Comparison of Connective Tissue Invaded by Lewis Lung Carcinoma to Healthy Connective Tissue by Means of Micro-Raman Spectroscopy

by

Garnet Martin Cluff
B.Sc., Simon Fraser University, 2002

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE
in the Department
of
Physics

Simon Fraser University
Spring 2007

APPROVAL

Name: Garnet Martin Cluff

Degree: Master of Science

Title of Thesis: Comparison of Connective Tissue Invaded by Lewis Lung Carcinoma to Healthy Connective Tissue by Means of Micro-Raman Spectroscopy

Examining Committee:

Dr. Karen Kavanagh
Professor, Department of Physics (Chair)

Dr. Michael Chen, Senior Supervisor
Senior Lecturer, Department of Physics

Dr. Jenifer Thewalt, Supervisor
Associate Professor, Department of Physics

Dr. Martin Zuckermann, Supervisor
Adjunct Professor, Department of Physics

Dr. Andrew DeBenedictis, Examiner
Senior Lecturer, Department of Physics

Date Approved: December 7, 2006

SIMON FRASER UNIVERSITY library

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, BC, Canada

Abstract

It has been found that tissue invaded by cancerous cells gives rise to notable changes in the macro-Raman spectrum as compared to healthy tissue. To gain further insight, studies of the micro-Raman spectrum of connective tissue injected with Lewis Lung Carcinoma were undertaken. Raman images on the order of tens of microns were taken for the invaded connective tissue and a controlled connective tissue sample, with results showing increased nucleic acid and fatty acid/lipid concentrations in the invaded tissue sample, as well as lower collagen concentrations. The Raman images suggested a much higher cell density in the invaded connective tissue. A categorical principle component analysis was undertaken on Raman spectra sampled over one hundred μm^2 areas, to ascertain if one could use this method to classify a given tissue sample as cancerous. The results show a separation between the majority of cancerous tissue spectra as compared to the control tissue.

Keywords:

Micro-Raman spectroscopy of cancerous tissue

To my family and friends, whose support and understanding have always given strength.

Acknowledgements

I would initially like to thank Dr. Michael Short, whose insight and experience have proven to be an extremely valuable asset as I've traveled along the path to completion. His willingness to discuss and to occasionally stand in as a sounding board greatly helped during times of confusion. I would also like to thank my senior supervisor Dr. Michael Chen, who allowed me the freedom to explore on my own. Admittedly I took a decent amount of wrong turns and had at times some frustration as a result, but it gave me much more knowledge into the process of research as a whole.

Numerous other people also helped in the completion of my research. I would like to thank my committee members Dr. Jenifer Thewalt and Dr. Martin Zuckermann for their suggestions and advice; they have added to the clarity and quality of my final works. I would also like to thank Michelle Yeng and Sara Yuqiz for their aid in helping me prepare and stain my tissue samples. As well I would like to thank Haishan Zeng for providing me with my tissue samples, and Dr. Harvey Lui and Dr. David McLean for their support. Finally, I would like to thank Dr. Andrew deBenedictis for his role as internal examiner, agreeing to take the position on quite short notice (and doing a great job regardless).

Table of Contents

List of Figures

Introduction

Clinically speaking, numerous methods have been attempted in an effort to more accurately, cheaply, and quickly identify the presence of a cancerous or precancerous tumour in an otherwise normal and healthy region of tissue. Current processes consist of such methods as visual identification based on details inherent to a given type of cancer, performing biopsies, cutting and staining a given sample and identifying the large-scale detail of a tumour in that manner. The first method is quite inefficient, giving a lower than desired proper identification of tumourous regions, while the latter is rather time-consuming. It is the hope that by using Raman spectroscopy, one can provide a means to more quickly and accurately identify cancerous or precancerous regions, with the desire being that the tumour (and its actual boundaries) can be extracted with a minimum effect on the surrounding healthy tissue. Macroscopic Raman spectroscopy (studying regions of say a half mm by a half mm or larger) research has indeed shown that differences can be seen in cancerous and noncancerous tissue, suggesting that it may indeed provide a useful procedure for accurate identification of tumours [1]. In this study we intend to take this research a step further, to use micro-Raman spectroscopy (down to a 0.5 micron dot) in the hopes of being able to explore in better detail the differences that result in the altered Raman spectra seen in macro-level comparisons between cancerous and noncancerous tissue. As well, we intend to study the distribution of certain molecules that show higher (or lower) concentrations in cancerous tissue, and gain data and insight into these varied concentrations. It is the hope that any notable details seen can prove useful in identifying regions of cancerous tissue, be it a large-scale region or a region on the order of tens of microns.

Raman spectroscopy is dependent on inelastic scattering of photons incident upon a given molecule, with the aforementioned shift in frequency of the scattered light dependent on transitions of vibrational modes of the molecule.

The effect was first theorized by Austrian physicist A. Smekal in the early part of the 20th century [2]. The actual first empirical evidence of this inelastic scattering was first documented in 1928 by Sir C.V. Raman and his colleague K.S. Krishnan, who noted the existence of inelastic scattering through the use of filtered sunlight [2]. Two Moscow physicists, Mandelstam and Landsberg, almost simultaneously documented the results as well [2].

Raman and Krishnan verified the existence of inelastic scattering through the use of simple sunlight. The two would shine filtered sunlight through various gases and use coloured filters as monochromators, detecting shifts with nothing more than the naked eye [2]. The results of these experiments would eventually lead to the inelastic scattering being termed the Raman Effect, and would eventually lead to Raman winning the Nobel Prize in 1930. It is rumoured that Raman was so confident about winning the prize that year that he actually booked train tickets to Stockholm before the winner of the award was actually announced.

In the 1930's the notion that Raman scattering could be prove useful for chemical analysis came into acknowledgment. However, the fact that inelastic scattering was extremely weak when compared to elastic scattering always presented a key difficulty, resulting in lengthy sampling times before a relatively strong signal could be achieved. Initial light sources consisted of monochromatic mercury lamps, and while these proved quite useful, the eventual discovery of lasers proved advantageous, allowing for the potential to greatly reduce fluorescence, and also allowed for more accurate studies of low frequency shifts.

Micro-Raman spectroscopy came into being in 1966 [2], when it was suggested that the intensity of inelastic scattering should be independent of the sample volume; that is, that the ratio of Rayleigh scattering to Raman scattering should not be dependent on sample size.

It has been only recently that Raman spectroscopy has been considered for biological purposes. The problem was as always the weak intensity of the Raman scattered light as compared to the large fluorescence given off by biological tissue, usually resulting in a spectrum that has the Raman component entirely washed out by fluorescent radiation. Even with the use of new CCD cameras and the help of more powerful computers, macro-Raman spectroscopy still shows weak Raman peaks against a large fluorescent background when studying biological tissue (though the peaks are usually strong enough to provide insightful data). Micro-Raman spectroscopy works to further reduce this fluorescent noise by sampling a much smaller volume of material, and at the same time allows us to take spectra from micron scale sampling sizes.

The first three chapters will outline the basic theory behind not only Raman spectroscopy, but also some basic biological theory, and a quick look at the nature of cancer as a disease and how it develops. Chapter four will study the procedural details of how our micro-Raman spectrometer works, while chapter five looks at the experimental details used in our setup. Chapter six is a small section looking and comparing the tissue samples used for the experiment, while chapters 7 and 8 pertain to the analysis and results of the data taken (the good stuff).

1. Raman Scattering

As stated, Raman spectroscopy is based on the allowable vibrational energy modes of given molecules. Energy in the form of photons is incident upon a given molecule, whereby some of the energy is absorbed to increase the energy level of a given vibrational mode. This loss of energy can be seen when the photons are basically "re-emitted" by the molecule at a slightly lower energy level, shifted by the amount of energy required to cause the transition in the given vibrational mode. The desire to study these allowable vibrational modes is dependent on an understanding of how molecules vibrate in the first place.

1.1 Oscillating Molecules

An oscillation is the case where a system of particles undergoes periodic motion in which the center of mass of the system does not change, and the particles themselves pass through their equilibrium positions at coincident times. Two atoms oscillating due to a shared bond follow the basic kinetic principles of anharmonic vibrators, namely their motion is dictated by modeling their potential energy using a Morse curve (figure 1.1). For a given molecular bond, we can

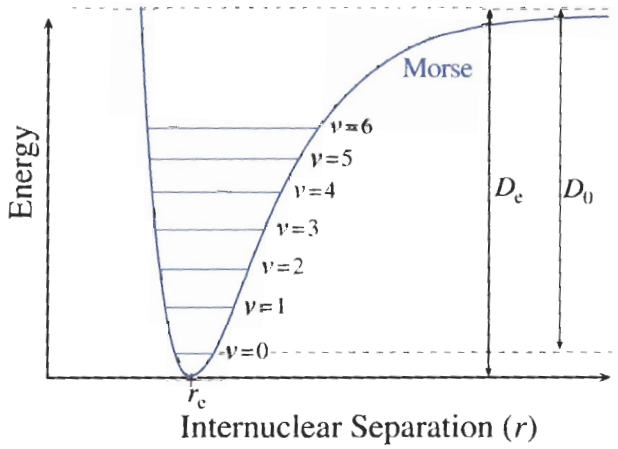

Figure 1.1: diagram of the Morse potential of a diatomic bond. Re is the equilibrium distance of the bond between the given atoms. Source: [3], by permission.

see a given distance where the potential energy is lowest (the equilibrium point). This is the defined bond length of the given bond between the atoms. Near this point, the energy scales upwards parabolically, and we can describe the motion using Hooke's Law,

$$\overline{F} = -k\overline{x}, \tag{1.1}$$

where *k* is a given constant and **x** is the displacement from the equilibrium point. Using this force law, the atoms will oscillate at a frequency of

$$\omega = \sqrt{\frac{k}{\mu}}, \tag{1.2}$$

where μ is the reduced mass of the two atoms. *k* depends on the bond and the atoms; weak bonds and heavy atoms give rise to lower frequencies (lower *k*) than strong bonds and light atoms. As the internuclear distance of the atoms gets larger, we see that the potential energy eventually flattens out; at this point the bond between the two atoms has become severed, and we no longer have periodic oscillations. As the distance between the atoms decreases past the equilibrium point, the potential energy of the system greatly increases.

The Morse potential energy is of the form [4]

$$E = D(1 - e^{a(R_{eq} - R)})^2, \tag{1.3}$$

where *D* is the dissociation energy of the bond, *a* is constant based on the atoms sharing the bond, *Req* is the equilibrium bond length, and R is the distance

between the atoms. The total vibrational energy of the molecule (potential and kinetic) is quantised, and is given by the equation [5]

$$E = h((V + ½)V_n - (V + ½)^2 X_e V_n), \qquad (1.4)$$

where V_n is the vibrational frequency, V is the given quantum number (0,1,2,..), h is Planck's constant, and X_e is the anharmonicity constant (based on how strong the given bond between the atoms is).

Number of Vibrational Modes

Of course, molecules aren't necessarily made up of only two atoms with one bond; we tend to have numerous atoms with equally numerous bonds. As such, instead of having only two atoms vibrate about one bond, we have numerous atoms vibrating about numerous bonds. The result of this is an increase in the number of ways a given molecule can vibrate at a given frequency, a frequency based on the size of the atoms, the strength of the bonds, and the geometrical distribution of said atoms.

In studying the different ways a molecule can vibrate, we need to ignore the rotational or translational location of the molecule. In other words, we want to isolate the energy of the vibration of the molecule away from translational or rotational velocity inherent to all the nuclei in the molecule; these are not vibrations. An example of this would be a tri-atomic molecule rotating about say one atom; the molecule has energy (kinetic), but not vibrational energy. Vibrational energy has to do with how the atoms in the molecule move with respect to each other; if the total velocity of the atoms with respect to the other atoms is zero, we don't have any vibrational energy.

The above seems to suggest we should isolate how many ways a given molecule can vibrate (the vibrational modes). Each given atom has 3 degrees of freedom in terms of its motion, namely movement in the x, y and z directions (we ignore the rotational moment of each atom, as again we are studying vibrational modes only, and in that regard whether the atom is spinning or not doesn't really matter). However, we want to ignore simple translational modes where all the atoms in the molecule are displaced equally, which means we remove three degrees of freedom (again, corresponding to 3 dimensional translational motion where all atoms displace equally). As well, we also need to subtract degrees of freedom where the entire molecule is rotating, as again these don't correspond to vibrational modes. Thus we need to remove three degrees of freedom here (corresponding to rotation in the orthogonal x, y, and z directions); two degrees for linear molecules (as linear molecules can only rotate in two ways where the molecules actually become displaced).

Consequently, we see that for a molecule with n separate atoms, the molecule can vibrate in $3n-6$ different ways, $3n-5$ if the molecule is linear. A simple diatomic molecule can only vibrate in one way (all diatomic molecules being naturally linear), with the atoms directly away from each other, then coming back (called a stretching mode). Molecules with three atoms can vibrate in 3 ways, two if the molecule is linear. Molecules with 4 atoms have 6 types of ways to vibrate, 5 atoms 9 vibrational modes, etc.

Quite naturally, the frequencies at which the molecules will vibrate can be found by calculating the normal modes allowable to the molecule. Recall from basic coupled oscillation theory that the vibrations caused by different bonds on a single atom will interact, giving rise to new vibrational modes seen (eg asymmetric stretching modes). An example here would be say a hypothetical linear H_3 molecule, 3 hydrogen atoms in a row. Instead of oscillating at

$$\omega = \sqrt{\frac{k}{\mu}}, \tag{1.1}$$

The system of atoms will oscillate at the frequencies of,

$$\omega = \sqrt{\frac{k}{m}} \quad \text{(symmetric stretching)}, \tag{1.5}$$

and

$$\omega = \sqrt{\frac{3k}{m}} \quad \text{(asymmetric stretching)}, \tag{1.6}$$

where m is the mass of hydrogen. In this way we can see then how multiple bonds on a given atom can give rise to new vibrational modes of differing frequencies and correspondingly different energies. In fact, these coupled oscillations can be used to more accurately identify the material being studied, by noting how much a given mode shifts in frequency as a result of other vibrating bonds. Butanol and ethanol, for example, have a C-C stretching vibrational mode that exists at slightly different frequencies, due to the slightly different makeup of each molecule resulting in differing coupled oscillations.

Types of Vibrations

There are two basic types of molecular vibrations, stretching and bending. Stretching modes consist of symmetric and asymmetric stretching, modes which are quite self-explanatory; for say O=C=O, symmetric is when the O's move

outward at the same time, and inward at the same time. For asymmetric stretching, the motion of the O's is 180 degrees out of phase; one O moves inward as the other moves outward (figure 1.2 shows these two stretching types).

Bending modes are more varied, and numerous types can be found, such as twisting, rocking, scissoring, breathing (ring molecules, where all the atoms move outward and inward at the same time), etc. Bending modes usually have more to do with changes in the angles of given bonds (with respect to a reference angle) than they do with a change in bond length, and consequently depend significantly on the geometry of the molecule. Since the bond lengths between the given nuclei don't change as much (if at all), the corresponding change in the energy levels of these vibrational modes is smaller, resulting in Raman scattered light with correspondingly lower frequencies. For a given *n* atom molecule, there are *n-1* stretching vibrations and *2n-5* bending vibrations.

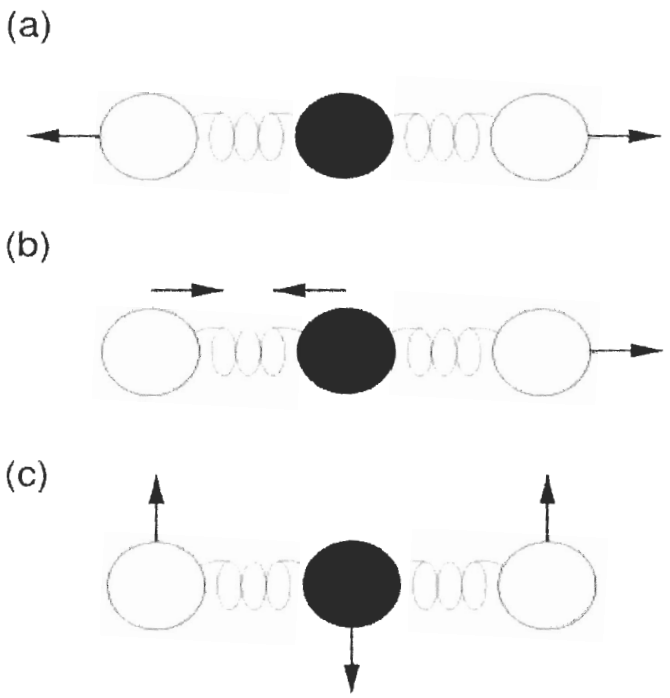

Figure 1.2: diagram of the vibrational modes of the linear molecule CO_2. A) shows the symmetrical mode (both O molecules move in and out at the same time), B) shows the asymmetric mode, and C) shows the bending mode. Source: [6], by permission.

Group Contribution of Vibrations

Generally speaking, organic based molecules can have lots of atoms, on the order of hundreds of atoms or more. Consequently, it would be impossible to try and identify every single type of allowable vibrational mode of the given molecule. In this case it is best to look at any chemical groups that constantly repeat themselves in the molecule, and classify each of these repeating groups as independently contributing to a given mode, based on the modes found in that repeating group. The most important assumption behind this idea is that the vibrations of these individual groups of molecules is largely unaffected by bonds that attach these groups to other groups. While this isn't always the case, it is usually quite valid, most notably when the bonds that interconnect the given groups of molecules oscillate at greatly differing frequencies than the frequencies of oscillations inside the molecular group (recall oscillations couple and affect the frequency of oscillations more so when the frequencies of the said oscillations are close in magnitude, and coupling effects drop off when the oscillations are farther apart in frequency). In this way then, one can focus on the frequencies of the vibrational modes of the given molecule groups, modes that hopefully will provide very intense Raman scattering simply by means of not only a high probability of scattering, but also simply due to the large number of recurring molecular sub-groups that exist inside the total molecule. An example here would be the O-P-O stretching mode seen in DNA chains; the phosphate molecule that comprises the backbone of the DNA chain occurs with regularity, giving rise to numerous O-P-O molecules that vibrate in the same way. These numerous vibrations give rise to a very intense Raman scattering event. Figure 1.3 shows a group vibrational mode known as breathing, where the individual atoms move inwards and outwards in tandem.

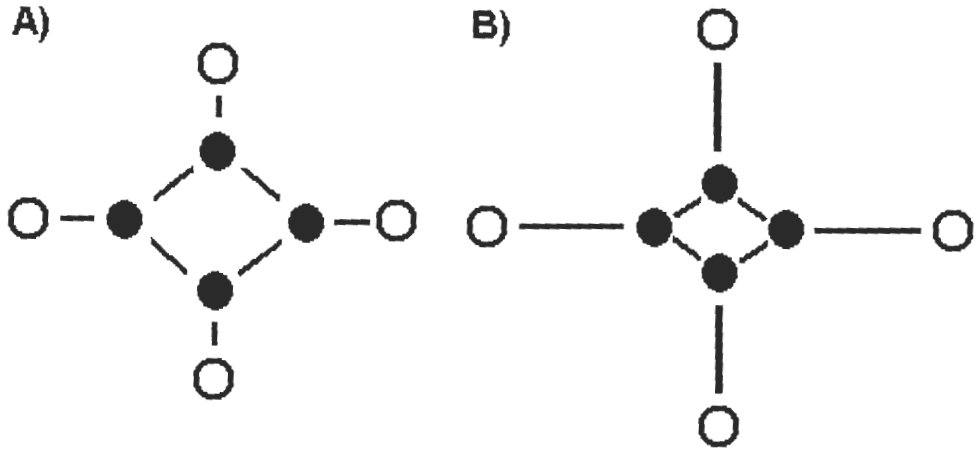

Figure 1.3: diagram of a group vibrational mode, otherwise known as breathing. The inner atomic nuclei form a ring, one that symmetrically expands and collapses due to the bonds with the outer atomic nuclei. A) shows all the inner atomic nuclei at their farthest apart from each other, B) shows them at their closest.

Vibrations and the Electron Cloud[1]

An important point to note when discussing Raman scattering is the distribution of electrons over the given molecule. The distribution and resultant varying density (over the molecule) of the electron cloud of a given molecule are based on the atoms and geometrical makeup of said molecule. If the atoms inside the molecule vibrate, the distribution of the electron cloud will undoubtedly alter, possibly resulting in a change in the dipole moment or polarizability of the given molecule. Polarizability is a measure of the ease by which the electron cloud can be induced to alter its motion under an applied electronic field. By changing the geometrical makeup of the molecule (as the atoms in the molecule vibrate), we potentially change how quickly or easily the electron cloud can alter its distribution in the presence of an electric field. In our case, the polarizability measured is the electron cloud's ability to basically polarize itself in accordance

[1] The theory pertaining to the qualitative and quantitative analysis of the Raman Effect can be found in numerous books, such as P. Hewdra, C. Jones, G. Warres, *Fourier Transform Raman Spectroscopy*, Ellis Howard, 1990.

with the oscillating electric field of an incident photon (as seen by looking head-on at the photon). This change in polarizability is vital for Raman scattering, as discussed below. Generally speaking, the vibrations of highly polar bonds are quite small, and have a resultant lesser degree of change in polarizability (as the electron cloud distribution does not change significantly over the vibration).

1.2 Photon Interaction

Qualitative View

For the most part, radiation incident upon an atom or a molecule interacts predominantly with the electron cloud around the nuclei of said atom or molecule (if it interacts at all). The photon polarizes the cloud (through EM interactions), causing the system of electrons to briefly adopt a highly unstable state called a virtual state. Normally, this virtual state quickly degrades, re-radiating the photon away at close to the same frequency (slight alterations in frequency occur due to the electron cloud distortion, though these frequency changes are tiny). This process is called elastic or Rayleigh scattering, and accounts for the majority of light scattered from the atom or molecule.

There is the possibility, however, that the photon interaction not only causes distortion of the electron cloud, but through distorting said electron cloud can also induce motion in the atomic nuclei of the molecules. If this occurs, the vibrational mode of the atom becomes altered (increases in energy), resulting in the photon being re-emitted at a lower energy level. The occurrence of this is quite rare when compared to Rayleigh scattering (one photon in 10^6 to 10^8), but when it does occur we have Raman scattering. By measuring the frequency shift of the Raman scattered light, we can successfully attain information on the difference in energy of the given vibrational mode of the molecule.

To induce a change in the energy state of the vibrational mode of the molecule, the molecule must be able to undergo a polarizability change in its makeup. If no polarizability change can occur (in other words, a change in how easy or hard it is for the electron cloud to distort), the oscillating EM field of the photon has no means by which to induce motion in the atom of the given molecule, and no means by which to change its vibrational state. We should note here that if the molecule already has a dipole moment (not one induced by altering the motion of the electron cloud, but one inherent in the atomic makeup of the molecule), it can directly absorb photons corresponding to the energy difference between two vibrational modes (where a difference dipole moment is induced); this is infrared absorption spectroscopy, however, and not Raman scattering.

In summary, for Raman scattering to occur, photons interact with the electron cloud of a given molecule, resulting in the rare case where the nuclei in the molecule become altered in terms of their motion. The photon becomes re-emitted at a lower energy level, while the molecule itself shifts into a higher energy vibrational mode. This loss of energy by the incident photon, resulting in a scattered photon with a lower frequency, is known as Stokes Raman scattering. It may also be the case that, if the molecule isn't in a ground state vibrational mode, that the scattered light will be higher in frequency; the interaction of the photon with the molecule again forms a virtual state, though one higher in energy (the energy of the photon plus the energy of the excited vibrational mode). This virtual mode can then degrade to the ground state vibrational mode of the molecule, giving off a photon with a corresponding higher frequency than the initial photon. This type of scattering is called anti-Stokes Raman scattering. The type of scattering dominates (Stokes or anti-Stokes) depends on the population of the higher energy state as compared to the ground state; normally in our samples (at the temperatures we use) the population of the ground state is much higher, resulting in stokes scattering being measured (we measure the frequency

shifts of photons with energy lower than the initial photons). Figure 1.4 shows the various energy shifts for Rayleigh, Stokes, and anti-Stokes Raman scattering.

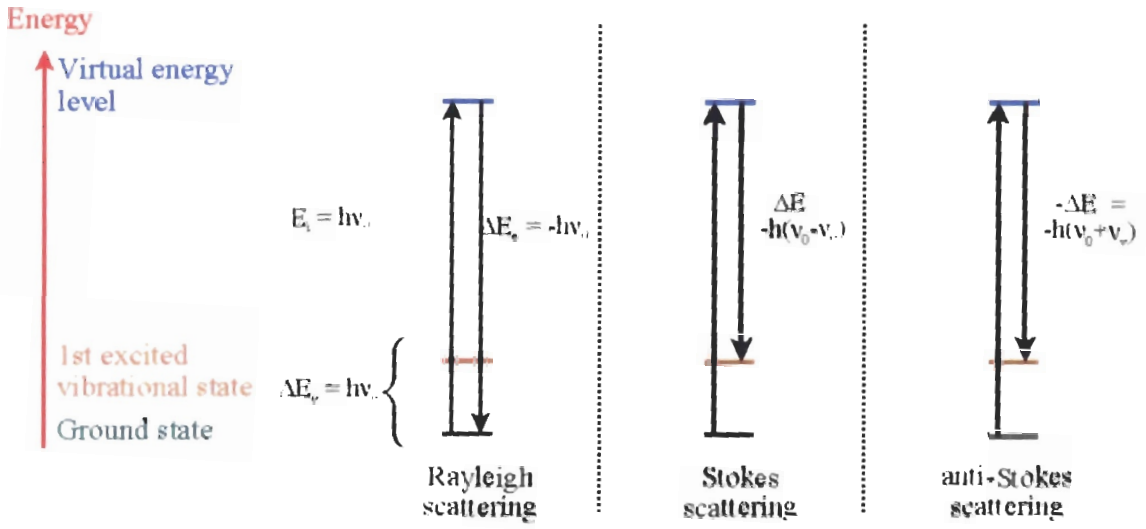

Figure 1.4: diagram of Rayleigh, Stokes, and anti-Stokes scattering. An initial photon of energy hv_o is incident upon the molecule, where it excites it to a virtual energy level. Depending on the selection rules and the initial vibrational state of the molecule, the molecule can spontaneously degrade in energy, giving rise to Rayleigh scattering (no energy change), Stokes scattering (photon re-emitted at lower energy), or anti-Stokes scattering (photon re-emitted with more energy). Source: [7], by permission.

Quantitative View

As discussed above, polarizability is a measure of how easy it is to alter the motion of the electron cloud in a given molecule. If we apply an electric field, we can form a dipole moment of the molecule (by distorting the electron cloud of the molecule), ***P***, such that

$$\overline{P} = \alpha \overline{E}, \tag{1.6}$$

where ***E*** is the electric field and α is the polarizability constant of the given molecule. The electric field causes the electrons in the molecule to oscillate,

becoming a source of secondary radiation. Now, it is not always the case that the induced dipole moment is actually in the same direction as the applied field, the moment may be in a different direction, depending entirely on how easy it is to move the electron cloud in one direction as compared to another. The polarizability of a molecule may allow electrons to move easier in one direction than say another, depending on the geometrical makeup (and strength) of the given atoms in the molecule. In matrix form, we have

$$[\alpha_{jk}] = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix}. \tag{1.7}$$

This then results in

$$P_i = \sum_j \alpha_{ij} E_j . \tag{1.8}$$

This matrix is symmetrical, and depends on the special orientation of the given molecule, but not on the applied electric field. As such, if the spatial orientation of the molecule should change, then the elements of the polarizability matrix may also change.

In the case of photons, where the field is shifting over time with a frequency F_o, we have

$$P = \alpha E_A \cos(2\pi F_0 t), \tag{1.9}$$

where E_A is the maximum amplitude of the applied field and t is the given elapsed time. We can also assume that the molecule itself is vibrating (recall a

molecule will vibrate even at its ground state), with the displacement position of the molecule from equilibrium being given by $x(t)$, with x_0 being the molecule's maximum displacement from equilibrium. In this case, we can model the oscillations away from the equilibrium point as

$$x(t) = x_0 \cos(2\pi\omega_v t), \qquad (1.10)$$

where ω_v is the frequency of oscillations about the equilibrium point.

One item we have to take into consideration is the fact that vibration of the molecule may change the polarizability of the given molecule; altering the geometrical distances of the nuclei in the molecule can affect the polarizability constants of the given molecule, as these constants are an inherent trait of the geometrical makeup of the molecule in the first place. Now, if the amplitude of the displacement from vibration is small (and linear), we can use a Taylor expansion and approximate the new polarizability constant as

$$\alpha = \alpha_0 + x(t)\left(\frac{d\alpha}{dx}\right)_0 + \cdots \text{ (higher orders small).} \qquad (1.11)$$

In other words, we can approximate α by multiplying the total displacement x by how much α changes over that given x (with the change in α over this interval being approximated as the derivative at $x = 0$). Obviously this will only work for small displacements, but in our model this will suffice, as we're focusing on initially being in the ground state and transitioning to the first excited state. Joining equations (1.10) and (1.11), we have the variation in polarizability as a function of time,

$$\alpha = \alpha_0 + x_0\left(\frac{d\alpha}{dx}\right)_0 \cos(2\pi\omega_v t). \qquad (1.12)$$

Putting equation (1.12) into equation (1.9), we have,

$$P = \alpha_0 E_A \cos(2\pi F_0 t) + E_A x_0\left(\frac{d\alpha}{dx}\right)_0 \cos(2\pi\omega_v t)\cos(2\pi F_0 t). \qquad (1.13)$$

Looking at the trig identity,

$$\cos\vartheta\cos\beta = \tfrac{1}{2}((\cos(\vartheta+\beta)+\cos(\vartheta-\beta)), \qquad (1.14)$$

we can then alter (1.13) to the form,

$$P = \alpha_0 E_A \cos(2\pi F_0 t) + \tfrac{1}{2} E_A x_0\left(\frac{d\alpha}{dx}\right)_0 (\cos(2\pi(F_0+\omega_v)t) - \cos(2\pi(F_0-\omega_v)t)). \qquad (1.15)$$

We see then that the induced dipole moment will not only vary with the frequency of the incident photon (F_0), but also as $F_0 + \omega_v$ and $F_0 - \omega_v$. As the molecule spontaneously relaxes from the higher energy induced dipole state (the virtual state), it will release energy, in the form of a photon with a frequency F_0 (Rayleigh scattering), $F_0 + \omega_v$, and $F_0 - \omega_v$. Obviously, for the photon to be emitted at a frequency $F_0 + \omega_v$, the molecule must impart some vibrational energy; it must have been in a higher order vibrational state than the ground state. $F_0 - \omega_v$ comes from energy being absorbed from the incident photon to increase the molecule vibrational energy by one level, $F_0 + \omega_v$ comes from a molecule in an excited vibrational mode falling to a lower vibrational mode. F_0 -

ω_v is again called Stokes Raman scattering, $F_0 + \omega_v$ is anti-Stokes. Again, as stated above, the majority of molecules in our study (based on our experimental setup) are initially in the ground state, so we study Stokes Raman scattering.

Equation (1.15) also gives some insight into why inelastic scattering is so low when compared to Rayleigh scattering; the intensity of scattered light, being proportional to $|P^2|$ [8], is consequently dependent on the coefficients of the given frequency of oscillation (F_0, $F_0 + \omega_v$, $F_0 - \omega_v$). The coefficients for $F_0 + \omega_v$ and $F_0 - \omega_v$ turn out to be significantly smaller than for F_0, with the result being that only one photon in 10^9 is inelastically scattered.

2. Biological Theory

The following is a review of biological theory as it pertains to our experimental study of the Raman spectra of cancerous and noncancerous tissue types. As such, the theory below, while being quite elementary, provides insight into the spectra and resulting differences we see in our samples[2].

2.1 DNA and RNA

DNA is a long polymer consisting of four basic subunits. Its basic nature seems quite simple, yet it allows for all the information needed for a cell to formulate the myriad of proteins necessary for survival. DNA is two stranded, and is in the form of a double helix. These strands consist of the four subunits (nucleotides) as well as sugar phosphate molecules. These phosphate molecules form the backbone of the DNA chain, with one phosphate molecule being attached to the next one through a covalent phosphodiester bond. Attached to each phosphate molecule is also one of the four basic nucleotides – adenine, cytosine, guanine, and thymine. These nucleotides form a bond with only *one* other type of nucleotide – guanine only binds with cytosine (and vice versa), and adenine only bonds with thymine. This fact, that of each nucleotide bonding uniquely to one other type, is the most important aspect of the structure of DNA; it is the key that allows for accurate DNA transcription and replication.

