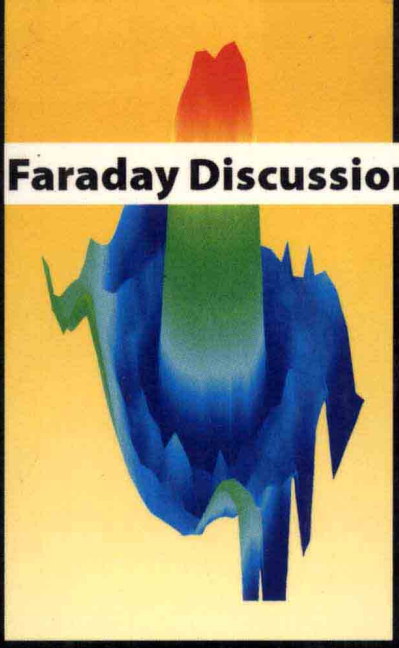


Faraday Discussions

RS•C

Volume 126
2004
www.rsc.org/faraday_d

Applications of Spectroscopy to Biomedical Problems

Applications of Spectroscopy to Biomedical Problems

University of Nottingham
September 1–3, 2003

FARADAY DISCUSSIONS

Volume 126, 2004

RS•C

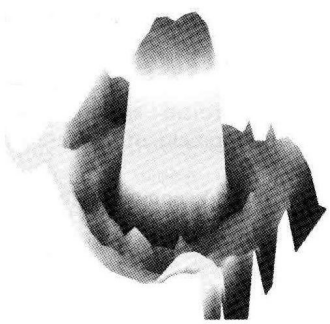
Organising Committee
Professor M. A. Chesters (*Chairman*)
Professor L. D. Barron
Dr P. Dumas
Dr L. R. Fisher
Dr M. W. George

ISBN: 0-85404-977-0
ISSN: 1359-6640
Typeset by Santype International Ltd., Netherhampton Road, Salisbury, Wiltshire and printed and bound in Great Britain by Black Bear Press, Cambridge, UK.

A General Discussion
on

Applications of Spectroscopy to Biomedical Problems

A General Discussion on Applications of Spectroscopy to Biomedical Problems, was held at the University of Nottingham, UK on 1st, 2nd and 3rd September 2003.



Cover
See P. Dumas, pp. 289–302.
Chemical image of lipid distribution across a human skin section with a hair embedded. This chemical image was obtained by using synchrotron infrared microscopy, with a 6×6 μm^2 aperture, over an area of 150×150 μm^2.

Image kindly supplied by Dr Paul Dumas, LURE–CNRS, Centre Universitaire Paris Sud, Orsay, France.

Contents

Introductory Lecture

From biomolecules to biodiagnostics: Spectroscopy does it all

Michael Jackson

Institute for Biodiagnostics, National Research Council Canada, 435 Ellice Ave., Winnipeg, Manitoba, Canada R3B 1Y6

Received 28th October 2003, Accepted 28th October 2003
First published as an Advance Article on the web 2nd December 2003

Optical spectroscopy and imaging have been used in medicine since medicine was first practised. However, the more sophisticated instrumental methods now under development have made little impact on clinical medicine. In this paper a brief overview of the development of optical diagnostics is presented, highlighted by some successful pre-clinical applications. The reasons for the slow penetration of optical diagnostics into clinical practice are discussed.

The past

Spectroscopy and imaging are, and always have been, at the core of medicine and biology. The earliest, and still the most common, medical diagnostic instrument is actually a highly advanced spectroscopic imaging system. This spectroscopic imaging system is capable of interrogating skin and determining the extent of blunt traumatic injury and alterations in blood flow, metabolism and tissue hydration (oedema). The imaging system has a self-adjusting aperture to control light levels and a self-focussing lens. The light-sensing device is an intricate arrangement of four types of exquisitely sensitive photosensors in a two-dimensional array of hundreds of millions of pixels covering an impressive spectral range. Spectroscopic images are acquired and transmitted to a control unit in real time, where the data is processed using a complex neural network. This spectroscopic imaging system can be viewed as the prototype for all modern optical spectroscopic imaging systems.

Of course this spectroscopic imaging device is the human visual system. The human visual system allows us to evaluate the degree of redness of skin, which is directly related to local blood flow. Changes in blood flow may be related to emotional factors (rage, embarrassment), hormonal changes (for example in menopause) or illness/injury (inflammation). The same system allows us to evaluate oedema (accumulation of tissue water) and the extent of blunt trauma injury (the severity and extent of bruising). Using this spectroscopic imaging system, humans are able to determine the emotional state of a potential enemy, the hormonal state of a potential mate and to assess the degree and extent of injury.

Impressive though the human visual system is, it has four fundamental limitations: spectral range, spatial resolution, sensitivity and specificity. As humans evolve and want to "see" and understand more of the world around us, these limitations mean that the human visual system can no longer provide the answers we are looking for. Nowhere is this more obvious than in the fields of biology and medicine. To overcome these fundamental limitations and allow more advanced diagnosis, new, enabling technologies are required.

The first of these enabling technologies to be developed was thermal imaging. The ancient Egyptians practised a crude form of "thermal medicine", in which areas of disease were identified based upon temperature differences sensed by running the fingertips over the body. The relationship between body heat and disease was also known to the ancient Greeks, including the Greek physician Hippocrates who wrote in 400 BC "in whatever part of the body excess of heat or cold is felt, the disease is there to be discovered." The ancient Greeks improved upon the Egyptian diagnostic technique with the development of a crude form of thermal imaging. The Greeks immersed the body in wet mud and areas that dried more quickly than others, indicating a warmer region, were considered to be diseased.[1] Primitive though this technique is, it should be considered a considerable advance, as it one of the earliest documented uses of contrast enhancement in medicine.

