g for 10 minutes, and serum was pipetted and stored at -80°C until analysis. The control healthy individual sera were from the baseline collection of healthy women (age 41-58 years) participating in an independent intervention study under approval number 13317, Edith Cowan University, Perth, Australia. All participants provided informed consent. The sera were thawed, aliquoted in small volumes, and stored at -80°C until analysis. Information about the BC patients is provided below in appendix.

In this study, the serum samples from healthy controls were provided by Dr. Marlina Catherine Kruger at the Department of Nutritional Science, School of Food and Advanced Technology, College of Sciences, Massey University, Tennent Drive, Palmerston North 4442, New Zealand.

2.3 Spectral Response Measurements

The surface of ATR crystal was first cleaned with sterile phosphate-buffered saline and by ethanol. The cleanliness test was then conducted to ensure there are no signal peaks higher than the environmental noise level. Background measurement, showing the presence of any

environmental residue on the crystal surface or in the light path, was performed on a clean surface before each measurement. It allows us to get high-quality spectra using background corrections. Serum samples (1 μL) are deposited in order to fully cover the crystal surface and allowed to air dry (~10 minutes) at room temperature. An evanescent wave produced at the interface by the internal reflection of light interacts with the layer of the sample within the penetration depth[134] (see Appendix A.2 for detail). Each sample was scanned multiple times to get ten or more high-quality (exactly overlapping) spectral data within the mid-infrared range 400 to 4000 cm^{-1} and 6 reads of the 100 co-added scans for each sample (total of 600 scans) were averaged. Using OPUS 6.5 software, all the spectra were min-max normalized by scaling the range 1800 to 900 cm^{-1} . This range comprises a significant biochemical fingerprint of the biological material and the amides (I and II) bands[24]. Vector normalization[135] is another popular protocol where the average absorbance is subtracted from each data point, and the spectrum is divided by the square root of the sum of squares of absorbance. In general, the normalization protocol[24] can be chosen to best serve the purpose of comparison. In this study, min-max normalization to the amide I peak is employed as it emphasizes the deviation between the colitis and control sample spectra. Min-max normalized spectra were then sectioned within the range of 1800 to 900 cm^{-1} for further analysis. The vector normalization is also used while comparing the second derivatives spectra.

2.4 Data Analysis

2.4.1 Hierarchical Cluster Analysis (HCA) and Dendrogram Plot

Our analysis includes the dendrogram of HCA, which is commonly employed to identify the similarities between the FTIR spectra by using the distances between frequencies and aggregation algorithms [63]. Using the “PAST (Paleontological STatistics[136]) 4 - the Past of the Future” software and the vector normalized second derivative curve of the absorbance spectra as input data

(variables), HCA is performed and the dendrogram is plotted. Dendrogram tree diagram performed, using Ward's algorithm and Euclidian distance measurements, allow us to visualize the overall classification.

2.4.2 Principal Component Analysis (PCA)

PCA, a useful statistical analysis[137], is first performed to explain the holistic evaluation biomolecular content variations reflected in infrared absorbance data. Using the "PAST (PAleontological STatistics) 4 - the Past of the Future" software and the vector normalized second derivative curve of the absorbance spectra, as input data variables, we analyzed the variance-covariance matrix with the pairwise exclusion of missing values to get the component plots. The scatter plot of PC1 and PC2 were used to visualize the clustering of groups together with different magnitudes and directions. The scree plot is also used to check the total variance presented by PC1 and PC2 is significant. These findings from the PCA analysis led us to investigate spectral signatures useful in the clinical domain.

2.4.3 Spectral Fitting (Deconvolution) for Protein Conformational Studies

We perform quantitative analysis of the absorption band values at different spectral markers' positions, which follows discrimination of secondary structures by deconvolution of the spectra in the amide I region. During the deconvolution process, the computed curve that best fit with the experimental spectrum is obtained from the superposition of Gaussian function energy bands (GFEB). The individual bands from deconvolution represent proteins' secondary structures as discussed in the protocol papers, and similar studies in the field[108, 116-118, 120, 126] which were used to verify the applicability of this technique over complementary spectroscopic methods (X-ray crystallography, and nuclear magnetic resonance).

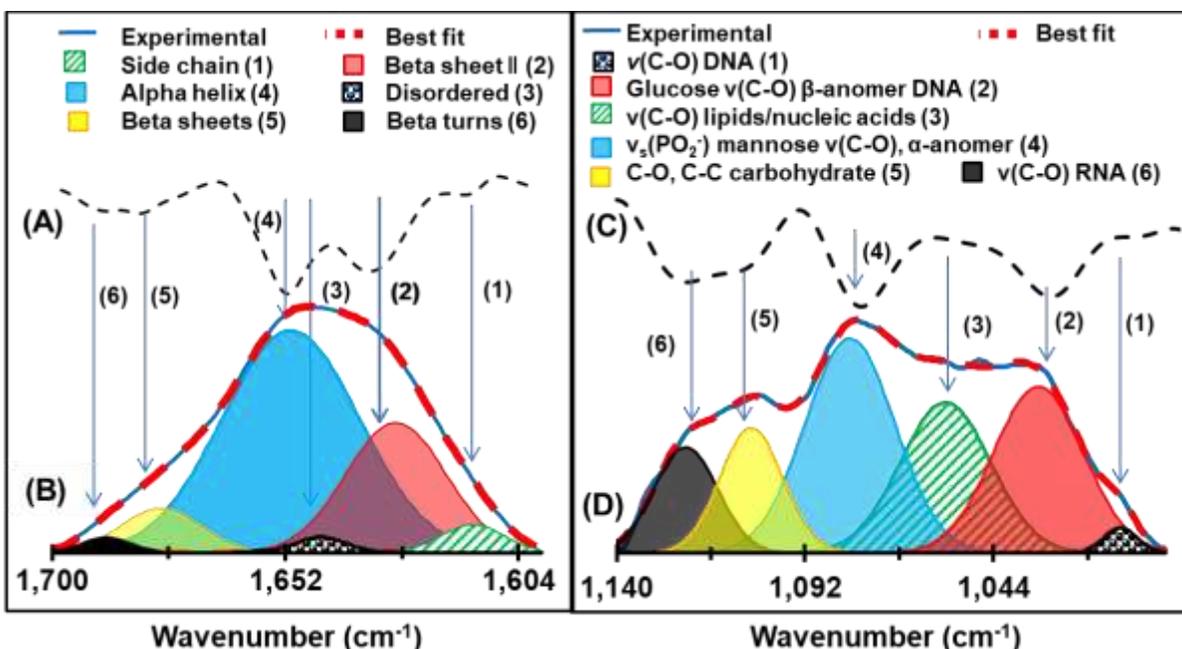


Figure 2.1 The process of spectral deconvolution[6]. (A) Second derivative spectra of amide I absorbance curve 1600-1700 cm⁻¹. (B) Gaussian function energy bands used to obtain a curve that fits the experimental absorption curve. (C) Second derivative spectra of nucleic acids and carbohydrates, complex band 1000-1140 cm⁻¹. (D) Deconvolution of the complex band by estimating the number and position using the minima of the second derivative curve.

The process of deconvolution includes sectioning of min-max normalized spectra followed by rubber-band correction with two baseline points, such that absorbance values at extreme ends of the selected region will be zero. These spectra were then fitted with GFEBS by approximating the position and the number of bands from the minima of second derivative spectra (Figure 2.1 (A)) and also by the minimization of root mean square error via the Levenberg-Marquardt[138] algorithm. In the fitting process position, height, width, and hence integral varied such that the simulated curve best fits the experimental curve. Herein, the goodness of the fit is determined by the convergence of the root mean square error or the residual to a singular minimum value. Figure

2.1 (B) shows the spectral fitting within amide I. Using six bands with corresponding biological components; side chain, beta-sheet, alpha helix, disordered structure, and beta turns, we found that fitting solution is unique as determined by the second derivative. The minima (position and number) of the second derivatives of the spectral curves reflect the positions and the number of bands needed to fit the spectral curve. The minor variation of the positions of the six bands are all within the acceptable ranges for those biological components. Similarly, by choosing the number and positions of bands indicated, excellent root mean square (RMS) error values less than 0.0035 were obtained. It further allows us to overcome the challenges[139] of deconvolution process for quantitative analysis. Herein, by normalizing and sub-sectioning, the amide-I region of absorption spectra followed by baseline correction and spectral fitting, the effects of the bands describing individual biological components allow us for a more precise analysis of the spectra, thereby increasing confidence while comparing two states.

Similarly, the complex region[140] of carbohydrates and nucleic acids $1140\text{-}1000\text{ cm}^{-1}$ is deconvoluted with six GFEB, describing individual biological components is as shown in figure 2.1 (D). The number and positions of GFEBs are again approximated by using the minima of second derivatives (figure 2.1 (C)). The sum of the integral areas covered by six bands was further statistically analyzed.

2.4.4 Statistical Analysis

Quantified values of these identifying spectral signatures are statistically analyzed to see the colitis-associated alteration in the blood serum, and their stabilization after biological therapy. Our statistical analysis mainly includes finding the sensitivity and specificity of the signatures for their discrimination. Sensitivity and Specificity of a diagnostic test are often used to describe the diagnostic accuracy/performance of the analysis in biomedical research[141]. The discriminating

potential of a diagnostic regimen can be quantified by the Youden index and the area under the receiver operating characteristic (ROC) curves [142]. The ROC curve is plotted to find the area under curve (AUC). The optimal cutoff value calculated based on the Youden index for each spectral marker is used to select the positivity/negativity of the disease and to estimate the sensitivity and specificity.

3 RESULTS

The result section systematically explains the identifying spectral signatures in mouse models and human patients in a stepwise pattern. The quantitative information reflected in identifying spectral signatures; are further analyzed by using statistical analysis techniques.

3.1 Identifying Infrared Spectral Signatures of IBD in Serum

The serum sample extracted from the experimental models of IBD: IL10^{-/-} and DSS mouse shows diagnostic accuracy with 80-100% sensitivity and specificity values. Importantly, the findings of human IBD patients' study also resemble these proofs-of-concept investigations using mouse serum. Maximum values of sensitivity and specificity are 100% and 86%, respectively, in human samples.

3.1.1 Spectral Signatures in Serum of DSS Mouse

The encouraging results of the earlier studies using DSS mouse[4, 5] have motivated us further elaboration by monitoring the temporal variation of molecular composition in the serum along with the colitis aggravation[11]. We examined samples at days 0, 3, and 7 of the chemical feeding. Six mice (S1, S2, S3, S4, S5, and S6) were fed 3% DSS dissolved in drinking water. The serum samples were tested for the disease progression using the spectral response analysis. After analyzing the absorbance spectra within spectral range 1000-1140 cm^{-1} , we learned the concentration of carbohydrates such as glucose and mannoses was elevated due to colitis as judged in previous studies[4, 5]. These studies have shown that the elevation of the spectral band representing the Mannose presence in the carbohydrate region. Similarly, increased presence of β -pleated sheet protein secondary structures and decreased the presence of α -helical structures in the Amide I region (1600-1700 cm^{-1}) is also evaluated.

Figure 3.1 (A) shows the average ($n=6$) of max-min normalized spectra of mice sera extracted from three different time periods before (days 0, 3, and 7 of chemical feedings) the DSS mouse model, with a Biomolecular assignment at different wavenumber regions. Our study is primarily focused on testing the temporal variation in serum components along with the colitis progression by analyzing identified unique spectral signatures of colitis (shaded regions in figure 3.1). The second derivative curve of the absorbance spectra is shown in figure 3.1(B). Visual inspection of each type of spectrum shows the alteration of disease Biomarkers due to colitis. This variation can be seen on day 7 of DSS feeding, but not on day 3.

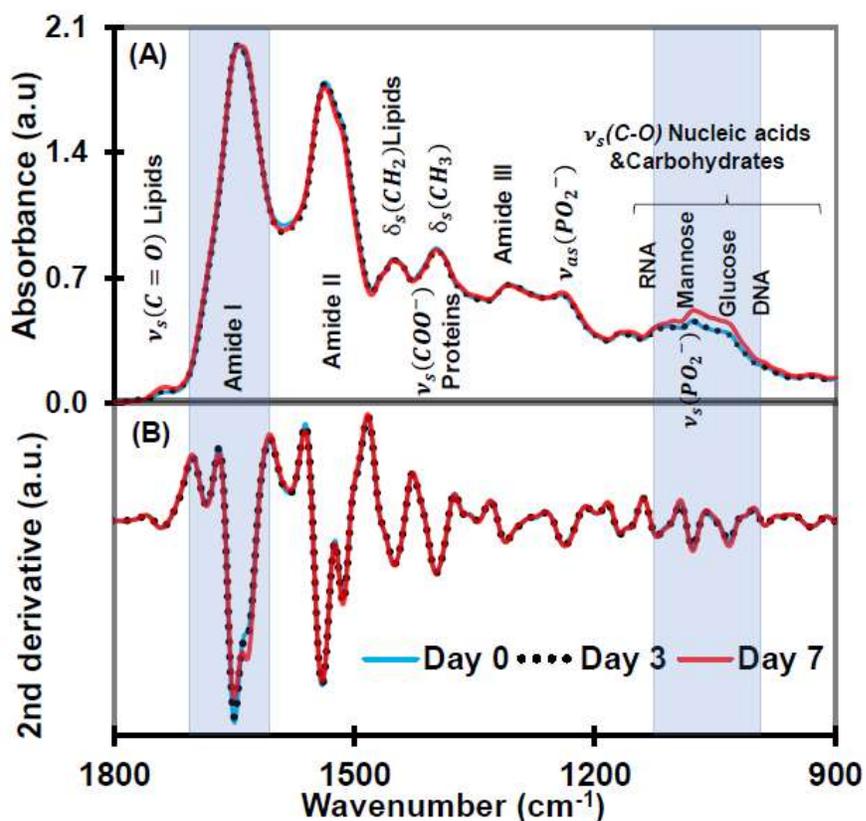


Figure 3.1 (A) Averaged ($n=6$) ATR-FTIR spectra of serum samples derived from DSS mice on day 0, day 3, and day 7 of the chemical feeding. The variation in infrared absorbance on day 7 is higher in the complex region of carbohydrates and nucleic acids compared to on day 3.

(B) Second derivative curves of the absorbance spectra. Second derivative spectra show the variation on day 7 in the amide I region. Herein, shaded regions show two major regions of interest[11].

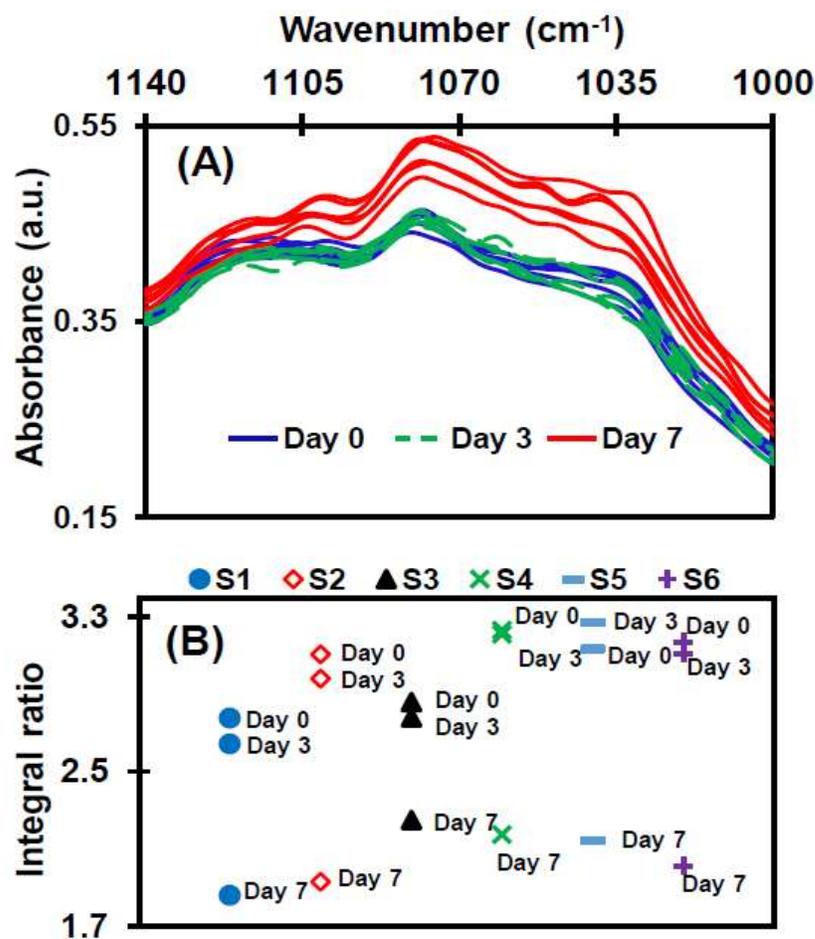


Figure 3.2 Longitudinal analysis of serum samples of DSS induced colitis mice[11]. (A) Normalized absorbance values in the range 1000-1140 cm⁻¹ showing changes in spectral signatures representing carbohydrates and nucleic acids are significant at day 7, but not at day 3. (B) The integral ratio of Gaussian function energy bands representing α -helix and β -sheet secondary structures of proteins in serum samples extracted from six mice (S1, S2, S3, S4, S5 and S6). The secondary structure analysis also shows the significant changes only at Day 7.

Figure 3. 2 (A) shows the normalized absorbance spectra covering the complex region of carbohydrates and nucleic acids ($1000-1140\text{ cm}^{-1}$). IR absorbance values of this region are more on day 7, where mice develop fully colitis. In the normalized data, absorbance values of the spectral position, 1076 cm^{-1} representing mannose, for the control (day 0) are between 0.44-0.47 a.u., colitis (day 7) are 0.49-0.53 a.u. that the values for the intermediate (day 3) stage lies within the range of control. Similarly, the ratio of the integral areas taken by GFEBs representing α -helical and β -pleated sheet protein secondary structures for each individual mouse is shown in figure 3.2 (B). The integral ratio of α -helix and β -sheet protein structures also show a clear difference in the spectral data representing day 0 with those representing day 7 of chemical administration, while there is no significant difference in day 3 data. In these two regions of interests (first: Glucose, Fructose, Mannose, and Endocyclic C-O-C vibration and second: α -helical protein secondary structure of amide I and β -pleated sheet structures) show alteration with higher statistical significance (student's t-test, two-tailed unequal variance p-value < 0.05).

3.1.2 Spectral Signatures in Serum of IL10^{-/-} Mouse

This study is the follow up for an earlier study from our lab[5]. Addition to earlier studies, we have increased the number of mice (samples) and therapeutic effect of anti-TNF α is also evaluated[6].

Discrimination of Absorbance Values: To monitor the therapeutic response of colitis, we examine the chemical composition of serum samples derived from normal (n=16), IL10^{-/-} induced colitis (n=9) and with anti-TNF α therapy IL10^{-/-} mice (n=9). Figure 3.3 (A) shows average representations of the normalized absorption spectra of serum samples derived from each mice type. Herein, we have monitored previously identified IR spectral signatures for colitis in diseased species with anti-TNF α antibody therapy, which displays exciting results. The alteration

in the IR absorption values (at identifying spectral signatures) due to colitis in IL10^{-/-} mice come close to control mice in the anti-TNF α treatment treated cases.

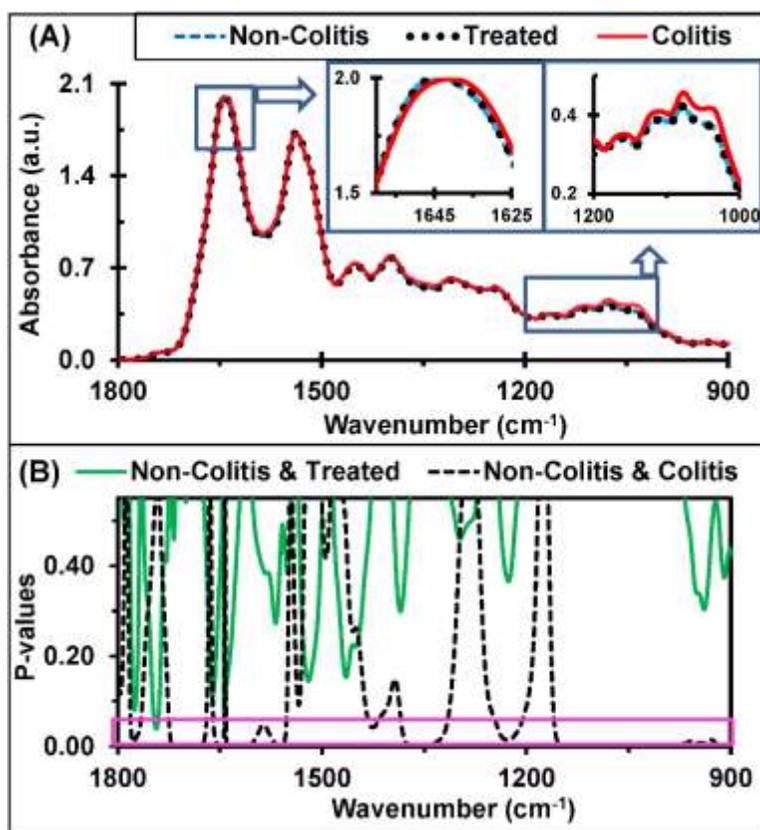


Figure 3.3 Normalized absorbance curve and p -value calculation results[6]. (A) Averaged ATR-FTIR spectra of serum samples, derived from control (Non-Colitis, $n=16$), treated with anti-TNF α antibody (Treated, $n=9$) and colitis untreated (Colitis, $n=9$) mice. These spectra show proper anti-TNF α therapy ameliorate the colitis condition. Inset shows magnification within 1200-1000 cm^{-1} and 1660-1625 cm^{-1} , which clearly shows how the absorbance curve of anti-TNF α treatment comes close to the control levels. (B) P -value calculation of colitis and treated conditions with control types. The p -value is less than 0.05 at various spectral bands while comparing non-colitis and colitis, but it is always greater than 0.05 while comparing non-colitis and treated. The region highlighted by the pink box is the region with $p \leq 0.05$.

The student t-test p-values on absorption spectra representing normal, colitis, and antibody-treated colitis mice are as given in Figure 3.3 (B). It shows how absorption values vary with control Vs treated and the control Vs colitis within the entire wavelength region 900-1800 cm^{-1} . Throughout the whole wavelength region, the differentiating signatures better than 95% (or higher) are highlighted by the shadowed area. These spectral regions with biomolecular assignments and bond vibrations having a significantly better than 0.05 are assigned as C=O of lipids (1766-1780 cm^{-1}), amide I of proteins (1670-1710 cm^{-1}), amide I and amide II of proteins (1548-1649 cm^{-1}), CH_3/CH_2 bending (1317-1383 cm^{-1}) and asymmetric PO_2 stretching of phosphate lipids 1208-1244 cm^{-1} , and the mixed region of carbohydrates, and nucleic acids in 900-1159 cm^{-1} . Due to the complexity of biological systems, biomolecular and bonding vibration assignments are tentative and are based on numerous studies as shown in table 3.1. The colitis associated changes in all these spectral bands come to the level of control in anti-TNF α antibody therapy mouse serum.

Table 3.1 Discriminatory Infrared spectral bands of UC, with biomolecular assignments and their bond vibrations[6].

Band (cm^{-1})	Assignment and vibrations
900-1158	Carbohydrates[143, 144] (Glucose, Mannose, Fructose) and nucleic acids (Deoxyribose/Ribose DNA, RNA)[145]: C-O, C-C stretch, C-H bends, Endocyclic C-O-C vibration and, $\nu_s(\text{PO}^{2-})$ [146]
1208-1244	Amide III, $\nu_s(\text{PO}^{2-})$ [147]
1317-1382	Collagen: CH_2 wagging, the vibration of α , and β anamor[148].
1420-1430	Polysaccharides, $\nu_s(\text{COO}^-)$, $\delta(\text{CH}_2)$
1480-1580	Amide II of proteins: (α -helical, β -pleated sheet, unordered conformation structures), $\delta(\text{N-H})$, $\nu(\text{C-N})$ [25].
1600-1700	Amide I of proteins: (α -helical, β -pleated sheet, β -turns, random coils, and side-chain, β (anti- +turn) structures), $\nu(\text{C=O})$, $\nu(\text{C-N})$, CNN[25].
1720-1750	Lipids C=O stretching [149]

Identified spectral signatures: In this study, identified spectra signatures are: (a) absorption values at wavenumber 1033 cm^{-1} (I_{1033}), primarily due to glucose presence; (b) absorption values at wavenumber 1076 cm^{-1} (I_{1076}), representing the mannose as well as phosphate[23] presence; (c) the ratio of absorbance at wavenumber 1121 cm^{-1} , associated with RNA presence, to its value at 1020 cm^{-1} , associated with DNA presence (I_{1121}/I_{1020}); (d) the ratio of absorbance at wavenumber 1629 cm^{-1} , indicating the protein presence, to 1737 cm^{-1} signaling the presence of lipids (I_{1629}/I_{1737}); (e) the ratio of Gaussian function energy bands representing α helix and β sheet protein secondary structures as obtained from the deconvolution of amide I region; and (f) the sum of the integral area of GFEBs used to fit the experimental curve within the complex band of carbohydrates, and nucleic acids $1000\text{-}1140\text{ cm}^{-1}$.

Earlier, various serological markers were analyzed for IBD diagnosis[43] and discrimination of IBD types[150]. These studies have reported abnormalities in lipids, amino acids, and energy metabolisms in the serum samples of IBD patients[151]. Associations between fatty acids and inflammatory cytokines or the protein abundance and the epigenetic alteration in the samples of IBD patients have been also established[152].

Similarly, proteomic analysis in serum samples of IL10^{-/-} mice shows alterations of various proteins[54] under inflammatory conditions. The protein secondary structure alterations could be due to a melded manifestation of these protein variations. As such, variations in these reported serological markers are most likely the primary reason for the IBD induced changes in absorption frequencies of functional groups of proteins, carbohydrates, nucleic acids, and lipids.

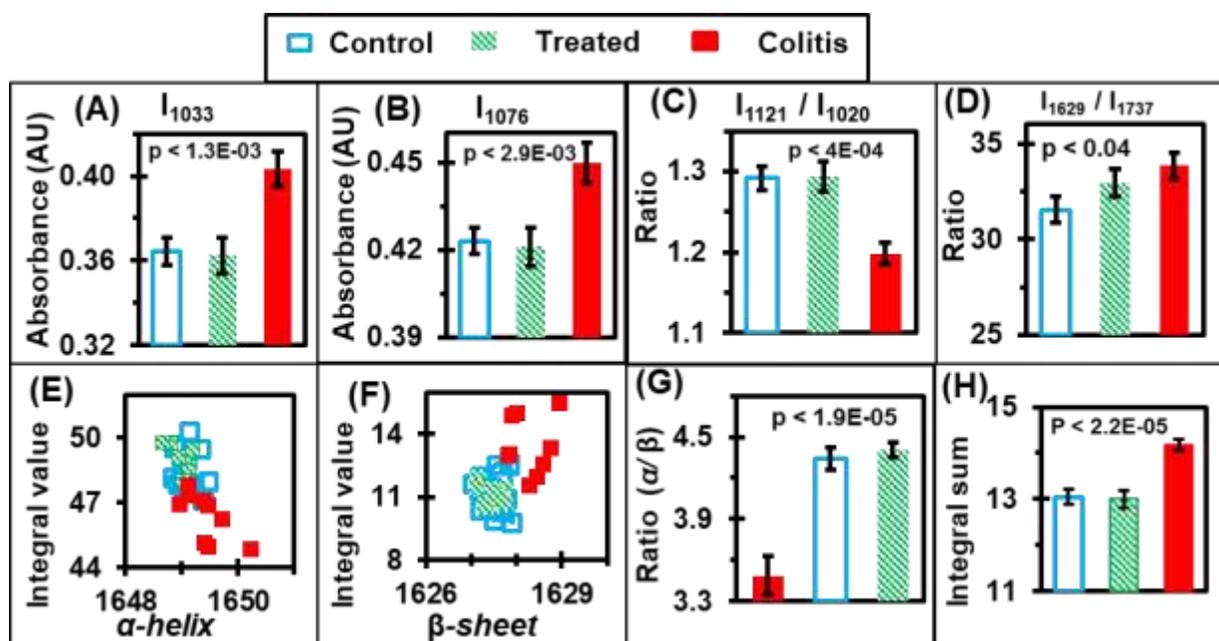


Figure 3.4 Representation of identifying spectral signatures[6]. (A) Ensemble average representative of absorbance at wavenumber position 1033 cm^{-1} . (B) Ensemble average representative of absorbance at wavenumber position 1076 cm^{-1} . (C) The ratio of absorption values at wavenumber 1121 cm^{-1} to 1020 cm^{-1} . (D) The ratio of absorbance at wavenumber 1629 cm^{-1} to 1737 cm^{-1} . (E) The integral values of Gaussian function energy bands representing the α helical structure of protein secondary structure. (F) The integral values of Gaussian function energy bands representing the β pleated sheet structure of protein secondary structure. (G) The integral ratio of α -helix and β -sheet protein secondary structures. It shows the increased level of β sheet structures and a decrease in α helix structures due to colitis and their resettlement in anti-TNF α therapy mice. (H) The sum of the integral area of GFEBs used to fit the experimental curve in the complex band $1000\text{-}1140 \text{ cm}^{-1}$.