RNA, also known as ribonucleic acid, has a very similar structure to DNA, and consists of four base nucleotides and sugar phosphate molecules. The phosphate molecules again form the backbone of a long chain, with one nucleotide molecule again attached to each phosphate molecule. RNA differs from DNA in that it is only a single stranded helix, not the double-strand structure

---

[2] The biological theory presented in this chapter can be found in any first year university textbook.

found in DNA. The nucleotides in RNA are the same as in DNA, save for thymine, which is replaced by the nucleotide uracil. Figure 2.1 Shows the Raman Spectra for both DNA and RNA; the two are very similar.

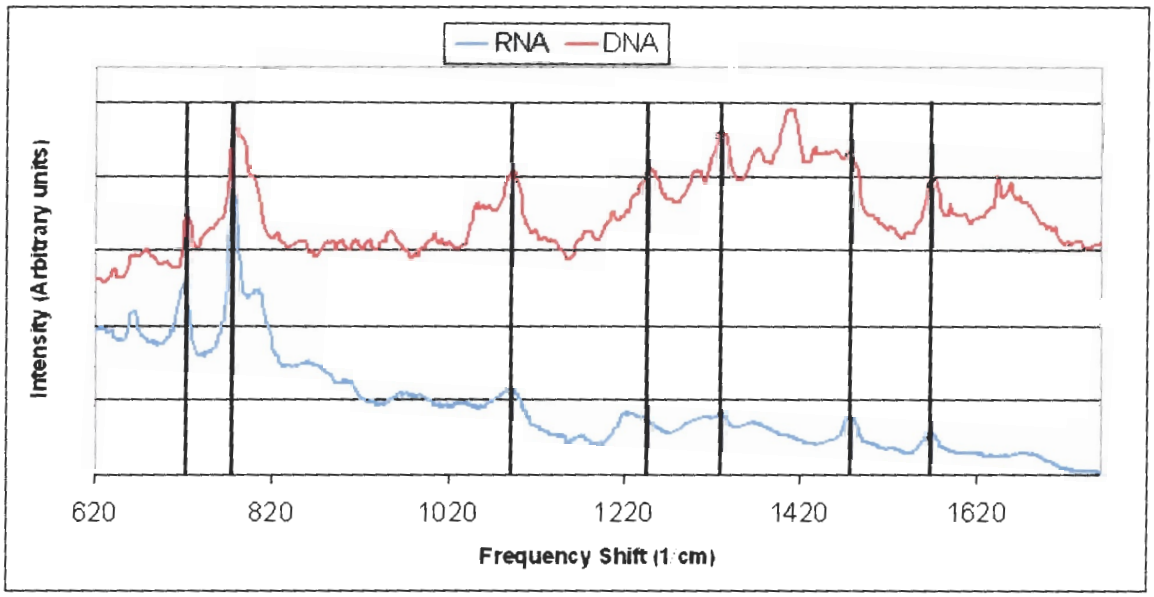

Figure 2.1: diagram of the Raman spectra of both DNA and RNA. Note the location of important peaks for both, they are equivalent. Also note the two spectra have differing backgrounds, and the DNA spectrum does show a few peaks resulting from chemicals used during processing. The largest difference seen between RNA and DNA is based on relative intensities of the given peaks found; RNA will have some peaks with differing intensities than DNA when relative intensities between given peaks are studied.

2.2 Amino Acids

In the studying of tissue and cells through the use of Raman spectroscopy, one finds that a large portion of intense spectral peaks are a result of vibrations of the molecules that make up amino acids (or the vibrations of the amino acid as a whole). Amino acids are molecules comprised of a carboxyl group and an amino group (technically the amino acids used by the cell are called alpha amino acids, where the carboxyl and amino groups are attached to the same carbon; we will just stick to calling these amino acids), and are used by

cells to create proteins. The actual geometry of the amino acids can be seen in figure 2.2. In this case, R is the side string of elements or molecules pertaining to each given amino acid. For example, in tyrosine the side chain is C_6H_7O (in the form of a ring).

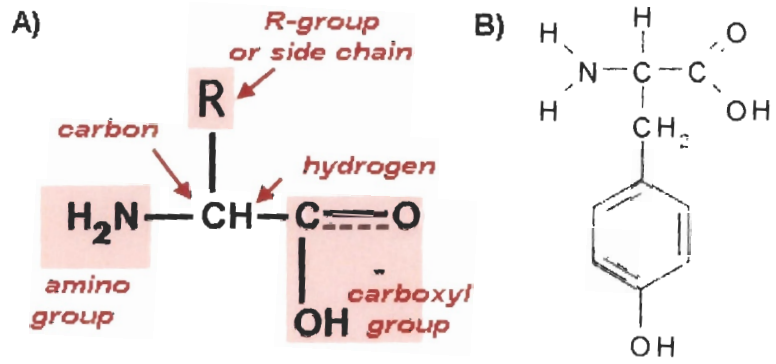

Figure 2.2: diagram of the standard form of an amino acid (A), and a diagram of tyrosine (B). Note the ring like structure in tyrosine; this gives rise to a breathing vibrational Raman mode at 853 cm^{-1} [9]. Source: [10], by permission.

Unsurprisingly, differing amino acids will provide different vibrational frequencies, such as the C-C twisting mode that can be seen. The frequency shift due to this mode in tyrosine occurs at 645 cm^{-1}, while in phenylalanine this mode occurs at 623 cm^{-1} [9]. Tyrosine's chemical formula is $C_9H_{11}NO_3$, while phenylalanine is $C_9H_{11}NO_2$. It is interesting to see how the presence of a single oxygen molecule can create such a difference in the frequency of a given vibrational mode (due to the aforementioned coupling of the vibrating bonds of the given molecule). Of the numerous amino acids, the ones that we see prominently as important contributors to a Raman spectrum (in biological tissue anyway) are phenylalanine (figure 2.3), tyrosine, and tryptophan.

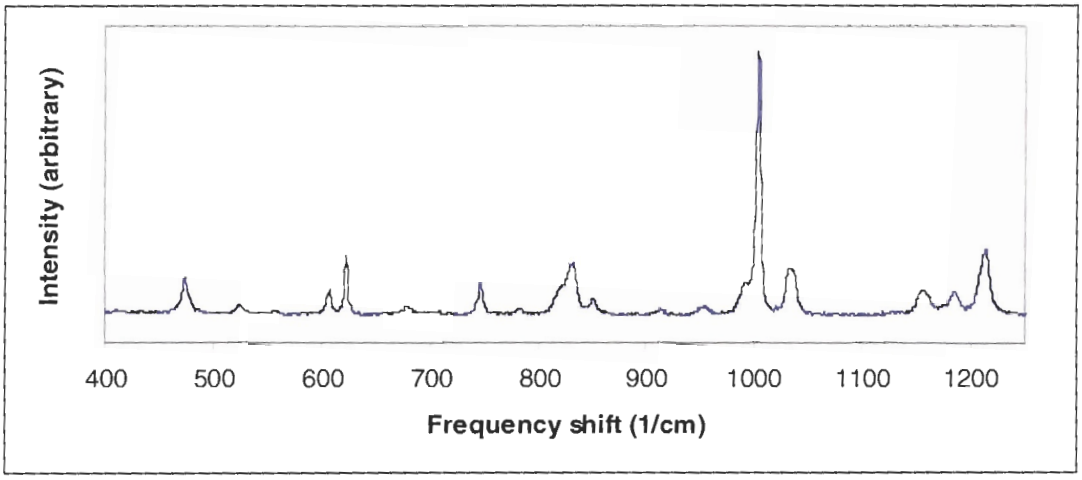

Figure 2.3: plot of the Raman intensity versus frequency shift of the incident laser light for phenylalanine. Note the large peak at approx 1000 cm^{-1}; this is due to phenylalanine's ring breathing vibrational mode.

2.3 Proteins

Proteins are as fundamental as DNA in the existence of living cells, and perform almost all the tasks required to promote cell life and cell division. Raman spectroscopy has the ability to identify some individual proteins, and can look at protein concentrations as a means to attempt to provide information.

Through the use of messenger RNA and T-RNA the 20 standard amino acids are linked together to form proteins. The amino acids are attached to each other through peptide bonds (bonds between the carbon atom found in the carboxyl group of one amino acid and the nitrogen atom of another amino acid) to form a protein chain. The backbone of this chain is comprised of carbon, nitrogen, and hydrogen, while the side chains end up being the key to the three dimensional structure of the protein, as they determine how the protein basically "folds in" on itself.

Protein Structures

The *primary structure* of a given protein is the actual sequence of amino acids in any part of the protein chain. The primary structure of an amino acid chain can be found by following the backbone of the protein and sequencing the amino side chains in simple linear order.

The *secondary structure* of a protein is defined as highly patterned and localized sub-structures that occur due to hydrogen bonding. This hydrogen bonding occurs between the C-O and N-H groups found in the polypeptide bonds of the protein backbone. These hydrogen bonds can result in the formation of a geometric pattern or shape forming between two parts of the protein chain. The secondary structure is localized on the chain - more than one type of secondary structure can occur in a given protein (usually there are a lot).

Two common secondary structures that are seen in proteins are the alpha helix and beta sheets (the beta sheet structure is seen in figure 2.4). Both geometric structures give rise to molecular vibrations of differing frequencies that can be identified through Raman spectroscopy. The interval of 1231 cm^{-1} to 1284 cm^{-1}, for example, has peaks that correspond to vibrations that arise out of the secondary structures of sheets and helices [9]. While these structures occur only in proteins, allowing us to identify a protein concentration through their Raman peak, they are common enough that they don't normally allow one to identify what exact protein is being studied.

The *tertiary structure* is the general shape of the entire protein, or in other words, the actually 3 dimensional relationship between the various secondary structures in a given single protein.

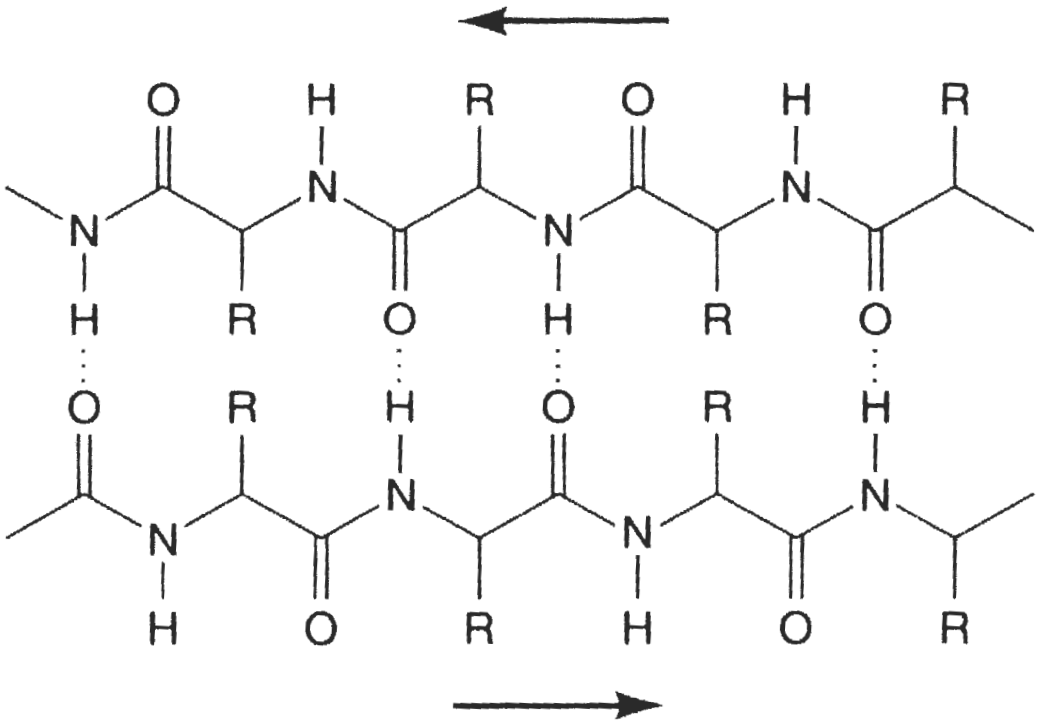

Figure 2.4: diagram of the beta sheet secondary structure. The resultant bonds creating the secondary structure give rise to prominent peaks from 1231 cm^{-1} to 1284 cm^{-1}. Source: [11], by permission.

2.4 Lipids

The term lipid covers a wide, diverse range of hydrocarbon based chemicals that have one major item in common – they are predominantly nonpolar, or hydrophobic. In other words, these chemicals are not very soluble in polar solvents, most notably water. Many lipids, though not all, also consist of chemical constituents that are also hydrophilic (polar) that do like to interact with polar solvents. Many lipids consist of significant chains of hydrocarbon molecules, such as chains of CH_2 molecules.

Various vibrational modes of some lipid-based compounds usually give quite strong Raman peaks, usually due to the long chains of hydrocarbon molecules that are found in said lipids. These chains allow for numerous

vibrations of the same mode (and frequency), giving intensely scattered Raman light at the corresponding frequency shift. CH_2 twisting vibrations, for example, occur as a result of chains of CH_2 seen in say a lipid bilayer, and give an excellent Raman peak at a frequency shift of 1301 cm^{-1}[9].

2.5 Cell Division

The basics of cell division are naturally important topics to review when talking about cancer. Firstly, cancer occurs as a result of mutations to DNA when the cell itself divides. Secondly, one of a cancerous cell's inherent traits is the ability of the cell to divide autonomously, something that doesn't usually occur in multicellular organisms.

Cell Division Cycle

Cell division is comprised of four main stages, in the following order: gap1, synthesis, gap2, and mitosis. In the gap1 stage, the cell produces RNA and proteins, and grows in size. At the end of this stage, there is a checkpoint to ensure everything is set to go for the next stage, synthesis. It is during the synthesis or S stage that DNA is replicated in preparation for the splitting of the cell into two daughter cells. The S stage is then followed by the gap2 stage, where the cell continues to grow and new proteins are produced for use in cell division. Again, at the end of this stage there is another checkpoint to ensure everything is in order before the next stage occurs. The final stage is mitosis, where the actual division process occurs. In this stage the entire focus is on cell division, and new protein production and cell growth are halted. Mitosis itself is broken into a few distinct phases, with another important checkpoint occurring partway through to again ensure the cell division process is going as planned. The entire mitosis phase is usually quite short, only lasting one to two hours.

Cell Division Checks and Balances

The cell not only has genes and proteins that are used to promote cell growth and cell division, but it also has genes and corresponding proteins used to check cell division, and to halt the process if a given stage is incomplete or if damage has occurred. These genes check to see if all the chromosomes are aligned on the mitotic spindle before mitosis occurs, DNA replication is complete, etc. P53 for example, is a protein that can halt the cell from progressing into the S phase if DNA damage is evident. P53 is normally inactive, though it becomes active in the presence of DNA damage sensed by checkpoints in the cycle of the cell, and in the presence of some DNA damaging chemicals and UV radiation (and even some carcinogens). P53, once active, initiates the transcription of several genes, including ones that work to inhibit cyclin complexes, preventing the G1 to S transition. In this way it stops cell division until DNA damage can be repaired. P53 itself also activates proteins responsible for DNA damage repair, and if needed, can initiate premature cell death if the DNA damage is too extensive. P53 is in fact quite vital to prevention of DNA mutation when cell division occurs, and as a result is a major protein in preventing cancer. In fact, if the P53 gene itself becomes mutated such that the P53 protein ceases to function properly, the rate of DNA mutation greatly increases, and the disease known as cancer becomes much more prevalent. It is no surprise then that P53 mutations are noted in more than 50% of all cases of cancer.

2.6 Connective Tissue

The relevance in discussing the nature of connective tissue lies in the fact that, upon studying our cancerous samples, it appears that the prominent region where our tissue becomes invaded by cancerous cells is in fact a region of connective tissue. Consequently, a brief discussion concerning this type of tissue is in order.

Connective tissue is predominantly made up of an extracellular matrix, comprised mainly of collagen fibers (figure 2.5). These fibers are quite strong, made up of subunits of coiled strands that provide collagen with a lot of strength. Collagen is also made up predominantly of the amino acids glycine and hydroxyproline.

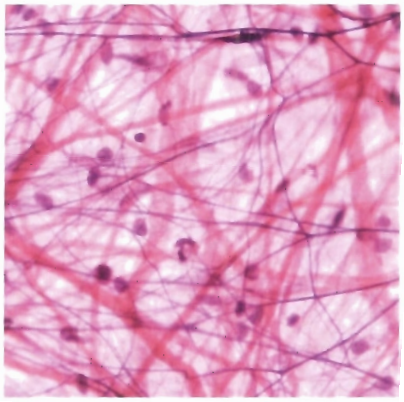

Figure 2.5: Loose connective tissue (100X magnification). The thick tubes seen are collagen tubules, while the thinner ones are elastin. The dark spots correspond to stained fibroblast cells.

Because of the rather homogeneous nature of collagen, it has a very strong Raman signature, with peaks corresponding to various amino acid and secondary structure protein vibrations (such as the amide-3 helix). In effect, identifying collagen concentrations using Raman spectroscopy is quite easy (see figure 2.2).

Connective tissue can also be made up of elastin, smaller fibres that provide elasticity to the connective tissue region. In conjunction with collagen, they provide tissue with tensile strength and resiliency. Elastin has a Raman spectrum dissimilar to that of collagen, so the two can be easily identified. Elastin does, however, have a spectrum that has only subtle differences when compared to some other types of proteins, such as actin. Fibroblasts, cells that form connective tissue, naturally appear with regularity in connective tissue. Their role

is to create collagen and elastin as needed, as well as other chemicals that are found in the extracellular matrix.

In effect then, we see that connective tissue is usually quite homogeneous, consisting of predominantly fibrous tissue with cells embedded with regular density. Its Raman spectrum, therefore, should also show little variation in its signature.

3. Cancerous Tissue

Cancerous cells (otherwise known as neoplastic cells) are the potential result of damage done to the DNA based chromosomes and genes found in the nuclei of a given type of cell. For example, basal cell carcinoma is the result of damage done in some way to basal cells, resulting in the new cell acting in an abnormal manner, as defined below. This unusual manner may result in a variety of changes in how the cell works/interacts with the surrounding tissue, though in the case of cancerous tissue a well-defined set of guidelines exist that must be met before the tissue or cell is considered cancerous or neoplastic. These guidelines will be listed below[3].

An important point to note is that the underlying mechanisms between a normal cell and a cancerous cell is still pretty much the same- it's only really the control and procedures undertaken by the genes that have changed. A loose analogy would be say a computer software virus; the hardware that makes up the computer is still the same, rather the software and resulting controls affected by the software are the items that have been altered.

The following will provide the basic theory behind the disease of cancer and how it progresses, as well as the notable differences between a neoplastic cell and its surrounding tissue, as compared to the workings of normal cells and tissue. The differences discussed will focus on changes that are vital to the workings of a neoplastic cell, as well as changes that may be viewable through the use of Raman spectroscopy.

---

[3] Again, the theory present in this chapter is quite common, and can be found in numerous books, such as L.M. Franks, N.M. Teich, *Introduction to Cellular and Molecular Biology of Cancer*, Oxford Press, 1997.

3.1 Defining a Cancerous State or Disease

According to the American Cancer Society, cancer is defined as "a group of diseases characterized by uncontrolled growth and spread of abnormal cells." The following must occur for a cancerous condition to be applied:

1) the cells must gain autonomous control over their ability to divide,

2) the occurrence of angiogenesis, the ability to form vesicles and other biological constituents to allow the tumour to continue to survive,

3) The ability of the neoplastic cells to invade surrounding tissue, and the ability of the neoplastic cells to metastasize, or move throughout the body and form new proliferations of tumours in regions not adjacent to the original tumour.

3.2 Causes of Cancerous Cells

There are numerous carcinogens (cancer causing agents) that can result in the transformation of a cell from a normal operating mode into a cancerous mode. Exposure to certain chemicals or materials in high concentrations can result in cancer, as can certain viruses. As well, radiation (such as X-rays) can affect changes and damage DNA, resulting in the mutations required to form a neoplastic cell. Insofar as something can cause the DNA of a given type of cell to become damaged and/or mutated, that material has the possibility of causing cancer. Indeed, the actual hydrolytic interaction between water inside the nucleus and DNA itself admits a small probability of mutation in any cell. Usually the cell is able to identify and either repair or stop the mutation, but in the case of eventual cancer cells this is obviously not the case.

3.3 The Development of Cancer as a Disease

It is important to note that not all of the above steps must occur at the initial onset of the transformation of the cell into a precancerous cell upon DNA mutation (by precancerous we mean a cell that is moving through the process to potentially become cancerous). It is entirely possible (and usually more probable) for the eventual neoplastic cell to possibly gain the ability to divide on its own (autonomous control), and then later for the resultant benign growth to gain the ability to become invasive and/or metastasize.

In carcinomas, it is thought to take 4 to 5 serious changes to gene expression for the resultant condition to be deemed cancerous (as outlined above). In a sense then, going from a normal cell to a malignant tumour can be seen as sort of an evolutionary process, with the "correct" genetic alterations being those that maximize (or at the very least allow) the tumour to perform as outlined above. Other mutations that do not allow for the above processes to occur will result in either insignificant alterations (at least by our above standard on what comprises cancer) or alterations that act as a detriment to that cell (and its corresponding daughter cells) from forming a cancerous condition.

If the DNA mutation is such that certain genes that govern DNA duplication have been damaged, it's quite possible that further mutations will occur at a much more rapid pace than seen in normal cellular growth. This increase in mutation rate will then allow for a larger possibility that a percentage of the precancerous cells formed can then mutate into cells allowing for the possibility of angiogenesis and mestastatic growth, among other things. Indeed, studies of cancerous tumours have shown that the cells that make up the tumour are quite heterogeneous and show a varied difference in structure and behaviour. This provides evidence then that, even though these cells are thought

to come from a singular initial cell, a significant amount of mutations have occurred that allow for the variance seen in the tumour cells.

3.4 Autonomous Control over Cell Division

There are a few ways that an individual cell can gain the ability to divide under its own accord, based on the mutated DNA that now controls its functions. With respect to uncontrolled growth and division of cancer cells, it is common to find that at least one (usually more or all) of seven types of proteins have undergone a mutation in some form, protein types that affect cell division in some regard or other. These protein types include growth factor proteins, growth factor receptor proteins, signal proteins, transcription factors, DNA-repair enzymes, cell-cycle control proteins, and apoptis regulating proteins. The possibility of mutation resulting in uncontrolled division are broken into two main classifications, namely mutations that result in an activation or over activation of some process that is normally inactive or strictly controlled (and is used during cell division), or mutations that serve to inactivate or disable proteins that regulate cell division in some manner. In other words, proteins involved in cell division are usually over expressed or mutated so that they provide positive signals for cell divisions, while proteins involved in preventing cell division in some manner are usually mutated so that they become ineffective. The genes that produce these cell division initiating proteins are also known as proto-oncogenes (oncogenes if they mutate into a form that increases the malignancy of a cell), while the genes that produce the latter proteins are known as tumor suppressor genes.

3.5 Angiogenesis

Angiogenesis is the ability of the body to form new blood vessels from existing vessels, based on the demand of cells in the localized areas. With respect to tumours, the ability of the tumour to initiate angiogenesis will result in chemical changes to the tissue surrounding the tumour.

Tumours can normally reach only a size of about 1-2 mm in diameter before being unable to sustain larger growth due to a lack of nutrients. To achieve a larger size, the tumour must have access to nutrients and the ability to remove waste products from both the cells and surrounding tissue. To facilitate these requirements, the tumour needs access to the transportation infrastructure of the body, namely the circulatory system. It does this by creating new blood vessels from existing ones, creating a system of capillary vessels to further support the tumour. These vessels are created as a result of the release of growth factors that affect endothelial cells (cells that form the walls of blood vessels). The most important of these growth factors is VEGF, or vascular endothelial growth factor. VEGF is created in the cell, emitted into the surrounding extracellular tissue, and moves through the body until it reaches the endothelial cells, at which point it interacts with growth factor receptors on the membrane. This sets up a chain reaction, telling the endothelial cells to proliferate and create chemicals used to create new blood vessels. VEGF tells the endothelial cell to produce enzymes called metalloproteinases, proteins used to break down the extracellular matrix materials found in the spaces between cells. Once this matrix has been broken down, the endothelial cells can move into the region and proliferate, eventually forming new blood vessels that attach to existing vessels. The resulting difference of the chemical composition of this altered tissue has the potential to be identified by Raman spectroscopy, especially in cases where the chemical compositions are quite different due to very increased concentrations of proteins normally found in low concentrations).

3.6 Invasion and Metastasis

A single primary tumour, alone in a multi-cellular body, is usually not fatal. It is the eventual ability of the tumour to spread throughout the body and disrupt vital processes that cause cancer to be deadly. To facilitate this spreading through the body to form new tumours (otherwise known as metastasis), the cells need to have the ability to invade through tissue and to have increased motility.

Tissue invasion is necessary to allow the neoplastic cell the ability to gain access to either the lymphatic system or the circulatory system. Without the ability to degrade the extracellular matrix found in tissue outside cells, and the ability to slip through the basal lamina (a thin membrane of type 4 collagen that surrounds endothelial cells making up the walls of blood vessels, as well as surrounding the vessels of the lymphatic system), neoplastic cells have no hope of spreading throughout the body. To destroy the extracellular matrix (which, again, consists of collagens, gelatins, laminins, etc.), the cell must emit collagenases, gelatinases (otherwise known as matrix metalloproteinases), and other enzymes that, under normal circumstances, it does not emit. As such, this again entails a required genetic mutation to either initiate proteins or processes that activate these enzymes, or to mutate genes that result in the inhibition of these processes. The resultant degradation of the matrix grants the neoplastic cell the initial ability to leave the primary tumour, enter a circulatory system and move to another region.

While the extracellular matrix must be degraded to allow for the passage of neoplastic cells, it can not be totally destroyed, as the matrix also provides traction for the movement of the cells; it is the matrix that the cell uses to move through tissue, grasping and pulling itself along. Consequently, for metastasis to occur, a balance of degraded and still existing matrix components must exist.

Matrix destroying enzymes and the resultant affect they have on surrounding tissue will constitute a chemical and molecular change, as compared

to the usual tissue composition that would exist in normal tissue. Both the increased proliferation of metalloproteinases and their destructive effect on the ECM components of collagen, etc., result in potential differences that may be visible through Raman spectroscopy, and allow us another tool by which to clinically test for the existence of a cancerous tumour. As well, it may be the case that the surface distribution and concentration of integrins seen on neoplastic cells will result in viewable differences. It may even be the case that the differences seen will allow us enough detail to find out if a given tumour has reached a stage of metastasis yet, or is still in a benign, noncancerous state.

3.7 Notable Differences

Neoplastic cells also have a few other notable changes that occur, changes that while not necessarily affecting the cell's ability to promote angiogenesis, metastasis, etc., are important from a viewpoint concerned with noting differences between cancerous and normal tissue.

A very significant difference is the existence of a condition known as anueploidy that quite often exists in neoplastic and precancerous cells [12]. This is the condition where the cell nucleus contains an abnormal number of chromosomes (either more or less than the regular number of 2), resulting in an altered amount of DNA content. It is usually the case that the cell will contain more chromosomes, resulting in higher DNA concentration. This concentration not only results in darker staining when using dyes that bond to nucleic acids, but should also be viewable through Raman spectroscopy as higher DNA peaks in the given spectra. The ratio of nucleus to cytoplasm may also change in a cancerous cell, as can both the shape of the nucleus and cell itself, usually forming irregular shapes, which again should be visible using Raman spectroscopy to create tissue images.

Cancer cells also lose what is known as contact inhibition, a condition that prevents cell density from becoming high. In normal conditions, when the cells achieve a certain density, contact inhibition prevents further proliferation of cells. Neoplastic and precancerous cells do not adhere to this, and will continue proliferation, even in locations with high cell density (see figure 3.1). Again, this provides another means of verification of the existence of a cancerous condition, as a higher density of cells will result in a differing Raman spectrum, one where chemical concentrations attributed to cells (instead of outlying tissue) dominate the spectra.

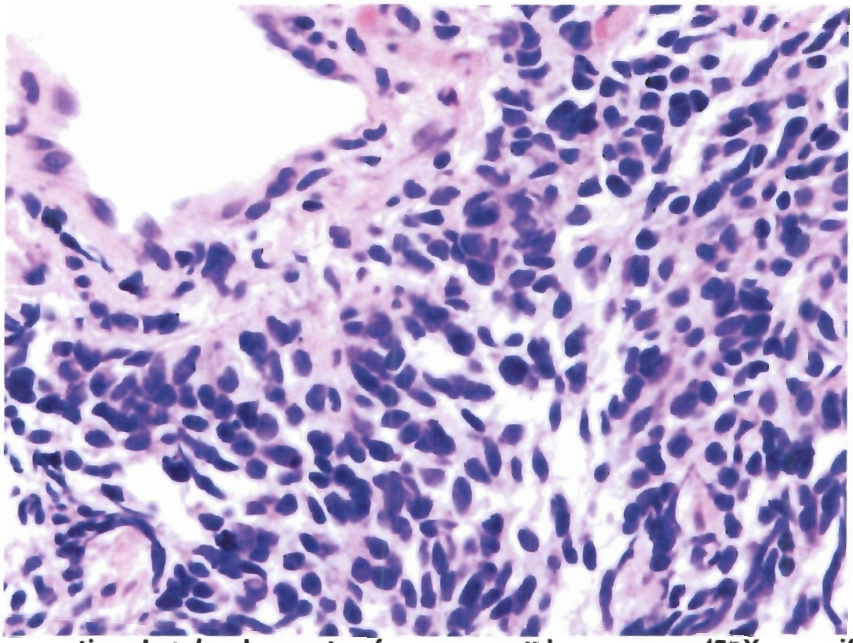

Figure 3.1: cross-sectional stained sample of a non-small lung cancer (50X magnification). Note the high, disorganized density of cells inside the main tumour region, as compared to the tissue regions where the tumor has yet to fully infiltrate. As well, note the irregular shapes of the cell nuclei (stained dark blue); they are not generally elliptical in shape. Source: [13], by permission.

3.8 Body's Reaction to Cancer

Chemical compositional changes also result from the body's reaction to a cancerous condition. The body normally does not sit idly by while cancer develops, but rather works to prevent it. However, for the body's immune system to begin attacking tumour cells, it must first recognize the cells as being altered

from normal tissue (e.g. said cells must look like invasive cells). The main defense mechanism used by the body against tumour cells are T cells, a type of lymphocyte cell. T cells normally watch to attempt to detect the presence of invasive cells, doing so by studying cell surface proteins called Major Histocompatibility Complex (MHC) class 1 proteins. These are proteins with structures unique to a given individual, allowing for the immune system to ascertain whether a cell is foreign or not. If the cell is normal, then the T cell takes no action, if the cell is abnormal, the T cell will emit chemicals in an attempt to kill the cell. It is apparent that cancer cells can show MHC proteins signifying an abnormality (due to DNA mutations), which can result in the T cell attacking that cell. However, as is often the case, there are also usually cancerous cells that do not show significant levels of these abnormal MHC proteins (recall cancer tumours normally have cells with differing mutations), which results in the T cell not attacking said cell. Consequently you have T cell lymphocytes present and attacking cancer cells, but not attacking all the cancer cells. Of course, an unfortunate occurrence here then is that cancer cells that do not show these abnormal MHC proteins become selected to continue to exist and reproduce, so to speak, making it harder for the T cells to continue to destroy a given tumour.

Once a lymphocyte recognizes a problem, it will quickly undergo repeated divisions, thus greatly increasing the proliferation of lymphocytes. These lymphocytes then continue to attack abnormal cells and reproduce until such a time as the threat has appeared to be concluded. As such, in areas of high concentration of abnormal cells, one can expect to find a significant concentration of lymphocytes (insofar as said lymphocytes can identify said abnormal cells).