Medical imaging remained limited to simple visual inspection for two millennia, until the publication of *Micrographia* by Robert Hooke in 1665.[2] Hooke was an accomplished maker of compound microscopes, invented in 1595, and *Micrographia* was a detailed illustrated record of Hooke's microscopic observations of insects and plants. The most famous of Hooke's observations is his studies on thin slices of cork. In "Observation XVIII" of the *Micrographia*, he wrote: "I could exceedingly plainly perceive it to be all perforated and porous, much like a Honey-comb, but that the pores of it were not regular....these pores, or cells,...were indeed the first microscopical pores I ever saw...". Hooke's observations intrigued a Dutch tradesman Antony van Leeuwenhoek, who began building his own simple microscopes. Leeuwenhoek's skill at grinding lenses enabled him to build microscopes that magnified over 200 times, allowing him to see clearer and brighter images than any of his contemporaries, including Hooke, could achieve. With these incredibly simple instruments (essentially a single lens and a spike as a sample holder) Leeuwenhoek discovered protozoa (in 1674) and bacteria (in 1683).[3]

The microscope represents the first truly enabling (and transformational) technology in medical imaging and spectroscopy. The discovery of cells by Hooke and bacteria and protozoa by Leeuwenhoek were key discoveries that lead to the foundation of pathology as a modern discipline and the postulation of the Schleiden–Schwann cell theory,[4] a fundamental tenet of modern biology.

The microscope clearly alleviated one of the four fundamental limitations to visual diagnosis: spatial resolution. However, the problems of spectral range, specificity and sensitivity still remained. Specificity and sensitivity to disease states were addressed to some extent by the development of histological staining techniques. Essentially, staining improves sensitivity and specificity based upon reactions between exogenous agents and tissue components to produce differential staining of cell and tissue components. The most widely used staining technique is that of haematoxylin and eosin, (H&E), in which basophilic structures such as the nuclei of cells are stained blue, black, purple or grey by the haematoxylin whilst the acidoiphilic structures are coloured pink or purple by the eosin. Haemotoxylin staining was first demonstrated over a hundred years ago by Waldeyer using an extract of logwood[5] (*Haematoxylon campechianum*). The reliance on natural dyes diminished considerably with the synthesis of the first aniline dye (aniline purple or mauve) and staining techniques for tissue began to proliferate with advances in synthetic capabilities. For example Gram produced his stain for microorganisms in 1884[6] followed by Daddi's 'Sudan' stain for lipids in 1896.[7]

The development of microscopy and staining techniques heralded a period of rapid advances in tissue and cellular characterisation. However, medicine still relied exclusively upon the visible region of the electromagnetic spectrum. This reliance precluded the direct acquisition of information concerning the biochemistry of disease states and severely limited the nature of samples that could be analysed. Simply put, visible microscopy suffered from the need to have very thin sections of tissue (allowing only *ex vivo* studies) and a lack of visible chromophores (relying upon the addition of stains to be able to deduce and biochemical information). Two key developments in the 19th century changed this. The first was the discovery of infrared light in 1800 by the British astronomer Sir William Herschel. Herschel discovered the existence of infrared radiation when he tried to measure the heat produced by separate colors of a rainbow spectrum.[8,9] He noted that the highest temperature fell beyond the red end of the spectrum, implying the existence of invisible light beyond the red. Herschel termed this light "calorific rays", a term soon abandoned in favour of infrared (literally meaning beyond the red). Experiments continued with infrared light, leading to the construction of the first infrared spectrometer in 1835.[10] In 1840 Herschel's son, Sir John

Herschel, acquired the first thermal image of the sun on a piece of paper (using carbon particles suspended in alcohol).[11] This image, which he termed a thermogram, formed the basis of modern thermal imaging. Infrared spectroscopy then expanded rapidly, thanks to renowned spectroscopists such as Coblentz and Ångstrom, whose pioneering work between 1850–1900 established IR spectroscopy as an important tool for chemical characterisation and astronomy.

The second major event of importance to medical image in the 19th century was the discovery of X-rays by Wilhelm Röntgen in 1895.[12] Röntgen was investigating the range of cathode rays in air when he noticed that a screen coated in the fluorescent material barium platinocyanide would glow while the rays were being produced. He assumed unknown rays, which he called X-rays, were being emitted from the walls of the tube. To his surprise, Röntgen found that the rays could pass through his hand and cast shadows of his bones on the fluorescent screen. In this way X-ray imaging was born (with help from Thomas Edison, who developed the X-ray imaging screen within a year of Röntgen's discovery of X-rays).

The next major development in the field of spectroscopy and imaging that has relevance to biology and medicine is the discovery of Raman scattering in 1928 by Sir Chandrasekhara Venkata Raman.[13] who was awarded the Noble prize for the discovery in 1930. Raman had a strong interest in scattering, having previously demonstrated that the colour of the sea was dictated by scattering. In further experiments Raman demonstrated that when sunlight was focussed through the optics of a telescope and through a series of filters (to produce monochromatic light) onto a sample, most of the scattered light is observed at the same frequency as the incident light. However, a small fraction of the scattered light is shifted in frequency, a process now known as Raman scattering. This frequency shift is now known to be the result of absorption of photons to promote molecular vibrations and re-emission of photons at a frequency corresponding to the difference in energy between the photon absorbed and the energy required to promote the molecular vibration.

Although the Raman effect was demonstrated in 1928, it took many years for Raman spectroscopy to become a routine analytical tool. The main reason for this lack of utility is the weak nature of the Raman effect, only a very small fraction of light (about 1 photon in 1 million) impinging upon a sample is scattered in this way. Thus, intense monochromatic light is required to ensure that sufficient scattered light is produced to allow detection. Routine use only became feasible with significant instrumental developments such as development of the laser in 1958 by Schawlow and Townes.[14] The use of laser illumination allowed stimulation of the Raman effect at detectable levels routinely, resulting in the first laser Raman spectrum of a protein in 1970.[15] Development of a method for microscopic sampling was crucial to the emergence of Raman spectroscopy in the biological arena[16] in 1975. However the technique of Raman microscopic imaging really made major strides with the integration of the charge coupled device (invented in the 1970s) into Raman imaging systems, allowing detection of the extremely weak Raman effect with high sensitivity.