The representation of ensemble averages for identifying five spectral signatures, I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , integral ratio (α/β), and the sum of the area of GFEBs used to fit $1000\text{-}1140 \text{ cm}^{-1}$ are shown in Figure 3.4. Absorption values representing the glucose peak position at 1033 cm^{-1} are shown in Figure 3.4. Absorption values representing the glucose peak position at 1033 cm^{-1} is shown in figure 3.4 (A), and this value at the mannose peak position 1076 cm^{-1} is shown in

figure 3.4 (B). The average ratio of the RNA peak to DNA peak (I_{1121}/I_{1020}) is as indicated in figure 3.4 (C), and the average ratio of the peak representing protein to lipid peak is shown in figure 3.4 (D). Figure 3.4 (E) and figure 3.4 (F) shows the integral values of the α -helix and β -sheet protein secondary structures, respectively. Similarly, the average integral ratio of α -helix to the β -sheet structure is shown in figure 3.4 (G). The sum of integral values of Gaussian bands used to fit the experimental curve within 1000-1140 cm^{-1} are shown in figure 3.4 (H). These absorption and ratio values are also tabulated in Table 3.2.

Although, the precise mechanism of anti-TNF α treatment in IL10 $^{-/-}$ mice that helps to regain altered biomolecules into the control level is still unknown[128]. Specifically, proteomic marker (α/β) could be due to either due to anti-TNF alpha or due to the downstream expression of other proteins. However, our experimental demonstration verifies the efficacy of IR spectroscopy to monitor the signature of anti-TNF α therapy for inflammatory bowel diseases. Since the focus of this work is to identify the IR signatures, regaining of altered spectral signatures of IBD close to the control level in anti-TNF α treatment is an important finding.

Table 3.2 Quantified values of identifying spectral signatures: I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , integral ratio (α/β), and the sum of the integral area of GFEBs used to fit the experimental complex band 1000-1140 cm^{-1} [6]. Tabulated p-values represent a comparison between control and colitis.

Features	Glucose (1033 cm^{-1})	Mannose (1076 cm^{-1})	RNA/DNA (1121/1020)	Protein/Lipid (1629/1737)	Integral Ratio (α/β)	Integral sum Complex
Control	0.36 \pm 0.01	0.42 \pm 0.01	1.29 \pm 0.02	31.54 \pm 0.69	4.34 \pm 0.09	13.04 \pm 0.16
Colitis	0.41 \pm 0.01	0.46 \pm 0.01	1.19 \pm 0.01	33.85 \pm 0.69	4.40 \pm 0.06	14.17 \pm 0.12
Treated	0.36 \pm 0.01	0.42 \pm 0.01	1.29 \pm 0.02	32.98 \pm 0.73	3.49 \pm 0.14	13.0 \pm 0.18
P-value	1.3E-03	2.9E-03	4E-04	0.04	1.9E-05	2.1E-05

Statistical Analysis: The heterogeneity between each study type is further tested by using quantitative analysis that involves the calculation of Youden's index, the area under the receiver operating characteristic (ROC) curves, sensitivity, specificity, and the calculated p-value [142]. Statistical measures[141] of these discrimination techniques also reveal the feasibility of the studied monitoring regimen in the clinical domain. The ROC curves are plotted to find the area under the curve (AUC); for each of these features, the threshold values were used to estimate the sensitivity and specificity. Then, the Youden index, (the maximum of sensitivity + specificity – 1), was used to find the optimal cutoff values among the threshold values. Herein, figure 3.5 (A) shows the confirmation of the effect of anti-TNF α therapy in studying mouse groups evaluated by analyzing expression levels of TNF α in the total RNAs extracted from colons. ROC plots for the (α/β) signature are shown in figure 3.5 (B).

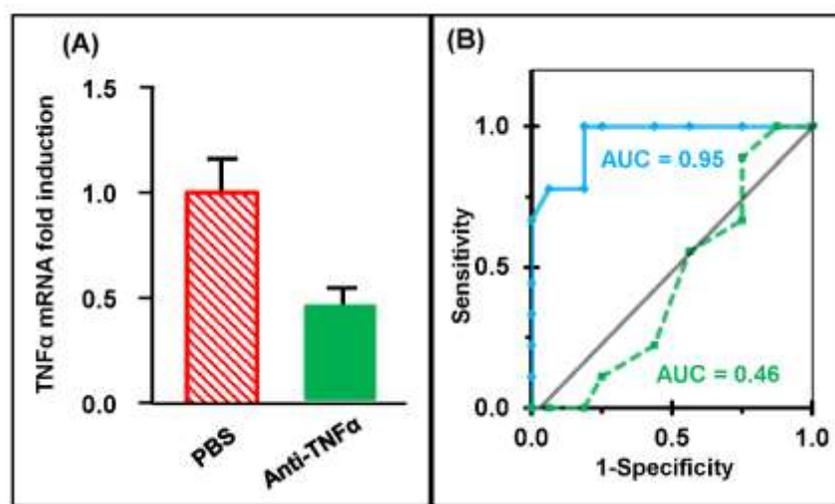


Figure 3.5 (A) The expression levels of TNF α in the Total RNAs extracted from colons. The level of TNF α is quantified by qPCR and it confirms the effect of anti-TNF α therapy. **(B)** ROC curve for the data obtained from the ratio of the integral area of energy bands representing α -helix and β -sheet protein secondary structures. AUC is 0.95 while comparing colitis and non-

colitis, but its value is 0.46 while comparing non-colitis and the treated. High diagnostic accuracy (with AUC = 0.95) can be seen for colitis, but poor diagnostic accuracy (with AUC =0.46) for the treated[6].

The maximum sensitivity and specificity of this particular signature were found to be 100% and 81%, respectively. Sensitivity and specificity are major concerns in such studies and analysis led to the identification of statistically significant differences in IR spectral markers between non-colitis, IL10-/- induced colitis, and with anti-TNF α treated mice experiments. Sensitivity, specificity, and AUC values are calculated for all the identifying spectral signatures as shown in table 3.3. Statistically, significant differences can be seen between non-colitis and colitis conditions for identifying signatures. However, while comparing control and the treated conditions, approximately 50% (0.5) AUC values can be obtained, showing there is not statistically difference between control and treated groups.

Table 3. 3 *Sensitivity, specificity and AUC value calculations of identifying spectral signatures of UC in IL10-/- mouse[6].*

Specifications	Signatures					
	Glucose (1033 cm ⁻¹)	Mannose (1076 cm ⁻¹)	RNA/DNA (1121/1020)	Protein/Lipid (1629/1737)	Integral Ratio (α/β)	Integral sum complex band
Sensitivity	89	89	100	90	100	100
Specificity	88	88	75	56	81	75
AUC (control & colitis)	0.88	0.86	0.90	0.73	0.95	0.92
AUC (control and treated)	0.54	0.56	0.49	0.66	0.46	0.55

As seen in table 3.3, AUC values while comparing control and treated are 0.54, 0.56, 0.49, 0.66, 0.46 and 0.55 for the spectral signatures: I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , integral ratio (α/β), and the integral sum, respectively. Since the AUC values of colitis and non-colitis are higher (close to 90%) and that of non-colitis and treated is close to 50%, showing these spectral signatures can be applied to drug signature monitoring. Having shown the discriminating potential of the present diagnostic regimen; it is essential to study many samples producing acceptable sensitivity and specificity values. Discriminating temporal variation during disease progression or its suppression is also a critical follow-up step.

The present study provides a detailed insight into the molecular structural changes in serum samples of the IL10^{-/-} IBD mouse model. Infrared spectral signatures representing proteins, carbohydrates, nucleic acids, and lipids are identified as the potential signatures reflecting molecular changes. The success of the present technique in measuring the effect of anti-TNF α on the identified spectral signatures will provide an additional level of information about the treatment option of IBD patients and will increase the possibility of adapting this technique for disease status monitoring. Addressing the cost and hassle of colonoscopy that discourages people from being screened timely, doctors, particularly primary care physicians can make their patients more aware of IR spectroscopic analysis of blood serum. If the test shows an indication of diseases, colonoscopy can be performed. However, at its present form, the molecular mapping in serum samples as the IR signatures of the disease cannot replace gold standard colonoscopy tests but will provide additional information about the IBD patients before colonoscopy. Therefore, the feasibility of FTIR spectroscopy to extract a snapshot of cumulative molecular interactions within mouse serum samples for IBD study warrant a thorough investigation, as enabled by interdisciplinary collaborations between spectroscopists, biologists, and clinicians.

Using a statistically significant number of experimental mouse models with different stages of disease progression and healing can establish an association between identifying spectral signature changes with the disease stage. Mouse model study helps us to establish an association between spectral signatures of IBDs with histological findings of intestinal tissues. Understanding the association between identifying spectral signatures and the abolition or restoration of tissue functions in the intestinal tract is important because lymphocytes infiltrations and erosion of crypt can be seen in the diseased IL10^{-/-} mice.

Additionally, by building a logistic regression model for the spectral features and doing cross-validation, a correlation between the degree of variation in spectral signatures and the level of Lcn-2 in their stool or myeloperoxidase (MPO) activity of the distal colon can be established. Research into samples of mice with intermediate stages of the disease will provide more insights into the evolution of healing and add confidence to the analysis and the applicability of the technique for early diagnostics. However, the animal models are not enough for preclinical human disease studies and for confirming the potential clinical effects in humans. Therefore, a follow-up step would be to analyze a statistically significant number of human samples to examine covariate effects of demographic and clinicopathological variables like age, weight, ethnicity, diet, disease stage, and comorbidities on the spectral signatures.

3.1.3 Spectral Signatures in Serum of Human Colitis Patients

Discriminate Control and UC Sera Using Absorbance values: Figure. 3.6 (A) shows a typical min-max normalized[24] spectrum covering the fingerprint region for biomolecules comprised of various biological functional groups of lipids, proteins, nucleic acids, and carbohydrates. The spectral difference (UC-Control), as shown in the solid green line (—) of Figure. 3.6 (B), is significantly higher than the difference obtained from the multiple

measurements of the same sample for both Control (— —) dashed and UC (••••) dotted lines, almost completely overlapping). Figure. 3.6(C) shows the p-values of the Student's two-sample t-test (with two-tailed unequal variance) for the absorbance representing 28 healthy volunteers (controls) and 18 UC patients. P-values less than the nominal significance level of 0.05 indicate the most discriminatory features of the spectrum. P-values less than 0.025 show the validity of all the features, except for regions of lipid (pink line — —).

The prominent, discriminatory regions include C=O/C-N stretching and N-H bends in amides, C=O in stretching lipids, RNA/DNA nucleotides and C-O vibrations of carbohydrates[153]. Because of the complexity of biological systems, biomolecular and bonding vibration assignments are tentative and are based on numerous studies[23, 25, 147, 154-156]. The spectral region, 1700-1750 cm^{-1} , is primarily recognized by its C=O stretching band in lipids[23, 147]. The amide I (1600-1700 cm^{-1}) vibration is known to be sensitive to the secondary structure of proteins[25], which mainly arises due to C=O stretching vibration with minor contributions from out-of-phase C-H stretching vibrations, C-C-N deformation and N-H in-plane bends. Spectral band around 1578 cm^{-1} is due to C=N adenine[147] while the region 1136-1153 cm^{-1} is due to the ring vibration mode of C-O-C including C-O-H, C-O and C-C stretch[154]. The spectral band 900-1091 cm^{-1} results from C-O, C-C stretch, C-H bend, deoxyribose/ribose DNA, RNA, is(PO_2^-) [154] of carbohydrates and nucleic acids.

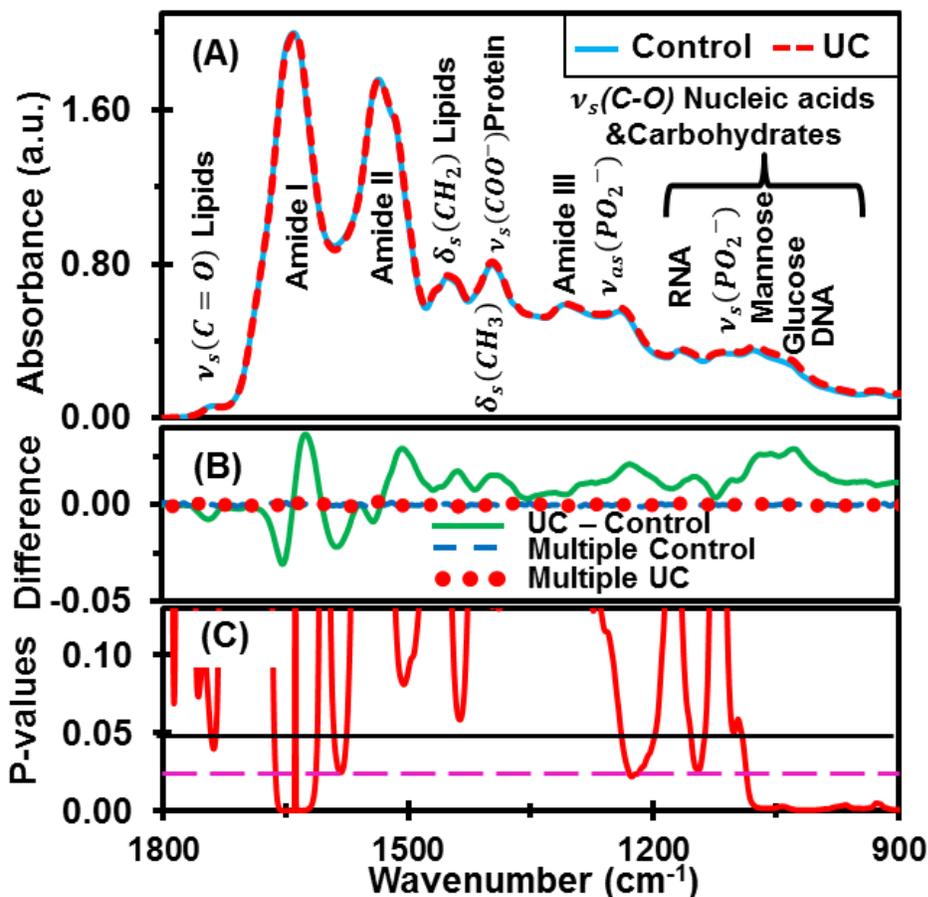


Figure 3.6 Infrared absorption spectra of sera to discriminate control and UC conditions.

(A) Ensemble average representative of min-max normalized absorbance spectra of 28 control serum samples (healthy volunteers) and 18 UC patients indicating biomolecular assignments. (B) The spectral response difference between UC and control. The average absorbance values of UC conditions are elevated at carbohydrate and nucleic acids regions and fluctuating up and down in the Amides region. The difference between repeat measurements of the same sample (Multiple Controls and Multiple UCs) shown by the (— — the blue dashed and •••• the red dotted lines respectively) lines are negligible compared to the difference spectra between the average UC and average control (— green; UC-Control). (C) p-values of the Student's t-test (with two-tailed unequal variance) for the absorbance representing healthy volunteers and UC patients. p-values

less than 0.05 are used to highlight the strongest discriminatory features of the spectrum. p-values less than 0.025 (horizontal  pink-dotted line) show the validity of all the features except lipids.

Studies have investigated various serological markers[43] for UC diagnosis and the discrimination between UC types with their clinical feasibility[150]. These studies have reported abnormalities in lipids, amino acids, and energy metabolisms in the serum samples of IBD patients[151]. Associations[152] between fatty acids and inflammatory cytokines or the protein abundance and the epigenetic alteration in the samples of UC patients have been established. As such, variations in these reported serological markers are most likely the primary reason for the UC induced changes in absorption frequencies of functional groups of lipids, proteins, nucleic acids and carbohydrates.

Protein Secondary Structure Analysis Via Spectral Deconvolution: The amide I band[25], $1600\text{-}1700\text{ cm}^{-1}$, has been used for the analysis of protein secondary structures, especially to observe their alteration due to changes in health conditions. We have also implemented a spectral deconvolution technique within the amide I band to find the protein secondary structure alteration in sera due to UC. The spectral band $1600\text{-}1700\text{ cm}^{-1}$ is deconvoluted into six Gaussian band energy profiles representing the side chain ($\sim 1610\text{ cm}^{-1}$), β -sheet ($\sim 1630\text{ cm}^{-1}$), random coil ($\sim 1645\text{ cm}^{-1}$), α -helix ($\sim 1652\text{ cm}^{-1}$), β -turn ($\sim 1682\text{ cm}^{-1}$), and β -sheet anti-parallel ($\sim 1690\text{ cm}^{-1}$) structures. Herein, the minima of the second derivative spectra are used to find the position and the number of Gaussian function energy bands enough to fit the experimental absorption curve. The integral area covered by energy bands representing the α -helix and β -sheet is larger compared to other structures: side chain, antiparallel beta-sheet, disordered structure, and

beta turns. In addition, the UC-associated alteration in the integral bands is also significantly larger in the α -helix and β -sheet structures, hence only these two structures will be analyzed.

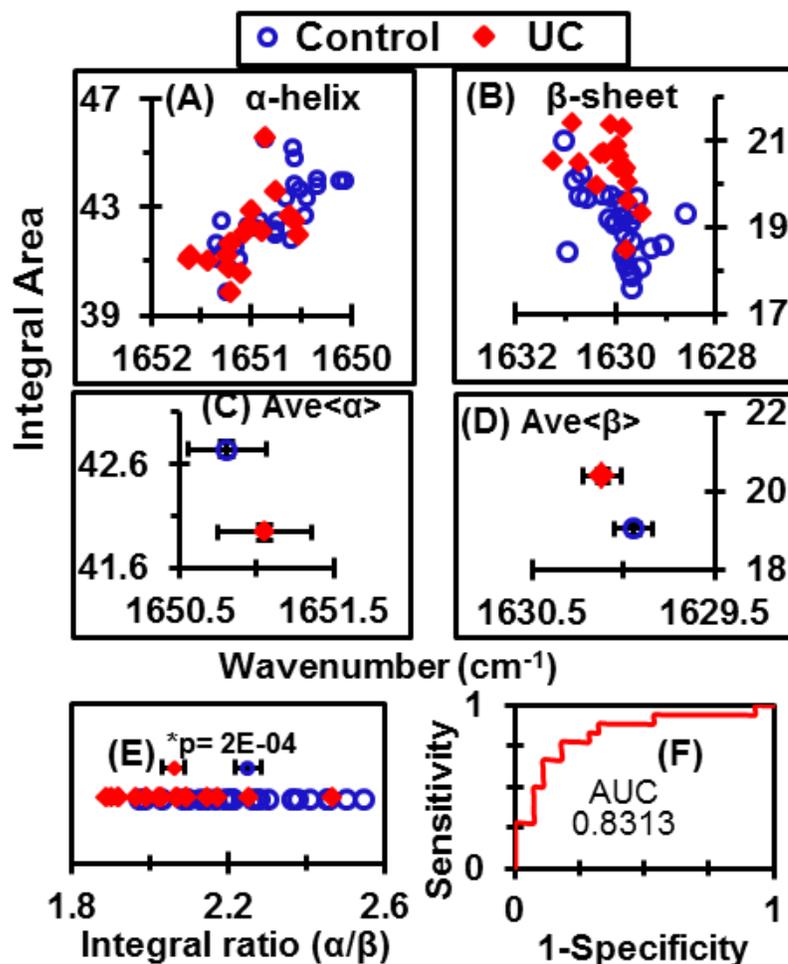


Figure 3.7 Plots of integral values of Gaussian function energy bands representing α -helix, β -sheet protein secondary structures in human sera. (A) Quantified integral values of α -helix components. (B) The integral values of β -sheet components. (C) Ensemble average representation of integral values of α -helix. P-values for the integral area of α -helix is 0.05 and for the mean position is 0.03. (D) Ensemble average representation of integral values of the β -sheet. P-values for the integral area of β -sheet is $9E-07$ and mean position is 0.3. (E) Representation of integral

ratios between corresponding α -helix and β -sheet structures. Average values of control and UC cases for the integral ratio of α -helix to β -sheet (Integral (α/β)) shows a statistically significant difference with a p-value of 1E-04. (F) The ROC curve of the Integral (α/β) and the corresponding AUC.

Figure. 3.7 shows a plot of the integral values of Gaussian energy bands representing the α -helix and β -sheet protein secondary structures. The integral values of the α -helix components are less for UC cases compared with controls (Figure. 3.7 (A)), while the integral values of the β -sheet components are higher in UC cases compared with controls (Figure. 3.7 (B)). A significant difference between the integral ratios of the α -helix and β -sheet structures can be obtained between the control and UC groups in this study (Figure. 3.7 (C)). The mean wavenumber position, 1650.81 ± 0.07 , of the energy band representing the α -helix structure has significantly moved towards the higher wavenumber position, 1651.05 ± 0.32 (with a p-value = 0.03), as shown in Table 3.5. Similarly, the integral area covered by the energy-band representing the β -sheet structures are different (p-value = 9E-07) between the control (19.05 ± 0.15) and UC (20.40 ± 0.17).

Table 3.4 Quantified information about average integral values and average position of α -helix and β -sheet secondary structure, energy bands in human serum samples.

Features	Types	Integral \pm std. error	Significance (p-values)	Wavenumber \pm std. error	Significance (p-values)
α -helix	Control	42.74 \pm 0.25	No (> 0.05)	1650.81 \pm 0.07	Yes (< 9E-07)
	UC	41.94 \pm 0.08		1651.05 \pm 0.32	
β -sheet	Control	19.05 \pm 0.15	Yes (< 0.03)	1629.94 \pm 0.11	No (> 0.3)
	UC	20.40 \pm 0.17		1630.12 \pm 0.11	

Identified Spectral Signatures: Identified spectral signatures in this study are I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , and α/β secondary structure analysis. Histogram representation of ensemble

averages for four spectral signatures, I_{1033} , I_{1076} , I_{1121}/I_{1020} , and I_{1629}/I_{1737} , are shown in Figure. 3.8 (A), 3.8 (B), 3.8 (C) and 3.8 (D), respectively.

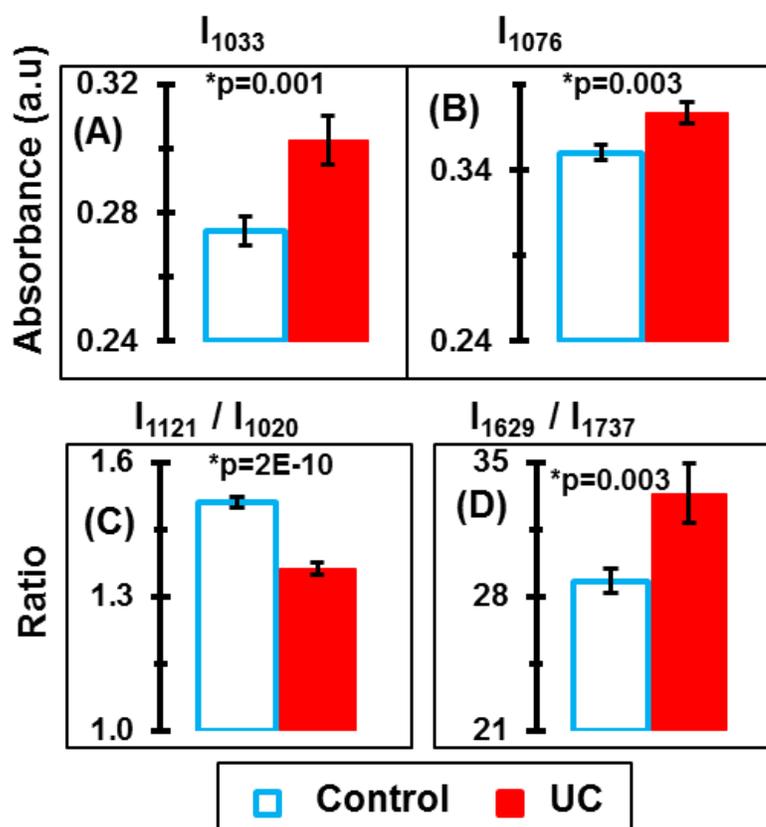


Figure 3.8 Histogram representation of identifying spectral signatures. (A) Ensemble average representative of absorbance of wavenumber position 1033 cm^{-1} . (B) Ensemble average representative of absorbance of wavenumber position 1076 cm^{-1} . (C) The ratio of absorbance values (I_{1121}/I_{1020}) at wavenumber 1121 cm^{-1} to 1020 cm^{-1} . (D) The ratio of absorbances (I_{1629}/I_{1737}) at wavenumber 1629 cm^{-1} to 1737 cm^{-1} .

The altered pattern between UC and control groups can be seen by comparing the range of data points and their average values in the identified spectral signatures as tabulated with standard error in the unshaded region of table 3.6. The average absorbance value of the glucose peak for

control and UC is 0.274 ± 0.004 and 0.303 ± 0.007 and the average for the mannose peak (for control and UC) are 0.349 ± 0.004 and 0.373 ± 0.006 , respectively. Similarly, the average ratio of the RNA peak to the DNA peak is 1.510 ± 0.011 (control) and 1.362 ± 0.015 (UC), the average ratio of the peak representing protein to lipid peak is 28.804 ± 0.642 (control) and 33.373 ± 1.519 (UC) and the average integral ratio of α -helix to β -sheet structure is 2.250 ± 0.030 (control) and 2.061 ± 0.033 (UC).

Sensitivity and Specificity: The discriminating potential of any diagnostic technique is further evaluated by sensitivity, specificity, ROC, AUC. Figure 3.7 (F) shows the ROC curve of the integral ratio (α/β) and the corresponding AUC. Similarly, for each feature (Table 3.6) optimal cutoff values, AUC values, sensitivity and specificity were calculated with the aid of ROC curves. Quantified information about these features for corresponding spectral signatures are shown in the shaded region in Table 3.6. The maximum values of sensitivity and specificity values of each feature describe the differences between the UC versus the control groups.

Table 3.5 Quantified information about identifying discriminatory bands. These include absorbance values at 1033 cm^{-1} and 1076 cm^{-1} , the ratios of absorbance at 1121 cm^{-1} to 1020 cm^{-1} , 1629 cm^{-1} to 1737 cm^{-1} . The integral ratios of Gaussian function energy bands representing α -helix and β -sheet protein secondary structures are also discussed. Statistical measures of discriminatory infrared spectral signatures for UC using sera shown in the shadowed region of the table. The optimal cut-off and the corresponding sensitivity and specificity are calculated based on the Youden index. P-values are calculated based on the student's t-test with a two-tailed unequal variance.

Types		I ₁₀₃₃	I ₁₀₇₆	I ₁₁₂₁ /I ₁₀₂₀	I ₁₆₂₉ /I ₁₇₃₇	Integral(α/β)
Control	Range	0.244- 0.328	0.325- 0.405	1.482- 1.611	21.124- 34.066	1.973- 2.547
	Average \pm std. error	0.274 \pm 0.004	0.349 \pm 0.004	1.510 \pm 0.011	28.804 \pm 0.642	2.250 \pm 0.030
UC	Range	0.257- 0.375	0.329- 0.416	1.231- 1.444	22.261- 48.055	1.896- 2.469
	Average \pm std. error	0.303 \pm 0.007	0.373 \pm 0.006	1.362 \pm 0.015	33.373 \pm 1.519	2.061 \pm 0.033
Cut-off value		0.273289	0.351057	1.44565	32.8145	2.09855
AUC		0.7817	0.7619	0.9444	0.7619	0.8313
Sensitivity (%)		89	78	100	61	78
Specificity (%)		68	68	86	93	82
p-value		0.001	0.003	2E-10	0.003	0.0001

3.2 Spectral Signatures of Lymphoma and Melanoma in Serum Samples of Mouse Models

Discrimination of Absorbance Values: Figure 3.9 (a) shows the average normalized ATR-FTIR spectrum of air-dried serum samples extracted from tumor-bearing mouse models of EL4 lymphoma (n=8) and B16 melanoma (n=8) in wild types and corresponding control types (n=15). Using the student's t-test, p-values (two-tailed unequal variance), the most discriminatory

features of the spectrum within the spectral range $1800\text{-}900\text{ cm}^{-1}$, were extracted (figure 1(b)). Interestingly, the features observed for different groups enable the classification between control cases and malignant cases and between the two malignant cases of lymphoma and melanoma. Molecular assignments[33, 157-161] of five spectral bands showing discrimination of EL4 lymphoma from their control types, with higher significance (i.e. $p\text{-values} < 0.05$) are presented in Table 1. These are the bands originating from (i) amide I of protein, (ii) amide II of protein (iii) C-H deformation of CH_3/CH_2 groups, (iv) asymmetric phosphate I, and (v) Carbohydrates and nucleic acids. Similarly, two spectral bands showing the significant difference between B16 melanoma and their control types are also shown in the shaded regions of the table. Significant alteration in the amide I band, and the complex band of carbohydrate and the nucleic acids are observed for B16 melanoma. The difference in the $p\text{-values}$ observed between lymphoma and melanoma could be attributed to the difference in mechanism of each type of tumor development, while similarity could be attributed to common etiology[162].