Both the lymphocytes themselves, and the chemicals they produce, may provide information to our attempt to notice differences in normal tissue and cancerous tissue through the use of Raman spectroscopy. The lymphocytes themselves are normally very small in size yet contain high concentrations of nucleic acids. Thus, if a large number of lymphocytes are present in a tumour, a

Raman image showing numerous smaller cells with high nucleic acid concentrations may be used as a means of identifying a cancerous region.

3.9 Chapter Summary

As we've seen, significant changes must occur for a cell to become neoplastic and result in cancer. These changes, though subtle, result in potentially large chemical changes, not only for the cell but also the tissue surrounding a cancerous tumour. Increased concentrations of proteins, altered structure of the nucleus and cell itself, destruction of tissue outside the cell, increased capillary concentration, etc., all have the ability to be detectable through the use of Raman spectroscopy. Raman images based on lipid and membrane concentrations can show us not only the shapes of cells, but also the density of cells, letting us know if something is out of the ordinary. Collagen and other extracellular matrix components may show significantly lower concentrations in cancerous tissue, while concentrations of chemicals signifying increased cell density may be much higher (e.g. DNA/RNA concentrations), something we can again pick up. As well, the aneuploidal nature of neoplastic cells may give us signatures with higher DNA content than we would normally expect. These variations should allow one to discern, with Raman spectroscopy, not only the existence of a cancerous tumour, but also the ability to note the boundary of that tumour's influence.

4. Apparatus Detail and Empirical Concerns

In this section we will give a brief overview concerning the main components of the micro-Raman spectrometer. Particularly, we will focus on the constituents that provide the key processes to the Raman spectrometer.

4.1 Raman Spectrometer

One of the key elements of the spectrometer is the confocal light microscope, allowing one to gather scattered light directly from a desired spot of the given sample while ignoring light scattered from other regions of tissue that are in the path of the laser beam. Laser light passes through the objective, is incident upon the sample, and returns back through the objective up to the confocal aperture. White light used for visual identification is also reflected from (or transmitted through) the sample and into the same objective (obviously not at the same time as the objective is capturing Raman light naturally). The confocal aspect of the microscope works by having gathered light incident upon a pinhole diaphragm, a blocking piece of material with a small pinhole in the middle. This pinhole is lined up such that only photons gained from the axial region of tissue at the focal point of the microscope is in focus at the pinhole; light outside of this region is not in focus, and does not hit the blocking diaphragm at the pinhole (see figure 4.1). Naturally, light from any region traveling directly along the optic axis of the objective lens will pass through the pinhole, but this is a mere fraction when compared to the total amount of light that is subtended by the objective lens. In a typical measurement, the pinhole is set such that the axial depth resolution gives a 5 micron full width half maximum; in other words the captured scattered light from points 2.5 microns away from the focal point is half of that captured from points directly in focus. The pinhole also allows for slightly better lateral resolution as well, again due to the pinhole nature's ability to block out photons not on a correct geometrical path. In terms of lateral resolution, the

sampled region is an approximate 0.5 micron dot, though vibrations and motion of the spectrometer suggest the limit of resolution is about 1 micron.

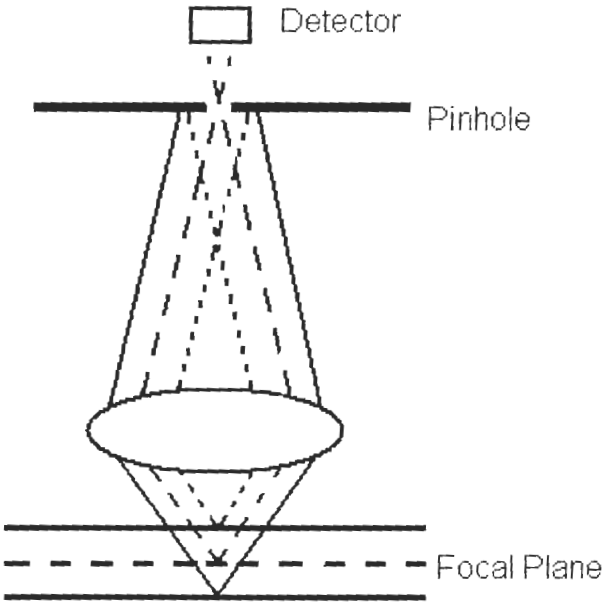

Figure 4.1: diagram of a confocal pinhole. The pinhole acts to only allow light through from areas found on the focal plane; light rays from regions not in the focal plane do not hit the pinhole diaphragm at the location of the pinhole (they hit to the right or left of it), and are subsequently stopped from continuing along the optical path.

Before hitting the confocal diaphragm the collected light is first incident upon a notch filter, used to remove as much of the initial laser light beam as possible (recall Rayleigh scattered light is orders of magnitudes more intense than Raman scattered light). This allows us to better see the frequency shifted scattered radiation. The notch filter works through the use of thin multiple layers of materials with differing indices of refraction that transmit and reflect a given percentage of the incident light. This sets up an interference pattern for a given wavelength of light, and if the thickness of the layers and the angle of incidence of the light are correct can result in almost complete destructive interference in a desired path. In our case, the filter is chosen so that our Rayleigh scattered laser

light will be fully rejected by the notch filter (said laser light feeling full destructive interference along the path of transmission), while our Raman scattered light is allowed to pass through without a change in beam path. The notch filter attenuates the Rayleigh scattered laser light by about 6 orders of magnitude, while attenuating the remainder of the shifted scattered light by a neglible amount. Even with the notch filter the Rayleigh scattered laser light presents a huge peak when compared to the peaks seen as a result of Raman scattering, and consequently any peaks with a low enough frequency shift become impossible to see, being swamped by the Rayleigh scattered light.

Once the collected scattered light passes through the notch filter, it is then incident upon a diffraction grating. The diffraction grating's purpose is to allow us to study the intensity of the scattered light as a function of frequency, allowing us to formulate peaks in our measured data. The diffraction grating causes interference to take place, with constructively interfered light being emitted from the grating at angles corresponding to the given wavelength of the light

$$D(\sin\theta + \sin\beta) = M\lambda, \qquad (4.1)$$

Where D is the distance between grating spaces, θ is the angle of incidence made with the normal of the diffraction grating, β is the angle of diffraction, M is an integer corresponding to the given order of the bright spots, and λ is the wavelength. In the case of our spectrometer, the first order bright spot is used to provide information; remaining orders are ignored.

After the grating, the scattered light, no longer traveling along the same path as a result of the grating, travels through geometric optics until it arrives at a cooled (approx 220 Kelvin) CCD camera. The cooling acts to reduce dark noise that can cause random photonic activity to interact with the CCD camera. Dark noise arises from simple thermal effects, such as an increase in random photons (e.g. photons emitted from the CCD camera housing hitting said CCD camera).

Dark current, an effect where temperatures give rise to thermally generated electrons that cause random noise in the measurement of a given pixel, is also reduced by a reduction of temperature. Cooling the CCD camera by a 20 degree interval acts to reduce noise approximately tenfold. The laser light hits the CCD camera along a thin line of pixels 3 pixels wide and about 997 (27 pixels are lost due to noise and are not used) pixels across, giving us 997 separate bins to collect the laser light. The scattered light hits the camera at locations based on the path of the said light; thus light scattered at different angles, denoting different frequencies, hit the CCD camera at different locations along the line (thereby hitting different CCD pixels). The CCD camera collects light over a given amount of time, allowing us to study the intensity of the scattered light (by collecting more photons, with the photon count directly correlated to light intensity as a function of frequency). This allows us a measuring tool to study the intensity of the light as a function of frequency, turning our photons into electronic signals, with pixels collecting higher amounts of light resulting in a higher stored current (eventually resulting in a higher voltage that is used to quantify the initial radiation intensity). Our scattered light intensity pattern has now been transformed into a digital signal, with each pixel on the CCD camera correlating to a given frequency range (of finite width Δf), and the current/voltage correlated with that pixel giving us the intensity of the scattered light within the frequency range Δf.

It should be noted that the range Δf of each pixel is not constant, but actually varies over the CCD camera. This is merely due to the geometry of the setup; light incident at differing pixels on the CCD camera comes from light traveling along different path lengths (as a result of the angular dispersion that took place at the diffraction grating), and hits the CCD camera at different angles. As a result, the range of frequencies hitting a given pixel will vary. A given cross-sectional area near to the CCD camera will encapsulate a given frequency range; as you travel farther along the same path said frequency range will now occupy a larger cross-sectional area as light of differing frequency diverges. As such,

pixels that require a longer travel distance will encapsulate a smaller frequency range. Figure 4.3 shows the basic layout of the equipment and path traveled by the laser as it travels through the spectrometer.

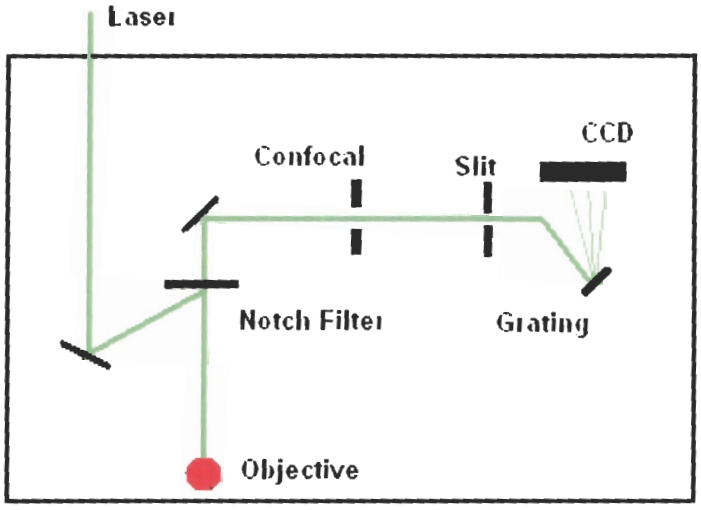

Figure 4.2: diagram of the important aspects of the Raman spectrometer used in the experiment. The laser initially hits the notch filter, is reflected to the objective, whereupon the scattered light returns along the same path. Once hitting the notch filter again, the laser light is discarded while the Raman scattered light passes through. After this, the light is incident upon a confocal pinhole, passes through the entrance slit, hits the grating, and finally arrives at the CCD detector.

The spectrometer software then comes into play, allowing us to define the given frequency ranges seen on the CCD camera. By comparing the user defined initial laser frequency the software designates the position of a given pixel (or pixels, depending on resolution) with its corresponding frequency shift, based on the diffraction grating used. Calibration is required to allow for the correct positioning of the initial laser beam, and the software labels each pixel range with the corresponding frequency shift (based on the first order peaks from the scattered light). This frequency shift then provides us with an axis to compare intensities on a graphical analysis of the frequency versus intensity of the incident scattered laser light.

4.2 Fluorescence

Unfortunately, it's not only Raman scattering that is incident upon the CCD camera, we also have fluorescence to cope with. Fluorescence is photon emission due to molecules dropping from electronically excited states to electronic states with lower energy. The given molecule is usually initially excited to some higher electronic energy state, with vibrational and rotational energy levels also being higher than their respective ground states. Through collisions and interactions with other molecules, said molecule loses energy until its vibrational and rotational energies are at their ground states of the respective excited electronic state. It usually takes about 10^{-12} to 10^{-14} seconds for a molecule to decay to its given ground vibrational state [14]. The molecule, still existing in the excited electronic state (but with vibrational and rotational states at ground for this excited electronic state), then decays into its ground electronic state, decaying into a variety of vibrational and rotational energy states of said electronic ground state (it usually takes around 10^{-8} to 10^{-10} seconds for the decay from an excited electronic state to a ground electronic state [14]). This decay results in fluorescence, emitted photonic radiation that occurs in the visible spectrum of light for outer electron transitions. Since our measurement is studying light with a scattered frequency of V_O - V_f, where V_O is in the visible spectrum and V_f is small in comparison, we see that noise from fluorescence is potentially significant.

Fluorescence shows itself in our Raman spectrum not as a peak but rather as a simple increase in intensity, normally showing up as an uneven background upon which Raman peaks are seen. This is due to the fact that fluorescence results in peaks that are quite broad in comparison to Raman peaks; Raman peaks are on the order of 10 – 100 cm^{-1} wide, while fluorescent peaks can be on the order of 1000 cm^{-1} wide or larger.

Fluorescence not only comes from the tissue sample, but also the substrate the sample is on (much like Raman scattered light from the substrate). While the confocal microscope acts to prevent this, fluorescence from the substrate can also be incident upon the CCD camera, due again to the fact that the Raman signal from the tissue sample is quite weak, and the small amount of fluorescent radiation that makes it through the confocal hole can be comparatively large. The noise in a micro-Raman is, however, significantly less than in a macro-Raman setup, where the Raman sampling volume is much smaller than the volume that contributes to the fluorescence spectrum. The result here is that in macro-Raman spectra, the Raman peaks can be very easily washed out by the much more intense fluorescent signal; in micro-Raman spectra Raman peaks can be much more easily distinguished.

If desired, this background noise can be removed by a few methods, to allow for a more accurate viewing of the Raman scattering peaks themselves and to more accurately identify the actual intensity of each Raman peak. The spectrometer software allows a few methods for performing this background subtraction, the most useful method being the use of a multi-order polynomial to define the background, which can then be subtracted. It is important to note, however, that the background is quite broad and slowly varying, and does not usually affect the ability to see the individual Raman peaks; it only tends to make it more difficult to directly ascertain accurate Raman peak intensities.

4.3 Instrumental Resolution and Spectral Calibration

The resolution of our spectrometer (the distance that two peaks must be separated by before one can accurately discern them as independent peaks) is predominantly controlled by the width of the entrance slit, and the actual quality of dispersion of the diffraction grating (the change in the angle of diffraction with respect to the change in the frequency of the incoming light, $d\beta/d\lambda$). The latter is

unchangeable for our setup, but the former can be altered. A larger slit width corresponds to more light reaching the grating, giving a more intense signal, but also corresponding to worse resolution (increasing the amount of stray light that hits the diffraction grating at an angle say $\theta + d\theta$ instead of θ). Hence, a middle ground is usually desired, a slit width that allows a decent amount of light through but doesn't overly affect spectral resolution.

Spectral calibration (calibrating the software so that the measured frequency shift is correct) is usually done by sampling a material with a well-defined Raman peak. In our case silicon was used, as it has a very prominent and sharp peak at 520.7 cm^{-1}. As such, using the software we positioned the frequency shift so that this peak was set to 520.7 cm^{-1}. The software then recalibrates the frequency shift axis accordingly to give the correct frequency shift range (as given in wave numbers, cm^{-1}).

4.4 Raman Imaging

As we eventually found, it was impossible to ascertain small scale visual detail on our unstained samples. This made it extremely difficult, if not impossible, to be 100% sure of what we were sampling for our Raman measurements. Fortunately, the spectrometer allows one to perform Raman imaging, where the spectrometer can collect spectra from much larger sample areas, one micron dot at a time. Not only does this allow for a more accurate idea of what we are sampling (by allowing us to later compare our sampled regions with stained images of the same region), but it also allows us to investigate the distributions of different kinds of molecules and their variation for differing tissue types. To perform this Raman imaging, one merely has to specify the area size, the length of sampling time for each single spectra, and let the machine run. The software then automatically moves the sample table by displacement increments as defined by the user. Depending on the time and area desired, total spectral gathering could take anywhere from minutes, upwards to hours or even days. As

an example, a 50 by 50 micron area, sampled for 30 seconds at each point (2500 points in all), would take 1250 minutes, or 20 hours, 50 minutes. What this method loses in time, it makes up for by allowing the controller the ability to leave and sleep, keeping the machine running over night. It also allows for the ability to form very accurate Raman spectral images of the tissue, allowing us to study details on the order of a few microns and up. One can also average over a number of the individual Raman spectra taken, allowing one to gain the average Raman spectra over a given area, such as say a 10 by 10 micron square. All in all, the image technique was an essential tool in our experimental process, allowing for the possibility to study larger areas in a decent amount of time (which further allowed for the identification of important details not easily seen, and as a whole significantly more data to work).

The images themselves are formulated by specifying a given wave number interval of the Raman spectrum. The interval used is usually centred on a prominent Raman peak, with the ends of the intervals coinciding with the intersection of the peak with the background spectrum. As such, our images then convey the intensity difference as a function of map location for only a single given peak. Ideally by imaging this peak we can gain insight into the concentration and distribution of the molecule that gives rise to this studied Raman peak. This should allow us to note the location of cell nuclei, increased concentrations of say fatty acids and the existence of collagen (and potentially degraded collagen). Figure 4.3 is an example of Raman imaging, showing an image based on the intensity of a peak seen at 1376 cm^{-1} [9], a peak corresponding to nucleic acids.

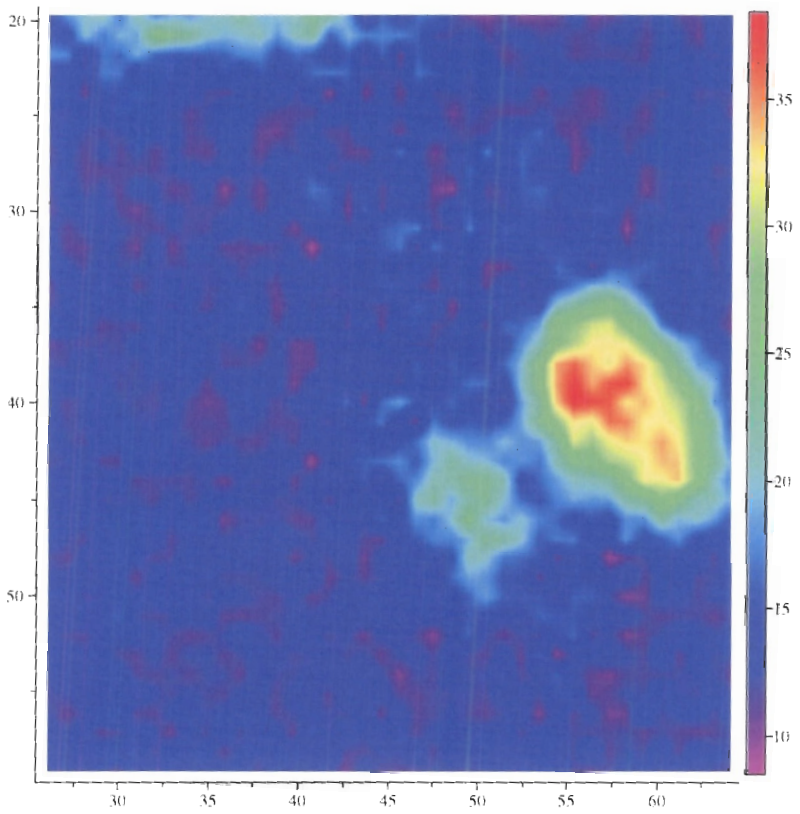

Figure 4.3: Raman image of a 40 by 70 micron area, taken based on the Raman peak seen at 1376 cm^{-1}. The region shows a low intensity peak for everywhere save in the centre - right quadrant; the more intense signal here, and the shape of this more intense region, suggests the presence of a cell nucleus. The lower and left hand axes indicate position in microns, the right hand axis indicates intensity.

5. Experimental Details

5.1 Tissue Sample Preparation

In our experiment, the desire was study tissue that could be left as unaltered as possible, to prevent the Raman spectra from being possibly contaminated by chemical processes (such as the use of paraffin to maintain tissue structure and avoid tissue damage while cutting the tissue). As a result, we opted to have tumour cultures of lewis lung carcinoma injected into lab mice. These cultures, once matured, formed large-scale tumours that would not only be of significant size to allow for easier study, but would also give the tumours time to affect surrounding tissue and time for the organism to initiate a response in the form of lymphocytes. By using an already existing cancer culture we can assume that said culture has all the necessary traits of cancer; the ability to metastasize, the ability to induce vascularisation, etc. In our experiment the injection point was in a connective tissue layer, under the hypodermis layer of the skin. While lewis lung carcinoma is formed initially from epithelial lung cells, the injection of the cancer culture into the connective layer still allows for an excellent comparison based on the type of tissue studied. Lung tissue has a connective tissue component, one that is arguably quite similar to the connective tissue layer we studied. Once lewis lung carcinoma begins spreading it is not uncommon for it to reach and invade the connective tissue layer found in lungs, or to other regions of connective tissue in the body [15]. We can also hypothesise that some of the differences seen in our Raman spectra due to the main characteristics of many carcinomas, such as high cell density, can be attributed to all carcinomas, and not just lewis lung carcinoma. Consequently, for us to study and compare the connective tissue of a cancerous induced sample to that of a normal tissue sample seems quite relevant, regardless of fact that we're studying connective tissue below the dermis of the skin layer, as opposed to the connective tissue found in a lung.

Once the affected tissue area was removed from the mouse, the samples were stored at minus 80 degrees for a few weeks, until a cryostat cutter could be used. The cryostat cutter (kept at a temperature of around minus 20 Celsius) was used to slice the tissue into 10 micron thick samples, which were the placed onto slides. OCT (a glue-like chemical that freezes solid at below freezing temperatures) was used to aid in the cutting by providing a much more stable and thick "brick" to cut, as opposed to just the piece of tissue itself. It is important to note that OCT was the only other chemical used on the tissue; no other chemicals were used on the samples marked for Raman measurements. OCT itself resembles a glue-like substance at room temperatures, but hardens solidly at lower temperatures (e.g. below zero).

After the samples were cut half of the samples were left as is, while the other half (alternating samples) were stained to allow for a visual view of large-scale structure, and to see how the tissue fared through the entire growing/cutting process. These stained samples were not used to take Raman measurements; rather they were only used as a visual aid for the unstained samples, where detail was significantly harder to make out. Staining was done using hematoxylin (stains nucleic acids such as DNA/RNA [16]). One important item that did need to be verified after cutting was the condition of the tissue samples; we needed to verify what type of damage, if any, was done to our samples during the entire freezing/cutting process; this will be discussed later.

5.2 Experimental Procedure

Unstained samples were irradiated with a 515 nanometer Argon Ion laser, with a stated initial power output of 200 milliwatts. Laser light was incident upon various optical mirrors, used to direct the beam into the spectrometer. Upon entering the spectrometer the beam passes through a simple neutral density filter to allow for a further reduction in beam intensity, if desired. In our case, 50 and 75 percent attenuation filters were used to reduce intensity to avoid possibly

burning the sample and ruining our spectral measurement. Actual power measured at the sample site with an initially listed 200 milliwatts of power from the laser was seen to be 6 milliwatts at the sample, spread over a 0.5 micron diameter dot through a 100x magnifying objective. The 50% filter was predominantly used, reducing our power to 3 milliwatts over the one micron dot. On occasion, however, we used the 75% attenuation filter in the few cases where the sample area burned at 50% (it appears some tissue types or regions were more capable of withstanding higher laser intensity before burning than other regions). In the cases that the 75% filter was used, the intensity measured was later increased by a multiple of 2 to compensate for the decrease in overall power from the laser, so that we could more accurately compare our spectra.

Once the sample was irradiated, scattered light was collected back through the microscope objective, passed through the notch filter, and was incident on the diffraction grating. In our case the diffraction grating used was 1800 lines per centimetre, giving us an angular difference in the path direction of light based on its frequency. In our experiments we sampled the region between 406 cm^{-1} to 1937 cm^{-1}, which based on the 515 nm lasing frequency of the argon laser set each pixel to correspond to an approximate wavenumber range of 1.6 cm^{-1}. This gives us an idea of the potential resolution limit of our measured spectra and how much detail we can gain. Again, recall that the spectral width of the pixels vary due to the variation in path length that results due to the angular dispersion of the diffraction grating. In our case, at 406 cm^{-1} each pixel had a wavenumber range of 1.69 cm^{-1}, while at 1937 cm^{-1} each pixel had a wavenumber range of 1.40 cm^{-1}.

The CCD camera was set to collect photons for a 30 second duration, with the results of the number of counts of photons at each pixel being output to the computer. At this point a shutter would very briefly inhibit further collection of light, basically re-initializing the camera and allowing it to again count photons from the required starting point (the data then being output to the computer).

6. White Light Images

The following chapter takes a descriptive look at the white light images taken of both our cancerous and control samples. These images were used to ascertain the location of any regions of connective tissue highly invaded by neoplastic cells, and to look at visual differences between the two samples. As well, by studying these samples visually we were able to get a general idea of how much damage the tissue underwent as a result of the cutting and preparation process. Again, recall that these samples were not prepared with alcohol and paraffin, and were frozen containing water, resulting in the potential for damage (such as the bursting of cellular membranes due to the phase transition water undergoes during freezing [17]).

Figure 6.1 depicts a cross-sectional image of both our cancerous sample and our control sample. If studied closely, there are significant differences between the two; in the cancerous sample, there appears to be significantly more tiny spots denoting cells (though they are tough to make out at this magnification). This invasion of cells also seems to have an affect on certain tissue layers, as the skin layers seen in the control sample are much well more defined (and look quite darker). The cancerous sample shows significant less hair as well (the very dark lines seen in the skin layer, protruding outward from the tissue), probably a result of the invaded cancerous cells affecting the tissue.

The region we studied, the connective tissue layer below the skin layer, shows large visual differences, most notably the fact that the cancerous connective tissue layer appears to have a more 'bumpy' texture, while the healthy connective tissue appears more homogeneous (save for the dark spots, which are most likely contaminants).

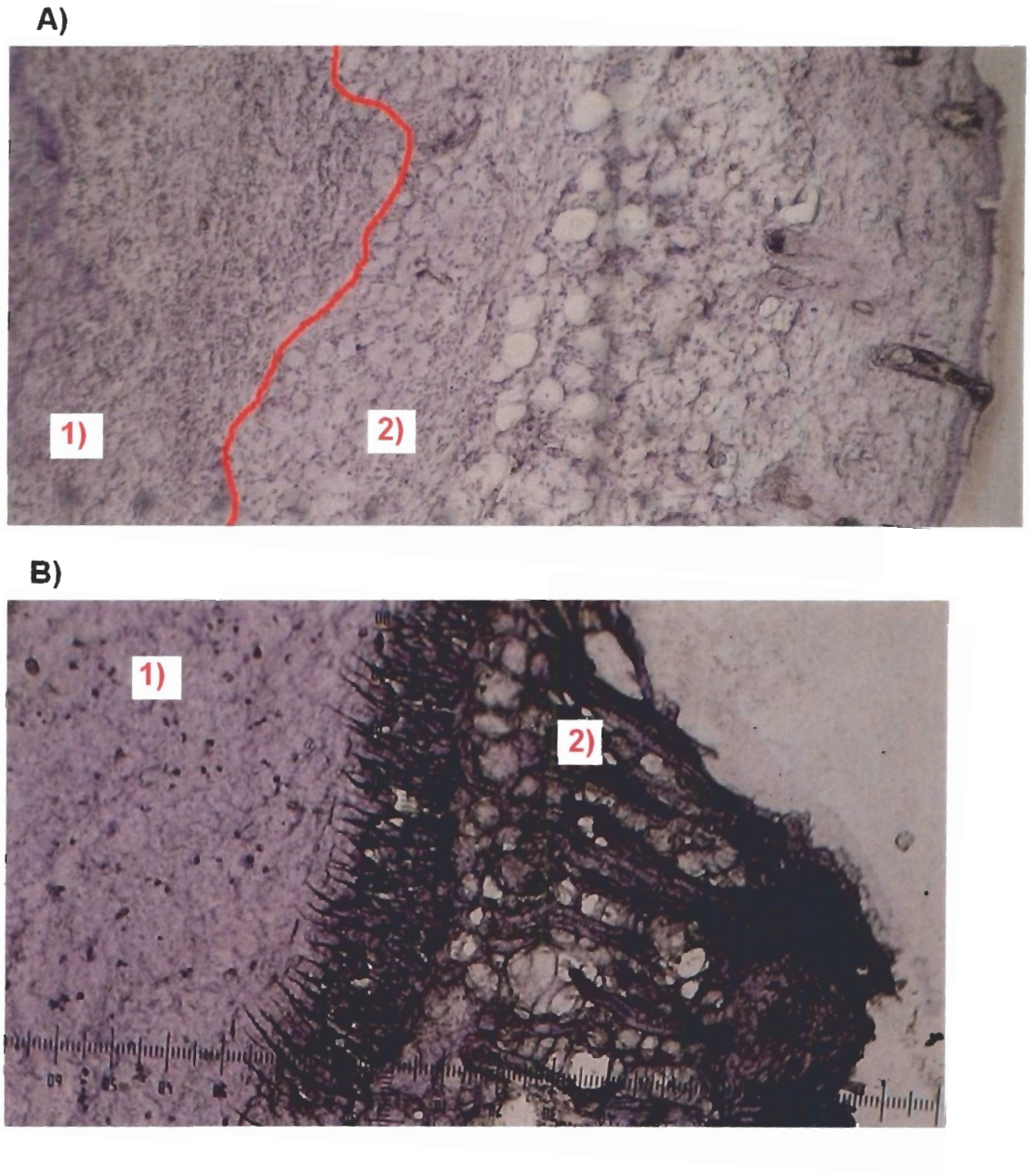

Figure 6.1: white light images of both our cancerous sample (A) and our control sample (B). Region 1 depicts the connective tissue region, region 2 depicts the layers of skin, namely the epidermis and dermis (the red line in A is the approximate boundary between the two regions). Note the significant differences between the two regions; the small spots in the cancerous sample most likely depict cell nuclei, suggesting a much higher proliferation of cells in the cancerous sample. The connective tissue region in the cancerous sample appears much more textured than the same region in the control sample, suggesting significant differences between this region in the cancer and control samples.

Save for the tiny elliptical spots signifying the nuclei of cells (which again, are barely viewable at this magnification), not much more detail can be made of said cells at this power. Cell membranes can't be visually seen, nor can any inner organelles (or strands of collagen or elastin for that matter). The large scale detail, most notably the varying levels of tissue in the skin epidermis and the boundary layers, can be discerned quite well.

Figure 6.2 shows the connective tissue regions for what appears to be a highly invaded region of connective tissue from our cancerous sample and a region of connective tissue from our control sample, at high power. Again, not much detail can be seen; individual strands of collagen or elastin can't be seen, nor can any cell membranes or organelles. Indeed, if one looks at the connective tissue region from the control sample, you don't see any significant detail whatsoever. In the connective tissue of the cancerous sample, however, one can see a large concentration of black spots, again denoting regions of high nucleic acid concentration (suggesting cell nuclei). This is an important detail, and as expected for a region of tissue invaded by aggressive neoplastic cells and lymphocytes. Normal connective tissue itself does contain cells, though the density is usually quite small, resulting in our hematoxylin staining showing little to no detail (the darker regions seen in the upper right quadrant of our control sample has potential to be from nucleic acids). This image taken from our control sample shows little cell population, though images taken from other regions in the control sample do show a few cells (no where near the density seen in our cancerous sample however). Again, this provides evidence that the region we're studying contains a large concentration of neoplastic cells and potentially lymphocytes. The fact that the dots are rather well-defined suggest that the majority of nuclear membranes (at the very least) remained intact during the cutting and processing of our samples; if these membranes were to have ruptured, these spots would most likely be less well-defined.

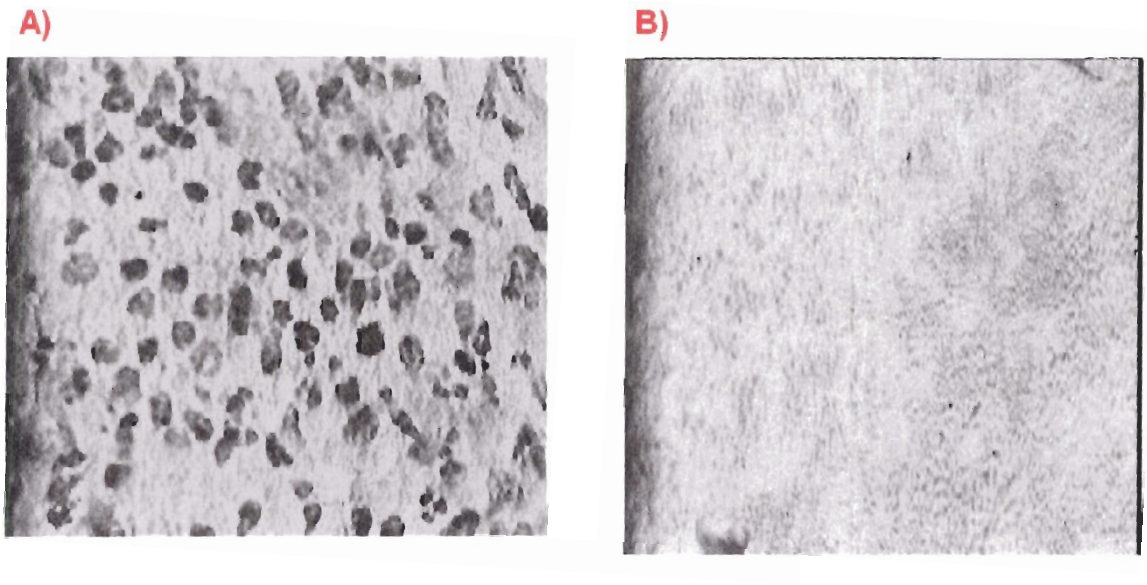

Figure 6.2: White light images of the highly invaded region of connective tissue for our cancerous sample (A) and a region of connective tissue from our normal sample. The most significant difference is the apparent large cell concentration seen in our cancerous sample, which is as expected for a region of tissue invaded by neoplastic cells. Each image is approximately 100 microns across.