While Raman techniques were advancing, infrared techniques were also evolving. Studies by Elliot and Ambrose[17] in 1950 demonstrated that protein conformation could be studied by IR spectroscopy. Around this time Blout and Mellors[18] and Woernley[19] were beginning to analyse human and animal tissues to determine if discrete spectroscopic signatures could be identified from various tissues. These studies met with limited success due to a combination of the relatively unsophisticated instrumentation available, little knowledge of spectroscopic properties of biological molecules and the complexity of the samples under investigation. However, at about the same time as Ambrose, Elliot, Blout and Woernley were conducting their studies, a new piece of instrumentation that would revolutionise the application of IR spectroscopy in biology was demonstrated. Barer[20] and colleagues demonstrated that it was possible to couple a reflecting microscope to an infrared spectrometer and obtain infrared spectra of small particles (crystals and fibres). The applications of this new technique in biology are obvious: studies of tissues with very high spatial resolution. However, despite this instrumental advance, studies on tissues were essentially abandoned in favour of studies of the isolated building blocks of tissues.

Of course it was not just optical spectroscopic and imaging techniques that were undergoing dramatic advances. New techniques based upon nuclear magnetic resonance were also developed during this time period. In 1946 Felix Bloch *et al.*[21] and Edward Purcell *et al.*[22] independently demonstrated that when certain nuclei were placed in a magnetic field they absorbed energy in the radiofrequency range of the electromagnetic spectrum, and re-emitted this energy when the nuclei

returned to their original state. This discovery laid the foundations for the development of the technique of nuclear magnetic resonance (NMR) spectroscopy, which rapidly became an important analytical method in biology, making enormous contributions in the field of protein structure. Twenty-five years later Raymond Damadian demonstrated that the nuclear relaxation times of tissues and tumours differed, suggesting that NMR techniques could be used for tissue characterisation. In 1973 Paul Lauterbur presented images of test tubes of water obtained using magnetic resonance effects, results that would pave the way for the development of magnetic resonance imaging (MRI) as a medical diagnostic modality and revolutionise medicine.[23] Two years later Richard Ernst proposed magnetic resonance imaging using phase and frequency encoding and the Fourier transform,[24] an approach that forms the basis of current MRI techniques. The field developed rapidly, and by 1987 echo-planar imaging was being used to perform real-time imaging of the cardiac cycle.

The above brief (and by no means comprehensive) walk through the evolution of biological/medical imaging and spectroscopy highlights a number of important issues. Firstly, the development of any imaging or spectroscopic technique requires input from a large number of disciplines and an even larger number of laboratories. For example to develop the infrared microscope in it's current form required at a minimum contributions from Herschel (discovery of IR light), Fourier (the mathematics required to process interferograms), Melloni (constructed the first IR spectrometer), Michelson (developed interferometry), Barer (the first IR microscope) and Schawlow and Townes (inventors of the laser). Equally importantly, this example show the long lead times that are required to develop and establish new techniques, in this case 200 years. Of course occasionally so-called transformational technologies, of which MRI is a clear example, have a much-compressed time frame of development and implementation. In the case of MRI, a new technique required less than 50 years from discovery of the underlying phenomena to full and widespread clinical acceptance.

The rapid development and acceptance of MRI as a tool of profound clinical and biological importance raises a provocative question: why have optical spectroscopy and imaging not been able to match this success despite a much longer (and some may say richer) history? The remainder of this article will outline some of the successes of optical spectroscopy and imaging in medicine and suggest some reasons why these powerful tools have yet to make a clinical impact

The present

Before outlining the success of optical spectroscopy and imaging it is perhaps useful to summarise the advantages of spectroscopic and imaging approaches, and to highlight the strengths and weakness of the major spectroscopic and imaging tools that may be considered for biological and clinical use.

Spectroscopy *vs* imaging

Spectroscopy. For the purposes of this article spectroscopy may be broadly defined as the study of the absorption characteristics of materials following interrogation with a broadband source of electromagnetic radiation. Spectroscopic techniques that have found a place in biology and medicine include NMR, infrared, Raman, UV/Vis circular dichroism and fluorescence spectroscopy. With respect to biological and clinical situations, spectroscopy may be performed in three ways:

1 Ex vivo spectroscopy. Ex vivo spectroscopy can provide information on individual materials or assemblies. The major advantages to this approach are that systems may be easily manipulated (temperature, pH *etc*) and that materials may be studied in isolation, without spectroscopic interferences. However, the physiological relevance of information obtained in such experiments is always questionable.

Ex vivo spectroscopy is mainly used in biology to study the structure and dynamics of proteins. NMR spectroscopy is undoubtedly the technique of choice for soluble protein analysis, providing 3-D structure at atomic resolution. However, at present NMR techniques are limited to relatively small proteins (although the mass that can be studied is constantly increasing) and cannot be applied to membrane proteins due to the low mobilities of these proteins. Furthermore, NMR

spectrometers may be prohibitively expensive for many labs ($400 K–1 M). In contrast IR and Raman spectroscopy are excellent tools for studying all types of proteins, including membrane proteins, although the information obtained is limited to secondary structure. In addition to studying static structures, both infrared and Raman can be used to study protein dynamics, through reaction induced difference spectroscopy and resonance Raman spectroscopy respectively.[25,26] Furthermore, IR and Raman spectroscopies can be used to study lipids, nucleic acids, carbohydrates *etc.*, and have been used extensively to study complex mixtures of these materials.[27]

UV/Vis circular dichroism (CD) spectroscopy has also been used to study proteins. As with IR and Raman techniques, CD spectroscopy only provides information on 2° structure of proteins, however it has the advantage that it requires concentrations approximately two orders of magnitude lower than IR and Raman techniques. Light scattering effects limit the usefulness of CD spectroscopy to the study of membrane proteins, unless they have been solubilised with the use of detergents. The range of biological materials that can be studied with CD spectroscopy is limited, as this technique is only useful for chiral molecules.

Fluorescence spectroscopy is generally the most limited of the optical spectroscopic techniques used for *ex vivo* analysis, and can be technically challenging. In general, fluorescence techniques require the addition of exogenous probes to allow information to be extracted. Typically this information relates to properties (hydrophobicity *etc*) of specific regions or domains within proteins or provides information on protein dynamics (for example through techniques such as fluorescence resonance energy transfer).