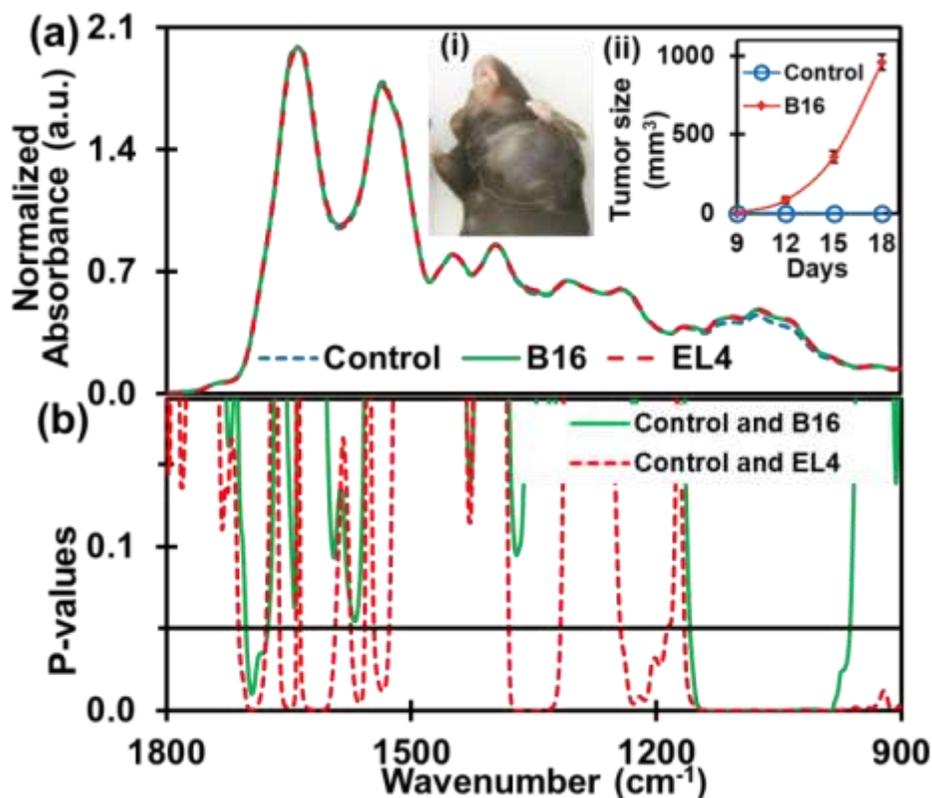


Figure 3.9 The discriminatory region of infrared absorbance spectra. (a) Average normalized ATR-FTIR spectra of serum samples extracted from EL4-lymphoma ($n=8$), B16-melanoma ($n=8$) mouse models in wild types and corresponding control types ($n=15$). The inset (i) shows B16-melanoma mouse with tumor size approximately 1000 mm^3 (day 18). An increase in volume of the tumor from day 9 to day 18 of tumor inoculation in B16-melanoma mouse is as in inset (ii). Like the B16 mouse, the elevation of tumor size is also monitored in EL4-lymphoma. The serum sample is extracted for both types of mice when tumor size becomes bigger than 1000 mm^3 . (b) Student's *t*-test (two-tailed unequal variance) *p*-values of absorbance. Discriminatory region for lymphoma with higher significance ($p < 0.05$) are amide I of protein, amide II of protein, C-H bends of CH_3/CH_2 groups in α - and β - anomers, asymmetric phosphate I, and carbohydrates with predominant contributions nucleic acids (DNA/RNA via PO_2^- stretches). Discriminatory regions of melanoma are amide I and carbohydrates with predominant contributions of nucleic acids[12].

Table 3.6 Discriminatory infrared spectral bands of lymphoma and melanoma serum with biomolecular assignments[12].

Wavenumber region (cm ⁻¹)		Assignments (taken from references[33, 157-161])
i	1700-1600	Amide I of proteins: (α -helical, β -pleated sheet, β -turns, random coils and side-chain structures), $\nu(\text{C=O})$, $\nu(\text{C-N})$, CNN.
ii	1480-1580	Amide II of proteins: (α -helical, β -pleated sheet, unordered conformation structures), $\delta(\text{N-H})$, $\nu(\text{C-N})$.
iii	1325-1380	C-H deformation: due to CH_3/CH_2 bending (groups in α and β anomers) of lipids and proteins.
iv	1190-1240	Asymmetric phosphate I: $\nu_{\text{as}}(\text{PO}_2^-)$ of lipid phosphates.
v	1000-1140	Carbohydrates and nucleic acids: C-O, C-C stretch, C-H bends, deoxyribose/ribose DNA, RNA, $\nu_{\text{s}}(\text{PO}_2^-)$.

Protein Secondary Structure Analysis by Deconvolution of Amide I Band: In order to demonstrate alterations in structural components due to malignancy, integral values of α -helical and β -sheet structures and their ratios were statistically analyzed. Figure 3.10 (a) and 3.10 (b) show the cluster plots of the integrals of α -helical and β -sheet structures respectively for the control, B16 and EL4 mice. These figures clearly demonstrate a separation between the corresponding integral values for the control and tumorous groups for β -sheet and α -helix. Furthermore, the ratio of integral values α -helix to β -sheet (figure 3.10 (c)) is always less than the control values for both mouse models with greater than 99% significance.

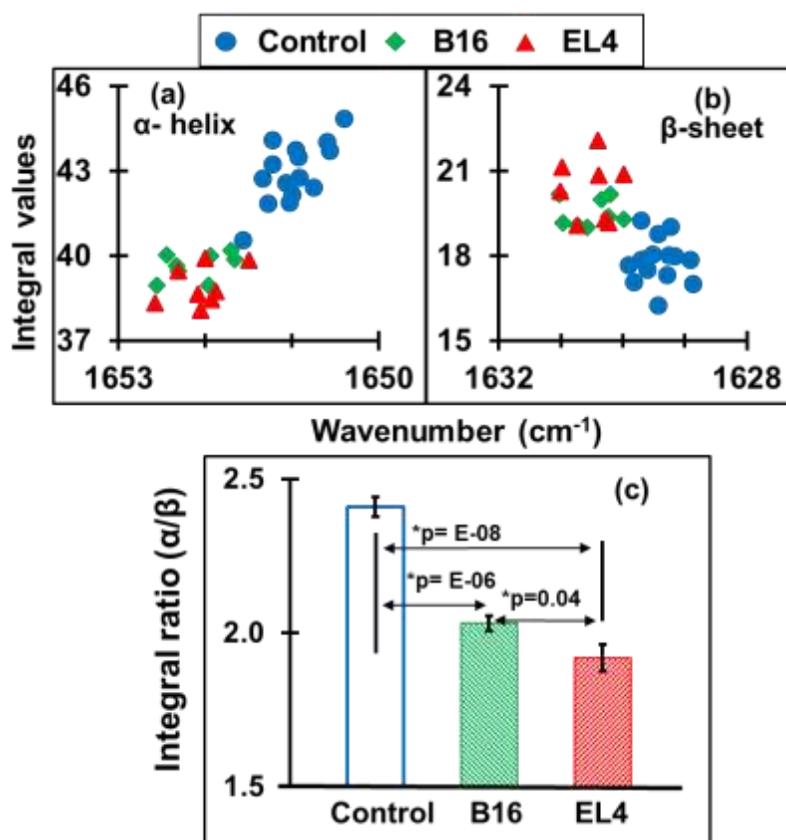


Figure 3.10 Plots of the protein secondary structures (α -helix, β -sheet) and their ratio [12].
(a) Quantified integral (area covered) values of α -helix components are less for tumorous cases compared to control. **(b)** Integral values of β -sheet components are higher for tumorous cases compared to control **(c)** Bar graph representation of average integral ratios between α -helix and β -sheet for the control, B16 and EL4. Significant alteration in integral ratio (α -helix/ β -sheet) is found between control and tumorigenic case.

Amide I and II Absorbance Values: Amide I and Amide II are the two major bands of the infrared spectrum for protein interrogation in biological materials [157, 158]. The intensity and position of these bands, determined by backbone conformation of the hydrogen bonding pattern change with malignancies [62, 163]. Amide I band position shifts towards the lower wavenumber due to malignancy. The average position of amide I representing control is at 1641 cm^{-1} , B16 is

1640 cm^{-1} and that of EL4 is 1638 cm^{-1} , but the position of amide II stays the same at 1538 cm^{-1} for all three types. Altered position of amide I is statistically significant for EL4 (*p = 0.001) while that of B16 is not significant (*p = 0.2). Similarly, the altered ratio between amide I and amide II absorbance values are significant (*p = 0.01) for EL4 lymphoma, but not (*p = 0.3) for B16 melanoma in comparison to the control groups.

Nucleic Acids and Carbohydrate Analysis: In the region 1140-1000 cm^{-1} , there are plenty of overlapping vibrational modes of biological macromolecules[23] with the major contribution of nucleic acids and carbohydrates[24]. Bands approximately at 1121 cm^{-1} arise from RNA absorbance, whereas the band at 1020 cm^{-1} arises from DNA absorbance[145]. The spectral band near 1080 cm^{-1} is due to $\nu_s(\text{PO}_2^-)$, and the band approximately at 1056 cm^{-1} corresponding to the $\nu_s(\text{PO}_2^-)$ absorbance of phosphodiester of nucleic acids and the O-H stretching coupled with C-O bending of C-OH groups of carbohydrates[164]. Similarly, absorbance near 1033 cm^{-1} and 1076 cm^{-1} are due to the presence of glucose (C-O stretching carbohydrate, β -anomer) and mannose (C-O stretching carbohydrate α -anomer). Alteration in the concentration of two sequences of basic genetic materials- (a) RNA (which play an active role in protein synthesis) and (b) DNA (which is primarily involved in the storage, copying and transferring genetic information), has been already reported from the tissue analysis of NHL[145] and subcutaneous melanoma[165]. Due to the fluctuation in these biomolecules, there is a dissimilarity between malignant groups from their control types. In order to verify these dissimilarities, we have used Hierarchical Cluster Analysis (HCA) along with spectral deconvolution within this spectral range.

Dendrogram of hierarchical cluster analysis: HCA is commonly employed to identify the similarities between the FTIR spectra by using the distances between spectra and aggregation algorithms[63]. The dendrogram of HCA is performed with ATR-FTIR spectra of control, B16,

and EL4 mice are shown in figure 3.11. Dendrogram tree diagram performed using spectral region of nucleic acids and carbohydrates, $1140-1000\text{ cm}^{-1}$, using Ward's algorithm and squared Euclidian distance measurements, allow us to visualize of overall grouping structure, including the sub-groups. The distinct cluster for the control spectra which are grouped together, describing a high degree of similarity within the groups. Similarly, there is a distinct clustering in the cancer spectra showing the higher degree of heterogeneity between spectra of cancerous groups.

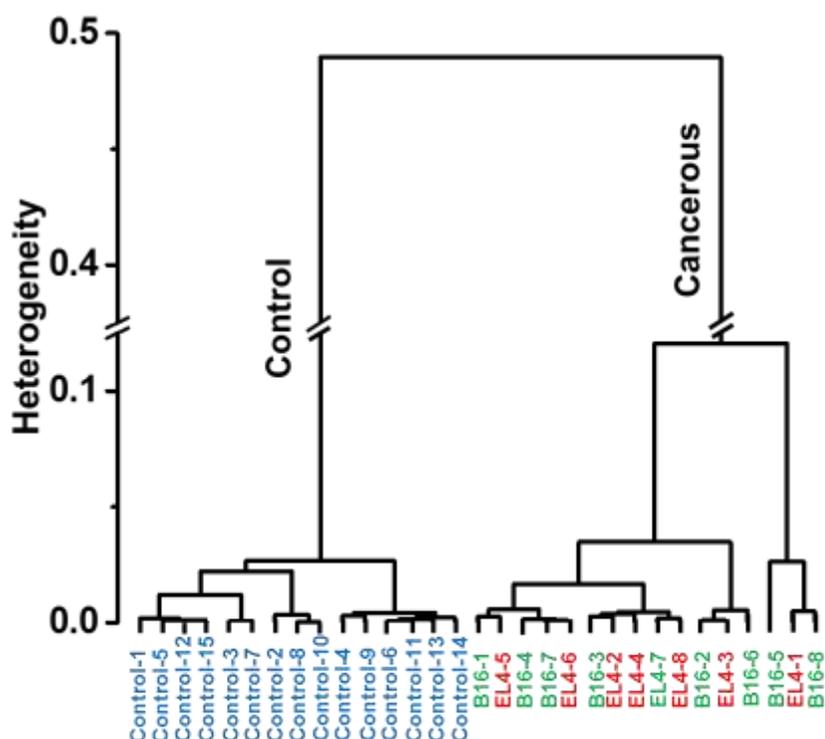


Figure 3.11 Dendrogram of hierarchical cluster analysis[12]. The dendrogram tree diagram performed within spectral range $1140-1000\text{ cm}^{-1}$, by using Ward's algorithm and squared Euclidian distance measurements. The spectra are correctly classified. Control spectra appear grouped together, which describes a high degree of similarity within the groups. Similarly, there is a distinct clustering in the cancer spectra in two subgroups showing the higher degree of heterogeneity between cancerous spectra.

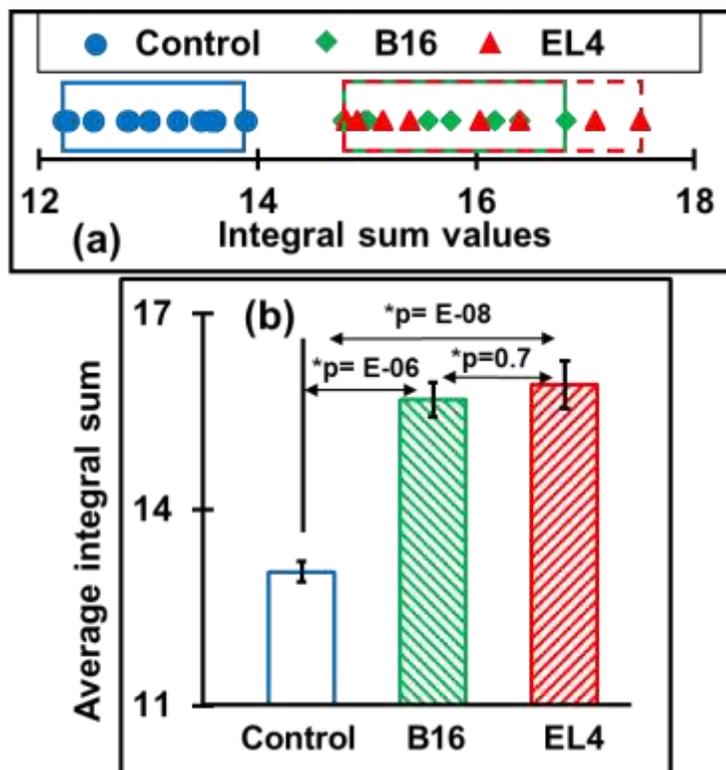


Figure 3.12 The integral sum of Gaussian energy profiles used to fit experimental curves within $1000-1140\text{ cm}^{-1}$. (a) The sum of integral values of control groups cluster within the approximate range 12-14, B16 covers the range 15-17 and EL4 covers 15-18. (b) Bar graph representation of the average value of the integral sum which shows significant differences between control and tumorigenic case[12].

Furthermore, to quantify tumor-associated alteration within this complex spectral region of $1140-1000\text{ cm}^{-1}$, deconvolution of experimental spectra into Gaussian function band profiles is further employed. Six Gaussian function energy band profiles (figure 5(b)) are used to fit the spectra by approximating number and position using the minima of second derivatives (figure 5(a)). The sum of the integral areas covered by six bands (integral values) is then statistically analyzed to evaluate the tumor-associated alteration in the serum. A calibration curve is obtained,

as shown in figure 3.12(a) between control and tumorous groups. A clear separation between control (12-14) and cases of tumorigenicity B16 (15-17) and EL4 (15-18) is found while adding integral values. The bar graph representation of these values with significance greater than 99% is shown in figure 3.12(b).

Table 3.7 Quantified values of discriminatory features. Clear separation can be seen between control and cancerous cases (both B16 and EL4) while comparing integral values of α -helix, β -sheet structure components, and their ratios. Similarly, the altered position of amide I peak, amide I/amide II ratio and absorbance values at 1212 and 1335 cm^{-1} show significant difference only between EL4 and control.

Feature	Spectral deconvolution				Amide I and II		Absorbance	
	Amide I 1600-1700 cm^{-1}		Ratio (α/β)	Mixed region 1000-1140	Position amide I	Ratio: abs Amide I/II	ν_{as} (PO_2^-) ~1212 cm^{-1}	C-H def. ~1335 cm^{-1}
	Integral values			Integral sum				
	α -helix	β -sheet						
Control	40.6- 44.9	16.3- 19.3	2.2- 2.7	12.2-13.9	1640- 1645	1.10-1.13	0.42- 0.44	0.54- 0.59
B16	38.9- 40.2	19-20.2	1.9- 2.1	14.8-16.8	1638- 1644	1.10-1.15	0.42- 0.45	0.54- 0.59
EL4	38.1- 39.9	19.1- 21.1	1.7- 2.1	14.8-17.1	1636- 1642	1.12-1.15	0.43- 0.46	0.58- 0.59

3.3 Spectral Signatures of BC in Serum of Human Patients

Using absorbance spectral data of serum samples (using $n = 10$ each BC and controls), we investigate the applicability of FTIR spectroscopy to discriminate between the control and cancer sera. ATR sample mode of FTIR spectroscopy is used, and the discrimination between the control and test groups were done using various data analysis techniques. The investigation involves

multivariate analysis, p-value calculation, quantification by spectral deconvolution and is followed by statistical analysis.

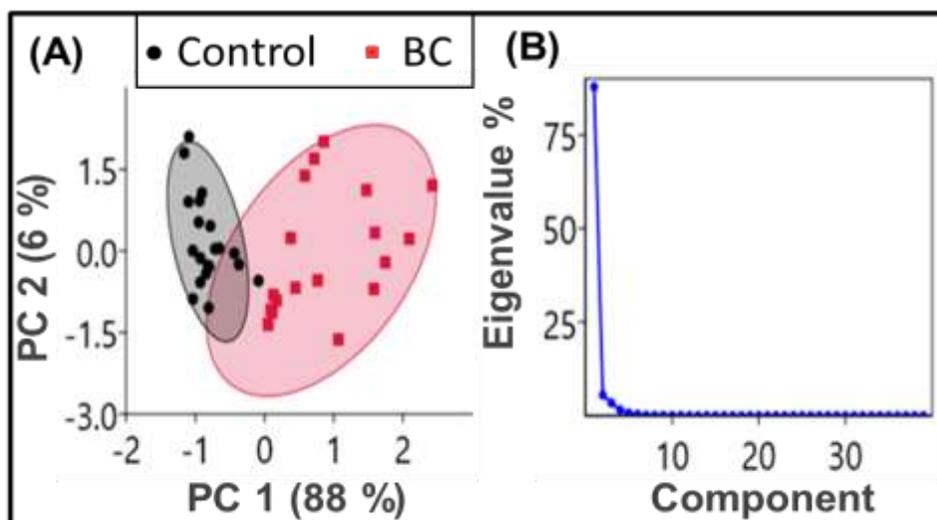


Figure 3.13 PCA analysis of second derivatives curves of FTIR absorbance spectra[13]. (A) PCA scores plots (PC1 x PC2) with 95% confidence ellipse. The data related to control groups (•• black dots enclosed by a black shaded ellipse) and BC (▪▪ red dots surrounded by a red-shaded ellipse) are clustered together with different magnitudes and directions. (B) Scree plot of eigenvalues showing the percentage variance of components one and two is significant compared to others.

Principal Component Analysis: (PCA), a useful statistical analysis[137], is first performed to explain the holistic evaluation of protein structural content variations reflected in amides (amide I and II region, $1480\text{-}1600\text{ cm}^{-1}$). Herein, each of the ten samples is measured twice (measurement replicates) to obtain 20 spectral data of BC and 20 of control. Using the “PAST (PAleontological STatistics) 4 - the Past of the Future” software and the vector normalized second derivative curve of the absorbance spectra within $1480\text{-}1600\text{ cm}^{-1}$, as input data variables, we

analyzed the variance-covariance matrix with the pairwise exclusion of missing values to get the component plots. The output of the component plot with 95% ellipses is showing (Figure 3.13 (A)) a clear separation between each studied group. The scatter plot of PC1 (variability 88%) and PC2 (variability 6%) shows the data related to control and BC groups are clustered together with different magnitudes and directions. Figure 3.13 (B) is the scree plot showing the total variance presented by PC1 and PC2 are significant. These findings from the PCA analysis of amide bands have led us to investigate spectral signatures useful in the clinical domain.

Discrimination of Average Absorbance: The average of normalized absorbance spectra for both control and BC sera that includes the fingerprint region of biological functional groups (lipids, proteins, nucleic acids, and carbohydrates) is shown in Figure 3.14(A). Solely by looking at the FTIR spectra, it is difficult to discriminate between the absorbance of the functional components of the two groups. However, the comparison of the absorbance spectra between the two groups using student's t-test (with two-tailed unequal variance) revealed the discriminating potential at amide regions ($1541\text{-}1656\text{ cm}^{-1}$) and the mixed regions of carbohydrates and nucleic acids ($1018\text{-}1076\text{ cm}^{-1}$), as highlighted by the red ellipses in ($p < 0.05$) Figure 3.14 (B). The prominent, discriminatory regions include C=O/C-N stretching, and N-H bends in amides, RNA/DNA nucleotides, and C-O vibrations of carbohydrates[153], as reported in previous studies[89]. Previous studies using principal component analysis-linear discriminant analysis (PCA-LDA) of FTIR spectra have shown that healthy and cancerous serum samples had different characteristics[89].

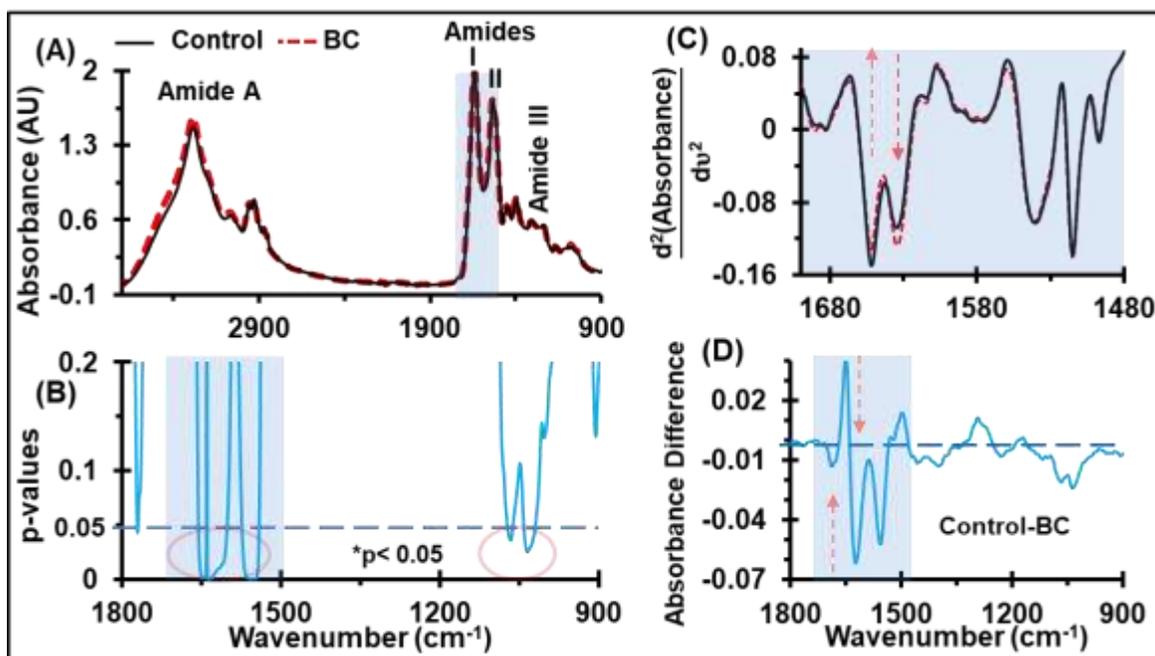


Figure 3.14 Identification of discriminatory bands. (A) Ensemble averages of normalized serum spectra derived from control, $n = 10$ and BC, $n = 10$. This wider range of spectra is presented to show the quality of the spectra, which overcomes the noise and atmospheric contamination while measuring them at resolution 4 cm^{-1} . (B) Corresponding student t -test p -values for the control and BC. (C) The second derivative absorbance spectra confined to the amide-I region, covering $1600\text{-}1700 \text{ cm}^{-1}$. (D) Difference between the absorbance spectra of control and BC, indicating up- and down-regulation of proteins, carbohydrates, and nucleic acids in the serum of BC patients. Figure is taken from our own publication[13].

Molecular assignments of major spectral bands showing discrimination between control and BC, with higher significance (i.e., p -values < 0.05), are also presented in Table 3.8. These are the bands originating from amides of protein, carbohydrates, and nucleic acids. The amide vibrations are mainly arising from C=O stretching vibration, with minor contributions from out-of-phase C-H stretching vibrations, C-C-N deformation, and N-H in-plane bend[25]. Similarly, the mixed

regions of carbohydrates and nucleic acids result from C-O/C-C stretching, C-H bends, and $\nu_s(\text{PO}_2^-)$ [154]. The second derivative spectra of these absorbance curves revealed that the absorbance at the minima positions at wavenumber 1629 cm^{-1} and 1652 cm^{-1} differ between healthy individuals and BC patients (Figure 3.14 (C)). Elevation of absorbance values at the energy band $1018\text{-}1076\text{ cm}^{-1}$ (Figure 3.14(D)) suggests differences in glycomic profiling[166] and circulating DNA[167] in the blood components. Circulating DNA and glycomic profiling has proven to be critical molecular markers[166, 167] in several tumor entities.

***Table 3.8** Discriminatory IR bands for BC serum samples from controls, and primary biomolecular assignments, giving rise to the main contributions for the absorbance. Amide regions and the complex region of carbohydrates and nucleic acids show the discriminating potential[13].*

Wavenumber (cm^{-1})	Biomolecular Assignments
1700-1600	Amide I: Sensitive to protein secondary structures of proteins, which arises mainly due to C=O stretching vibrations and by the C-N groups.
1580 -1480	Amide II: Sensitive for protein conformation, originates mainly from the in-plane N-H bending mode along with C-N and C-C stretching vibrations.
1140-1000	Carbohydrates: Sensitive to C-O, C-C stretch, C-H bends and Nucleic acids: Sensitive to deoxyribose/ribose DNA, RNA, $\nu_s(\text{PO}_2^-)$.

Discrimination of Protein Secondary Structures: Figure 3.15 (A), the average of the second derivative spectra at the amide I absorbance region is shown. The minima of the second derivatives of spectra allow us to approximate the positions and numbers of Gaussian function energy profiles required to fit the experimental curve. The amide I band of each spectrum was deconvoluted so that the baseline-corrected spectra were fitted with 6 GFEB profiles by estimating the number and

position of the minima of second derivatives, which was simulated (•••) to fit the experimental curve (—) as shown in Figure 3.15(B).

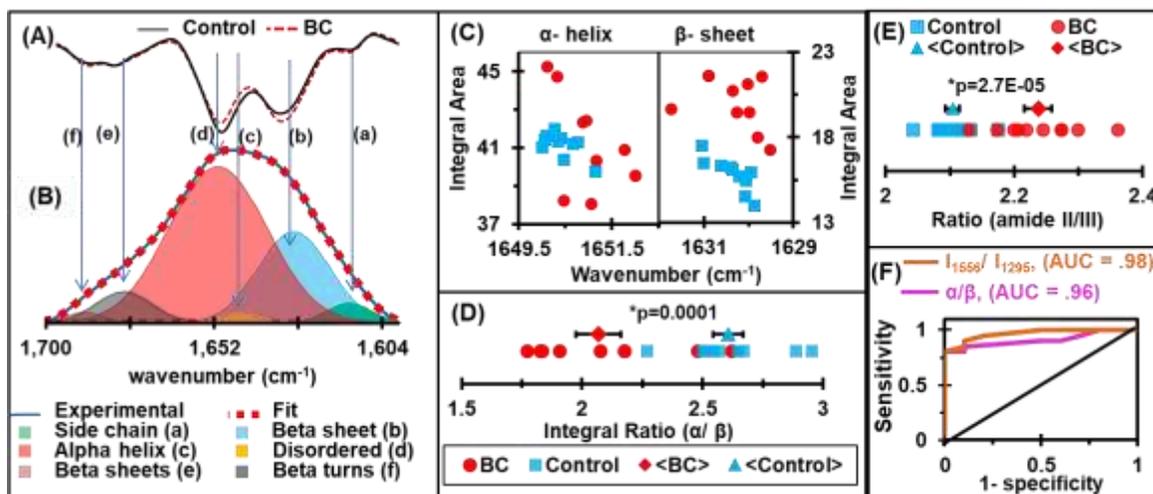


Figure 3.15 Protein secondary structure analysis[13]. (A) Representative second derivatives of absorbance spectra at the amide-I absorbance region. (B) Deconvolution of the amide-I region: the baseline-corrected spectra fitted with 6 GBEF by approximating number and position of minima of second derivatives, which simulated fits (•••) to the experimental curve (—). (C) Integral area of GBEF representing α helix and β sheet. (D) The ratio of α helix and β sheet energy bands, which proves an elevation of β sheet and drop off α helix structures due to malignancies. (E) The ratio of IR absorbance at amide II (I_{1556}) to its value at amide III (I_{1295}). (F) ROC curves for the ratio of the integral area of energy bands representing α -helix and β -sheet protein secondary structures and the respective absorbance at amide II and amide III. The maximum values of sensitivity and specificity are 90% and 90% for signature α/β , while these values are 100% and 80% for signature I_{1556}/I_{1295} , respectively.