Figure 6.3 shows our unstained cancerous sample, the sample used for the actual Raman measurements (the image taken is after the sample was stained, to allow for better visual detail). If one again looks at the connective tissue region, we can see a large area that appears significantly darker; this region corresponds to the same region seen in our stained sample that showed the highest concentration of cells. This would then suggest that this region in the unstained sample also contains the highest concentration of cells, and has been highly invaded by neoplastic cells or lymphocytes. The lighter regions of connective tissue in or cancerous sample would appear to be less invaded by cells. This dark region was the area of primary study for our Raman measurements, and numerous spectra were taken from here. Spectra were also taken near the boundary of this region, and farther away, where the connective tissue appeared to be less invaded and visually looked more like the connective tissue of our control sample.

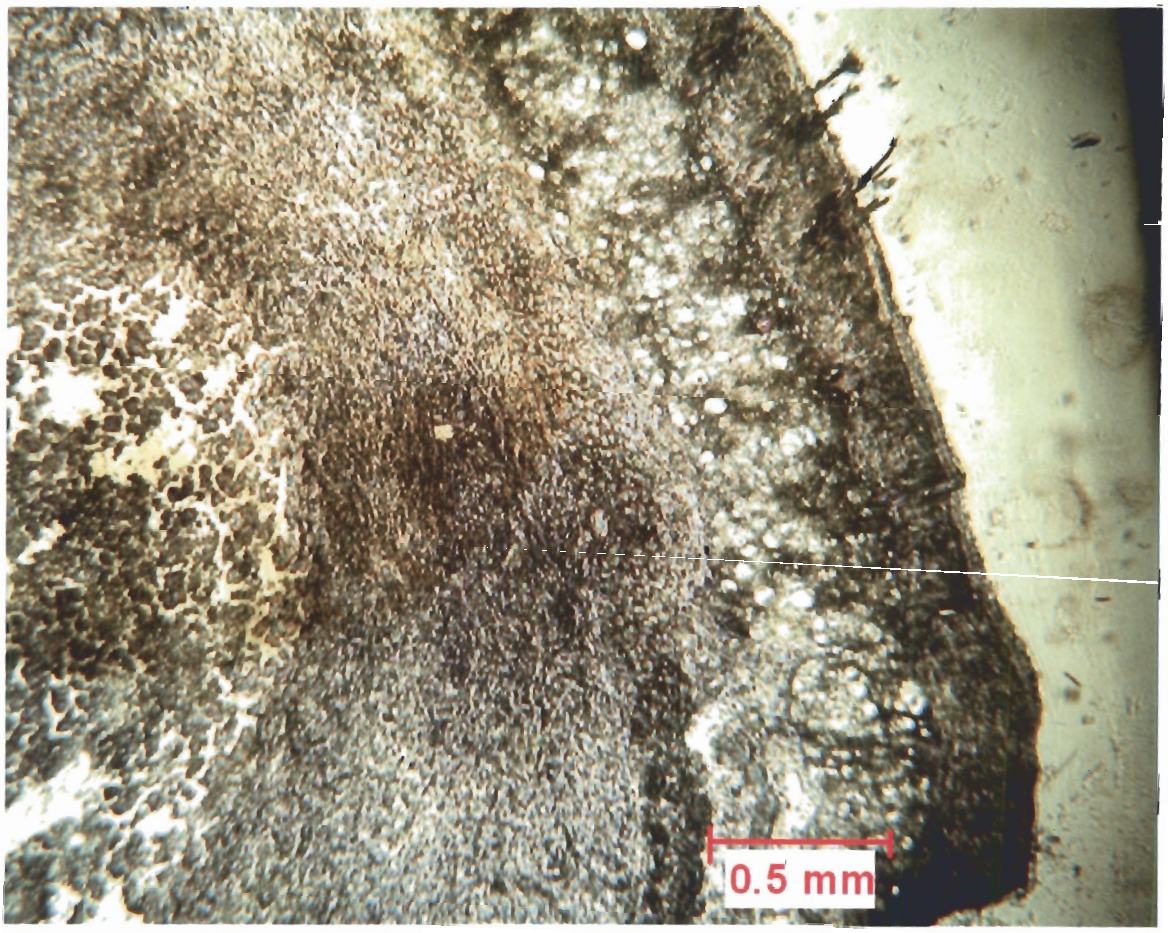

Figure 6.3: White light image of sample used for Raman measurements (originally unstained for measurement purposes, then stained to allow for more visual detail). The dark region seen in the connective tissue layer (the middle of the image) corresponds to the very highly invaded region seen in the stained sample (the little white block is a region that was burned during one attempted Raman measurement). This was the primary region of study for our Raman measurements, with spectra also taken from the apparent "boundary" of this region and connective tissue regions farther away.

After completion of Raman spectra from both the cancerous and control sample, said samples were stained to allow for more visual detail. Said staining did allow for easier identification of large scale detail (as can be seen in 6.3); however when looking at the samples at high power the amount of visual detail was unimpressive. This may be the result of the samples being out in the open air for a few weeks during measurements, resulting in reducing the effectiveness of the staining process.

7. Raman Images

The first part of the data to be studied will be the distribution of the molecular constituents in the highly invaded region seen in the cancerous sample (based on a significant increase in cell density), as compared to normal connective tissue seen in the control sample. The regions studied are loosely confirmed by looking at the corresponding stained samples; in other words, the regions studied in the unstained samples are based on the findings seen in the stained samples. In studying the images, particular attention will be paid to the homogeneity of the chemicals found (the variation in the given chemical components) throughout the images, and the differences seen between the images of normal loose connective tissue and the highly invaded region in the cancerous sample will be identified. It is hoped that these images will provide insight and usefulness in an attempt to separate and classify normal loose connective tissue from tissue that has been invaded with neoplastic cells.

We will be looking at four images, two from the cancerous sample and two from the control sample. As stated, both will be from the loose connective tissue layer seen in each sample, and both are quite indicative of the images taken from these given regions. The region studied in the cancerous sample will be from a region of connective tissue that appears to be highly invaded by the introduced cancerous cells; the region appears to have a high cell density when studying the corresponding stained samples. Images of normal loose connective tissue have been taken from the control tissue samples where no cancer culture was introduced (a few images were also taken of the connective tissue farther away from the highly invaded regions seen on the cancerous samples).

7.1 Spectral Peaks used for Imaging

Figures 7.1 and 7.2 correspond to white light images of the regions imaged (7.1 is the invaded region, 7.2 normal connective tissue); as can be seen, neither image shows significant cellular detail. Figure 7.3a shows an image (each spectra lasting 30 seconds and encompassing a 1 μm^2 region) based on the Raman peak at approximately 788 cm^{-1}, while 7.3b shows the variation of the peak at 788 cm^{-1} throughout the image (the plots formed are based on the average spectra taken from a 9 μm^2 region to reduce noise) [9][4].In particular, the peak at 788 cm^{-1}corresponds to a vibrational mode of the O-P-O string that forms the backbone of DNA and RNA strands. The image outlined in figures 7.4 (invaded region) is based on the Raman peak found at approx 1301 cm^{-1}. This rather intense peak is mostly indicative of lipid molecule and fatty acid concentrations. The peak itself is a result of a CH_2 twisting vibrational mode, a mode that is quite intense due to the long chains of CH_2 molecules found in fatty acids and lipids made up of said fatty acids. This peak does potentially have small contributions resulting from proteins, in our case however we are studying the differences in peak intensity over our image range, which appears to be due to lipid and fatty acid variation. This can be verified by looking at smaller peaks that have lipids as their only contributor (such as the peak at 1743 cm^{-1} due to carbon - oxygen ester bonds). In regions where the other peaks signifying lipid/fatty acids are nonexistent, the peak at 1301 cm^{-1} is small; in regions where these peaks are at their most intense, the peak at 1301 cm^{-1} is also at its most intense. Suffice to say then, with regard to the fact that we are studying the variation in intensity of the peak at 1301 cm^{-1} (from its lowest point to highest), it appears we can make an excellent approximation by regarding this variation as predominantly due to variation in lipid/fatty acid concentration. The two peaks at 788 cm^{-1} and 1301 cm^{-1} were also used for the corresponding Raman images of our normal connective tissue region, as we will see later.

---

[4] It should be noted that all listed Raman peak values have been cited from this given reference.

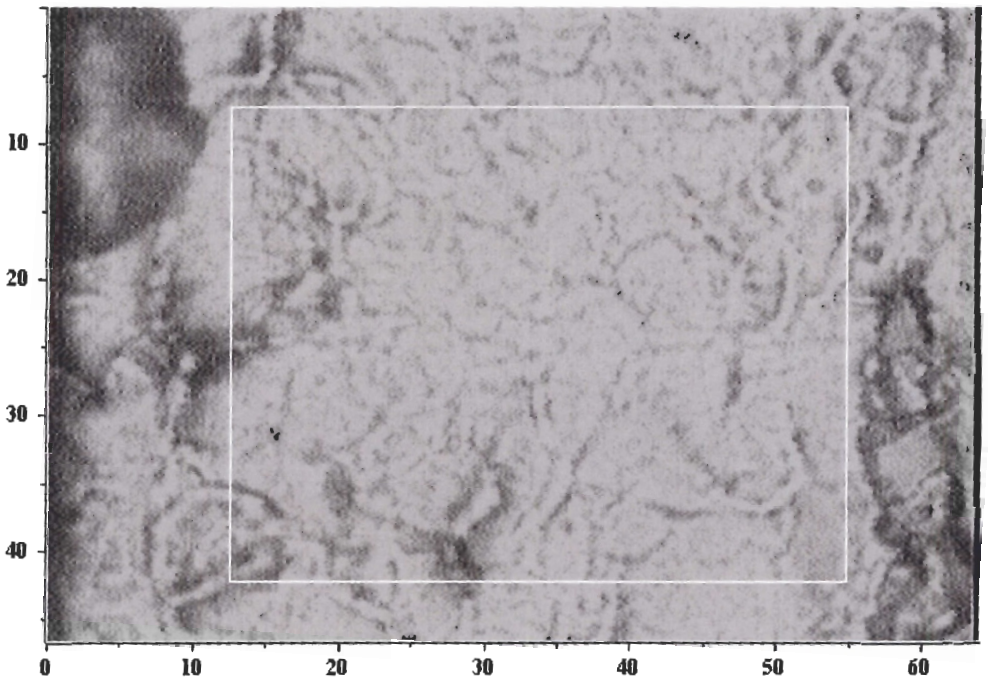

Figure 7.1: white light image of the region imaged in the invaded region of the cancerous sample (boundary of the image outlined in white). Note the lack of detail, save for a few "channel-like" structures; it is impossible to note any cell or cell nuclei boundaries. The scale is in microns.

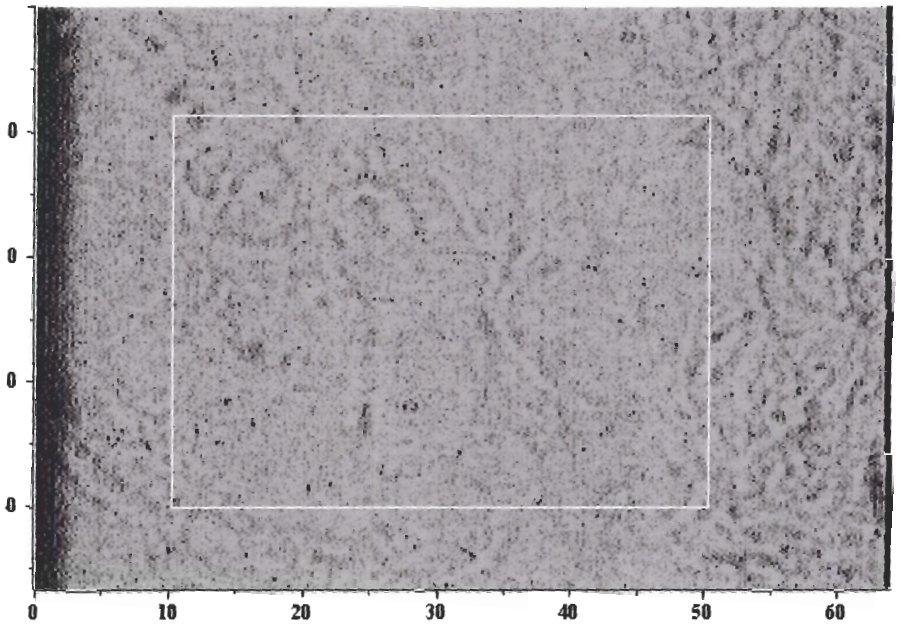

Figure 7.2: white light image of the region imaged in the connective tissue region of our control sample. Again, note the lack of detail, even more so than figure 7.1. The scale is in microns.

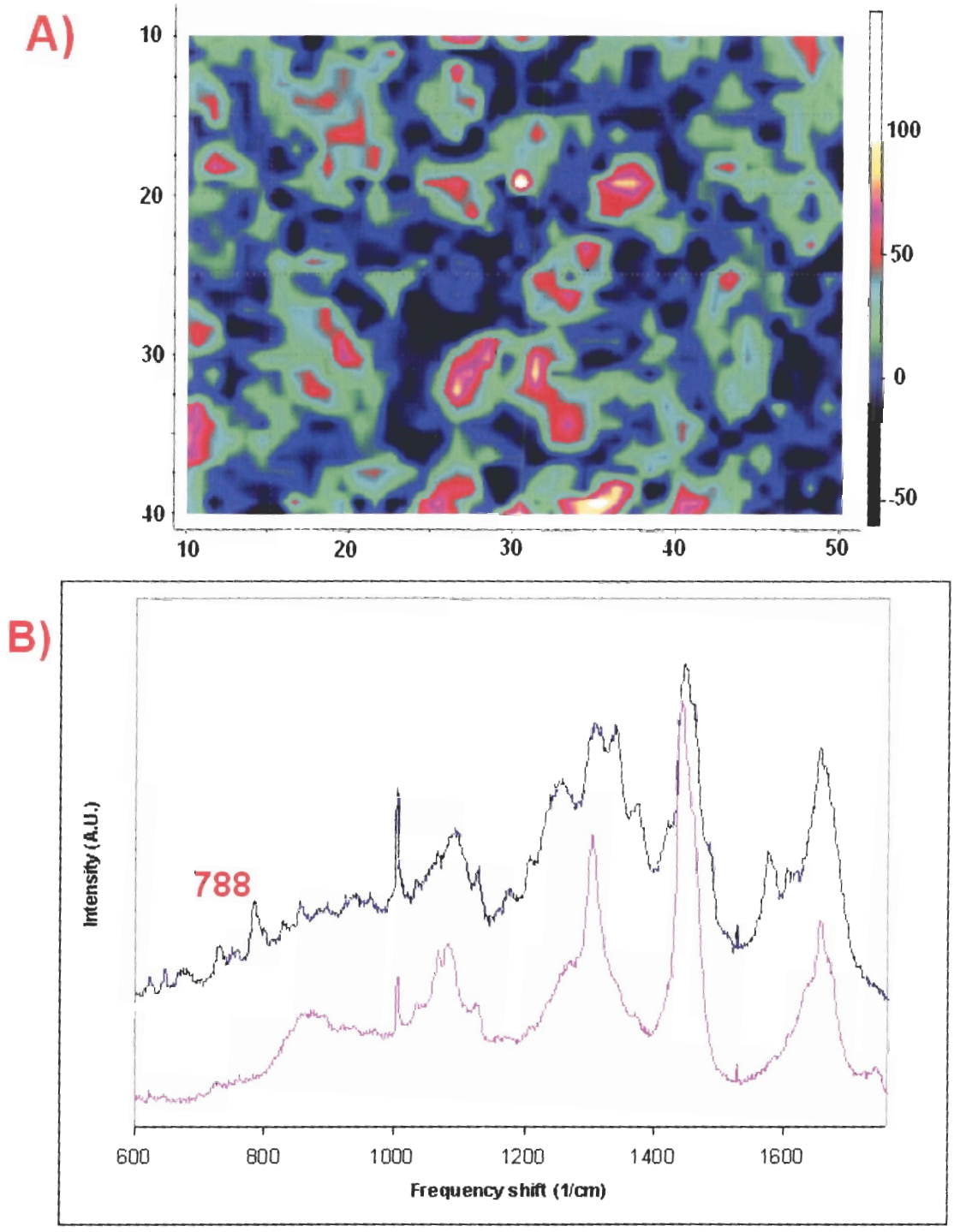

Figure 7.3: Raman image of the invaded region of the cancerous sample, based on the peak at 788 cm^{-1}(A) (corresponding to nucleic acid concentration). The left and bottom axes indicate position in microns, the right axis indicates the intensity of the Raman signal. (B) is a graph showing the variation in said peak seen in (A). Note the image is based on the peak at 788 cm^{-1}; the remaining spectra may or may not show significant variance.

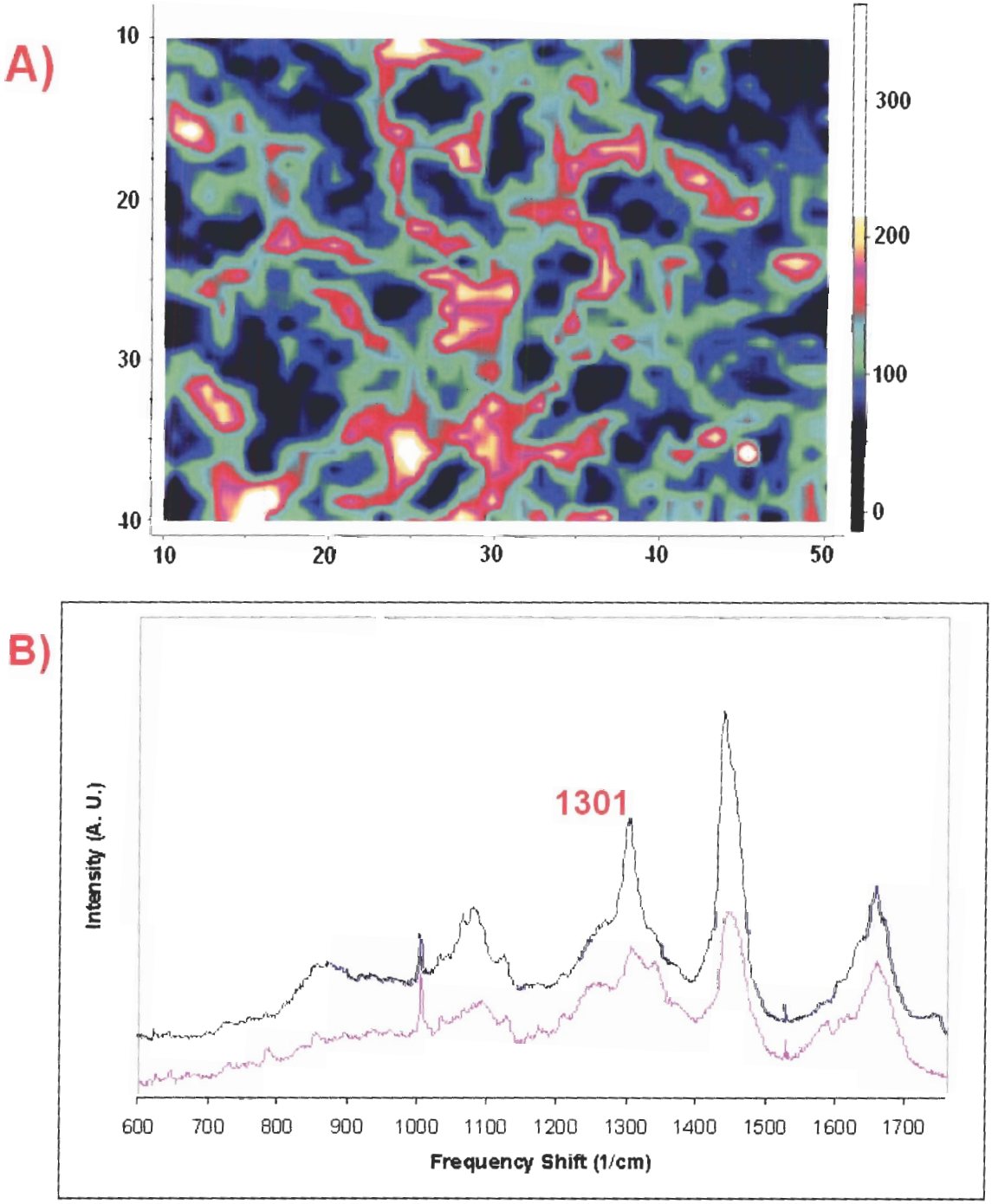

Figure 7.4: Raman image of the invaded region of the cancerous sample (axes indicate position in microns and intensity), based on the peak at 1301 cm^{-1}(A) (corresponding to lipid/fatty acid concentration). (B) shows the corresponding variance seen in said peak for (A).

7.2 Highly Invaded Region Images

Initially, images from figures 7.3 and 7.4 will be compared to study the distribution of DNA/RNA as compared to the fatty acid/lipid concentration (a direct comparison can be seen in figure 7.5). Again, recall that the DNA/RNA concentration is based on the peak seen at 788 cm^{-1}, while the lipid/fatty acid image is based on the peak seen at 1301 cm^{-1}[5]. Other peaks indicating equivalent molecular constituents could be used to form similar images (such as 1449 cm^{-1} for the lipid or 828 cm^{-1} for DNA/RNA); these peaks were studied and empirically found to give approximately equivalent images.

By studying the comparison we can see that when a given region has a high concentration of DNA/RNA it does not have a corresponding high concentration of lipids/fatty acids, and vice versa. In fact, save for a few regions seen in the images where both DNA/RNA and lipid/fatty acid concentrations are low, the two images seem almost exclusive; regions where DNA/RNA are low correspond to high lipid/fatty acid concentration, and regions where lipid/fatty acid concentrations are low correspond to high DNA/RNA concentration. The regions where both concentrations are low prove valuable, as they suggest that there isn't some negative correlation between the images resulting from some sort of software anomaly (e.g. an inability to discern the 1301 cm^{-1} peak from that of a DNA/RNA peak found close to it). These results are unsurprising, as we expect the largest concentration of lipids and fatty acids to be outside of the nucleus in the cell cytoplasm (and membranes), with the largest expected concentration of DNA/RNA to be inside the nucleus.

[5] Individual reference spectra for DNA, Collagen, Keratin and Oleic Acid (fatty acid) can be found in Appendix A.

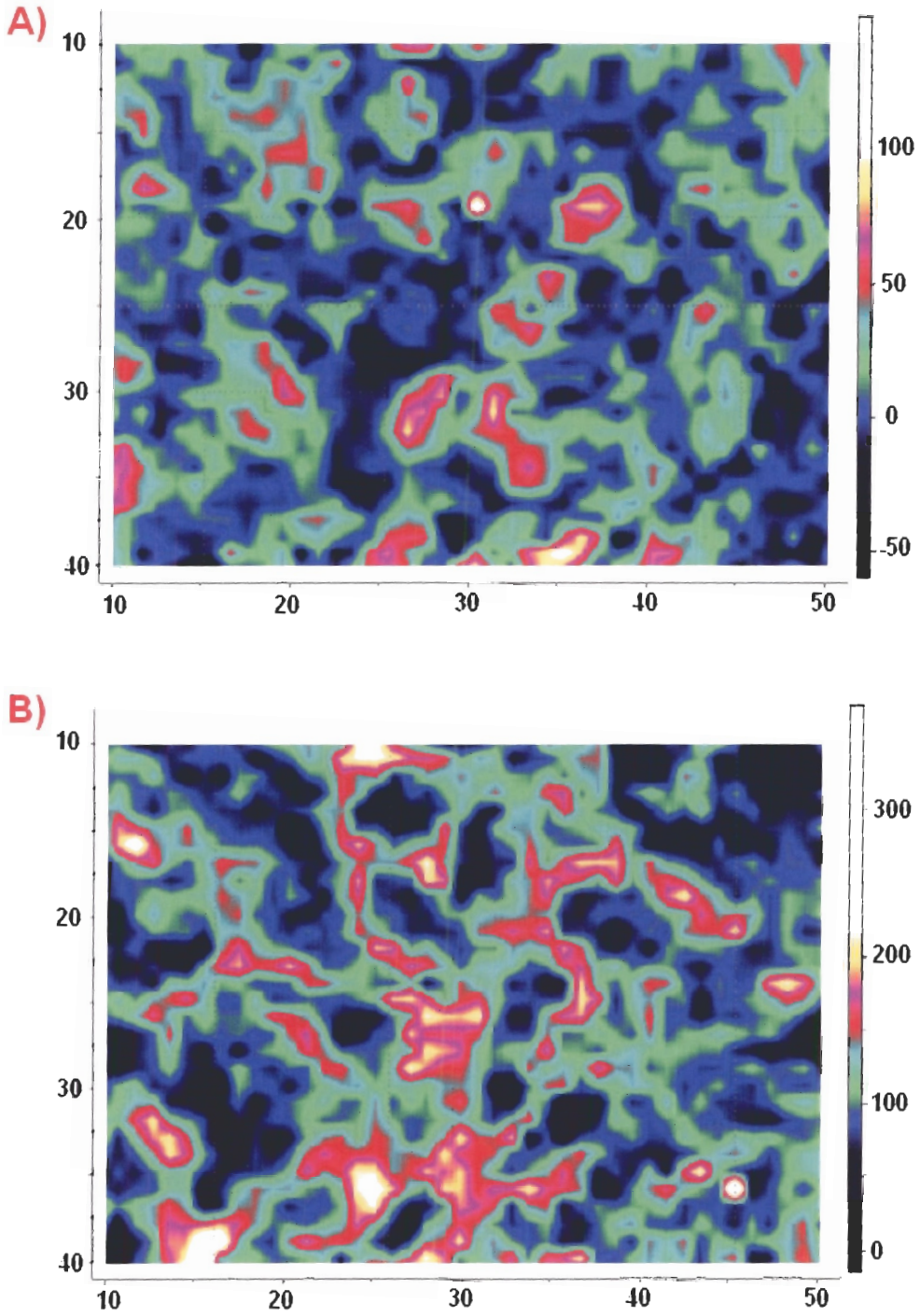

Figure 7.5: Raman images based on concentrations of nucleic acids (A) to lipids/fatty acids (B) for the highly invaded region of connective tissue in the cancerous sample. The left and bottom axes show position in microns, the right axis indicates the intensity of the Raman signal. Note the differing intensity scales to show better contrast. As can be seen, regions where we have a high concentration of lipids/fatty acids show a low concentration of nucleic acids, and vice versa.

As well, if one looks at the overall shapes given by the concentrations, we see that the DNA/RNA high concentration locations are clumped into individual "shapes" (the high concentration DNA/RNA regions show more separation), while the lipids and fatty acids are more strung out much like in a network-like distribution, taking up most of the space left over by the DNA/RNA high concentration regions. These results are as expected, as they seem to point to a heavy concentration of cells inside the sample region studied. Again, recall that when the density of cells becomes high, the cell walls become pushed together, and the cell nucleus itself takes up the majority of space inside the cell. This is what we appear to see here, a large number of oddly shaped DNA/RNA concentrations pointing to cell nuclei, with networks of lipids and fatty acids occupying the space in between said nuclei.

We should remember that there is the potential for tissue damage resulting from the cutting and freezing process undergone by the tissue. Undoubtedly there will be parts of the cell membranes that have burst, and possibly parts of the nuclear membranes as well (though the stained samples seem to suggest the nuclei remained more or less intact, as suggested by the loosely shaped elliptical nuclei seen in the stained control samples). This tissue damage may lead to the oddly shaped concentrations of high DNA/RNA. It is possible, however, that the odd shapes are also due to RNA concentrations outside of the nucleus (giving us the distribution we see), or potentially due to cutting through more than one nucleus at the given surface, giving us a few oddly shaped high concentrations of DNA/RNA (like the region seen inside the intervals $x = 35$ to 40 and $y = 30$ to 35 of figure 7.5 (a)). The possibility of cell membranes bursting is also difficult to verify from looking at the fatty acid/lipid image, as the cytoplasm of the cells seen may simply be filling up allowable space based on the location of other cells.

A majority of the high DNA/RNA spots seem quite small and highly concentrated, while some regions of higher DNA/RNA don't have quite as high a concentration, but do take up a larger area. It is possible that these smaller regions correspond to lymphocyte nuclei, while the larger regions of high DNA/RNA may correspond to the nuclei of the neoplastic cells themselves. While this is a good hypothesis, it is impossible to say with absolute certainty if this is indeed accurate. Again, one can not be 100% certain how much of the cell nuclei we are looking at; we may have a whole nucleus being sampled by the laser's penetrating depth, or we may be sampling a smaller percentage of a nuclei, or a smaller percentage of two nuclei very close together. Studying the distribution of fatty acids and lipids, however, should at the very least allow us to better hypothesise which is indeed the case. Again, for example, looking at the region of x = 35 to 40 and y = 30 to 35 of figure 7.5(a), seeing the small high concentrations of DNA/RNA and noting the thin string of lipids/fatty acids separating the two, the hypothesis that this is two lymphocytes close together would not be unreasonable.