2 In situ spectroscopy. In situ spectroscopy can provide information on complex systems (cells and tissues), and allows studies of materials in a more natural environment. However, the spectra obtained tend to be very complex with significant spectral overlap from many materials, making assignment to individual species difficult. Spectra may in fact be so complex that little if any direct chemical information can be obtained. In addition, spatial information (a key factor in clinical spectroscopy) is often not obtained.

Many of the tools used to study biological materials in isolation can be applied to studies *in situ*. For example ^{1}H NMR spectroscopy provides information on low molecular weight metabolites such as lactate, choline, amino acids and simple sugars such as fucose, and is useful for studying tissue biopsies and biological fluids.[28] Unfortunately, the information obtained is a spatial average, and no information on larger macromolecules (membrane lipids, structural proteins, nucleic acids) is obtained. In contrast, IR and Raman techniques can be used to provide either a spatial average or to assess microscopic regions of tissues. These techniques are excellent tools for analysis of structural proteins, membrane lipids and nucleic acids *in situ*, with a spatial resolutions of 25 μm easily obtained for infrared techniques (6 μm with a focal plane array, 3 μm with a synchrotron) and 1 μm possible with Raman techniques.[29]

3 In vivo spectroscopy. In vivo spectroscopy allows information to be obtained from intact animals, and provides information on all spectroscopically active materials present at the measurement site, weighted to concentration. It provides true physiological information, but again spectra are complex and assignment of absorption features may be difficult. In addition problems with light scattering can be significant. Again, most of the techniques described above can be used *in vivo*. Fluorescence spectroscopy/endoscopy has been used to analyse the endogenous fluorescence of tissues (for example NADH, tyrosine and tryptophan fluorescence) of a variety of tissues *in vivo*, with limited success to date.[30] This limited success is largely due to the low intensity exogenous fluorescence from most tissues, coupled to the low specificity of changes in fluorescence for particular disease states. Near infrared and Raman spectroscopy provide more specificity and more chemical information, and are thus potentially more useful as *in vivo* diagnostic tools. Potential diagnostic utility is further enhanced by the favourable light transport characteristics of tissue in this spectral range.

Imaging. In contrast to spectroscopic techniques, imaging techniques are based upon an assessment of the interaction of light with a particular chromophore or type of chromophore (rather than all chromophores that may be spectroscopically active) and assessment of the distribution of the chromophore in 2 or 3 dimensions. As discussed above, the most valuable of the current imaging modalities is undoubtedly MRI. Diagnostic MRI relies very heavily upon proton imaging (T_1, T_2, DW (diffusion weighted), contrast enhanced imaging *etc*), and is based upon

differences in relaxation rates of protons in tissue components, tissue water and bulk water. Such techniques provide medium image contrast and biochemical specificity is low, but the over-riding advantage of MRI is production of medium resolution 3 dimensional images of tissues.

Optical imaging techniques that have been demonstrated include near infrared and fluorescence imaging. Clinical near infrared imaging techniques are based upon the differential absorption of near infrared light by oxy and deoxyhaemoglobin and water. Such techniques have slightly higher biochemical specificity than MRI and medium contrast. Images are restricted to medium spatial resolution images in 2 dimensions. Fluorescence imaging can be used with exquisite biochemical specificity with the appropriate choice of fluorophore (for example a labelled antibody), but again images are restricted to medium spatial resolution images in 2 dimensions

Spectroscopic imaging. Clearly both spectroscopy and imaging have clinical potential. However, while spectroscopic techniques are generally high information content techniques, they provide little spatial information. In contrast imaging techniques provide spatial information but convey little biochemical information. A combination of the two approaches, spectroscopic imaging, should in principle combine the main advantages of the two techniques and provide both biochemical and spatial information. Conceptually this may be thought of as simply simultaneously interrogating samples with multiple wavelengths and simultaneously assessing the distribution of multiple materials. In magnetic resonance spectroscopic imaging the 3 dimensional distribution of multiple proton-containing species (lactate, amino acids, fucose) or multiple phosphorous-containing species (ATP, ADP, AMP, Pi, PCr, *etc*) can be imaged, providing images with very high metabolic specificity, medium contrast and medium spatial resolution. Optical spectroscopic imaging may be used to simultaneously assess the distribution of multiple endogenous chromophores such as NADH, oxy and deoxyhaemoglobin and water *etc.*,[31] or to monitor exogenous chromophores such as labelled antibodies, green fluorescent protein and luciferase. Optical techniques can thus provide haemodynamic and metabolic specificity with medium contrast.

Some successes

The development of optical spectroscopy and imaging is a broad field, and it is not the intention of this article to describe in depth the advances that have been made in these areas. Rather the main aim is to highlight the nature of the information that may be obtained and some of the pitfalls that must be overcome. This article will therefore concentrate on one specific example; work conducted in our laboratories aimed at developing a rapid diagnostic test for skin cancer will be outlined. These studies began with *ex vivo* analysis of a range of tissues to allow an understanding of the spectroscopic properties of tissues to be developed, and advanced to *in vivo* spectroscopic studies on volunteers and preliminary imaging studies on rodents. This example therefore provides an overview of many of the difficulties and potential solutions encountered using a variety of techniques.

Why skin cancer?. Skin cancer is the most common cancer in most countries. In Canada in 1992 alone 50 300 new cases of skin cancer were reported, compared to 19 300 cases of lung cancer, 16 200 cases of colorectal cancer and 15 700 cases of breast cancer. Skin cancer is thus as prevalent as the next three leading types of cancer combined. The incidence continues to rise, with 64 200 new cases in 1997, an increase of 14 000 cases annually in five years. Fortunately, the malignancy is readily accessible and can be removed in a minor surgical procedure. In fact, if caught early enough skin cancer has a cure rate of 100%. Unfortunately, early diagnosis of skin cancer remains difficult. Skin cancer is not one disease but a class of related conditions, many of which appear similar when inspected visually. Diagnosis based upon visual inspection is therefore subjective.