In order to assess any alterations in structural components associated with malignancy, the integral values of α -helix and β -sheet structures and their ratios were analyzed. Due to the fact

that the intensity of the GFEB has a linear relationship with the concentration according to the Beer-Lambert law[168], the width of GFEB and full width half maximum (FWHM), is inversely related to the vibrational mode lifetime, which is a function of the ‘rigidity’ of the vibrating bond[103]. The interaction of the molecule with its immediate environment also affects the width of GFEB[169]. If a molecule transfers energy to its surroundings, the spectral peak has a broader line width and reduced intensity, even though the concentration of the molecule remains unchanged. In such cases, the integral area under the curve is a better indicator of the concentration than the intensity alone. Interestingly, we found that even though the levels of most structures did not differ between the samples from BC patients and healthy individuals, BC samples had an increase in β -sheet structures, while the levels of α -helix structures were decreased (Figure 3.15 (C) and 3.15 (D)). Furthermore, the amide II region is used to report on protein unfolding based on the extent of hydrogen exchanged. Because of the lack of water interference amide III region is also considered as a promising region to analyze protein structures. Herein, we have also used a ratio of IR absorbance at the amide II (I_{1556}) to its value at the amide III (I_{1295}) for analysis of BC associated protein alteration. The dot plots of these amides ratios are shown in figure 3.15(E).

ROC Curves and AUC Values: Sensitivity and specificity of a diagnostic test are often used to describe the diagnostic accuracy/performance of the analysis in biomedical research. The discriminating potential of a diagnostic regimen can be quantified by the AUC values of ROC curves [142]. The ROC curve is plotted to find the AUC, as in Figure 3.15(F). The optimal cutoff value calculated for each spectral signature is used to select the positivity/negativity of the disease and to estimate the sensitivity and specificity. Strong discrimination between diseased and control serum can be seen with 90% sensitivity and 90% specificity for signature α/β , and these values are

100% and 80% for I_{1556}/I_{1295} , respectively. The results indicated that the spectral signatures in the specified bands have high diagnostic accuracy.

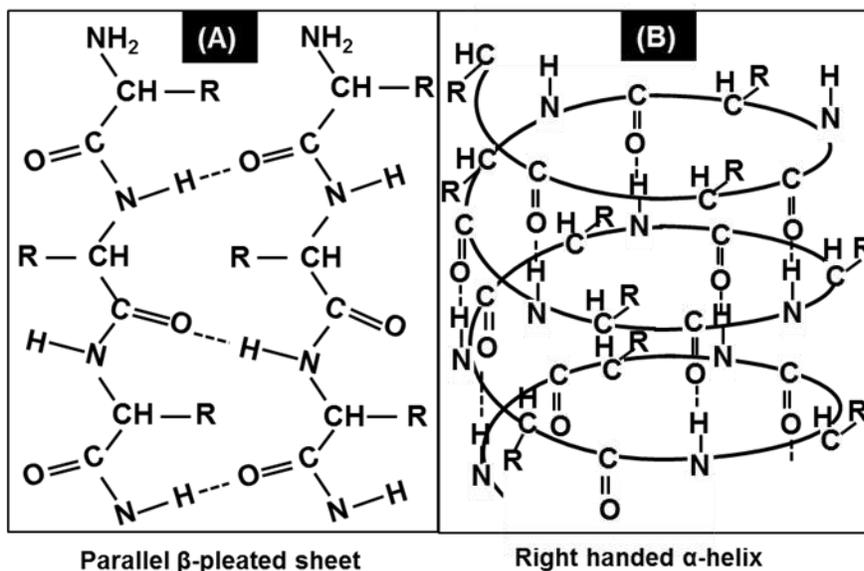


Figure 3.16 Understanding protein secondary structures and physics of IR interaction. (A) Parallel β -pleated sheet structure of proteins. N-H groups in the backbone of one strand form hydrogen bonds with the C=O groups in the backbone of the adjacent strand to form a β -sheet. (B) Right-handed α -helix structures of proteins. The backbone N-H group donates a hydrogen bond to the backbone C=O group, contributing to the helical structure of the α -helix[13].

As shown in Figure 3.16, the backbone N-H group donates a hydrogen bond to the backbone C=O group to form the helical structure of the α -helix (Figure 3.16 B). In contrast, the backbone N-H groups of one strand can form hydrogen bonds with the backbone C=O groups of the adjacent strands, resulting in β -sheet structures (Figure 3.16 A). Therefore, the cancer-associated alterations in the integral ratio of α -helix and β -sheet protein secondary structures suggest that protein conformational alterations accompanying changes in their biological function

might be a key event during the development of cancer. Several studies have shown that proteins in serum change during BC[88-90, 170]. The alterations in the conformational compositions are presumably due to alteration in the concentration of cancer embryonic antigen (CEA) proteins[84].

4 DISCUSSION

In this section, the results, possible clinical applications, optimization of the technique for clinical applications are discussed. This includes several projects, that have been working on while preparing this dissertation.

4.1 Discussion Based on DSS Study

This study demonstrates the application of the IR spectroscopic technique for monitoring colitis progression on days 0, 3, and 7 of chemical administration on DSS mouse models of colitis. It also reveals as a preliminary step in anticipating the validity of the test for humans and the development of a portable prototype which can facilitate an ASSURED, IBD screening. Monitoring alteration in the unique spectral markers [such as the concentration of carbohydrates (glucose and mannose) and the alpha-helix to beta-sheet ratio of the protein secondary structure] reflected in the IR spectra of serum, for colitis during disease progression and suppression will allow us to understand the feasibility of the presented diagnostic regimens in the clinical domain. Also, the DSS mouse started to behave differently at day 3 of the chemical feeding, so the present study using samples on days 0, 3, and 7 of chemical administration, is the foundation for further research with an increased number of sample size. Further work is in progress to investigate dose-dependent aggravation, and disease progression of intermediate stages at day five and its suppression after twelve days, after stopping the chemical feeding. Herein, the DSS mouse model eliminates few challenges of the human diseases studies and allows us to conduct disease research on the controlled environment, but there are several other challenging aspects to translate these findings into humanistic studies. The change in spectral signatures along with disease progression and suppression needs to be confirmed with the use of samples of human patients. Thus, in our future work, spectral signatures will be further confirmed by using serum samples from UC

patients. Once the unique spectral signatures are identified and confirmed, an ASSURED portable prototype can be developed in which measurements and data analysis techniques can be automated in the software, so that a technician will be able to perform the test from depositing the specimen onto the sample holder to the positive or negative results with a screen touch. This rapid, simple, cost-effective, and the minimally invasive technique would allow the assessment of disease status and personalized drug management in the future health care of IBD patients.

4.2 Discussion Based on IL10-/- Study

The application of ATR-FTIR spectroscopy to identify the treatment effect of anti-TNF α in IL-10-/- genotype IBD mouse model is demonstrated. By measuring spectra of serum samples of untreated, anti- TNF α treated and normal mice, we have compared the statistical significance difference between the groups and identify the discriminatory features. Identifying IR spectral signatures I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , α -helix/ β -sheet, and integral sum verify antibody therapy markedly ameliorated the disease as judged by earlier studies[58]. These data justify more initial findings and support the feasibility of this regimen for drug management during its treatment. The study of ATR-FTIR spectroscopy of serum samples thus needs to extend with a statistically significant number of experimental models and human participants.

Importantly, the standardization of technique for the acquisition of spectral signatures applicable in the clinical domain is explored. The improved standardization of spectral signature acquisition is needed to assure the potential clinical application of the technique for IBD screening. The standardization includes reducing the unwanted experimental and data analysis specific variances. Herein, using a trial-and-error based approach and the clinically relevant data sets on ATR-FTIR spectroscopy of serum samples, most optimized parameters were used for data taking

and their analysis. Diagnostic performance parameters, such as sensitivity, specificity, and AUC under ROC curves prove the feasibility of the studied technique.

All in all, the present study provides evidence that ATR-FTIR spectroscopy of serum samples accompanied by data analysis techniques could be a tool to evaluate anti-TNF α therapy in IBD. The study also provides a detailed insight into technical and instrumentation for its clinical applications. Understanding the drug signature efficacy by using IR spectroscopy of serum samples further strengthens the idea of the potential applicability of the technique. It is because; the access to this screening technique would increase the likelihood of compliance with screening recommendations for those who are otherwise reluctant to undergo the existing endoscopic tests.

4.3 Discussion Based on Colitis Human Patients Study

Early diagnosis and treatment of IBD is crucial because of its negative effects on one's overall personal well-being. Nonetheless, as for earlier diagnosis of IBD, the gold standard colonoscopy diagnostic tests are not easily available and/or are not attractive for all eligible populations. These are expensive, invasive, risky, and uncomfortable for patients[171]. Limited studies have reported alternative approaches[43], which could make way for potential advances in the medical screening of IBD[44]. However, these studies are not yet able to provide a low cost, quick, and easy early screening test as an alternative to colonoscopies.

The present, minimally invasive technique identifies spectral signatures as biomarkers associated with the development of disease and could become applicable for UC screening. This present technique will not replace traditional colonoscopy techniques but will provide an earlier indication of disease or an additional level of information about UC patients. Importantly, it can be regularly performed easily at a lower cost at the physician office during annual checkups to monitor the health of the patient with the possibility of an early indication of UC. Molecular

mapping in the form of IR signatures has the potential to become a routine screening regimen for the continuous assessment of patients. Further evaluation and resolution of UC can be performed afterward using established colonoscopy tests.

We identified I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} , and integral α/β structures as spectral signatures for UC, and the corresponding differences between them were statistically analyzed. Note that experimental mouse model studies also showed that these peaks are unique to UC when compared with metabolic syndrome and arthritis. The ROC curve, sensitivity, specificity, and p-value calculations further confirm the validation of our performance testing. All these spectral signatures' AUC values range from 0.76 to 0.94. These values serve as important pieces of evidence to demonstrate the validity of the classifier. The sensitivity and specificity pair[141] and the p-values of the two-sample t-test also verify that the identified signatures are valid for diagnostic purposes. Using this quantified information of identified spectral signatures, the required number of controls and UC cases applicable in medical diagnostic trials[172] can also be calculated. The sample size estimation for high diagnostic accuracy is discussed in Appendix A.3 in detail.

Overall, these results of preliminary research using human samples resemble our earlier findings using experimental mouse models of disease and establish this technique as a proof of concept for further investigation using a larger sample size followed by clinical studies. Following the same protocol for the sample care, spectral measurements and data analysis, calibration curve applicable for initial diagnosis, differentiate IBD types, the monitoring of therapeutic response, the prediction of relapse and the ruling out of any flare-ups can be accomplished. Our study further paves the way for the development of new diagnostic regimens in the physician's office for IBD screening as a preventive step. This simple, risk-free, and cost-effective technique with minimal

sample preparation and lack of discomfort will be attractive to most of the patient population in the routine screening for UC and other IBDs. This technique can be carried out by primary care physicians, as a simple & cost-effective preliminary indicator for UC so that the GI can perform the colonoscopies for the final verification.

This novel proposed technique not only has the potential as an alternative for early diagnosis of IBDs but also for identifying the progression status of IBDs, specifically UC and CD. The future work includes the exploration of spectral signatures for CD differentiation between UC and CD. It is because, patients undergoing an initial evaluation for IBD will often undergo a series of diagnostic tests, and a large amount of clinical information obtained may make a physician uncertain as to whether to label the patient as having CD or UC[37]. In many cases, similar drugs are used to treat UC and CD, but treatment is not equally efficient and responds differently. It is known that UC mostly presents with rectal inflammation and continuous lesions, while CD presents with discontinuous inflammatory lesions and frequently involves the ileocecal area. Studies further show, the location of inflammation (in UC colon is the only site that is affected and in the CD it affects anywhere along the digestive tract), the efficacy of drug treatment, aggravation to the other body parts[173] are also different. One can undergo a single surgery to treat UC, whereas CD may need multiple surgeries. It was reported that concentrations of some chemicals were different in UC and CD patients[174, 175], thus differentiating IR spectral markers are expected.

4.4 Potential Prototype for the Portable Device Development

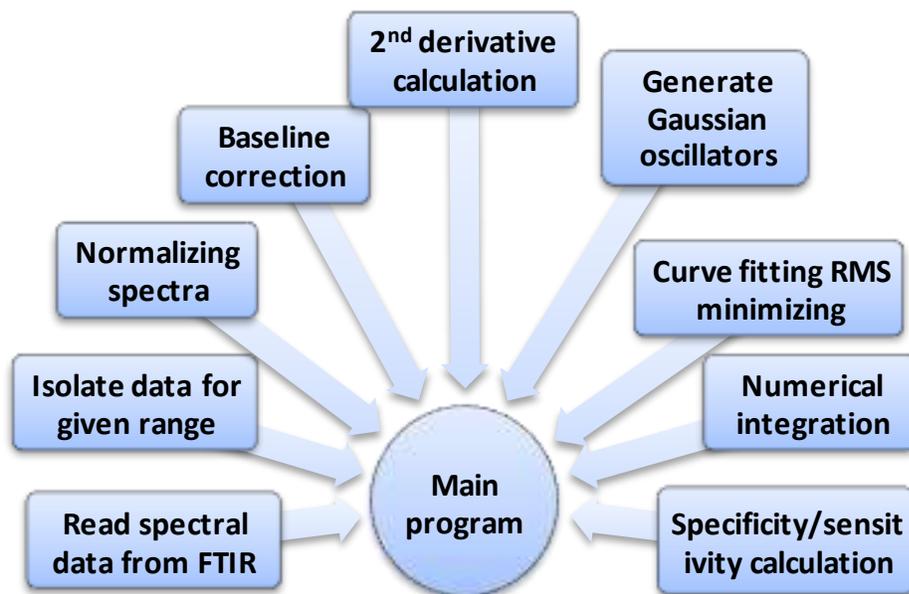


Figure 4.1 Structure of the software program showing 9 subroutines to perform each specific task[11]. Each of these specific programming tasks was individually carried out in our preliminary work in order to get the conclusion. In the portable device development, all those will be automated into one program which will be done at a single touch command.

Our study is also directed toward designing and developing a prototype of a portable tool for facilitating such measurements. As the first stage of the portable device development, we have studied the possible use of a commercially available portable FTIR (Bruker ALPHA & Perkin Elmer 2) with ATR setup. The communication with the commercial software from the FTIR manufacturer will be critical to develop a user-friendly prototype, so accompanying software can be developed to completely automate the process.

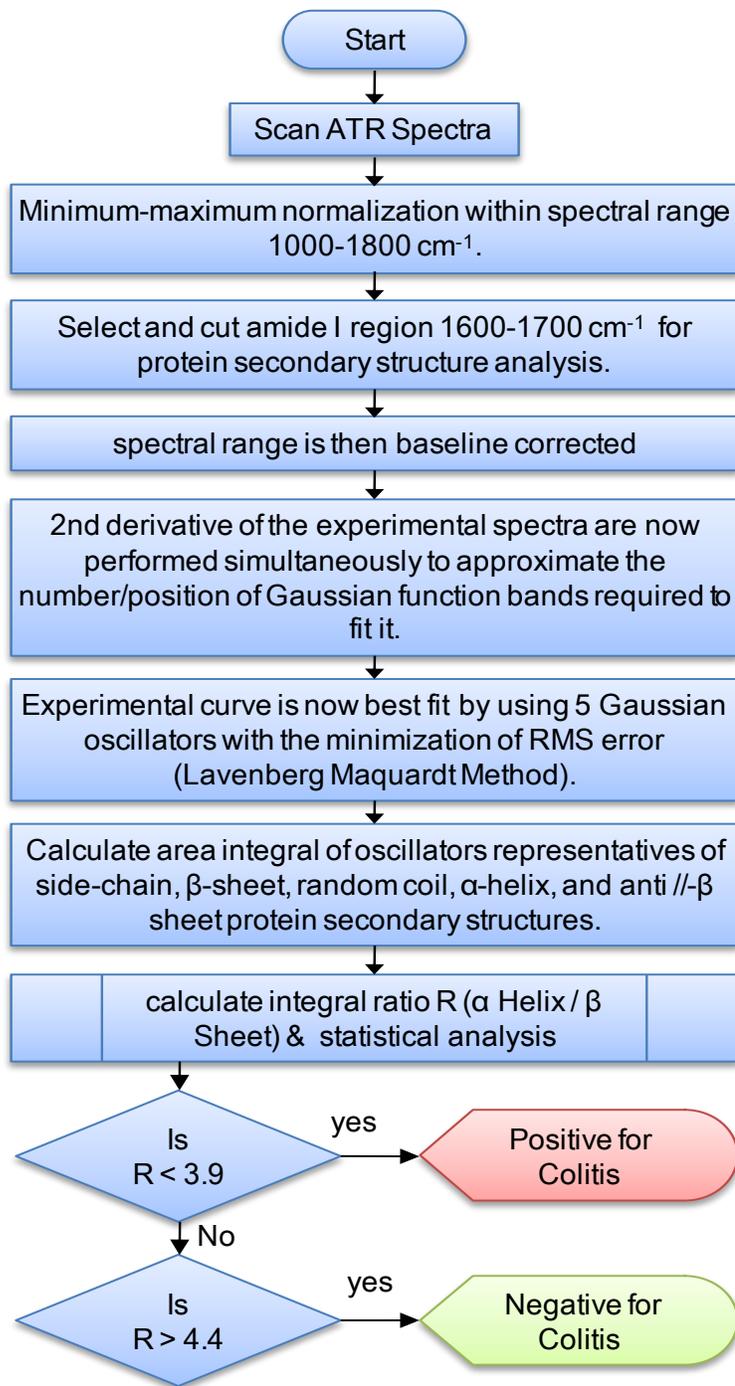


Figure 4.2 Flow chart for the programming to study protein secondary structures[11]. Only the α -helix and β -sheet ratio analysis is shown in the flow chart. Similarly, other biomarker analysis will also be in the program as needed to achieve the needed confidence limit.

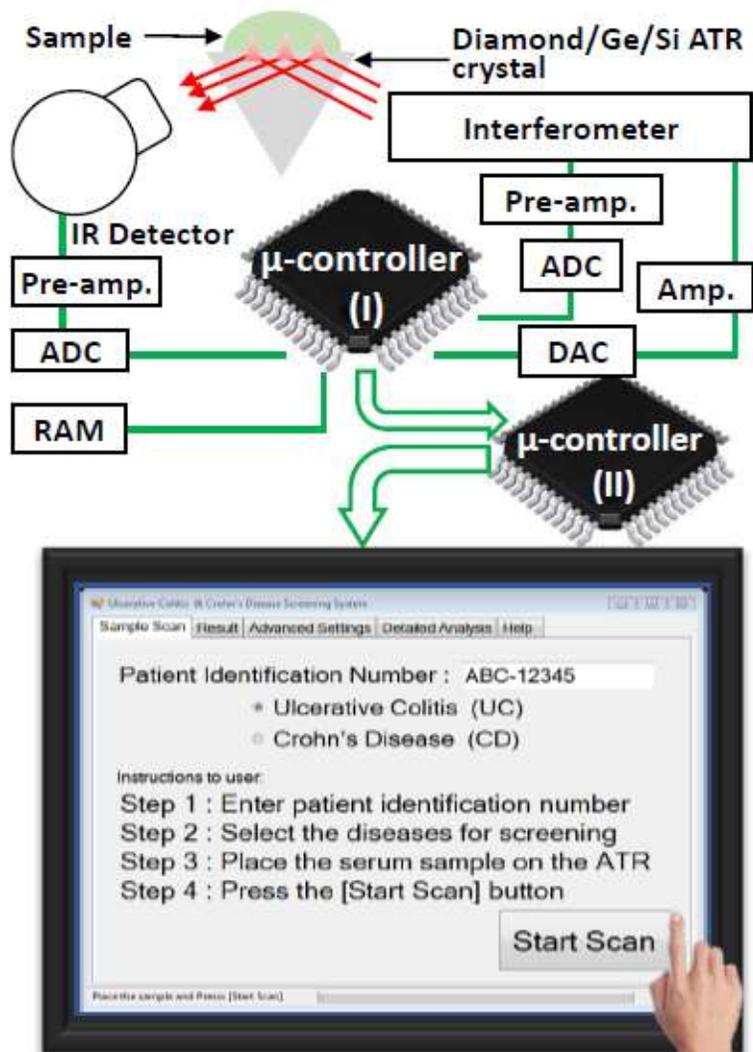


Figure 4.3 Schematic of measurements and data analysis tool permanently integrated into the anticipated modified spectroscope. "Sample Scan" tab allows users to enter the patient ID, select the disease followed by pressing "Start Scan" to initiate the program[11].

Fully automating the spectral measurements and analysis presented above is the second step to check the feasibility of the proposed portable device. Minimization of the instrumentation by integrating it with a simple touchscreen (e.g. Windows tablet) will make it a user-friendly desktop unit. Spectral measurements and data analysis procedures will be automated into a single

step so that a technician can deposit the sample onto the sample holder and start the measurement with a simple click to get the result and if needed the biochemical information.

The software program will include 9 subroutines as shown in Figure 4.1 and each will do a specific task in the data extraction or analyzing stage. Reading a spectral data file from the FTIR, extract data for a suitable biomarker range depending on the user-selected disease, normalizing and baseline correction of spectral data subroutines will have simple loops, condition check, and basics mathematical calculations. The second derivative will be calculated by using divided difference formulas for discrete data. After finding the number of minimums and their positions, the program will assign parameters for Gaussian oscillators. The curve fitting subroutine will iteratively vary the Gaussian curve parameters to minimize the RMS error between the experimental absorption curve and the corresponding summation of the Gaussian curves. The Levenberg Marquardt algorithm will be used in this subroutine. The standard numerical integration technique will be used to find the area under each Gaussian oscillator.

As an example, the flow chart for protein secondary structure analysis is shown in figure 4.2. Other biomarkers already identified and new ones to be identified in the process can also be included in the program. For debugging and troubleshooting reasons, it is important to be able to see the output at each stage in addition to the outcome. Also, this allows anyone interested to see the important data at each stage of the process and to try other changes.



Figure 4.4 Anticipated portable device with the touch screen showing "Test Report". The test report can be printed as a hard copy or save the patient record in the hospital database[11].

A schematic of this anticipated portable prototype for the colitis screening by using ATR-FTIR spectroscopy is shown in figure 4.3, where the microprocessor (I) controls all the functions in the instrument, and the software (spectral measurement and data analysis) will run on a microprocessor (II). The software package developed can be installed on the Windows tablet PC attached to the portable FTIR. Herein, the "Sample Scan" tab allows users to enter the patient ID, select the disease, followed by the user pressing "Start Scan" to initiate the program. Pressing "Start Scan" after depositing a sample on the holder can automatically perform the full data analysis and will display a laboratory test report as seen in figure 4.4.

4.5 Discussion Based on Lymphoma and Melanoma Study

The results of the present study shown remarkable differences between the ATR-FTIR spectra of serum samples representing tumor-bearing mouse models of melanoma (n=8) and NHL (n=8) from their control (n=15) types. The differentiating signatures between spectra are obtained by observing (i) p-values comparison, (ii) the spectral position, and ratio analysis of amide peaks (iii) the fit of the experimental spectra and (iv) the employment of multivariate analysis (HCA). This difference between control and tumorous cases is evident through the gradual changes in the intensities of the absorption of mainly proteins, carbohydrates and nucleic acids in the serum. It is noted that serological tests show the alterations of certain proteins, peptides, and nucleic acids (DNA, mRNA) for patients with melanoma[176] and lymphoma[177]. Manifestations of these alterations in biomolecules (serological markers) are most likely due to the tumor-induced alterations directing towards identifying the spectral signatures.

Herein, this is an experimental demonstration of the rapid and reliable spectroscopic techniques for the discrimination of B16 melanoma and EL4 lymphoma mice from their control types. B16 murine tumor model remains indispensable for metastasis and therapeutic studies of human melanoma skin cancer[178]. Similarly, development of EL4 murine tumor model considered a huge benefit to the human NHL research cancer. This work is thus expected to lay a foundation for further research which could lead to the development of diagnostic techniques for future health care of cancer patients of melanoma and lymphoma using body fluid samples that can be collected with relatively low risks. It is thus critical to extend the present study to human patients for the assessment of disease status and personalized drug management. Furthermore, the study of temporal variation in spectral marker signatures is important for tumor grading, subtyping, and assessing heterogeneity. Further work is in progress (i) to investigate temporal

variation in serum components along with the progression of the disease by increasing sample size, (ii) identify the alteration in spectral signatures using human patients, and (ii) to integrate data analyzing software into the narrow multiband detector. After setting a calibration curve of unique spectral signatures for NHL or subcutaneous melanoma, bulky instrumentation will be avoided using specific multiband IR detectors capable of simultaneous detection in the expected narrow bands. Recent advances in IR technology allow the operation of multiband detectors at room temperature[179]. Complex statistical analysis of identifying spectral signatures of NHL or melanoma can also be integrated into the clinical tool as a software application into the computer program. In terms of clinical application, we can anticipate that the potential technology can be further developed into a personalized diagnostic tool in which patient-to-patient and within a patient over time (due to health conditions or other factors) differences in molecular signatures would allow the assessment of disease status and personalized drug management. To be used as a patient to the patient screening test, a normal range of spectral signatures unique to the particular disease should be set by using a statistically significant set of normal serum samples. These average normal values can be incorporated into the program which can identify the deviations of the test sample from the average values. Technological advancement of ATR-FTIR spectroscopy of serum sample to discriminate normal and tumorous conditions will thus support to increase compliance rate eligible population for tumor screening and to make physician decision for advanced histological examination using biopsy.

4.6 Discussion Based on BC Study

Protein analysis is considered as a promising technique for understanding the progression of cancers. Similarly, FTIR spectral analysis is one of the accepted paradigms for the holistic evaluation of protein structural content at the molecular level in biological samples. Several studies

have introduced the applicability of FTIR spectroscopy in serum samples accompanied by spectral analysis techniques for BC discrimination. Reports[89, 97] show the potential application of FTIR spectroscopy for protein analysis in serum samples from the BC patients. However, cancer initiation, progression, and response to therapy depend on an array of complex interactions between constituent biomolecules (proteins, lipids, nucleic acids, and carbohydrates) and not only at the level of the single (biomarker or target) molecule. Therefore, the feasibility of FTIR spectroscopy to extract a snapshot of cumulative molecular interactions within serum samples warrant a thorough investigation, as enabled by interdisciplinary collaboration between spectroscopists, biologists, and clinicians. It is noted that the evaluation of serological biomarkers (CA15-3, HSP90A, and PAI-1) do not show consistent differences between BC cases and controls that can lead up to diagnosis [84]. Our data show alterations in biochemical, and structural, information of the constituent components of the sample medium. Such, holistic evaluation of biochemical details with the use of IR spectroscopy can thus have an immense potential for BC discrimination analysis in the clinical domain.

Deconvolution of spectral range $1140\text{-}1000\text{ cm}^{-1}$: To analyze the snapshot of alterations reflected in our FTIR spectral data, the complex region[140] of carbohydrates and nucleic acids, $1140\text{-}1000\text{ cm}^{-1}$, was deconvoluted. The BC-associated alterations in the DNA and RNA are reflected in this region. The circulating DNA and protein markers are generally evaluated to track the biomolecular events of cancerous patients[180]. Herein, this spectral range is deconvoluted with six GFEB (Figure 4.5 (A)) by approximating numbers and positions using the minima of second derivatives. The sum of integral areas covered by six bands (integral values) of control, samples range from 11.4 to 13.2, while these values in BC samples are from 13 to 14.8. This quantified information was further statistically analyzed (Figure 4.5(B)), and the result shows a

clear separation between control and BC. Similarly, figure 4.5 (C) shows the histogram of the average values of absorbance at wavenumber 1020 cm^{-1} . Absorbance at this energy band is found to be due to the presence of circulating DNA.