7.3 Comparison between Images Taken from the Highly Invaded Region and Normal Connective Tissue

Figures 7.6 show the Raman images based on the peaks 788 cm^{-1} for the connective tissue region of our control sample, along with the variation seen in said peak. Figure 7.7 directly compares the nucleic acid and fatty acids/lipid concentrations for our connective tissue sample. The comparison made between the loose connective tissue and invaded tissue images shows some significant differences, not only in spectra intensity differences and differences in relative concentrations, but also in the homogeneity of molecular concentration distributions. The normal tissue image shows a much more homogeneous distribution of concentrations in terms of the DNA/RNA concentrations (also, not shown is the collagen peak intensities; they show consistent peak intensities

throughout the normal connective tissue image). Initial analysis suggests that the DNA/RNA Raman peaks are more intense and show more variation in the

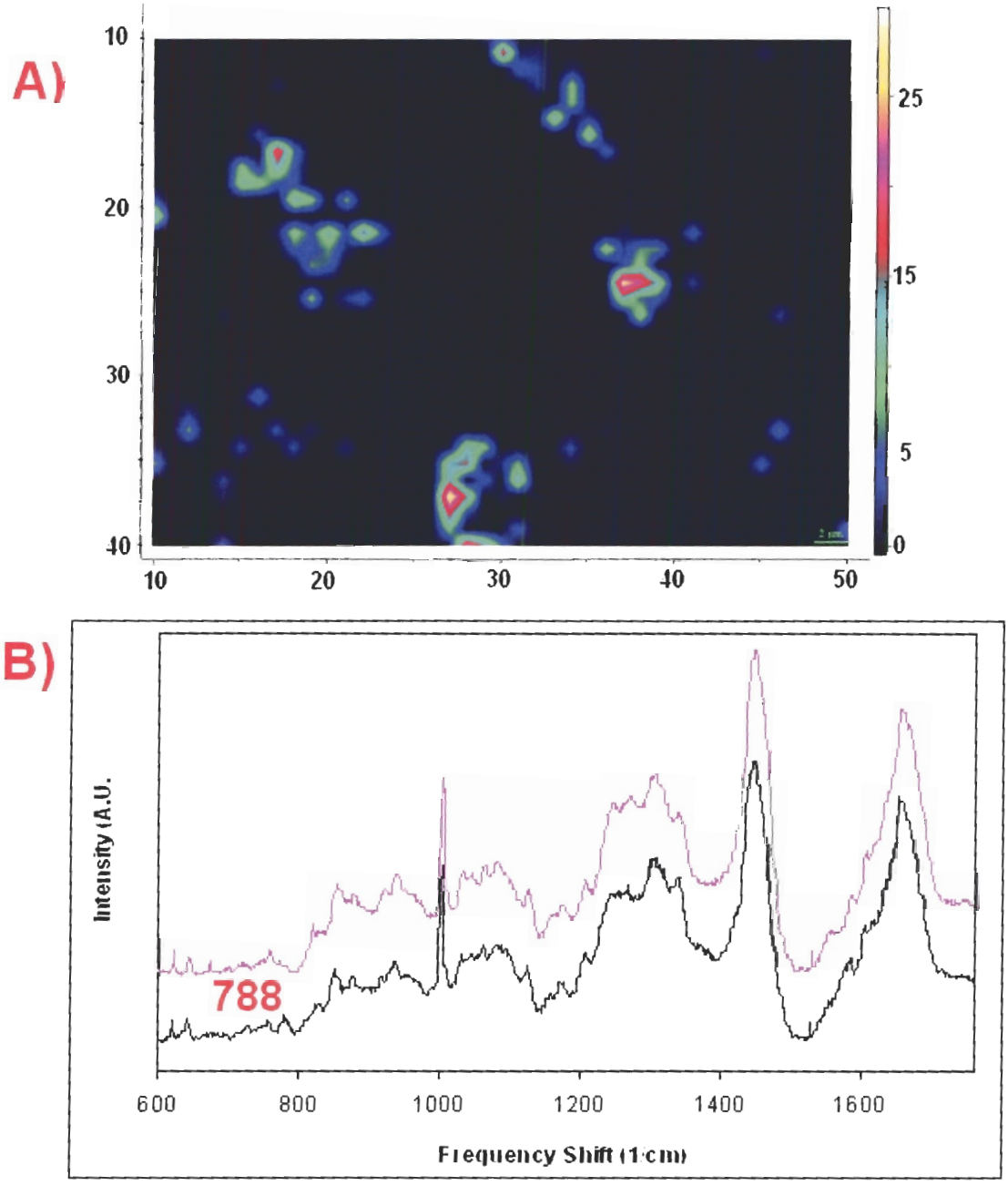

Figure 7.6: Raman image of normal connective tissue from the control sample based on the peak at 788 cm^{-1}(A), and the corresponding variation in said peak (B).

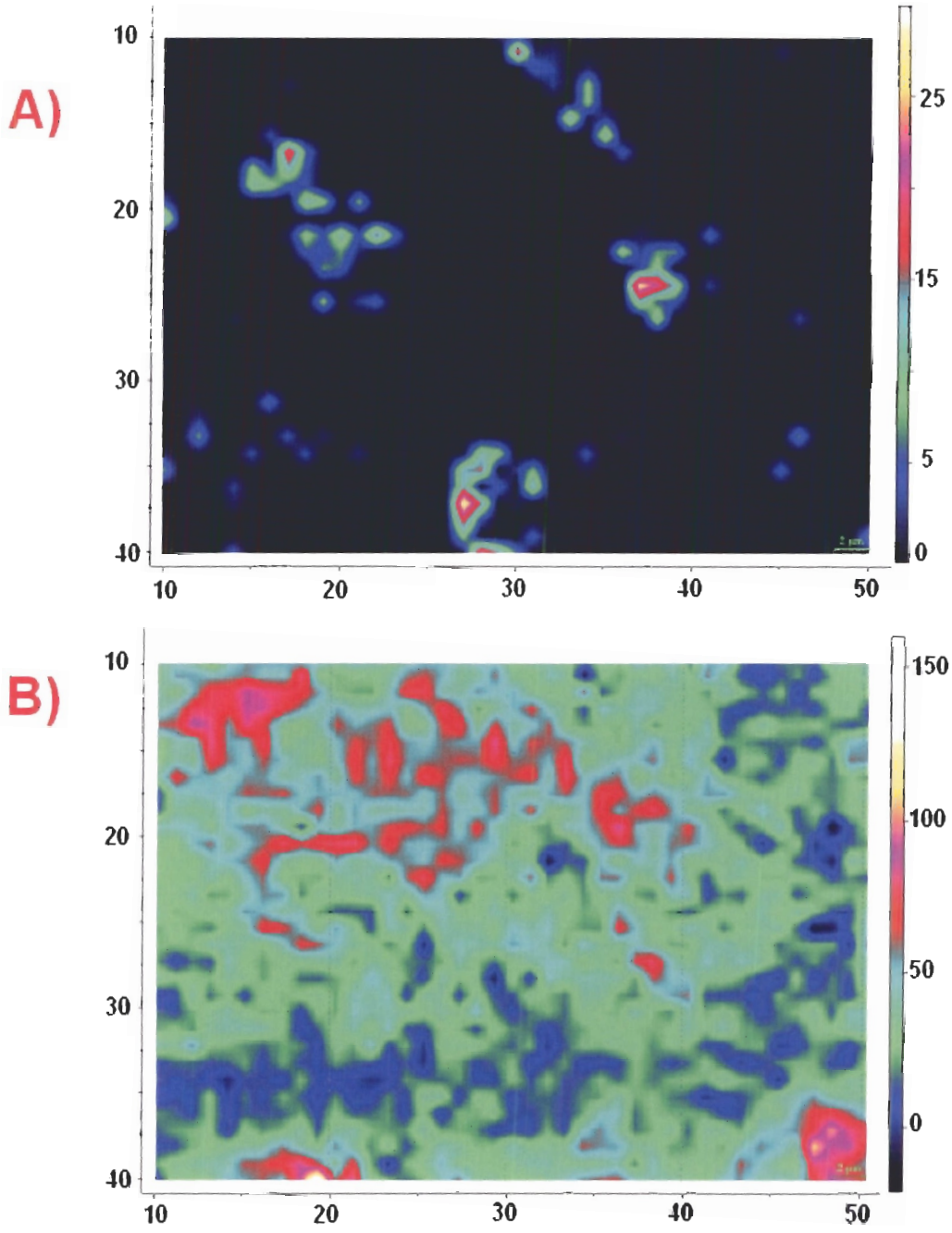

Figure 7.7: Raman images based on concentrations of nucleic acids (A) and lipids/fatty acids (B) for the normal connective tissue region of our control sample. Note the relative homogeneity seen in both images as compared to figure 7.5.

invaded tissue region. The lipid/fatty acid images are similar, showing not only more intense peaks but also much more variation in the invaded tissue region as compared to the unaltered connective tissue region. As well, one can note that there are fewer regions of variation in both images done in normal tissue (the variation over concentrations appears to happen more slowly). In the lipid/fatty acid image of normal tissue, we can see that the area of higher concentration is more or less in one region, while the remaining region is low concentration; we don't see the network-like distribution that we do in the cancer image. This is not surprising; normal loose connective tissue is usually quite homogeneous in molecular make-up (on the given scale measured), while cancerous tissue, with everything that is going on (increased cell density, increase in fatty acids and lipids as a result of increased cell density, the body's response to the neoplastic cell proliferation, etc.), is usually quite inhomogeneous. This variation in homogeneity could itself be used as an indicator of what type of tissue is being sampled.

It comes as no surprise that one sees significant differences between the invaded tissue and loose connective tissue, simply due to the significantly increased density of cells one sees in the invaded region. One could then question whether this method of analysis would be as useful when comparing the highly invaded region to a tissue type that normally also has a large cell density. These images would suggest that this type of classification using Raman spectroscopy would still be viable, as one can study the actual distribution of cells and nuclei (based on nucleic acid concentrations) and compare; normal tissue should show a much more organized cell distribution, due to the proper workings of extracellular proteins used for this very purpose. Concentrations of neoplastic cells usually show a much more skeltered distribution and usually have odd cell shapes (not to mention the presence of lymphocytes) when compared to normally occurring cells found existing in regions of higher cell density. All of these differences, while not necessarily being visible in a single

Raman spectrum, can quite possibly be seen when we map the Raman spectra over a sizable region.

7.4 Nucleic Acid Concentrations

When directly comparing the nucleic acid concentrations between the normal and cancerous tissue (figure 7.8), numerous differences can be seen. The cancerous tissue images not only show significantly more spots of nucleic acids (signifying more nuclei), but the intensity of the peaks pertaining to the nucleic acids are also significantly higher. The normal tissue image shows 3 larger "spots" of nucleic acids, with significant distance between them. The highly invaded image, on the other hand, shows numerous spots with less distance in between. The larger nucleic acid spots seen in the cancerous image are most likely neoplastic cells, if they indeed are individual cells and not two lymphocytes close together.

As well, one sees that the overall nucleic acid concentration is higher in the cancer image than in the normal image across the entire area sampled (not only are the lower concentration regions seen in the cancer image higher than in the normal image, but so are the high nucleic acid concentration regions). The higher nucleic acid concentrations seen in the cancerous sample may be due to the increased DNA one finds in lymphocytes as compared to normal cells, the aneuploidal nature of cancerous cells (having more than two sets of chromosomes), or may just be due to the higher statistical probability of cancerous cells potentially existing in the divisionary state of their life cycle.

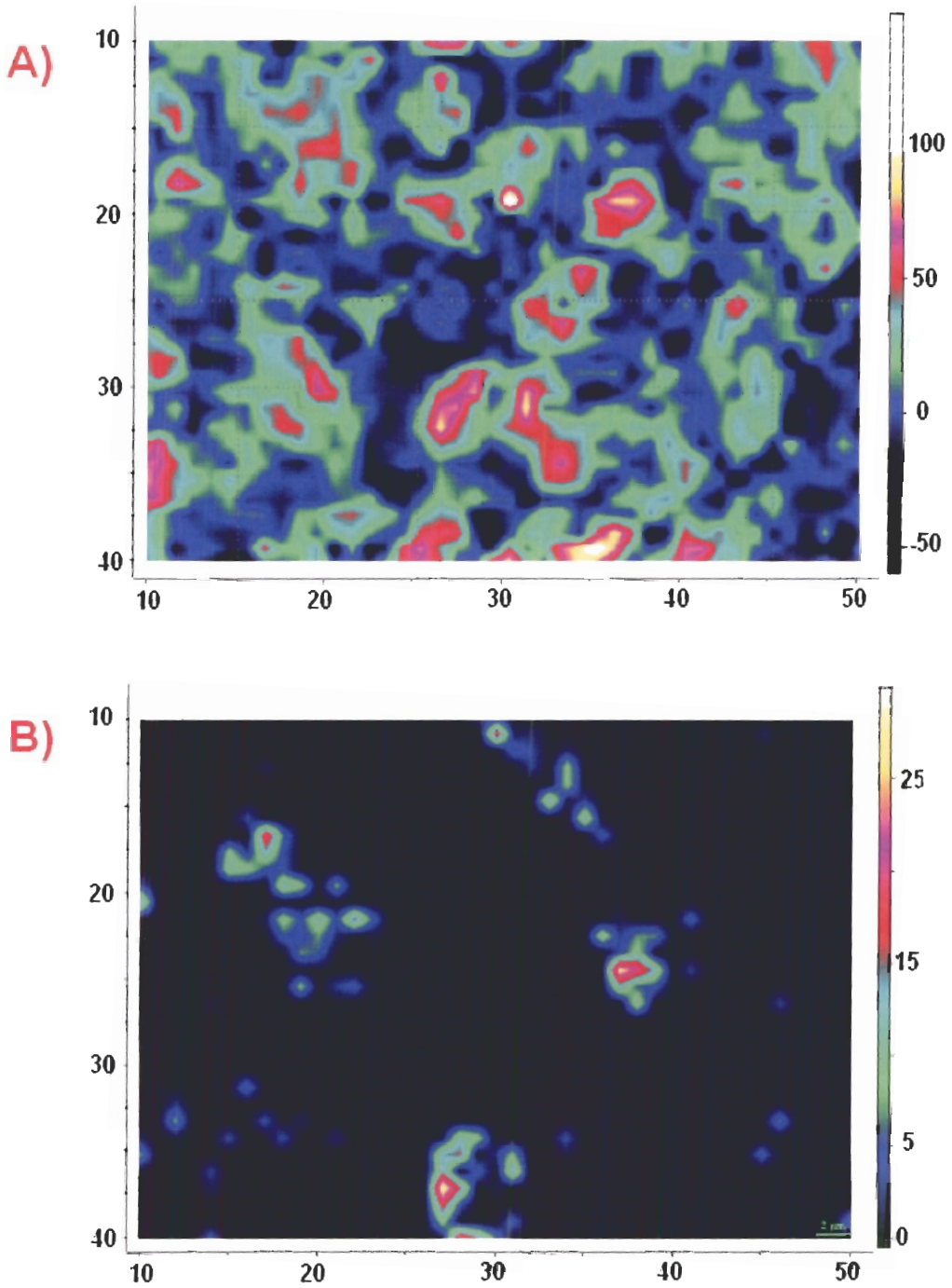

Figure 7.8: comparison of nucleic acid concentrations Raman images for the invade region of our cancerous sample (A) and normal connective tissue from our control sample (B). Note the difference in intensity scales. The cancerous sample shows significantly more regions of high concentration, as well as higher concentration over said regions.

7.5 Lipid/Fatty Acid Concentrations

When comparing the "1301 cm^{-1} peak" invaded region image to that of the control connective tissue image, it was found that the overall intensity of the peak was higher in the former (see figure 7.9). In terms of areas with the highest intensity, the 1301 cm^{-1} peak was seen to be the most intense in the invaded region image, denoting regions of higher concentrations of lipids/fatty acids than in the normal connective tissue image. We expect higher lipid/fatty acid concentration in the cancerous tissue, due to the high increase in cell density (and consequently cell membranes and fatty acids in the cell cytoplasm), so it is again not too surprising to measure this result. Areas of lipid/fatty acids concentration in normal loose connective tissue are expected as well, due to the few cells that do exist in that region, and potentially due to fatty acids that can be found present in connective tissue. The concentrations of both should be lower in connective tissue than in regions of high cell density, so our empirical results again match our expectations.

Also of importance is again how often we see the 1301 cm^{-1} peak vary in intensity in the cancerous tissue as compared to normal tissue; the variation from regions of lower intensity to that of higher occurs more gradually in the normal tissue than it does in the invaded tissue region, and the corresponding regions are quite large in size; the network-like distribution is not seen in the normal tissue image of the 1301 cm^{-1} peak. Suffice to say, one of the most vital pieces of information seen from the 1301 cm^{-1} peak images would undoubtedly

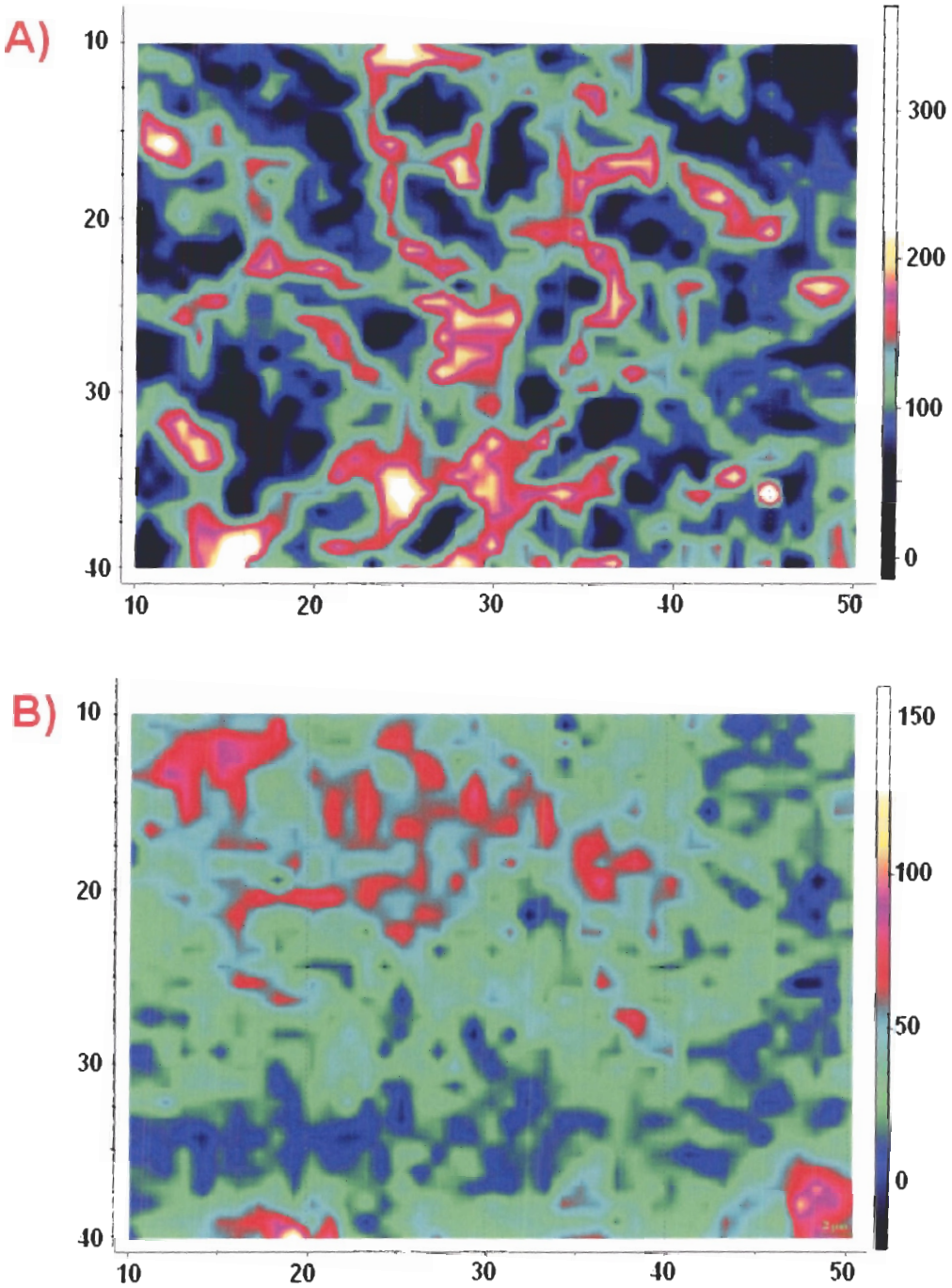

Figure 7.9: Raman images of the invaded region (A) as compared to normal connective tissue (B), based on concentrations of lipids/fatty acids. Note the invaded region shows a higher intensity over all (the intensity scales differ for (A) and (B)), as well as more variation in the distribution of lipids and fatty acids.

be the differences in distribution that we see; the network like structure of the cancerous tissue as compared to the larger, more gradual variation in intensity seen in normal tissue. This variation seen in the invaded tissue image suggests regions of varying chemical composition on a scale that would be expected for a region of high cell density (low concentrations for cell nuclei, high concentrations for cell cytoplasm).

7.6 Further Regions of Interest

Another region that shows significant difference between the unaltered connective tissue and the invaded region is the spectral region between 800-1000 cm^{-1}. Looking at this portion of the spectra, we see that the invaded region shows less intense peaks. In normal connective tissue, the most significant contribution to these peaks appears to be from collagen (these peak intensities are quite homogeneous throughout the connective tissue image), while in the cancerous tissue samples the most significant contributor to this region may possibly be due to a lower concentration of collagen (expected as the presence of a large number of neoplastic cells would most likely displace or degrade said collagen), a spectrum resulting from collagen that has been diminished due to the actions of proteins emitted to destroy the extracellular matrix, or may in fact be not due to collagen at all, but rather due to proteins that give a similar spectra in this region (with a few noticeable differences). This region will be studied in further detail later on.

We should also note that the region of 1200-1350 cm^{-1} also shows significant variation; this variation at first glance appears to occur not only between normal connective and invaded tissue regions, but also between differing regions within the two tissue types as well (including the 1301 cm^{-1} peak). Whether or not a pattern within these variations can be attributed to strictly normal connective or invaded tissue is not immediately apparent, save perhaps

for the much increased intensity seen at the 1301 cm^{-1} peak from the invaded tissue region in the cancerous sample.

7.7 Comparison of Raman Image with Expected Cell Density

As stated previously, our unstained samples were stained after completion of collecting Raman spectra, in the hopes of a direct comparison of the location of individual cell nuclei with the regions that showed high concentration of DNA/RNA in our Raman images. This proved impossible to do, as the staining process provided little extra visual detail, again potentially due to our samples being left in the open air for a few weeks during the Raman sampling. However, looking at our stained samples we can compare the cell nuclei density of our white light images with our Raman images, to allow for a verification of concentrations. Figure 7.10 compares the white light image of our highly invaded region (at high power) with the Raman image from the corresponding image. As can be seen, the cell nuclei density and nuclei size is quite comparable. Some of the nuclei in the white light image also don't appear as dark, suggesting slightly a slightly lower concentration of nucleic acids; these spots may correspond with some of the regions in the Raman image where the nucleic acid concentration isn't as intense. Suffice to say, when studying the peak at 788 cm^{-1} (corresponding to nucleic acid concentration), it appears we get an excellent correlation between the number and size of regions of high intensity of said peak, and the number and size of cell nuclei seen in the white light image.

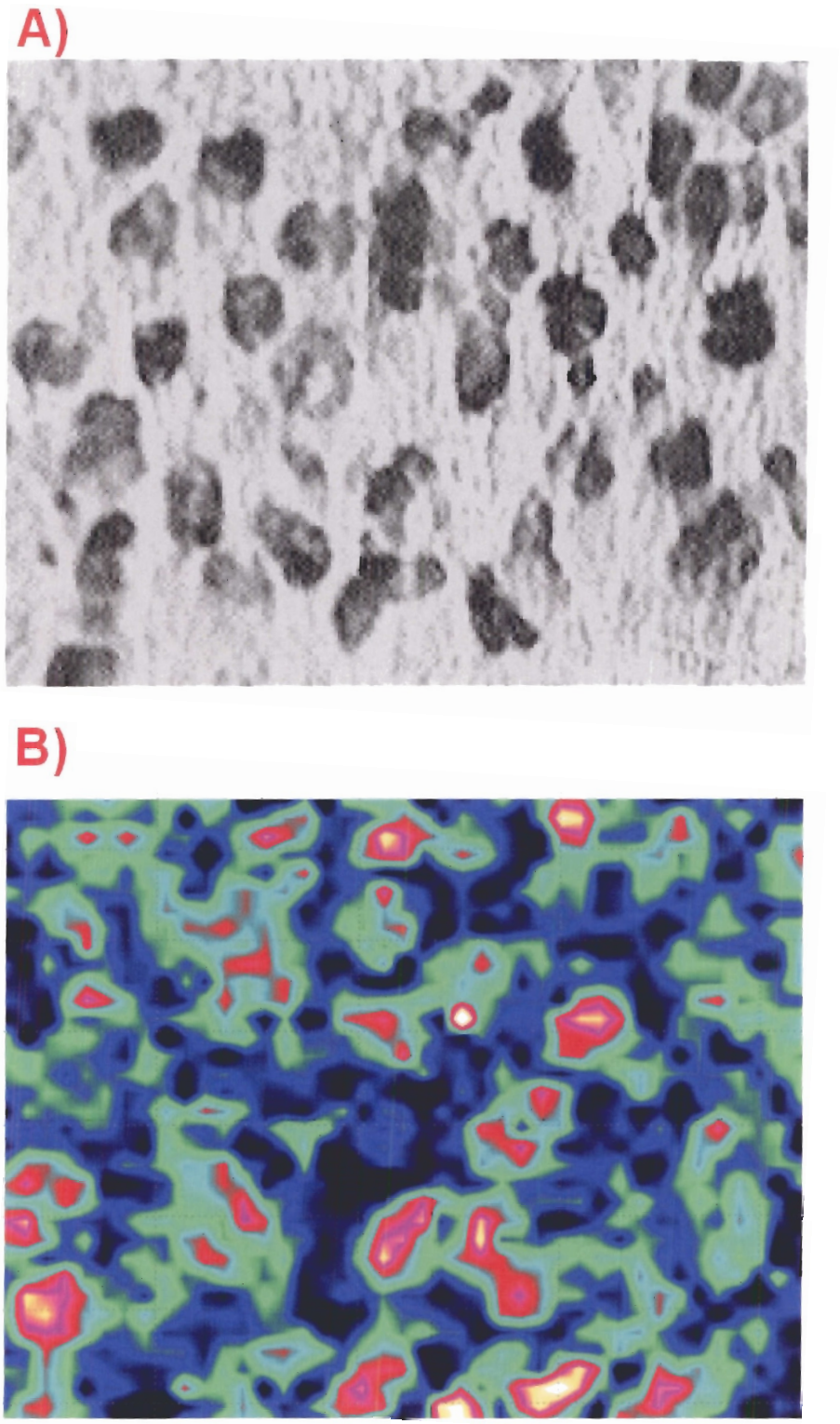

Figure 7.10: Comparison of a white light image (A) and a Raman image (B) from the highly invaded connective tissue region (each image is about 45 microns across). The concentration and size of regions of high nucleic acid concentration in the Raman image correlate quite nicely with the size and concentration of dark spots seen in the white light image.

8. Results of ten by ten Micron Spectra Compilation

In this section analysis of the averaged spectra over ten micron by ten micron areas will be discussed. This analysis will first comprise a basic comparison of the spectra from each given region, namely the highly invaded region in our cancerous sample as compared to the normal connective tissue of our control sample. After this, statistical analysis will be performed to see if our data "clusters" together as we would expect (based on the differing tissue compositions), and subsequently if this statistical analysis provides a means of identifying and classifying said tissue types.

8.1 Preliminary Study of Data

Figure 8.1 is a comparison of the average of the 10 micron by 10 micron spectra compilations (minus fluorescent background) of normal connective tissue, as compared to spectra taken from the highly invaded region of connective tissue seen in the cancerous sample. The invaded region consists of 70 separate 10 micron by 10 micron spectra, while the average of the normal connective tissue spectra is composed of 37 said spectra. The figure also shows the limits of one standard deviation for both average spectra.

As can be seen, the highly invaded region in the connective tissue gives a more intense spectra in the majority of regions, save for the region between 600-1000 cm^{-1}, where the intensity appears equivalent *at first glance*. It is interesting to note that between 1200 -1500 cm^{-1} the lower standard deviation limit of the highly invaded region coincides quite well with the higher limit of the normal connective tissue spectra. This region predominantly corresponds to spectra resulting from protein and lipid/fatty acid concentrations. Figure 8.2 shows the resultant difference between the two averages seen in figure 8.1, as well as the

suspected error range (using basic error propagation theory) resulting from the standard deviations seen in the same figure.

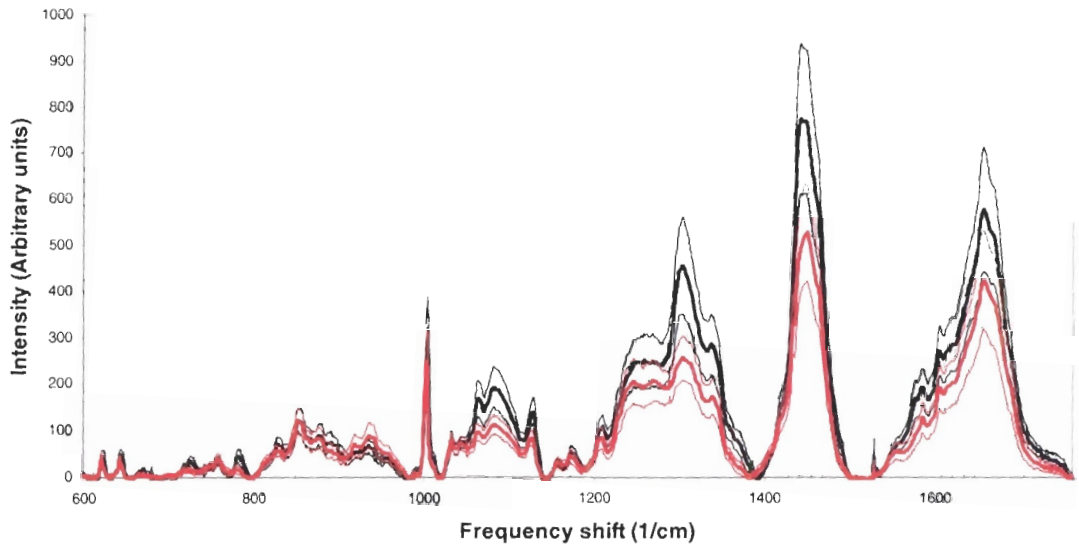

Figure 8.1: Plot of the average spectra of data taken from the highly invaded region in the cancerous sample (black) as compared to normal connective tissue (red). The thin lines denote one standard deviation from both the cancerous sample (black) and the connective tissue region (red).

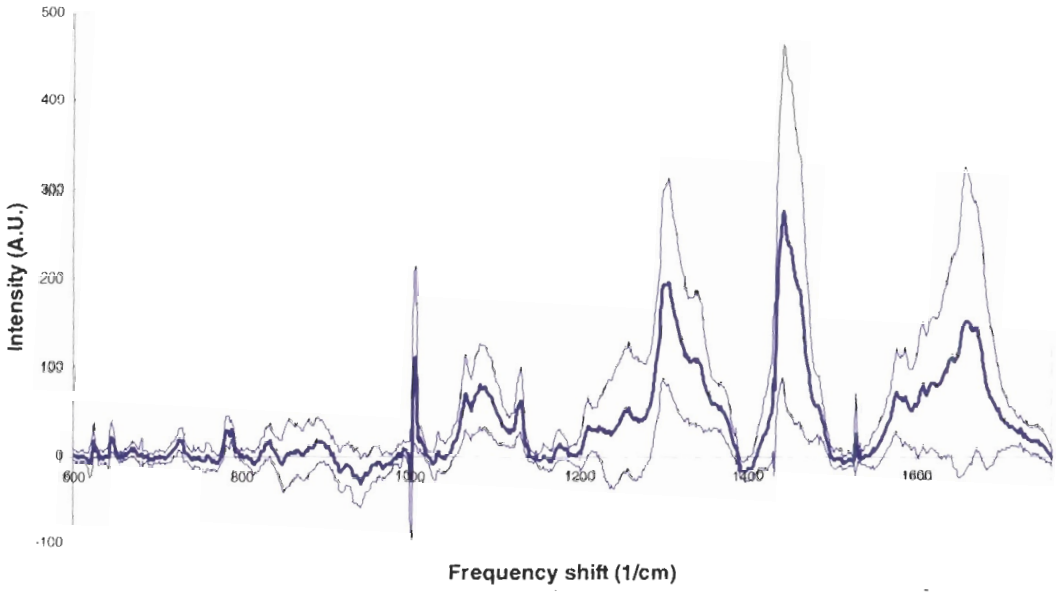

Figure 8.2: Plot of the difference between the average spectra of the highly invaded region in the cancerous and normal connective tissue. The thick line denotes the mean difference; the thin lines denote error based on the standard deviations as seen in figure 8.1.

Looking at the region between 600 - 800 cm^{-1}, we can see a few noticeable peaks; 625 cm^{-1} and 646 cm^{-1} are twisting vibrational modes of phenylalanine and tyrosine respectively [9][6]. The peaks occurring at approximately 730 cm^{-1} and 788 cm^{-1} are both indicative of nucleic acids; the 730 cm^{-1} peak indicates the presence of the base argine, while the 788 cm^{-1} peak is due to a stretching mode of the phosphate background found in DNA and RNA[7]. Hence, we seem to have the suggestion that there is more nucleic acids present in the highly invaded region; no real surprise, as we'd expect a higher concentration of nucleic acids present as a result of the much increased cell concentration found in the invaded region as compared to normal connective tissue.

The region between 800 – 1000 cm^{-1} is rather difficult to ascertain with accuracy; it is a region that shows striking similarity between numerous types of proteins. The peak found at approximately 897 cm^{-1} is again indicative of nucleic acids. The negative peak seen at 940 $cm^{-1}$is a protein dependent peak, due to a C-C vibrational mode found in alpha helices. The fact that this peak is negative stipulates that its concentration is higher in the normal connective tissue sample than in the highly invaded region studied in the cancerous sample. As this peak is a prominent peak found in collagen (usually more prominent in collagen than other protein types when compared to the peak seen at 850 cm^{-1}), this would seem to indicate a higher concentration of collagen in the normal connective tissue region. The small negative peak seen at approximately 850 cm^{-1} would provide further evidence for this, as it also a peak seen in collagen. Whether this decreased collagen content is due to less collagen in the invaded area (resulting from the increased cell density) or due to collagen that has been degraded cannot be determined.