To understand this subjectivity, consider an abnormal skin growth. The growth may be pigmented or non-pigmented. If it is non-pigmented and malignant, then it is probably basal cell carcinoma or squamous cell carcinoma. However, the clinical course of these two forms of skin cancer is very different. Basal cell carcinoma spreads laterally across the surface of the skin, without penetrating the deeper layers of the skin. Thus, although disfiguring, basal cell carcinoma rarely metastasises and is rarely fatal. However, squamous cell carcinoma metastasises readily, and is often fatal. It is therefore important to be able to distinguish between the two types of skin cancer. Visually, discrimination between these two types of cancer is difficult.

If the growth on the skin is pigmented it may be one of the many types of pigmented lesion commonly referred to as moles and technically referred to as naevi. Types of naevi include congenital naevi, acquired naevi, blue naevi and dysplastic naevi. Alternatively, malignant melanoma may be suspected. Initially, malignant melanoma is suspected based upon the unusual appearance or behaviour of naevi. The so-called "ABCD" warning signs guide assessment of the likelihood of malignant melanoma: **A:** *A*symmetry in shape; *A*ppearance of a new mole. **B:** *B*orders that are notched and/or, irregular; *B*leeding,.**C:** *C*olour is variable or contains blue, grey, white, pink or red; *C*hange in shape, size or colour; *C*oncern of patient. **D:** *D*iameter exceeding 6 mm. Diagnosis of malignant melanoma based upon such an approach is obviously subjective. Complicating matters further, basal cell carcinomas may be pigmented, and appear similar to malignant melanoma. Even if a diagnosis of malignant melanoma or pigmented basal cell carcinoma is ruled out, abnormal growth of naevi is clinically relevant as some types of naevi carry a definite risk of progression to malignant melanoma.

Upon initial presentation the clinical problem is therefore difficult, requiring classification of an abnormal skin growth as squamous cell carcinoma, pigmented or non-pigmented basal cell carcinoma, malignant melanoma, or an acquired, congenital of dysplastic naevus. Ultimately some additional information concerning grade (*i.e.* degree of aggressiveness, depth of penetration *etc.*) is required, which is even more difficult to ascertain visually. Once an initial diagnosis has been made a more definite classification requires excision of tissue and histological analysis. It is only at this point that the diagnosis becomes definitive. A non-invasive technique that could remove the subjectivity in the characterisation of skin lesions and could be performed in the physicians/dermatologists office would be of practical benefit.

We have taken a multi-step approach to skin cancer diagnosis. The first step was to determine the optical technique and spectral region of choice. Based upon skin transmission characteristics we chose the near infrared spectral region as that most likely to allow *in vivo* diagnosis. The second step was to gain a broad understanding of the fundamental spectroscopic properties of tissues, and to determine whether or not information concerning skin cancer could be extracted from biopsies. We chose to study biopsies using mid infrared spectroscopy rather than near infrared spectroscopy, as the information content of mid infrared spectra is higher than (although related to) that of near infrared spectra. The rationale for this was simple: if we could not understand the optical properties of biopsies and obtain clinically useful information from mid infrared spectra, it would be highly unlikely that we could obtain such information *in vivo* from near infrared spectra. The third step was to transfer our knowledge of *ex vivo* specimens to *in vivo* lesions, and develop an *in vivo* spectroscopic technique for classifying skin lesions. The final stage was to develop an imaging modality for skin cancer diagnosis.

Step 1. *Ex vivo* characterisation of biopsies. Our studies began with attempts to understand the infrared spectroscopic properties of tissues. The wavelengths of infrared light absorbed by a sample (called its infrared spectrum) provide information relating to the composition and structure of the sample. In other words, the infrared spectrum is a molecular fingerprint. As the molecular fingerprint of tissue changes during a disease process, differences should be seen in the infrared spectra of normal and diseased tissues. We therefore undertook a major study investigating the spectroscopic properties of a wide range of tissues.

Fig. 1 shows spectra of breast and liver tissue obtained using macroscopic sampling techniques. Spectra of malignant breast tissue were acquired by placing 10 μm sections of breast tissue between CaF_2 windows without allowing the tissue to air dry. Spectra of liver tissue were obtained from 10 μm section of liver tissue that were placed onto CaF_2 windows and allowed to air dry. Obvious differences are apparent between spectra of breast and liver, and between the different breast and liver samples. It may be surmised that the spectral differences between the three sections of breast tissue are in some way related progression of the disease. However, each section was characterised by a pathologist as arising from invasive ductal carcinoma of the breast at the same stage of development. Spectral differences are unlikely therefore to reflect differences in disease progression. In fact, these differences are simply reflective of the normal variations that are to be expected in animal tissues. The upper trace is dominated by absorptions from lipid-like species, as indicated by prominent absorptions from CH_2, C=O and C–O–C groups and lacks an amide I absorption. As the amide I absorption arises from cellular and structural proteins, this section does not contain

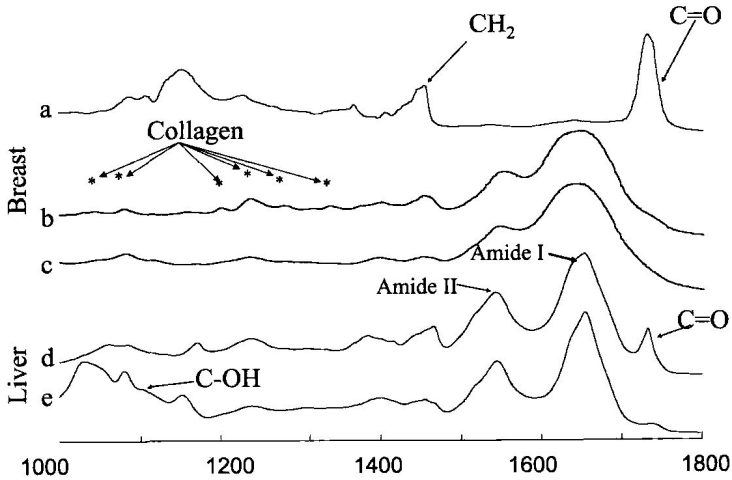


Fig. 1 Mid infrared spectra of macroscopic sections of breast (a–c) tumours and liver tissue (d–e)

connective tissue of epithelial/tumour cells. In fact, this tissue section is dominated by adipose tissue. In contrast, the section giving rise to spectrum (b) is dominated by absorptions from protein, specifically collagen. This section is therefore mainly composed of connective tissue. The section giving rise to spectrum (c) is also dominated by protein, but does not exhibit absorptions from collagen. This section is therefore dominated by absorptions from epithelial/tumour cells.