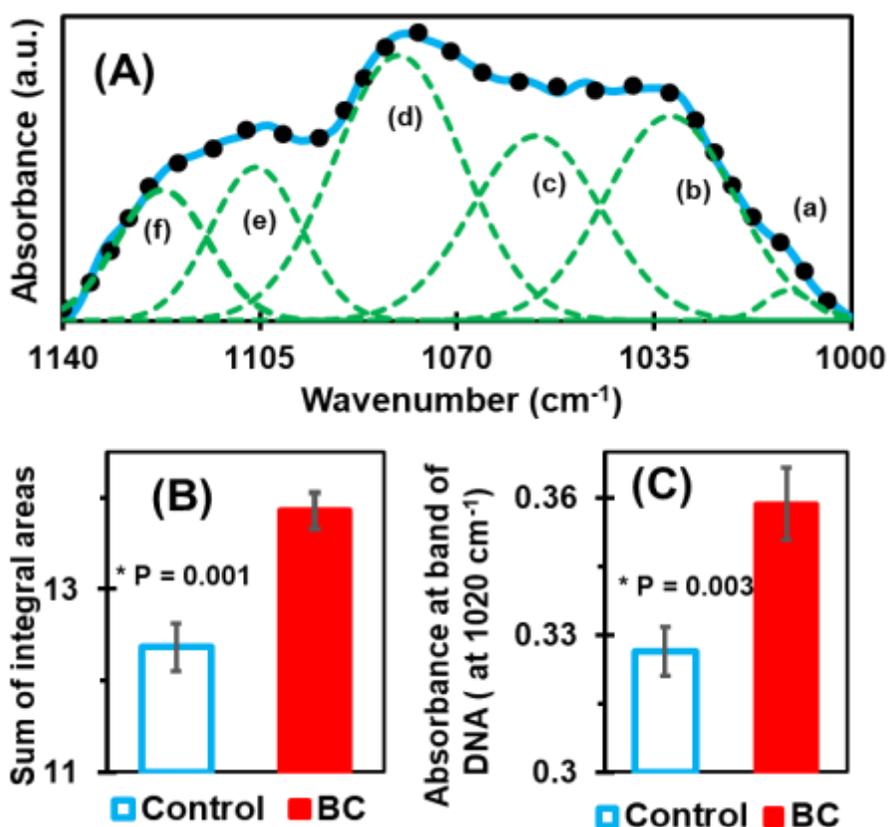


Figure 4.5 (A) Deconvolution of the complex band of carbohydrates and nucleic acids $1000\text{-}1140\text{ cm}^{-1}$. The number and position of six bands used to fit the experimental curve were determined by using the minima of secondary curves as in amide I case. (B) Bar graph representation of the average value of the integral sum which shows a significant difference between control and BC case. (C) Bar graph of average absorbances at wavenumber position 1020 cm^{-1} , which is mainly due to the presence of DNA. It also shows a significant difference between control and BC cases[13].

Potential prototype for clinical application: Moreover, a prototype for our presented diagnostic regimen for clinical use can be developed. Spectral measurements and data analysis procedures will be automated into a single step so that a technician can deposit the sample onto the sample holder and start to measure with a simple click to get the result and, if needed, the biochemical information easily as shown in Figure 4.6. ATR-FTIR spectroscopy (that is reliable for body-fluids analysis) integrated with two micro-controllers, where micro-controller A controls all the functions in the FTIR, while the controller B controls software for data analysis. The software program will include several subroutines as reading spectral data from the FTIR; extract data for suitable spectral signatures in the measured range; normalizing and baseline correction of spectral data subroutines will have simple loops, condition checks, and basics mathematical calculations.

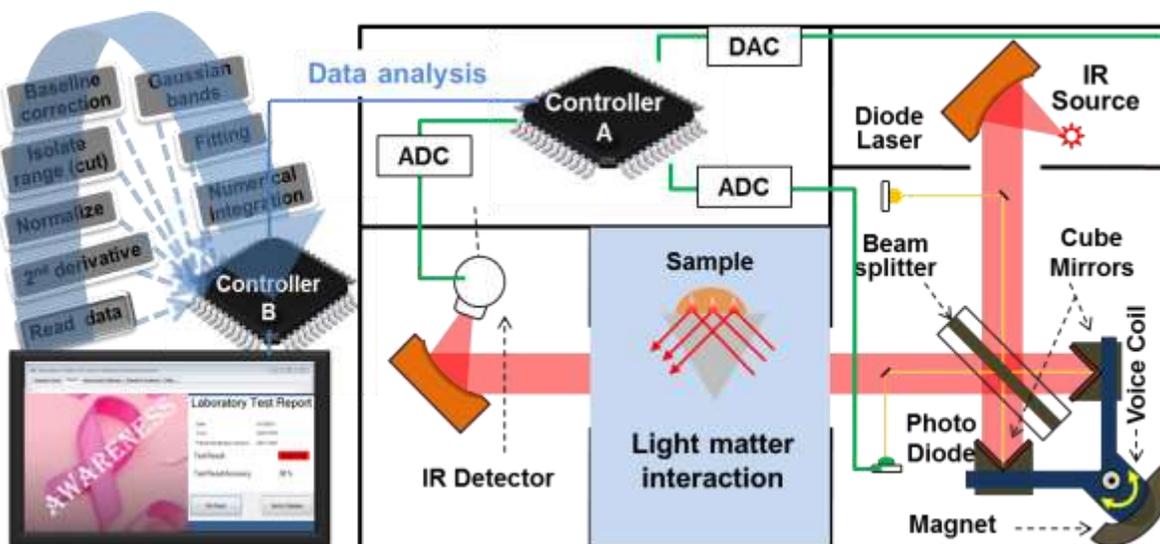


Figure 4.6 Schematic of ATR-FTIR spectrometer integrated with two micro-controllers (micro-processors) A and B. The controller A extracts the information about the signal-sample interaction, while controller B stores the spectral analyzing software application in the clinical domain[13].

The second derivative will be calculated by using divided difference formulas for discrete data. After finding various minimums and their positions, the program will assign parameters for Gaussian energy bands and select settings for bands to minimize RMS error (Levenberg Marquardt algorithm) between experimental data and fitted curves. The standard numerical integration technique will be used to find the area under Gaussian bands and the ratio. Additionally, combining all identified multiple spectral signatures into a single diagnostic index using them as the discriminating signature marker, a portable device integrated with the user-friendly desktop unit (can automatically perform the full data analysis and will display laboratory test report) can be prepared.

All in all, FTIR spectroscopy of serum samples could be a promising technique for an ASSURED regimen for evaluation of BC associated molecular level of alteration in constituent protein structures. Our study holds value as available techniques such as mammograms, MRI, and ultrasonography have their limits and may not be 100% accurate[81, 181-183]. Among them, MRI achieves a high sensitivity of 70%–100% in the initial screening (prevalence), compared at 40% or less for mammography in patients with high risk to develop BC[182, 183], but the specificity of MRI is hampered by its difficulty while distinguishing the overlapping features of benign and malignant lesions, leading to higher false-positive rates[181]. Ultrasonography also fails to detect micro-calcifications and has poor specificity. Therefore, the present diagnostic regimen of BC having the potential to promote timely onward referral of patients for further testing and detection of recurrent disease, “enabling serial sample and testing with less cost, resource and radiation exposure” could be beneficial for several patients.

5 CONCLUSION

The findings presented in this dissertation confirm the potential application of FTIR spectroscopy of serum samples, accompanied by data analysis techniques, promise for use as a clinical tool to differentiate IBD and cancer patients from healthy individuals. Importantly, it is evident through the use of experimental mouse models and human patients. However, the utility of this technique as a clinical tool requires further investigation and validations. Therefore, prospective studies evaluating specificity, sensitivity and accuracy of the technique on a large and carefully case-defined population is warranted. Accurate and objective classification, staging and grading for cancer management as opposed to the present gold standard, histopathological diagnosis, to guide treatments and predict patient prognosis is crucial to accelerate point-of-care decisions and potentially revolutionize cancer diagnostics in personalized medicine by using FTR spectroscopy.

In the IBD study, the experimental models: IL10^{-/-} and DSS shows diagnostic accuracy with 80-100% sensitivity and specificity values. In human serum samples, our study shows sensitivity and specificity values are 100% and 86%, respectively. Similarly, in cancer studies, mouse models of NHL and subcutaneous melanoma, emphasizes the diagnostic potential of this approach as a screening technique. Similarly, the BC-associated protein conformational alteration in the serum samples shows the sensitivity and the specificity of identifying spectral signatures were both 90%. Various spectral signatures were analyzed as shown in Table 5.1. These include (a) absorption values at wavenumber 1033 cm⁻¹, (b) absorption values at wavenumber 1076 cm⁻¹; (c) the ratio of absorbance at wavenumber 1121 cm⁻¹, to its value at 1020 cm⁻¹, (d) the ratio of absorbance at wavenumber 1629 cm⁻¹, to 1737 cm⁻¹. (e) the ratio of GFEB representing α helix and β sheet protein secondary structures as obtained from the deconvolution of amide I region, (f)

the sum of the integral area of GFEBs used to fit the experimental curve within the complex band of carbohydrates, and nucleic acids 1000-1140 cm^{-1} (g) absorbance at wavenumber 1212 cm^{-1} , and the absorbance at wavenumber 1335 cm^{-1} .

Table 5.1 Summary of identifying spectral signatures.

Studied types		Discrimination from the controls									
		I ₁₀₃₃	I ₁₀₇₆	I ₁₁₂₁ / I ₁₀₂₁	I ₁₆₂₉ / I ₁₇₃₇	α/β	Integral sum	Amide I/II	Amide II/III	I ₁₂₁₂	I ₁₃₃₅
DSS Colitis		Yes	Yes	Yes	Yes	Yes
IL10-/-	Colitis	Yes	Yes	Yes	Yes	Yes	Yes
	Treated	No	No	No	No	No	No
	Control	No	No	No	No	No	No
Human Control		No	No	No	No	No
Human UC		Yes	Yes	Yes	Yes	Yes
Lymphoma		Yes	Yes	Yes	Yes	Yes
Melanoma		Yes	Yes	No	No	No
Control mice		No	No	No	No	No
BC		Yes	Yes	Yes	Yes	Yes
Human Control		No	No	No	No	No

REFERENCES

- [1] M. S. H. Akash and K. Rehman, "Introduction to Spectrophotometric Techniques," in *Essentials of Pharmaceutical Analysis*: Springer, 2020, pp. 19-27.
- [2] S. Mukamel, *Principles of nonlinear optical spectroscopy*. Oxford university press New York, 1995.
- [3] B. C. Smith, *Fundamentals of Fourier transform infrared spectroscopy*. CRC press, 2011.
- [4] J. Titus, H. Ghimire, E. Viennois, D. Merlin, and A. J. J. o. b. Unil Perera, "Protein secondary structure analysis of dried blood serum using infrared spectroscopy to identify markers for colitis screening," vol. 11, no. 3, p. e201700057, 2018.
- [5] J. Titus, E. Viennois, D. Merlin, and A. J. J. o. b. Unil Perera, "Minimally invasive screening for colitis using attenuated total internal reflectance fourier transform infrared spectroscopy," vol. 10, no. 3, pp. 465-472, 2017.
- [6] H. Ghimire, X. Hu, G. Qin, and A. U. J. B. O. E. Perera, "Optimizing infrared spectral discrimination to enhance disease diagnostics: monitoring the signatures of inflammatory bowel diseases with anti-TNF α therapy," vol. 11, no. 8, pp. 4679-4694, 2020.
- [7] A. U. Perera and H. Ghimire, "Quantification of infrared spectral markers for ulcerative colitis using sera," in *Biomedical Vibrational Spectroscopy 2020: Advances in Research and Industry*, 2020, vol. 11236, p. 112360T: International Society for Optics and Photonics.
- [8] A. U. Perera, H. Ghimire, K. Perera, E. Viennois, and D. Merlin, "Colitis screening using IR spectroscopy of serum samples," in *2017 Eleventh International Conference on Sensing Technology (ICST)*, 2017, pp. 1-4: IEEE.
- [9] H. Ghimire, K. Perera, and A. J. B. o. t. A. P. S. Perera, "Mid-IR Spectroscopy of Dried Serum Samples: An Application for Colitis Prescreening," vol. 62, 2017.
- [10] J. Titus, H. Ghimire, E. Viennois, D. Merlin, and A. U. Perera, "Infrared spectroscopy as a screening technique for colitis," in *Bio-MEMS and Medical Microdevices III*, 2017, vol. 10247, p. 102470C: International Society for Optics and Photonics.
- [11] H. Ghimire, P. Jayaweera, A. U. J. I. P. Perera, and Technology, "Longitudinal analysis of molecular alteration in serum samples of dextran sodium sulfate-induced colitis mice by using infrared spectroscopy," vol. 97, pp. 33-37, 2019.
- [12] H. Ghimire, M. Venkataramani, Z. Bian, Y. Liu, and A. U. Perera, "ATR-FTIR spectral discrimination between normal and tumorous mouse models of lymphoma and melanoma from serum samples," *Scientific reports*, vol. 7, no. 1, pp. 1-9, 2017.
- [13] H. Ghimire *et al.*, "Protein Conformational Changes in Breast Cancer Sera Using Infrared Spectroscopic Analysis," *Cancers*, vol. 12, no. 7, p. 1708, 2020.
- [14] M. H. Niemz, *Laser-tissue interactions*. Springer, 2007.
- [15] B. Stuart, "Infrared spectroscopy," *Kirk-Othmer encyclopedia of chemical technology*, 2000.
- [16] H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise, *Near-infrared spectroscopy: principles, instruments, applications*. John Wiley & Sons, 2008.
- [17] P. R. Griffiths and J. A. De Haseth, *Fourier transform infrared spectrometry*. John Wiley & Sons, 2007.
- [18] M. Blanco and I. Villarroya, "NIR spectroscopy: a rapid-response analytical tool," *TrAC Trends in Analytical Chemistry*, vol. 21, no. 4, pp. 240-250, 2002.

- [19] C. Kendall *et al.*, "Vibrational spectroscopy: a clinical tool for cancer diagnostics," *Analyst*, vol. 134, no. 6, pp. 1029-1045, 2009.
- [20] J. Dubois and R. A. Shaw, "Peer Reviewed: IR Spectroscopy in Clinical and Diagnostic Applications," ed: ACS Publications, 2004.
- [21] C.-M. J. F. s. i. Orphanou, "The detection and discrimination of human body fluids using ATR FT-IR spectroscopy," vol. 252, pp. e10-e16, 2015.
- [22] F. Parker, *Applications of infrared spectroscopy in biochemistry, biology, and medicine*. Springer Science & Business Media, 2012.
- [23] Z. Movasaghi, S. Rehman, and D. I. ur Rehman, "Fourier transform infrared (FTIR) spectroscopy of biological tissues," *Applied Spectroscopy Reviews*, vol. 43, no. 2, pp. 134-179, 2008.
- [24] M. J. Baker *et al.*, "Using Fourier transform IR spectroscopy to analyze biological materials," *Nature protocols*, vol. 9, no. 8, p. 1771, 2014.
- [25] H. Yang, S. Yang, J. Kong, A. Dong, and S. Yu, "Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy," *Nature protocols*, vol. 10, no. 3, pp. 382-396, 2015.
- [26] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott, and F. L. Martin, "Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives," *Analyst*, vol. 137, no. 14, pp. 3202-3215, 2012.
- [27] J. G. Kelly *et al.*, "Robust classification of low-grade cervical cytology following analysis with ATR-FTIR spectroscopy and subsequent application of self-learning classifier eClass," *Analytical and bioanalytical chemistry*, vol. 398, no. 5, pp. 2191-2201, 2010.
- [28] S. G. Kazarian and K. A. Chan, "ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems," *Analyst*, vol. 138, no. 7, pp. 1940-1951, 2013.
- [29] J. Titus, C. Filfili, J. K. Hilliard, J. A. Ward, and A. Unil Perera, "Early detection of cell activation events by means of attenuated total reflection Fourier transform infrared spectroscopy," *Applied Physics Letters*, vol. 104, no. 24, p. 243705, 2014.
- [30] A. W. Snyder and J. D. Love, "Goos-Hänchen shift," *Applied optics*, vol. 15, no. 1, pp. 236-238, 1976.
- [31] G. Clemens, J. R. Hands, K. M. Dorling, and M. J. Baker, "Vibrational spectroscopic methods for cytology and cellular research," *Analyst*, vol. 139, no. 18, pp. 4411-4444, 2014.
- [32] K. A. Chan and S. G. Kazarian, "Attenuated total reflection Fourier-transform infrared (ATR-FTIR) imaging of tissues and live cells," *Chemical Society Reviews*, vol. 45, no. 7, pp. 1850-1864, 2016.
- [33] M. J. Baker *et al.*, "Developing and understanding biofluid vibrational spectroscopy: a critical review," *Chemical Society Reviews*, vol. 45, no. 7, pp. 1803-1818, 2016.
- [34] D. N. Frank, A. L. S. Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace, "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases," *Proceedings of the National Academy of Sciences*, vol. 104, no. 34, pp. 13780-13785, 2007.
- [35] S. H. Itzkowitz and X. Yio, "Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation," *American journal of physiology-gastrointestinal and liver physiology*, vol. 287, no. 1, pp. G7-G17, 2004.

- [36] C. G. Loftus *et al.*, "Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940–2000," *Inflammatory bowel diseases*, vol. 13, no. 3, pp. 254-261, 2007.
- [37] A. Bousvaros *et al.*, "Differentiating ulcerative colitis from Crohn disease in children and young adults: report of a working group of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the Crohn's and Colitis Foundation of America," *Journal of pediatric gastroenterology and nutrition*, vol. 44, no. 5, pp. 653-674, 2007.
- [38] M. Z. Cader and A. Kaser, "Recent advances in inflammatory bowel disease: mucosal immune cells in intestinal inflammation," *Gut*, vol. 62, no. 11, pp. 1653-1664, 2013.
- [39] R. Yantiss and R. Odze, "Diagnostic difficulties in inflammatory bowel disease pathology," *Histopathology*, vol. 48, no. 2, pp. 116-132, 2006.
- [40] S. Vermeire, G. Van Assche, and P. Rutgeerts, "Laboratory markers in IBD: useful, magic, or unnecessary toys?," *Gut*, vol. 55, no. 3, pp. 426-431, 2006.
- [41] J. A. Wilson, "Colon cancer screening in the elderly: when do we stop?," *Transactions of the American Clinical and Climatological Association*, vol. 121, p. 94, 2010.
- [42] R. Xavier and D. Podolsky, "Unravelling the pathogenesis of inflammatory bowel disease," *Nature*, vol. 448, no. 7152, p. 427, 2007.
- [43] A. Tesija Kuna, "Serological markers of inflammatory bowel disease," *Biochemia medica: Biochemia medica*, vol. 23, no. 1, pp. 28-42, 2013.
- [44] R. Burakoff *et al.*, "Blood-based biomarkers used to predict disease activity in Crohn's disease and ulcerative colitis," *Inflammatory bowel diseases*, vol. 21, no. 5, pp. 1132-1140, 2015.
- [45] B. Mesko *et al.*, "Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets," *BMC medical genomics*, vol. 3, no. 1, p. 15, 2010.
- [46] J. Ollesch, M. Heinze, H. M. Heise, T. Behrens, T. Brüning, and K. Gerwert, "It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy," *Journal of biophotonics*, vol. 7, no. 3-4, pp. 210-221, 2014.
- [47] S. Kazarian and K. Chan, "Applications of ATR-FTIR spectroscopic imaging to biomedical samples," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1758, no. 7, pp. 858-867, 2006.
- [48] C.-M. Orphanou, "The detection and discrimination of human body fluids using ATR FT-IR spectroscopy," *Forensic science international*, vol. 252, pp. e10-e16, 2015.
- [49] E. A. Martin, *Concise medical dictionary*. Oxford Quick Reference, 2015.
- [50] L. L. Banez *et al.*, "Diagnostic potential of serum proteomic patterns in prostate cancer," *The Journal of urology*, vol. 170, no. 2, pp. 442-446, 2003.
- [51] A. Lembo, B. Neri, J. Tolley, D. Barken, S. Carroll, and H. Pan, "Use of serum biomarkers in a diagnostic test for irritable bowel syndrome," *Alimentary pharmacology & therapeutics*, vol. 29, no. 8, pp. 834-842, 2009.
- [52] A. Tagore *et al.*, "Interleukin-10 (IL-10) genotypes in inflammatory bowel disease," *Tissue antigens*, vol. 54, no. 4, pp. 386-390, 1999.
- [53] W. E. Ip, N. Hoshi, D. S. Shouval, S. Snapper, and R. Medzhitov, "Anti-inflammatory effect of IL-10 mediated by metabolic reprogramming of macrophages," *Science*, vol. 356, no. 6337, pp. 513-519, 2017.

- [54] E. Viennois, M. T. Baker, B. Xiao, L. Wang, H. Laroui, and D. J. J. o. p. Merlin, "Longitudinal study of circulating protein biomarkers in inflammatory bowel disease," vol. 112, pp. 166-179, 2015.
- [55] H. M. Van Dullemen *et al.*, "Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2)," vol. 109, no. 1, pp. 129-135, 1995.
- [56] A. Akazawa, I. Sakaida, S. Higaki, Y. Kubo, K. Uchida, and K. J. J. o. g. Okita, "Increased expression of tumor necrosis factor- α messenger RNA in the intestinal mucosa of inflammatory bowel disease, particularly in patients with disease in the inactive phase," vol. 37, no. 5, pp. 345-353, 2002.
- [57] K. Aden *et al.*, "Metabolic functions of gut microbes associate with efficacy of tumor necrosis factor antagonists in patients with inflammatory bowel diseases," vol. 157, no. 5, pp. 1279-1292. e11, 2019.
- [58] L. Peyrin-Biroulet, "Anti-TNF therapy in inflammatory bowel diseases: a huge review," *Minerva gastroenterologica e dietologica*, vol. 56, no. 2, pp. 233-243, 2010.
- [59] R. Kühn, J. Löhler, D. Rennick, K. Rajewsky, and W. Müller, "Interleukin-10-deficient mice develop chronic enterocolitis," *Cell*, vol. 75, no. 2, pp. 263-274, 1993.
- [60] P. J. Koelink *et al.*, "Anti-TNF therapy in IBD exerts its therapeutic effect through macrophage IL-10 signalling," *Gut*, pp. gutjnl-2019-318264, 2019.
- [61] T. Scheinin, D. M. Butler, F. Salway, B. Scallan, and M. Feldmann, "Validation of the interleukin-10 knockout mouse model of colitis: antitumour necrosis factor-antibodies suppress the progression of colitis," *Clinical & Experimental Immunology*, vol. 133, no. 1, pp. 38-43, 2003.
- [62] B. Wood, L. Chiriboga, H. Yee, M. Quinn, D. McNaughton, and M. Diem, "Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium," *Gynecologic oncology*, vol. 93, no. 1, pp. 59-68, 2004.
- [63] P. D. Lewis *et al.*, "Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum," *BMC cancer*, vol. 10, no. 1, p. 640, 2010.
- [64] J. Backhaus *et al.*, "Diagnosis of breast cancer with infrared spectroscopy from serum samples," *Vibrational Spectroscopy*, vol. 52, no. 2, pp. 173-177, 2010.
- [65] C. A. Lima, V. P. Goulart, L. Côrrea, T. M. Pereira, and D. M. Zzell, "ATR-FTIR spectroscopy for the assessment of biochemical changes in skin due to cutaneous squamous cell carcinoma," *International journal of molecular sciences*, vol. 16, no. 4, pp. 6621-6630, 2015.
- [66] N. Fujioka, Y. Morimoto, T. Arai, and M. Kikuchi, "Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy," *Cancer Detection and Prevention*, vol. 28, no. 1, pp. 32-36, 2004.
- [67] E. Gazi *et al.*, "Applications of Fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. A pilot study," *The Journal of pathology*, vol. 201, no. 1, pp. 99-108, 2003.
- [68] B. Rigas, S. Morgello, I. S. Goldman, and P. Wong, "Human colorectal cancers display abnormal Fourier-transform infrared spectra," *Proceedings of the National Academy of Sciences*, vol. 87, no. 20, pp. 8140-8144, 1990.
- [69] G. Theophilou, K. M. Lima, P. L. Martin-Hirsch, H. F. Stringfellow, and F. L. Martin, "ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer," *Analyst*, vol. 141, no. 2, pp. 585-594, 2016.

- [70] A. J. Sommer, L. G. Tisinger, C. Marcott, and G. M. Story, "Attenuated total internal reflection infrared mapping microspectroscopy using an imaging microscope," *Applied spectroscopy*, vol. 55, no. 3, pp. 252-256, 2001.
- [71] M. Lens and M. Dawes, "Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 150, no. 2, pp. 179-185, 2004.
- [72] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a cancer journal for clinicians*, vol. 66, no. 1, pp. 7-30, 2016.
- [73] C. Garbe and U. Leiter, "Melanoma epidemiology and trends," *Clinics in dermatology*, vol. 27, no. 1, pp. 3-9, 2009.
- [74] S. G. Fisher and R. I. Fisher, "The epidemiology of non-Hodgkin's lymphoma," *Oncogene*, vol. 23, no. 38, pp. 6524-6534, 2004.
- [75] N. Howlader *et al.*, "SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. Bethesda, MD," ed, 2016.
- [76] A. F. Jerant, J. T. Johnson, C. Sheridan, and T. J. Caffrey, "Early detection and treatment of skin cancer," *American family physician*, vol. 62, no. 2, pp. 357-386, 2000.
- [77] M. Shipp *et al.*, "A predictive model for aggressive non-Hodgkin's lymphoma," *New England Journal of Medicine*, vol. 329, no. 14, pp. 987-994, 1993.
- [78] B. Stewart and C. P. Wild, "World cancer report 2014," *Public Health*, 2019.
- [79] S. McGuire, "World cancer report 2014. Geneva, Switzerland: World Health Organization, international agency for research on cancer, WHO Press, 2015," ed: Oxford University Press, 2016.
- [80] A. C. Society, "Breast cancer facts & figures 2017–2018," ed: American Cancer Society Atlanta, GA, 2017.
- [81] B. C. Yankaskas *et al.*, "Performance of first mammography examination in women younger than 40 years," vol. 102, no. 10, pp. 692-701, 2010.
- [82] A. B. Miller, C. Wall, C. J. Baines, P. Sun, T. To, and S. A. J. B. Narod, "Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial," vol. 348, p. g366, 2014.
- [83] D. E. Misek and E. H. Kim, "Protein biomarkers for the early detection of breast cancer," *International journal of proteomics*, vol. 2011, 2011.
- [84] A. Kazarian *et al.*, "Testing breast cancer serum biomarkers for early detection and prognosis in pre-diagnosis samples," vol. 116, no. 4, pp. 501-508, 2017.
- [85] G. C. Blobel, L. M. Obeid, and Y. A. Hannun, "Regulation of protein kinase C and role in cancer biology," *Cancer and Metastasis Reviews*, vol. 13, no. 3-4, pp. 411-431, 1994.
- [86] J. Bartkova, J. Lukas, H. Müller, D. Lützhøt, M. Strauss, and J. Bartek, "Cyclin D1 protein expression and function in human breast cancer," *International journal of cancer*, vol. 57, no. 3, pp. 353-361, 1994.
- [87] J. Jacquemier *et al.*, "Protein expression profiling identifies subclasses of breast cancer and predicts prognosis," *Cancer Research*, vol. 65, no. 3, pp. 767-779, 2005.
- [88] J. Li, Z. Zhang, J. Rosenzweig, Y. Y. Wang, and D. W. J. C. c. Chan, "Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer," vol. 48, no. 8, pp. 1296-1304, 2002.
- [89] F. Elmi, A. F. Movaghar, M. M. Elmi, H. Alinezhad, N. J. S. A. P. A. M. Nikbakhsh, and B. Spectroscopy, "Application of FT-IR spectroscopy on breast cancer serum analysis," vol. 187, pp. 87-91, 2017.

- [90] U. Zelig *et al.*, "Early detection of breast cancer using total biochemical analysis of peripheral blood components: a preliminary study," vol. 15, no. 1, p. 408, 2015.
- [91] E. Ostrovsky *et al.*, "Detection of cancer using advanced computerized analysis of infrared spectra of peripheral blood," vol. 60, no. 2, pp. 343-353, 2012.
- [92] T. Gao, J. Feng, and Y. J. A. c. p. Ci, "Human breast carcinomal tissues display distinctive FTIR spectra: implication for the histological characterization of carcinomas," vol. 18, no. 2, pp. 87-93, 1999.
- [93] R. Eckel, H. Huo, H.-W. Guan, X. Hu, X. Che, and W.-D. J. V. S. Huang, "Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue," vol. 27, no. 2, pp. 165-173, 2001.
- [94] D. J. Lyman and J. J. A. s. Murray-Wijelath, "Fourier transform infrared attenuated total reflection analysis of human hair: comparison of hair from breast cancer patients with hair from healthy subjects," vol. 59, no. 1, pp. 26-32, 2005.
- [95] S. M. Han *et al.*, "Synchrotron nanoscopy imaging study of scalp hair in breast cancer patients and healthy individuals: Difference in medulla loss and cortical membrane enhancements," vol. 79, no. 1, pp. 23-30, 2016.
- [96] D. C. Malins, N. L. Polissar, S. Schaefer, Y. Su, and M. Vinson, "A unified theory of carcinogenesis based on order-disorder transitions in DNA structure as studied in the human ovary and breast," *Proceedings of the National Academy of Sciences*, vol. 95, no. 13, pp. 7637-7642, 1998.
- [97] M. Kotkova, V. Sitnikova, T. Nosenko, T. Kotkova, M. Uspenskaya, and R. Olekhovich, "Spectroscopic Study of Blood Serum of Patients With Breast Cancer," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2018, pp. 657-660: IEEE.
- [98] A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin, and P. L. Martin-Hirsch, "Vibrational spectroscopy of biofluids for disease screening or diagnosis: translation from the laboratory to a clinical setting," *Journal of biophotonics*, vol. 7, no. 3-4, pp. 153-165, 2014.
- [99] M. Aaboe, B. V. Offersen, A. Christensen, and P. A. Andreasen, "Vitronectin in human breast carcinomas," *Biochimica et Biophysica Acta (BBA)-Molecular basis of disease*, vol. 1638, no. 1, pp. 72-82, 2003.
- [100] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Structural & Molecular Biology*, vol. 9, no. 9, p. 646, 2002.
- [101] J. Drenth, *Principles of protein X-ray crystallography*. Springer Science & Business Media, 2007.
- [102] K. Wüthrich, "NMR with proteins and nucleic acids," *Europhysics News*, vol. 17, no. 1, pp. 11-13, 1986.
- [103] A. J. B. e. B. A.-B. Barth, "Infrared spectroscopy of proteins," vol. 1767, no. 9, pp. 1073-1101, 2007.
- [104] A. J. J. L. o. B. M. Jabs, "Determination of secondary structure in proteins by fourier transform infrared spectroscopy (FTIR)," 2005.
- [105] J. A. Hering and P. I. Haris, "FTIR spectroscopy for analysis of protein secondary structure," *Biological and biomedical infrared spectroscopy*, pp. 129-167, 2009.
- [106] W. K. Surewicz, H. H. Mantsch, and D. Chapman, "Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment," *Biochemistry*, vol. 32, no. 2, pp. 389-394, 1993.