---

[6] Again, all listings of spectral wave numbers for given peaks have been cited from this source.
[7] Refer to Appendix A for reference spectra of collagen, DNA, and oleic acid.

The region between 1000-1200 cm^{-1} also shows significant differences. The large peak seen at 1003 cm^{-1} is a ring breathing mode of phenylalanine, and as can be seen is quite sharp. The error of said peak is quite large however, due most likely to subtle differences in the location of said peak; 1002 cm^{-1} on the normal tissue average and 1003 cm^{-1} on the invaded tissue average (within the acceptable uncertainty range). Consequently no final conclusion on the variation of the intensity should be made, though when looking at the averages of both the invaded region spectra and the normal connective tissue spectra the intensity appears slightly higher in the invaded region. The peaks seen at approximately 1069 and 1133 cm^{-1} correspond to C-N stretching modes in proteins and C-C stretching modes in lipids/fatty acids, again suggesting higher concentrations of either one of these materials (or both) in the highly invaded region of the cancerous sample. The broad peak seen at 1088 cm^{-1} is most likely a combination of vibrational modes resulting from lipids/fatty acids and nucleic acids, and the peak seen at 1182 cm^{-1} is most likely a vibrational mode resulting from tyrosine.

Between 1200 to 1400 cm^{-1} we see a large uncertainty range, signifying a lot of variation in the spectra of both the invaded region and the normal connective tissue region. This spectral range shows significant variation in both intensity and shape, making it difficult to ascertain an accurate average of the intensity of the peaks seen. Nonetheless, in this preliminary study we can at least stipulate that the peaks seen between 1300 cm^{-1} and 1361 cm^{-1} show increased intensity from the invaded region of the cancerous sample as compared to normal connective tissue. The peak seen at approximately 1301 cm^{-1} is due most prominently as a result of lipid/fatty acid concentration, though may potentially result from increased protein levels as well. The peak at 1343 cm^{-1} corresponds to the nucleic acid bases argine and guanine, while the peak at 1371 cm^{-1} is the result of a lipid/fatty acid stretching mode (CH_3 symmetrical stretching).

Between 1400 cm^{-1} to 1730 cm^{-1}, we again see significant changes, and a high level of variability. We see an increased intensity in the invaded region of the 1449 cm^{-1} peak, an intensity again suggesting a higher level of concentration of both proteins and lipids/fatty acids (C-H deformation vibrational mode). The shape of this major peak also becomes altered, with a much more "slanted" look on the right hand side, most likely due to a more intense peak at 1460 cm^{-1} in the invaded region sample (again correlating to proteins and lipids/fatty acids).The shoulder seen at 1491 cm^{-1} is unassigned, while the sharp peak seen at 1528 cm^{-1} is the result of an odd occasional glitch in the system and should be ignored. Between 1586 cm^{-1} to 1735 cm^{-1} we again see a large error range, suggesting increased variation in this part of the spectra in both tissue samples. The peak at 1586 cm^{-1} is most likely again due to argine and guanine, 1620 cm^{-1} is due to tyrosine and tryptophan, and 1668 cm^{-1} is a protein and lipid/fatty acid peak. This again provides evidence of either increased protein (save collagen) or lipid/fatty acid concentrations (or again both), and nucleic acid concentrations; all unsurprising given the expected increased concentration of cells.

Figure 8.3 shows both the highly invaded region as compared to the spectra taken from normal connective tissue, after being normalized to the average height of the peak seen at 1449 cm^{-1} for each tissue type, to allow for a direct comparison of the *average* Raman signature of each tissue type (each spectra was readjusted by multiplying the entire given spectra by a corresponding common multiple). Figure 8.4 shows the difference between the highly invaded region compared to the normal region after this normalization. The invaded region consists of seventy 10 micron by 10 micron spectra; two outliers were removed (their removal was found to have a neglible effect on both the average spectra and the standard deviation). The average of the normal connective tissue spectra again consisted of thirty-seven separate 10 micron by 10 micron spectra. Unsurprisingly, the standard deviation of the compilation of the spectra after being normalized to the 1449 cm^{-1} peak was either reduced

(e.g. between 1200 – 1500 cm^{-1}) or remained the same in magnitude (e.g. 600 – 800 cm^{-1}).

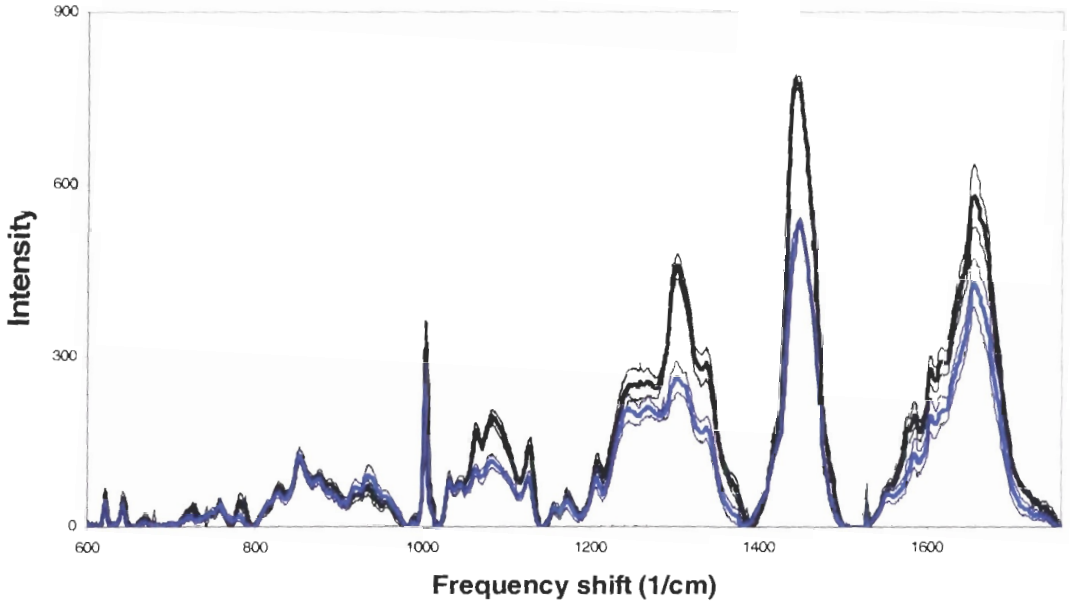

Figure 8.3: plot of the average spectra of the highly invaded region in the cancerous sample as compared to the average of normal connective tissue, with spectra being normalized to the average of the 1449 cm^{-1} peak respectively. The heavy black line denotes the highly invaded region in the cancer sample, while the heavy blue line denotes the connective tissue region. The thin lines denote the one standard deviation boundaries (black for the cancer sample, blue for the connective tissue sample). Note the significantly less error resulting from the normalization, as compared to figure 8.1.

There is also minimal difference between the mean difference between the normalized spectra of the invaded and connective region as compared to the unnormalised spectra; the only significant change is the apparent reduction in the error of said difference. Both figures denote the same changes, with the invaded region again showing a more intense spectra in all important regions save at approximately 950 cm^{-1}.

Figure 8.4 again lays claim to the fact that the concentrations of nucleic acids and lipids/fatty acids (and potentially non-collagen protein) is higher in the invaded region than in the connective tissue region, while concentration of collagen appears higher in the connective tissue region. Again, these results are as expected.

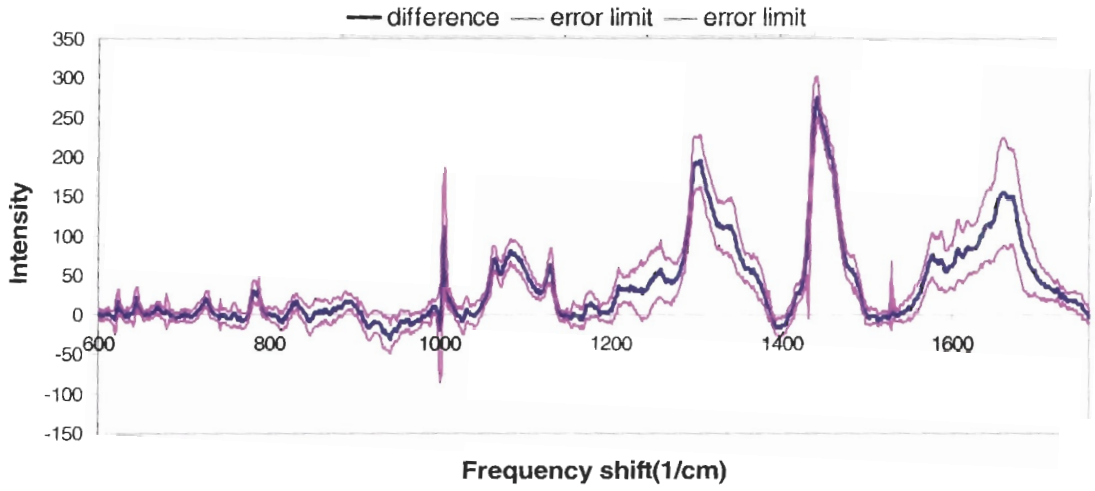

Figure 8.4: Plot of the difference of the average spectra as seen in figure 7.4 (between the highly invaded region and the normal connective tissue region), as well as the defined error limit as gained from the standard deviations found in figure 7.4. Note that the mean difference appears equivalent to that found in figure 7.3, save that the error range is much less.

While it is difficult to ascertain directly if some of the higher peaks seen in the invaded region are due to either an increase in the lipid and fatty acid or protein concentrations, cancer theory would seem to suggest the former plays the most dominant role. Connective tissue itself is comprised mainly of a few types of proteins (collagen, elastin, etc), while concentrations of lipids and fatty acids are found in much more abundance in cells and cell cytoplasm. Coming to this conclusion would also agree with the results we saw in chapter 7, where we saw a much more intense peak at 1301 cm^{-1} (predominantly lipid/fatty acid based) in the invaded tissue region than in the unaltered connective tissue region. Consequently then, one would suggest that the majority of the increases

in molecular concentration one sees in the invaded tissue region would be due to nucleic acids and lipids/fatty acids, while normal connective tissue shows a higher amount of collagen. Exact individual protein concentrations are undoubtedly different between the two tissue types, though the exact changes seen are quite subtle, due to quite similar spectral signatures between protein types (and the fact that the number of different protein types seen in cells is quite vast).

The same difficulty arises in attempting to figure out exactly which lipids/fatty acids have an increased concentration, as numerous fatty acids (and lipids) also show spectra with only subtle variations. RNA and DNA also share spectra that will appear equivalent, at least with the level of detail seen in our spectra. Nonetheless, even with these preliminary results we can see substantial differences in the Raman spectra of the two tissue samples, results that confirm our expectations and can be used as a classification means.

8.2 Categorical Principle Component Analysis

In a desire to clarify any trends in our data that would allow us to classify an invaded tissue region from normal connective tissue, a categorical principle component analysis (CATPCA) was run on our data. CATPCA is a process by which we transform our set of variables into a set of smaller variables, by looking for correlations between variables in our data. Once these correlations are found, the data is expressed in a new coordinate system, one based on the magnitude of the variational spread seen in the data set. Hopefully this will allow us to identify distinct patterns in our set of Raman spectra, allowing us to note similarities and differences in the data set. If all goes well, it should allow us to see a pattern by which spectra from the invaded tissue region in the cancerous sample and normal connective tissue can be separated.

In further detail, principle component analysis works in the following steps [18]. Firstly, we define our data set as a series of variables or vectors, defining each wave number bin as a vector with a magnitude based on the intensity seen at said wave number. In our case, our data set was reduced to 256 bins (over the range of 600 cm^{-1} to 1760 cm^{-1}), to allow for quicker computation. As such, we have a data set comprised of observations (each individual spectrum), with each observation being comprised of 256 vectors. Each vector (or wave number bin) is originally considered to be orthogonal, with our plan to reduce the number of total vectors by looking for relationships between each wave number bin (suggesting a lack of orthogonality). In other words, we want to look for relationships between these original vectors, and use this to create a new, orthogonal set of vectors that (hopefully) allows for a means to identify patterns.

Once we have our defined original vectors, we need to define a means to establish relationships between the separate vectors or wave number bins. This is established by calculating the covariance between given vectors, defined as

$$\mathrm{cov}(X,Y)=\sum_{\alpha=1}^{m}\frac{(X_\alpha-\overline{X})(Y_\alpha-\overline{Y})}{m-1}. \qquad (8.1)$$

$X\alpha$ and $Y\alpha$ are the respective values for the wave number bin of the given spectrum α, and $\overline{X}$ and $\overline{Y}$ are the mean values of the given wave number bin for the entire spectral set (the summation is completed over all m spectra in our data set).

Basically what we are doing here is depicting our data set of separate spectra as a series of vectors, each vector having its own labeled data set. The actual wave number vector values measured for a given spectrum are then compared by studying the distance from the mean value seen in the given wave number vector, and are then multiplied together. If it is the case that a correlation

exists between two individual wave number bins, then the total summation should be nonzero - a point a large distance away from the mean in one wave number vector correlates with a point a large distance away from the mean in the other wave number vector (creating a large value when the "two distances from means" are multiplied together). A zero value (or close to zero value) for the summation signifies that the distances from the mean vary randomly between the two studied wave number vectors, and consequently no correlation exists.

Since we have 256 wave number vectors, we have a corresponding large number of covariance values. To allow for easier computations, a matrix form is used for the covariance values,

$$C = \begin{bmatrix} \mathrm{cov}(1,1) & \mathrm{cov}(1,2) & \ldots & \mathrm{cov}(1,n) \\ \mathrm{cov}(2,1) & \mathrm{cov}(2,2) & \ldots & \mathrm{cov}(2,n) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{cov}(n,1) & \mathrm{cov}(n,2) & \ldots & \mathrm{cov}(n,n) \end{bmatrix}. \qquad (8.2)$$

The numerical values denote the separate wave number bin vector. This matrix is symmetrical, with values of say *cov(2,1)* equaling *cov(1,2)*. The values down the diagonal are the variance values of the given wave number vectors, a measure of the actual variational spread seen in the given vector (the square of the standard deviation of the total data set).

One of the benefits of studying the new variables (based on the transformation resulting from the old variables) is that it allows us to study our data by looking at variables that encompass the most amount of variation in the data set. This is gained by finding the eigenvectors of the above covariance matrix; the eigenvector with the highest eigenvalue corresponds to the principle component that accounts for the most variation seen in the dataset, the second highest eigenvalue accounts for the second most variation seen, etc. *The*

eigenvectors show the direction of the measured variation, the eigenvalues show the magnitude of variation projected along that line. In a sense then, we are transforming our coordinate system to a new coordinate system, one that uses a new variable definition to allow us to determine our axes in terms of variables that account for the variation seen in our original dataset (instead of our original variables, which may or may not be wholly independent from each other). Figure 8.5 shows an example of this for a 2 dimensional data set, showing not only the two eigenvectors, but also the original data. The original data shows an approximate linear relationship, so having our principle component along this linear direction is as expected, as this is the direction one sees the majority of variation in the dataset.

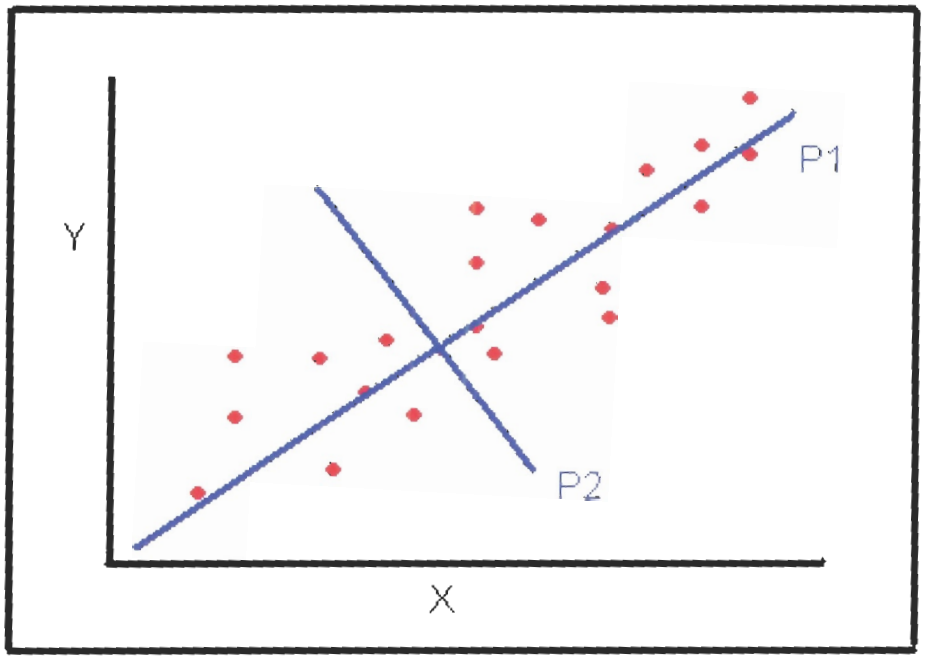

Figure 8.5: Sample plot of original 2 dimensional data (based on X, Y values), as well as the two principle components. Note that the P1 vector is along the direction of the majority of the variation seen in the data set, while P2, pointing orthogonally to P1, denotes that there is variation away from P2 (the original data is not in an exact line, but rather varies away from said line).

Categorical principle component analysis differs from the above in a slight yet important way; optimal scaling is used to account for potential nonlinear

relationships that may exist between our original variables. Optimal scaling procedures constitute methods used to accurately compare variables of differing measurement levels. The methods used can be quite extensive, and won't be discussed here.

8.3 Results of Categorical Principle Component Analysis

In an effort to see if a CATPCA could be useful as a means to identify invaded tissue from normal connective tissue our data was analyzed using tools in SPSS, a statistical software program. To further clarify our results, we broke our spectral data set into 4 groups – invaded tissue region, normal connective tissue region (as taken from our control sample), boundary invaded regions (where the tissue showed less cell density, suggesting the tissue wasn't as invaded by neoplastic cells), and finally regions far away from the invaded tissue region on the cancerous sample. It was our hope that the regions far away from the invaded tissue region clustered reasonably well with the normal connective tissue region from the control sample. As well, we were curious to see how the boundary region spectra clustered with the other regions, though it was suspected that it would show some clustering with both region types.

As stated, we reduced our wave bin number to 256 bins (from the approximate original 750), to allow for quicker and easier computing. When completing the CATPCA, it was found that of the first five principle components, plotting the first two (P1 versus P2) resulted in the most significant clustering of our data set. Plotting the other principle components (e.g. P1 versus P3) showed some signs of clustering, but also showed mixing between region types. Consequently, it appears that the significant differences seen between the given types of tissue regions are accounted for quite strongly by the first two principle components of our CATPCA, and the further principle components consist of

variation that is not indicative of any one type of tissue region (e.g. no viable clustering or noticeable patterns).

CATPCA was run not only on our original, raw data, but was also completed for our data normalized to the Raman peak seen at 1449 cm^{-1}. It's important to note here that in section 8.1 we normalised to the average peak height seen at 1449 cm^{-1} for *each* tissue type; here we are normalising *all* of our spectra to the same peak height. The idea behind this was to normalise to a peak to account for potential variations in intensities one can not only see from machine to machine, but also from tissue regions that may not be as properly focused as other regions, or regions of similar compositional makeup but appear to be less dense. Another peak save 1449 cm^{-1} could also be used as a normalization peak, as we are just looking at the differences in the total spectral signature (e.g. the spectral "shape"), though the 1449 cm^{-1} peak allows for hopefully a more accurate comparison, due to its signal to noise ratio. It is important to note that while this normalization doesn't affect the relative concentrations in the given spectra, it does affect the absolute concentrations, and should not be used as a direct comparison between tissue types in that regard (especially for spectra that have say significantly low or higher concentrations of the tissue component that corresponds to the peak used to normalize).

8.4 Results of CATPCA Performed on Raw Data

Figure 8.6 shows the plot of the first categorical principle component as compared to the wave numbers that it correlates to. This component actually accounts for 51.2% of the total variability seen in our spectral data set. The number of spectra used consisted of 70 spectra from the invaded tissue region, 27 from tissue far away from the invaded region, 37 spectra from connective tissue in the control sample, 34 spectra from "boundary" (lower cell density)

regions, and 5 spectra listed as unusual (173 in total). The location of border regions were approximated by studying the stained and unstained samples, looking at regions where there *appeared* to be less cell density.

As can be seen, significant positive weightings can be seen in numerous regions, regions that suggest increases in lipids/fatty acids, nucleic acids, and proteins as P1 increases. This would suggest then that the invaded tissue region (showing increases in all of these given molecular constituents) will occur for high values of P1, and lower values correlate to regions of normal connective tissue. The fact that the first principle component accounts for the significant variability seen in the above three tissue components simply verifies that the largest variation seen in our data set is due to these three components, as was expected.

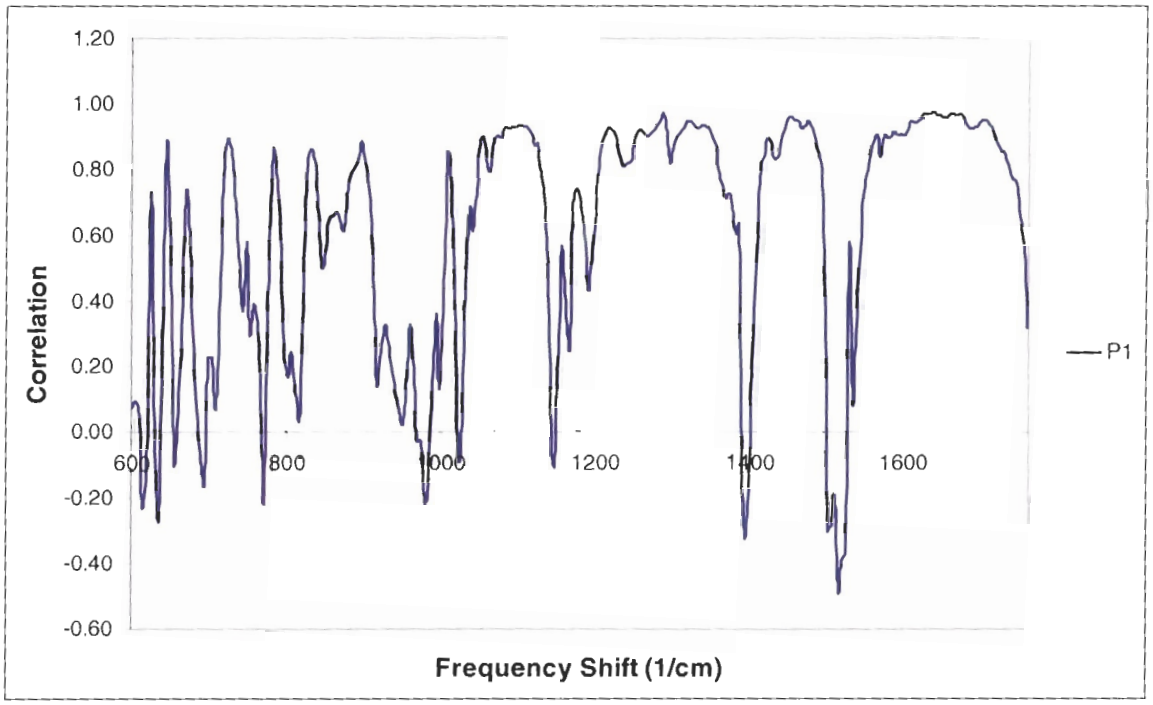

Figure 8.6: Plot of the relationship of P1 with respect to the frequency shift of our scattered Raman light. A positive weighting signifies a more intense signal at the given wavenumber as P1 increases, while a negative weighting shows a less intense signal. The above shows that signal intensity increases for wavenumbers that have contributions from nucleic acids, fatty acids/lipids, and proteins. The low weighting region seen between approximately 900 – 1000 cm^{-1} suggests no weighting along P1 for this region, giving rise to the conclusion that P1 has no weighting to collagen concentrations.

Again, it is important to note that the possibility exists that a higher density of material measured will result in an overall increased Raman spectra, something that our CATPCA does not account for. Of course, with the variations seen in the given chemical concentrations seen in the studied tissue, it's quite difficult to ascertain if we are looking at a region with an overall less intense signal, or one with simply a lower concentration of one given chemical concentration (say fatty acids for example). As stated above, we will initially just look at the absolute intensity values taken from our data set; in the next section we will normalize all our data to a given peak (1449 cm^{-1}) and compare in that manner. This allows us to study both possibilities, the current method letting us acknowledge the fact that differing tissue types may show differing densities of materials (resulting in stronger/weaker overall signals), while the latter allowing us to study only relative concentrations of given tissue types, ignoring absolute intensities. Suffice to say, in this current measuring method, fatty acids/lipids, nucleic acids, and potentially certain proteins (as shown by the increase in the amino acids tyrosine and phenylalanine) all show stronger peaks as P1 increases, for whatever reason.

We should also note that the region between 900-1000 cm^{-1} shows little weighting, suggesting that this region does not significantly vary in intensity over P1. This region potentially corresponds to a significant peak that may indicate collagen (940 cm^{-1}), so the fact that it does not vary over P1 may mean that collagen does not vary over P1 (or proteins in general), or that potential contributions from other proteins contribute to this intensity while collagen does vary. It is a little tricky to accurately identify which occurs, as certain proteins do provide to the spectral signal between this range (the peak at 940 cm^{-1} is due to an C-C stretching mode seen in alpha helices), though not usually to the same amount of intensity a given collagen concentration would (it's easier to identify collagen and other proteins when looking at more than the one peak at 940 cm^{-1}).

Figure 8.7 shows the plot of the second categorical principle component as again compared to wave number. This component accounts for 16.5 percent of the total variability seen in the data set (the third component accounts for 4.5 percent, the following components decrease rapidly). As can be seen, a significant positive weighting exists for the regions 920 – 980 cm^{-1}, as well as 1150-1285 cm^{-1}. These regions consists of a notable protein peaks (most likely collagen), and suggest that protein content (again, most likely collagen content) increases as P2 increases.

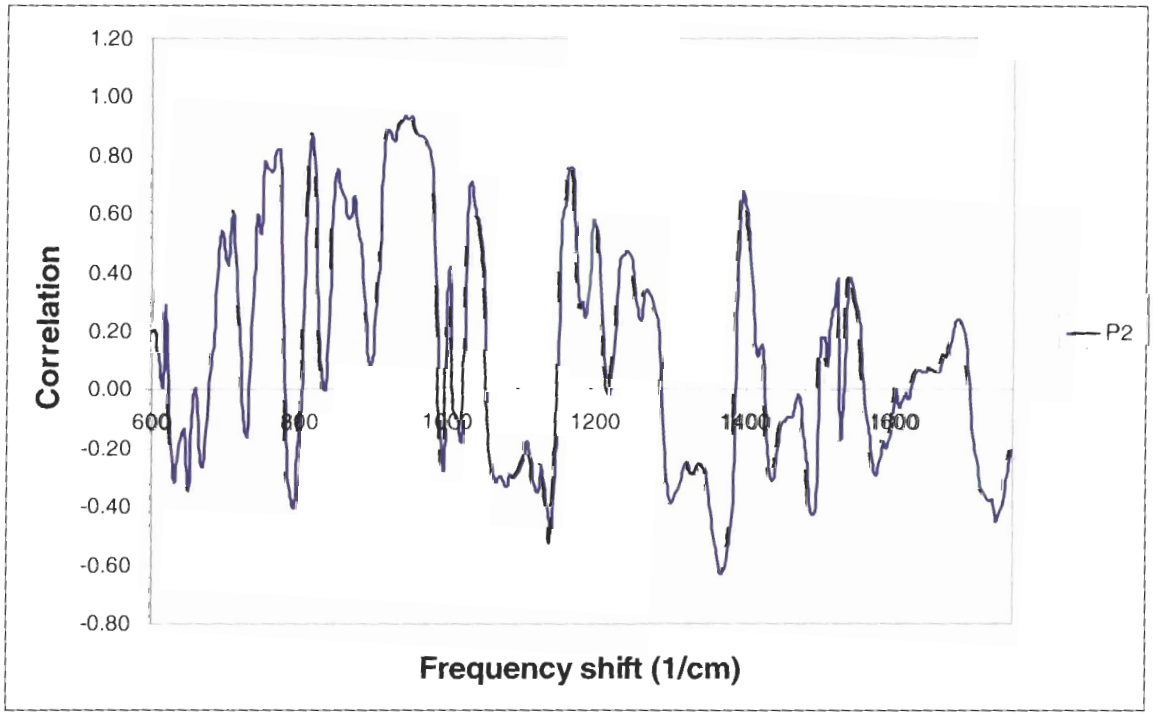

Figure 8.7: Plot of the relationship of the second principle component with respect to the given frequency shift. As can be seen, significant positive and negative weightings are present. The above plot suggests that collagen content increases as P2 increases, and that nucleic acids and fatty acids/lipids decrease as P2 increases.

Significant negative weightings also exist In P2, for the regions between 1050 -1150 cm^{-1}, 1280-1385 cm^{-1}, and 1710-1755 cm^{-1}.A small negative weighting also exists between 1405 – 1500 cm^{-1}. These regions correspond to

peaks signifying fatty acids/lipid concentration, so they again suggest said concentrations decrease as P2 increases. However, the significantly smaller negative weighting at 1405-1500 cm^{-1} (as well as the nonzero positive weighting seen between 1620 -1702 cm^{-1}) provides some interesting detail here. Beforehand we suggested that the large peaks seen at 1449 cm^{-1} and 1659 cm^{-1} were predominantly due to fatty acids and lipids; if this were indeed true, we'd expected both these regions to have similar magnitude negative weighting as the regions 1050-1150, 1290-1385, and 1710-1755 cm^{-1} (e.g., we'd expect the variation seen in the fatty acid/lipid concentration to be consistent when transformed into a given principle component, as it is in P1). Since this is not the case, one must conclude that the peaks seen at 1449 cm^{-1} and 1659 cm^{-1} also have significant contributions resulting from proteins. This is not at all surprising, but does allow us to get an idea of how much protein contributes to these peaks. Indeed, it lets us know that to study only lipid and fatty acid concentrations, these two peaks should not be included, as their signal intensity will have protein contributions.

Putting together the positive weighting in the regions 920-980 cm^{-1} and 1150-1280 cm^{-1} (as well as the shape of the correlation, noting the peaks at approximately 1240 cm^{-1} and 1260 cm^{-1}), along with the negative weighting at 1280-1385 cm^{-1} provides strong evidence that we have an increase in collagen as P2 increases.

If one looks closely as well, we can see a negative weighting at peaks seen at approx 728 cm^{-1}, 788 cm^{-1}, and 1578 cm^{-1}, suggesting a small negative weighting between nucleic acids and P2.

The results of all of the above suggest that by plotting P1 versus P2, we can form a plot that can show the desired variation in intensity levels of spectral peaks correlating to nucleic acids, fatty acids/lipids, and collagen. Theoretically, if invaded tissue shows higher concentrations of the first three listed chemical

types, its P1 values should be high, while its P2 values should be low/moderate. Similarly, connective tissue should have lower values of P1, while having higher values of P2 (due to increased collagen content and much lower fatty acid/lipid and nucleic acid content).

Figure 8.8 shows the results of the CATPCA on our raw, unaltered spectral data (a plot of the above principle components P1 and P2). As can be seen from said figure, our empirical results behave quite nicely with what we expected. Invaded tissue spectra occupy the lower, right hand side of the plot, suggesting increased values of fatty acids/lipids, proteins, and nucleic acids.