This brief discussion highlights an important drawback to the analysis of tissue using spectroscopic techniques: the lack of spatial information that may be obtained. Each of the tissues giving rise to spectra in Fig. 1 may have contained epithelial cells, but information on the cells can only be obtained from spectrum (c). This has important implications for those spectroscopists trying to develop spectroscopic techniques for diagnosis based upon macroscopic measurements. Consider the schematics of tissue sections presented in Fig. 2. In Fig. 2a we have two sections of tissue with the same number of adipose cells, but different numbers of other cells. One section has four epithelial cells, the other has two malignant cells. Now if we assume for the sake of argument that the tumour cells have twice the DNA content of normal epithelial cells, then spectroscopically we will

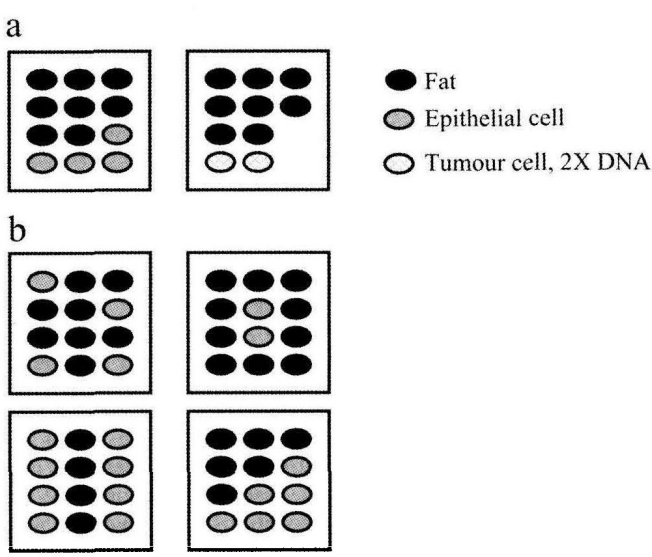


Fig. 2 Schematic representation of tissue heterogeneity. See text for description

detect the same amount of DNA in both sections. If we try and develop diagnostic tools based simply upon quantification of DNA in tissue sections (or from *in vivo* spectra), the tumour in section 2 will not be detected.

Fig. 2b further illustrates the importance of spatial information. This figure shows four sections which have epithelial cells randomly distributed, arranged as a focal group surrounded by fat, in organised striations or as a discrete structure within the tissue. A macroscopic assessment of these sections will not allow discrimination between the four tissue types, even though they are histologically distinct. Spectra alone do not provide sufficient information for characterisation; spatial information is required.

Step 2. *Ex vivo* characterisation of skin biopsies by infrared microscopy. IR microscopy is an essential tool for tissue studies as it allows both spectral and spatial information to be obtained.[29] An IR microscope focuses IR light onto a small area of a sample, allowing high quality spectra to be obtained from small samples. Coupling an IR microscope to a high precision computer controlled stage allows the automated collection of IR spectra at each position in a two-dimensional grid defined over a tissue section. Information collected at each point (pixel) is combined to produce an IR spectroscopic map of the tissue section. Large spectral maps are routinely collected containing thousands of spectra per map and enormous amounts of both spatial and spectral information.

We have used this approach to characterise skin tissue. A large number of skin biopsies were analysed by acquiring spectra from small regions (25 μm) of histologically identifiable structures within the sections. In this way we established a large database of spectra from the major histological structures in normal skin and the major types of skin cancer. Representative normalised spectra of epidermis, dermis, hair follicles and a basal cell carcinoma are shown in Fig. 3. Distinct differences are noted between major histological structures, even to the naked eye. It is clearly possible to distinguish the dermis, based upon the presence of unique absorptions from collagen at 1204,1240, 1280 and 1635 cm^{-1}. Hair follicle and sebaceous glands can be easily distinguished by an elevated lipid content (1740 cm^{-1}). Spectra of basal cell carcinoma and epithelium are more difficult to distinguish. Nevertheless, if normalised spectra of the two tissue type are superimposed (Fig. 4a), it becomes apparent that there is a significant difference in the apparent intensity of the absorptions arising from nucleic acids. This may be taken to be indicative of an increase in the nucleic acid content of the tumour cells in the basal cell carcinoma. However, if the spectra of the tumour and normal epithelium are overlaid without normalisation (*i.e.* raw data is presented,

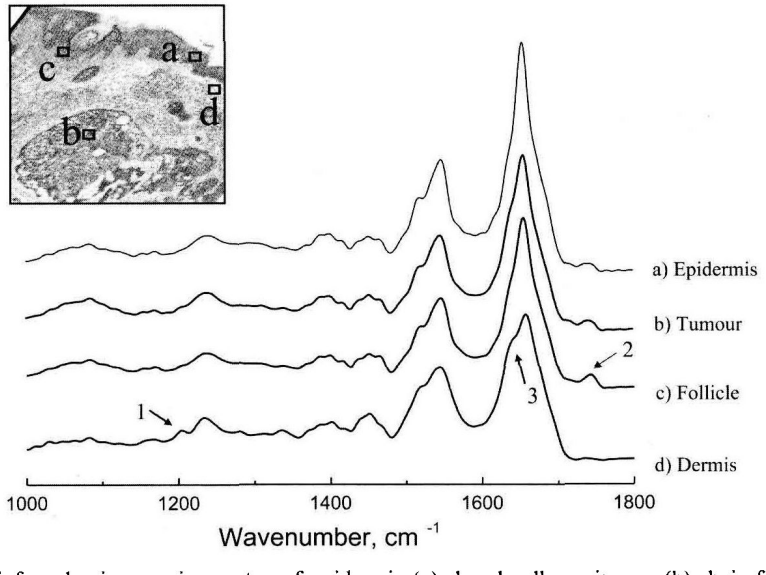


Fig. 3 Mid infrared microscopic spectra of epidermis (a), basal cell carcinoma (b), hair follicle (c) and dermis (d). Inset: regions of tissue from which spectra were acquired.