- [107] D. C. Lee, P. I. Haris, D. Chapman, and R. C. Mitchell, "Determination of protein secondary structure using factor analysis of infrared spectra," *Biochemistry*, vol. 29, no. 39, pp. 9185-9193, 1990.
- [108] R. Pribic, I. Vanstokkum, D. Chapman, P. I. Haris, and M. Bloemendal, "Protein secondary structure from Fourier transform infrared and/or circular dichroism spectra," *Analytical biochemistry*, vol. 214, no. 2, pp. 366-378, 1993.
- [109] L. M. Miller, M. W. Bourassa, and R. J. Smith, "FTIR spectroscopic imaging of protein aggregation in living cells," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1828, no. 10, pp. 2339-2346, 2013.
- [110] P. I. Haris and F. Severcan, "FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media," *Journal of Molecular Catalysis B: Enzymatic*, vol. 7, no. 1-4, pp. 207-221, 1999.
- [111] F. Garczarek and K. Gerwert, "Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy," *Nature*, vol. 439, no. 7072, p. 109, 2006.
- [112] J. Grdadolnik and Y. Maréchal, "Hydrogen–Deuterium Exchange in Bovine Serum Albumin Protein Monitored by Fourier Transform Infrared Spectroscopy, Part I: Structural Studies," *Applied spectroscopy*, vol. 59, no. 11, pp. 1347-1356, 2005.
- [113] S. J. Prestrelski, N. Tedeschi, T. Arakawa, and J. F. Carpenter, "Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers," *Biophysical journal*, vol. 65, no. 2, pp. 661-671, 1993.
- [114] R. Lu, W.-W. Li, A. Katzir, Y. Raichlin, H.-Q. Yu, and B. Mizaikoff, "Probing the secondary structure of bovine serum albumin during heat-induced denaturation using mid-infrared fiberoptic sensors," *Analyst*, vol. 140, no. 3, pp. 765-770, 2015.
- [115] H. Fabian, C. Schultz, D. Naumann, O. Landt, U. Hahn, and W. Saenger, "Secondary structure and temperature-induced unfolding and refolding of ribonuclease T1 in aqueous solution: A Fourier transform infrared spectroscopic study," *Journal of molecular biology*, vol. 232, no. 3, pp. 967-981, 1993.
- [116] M. J. Baker, E. Gazi, M. D. Brown, J. H. Shanks, P. Gardner, and N. W. Clarke, "FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer," *British journal of cancer*, vol. 99, no. 11, p. 1859, 2008.
- [117] L.-P. i. Choo, D. L. Wetzel, W. C. Halliday, M. Jackson, S. M. LeVine, and H. H. Mantsch, "In situ characterization of beta-amyloid in Alzheimer's diseased tissue by synchrotron Fourier transform infrared microspectroscopy," *Biophysical Journal*, vol. 71, no. 4, pp. 1672-1679, 1996.
- [118] M. Szczerbowska-Boruchowska *et al.*, "Biomolecular investigation of human substantia nigra in Parkinson's disease by synchrotron radiation Fourier transform infrared microspectroscopy," *Archives of biochemistry and biophysics*, vol. 459, no. 2, pp. 241-248, 2007.
- [119] A. Kretlow, Q. Wang, M. Beekes, D. Naumann, and L. M. Miller, "Changes in protein structure and distribution observed at pre-clinical stages of scrapie pathogenesis," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1782, no. 10, pp. 559-565, 2008.
- [120] P. I. Haris, "Probing protein–protein interaction in biomembranes using Fourier transform infrared spectroscopy," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1828, no. 10, pp. 2265-2271, 2013.

- [121] M. Jackson, P. I. Haris, and D. Chapman, "Fourier transform infrared spectroscopic studies of calcium-binding proteins," *Biochemistry*, vol. 30, no. 40, pp. 9681-9686, 1991.
- [122] M. Jackson and H. H. Mantsch, "The use and misuse of FTIR spectroscopy in the determination of protein structure," *Critical reviews in biochemistry and molecular biology*, vol. 30, no. 2, pp. 95-120, 1995.
- [123] W. K. Surewicz and H. H. Mantsch, "New insight into protein secondary structure from resolution-enhanced infrared spectra," *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, vol. 952, pp. 115-130, 1988.
- [124] H. Fabian and D. Naumann, "Methods to study protein folding by stopped-flow FT-IR," *Methods*, vol. 34, no. 1, pp. 28-40, 2004.
- [125] S. Zhang and A. Rich, "Direct conversion of an oligopeptide from a β -sheet to an α -helix: a model for amyloid formation," *Proceedings of the National Academy of Sciences*, vol. 94, no. 1, pp. 23-28, 1997.
- [126] J. Kong and S. Yu, "Fourier transform infrared spectroscopic analysis of protein secondary structures," *Acta biochimica et biophysica Sinica*, vol. 39, no. 8, pp. 549-559, 2007.
- [127] S. G. Kazarian and K. A. Chan, "Micro-and macro-attenuated total reflection Fourier transform infrared spectroscopic imaging," *Applied spectroscopy*, vol. 64, no. 5, pp. 135A-152A, 2010.
- [128] J. Pedersen, M. Coskun, C. Soendergaard, M. Salem, and O. H. Nielsen, "Inflammatory pathways of importance for management of inflammatory bowel disease," *World journal of gastroenterology: WJG*, vol. 20, no. 1, p. 64, 2014.
- [129] L. Hancock and N. Mortensen, "How often do IBD patients require resection of their intestine?," *Inflammatory bowel diseases*, vol. 14, no. suppl_2, pp. S68-S69, 2008.
- [130] G. Bouguen and L. Peyrin-Biroulet, "Surgery for adult Crohn's disease: what is the actual risk?," ed: BMJ Publishing Group, 2011.
- [131] S. Rakoff-Nahoum, L. Hao, and R. Medzhitov, "Role of toll-like receptors in spontaneous commensal-dependent colitis," *Immunity*, vol. 25, no. 2, pp. 319-329, 2006.
- [132] E. Viennois *et al.*, "Serum miRNA signature diagnoses and discriminates murine colitis subtypes and predicts ulcerative colitis in humans," vol. 7, no. 1, pp. 1-16, 2017.
- [133] M. Vijay-Kumar *et al.*, "Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5," *Science*, vol. 328, no. 5975, pp. 228-231, 2010.
- [134] J. Grdadolnik, "ATR-FTIR spectroscopy: Its advantage and limitations," *Acta Chimica Slovenica*, vol. 49, no. 3, pp. 631-642, 2002.
- [135] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201-1222, 2009.
- [136] Ø. Hammer, D. A. Harper, and P. D. J. P. e. Ryan, "PAST: Paleontological statistics software package for education and data analysis," vol. 4, no. 1, p. 9, 2001.
- [137] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37-52, 1987.
- [138] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical analysis*: Springer, 1978, pp. 105-116.
- [139] J. A. Pierce, R. S. Jackson, K. W. Van Every, P. R. Griffiths, and H. Gao, "Combined deconvolution and curve fitting for quantitative analysis of unresolved spectral bands," *Analytical Chemistry*, vol. 62, no. 5, pp. 477-484, 1990.

- [140] R. Sahu *et al.*, "Characteristic absorbance of nucleic acids in the Mid-IR region as possible common biomarkers for diagnosis of malignancy," vol. 3, no. 6, pp. 629-638, 2004.
- [141] J. B. Reitsma, A. S. Glas, A. W. Rutjes, R. J. Scholten, P. M. Bossuyt, and A. H. Zwinderman, "Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews," *Journal of clinical epidemiology*, vol. 58, no. 10, pp. 982-990, 2005.
- [142] A.-M. Šimundić, "Measures of diagnostic accuracy: basic definitions," *Ejifcc*, vol. 19, no. 4, p. 203, 2009.
- [143] C. Petibois *et al.*, "Determination of glucose in dried serum samples by Fourier-transform infrared spectroscopy," *Clinical chemistry*, vol. 45, no. 9, pp. 1530-1535, 1999.
- [144] A. Yamanaka, A. Hashimoto, T. Matsuo, M. Kanou, K.-I. Suehara, and T. Kameoka, "Analysis of kinetic uptake phenomena of monosaccharide and disaccharide by suspension TBY-2 cells using an FT-IR/ATR method," *Bioprocess and biosystems engineering*, vol. 30, no. 6, pp. 457-468, 2007.
- [145] P. G. Andrus and R. D. Strickland, "Cancer grading by Fourier transform infrared spectroscopy," *Biospectroscopy*, vol. 4, no. 1, pp. 37-46, 1998.
- [146] B. R. Wood, M. A. Quinn, F. R. Burden, and D. McNaughton, "An investigation into FTIR spectroscopy as a biodiagnostic tool for cervical cancer," *Biospectroscopy*, vol. 2, no. 3, pp. 143-153, 1996.
- [147] G. I. Dovbeshko, N. Y. Gridina, E. B. Kruglova, and O. P. Pashchuk, "FTIR spectroscopy studies of nucleic acid damage," *Talanta*, vol. 53, no. 1, pp. 233-246, 2000.
- [148] P. Ramalingam, Y. P. Reddy, K. V. Kumar, B. R. Chandu, and K. Rajendran, "Evaluation of metformin hydrochloride in Wistar rats by FTIR-ATR spectroscopy: A convenient tool in the clinical study of diabetes," *Journal of natural science, biology, and medicine*, vol. 5, no. 2, p. 288, 2014.
- [149] H. Fabian, M. Jackson, L. Murphy, P. H. Watson, I. Fichtner, and H. H. Mantsch, "A comparative infrared spectroscopic study of human breast tumors and breast tumor cell xenografts," *Biospectroscopy*, vol. 1, no. 1, pp. 37-45, 1995.
- [150] J. Barahona-Garrido *et al.*, "Serological markers in inflammatory bowel disease: a review of their clinical utility," *Revista de gastroenterologia de Mexico*, vol. 74, no. 3, pp. 230-237, 2009.
- [151] E. A. Scoville *et al.*, "Alterations in lipid, amino acid, and energy metabolism distinguish Crohn's disease from ulcerative colitis and control subjects by serum metabolomic profiling," *Metabolomics*, vol. 14, no. 1, p. 17, 2018.
- [152] D. M. Wiese *et al.*, "Serum fatty acids are correlated with inflammatory cytokines in ulcerative colitis," *PloS one*, vol. 11, no. 5, p. e0156387, 2016.
- [153] J. Ollesch, S. L. Drees, H. M. Heise, T. Behrens, T. Brüning, and K. Gerwert, "FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification," *Analyst*, vol. 138, no. 14, pp. 4092-4102, 2013.
- [154] L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszewski, G. D. Sockalingum, and M. J. Baker, "Investigating optimum sample preparation for infrared spectroscopic serum diagnostics," *Analytical Methods*, vol. 7, no. 17, pp. 7140-7149, 2015.
- [155] S. Yoshida *et al.*, "Fourier transform infrared spectroscopic analysis of rat brain microsomal membranes modified by dietary fatty acids: possible correlation with altered learning behavior," *Biospectroscopy*, vol. 3, no. 4, pp. 281-290, 1997.

- [156] K. A. Oberg, J. M. Ruysschaert, and E. Goormaghtigh, "The optimization of protein secondary structure determination with infrared and circular dichroism spectra," *European journal of biochemistry*, vol. 271, no. 14, pp. 2937-2948, 2004.
- [157] M. Meurens, J. Wallon, J. Tong, H. Noel, and J. Haot, "Breast cancer detection by Fourier transform infrared spectrometry," *Vibrational spectroscopy*, vol. 10, no. 2, pp. 341-346, 1996.
- [158] E. Gazi *et al.*, "A correlation of FTIR spectra derived from prostate cancer biopsies with Gleason grade and tumour stage," *European urology*, vol. 50, no. 4, pp. 750-761, 2006.
- [159] K. Gajjar *et al.*, "Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis," *Analytical Methods*, vol. 5, no. 1, pp. 89-102, 2013.
- [160] J. R. Hands *et al.*, "Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy," *Journal of neuro-oncology*, vol. 127, no. 3, pp. 463-472, 2016.
- [161] J. R. Hands *et al.*, "Attenuated total reflection Fourier transform infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples," *J. Biophotonics*, vol. 7, no. 3-4, pp. 189-199, 2014.
- [162] M. Lens and J. Newton-Bishop, "An association between cutaneous melanoma and non-Hodgkin's lymphoma: pooled analysis of published data with a review," *Annals of oncology*, vol. 16, no. 3, pp. 460-465, 2005.
- [163] Z. Hammody, R. K. Sahu, S. Mordechai, E. Cagnano, and S. Argov, "Characterization of malignant melanoma using vibrational spectroscopy," *The Scientific World Journal*, vol. 5, pp. 173-182, 2005.
- [164] E. Bogomolny, M. Huleihel, Y. Suproun, R. K. Sahu, and S. Mordechai, "Early spectral changes of cellular malignant transformation using Fourier transform infrared microspectroscopy," *Journal of biomedical optics*, vol. 12, no. 2, pp. 024003-024003-9, 2007.
- [165] S. Mordechai *et al.*, "Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma," *Journal of microscopy*, vol. 215, no. 1, pp. 86-91, 2004.
- [166] H. Ghazarian, B. Idoni, and S. B. Oppenheimer, "A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics," *Acta histochemica*, vol. 113, no. 3, pp. 236-247, 2011.
- [167] E. Heitzer, P. Ulz, and J. B. Geigl, "Circulating tumor DNA as a liquid biopsy for cancer," *Clinical chemistry*, vol. 61, no. 1, pp. 112-123, 2015.
- [168] D. J. J. o. c. e. Swinehart, "The beer-lambert law," vol. 39, no. 7, p. 333, 1962.
- [169] K. S. Schweizer and D. Chandler, "Vibrational dephasing and frequency shifts of polyatomic molecules in solution," *The Journal of Chemical Physics*, vol. 76, no. 5, pp. 2296-2314, 1982.
- [170] M.-C. W. Gast *et al.*, "Serum protein profiling for diagnosis of breast cancer using SELDI-TOF MS," vol. 22, no. 1, pp. 205-213, 2009.
- [171] A. R. Green *et al.*, "Barriers to screening colonoscopy for low-income Latino and white patients in an urban community health center," *Journal of general internal medicine*, vol. 23, no. 6, pp. 834-840, 2008.
- [172] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.
- [173] M. Moschandrew, U. Mahadevan, and S. Kane, "General health maintenance in IBD," *Inflammatory bowel diseases*, vol. 15, no. 9, pp. 1399-1409, 2009.

- [174] T. Castro Aguilar-Tablada, M. Navarro-Alarcón, J. Quesada Granados, C. Samaniego Sánchez, J. Rufián-Henares, and F. Nogueras-Lopez, "Ulcerative colitis and Crohn's disease are associated with decreased serum selenium concentrations and increased cardiovascular risk," *Nutrients*, vol. 8, no. 12, p. 780, 2016.
- [175] E. Andersson *et al.*, "Subphenotypes of inflammatory bowel disease are characterized by specific serum protein profiles," *PloS one*, vol. 12, no. 10, p. e0186142, 2017.
- [176] P. Vereecken, F. Cornelis, N. Van Baren, V. Vandersleyen, and J.-F. Baurain, "A synopsis of serum biomarkers in cutaneous melanoma patients," *Dermatology research and practice*, vol. 2012, 2012.
- [177] E. Legouffe *et al.*, "C-reactive protein serum level is a valuable and simple prognostic marker in non Hodgkin's lymphoma," *Leukemia & lymphoma*, vol. 31, no. 3-4, pp. 351-357, 1998.
- [178] W. W. Overwijk and N. P. Restifo, "B16 as a mouse model for human melanoma," *Current Protocols in Immunology*, pp. 20.1. 1-20.1. 29, 2001.
- [179] P. Jayaweera, S. Matsik, A. Perera, H. Liu, M. Buchanan, and Z. Wasilewski, "Uncooled infrared detectors for 3–5 μ m and beyond," *Applied Physics Letters*, vol. 93, no. 2, p. 021105, 2008.
- [180] H. Schwarzenbach and K. J. B. C. R. Pantel, "Circulating DNA as biomarker in breast cancer," vol. 17, no. 1, p. 136, 2015.
- [181] M. Kriege *et al.*, "Factors affecting sensitivity and specificity of screening mammography and MRI in women with an inherited risk for breast cancer," vol. 100, no. 1, pp. 109-119, 2006.
- [182] M. J. N. E. J. M. Kriege, "Magnetic Resonance Imaging Screening Study Group. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition," vol. 351, pp. 427-437, 2004.
- [183] M. S. G. J. T. Lancet, "Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS)," vol. 365, no. 9473, pp. 1769-1778, 2005.
- [184] J. M. Cameron, H. J. Butler, D. S. Palmer, and M. J. Baker, "Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying," *Journal of biophotonics*, vol. 11, no. 4, p. e201700299, 2018.
- [185] C. Hughes *et al.*, "Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum," *Journal of biophotonics*, vol. 7, no. 3-4, pp. 180-188, 2014.
- [186] K. Gajjar *et al.*, "Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer," *Analyst*, vol. 138, no. 14, pp. 3917-3926, 2013.
- [187] J. S. Koopmeiners and Z. Feng, "Asymptotic properties of the sequential empirical ROC, PPV and NPV curves under case-control sampling," *Annals of statistics*, vol. 39, no. 6, p. 3234, 2011.
- [188] J. A. Kimber, S. G. J. A. Kazarian, and b. chemistry, "Spectroscopic imaging of biomaterials and biological systems with FTIR microscopy or with quantum cascade lasers," vol. 409, no. 25, pp. 5813-5820, 2017.
- [189] M. Koç and E. J. A. S. R. Karabudak, "History of spectroscopy and modern micromachined disposable Si ATR-IR spectroscopy," vol. 53, no. 5, pp. 420-438, 2018.
- [190] J. D. Hanna, J. I. Scheinman, and J. C. J. P. C. o. N. A. Chan, "The kidney in acid-base balance," vol. 42, no. 6, pp. 1365-1395, 1995.

- [191] M. A. J. A. N. Napierala, "What is the Bonferroni correction," vol. 6, no. 4, p. 40, 2012.
- [192] R. Fluss, D. Faraggi, and B. J. B. J. J. o. M. M. i. B. Reiser, "Estimation of the Youden Index and its associated cutoff point," vol. 47, no. 4, pp. 458-472, 2005.
- [193] R. L. Baldwin and B. H. J. P. o. t. N. A. o. S. Zimm, "Are denatured proteins ever random coils?," vol. 97, no. 23, pp. 12391-12392, 2000.
- [194] M. Baldassarre, C. Li, N. Eremina, E. Goormaghtigh, and A. J. M. Barth, "Simultaneous fitting of absorption spectra and their second derivatives for an improved analysis of protein infrared spectra," vol. 20, no. 7, pp. 12599-12622, 2015.

APPENDICES

Appendix A

In this appendix, additional information on sample, study protocol and the discussion on feasibility of technique in clinical domain is provided.

Appendix A.1 Understanding the Clinical Feasibility of Technique for Disease

Screening

ATR-FTIR spectroscopy has proven to be an excellent alternative for observing the bio-molecular composition of body-fluids and their variations due to a myriad of pathologies[28]. Various studies have demonstrated the potential application of this technique in the forensic sciences for the routine confirmatory screening of biological evidence[48]. Consequently, several challenging aspects of the clinical applications of this technique have also been reported[184, 185]; thus, it is critical to have an in-depth understanding of these obstacles while developing a new diagnostic regimen.

One of the most prevalent issues in the clinical application[185] of this technique is the variation in the localized drying of the serum sample[184]. To address this, we have experimentally demonstrated the effect of the “Air Drying of Serum Samples” as in Figure A1. Major clinical feasibility parameters such as “Measurement to measurement” and “User to user reproducibility” during the spectral measurements are further tested and outlined in Figure A2. These studies show that the effect of localized drying phenomena of the sample on the ATR crystal surface is negligible compared to disease-associated changes as reported elsewhere. Note that the disease-associated alteration in the spectral data can be further optimized through the use of spectral classification techniques[186].

Furthermore, the user-friendly instrumentation, integrated with intelligent software packages for presenting the diagnostic regimen, is necessary for the acceptance of the technique as a possible technological development. A schematic of our instrumentation, accompanied with our standard spectral analysis process of chain functionality, data records, averages, normalizations, deconvolutions, and quantifications are briefly presented in our earlier study.

Air drying of serum samples: The drying phenomena of liquid droplets deposited on a solid surface depends on various, complex physical mechanisms, including fluid dynamics, properties of the liquid and contact surface, and heat transfer[184]. In the ATR sampling mode, the drying of serum samples results in complicated patterns[185]. We have tested the drying phenomena of the serum sample and their overall impact while comparing the composition within the control and colitis samples as shown in Figure. A1. The serum sample deposited on the surface was allowed to air dry at room temperature. Images (A-H) in Figure. A1 show the serum sample throughout the drying process continues.

Figure. A1 (A) is the image after 4 minutes of its deposition, and the following images (B-H) are at additional two-minute intervals. Figure. A1(I) shows the spectral data and their variation with the sample crystallization: Spectra 1 represents the first spectra measured after 4 minutes of sample deposition, 2 indicates the second spectra and 3 denote the third spectra. After 10 minutes of sampling and 5 spectral scans, the absorption curve displays promising overlap. The inset I(i) is the magnified view of the spectral range $3225\text{-}3350\text{ cm}^{-1}$, which is mainly comprised of water content and proteins. Similarly, inset I(ii) is the magnified view of spectral range $1610\text{-}1670\text{ cm}^{-1}$, a fraction of amide I region. The red dotted spectra show measurements before the sample has stabilized, while solid blue lines are these spectra after the sample becomes stable. These spectral

data and images clearly depict that the sample becomes stable on the crystal surface after approximately 10 minutes.

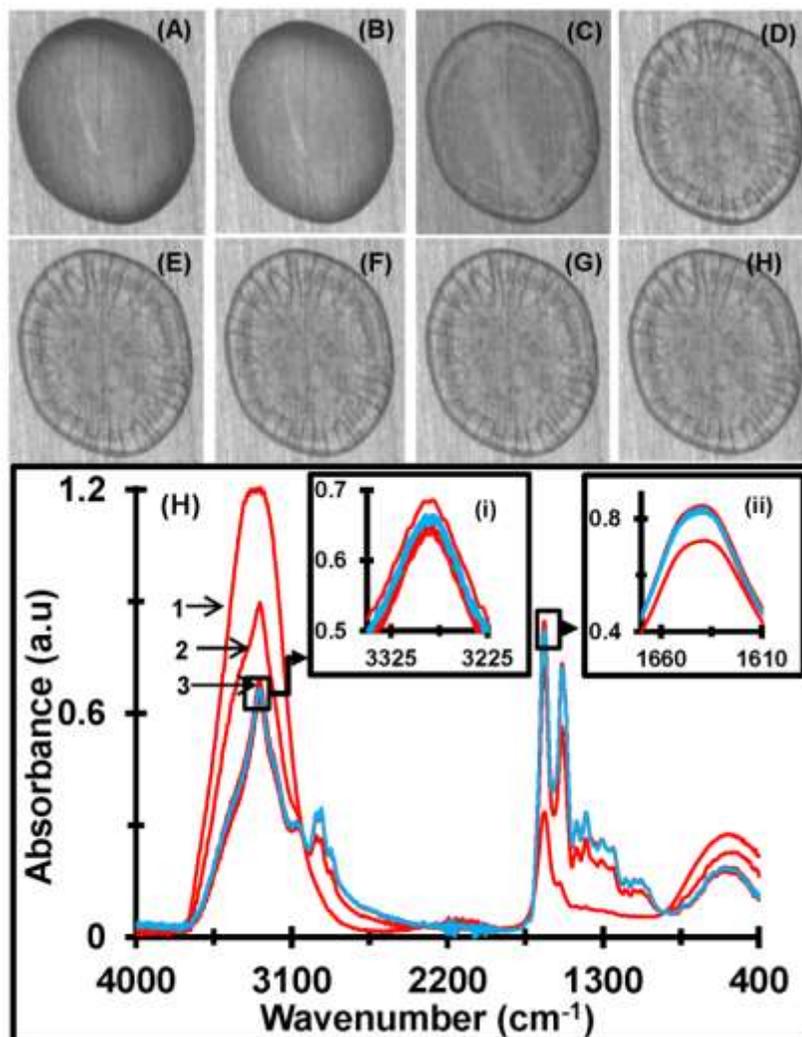


Figure A1 (A-H) Microscopic images of one microliter serum sample during different stages of drying at room temperature. Pictures (A) to (H) indicate the image of the serum in 2-minute intervals starting with (A) 4 minutes after its deposition and (H) 18 minutes after deposition. (I) FTIR spectra showing variation in absorbance during the air-drying of the serum samples. Spectra representing 1 is the first spectra taken after 4 minutes of sample deposition, 2 is second spectra (~ 5.2 minutes), 3 is third spectra (~ 6.4 minutes). Overlapping spectra (solid

blue —) indicate the stabilization of the sample. Inset I(i) and I(ii) show the magnified view of spectral range $3225\text{-}3350\text{ cm}^{-1}$ and $1610\text{-}1670\text{ cm}^{-1}$ to highlight spectral reproducibility nearly after 10 minutes of sample deposition.

Measurement-to-measurement and user-to-user reproducibility: We have further tested the measurement-to-measurement and user-to-user reproducibility as an important clinical feasibility parameter. Three researchers took sampling and spectral measurements to test user-to-user consistency. Likewise, the spectral measurements of twelve control and twelve diseased samples were taken to observe measurement-to-measurement reproducibility. Figure. A2(A) shows the p-value calculations, using the student's t-test (two-tailed, unequal variance) of the serum spectra, measured at two different times (the First and Second sets). Twelve measurements were taken using both the same sample and protocol. These results did not indicate any statistically significant wavenumber position differences from different measurements. Additionally, three different researchers (1st, 2nd, and 3rd) took 12 individual measurements and found the P-values as seen in Figure. A2(B). Similarly, we no statistically significant differences between spectral measurements were found. These results suggest that the ATR-FTIR technique of serum sampling/measuring offers a viable measurement-to-measurement reproducibility with minimal user-to-user variations. In conclusion, these results suggest that, despite the complicated drying phenomena of the serum samples, the measurement-to-measurement or user-to-user variation is negligible compared to the difference between control and disease samples. Hence, the overall spectral measurements affirm the technique as it verifies the clear similarity among the same types of samples while discriminating between colitis and control samples.

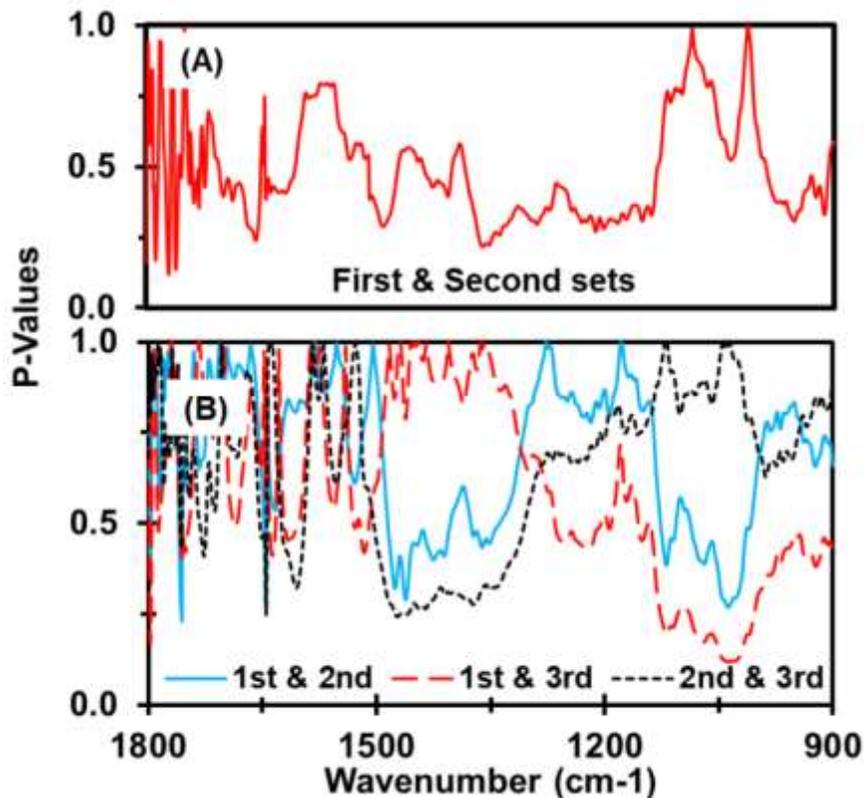


Figure A2 (A) Spectral discrimination by student's *t*-test (with two-tailed unequal variance) *p*-values of normalized absorbance using 12 measurements of the same sample at two different times. *P*-values $\gg 0.05$ clearly show the similarity between these spectra verifying better measurement-to-measurement reproducibility of the technique. (B) *P*-values calculated for the absorbance data measured by three different researchers. Significantly, similar *p*-values ($p \geq 0.18 \gg 0.05$) are found while comparing data from individual researchers, verifying the user-to-user consistency.