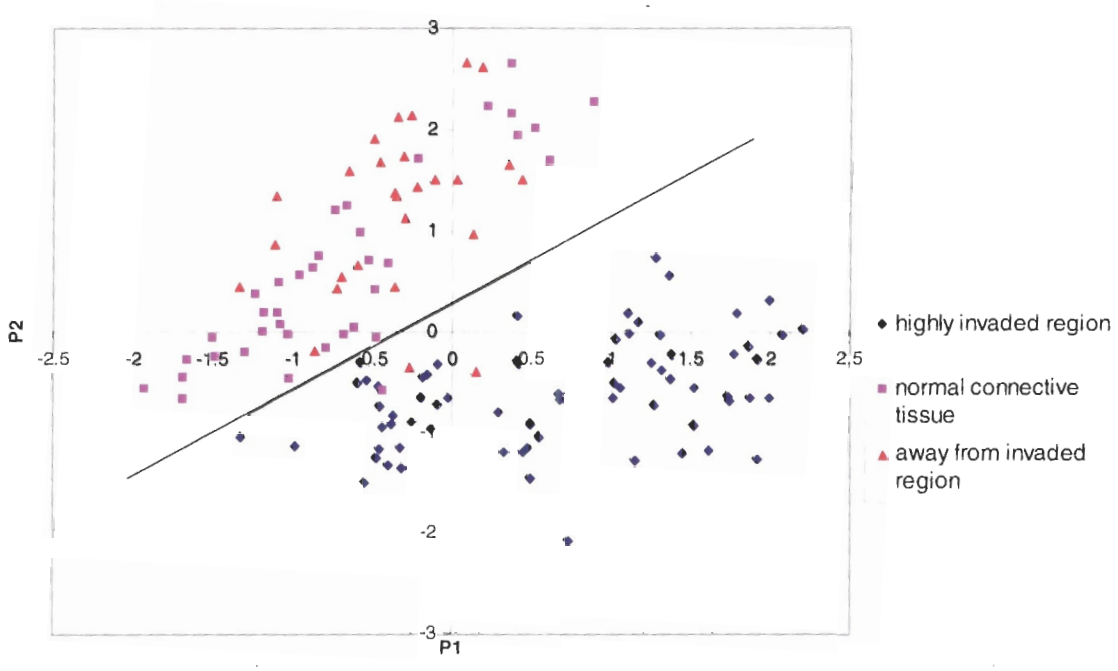

Figure 8.8: Plot of P1 versus P2. As can be seen, drawing a line along P2 = (0.872)P1 + 0.286 results in a division that separates the invaded tissue spectra from all but three noninvaded spectra, two of which are from the region outside of the invaded region in the cancerous sample. This provides evidence that this statistical method of classifying Raman spectra may prove useful in identifying a cancerous invasion in normal connective tissue.

Non-invaded tissue occupies the left hand, higher side, suggesting lower concentrations of the above tissue components, while having higher collagen

content. At first glance, one might worry about some of the invaded region spectra having lower P1 values than normal connective tissue, suggesting higher nucleic acid and fatty acid/lipid concentrations. However, recall the negative weighting seen in P2 for nucleic acids and lipids/fatty acids, allowing for our desired decrease in these tissue components for normal connective tissue. All in all, this plot is excellent, and the results are as expected based on P1 and P2. Spectra taken from the invaded tissue region are clustered together separate (save for a few spectra) from both connective tissue from the control sample, and the spectra taken from regions away from the invaded region in the tissue sample. As well, normal connective tissue and the region far away from the invaded sample cluster together quite well, suggesting similar molecular materials. This acts as a verification tool, allowing us to conclude that the two tissue regions studied are the same when far away from the invaded tissue region. Note that this clustering is not perfect, a few spectra from the connective tissue regions do appear to be closer to the invaded tissue type; this isn't much of a surprise due to the varying molecular nature of tissue in general. Drawing a line at P2 = (0.872)P1 + 0.286 results in 3 "positive" tests for invaded tissue, one from the connective tissue in the control sample (out of 34), and two from the tissue away from the invaded region (out of 27, suggesting slightly higher concentrations of cells). This is a very nice result, suggesting that using Raman spectroscopy (and this statistical method) can allow one to separate and identify invaded connective tissue from that of normal cancerous tissue.

Figure 8.9 shows the same CATPCA, including odd spectra and two regions bordering the highly invaded tissue region. As can be seen, these border regions tend to overlap both the invaded regions and the normal connective tissue regions. This would seem to suggest that these border regions show highly varying tissue, tissue that vary from being similar to invaded tissue to being similar to normal connective tissue (at least on the ten by ten micron scale studied, and as classified in our analysis). The first border region shows predominant similarities with the highly invaded tissue region, save for three

spectra that correlated with loose connective tissue. The second region showed more similarities with loose connective tissue than with a highly invaded tissue region. These results would seem to suggest ambiguity when trying to identify the type of tissue comprised of the bordering region (using CATPCA), though again this is not unexpected. While there undoubtedly may exist more subtle changes between regions bordering highly invaded connective tissue regions, said changes were not noticeable in our analysis.

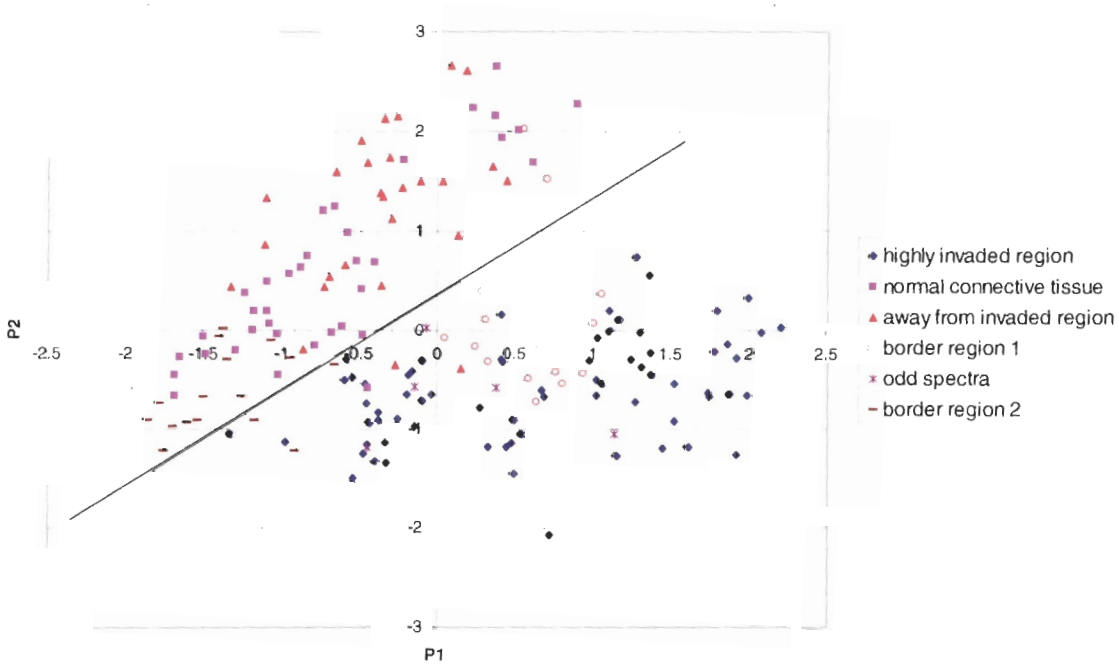

Figure 8.9: Plot of P1 versus p2, again for our raw data, showing other regions of interest. Regions bordering a highly invaded region of connective tissue show significant variation on the 10 by 10 scale, resulting in difficulty in classifying said regions as either invaded or not. This variation most likely is the result of decreased cell concentration in the region.

Figure 8.10 is a plot of P1 versus P2 (again for our raw data), used to denote clusters studied in the P1, P2 plot. The average of the spectra found in these clusters (given by the elliptical outlines) were taken and compared as a means to further identify differences, and see if the results agreed with our expectations from the weightings seen for P1 and P2. As well, we can study the

clusters closest to the border given by the separation line (again 0.872P1 + 0.286), to see how much of a difference actually exists between individual spectra found here.

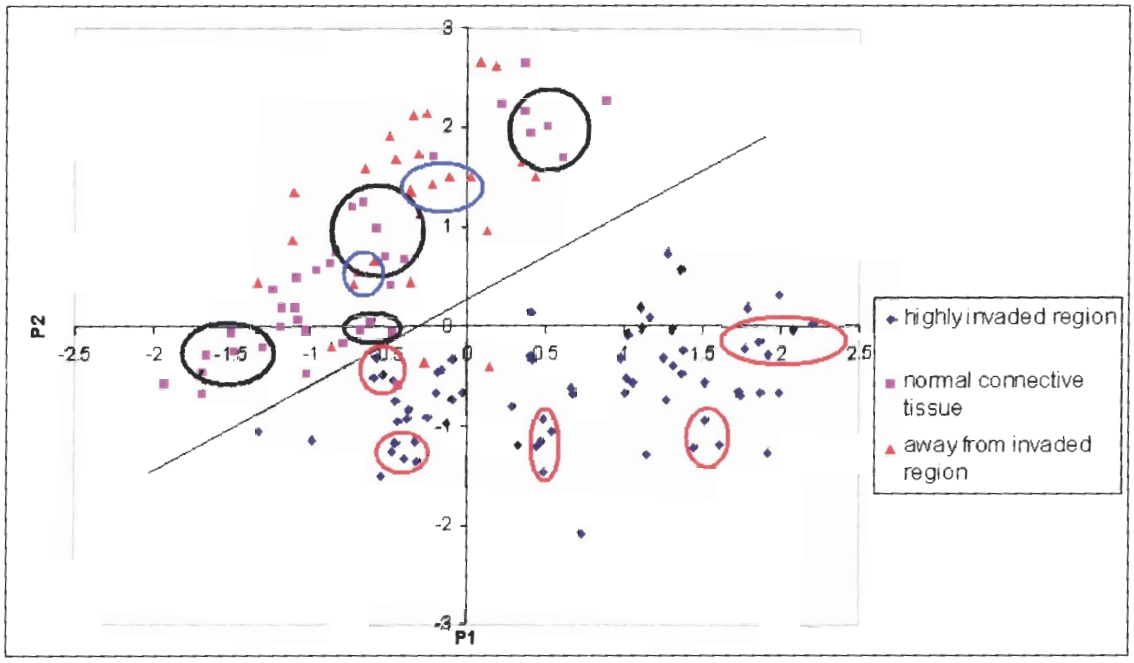

Figure 8.10: Plot of P1 versus P2, showing individual spectra clustered together, then averaged, as a means to compare regional data in the plot.

Figures 8.11- 8.14 denote comparisons of the averages of the spectra found in the aforementioned clusters. These cluster spectra were found to match in excellent agreement with each other, and taking the average without showing any given error was deemed acceptable.

Figure 8.11 is a plot of the averages for the clusters resulting from the invaded tissue spectra, and gives insight into how the spectra changes as we move along P1 and P2. We see that nucleic acid concentration and lipid/fatty acid concentration goes down while P2 increases, though at a smaller rate than when moving along P1. This is as expected from the component weightings, as

P1 has a significantly higher positive weight contribution for these peaks than does P2 a negative weight contribution. The subtle change in the spectral

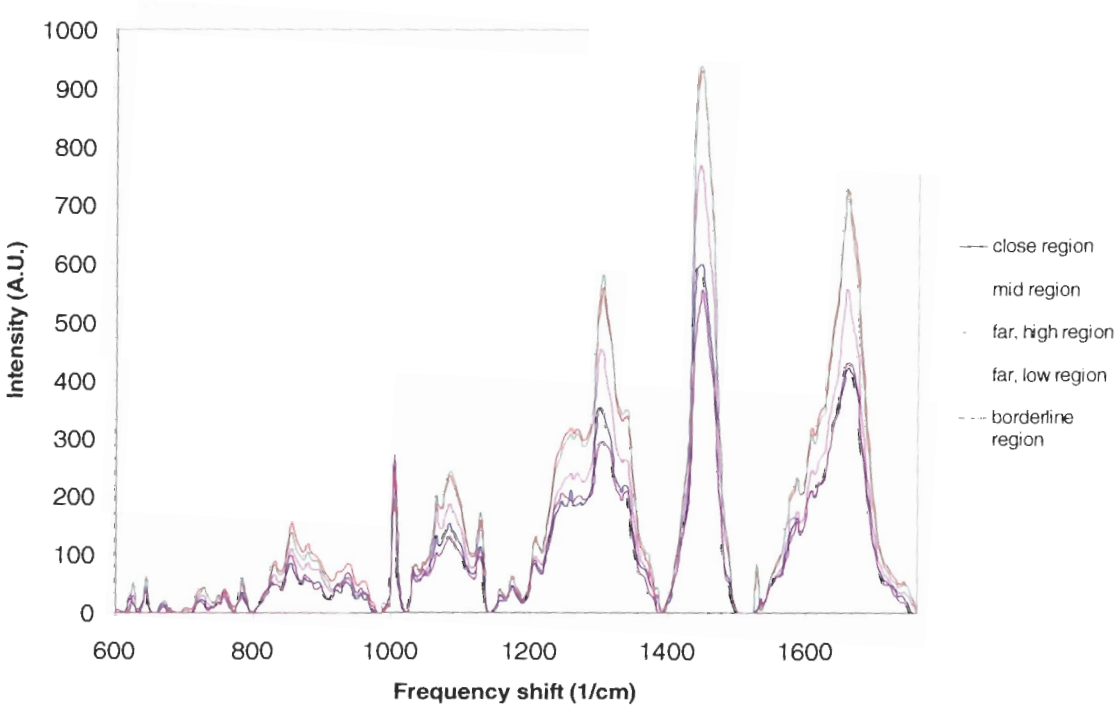

Figure 8.11: Plot of the averaged spectra for the given cluster region for the invaded tissue region. Note the increase in fatty acids/lipids and nucleic acids as P1 increases and P2 decreases. Also note the shape of the spectra in the 1225 – 1290 cm^{-1} range; the plateau like shape found near the "border" of the separated connective tissue and invaded tissue measurements suggest a lower relative concentration of fatty acids/lipids with respect to protein concentrations. The far, low region shows the highest amount of nucleic acids, while the far high region potentially shows the highest protein content. Whether this is degraded collagen or due to some other protein concentration was not verified.

signature seen between 1225 cm – 1290 cm^{-1} (the region is more curved and less of a plateau as P1 increases) also suggests an increase in fatty acids (the plateau shape seen in the 'boundary' region is more indicative of proteins, including collagen) as P1 increases. The relatively low intensity peak seen at 940 cm^{-1} when compared to the peak seen at 850 cm^{-1} seems to suggest that the protein concentration seen in this region is not in fact just due to collagen, but due to other protein concentrations as well (or potentially to some form of degraded collagen).

It does appear that the potential exists for the spectra to be quite equivalent if scaled to a given peak; though a closer look would suggest that the nucleic acid concentrations would still be different. The low, far region for example, shows a slightly higher peak at 788 cm^{-1} than does the far low region, which shows higher peaks between 825 – 975 cm^{-1}. It is difficult to say for sure however; again, this will be studied in the next section when we normalize the data to 1449 cm^{-1}.

Looking at figure 8.11, it would seem (if the overall intensities of the given individual spectra are equal) that spectra found at higher values of P1 (and lower values of P2 to a smaller degree) probably consist of measurements taken of 10 micron by 10 micron regions with slightly higher overall concentration of cell nuclei, and potentially cells as a whole in general (signifying higher nucleic acid and fatty acid/lipid concentration). Regions closer to the border line between the tissue types may consist of 10 micron by 10 micron regions not directly upon a cell or cell nuclei, thereby showing less concentrations of these given materials and potentially more concentration of collagen or other proteins such as elastin.

Figure 8.12 denotes the clusters for spectra resulting from normal connective tissue and tissue away from the invaded region. We can clearly see that for high values of P2, a significant dip occurs at approximately 911 cm^{-1}, suggesting the presence of high concentrations of collagen (when presented alongside the large peaks seen at approximately 850 cm^{-1} and 940 cm^{-1}). As well, note the rather small concentrations of nucleic acids and lipids/fatty acids, suggesting low concentrations of cells. We can see protein levels increase as P2 increases, with the spectral signature showing clear signs of high levels of collagen, levels that lower as P2 lowers. Note also that the peaks seen at 1301 cm^{-1} and 1449 cm^{-1} scale differently when P2 is varied, an effect that is most noticeable when looking at the border region and comparing it to the region taken from the middle of the normal connective tissue spectra. Thus it appears that fatty acid/lipid concentration (denoted by the 1301 cm^{-1} peak) decreases as our

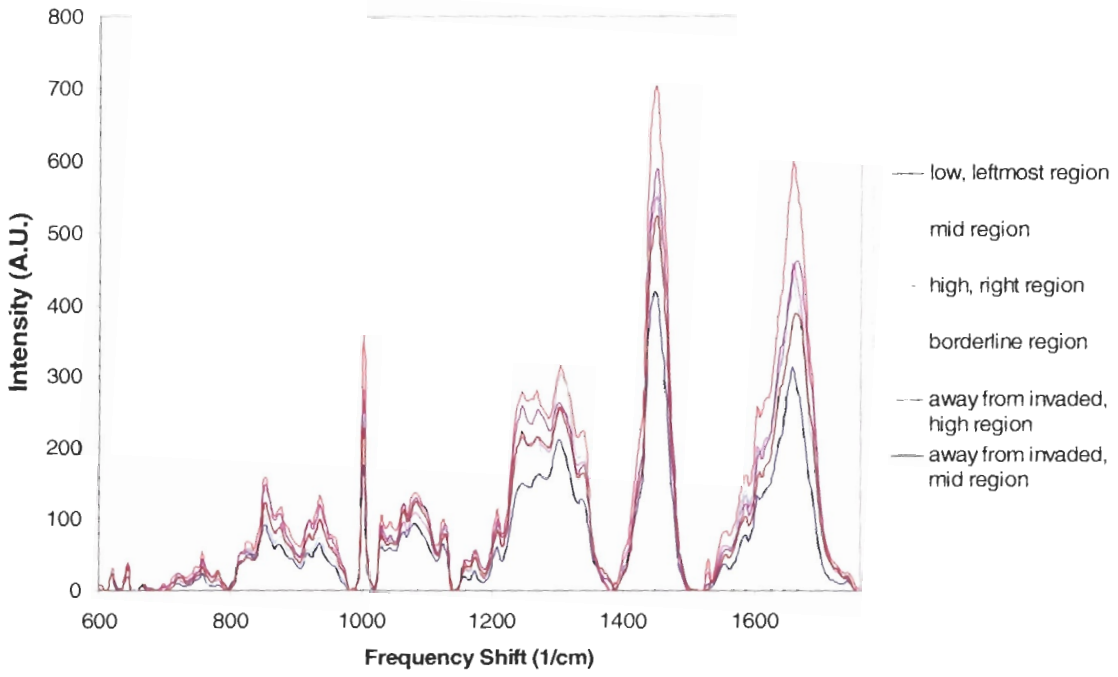

Figure 8.12: Plot of the averaged spectra for the given cluster region. Note the high collagen concentrations for regions with high P2 values, as shown by the large peak at 940 cm^{-1}. Also note the shape of the spectra between 1225-1290 cm^{-1}, also signifying high protein concentrations.

protein (most notably collagen) increases as P2 increases. Again, the region between 1225 – 1290 cm^{-1} also suggests a high protein concentration, due to its flatter signature (and more intense signal as compared to the fatty acid peak of 1301 cm^{-1}), as opposed to the curved shape it would have with a high fatty acid/lipid content.

As well, we can note that fatty acid/lipid concentration is highest at the "border" cluster region (given by the aforementioned line at 0.872P1 + 0.286), while collagen concentration appears to be lowest at this cluster. These results can be seen by studying the peaks at 1301 cm^{-1} and the region between 800-1000 cm^{-1}.

Figure 8.13 shows a comparison of the clusters found near the border line for the three main regions sampled – normal connective tissue, the highly invaded region in the cancerous sample, and connective tissue found away from the invaded region in the cancerous sample. It is interesting to study this area, and see if we can denote a potential means by which to quickly ascertain what tissue type we are looking at for spectra close to this border. As can be seen, the spectra are quite similar; the connective tissue region away

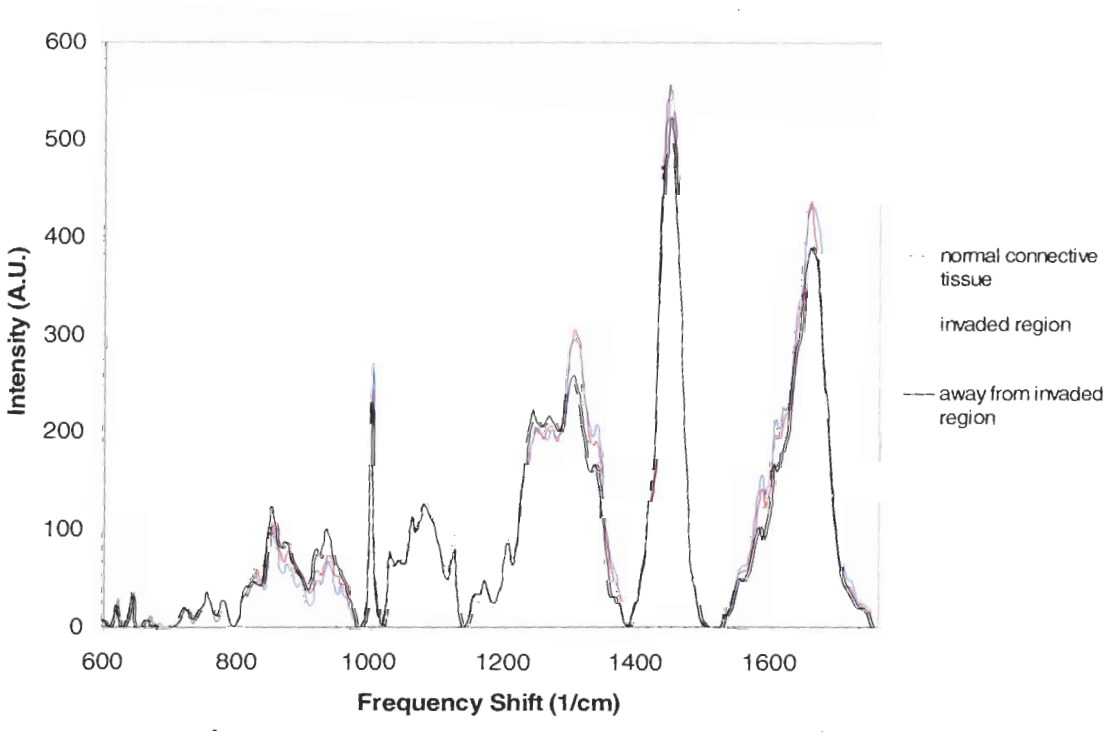

Figure 8.13: Plot of the averaged spectra for the given cluster region found near the border, as given by the aforementioned dividing line at 0.872P1 + 0.286. Note the almost equivalent spectra for the invaded tissue region and the normal connective tissue region; the spectra taken from the cancerous sample away from the invaded region show higher collagen levels (and less fatty acids/lipids), but is also farther away from the border line.

from the invaded region on the cancerous sample shows higher collagen levels, but is a bit farther away from the border line. The averaged spectra from the connective tissue of our control sample shows what appears to be higher levels of collagen levels than the invaded region sample (judging by the 800-1000 cm^{-1} region), while our invaded region shows slightly higher concentrations of nucleic

acids. The invaded region may also show slightly higher levels of non-collagen proteins, as suggested by the slightly higher phenylalanine content seen (resulting from higher peaks at 1003 cm^{-1} and 623 cm^{-1}). Again, we should note that all these differences are extremely subtle shifts in relative concentrations, and the spectra are quite similar. Any final conclusion resulting from this border region is difficult to say at best, simply because the clusters only consist of 4-5 spectra each, and since the separation distance between the clusters is so small, the actual positions of the individual spectra may play a significant role in affecting the average of the cluster, as the individual spectra are not spread out evenly in the clusters (as labelled in figure 8.10).

If we were to judge by these slight differences in concentration, it appears that the main spectral regions to study at this border would be the relative concentrations of nucleic acids to that of collagen, particularly the peaks at 788 cm^{-1} (nucleic acids) and 940 cm^{-1} (collagen). By looking solely at these relative concentrations we should be able to ascertain if a spectra result from normal connective tissue or from tissue that has been invaded by neoplastic cells. One would think we could also use lipids/fatty acids concentration as compared to collagen concentration to compare as well, though figure 8.13 provides no insight into this, as the lipid/fatty acids concentrations are again almost equivalent.

Figure 8.14, on the other hand, shows huge differences in the spectral signature of the compared clusters, that of the two regions farthest away from each other along P1. As can be seen, the invaded region cluster shows a more intense spectra signal for all peaks, save perhaps that at 940 cm^{-1}. The nucleic acid concentration of the our normal connective tissue averaged spectra is nonexistent;, as well the shape of the spectra between 1225-1290 cm^{-1} suggests the predominant contribution is due to protein (the relatively smallish peak at 1301 cm^{-1} suggesting low fatty acid/lipid concentration also verifies this). The normal connective tissue spectra may show a lower intense signal overall as a result of a slightly out of focus microscope, or it may indeed consist of a weaker

signal resulting from a lower concentration of material as a whole. However, the fact that the intensities appear almost equivalent at a few places, such as the collagen indicative peak at 940 cm^{-1}, prescribes to the fact that significant differences in relative concentrations of the given measured tissue types (proteins, collagen, fatty acids/lipids, nucleic acids) do occur. Looking at these spectra separated by large values of P1 suggest that at these extreme, the differences are extremely significant, and can be used to very easily identify the type of tissue studied.

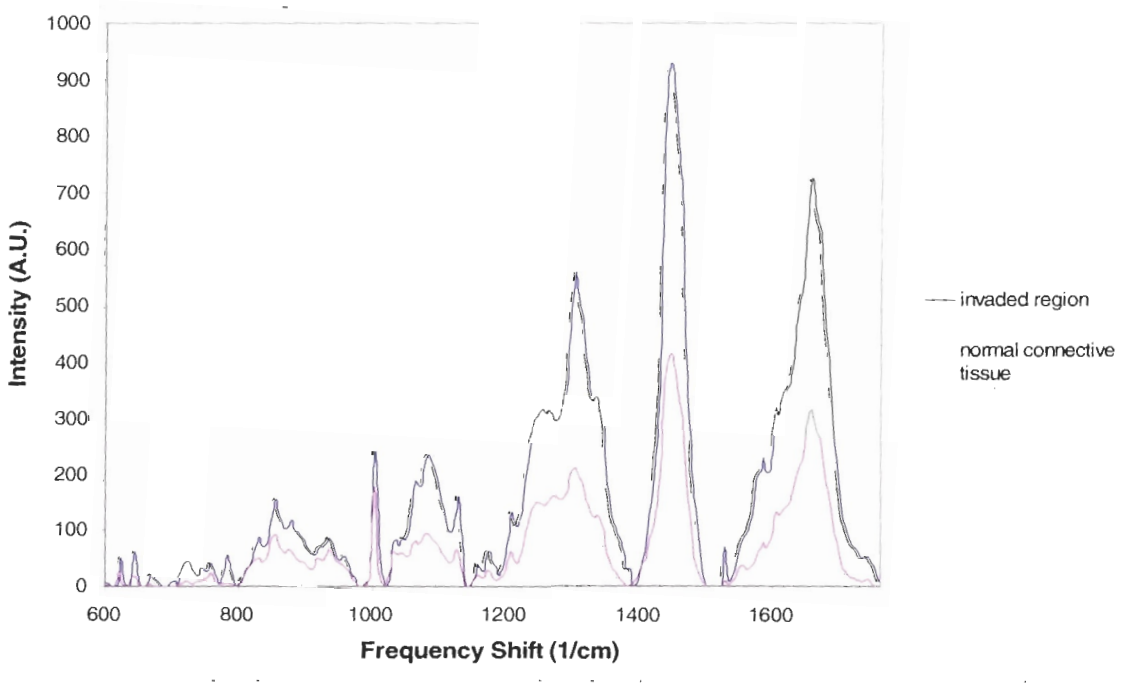

Figure 8.14: Plot of the averaged spectra for the two cluster regions at the extremes of P1. Note that while the normal connective tissue cluster shows what appears to be a less intense spectra overall, significant spectral differences occur to allow one to conclude that the relative concentrations of the regions sampled are quite different.

8.5 Results of CATPCA Performed on Normalised Data

As we've stated, one issue that can cause concern is spectra from regions with potentially the same compositional makeup, yet show (for whatever reason) a less intense spectra over all, possibly due to an out of focus microscope, or merely less of a concentration of material at a given point. While the above data may not necessarily be affected by these possibilities, it is not immediately clear, nor is there a viable way (at least for ourselves) to ascertain this. As another means of comparison then, we normalized our data to a given spectral peak, to allow for a means of comparing spectral signatures. Not only does this remove our worry of varying overall signal intensity, but we can use this normalization to act as a sort of standard among other measurements. If all our spectra are normalized to a given peak (whatever that peak may be), then we'd expect spectra from equivalent materials to give equivalent spectral signature.

Normalisation was done to the peak seen at 1449 cm^{-1} (chosen because it's the largest peak, and small inaccuracies will result in less of a percentage error of the overall peak size). We should note normalising to this peak resulted in the invaded tissue region spectra usually being lowered in intensity, while all normal connective tissue spectra had their overall intensity raised, as no connective tissue spectra had a peak at 1449 cm^{-1} as high as 800 arbitrary units. We also need to remember this peak undoubtedly varies in intensity over our given spectra (depending on concentrations of proteins and lipids/fatty acids), so *no* comparison of absolute values of individual tissue type concentrations should be made over differing tissue types (or tissue that results in differing spectral signatures). To do otherwise could result in erroneous comparisons, such as suggesting higher nucleic acid concentrations in one tissue type when in fact the opposite is true (e.g. for tissue types showing moderate nucleic acid concentrations but a small peak at 1449 cm^{-1}, as compared to a tissue type with a slightly higher nucleic acid concentration but significantly higher peak at 1449

cm^{-1}). This method does not allow one to compare concentrations of the given chemical components with respect to other tissue types, but it does allow one to study the spectral signature of the differing tissues and compare.

In this section we will save time by looking at the actual plots of P1 versus P2, and ignore the plots of P1 and P2 versus wave number, as all the required data from said plots can be found by said P1 versus P2 plot.

Figure 8.15 is a plot of P1 versus P2 for our spectral data normalized to the 1449 cm^{-1} peak (excluding regions on the boundary of the connective tissue found to be highly invaded by neoplastic cells). As can be seen, we again have excellent separation between spectra taken from normal connective tissue (as measured in the control sample) and connective tissue in the cancer sample away from our measured highly invaded region, and spectra taken from our invaded region in the cancerous sample. The invaded region is found at lower values of P1, suggesting higher concentrations of nucleic acids and lipids/fatty acids, while our connective tissue spectra are found to occur at higher values of P1, suggesting higher collagen content as P1 increases.

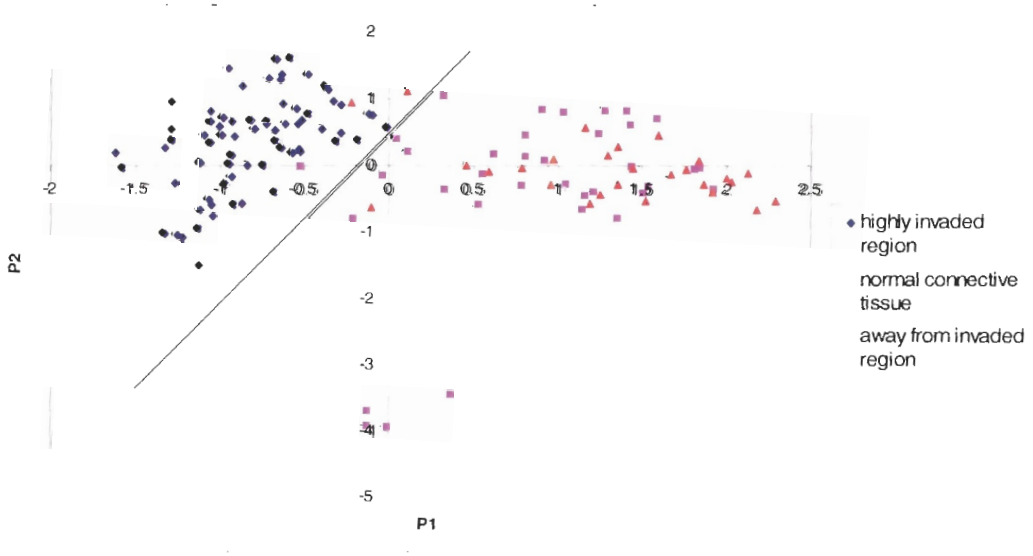

Figure 8.15: Plot of P1 versus P2 for our normalized data. Note the border line drawn at P2 = (0.254)P1+0.44 does an almost perfect job of separating our invaded tissue region spectra from our unaltered connective tissue spectra.