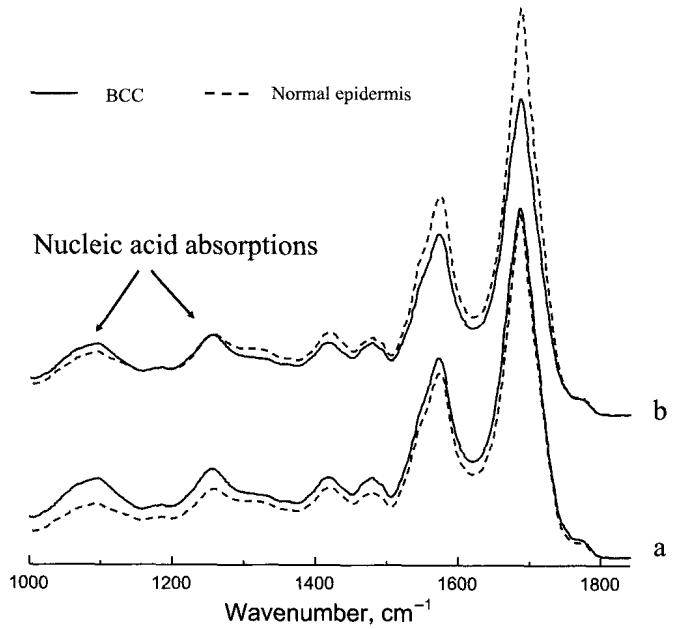


Fig. 4 Normalised (a) and raw spectra (b) of skin epithelium and basal cell carcinoma.

Fig. 4b), then the data suggest a different conclusion, in this case that there is a significant differences in protein content between the tumour and epithelial cells. Clearly the mode of presentation of the data can affect the conclusions drawn.

Which mode of presentation, and which interpretation, if either, is correct? Data is often normalised (for example to the amide I intensity) to ensure that differences in tissue thickness are not mistaken for differences in tissue composition. However, normalisation precludes the possibility of quantitatively assessing absolute variations in tissue composition. In reality, this means that both raw and normalised data should be analysed, and that data from the entire spectral map should be assessed rather than just a few spectra. In this example, examination of a large number of spectra suggest that the major biochemical differences between basal cell carcinoma cells and epithelial cells is a reduced protein content in tumour cells, probably as a result of nuclear enlargement.

Problems with spectral interpretation and data presentation illustrate the need for less subjective approaches to tissue characterisation, which led us to apply a number of pattern recognition techniques to infrared spectra of skin components. The simplest approach to pattern recognition analysis of spectroscopic data is to use techniques such as cluster analysis. Cluster analysis techniques compare spectra using some metric of similarity, and partition spectra into groups based upon this measure of similarity. Membership of groups is varied in an iterative process, for example to minimise the variance within groups while maximising the separation between groups, thus optimising the number of clusters formed. For spectroscopic maps, pixels belonging to the same cluster are displayed in the same colour, producing a coloured image showing cluster memberships. The results of a cluster analysis applied to a spectroscopic map from a section of basal cell carcinoma (Fig. 5a) are shown in Fig. 5b. Clusters corresponding to the epidermis, dermis, hair follicles and tumour are easily observed.[32]

Cluster analysis obviously shows groups of pixels that may allow some correlation with sample histology. However, this introduces an element of subjectivity. Other pattern recognition techniques make use of the fact that we often are in possession of information regarding clinical properties of a sample. For example in the case of skin lesions we have a clinical diagnosis (obtained by examination of a biopsy by a dermatopathologist). This information may be used to train techniques such as linear discriminant analysis (LDA) to recognise the particular combinations of peak frequencies, absorption band widths, relative intensities *etc.* in spectra that are characteristic of spectra from a particular clinical grouping. The trained LDA algorithm can then be applied to spectra of unknown

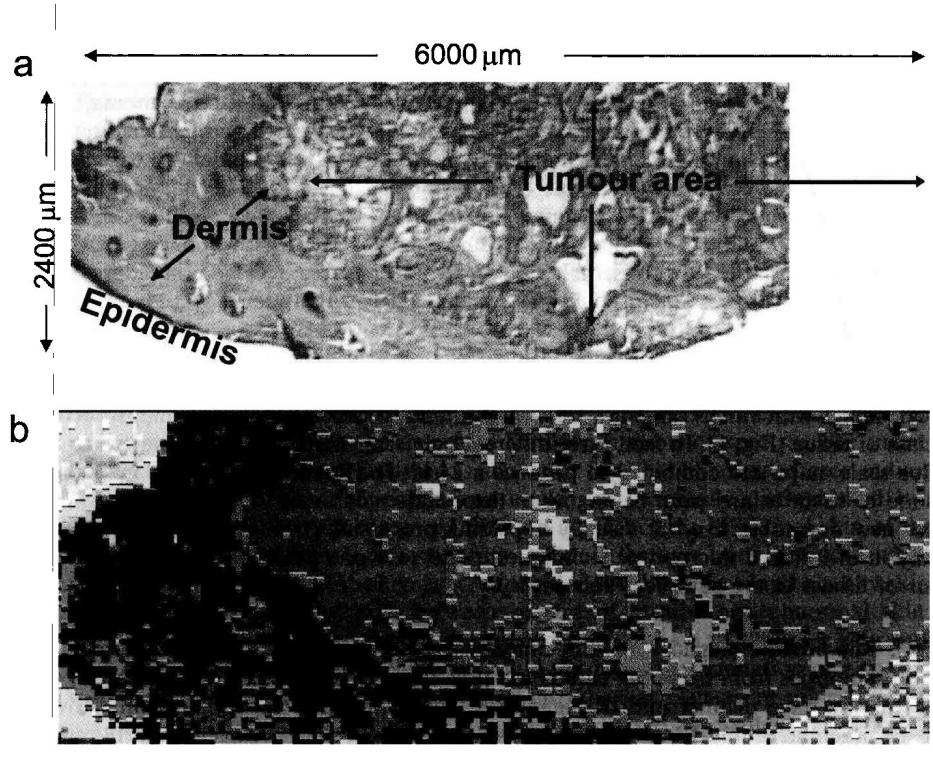


Fig. 5 Photomicrograph of a section from a basal cell carcinoma (a) and the results of fuzzy C-mean cluster analysis of spectra acquired by mapping the section using infrared microscopy (b). See text for more details.

origin, and the unknown spectra partitioned into one of these clinical groupings based upon the spectral pattern found.[33] The results of such an analysis are shown in Table 1.