Appendix A.2 Penetration depth of infrared light in the sera

In ATR-FTIR spectroscopy[28], the light is totally reflected inside the prism of the higher refractive index with an evanescent wave from the reflected light penetrating into the serum sample[30]. The penetration depth of evanescent waves on the sample medium is given by the

equation, $d_p = \lambda/2\pi(\mu_1^2 \sin^2\theta - \mu_2^2)^{1/2}$, which depends on the wavelength of incident radiation (λ), the refractive index of crystal (μ_1), the angle of incidence (θ), and the refractive index of the sample (μ_2). The energy of the evanescent wave is absorbed by the sample and the reflected light is then coupled back into the system as shown in Figure. A3(A). The output spectra represent a snapshot of the molecular component within the sample medium.

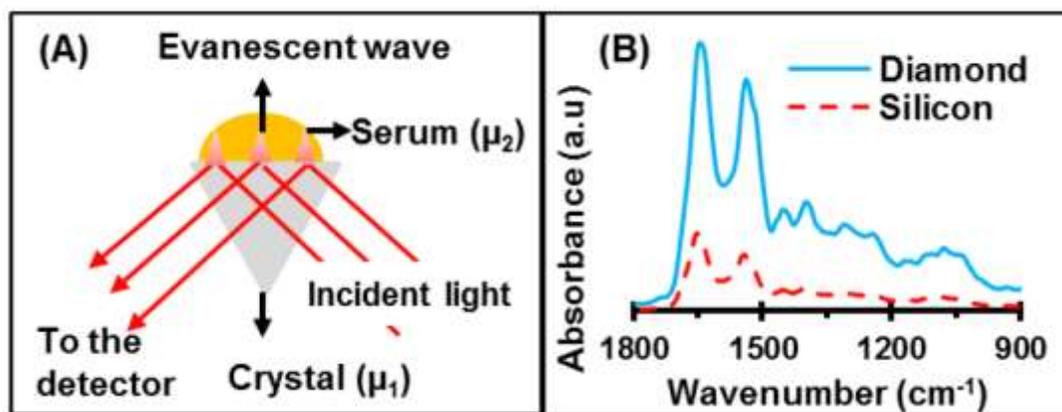


Figure A3 (A) Light-sample interaction and formation of evanescent wave in ATR sampling mode. (B) Output spectra of 1 μl of serum samples using diamond and Silicon crystals.

Specifically, for serum samples, penetration of around 2 μm can be achieved using diamond crystals and with an angle of incidence equals to 45 degrees. Fig. S3(B) shows two output spectra of 1 μl of the same serum sample using two different ATR crystals, Diamond and Silicon. These changes are due to the difference in the refractive indices of the materials of the two crystals. Spectra obtained with the use of the diamond crystal in the mid-infrared region is best for clinical application when compared with Silicon.

Appendix A.3 Sample Size Estimation for Diagnostic Accuracy in Human IBD Samples

Using 18 UC and 28 Control human samples, the sensitivity and specificity values are 78% and 82% for identifying spectral signature, ratio (α -helix/ β -sheet). These indicate that the spectral band ratio (α -helix/ β -sheet) has high diagnostic accuracy in detecting colitis and is thus a good signature while developing screening tool of colitis. We have compared different sample sizes by using the power analysis and simulation study that is common in medical diagnostic trials[172]. At a significance level $\alpha= 0.05$ and power levels $(1-\beta) = 0.90$, the required sample size is 70 each from control and disease groups (see table A1 as extracted from[172] and the initial values based on asymptotic theory[187]) to achieve the continuous-scale spectral signatures to have specificity 82% and sensitivity 90%. These 70 diseased samples can be divided into 18 initial stages, 18 aggravated stages, 17 treatment phases and 17 healing phases. In the table, false positive fraction (FPF) and true positive fraction (TPF) are used to find number of samples size.

Table A1 Estimated human sample sizes for continuous test using ROC (FPF0) as the basis for inference with $\alpha = 0.05$ and $\beta = 0.10$. Simulation studies were based on the binormal ROC curve with slope parameter 1 and the initial values based on asymptotic theory.

FPF0, FPF0	TPF1		
	.70	.80	.90
0.10,0.60	49	20	1
0.20, 0.70		44	0

Appendix A.4 Quantified Values at Spectral Signatures of UC Patients

Table A2 Quantified values of spectral signatures of UC patients. Shaded region of the table highlights the data extracted from male patients and unshaded are from female. Variability

of these signatures (I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} and α/β) does not show age, gender, height, weight, smoking and alcohol consumptions dependence in the uniform pattern, so a concluding remark cannot be made.

Gender	Age	Height (inch)	Weight	BMI	Smoke	Alcohol	I_{1033}	I_{1076}	I_{1121}/I_{1020}	I_{1629}/I_{1737}	α/β
Male	34	72	254	34.4	N	Y	0.2746	0.3518	1.4189	29.1055	1.9261
Male	41	72	240	32.5	N	Y	0.3252	0.3908	1.3069	37.5426	2.0303
Male	49	71	185	25.8	Y	Y	0.2741	0.3472	1.4072	32.7251	2.2569
Male	55	70	224	32.1	N	Y	0.3263	0.4058	1.3965	34.7197	1.9682
Male	63	69	166	24.5	N	N	0.2985	0.3715	1.3650	36.6690	2.0693
Male	71	68	189	28.7	Y	Y	0.2746	0.3537	1.4459	22.2489	2.4699
Female	21	65	136	22.6	Y	N	0.3583	0.4158	1.2559	33.7787	1.9947
Female	27	66	185	29.9	Y	Y	0.2769	0.3422	1.3803	36.8969	2.1782
Female	46	66	145	23.4	Y	Y	0.2856	0.3599	1.4062	29.5883	2.0761
Female	47	64	155	26.6	Y	Y	0.3103	0.3809	1.3763	33.0132	1.9962
Female	55	66	135	21.8	N	Y	0.2574	0.3288	1.4442	34.8734	2.0969
Female	59	64	170	30.1	Y	Y	0.3212	0.3953	1.3565	49.4526	1.9258
Female	60	66	205	33.1	Y	Y	0.2825	0.3603	1.4154	21.6636	2.0273
Female	63	67	187	31.1	N	N	0.2646	0.3409	1.4372	27.4452	1.8955
Female	65	65	163	27.1	N	Y	0.3276	0.3944	1.3203	46.3393	2.1512
Female	65	62	256	46.8	N	Y	0.2896	0.3639	1.3665	31.1652	1.9071
Female	82	66	200	32.3	N	Y	0.3225	0.3974	1.2887	30.2302	2.0975
Female	91	66	130	21	Y	N	0.3753	0.4161	1.1647	34.8415	2.0337

The quantified values of identifying spectral signatures; I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} and α/β of UC patients are shown in Table A2. These data do not depend on the age, gender, height, weight, smoking and alcohol consumptions of UC patients. However, while comparing with the corresponding values with control samples tabulated (Table A3), statistically significance difference can be seen, as discussed in the manuscript. These results show the identifying signatures are due to UC.

Table A3 *Quantified values of control samples at spectral marker signatures. Shaded region of the table highlights the data extracted from male and unshaded are from female controls.*

Variability of these signatures (I_{1033} , I_{1076} , I_{1121}/I_{1020} , I_{1629}/I_{1737} and α/β) does not show age and gender dependence.

Gender	Age (year)	I_{1033}	I_{1076}	I_{1121}/I_{1020}	I_{1629}/I_{1737}	α/β
Male	34	0.2439	0.3309	1.8472	32.5201	2.2639
Male	36	0.2724	0.3490	1.5322	20.6883	2.3639
Male	41	0.2986	0.3699	1.4209	31.1015	2.1678
Male	46	0.2516	0.3299	1.5741	28.0227	2.0996
Male	49	0.2687	0.3574	1.6699	24.1412	2.5469
Male	55	0.2698	0.3425	1.4936	30.7681	2.3792
Male	63	0.2994	0.3704	1.4425	31.7300	2.2136
Male	66	0.2567	0.3271	1.5818	31.9618	2.0859
Male	66	0.3056	0.3897	1.4194	24.839	2.4585
Male	68	0.2553	0.3282	1.5749	35.4579	2.1917
Male	71	0.2612	0.3346	1.4868	30.8804	1.9729
Female	21	0.3140	0.3897	1.3609	31.5131	2.2088
Female	27	0.2634	0.3419	1.4890	23.4095	2.1454
Female	35	0.2459	0.3314	1.5689	20.2486	2.4533
Female	46	0.2699	0.3459	1.5567	32.7598	2.2022
Female	47	0.2544	0.3302	1.4495	25.5301	2.1306
Female	49	0.2592	0.3249	1.5289	33.3139	2.0304
Female	53	0.2719	0.3613	1.6523	28.2439	2.4533
Female	53	0.2587	0.3249	1.5686	29.2131	1.9885
Female	54	0.2637	0.3283	1.5082	30.2426	2.0919
Female	55	0.2814	0.3503	1.5092	32.0153	2.2074
Female	56	0.3256	0.3883	1.2883	26.2872	2.2108
Female	57	0.2645	0.3363	1.5441	24.9430	2.2771
Female	58	0.2693	0.3432	1.4839	27.8657	2.4098
Female	58	0.2757	0.3474	1.4383	34.5593	2.3635
Female	59	0.2615	0.3389	1.4205	27.5615	2.3013
Female	60	0.3000	0.3776	1.4245	28.6436	2.2799
Female	60	0.3279	0.4052	1.5026	30.0617	2.5008

Appendix B

In this appendix additional information about the cancer study are presented.

Appendix B.1 Information for BC Patients

Table A4 Information about the BC patients: Biopsy results such as Invasive lobular carcinoma (ILC), Invasive ductal carcinoma (IDC), Ductal Carcinoma In-Situ (DCIS) confirmed the BC diagnoses. The tumor grading assessment was performed based on the national cancer institute (NCI) guidelines. Grade-2 denotes the intermediate grade or moderately differentiated tumor, whereas grade-3 denotes a high grade or poorly differentiated tumor.

ID	1	2	3	4	5	6	7	8	9	10
Age years	41.95	56.55	65.36	46.98	50.39	37.86	58.24	30.97	65.97	63.75
BMI	40.09	28.63	33.43	38.52	54.8	N/A	21.14	33.59	25.96	22.78
Menstrual Status	Pre-menopausal	Post-menopausal	Post-menopausal	N/A	N/A	Pre-menopausal	Unknown	Pre-menopausal	Post-menopausal	N/A
BC Discovery	Screening	Screening	Accidental	Accidental	Screening	Accidental	Self-exam	Self-exam	Self-exam	Pain
History	No	Yes	No	No	No	No	Yes	No	No	Yes
Tumor Size	.9 x .9 x 1.1cm	1 x .8cm	1.99x1.82x1.14 cm	5.3x3.9 cm	2.5 cm	6.6 x 3.9 x 4.2 cm	5 cm	2.4 x 2.1 cm	5.0 cm	7.0 x 5.3 x 3.7 cm
Laterality Biopsy	Right	Right	Left	Right	Right	Left	Left	Right	Left	Left
Biopsy Results	ILC	IDC	IDC	IDC	IDC	IDC	IDC	IDC	IDC	IDC
Malignant Grade	2	3	3	3	3	3	3	3	2	3
Lymph Nodes	Yes	Yes	Yes	No	Yes	No	No	Yes	No	No
Type of Surgery	Mastectomy	Mastectomy SLNB	Mastectomy	Partial Mastectomy (lumpectomy) SLNB	Mastectomy ALND	Mastectomy SLNB	Mastectomy SLNB	Partial Mastectomy (lumpectomy) SLNB	Partial Mastectomy (lumpectomy) SLNB	Mastectomy SLNB

Table A4. shows the BC patients' clinical information. The age of BC patients ranges from 30 years to 65 years and the age of control volunteers lies within this interval (41 to 58 years). The median age for BC patients and control volunteers remains similar; 53 years for BC and 52 years for controls. All BC patients were triple-negative (the hormone receptor subtype; estradiol receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2/neu) status was negative).

Appendix C

This appendix provides the details of the study protocol.

Appendix C.1 Research Protocol

Sample handling, measurement, data analysis was carried out in accordance with (university research) relevant guidelines and regulations. These regulations and the appropriateness of all the selected parameters are step wise discussed as the laboratory protocols while performing ATR-FTIR spectroscopy of serum samples for disease diagnostics. The corresponding original files are in the subfolder of L-drive (**L:\HemendraThesisMaterial\ ATR-FTIR study of serum samples**).

Spectrometer and parameter setting

1) **Spectrometer:** The Bruker Vertex-70 FTIR spectrometer, with a KBr beam splitter and Deuterated Tri-Glycine Sulfate (DTGS) detector, was used[3]. Here, the optimization of the application requires the selection of various components such as: source (Hg light), optics, electronics, beam splitter, and detector. The beam splitter transmits half of the radiation striking it and reflects the other half and KBr is the most widely used material for the IR beam splitters with coverage of $4000 - 400 \text{ cm}^{-1}$ mid-IR spectral region. DTGS detector is a sensitive room-temperature detector for mid-IR measurements. It uses temperature-sensitive ferroelectric crystals of deuterated triglycine sulfate. The working principle of DTGS detector involves, as the temperature and hence polarizability of the crystal's changes (due to the absorption of IR radiation) a charge is generated which is detected by two parallel electrodes. The deuterated form of the crystals is used because they have a higher Curie point.

2) **ATR accessory:** The MVP-Pro ATR accessory fitted with a diamond crystal (1 mm x 1.5 mm) configured to have a single reflection is used to provide high levels of output.

Diamond has a smaller acceptance angle cone, which allows having a good optical design to extract good spectra. Because of the strong, robust, and chemically inert feature of diamond, it is considered as the best ATR crystal for routine measurements on biological samples[127]. Germanium ATR crystal allows higher spatial resolutions (at high magnification) because of its higher refractive index ($\mu = 4$), but the use of high-magnification mode lowers the signal to noise ratio (SNR). Therefore, the germanium ATR crystal has not yet been used for the study of biological mediums[188]. Silicon ATR crystals are also not favored due to lattice vibrations within a biologically relevant region obscuring information below 1500 cm^{-1} wavenumber[189]. Conventional ATR crystals such as ZnS are not considered suitable for analyzing kidney excretions of patients with acid-base imbalance[190].

3) **Apodization:** It is an optical filtering technique and is widely used in recent cameras as well. As in the cameras, we can set apodization in our spectroscope. Please be sure that the application which helps us blur background and highlight only the object to be imaged (person who will be in photo) is apodization. It removes the airy disks caused by diffraction around the intensity peak and improves the focus. Herein, apodization is used to smooth the discontinuities at the end of recorded time. Medium Blackman-Harris 3-Term apodization function is employed to reduce noise without compromising resolution.

4) **Resolution and zero-filling factor:** It is the ability to separate two spectral lines that are very close in wavelength and it determines the maximum number of spectral peaks that our spectrometer can resolve. FTIR spectrometer provides different options to change resolution, and the change in resolution sets the change in travel distance of moving mirror. In the present study, A resolution of 4 cm^{-1} and a zero-filling factor of 4 to provide the best resolution ability (maximum signal-to-noise ratio). The 4 cm^{-1} resolution in the spectral data of range $400\text{-}4000 \text{ cm}^{-1}$

¹ means the system would be able to be resolving a maximum of 900 ($4000-400=3600/4$) individual wavenumbers across the spectrum. By doing so, we can get the protein secondary structure information and other spectral signatures appropriately. With a resolution of 8 cm^{-1} , it is found that this is not appropriate for secondary structural study, as second derivative spectra do not reflect the accurate position of disordered protein structures. By reviewing earlier studies, performing multiple studies, and acknowledging reviewers' recommendations we have used resolution of 4 cm^{-1} and a zero-filling factor of 4, so that the spacing between each data point is 1 cm^{-1} . We have performed repeated measurements and extract information from several studies, including references [23-25, 32], while selecting these parameters. This selection allows us to reduce noise, and to analyze the spectral data without losing information during spectral deconvolution. Resolution of 2 cm^{-1} gives noisy derivative spectra and of 8 cm^{-1} resolution gives rise to missing secondary structure information (spectral signatures show a lower degree of variability), while using diamond ATR crystal configured to have a single reflection of incident IR radiation. Therefore, a resolution of 4 cm^{-1} with a zero-filling factor of 4 to give a data spacing of 1 cm^{-1} is used.

5) **Zero-filling factor:** The zero-filling factor decreases the spacing between the data points as obtained from the resolution. Simply saying it increases the array-size of the spectral data. The zero-filling factor should follow (2^n , $n = 1, 2, 3, \dots$).

6) **Aperture size:** Aperture size is set to 2.5 mm for the detector response without saturation. Detector saturation needs to be avoided, otherwise it will lead to “starvation”, causing it to generate constant signals.

7) **Gas purging system:** A Parker-Ballston gas purging system was also used to maintain purified ambient air in the spectrometer.

Sample Management

1) While handling the biological samples in the laboratory, adherence to the university guidelines will be absolutely required ([Research Information - Georgia State Ahead \(gsu.edu\)](http://Research%20Information%20-%20Georgia%20State%20Ahead%20(gsu.edu))).

The exact date/time to pick the sample from the collaborator or suppliers (companies) should be scheduled earlier. The samples are stored in the freezer (at NSC room, Room Number 231) at a temperature -80°C upon received.

2) The availability of instruments, storage, and others is also checked before scheduling meetings with collaborators or before buying samples or performing the measurements. Necessary information was collected about the samples ahead so that we can set the measurement parameters accordingly and continue measurement immediately.

3) Samples were received in the labeled transparent storage containers kept systematically in a paper box, plastic, or dried ice containers depending on the nature of the sample. The storage condition is also checked with a colleague or with company representatives who had collected these samples, before the measurement.

4) Currently, the freezer at room 232 NSC (Dr. Hasting's lab) is used to store samples. Container tubes that are put in one column of paper or plastic box with the proper identification in the top rack of the refrigerator (Marked/labeled paper box with the name). These samples were stored properly (labeled/ marked) so that it could be safe while using the refrigerator by other research groups as well.

5) On the day of the measurements, sample storage container tubes with the sample were kept above the ice for thawing. (Use white LIFOAM Boxes and you can get ice in the 2nd floor of the Natural Science Center (NSC) (on the corridor back of Pete's office, shop manager) and 3rd (Undergraduate biology lab) floors of the same building. Since we are doing label-free

analysis, so environmental conditions for all the measurements should be the same (i.e. all samples should be dipped in ice immediately taking out from the freezer, before measurement). Multiple freezing and unfreezing alter the protein structures, so it is highly recommended to finish all measurements in one sitting).

6) Other issues like the use of cooling gases, (such as He & N₂), leaving the sample in air, crowded environment of the lab is also taken care of as they are not appropriate for our study.

Personal Security and Waste Management

We are analyzing biological (Hazardous) samples, so extreme care is taken while handling these samples and while managing related waste products including used gloves, cleaning wipes, Pipette tips, glass needles.

1) Gloves (sometimes masks and lab coats also) were used while handling samples, Broken parts, and wastes such as plastic pipette tips, used gloves, glass pipette, Berkshire paper, and others. Considering gloves could be the vectors for infection related to these diseases: colitis and carcinogenesis, these should not be used when touching anything other than the samples (while leaving the room and touching the door handle, mobile phones, computer, water bottle, etc.).

2) All the used gloves, glass pieces, serum tubes, and wipe-papers are properly managed in the appropriate container with biohazard red plastic bags which is kept in our lab (Optoelectronics laboratory, 146/148 NSC, GSU).

3) These waste items were collected in appropriate cardboard boxes as recommended by university guideline and are safely discarded.

Spectral Response Measurements

The spectral response measurements involve several steps as summarized in the video movie (<https://www.youtube.com/watch?v=vn1EQ8liI4g>) Also I college link ?//

1) **Cleaning:** The surface of ATR crystal was first cleaned with sterile phosphate-buffered saline (SALINE) followed by ETHANOL. For the longevity of diamond crystal, we are using Berkshire paper instead of other tissue papers. Wipe saline and ethanol by Berkshire papers in an x pattern. Wipe only once in each direction so that other dust molecules attached to the crystal surface will be absorbed.

Note: Change pipette needles fixed with rubber suction for every new experiment (sample) to avoid contamination, and gently press the rubber vacuum suction (pump) to pump SALINE and ETHANOL. Otherwise, it will come up to rubber surface from the pipette tip and increases the chance of contamination.

2) **The cleanliness test** was then conducted to ensure there are no signal peaks (indicating serum residue or other contaminant signatures) higher than the environmental noise level. (Just click on the cleanliness test option). Make sure the light signal path includes the ATR crystal.

3) **The Background:** Before any experimental measurement, background spectra were measured. The background spectrum allows us to get the true values of absorption peaks due to molecules present in the sample. Background shows the presence of any environmental residue on the crystal surface or in the light path. The spectrum was performed on a clean and empty surface before each measurement. It allows us to get high-quality spectra using background corrections.

4) **Background Measurement:** Open OPUS, by double clicking it. A small window will open and use User ID: UPLab1 and Password: OPUS to open.

Again, follow the video if there is any confusion during any step of the measurements.

<https://www.youtube.com/watch?v=vn1EQ8liI4g> Also I college

Press **Measure** and then **repeated measurement**. Select, **Advanced** and choose the path where you want to save the data. Remove **tick mark** on the ATR spectra box of data blocks, for background measurement. Fix other parameters like wave number, scan and go back to the **Basic** tab. You will see a **repeated sample single channel** (lower one). After pressing the **repeated sample single channel**, you will see the IR spectra. Whenever the graphs of spectra start to overlap stop the measurement by pressing **Abort** (right click the green line on the bottom of windows).

5) **Sample measurement:** Load the desired amount of sample (I found 1 μ L is enough to cover the ATR crystal) over the crystal plate by using Gilson micropipette carefully. In every measurement, sample should have a similar droplet dimension. Also, take care of other environmental factors like temperature, pressure, amount of light and other settings (all the background conditions) should be similar for every measurement).

Note: Serum samples (1 μ L) are deposited in order to fully cover the crystal surface (I tried with 2 μ L and 0.5 μ L samples and found 1 μ L is the best option).

After loading the sample, click the **Advanced** button and tick the **ATR spectrum**, select the path, you could change the name from the background you used previously to a new name clearly indicating sample identification information. Click on **Background** you will see an arbitrary previous file already selected. Clear (delete) this old background and select the last background you took (i.e. the curve just before pressing the abort button). After choosing the background, double click load background. Return to the basic button and start scanning by pressing the **Repeated sample single channel**, you will get the IR spectra. Stop your measurement

by pressing the **Abort** button, if you find spectra starts to repeat itself and you have more than eight overlapping spectra.

6) **Averaging:** Take the average of 6 spectra (Referring: Bayesian statistics), which were repeated (i.e. Exactly overlapped) after the serum dries measuring. Each sample was scanned multiple times to get ten or higher-quality (exactly overlapping) spectral data within the mid-infrared range 400 to 4000 cm^{-1} and 6 reads of the 100 co-added scans for each sample (total of 600 scans) were averaged.

7) **Number of Scans:** Generally, the trend is for the number of scans to go as 2^n ($n = 2, 3, 4, 5, \dots$), but we are taking 100 scans for each spectral data. Thus, in each spectrum that are selected for the final analysis represents (6x100) 600 scans. There is not any scientific reason for this selection of 100 scans, but my recommendation is use scan 64 instead of 100. (Our experience is by choosing 32 scans, will result into noisy spectra, and using 128 scans will take time, so my recommendation is to use 64 scans). Depending on the laboratory conditions, these numbers may be different. It is thus strongly recommended to check whether everything is ok before going to actual measurements.

8) **Spectra saving:** If you think you got the best spectra that can represent your sample, go to file, save as, mode Data Point Table (.dpt), File name and path, Open them in Excel you will get the a set of data, so select columns (A and B), go to Data, text to the column, next, comma, next, finish.

9) **Normalization:** Normalization is a crucial process while analyzing IR spectral data of the present study, as they were recorded under different sample loading, which is influenced by fluid dynamics of sample, and drying of sample conditions. Spectral data of the serum samples could have different intensity levels. Normalization takes care of variation in IR bands of serum

as the spectra were recorded under the same experimental parameters. Herein, Min-max normalization of absorbance spectra and the vector normalization of second derivatives spectra were performed. Using OPUS 6.5 software, all the spectra were min-max normalized by scaling the range 1800 to 900 cm^{-1} . This range is chosen as it comprises a significant biochemical fingerprint of the biological material and the amides (I and II) bands[24].

Vector normalization[135] is another popular protocol. We are using min-max normalization in absorbance spectra and vector normalization in second derivative spectra. In the following points, I have explained the basic concept of these normalization and mathematical formulation, so that one can use desired mathematical programming knowledge while normalizing these data.

10) **Min-max normalization:** Assuming absorbance values (before normalization) are $X = x_1, x_2, x_3, x_4, \dots, x_n$. Find the maximum of them and say X_{max} and minimum as X_{min} . Min-Max normalized values are now;

$$X_{normalized} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad i = 1, 2, 3, \dots, N.$$

The min-max normalized data using this formula is compared as in figure A4 (A) with normalized data using OPUS, which guarantees our formula can be used while using other programming languages.

11) **Vector normalization:** Assuming absorbance values (before normalization) are $X = x_1, x_2, x_3, x_4, \dots, x_n$.

First, find the square of each variable as $(x_1)^2, (x_2)^2, (x_3)^2, \dots, (x_n)^2$. After finding the square of variables find the sum of them as; $S = (x_1)^2 + (x_2)^2 + (x_3)^2 + \dots + (x_n)^2$. Again, find the square root of S, which is also called norm of variable of X. Now vector normalized data are now;

$$X_{normalized} = \frac{x_i}{\sqrt{S}} \quad i = 1, 2, 3, \dots, N.$$

The vector normalized data using this formula is compared as in figure A4 (B) with normalized data using OPUS, which guarantees our formula can be used while using other programming languages.

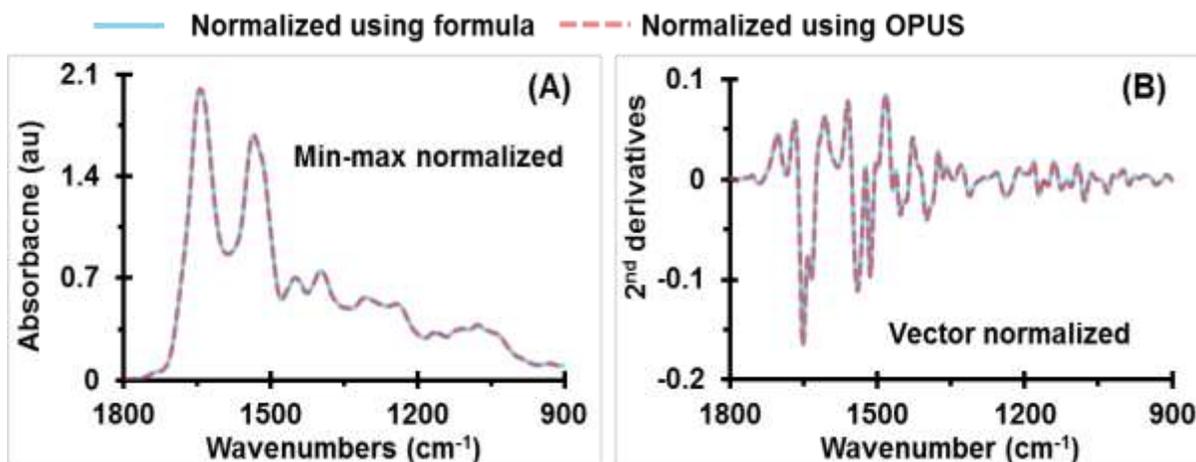


Figure A4 Verifications of mathematical formula for normalization. (A) Min-max normalization of absorption spectra. (B) Vector normalization of second derivative curves of the spectra. Spectra obtained from the formula overlaps with the corresponding spectra, as obtained from the OPUS software. It verifies, these formulas can be used while writing code in another programming language.

12) **Smoothing:** The second derivatives of Infrared spectra need smoothing.

The number of data in the absorption curve and the second derivatives curves are equal, so by smoothing the number of data points will be same. Let's say you have data $Y = (a,b,c,d,e,f,g,h,\dots,z)$ (i.e. total set of data =26). Now you want to have a smoothed set of data $Y' = (a', b', c', d', \dots, z')$. Here, $a' = (a)$, $b' = (a+b+c)/3$, $c' = (a+b+c+d+e)/5$, $d' = (a+b+c+d+e+f+g)/7$, $e' = (a+b+c+d+e+f+g+h+i)/9$, $f' = (b+c+d+e+f+g+h+i+j)/9$, $q' = (m+n+o+p+q+r+s+t+u)/9$, $r' = (n+o+p+q+r+s+t+u+v)/9$,, $v' = (r+s+t+u+v+w+x+y+z)/9$, $w' = (t+u+v+w+x+y+z)/7$ $y' = (x+y+z)/3$, $z' = z$. We have nearly

3500 sets of data and those crucial for our study lie in the middle, so no need to worry about these numbers in the extreme.