We also see that connective tissue away from the invaded region clusters quite well with the normal connective tissue region from our control sample, again suggesting equivalent makeup in tissue composition (as compared to our invaded region). Again, if we were to draw a border line at P2 = (0.254)P1+0.44, we see that only one spectra from our control sample and two spectra from connective tissue away from the invaded region would exist on the "invaded region" side of the border. These individual spectra are the same ones as found to be on the invaded region side of the border for our raw data plot, found in figure 8.8. Again, this result does suggest that the tissue away from the invaded region may show slightly higher signs of cellular density, at least in one of the regions studied away from the invaded region (both spectra found on the invaded region side of the border came from the same measured region). We can also see a large cluster of normal connective tissue for a low value of P2; these spectra will be studied later. Another interesting point to note is how the majority of the normal connective tissue regions spectra are quite far from the "border line", suggesting significant differences in the spectral signatures of the two separated regions.

The result of this plot and the ability to discern our invaded region from normal connective tissue regions suggest that the spectral signatures of the two tissue types are in fact different, even when normalized to a given peak. This gives further evidence that the tissue composition of the two regions is different, a difference that can be accurately identified by Raman spectroscopy.

Figure 8.16 is again a plot of P1 versus P2 for our normalized data, including spectra taken from regions at the supposed edge of the highly invaded region. Again, as seen before in figure 8.9, we see that some of the spectra in these boundary regions cluster with the invaded region, while others cluster with connective tissue, suggesting a sort of "transition" between the two tissue types.

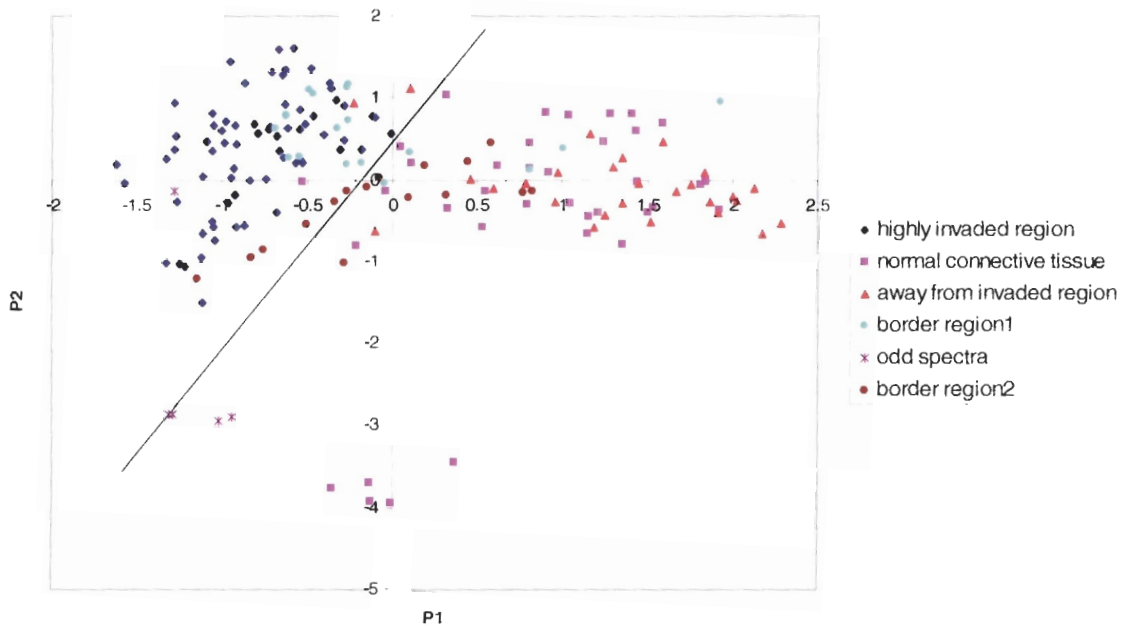

Figure 8.16: Plot of P1 versus P2 for our normalized data, also showing spectra taken from regions near the boundary of the highly invaded region, as seen on the stained and unstained samples. Note how these boundary regions show spectra that cluster with both the invaded tissue and connective tissue, suggesting highly varying regions.

Figure 8.17 shows individual spectra clustered together, with the average of each cluster being taken as a means to compare to spectra found in other regions of the P1 - P2 plot.

Figure 8.18 shows the average spectra taken from the given clusters found from the invaded tissue region. As can be seen, the variations are quite subtle, suggesting a significantly less total change in the intensity of a given spectral peak over the entire data set. In other words, it appears that by normalizing our data to the peak at 1449 cm^{-1}, we have significantly reduced the amount of variation seen for any given peak. This is why we see equivalent magnitudes of weighting in our principle components, yet see less overall variation over the same change in P1. Indeed, the spectra in figure 8.18 seem quite equivalent, with the regions closer to the border (again as given by the line P2 = (0.254)P1+0.44) showing slightly higher levels of protein (possibly collagen) and lower concentrations of fatty acids/lipids and nucleic acids.

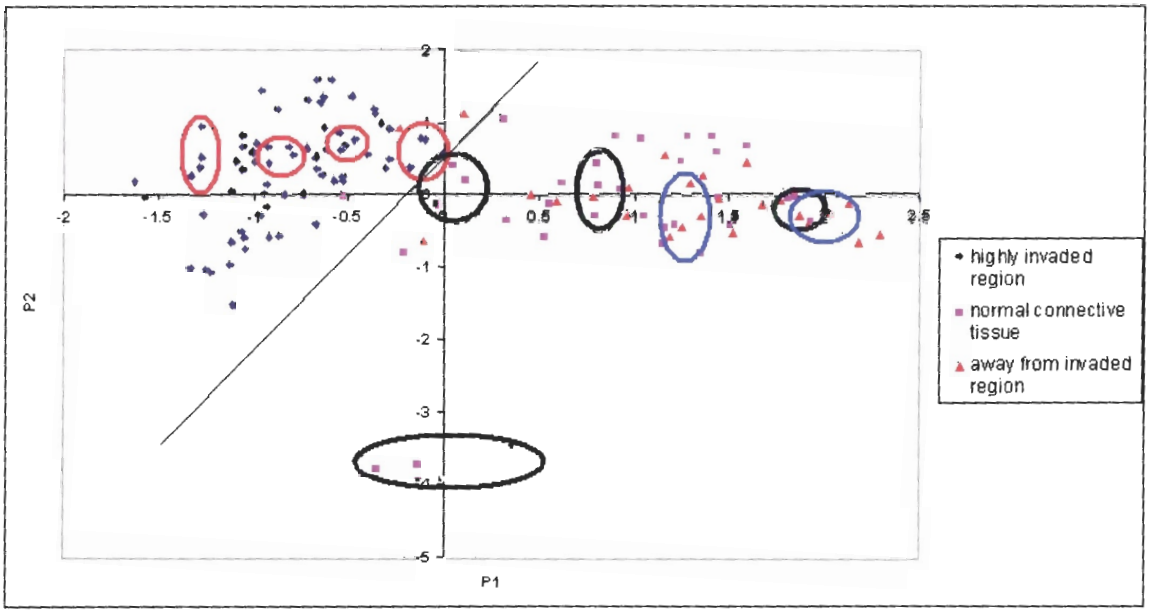

Fig 8.17: P1 – P2 plot showing regions clustered together (with the average spectra taken) as a means to compare to other regions on the P1 – P2 plot.

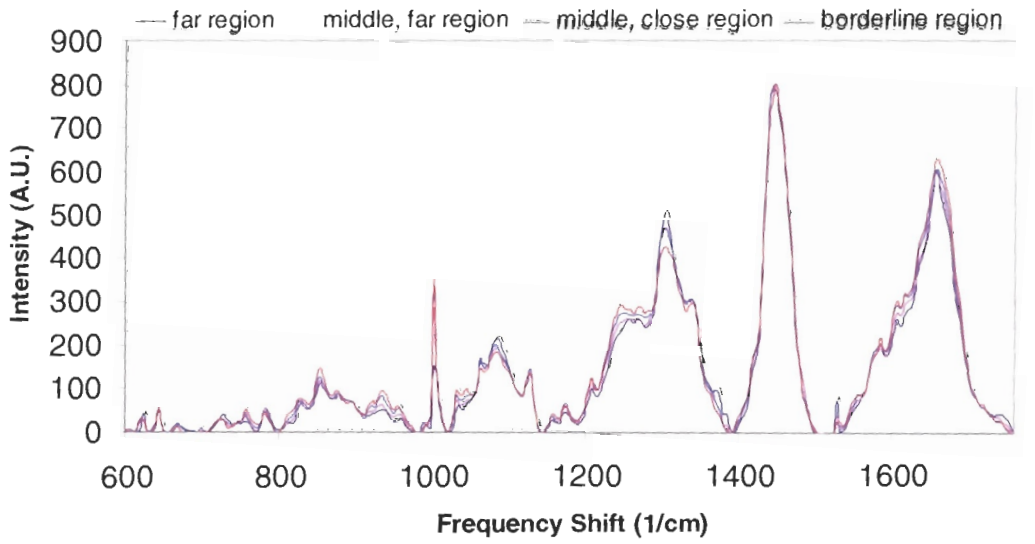

Fig 8.18: Plot of the average spectra taken from the clusters outlined in figure 8.19 for the invaded tissue region. Note that protein concentration (most probably collagen, as given by the region between 900 - 975 cm^{-1}, particularly the peak at approximately 940 cm^{-1}) increases as P1 increases (closer to the border line), while fatty acid/lipid concentration decreases as P1 increases. Nucleic acid concentration decreases slightly as P1 increases.

Figure 8.19 shows the average spectra taken from the given clusters found in the connective tissue region, consisting of spectra from our control sample. Spectra taken from connective tissue away from the invaded region of our cancerous sample were found to be equivalent, and were not included to allow easier viewing of the graph. Looking at the averaged spectra, we see significantly high levels of collagen concentration, while having lower levels of fatty acids/lipids, *save* for the lower cluster, which shows quite high fatty acids/lipid content. This region, clustered for the most part quite far away from the rest of the normal connective tissue regions, was actually taken near the boundary of the connective tissue region, and may show higher lipid content as a result of the next layer of tissue (the spectra in this cluster were all taken from the same region of our connective tissue sample). Otherwise, our results are as we expected, low concentration values of nucleic acids and fatty acids/lipids (getting lower as P1 increases), and high concentrations of collagen (again, increasing as P1 increases). In fact, the cluster from the far region shows a spectra consisting predominantly of collagen and other proteins, showing next to zero nucleic acid and low fatty acid/lipid concentrations. Again, recall that most of the spectra taken from the connective tissue regions were rather far from the border line given at, resulting in these spectra showing significant differences to the spectra taken from the invaded region of our cancerous sample.

Figure 8.20 shows the two clusters found near the border line, comparing the average spectra of each cluster. Again, as found in figure 8.13 for the same plot for the average spectra of the border regions for our raw data, we see that there is not much difference in the two spectra. We appear to note slightly higher nucleic acid concentrations in the invaded tissue spectra, while the normal connective tissue region shows slightly higher concentrations of protein, most likely collagen.

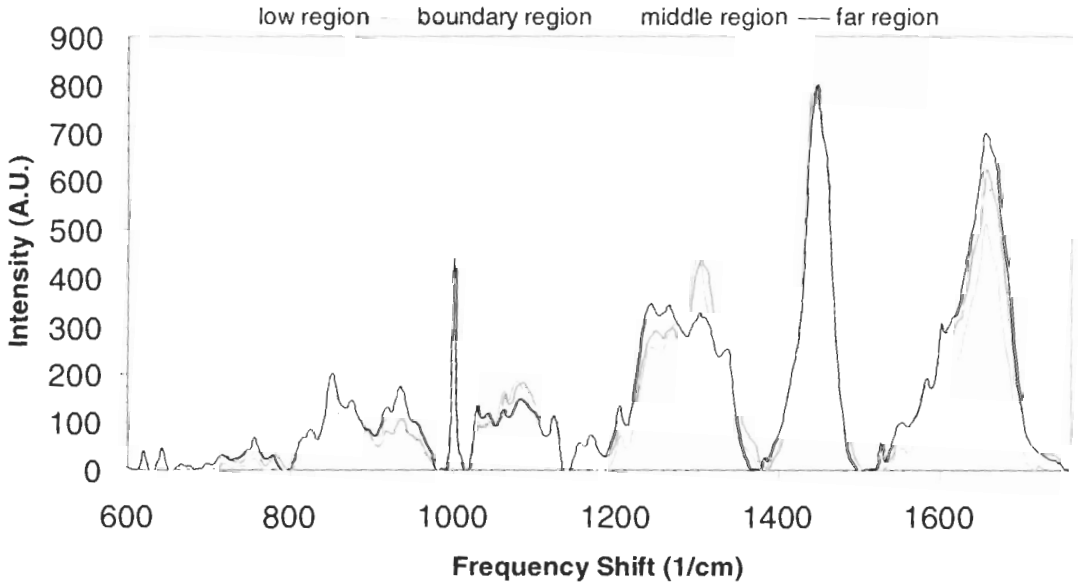

Figure 8.19: Plot of the average spectra take from various regions in the P1 – P2 plot for normal connective tissue from our control sample. Note the very low nucleic acid concentrations (the peak at 788 cm^{-1} looks higher resulting from a larger background due to the collagen peak seen at 760 cm^{-1}), low fatty acid/lipid concentrations, and high collagen concentrations.

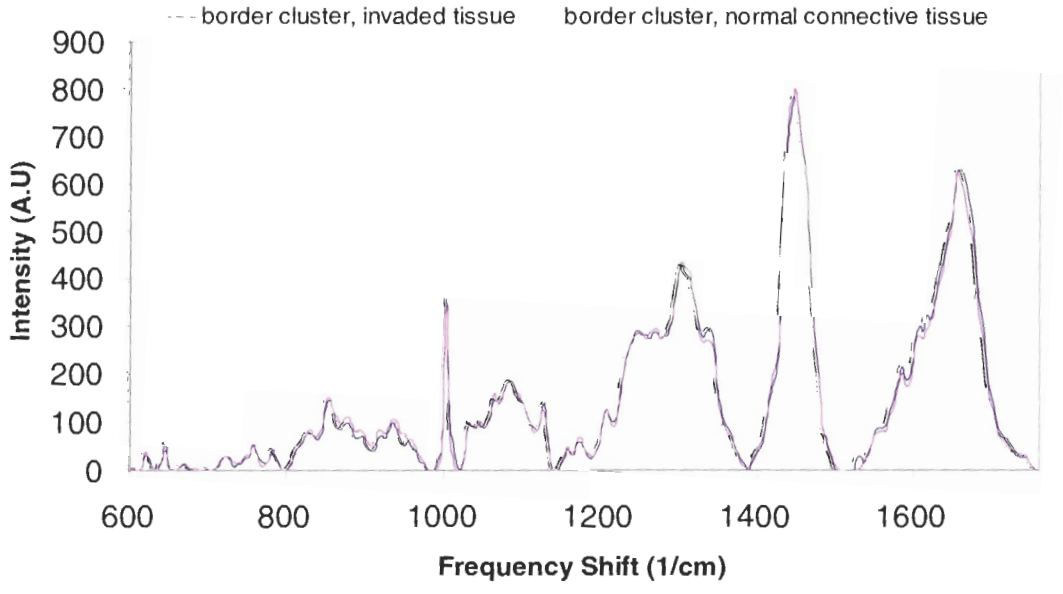

Figure 8.20: Plot of the averaged spectra taken for regions found near the border line given by P2 = (0.254)P1+0.44.

Figure 8.21 shows the average spectra from the clusters found at the two extremes of P1, as well as the average spectra taken from the invaded tissue cluster found near the border (for comparison purposes). While it was not as vital above, when comparing differing tissue types we again need to remember that the spectra we are studying here have been normalized to a given peak, so absolute values of other given peaks should not be compared; rather we should only look at the spectral signature of each tissue type as a whole. Otherwise we run the risk of coming to seriously erroneous conclusions, especially when we consider the fact that connective tissue usually showed lower concentrations of lipids/fatty acids, resulting in a correspondingly lower peak at 1449 cm^{-1}.

Keeping the above in mind, this plot shows that the relative collagen concentration is higher than the relative fatty acid/lipid concentration in connective tissue, as compared to our invaded tissue region. As well, when we compared nucleic acid concentration to collagen concentration (788 cm^{-1} to the peak at 940 cm^{-1}), we see that we again have a higher relative nucleic acid concentration in our invaded tissue region than in our normal connective tissue region. The connective tissue spectra here has a bit of a higher spectral background (the result of the falling edge of the peak at 760 cm^{-1}, a peak due to protein) between 700 - 800 cm^{-1}, which makes it more difficult to ascertain the differences in the peaks at 728 cm^{-1} and 788 cm^{-1} (peaks due to nucleic acids) with this plot. However, one can look at the peak at 1340 cm^{-1}, again due to nucleic acids, to again come to this conclusion. Figure 8.22 zooms in on the region between 700 cm^{-1} and 1000 cm^{-1} to provide further proof that our relative nucleic acid concentration to collagen concentration is higher in our invaded tissue sample. Indeed, this plot suggests that it may be possible to classify tissue types based solely on the ratio of the intensities of the two peaks at 788 cm^{-1} to 940 cm^{-1}; for a given ratio value and higher we may have invaded tissue (higher nucleic acid concentration to collagen concentration), while for a lower ratio value we can classify the sample as normal connective tissue.

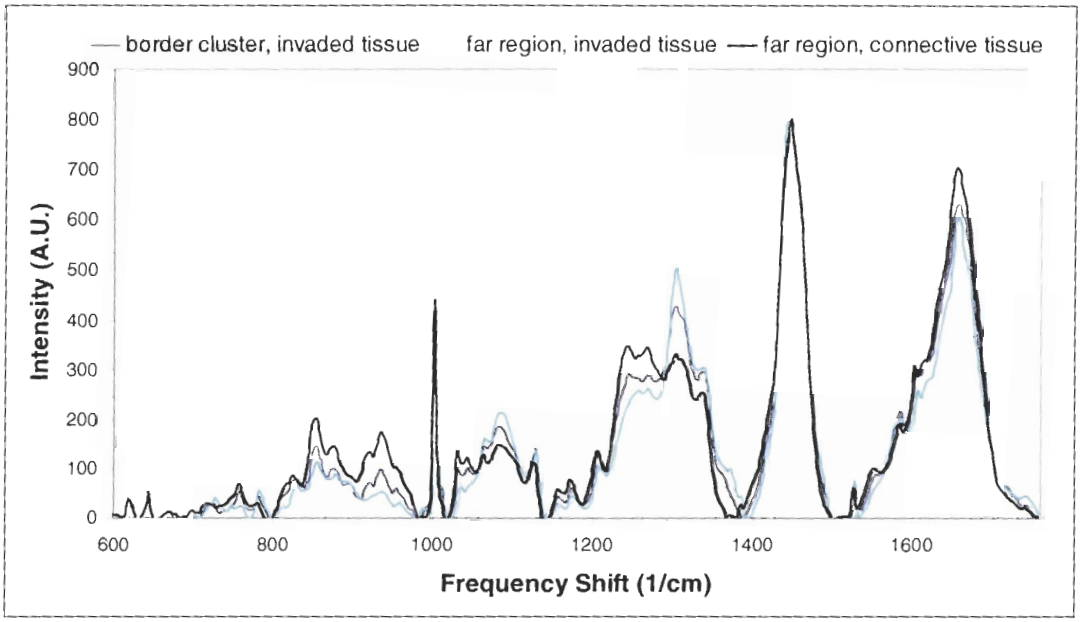

Figure 8.21: Plot of the averaged spectra for the cluster regions at the extreme ends of P1, as well as the cluster region found near the border line for invaded tissue. Looking at relative concentrations of given materials in each tissue type, it appears that we have a higher nucleic acid and fatty acid/lipid concentration in our invaded tissue region than we do in the connective tissue region of our control sample.

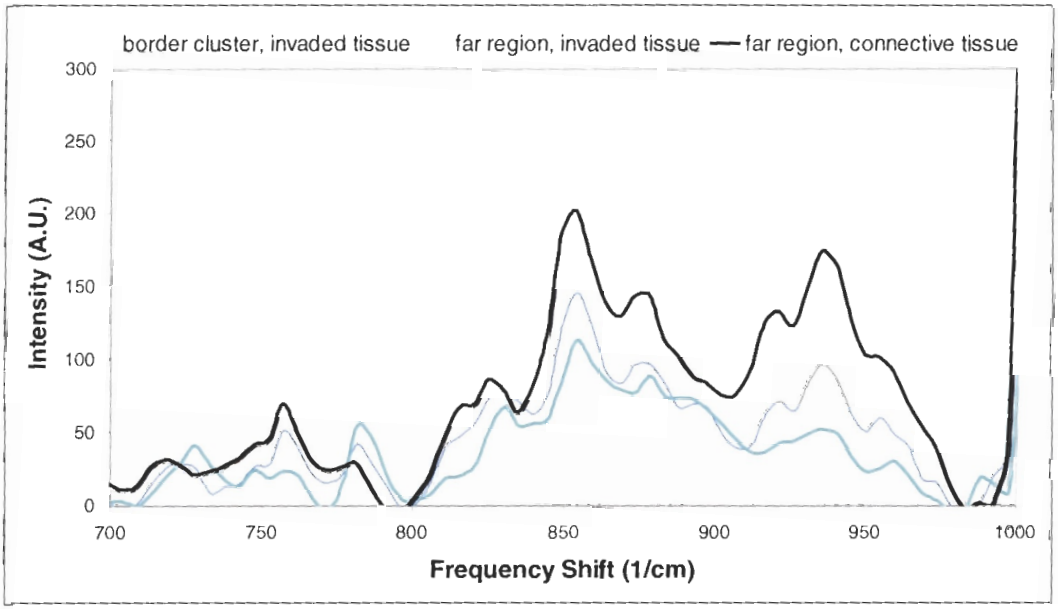

Figure 8.22: Plot of the spectra from the given clusters for the invaded tissue regions as compared to the normal connective tissue region. Note the higher peaks seen at approximately 728 and 788 cm^{-1} for the invaded region (as compared to the smaller bumps on a *higher background* signal for connective tissue), as compared to the lower peak seen at approximately 940 cm^{-1} for our invaded tissue region. This suggests a much higher relative concentration of nucleic acids (as compared to collagen) for the invaded tissue region.

The results of the analysis of our data when normalized to a given peak seem to suggest that one can use a peak as a sort of standard when comparing spectra taken from either differing regions, or potentially different machines. By normalizing to the given peak, we can accurately identify differing spectral signatures, suggesting different relative concentrations of materials in the given regions sampled. We don't have to worry about overall intensity issues (or varying sample times for that matter); rather we are looking only at differences in relative concentrations of the materials making up the regions sampled. In other words, we are looking at the spectral signature of the given tissue sample, nothing more. On the other hand, one should not discount the more intense overall spectra that can result from materials with a higher density, as this in itself can also be useful as a means of classification. This should prove invaluable when attempting to classify a given tissue type as either invaded or non-invaded.

9 Conclusion

9.1 Summary

All of the results seen in chapters 6 and 7 give rise to the conclusion that Raman spectroscopy indeed is a viable means to identify normal connective tissue from tissue that has been significantly invaded by cancerous cells (at least when studying said samples at the micron level). Our analysis indicated that the invaded tissue region show higher concentrations of nucleic acids and fatty acids/lipids (and lower collagen concentrations) than the connective tissue found in our control sample, as was expected based on the knowledge that these constituents are more prevalent as cell density increases in the given tissue region studied. As well, the differences seen were prominent enough to allow for separation and classification of our Raman spectra when using categorical principle component analysis, resulting in the clustering of our invaded tissue away from our normal connective tissue when only looking at the first two principle components. Further analysis of the actual relative concentrations of the given materials in the two tissue types (after normalization to the peak seen at 1449 cm^{-1}) suggest that classification may even be possible by comparing only the actual intensities of a few peaks, say the ratio of the nucleic acid peak at 788 cm^{-1} to the collagen dependent peak at 940 cm^{-1}. For a given ratio of nucleic acid to collagen (based on the intensities of these peaks), we may be able to stipulate whether the region sampled is unaltered connective tissue or tissue that has been altered by the presence of neoplastic cells. As well, these studies of the normalized spectra allow us to look solely at the differing Raman signatures of the two tissue types, letting us ignore potential variations in the actual overall intensity of the spectra seen, variations that may arise from out of focus optics, different machines or methods of Raman spectroscopy, etc.

In all of the results seen in the previous two chapters, the most prominent distinction between the tissue types has without a doubt been the increase in cell

density in the invaded tissue sample. This increase in cell density leads to higher nucleic acid and fatty acid/lipid concentrations (and the lower collagen concentrations, resulting in the significant changes in the spectral signatures of the two tissue types (more subtle results were not immediately recognizable). Studies of these three tissue constituents (collagen, nucleic acids, and lipids/fatty acids) appear to be all that is required in an attempt to ascertain the tissue type sampled through the use of Raman spectroscopy.

Appendix

Raman Spectra of Notable Tissue Components

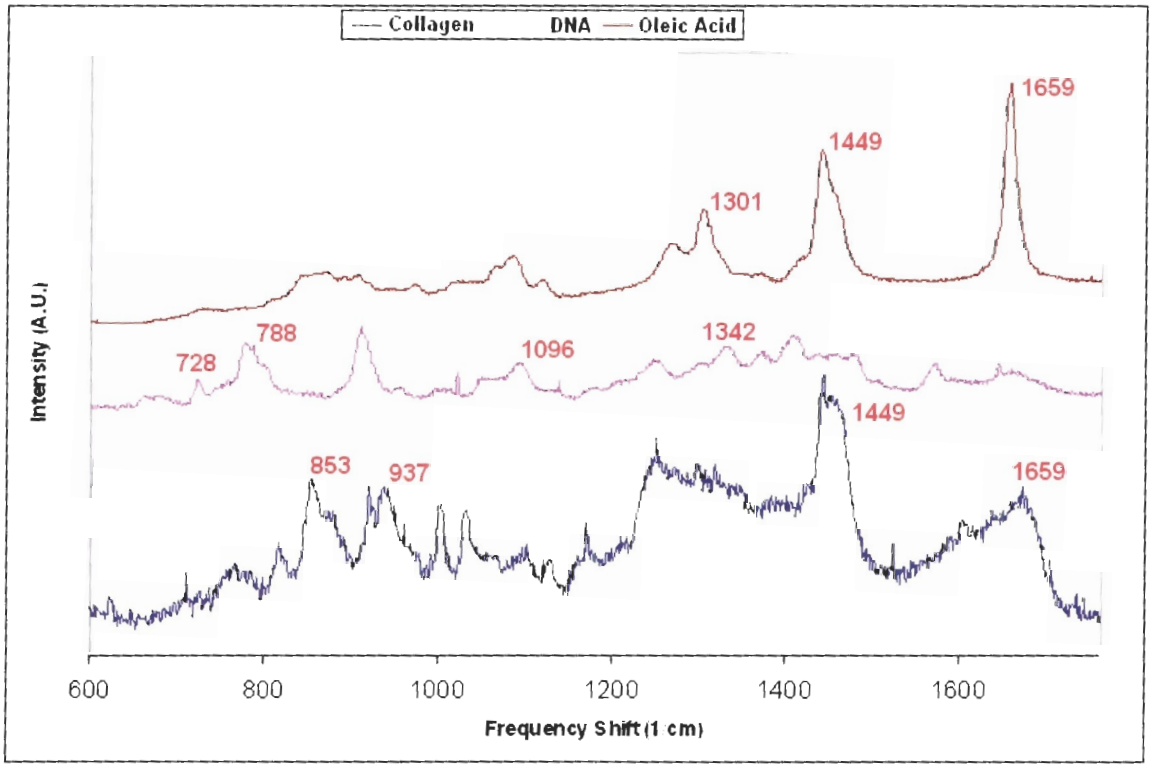

Figure A.1: plot of the Raman Spectra of DNA (pink line), human collagen (blue line), and oleic acid (brown). These plots show the significant peaks of the given tissue component, such as the large peaks at 728 cm^{-1} and 788 cm^{-1} for DNA, and the peaks at 856 cm^{-1} and 937 cm^{-1} for collagen (note for the DNA spectrum, the peak seen at approximately 900 cm^{-1} is due to the processing of the DNA, and is not a peak inherent to DNA). The spectra are not plotted to scale with respect to each other.

References

1. Z. Huang, A McWilliams, H. Lui, D McLean, S. Lam, H. Zeng, *Near-infrared spectroscopy for optical diagnosis of lung cancer, International Journal of Cancer*, Vol. 107 Issue 106, 1047-1052.

2. J. Javier, *an Introduction to Raman spectroscopy: Introduction and Basic Principles*,http://www.spectroscopynow.com/coi/cda/detail.cda?page=1&id=1882&type=EducationFeature&chId=6, 2005.

3. *Morse Potential*, Wikipedia.org, http://en.wikipedia.org/wiki/Morse_potential.

4. D. Hennig, J. Archilla, *Stretching and Relaxation Dynamics in Double Stranded DNA*, arxiv.org, 2003.

5. Shafee, Fariel, *Energy Level Sets for the Morse Potential,* http://arxiv.org/ftp/quant-ph/papers/0411/0411159.pdf.

6. *Carbon Dioxide*, Wikipedia.org, http://en.wikipedia.org/wiki/Carbon_dioxide.

7. *Raman Scattering*, Wikipedia.org, http://en.wikipedia.org/wiki/Raman_scattering.

8. *Raman Tutorial*, http://www.kosi.com/raman/resources/tutorial/index.html.

9. I. Notingher, S. Verrier, S. Haque, J.M. Polak, L.L. Hench, *Spectroscopic Study of Human Epithelial Cells in Culture: Living Cells versus Dead Cells*, BioSpectroscopy, Vol. 72, 230-240, 2002.

10. *Amino Acids*, Wikipedia.org, http://en.wikipedia.org/wiki/Amino_Acid.

11. *Beta Sheet*, Wikipedia.org, http://en.wikipedia.org/wiki/Beta_Sheet.

12. P. Duesberg, D. Rasnick, Aneuploidy, *The Somatic Mutation that Makes Cancer a Species of its Own*, Wiley-Liss, 2000.

13. *Blue Histology*, School of Anatomy and Human Biology, The University of Western Australia, http://www.lab.anhb.uwa.edu.au/mb140/.

14. A. Durkin, *Spectroscopy Systems and Spectroscopy Applications*, http://www.bli.uci.edu:3455/12/admin/download.html?attachid=17183, 2004.

15. G. Cox, W. Steward, K. O'Byrne, *The Plasmin Cascade and Matrix Metalloproteinases in non-small Cell Lung Cancer*, Thorax, Vol. 54, 1999.

16. D. Scheehan, B Hrapchak, *Theory and Practice of Histotechnology, Batelle Press*, 1980.

17. J. Bischof, *Quantitative Measurement and Prediction of Biophysical Response During Freezing in Tissue*, Annual Review of Biomedical Engineering, Vol. 2, 2000.

18. L. Smith, *A Tutorial on Principle Component Analysis*, http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf, 2002.

Bibliography

L.M. Franks, N.M. Teich, *Introduction to Cellular and Molecular Biology of Cancer*, Oxford Press, 1997.

P. Hewdra, C. Jones, G. Warres, *Fourier Transform Raman Spectroscopy*, Ellis Howard, 1990.

D. Long, *The Raman Effect*, J. Wiley, 2002.

A. Nijssen, T. Bakker Shut, F. Heule, P. Caspers, D. Hayes, M. Neumann, G. Puppels, *Discriminating Basal Cell Carcinoma from its Surrounding Tissue by Raman Spectroscopy, Society for Investigative Dermatology*, 2002.

G. Smith, G. Dent, *Modern Raman Spectroscopy*, J. Wiley, 2005.

J. Staal, *Raman Spectroscopy for Catalysis*, Van Norstrand Reinhold, 1990.

S. Verrier, I. Notingher, J.M. Polak, L.L. Hench, *In Situ Monitoring of Cell Death using Raman Microspectroscopy*, Wiley, 2003.