The results in Table 1 demonstrate that mid infrared spectroscopic analysis of skin lesions coupled to a supervised pattern recognition analysis methodology can be used to characterise

Table 1 Non-subjective analysis (IR predictions are in rows, pathological classification in columns, correct classifications are in bold)

(a) Classification of tissue types						
	BCC	Epidermis	Follicle	Accuracy	Specificity	Sensitivity
Basal cell carcinoma	**157**	0	0	100	100	100
Epidermis	0	**68**	0	100	98.1	96.4
Follicle	0	2	**51**	96.2	100	100
(b) Classification by tumour types						
	SCC	BCC	Melanocytic	Accuracy	Specificity	Sensitivity
Squamous cell carcinoma	**48**	2	0	96	95.5	91.4
Basal Cell carcinoma	5	**151**	1	96.2	95.1	90.7
Melanocytic	2	2	**30**	88.2	99.7	99.9
(c) Subdivision of BCC lesions						
	Nodular	Infiltrating	Superficial	Accuracy	Specificity	Sensitivity
Nodular	**53**	6	1	88.3	64.3	55.3
Infiltrating	8	**9**	1	50	95.0	83.3
Superficial	10	0	**27**	73	96.4	91

pathological lesions. Not only can the technique discriminate between basal cell carcinoma and normal tissue components, but it can also be used to distinguish between various types of skin lesions (basal cell carcinoma, squamous cell carcinoma and malignant melanoma, Table 1b). Furthermore, pattern recognition analysis of IR spectra may also be used to characterise various types of the same class of lesion, for example nodular and infiltrative basal cell carcinoma (albeit with reduced accuracy) as shown in Table 1c.

Importantly for clinical utility, pattern recognition of infrared microscopic data can be presented in a "physician friendly" manner. In essence, methods have been developed to present the data in a way familiar to physicians that resembles the way data is presented from standard histological analysis. In brief, techniques such as LDA return a coefficient (between 0 and 1) that indicates the probability that the spectrum at any pixel arises from a particular tissue type. If these values are plotted as a function of position within the spectroscopic grid, then a map showing the probability of tissue belonging to a particular class may be produced. This can be converted into an eight bit grey scale image with values ranging from 0 to 255, and so-called "likelihood maps" produced for each class of tissue (Fig. 6). In fact, three different grey scale maps may be converted to red, green and blue scale maps and combined to produce a 24 bit RGB (red, green, blue) map showing the likelihood of tissue at any pixel belonging to three classes of tissue.[34]

These new non-subjective but visually appealing processing tools coupled to the potential for automation of infrared microscope techniques, point to a potential role for infrared microscopic analysis of tissues in the pathology laboratory.

Step 3. Non-invasive diagnosis of skin cancer using near infrared/visible spectroscopy. We have shown that biopsies from basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and melanocytic tumors have distinct mid-IR signatures when compared to normal skin. However, the diagnostic potential of mid-IR spectroscopy *in-vivo* is limited, since complete absorption of mid-IR light results with samples greater than 10–15 μm in thickness. In contrast, near-IR light is scattered to a much greater extent than it is absorbed, making tissues relatively transparent to near-IR light, thus allowing the examination of much larger volumes of tissue and the potential for *in-vivo* studies.

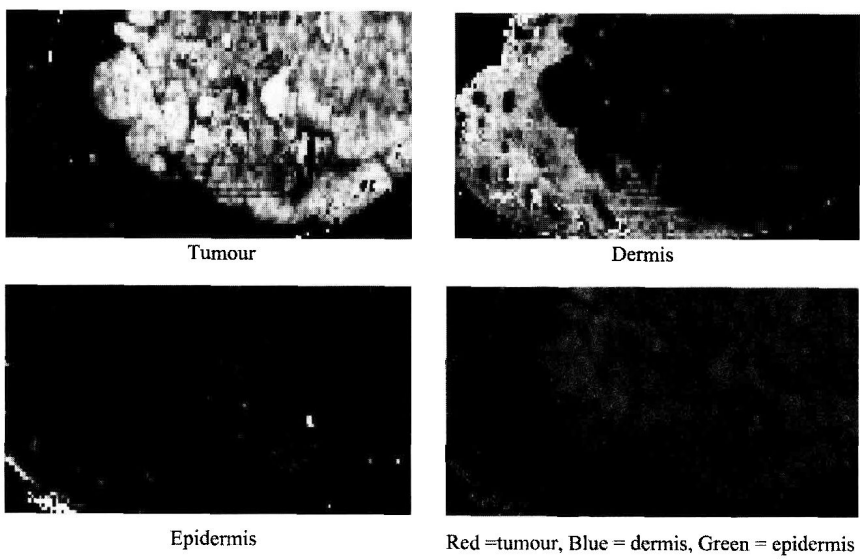


Fig. 6 Visual representation of the results of a linear discriminant analysis of spectra from the section shown in Fig. 5a. The LDA coefficient for each class (value between 0 and 1 representing the likelihood that the pixel corresponds to that tissue class) is converted to an 8 bit number. The 8 bit coefficient for tumour, dermis and epidermis for each pixel are then plotted as a function of position within the tissue as a greyscale image (black = 0, white = 255). The three greyscale images may be converted to red, green and blue 8 bit images and combined to produce a 24 bit RGB (red, green, blue) image showing the likelihood of spectra belonging to three classes simultaneously.