It is noted that we have to smooth only second derivative spectra. In OPUS software, while doing second derivatives, there are 9-point smoothing, 12-point smoothing, options. We cannot get second derivative spectra without smoothing in this software (Note: there is no option available to get second derivative spectra without smoothing). Herein, nine-point smoothing is used while taking the second derivative spectra. For clarification, a smoothed and unsmoothed spectral data using the vector normalized second derivative spectra is shown in figure A5. It is strongly recommended to use data before cutting 900-1800 cm^{-1} , as end points are affected during smoothing. We can use Adjacent averaging, as explained above or Savitzky-Golay smoothing function. The programming languages like Python are using Savitzky-Golay smoothing.

Look at the left side of the window and change the color of 2-3 last spectra so that you can see the difference clearly. If you think you got the best spectra that can represent your sample, go to file, save as, mode Data Point Table (.dpt), File name and path, open them in Excel you will get the sticking set of data, so select columns (A and B), go to Data, text to the column, next, comma, next, finish. The screen shot of each step is in the PPT of the folder.

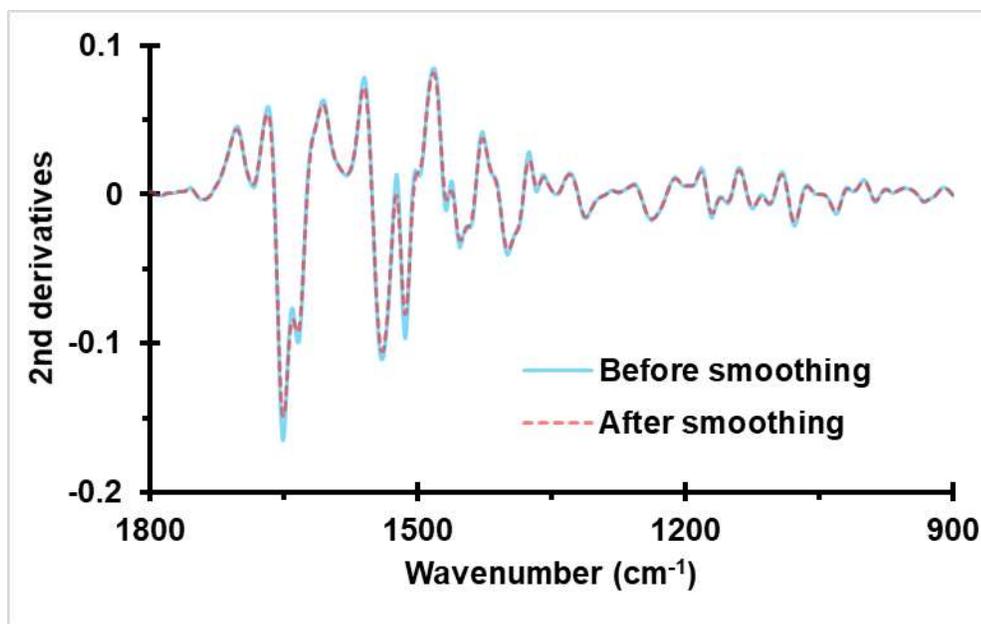


Figure A5 Spectral data before and after smoothing.

Data Analysis

The details of spectral analysis techniques are explained in the appropriate section of this dissertation.

1) First, the spectral data of the control and diseased samples within 400-4000 cm^{-1} wavenumber range. This mid-IR range includes bio fingerprint region (900-1500 cm^{-1}) and amide peaks (1500-1700 cm^{-1}).

2) Take the average of the final 6 spectra which are overlapped with each other (The sample size of at least 6 is needed for meaning full statistical analysis).

P-value calculation of min-max normalized absorbance: The spectral data of two groups (let's say control and diseased) were first compared using student t-test p-values. In this study we are using EXCEL to calculate the p-values and find the p-values at all the wavenumber positions.

3) Take the second derivatives with 9 point smoothing and do the vector normalize them to get the idea of minima and maxima of the data (that will be useful while fitting), for

Dendrogram tree diagram plot[63] and PCA[137]. Now vector normalized[135] the second derivatives curves.

4) Hierarchical Cluster Analysis (HCA) and Dendrogram Plot: Our analysis includes the dendrogram of HCA, which is commonly employed to identify the similarities between the FTIR spectra by using the distances between frequencies and aggregation algorithms[63]. Using the “PAST (Paleontological Statistics) 4 - the Past of the Future” software and the vector normalized second derivative curve of the absorbance spectra, as input data (variables), HCA is performed and the dendrogram is plotted. Dendrogram tree diagram performed, using Ward’s algorithm and Euclidian distance measurements, allow us to visualize the overall classification. Origin and MATLAB programming were also used for cluster analysis; however, use of PAST-4 is easy and has necessary inbuilt functions.

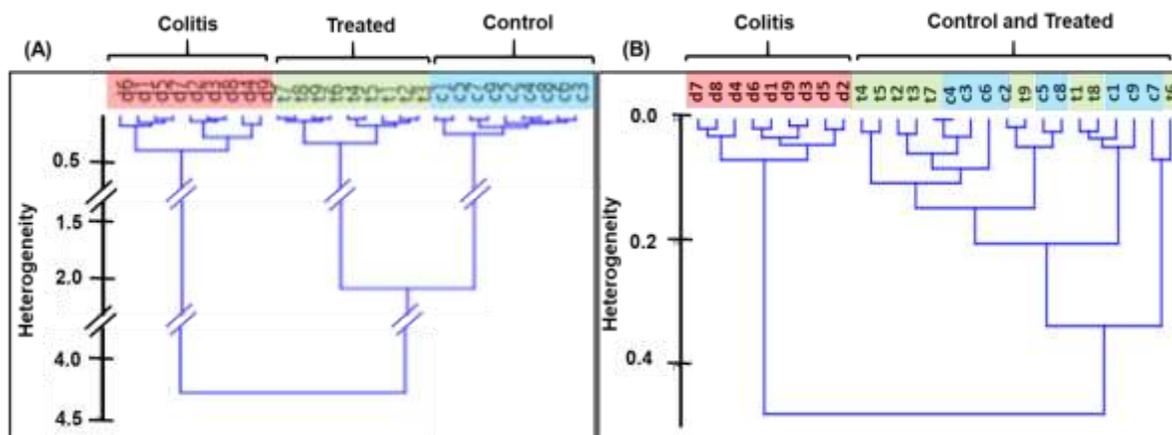


Figure A6 Dendrogram tree diagram performed, by using Ward’s algorithm and Euclidian distance measurements of vector normalized second derivatives curves within spectral range $1600-1700\text{ cm}^{-1}$. (A) A wide difference can be seen between spectra belonging to control and colitis mouse, compared to the degree of the dissimilarity between controls and treated while considering three groups constraints (1 for control: c1 to c9, 2 for treated: t1 to t9, and 3 colitis: d1 to d9). (B) Stratigraphic constraints of three groups showing colitis group has larger heterogeneity form

control and treated groups. While there is a similarity between control and treated groups in this classification.

Dendrogram tree diagram performed, using Ward's algorithm and Euclidian distance measurements, allow us to visualize the overall classification as shown in figure A6. In both analyses, spectral data from an equal number of mice were used (herein, we have randomly picked 9 controls). However, the degree of heterogeneity is less for treated groups from control compared with colitis while using group constraints as in figure 6A (A). Similarly, only colitis group shows strong degree of heterogeneity compared to control and treated groups in stratigraphic constraints. Herein, HCA analysis allows us to understand the preliminary idea about the potential application of the measurement techniques to segregate the groups.

5) PCA: PCA[137], a useful statistical analysis[137], is first performed to explain the holistic evaluation biomolecular content variations reflected in infrared absorbance data. Using the PAST 4 software vector normalized second derivative curve of the absorbance spectra, as input data variables, the variance-covariance matrix with the pairwise exclusion of missing values were analyzed to get the component plots. This can also be done easily by using Origin and SPSS software, but PAST-4 is easier one. The scatter plot of PC1 and PC2 were used to visualize the clustering of groups together with different magnitudes and directions. The scree plot is also used to check the total variance presented by PC1 and PC2 is significant. These findings from the PCA analysis led us to investigate spectral signatures useful in the clinical domain.

6) Spectral Deconvolution (fitting):

Watch the video first (<https://www.youtube.com/watch?v=zGHP2GBBdnE>) and look into section 2.4.3, figure 2.1 for details.

We perform quantitative analysis of the absorption band values at different spectral markers' positions, which follows discrimination of secondary structures by deconvolution of the spectra in the amide I region. During the deconvolution process, the computed curve that best fit with the experimental spectrum is obtained from the superposition of Gaussian function energy bands (GFEB). The individual bands from deconvolution represent proteins' secondary structures as discussed in the protocol papers, and similar studies in the field[108, 116-118, 120, 126] which were used to verify the applicability of this technique over complementary spectroscopic methods (X-ray crystallography, and nuclear magnetic resonance).

The process of deconvolution includes sectioning of min-max normalized spectra followed by rubber-band correction with two baseline points, such that absorbance values at extreme ends of the selected region will be zero as in figure A6. These spectra were then fitted with GFEBs by approximating the position and the number of bands from the minima of second derivative spectra and also by the minimization of root mean square error via the Levenberg-Marquardt[138] algorithm.

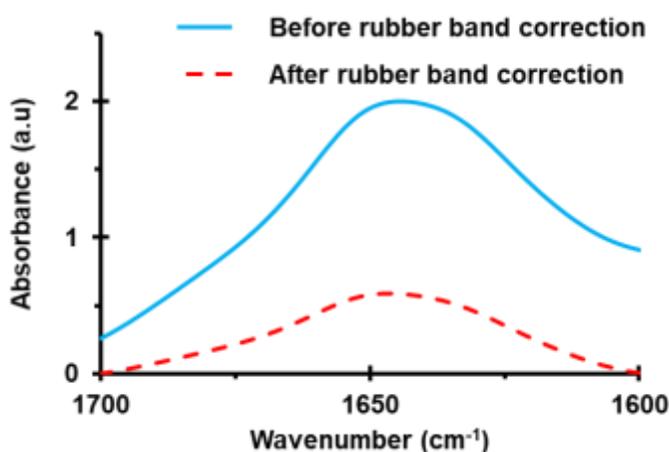


Figure A7 The process of rubber band baseline correction. The sectioned region 1600-1700 cm^{-1} is baseline corrected using baseline correction, such that the absorbance at the two extreme ends is 0.

7) Statistical Analysis: First, student's t-test is performed in EXCEL to calculate the p-value. P-value less than 0.05 (5% in Gaussian distribution) is considered as statistically significant in the null hypothesis. It is noted that Bonferroni correction[191] (new p-value, which is defined as the ratio of original p-value divided by the number of tests) is also performed to adjust type I error[191] (that has higher chances for a false positive; example includes analysis is showing differences in similar population due to problematic measurement or analyzing techniques) when conducting multiple analysis. It is noted that, using this new p-value, researchers calculate statistical power of the study.

Sensitivity, specificity the receiver operating characteristic (ROC) curves: Quantified values of these identifying spectral signatures are statistically analyzed to see the disease-associated alteration in the sample, and their stabilization after biological therapy. Our statistical analysis mainly includes finding the sensitivity and specificity of the signatures for their discrimination. Sensitivity and Specificity of a diagnostic test are often used to describe the diagnostic accuracy/performance of the analysis in biomedical research[141]. The discriminating potential of a diagnostic regimen can be quantified by the Youden index and the area under the receiver operating characteristic (ROC) curves [142]. The ROC curve is plotted to find the area under curve (AUC). The optimal cutoff value calculated based on the Youden index for each spectral marker is used to select the positivity/negativity of the disease and to estimate the sensitivity and specificity. Youden Index summarizes the displaying discriminatory accuracy of a diagnostic test in ROC curve while distinguishing two populations. It measures the effectiveness of test and enables for selection of optimal cut-off value[192].

Appendix D

In this appendix, information about the second derivative spectra is discussed.

Appendix D.1 Second Derivatives Spectra and Their Importance

In IR spectroscopy, molecules absorb light at resonance frequencies that are the characteristic of molecular structure. Where, the energy of vibration depends on molecular shape (potential energy), mass of atoms and the associated vibronic coupling (bonding)[3]. Therefore, the observed profile is a convolution of the line profiles of each individual mechanism of component constituents. The overall contribution of these individual vibrations is reflected in infrared absorbance spectra. Essentially, absorption spectrum is the analysis of the IR light interacting with molecules and is visualized in the graph of light absorption on the vertical axis vs wavenumber (or wavelength) in horizontal axis.

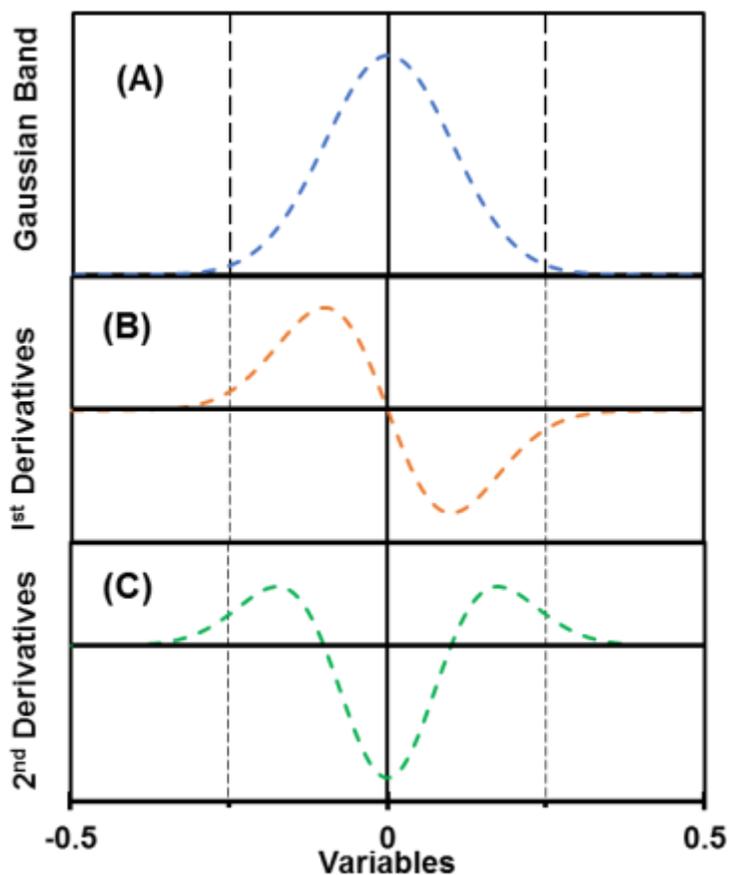


Figure A8 Derivative spectra and their importance for spectral analysis. (A) Gaussian function distribution. (B) First derivative spectra and (C) Second derivative spectra. The minima position of second derivative spectra exactly represents the maxima of Gaussian Distribution.

The line profiles, which have the shapes of Gaussian, Lorentzian and Voigt functions, represent the vibration of individual components and depends on the line position, maximum height and half-width. Therefore, curve fitting with these line shapes is sometime preformed either for interpolation, or to find the contribution of individual components. The Gaussian distribution (function) is used to explain the importance of second derivatives as in figure A7 (A). The first and second derivatives of the Gaussian curve are shown figure A7 (B) and figure A7 (C), respectively. The minima of second derivative curves exactly represents the maxima of Gaussian curve, so minima of second derivatives are used to approximate the number and position of GFEB's used to fit the spectra in our study. These second derivatives data also offer a practical and more specific method than routinely used absorption spectrum analysis to obtain the composition of proteins and nucleic acids. Therefore, second derivative curves are also used in PCA analysis and the dendrogram plots.

Appendix E

In this appendix, selection of the number of GFEBs (oscillators) for fitting the experimental absorbance curves is discussed. Second derivative spectra fitting reflecting the position of protein secondary structures is also discussed.

Appendix E.1 Number of GFEBs (oscillators)

In IR spectroscopy, the spectra were fitted in general with GFEBs by approximating the position and the number of bands from the minima of second derivative spectra and by the minimization of root mean square error. ATR-FTIR spectra taken at resolution 8 cm^{-1} , of air-dried serum sample at room temperature clearly depicts four valley positions [figure A9(A)] in the second derivative spectra showing significant presence of side chain, α -helix, β -sheet and β (anti-//+turn) structures. Random coil is an integral unit of a protein structures their contribution is significant in denatured proteins[193] analysis. However, the sera samples used in our study do not have denatured proteins, and our spectra were showing its impact on the infrared absorption is negligible compared to other four identified components. It is noted that the denaturation of proteins occurs at high temperatures or chemical interactions[193]. In order to be consistent, our analysis also includes an additional Gaussian band nearly at 1641 cm^{-1} reflecting random coil structures in protein secondary structures. Figure A9(B) and figure A9(C) shows the experimental curve fitting using GFEBs without ($n=4$) and with ($n=5$) the inclusion of random coil structures. This additional band does not alter the sensitivity or specificity of the identified distinguishing markers. Once the band at 1641 cm^{-1} is included, there is a slight variation in spectral signatures (specially, alpha helix/beta sheet) but the pattern of colitis in contrast to non-colitis remains the same as discussed, where absorbance spectra were measured at 8 cm^{-1} resolution[4, 5].

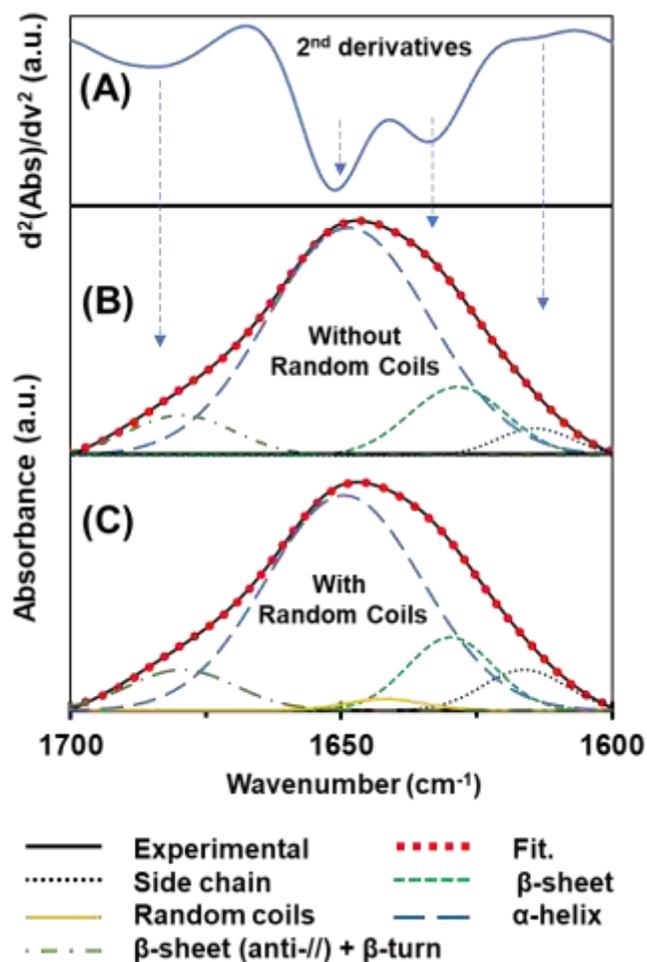


Figure A9 (A) Second derivative of ATR-FTIR absorbance of control sera at resolution 8 cm^{-1} , showing the deconvolved amide I region. (B) The individual secondary structure components were modeled using GFEB whose positions were determined from the second derivative of the absorbance to obtain simulated fits to the experimental curves. (C) Secondary structures were modeled using GFEB with the inclusion of random coil components approximately at wavenumber position 1641 cm^{-1} [4].

Importantly, with a resolution of 2 cm^{-1} gives noisy derivative spectra and of 8 cm^{-1} resolution gives rise to missing secondary structure information (spectral signatures show a lower degree of variability), while using diamond ATR crystal configured to have a single reflection of

incident IR radiation. By understanding of the positives and negatives of the techniques[122], and their optimizations protocols[123], we found that a resolution of 4 cm^{-1} allows us to overcome the challenges[139] of deconvolution process for quantitative analysis. At this resolution, two individual band positions for β -sheet (anti-//) and β -turn are reflected in the second derivative spectra. It further allows us to reduce the interspacing distance between bands. Increased spacing between individual bands is considered as one of the challenging aspect of fitting technique[139]. Figure A10 (A) shows the second derivative spectra of absorbance spectra of air-dried serum sample at resolution 4 cm^{-1} . The minima/valley positions in the second derivative spectra showing significant presence of side chain, α -helix, β -sheet, β -sheet (anti-//) and β -turn structures.

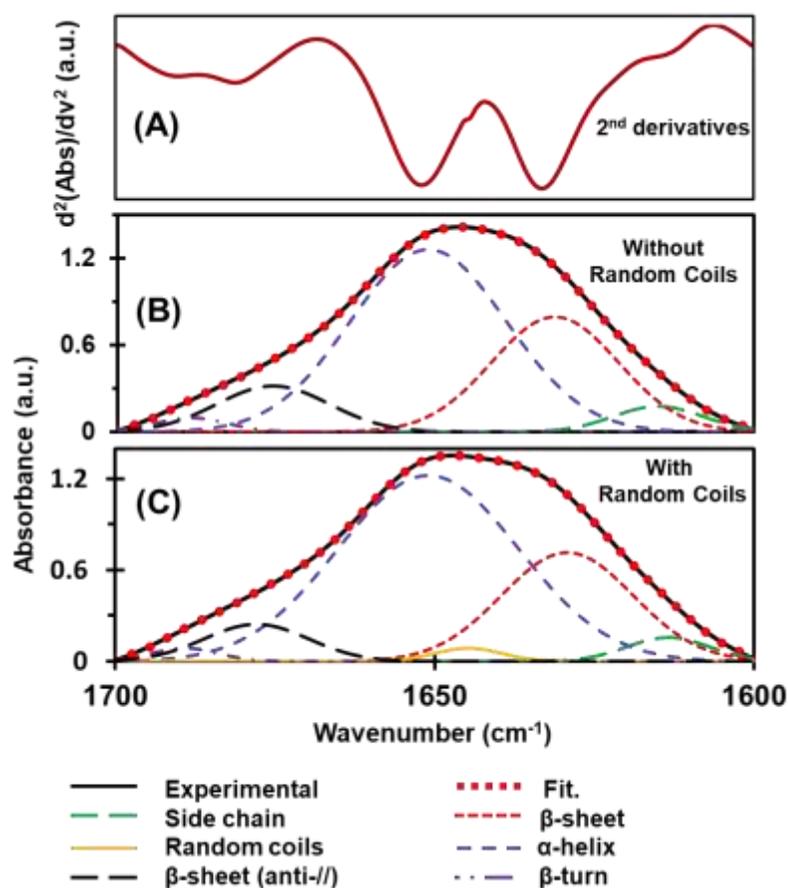


Figure A10 (A) Second derivative of ATR-FTIR absorbance of control sera at resolution 4 cm^{-1} , showing amide I, region fitting. (B) The individual secondary structure components were

modeled using GFEB whose positions were determined from the second derivative of the absorbance to obtain simulated fits to the experimental curves. (C) Secondary structures were modeled using GFEB with the inclusion of random coil components.

Experimental absorbance spectra taken at 4 cm^{-1} resolution were now fitted with (figure A10 (C)) and without (figure A10 (B)) random coil structures. The integral ratio of α -helix and β -sheet protein secondary structures with and without inclusion of random coil structure is analyzed using BC data presented on section 3.3 (Figures 3.13 to figure 3.15 above). Figure A11 (A) and figure A11 (B) are showing the ratio of the α -helix and β -sheet structures with and without including random coils in the fitting. Both figures show the level of β sheet structure increase and α helix decrease due to BC.

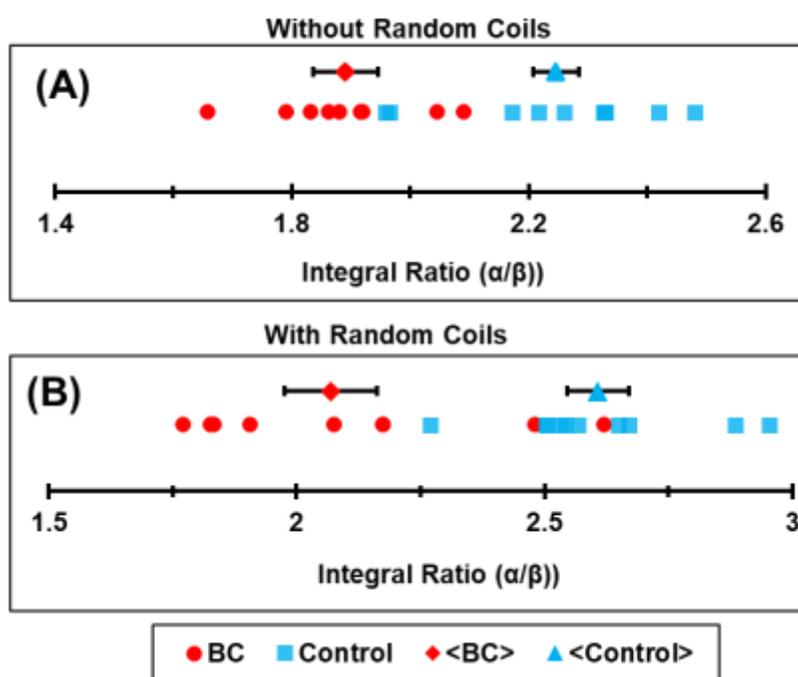


Figure A11 The integral ratio of α -helix and β -sheet protein secondary structures using five and six bands on experimental curves at resolution 4 cm^{-1} . (A) Ratio of integral values while fitting experimental curve without random coils structures. (B) Ratio of integral values while fitting experimental curves with including random coils structures.

Appendix E.2 Estimation of Bands' Position Using Second Derivatives Spectra Fitting

The second derivatives of absorbance spectra were also used to extract protein secondary structural information reflected in infrared spectral data[194]. The local minima position of the second derivative curves of the absorbance spectra provides the position of GFEB bands while analyzing absorbance data. Figure A12 shows the techniques for the estimation of the band's position using the second derivative fitting. Figure A12 (A) shows the second derivative spectra of the experimental absorbance spectrum. The values of the second derivative spectra were subtracted from 1 (Figure A12 (B)) at each wavenumber position and were rubber band corrected (Figure A12 (C)). Thus obtained rubber band corrected spectra are then fitted with Gussiand bands as shown in figure A12 (D). These second derivative spectral fitting offers a practical method to obtain supporting information about the chemical composition within the sample. However, the second derivative spectra may not provide all the biochemical components, like random coil structures; in the considered range as discussed above (figure A9-figure A11). It could be due to various reasons such as negligible presence of such protein structures or negligible contribution of the constituents for bonding vibrations.

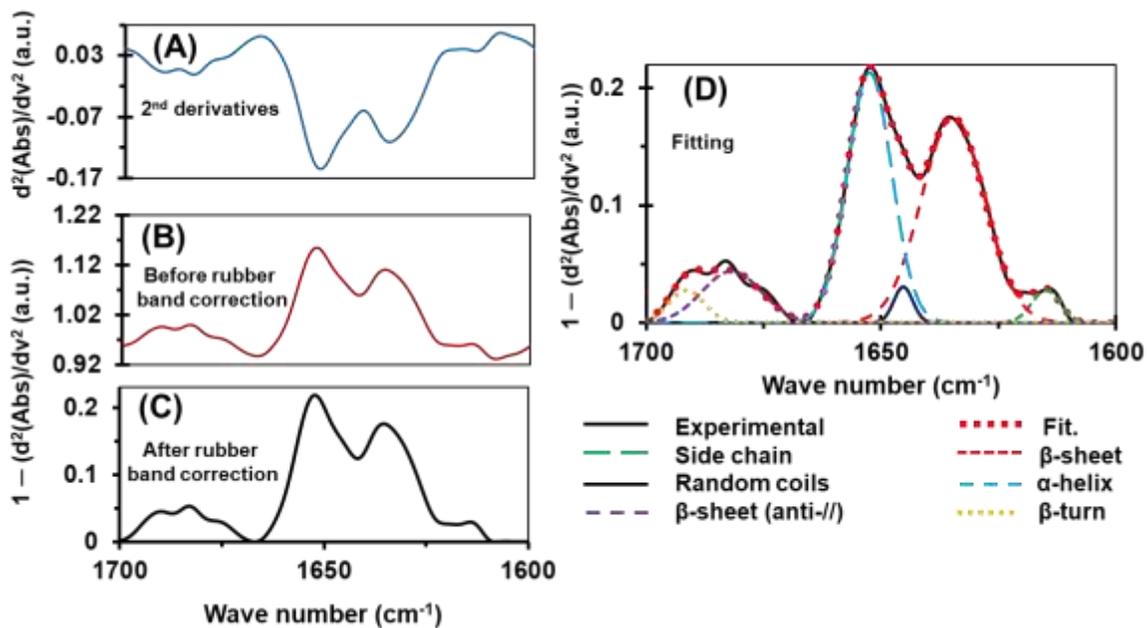


Figure A12 Second derivative curve fitting technique. (A) Second derivative spectra. (B) Values of second derivative spectra subtracted from one at each wavenumber position. (C) Rubber band correction of the curve obtained from one minus second derivative. (D) GFEBs fitting to estimate the position of the bands.