

## Choices for tissue visualization with IR microspectroscopy

Kathleen Gough<sup>a,\*</sup>, Margaret Rak<sup>a</sup>, Adam Bookatz<sup>a</sup>, Marc Del Bigio<sup>b</sup>,  
Sabine Mai<sup>c</sup>, David Westaway<sup>d</sup>

<sup>a</sup> *Department of Chemistry, Room 350 Parker Building, 144 Dysart Road, University of Manitoba, Winnipeg, Man., Canada R3T 2N2*

<sup>b</sup> *Department of Pathology, University of Manitoba, Winnipeg, Man., Canada R3E 0W3*

<sup>c</sup> *Manitoba Institute of Cell Biology, University of Manitoba, Winnipeg, Man., Canada R3E 0V9*

<sup>d</sup> *Centre for Research in Neurodegenerative Diseases, University of Toronto, Canada M5S 3H2*

Received 3 December 2004; accepted 15 February 2005

Available online 13 June 2005

### Abstract

In this paper, we address some of the issues associated with infrared (IR) imaging, with reference to our work on brain tissue from the TgCRND8 mouse, a transgenic model of Alzheimer's disease (AD). AD is the most common cause of dementia in the aging population. One of the characteristic hallmarks of this chronic neurodegenerative disorder is the accumulation of plaques in the brain, usually visualized with histochemistry and immunostaining. Although these methods are extremely useful, they illustrate only certain aspects of the sample, require a great amount of tissue processing, and are highly dependent on experimental protocols and reagent quality. IR imaging provides information on multiple components, with a minimal amount of sample processing. However, in order to interpret the data successfully, the issues of spectral acquisition parameters, pre-processing, and spectral artifacts need to be considered. The methods commonly used to process the data, such as uni- and bi-variate spectral analysis, and multivariate methods, such as hierarchical cluster analysis, and some issues concerning the use of second derivatives of IR spectra are discussed.

© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Infrared microspectroscopy; Alzheimer's disease; Spectral data processing; Second derivative

### 1. Introduction

The long-term goals for the use of infrared (IR) imaging range from application as a mainstream tool for rapid-throughput diagnostics to an important complementary tool in detailed pathology studies. There are several definable advantages for choosing IR over the many alternatives: (a) the spectra contain information about multiple components in situ, (b) the preparation of tissue sections, usually cryotomed from frozen, unfixed tissue blocks, is simple, (c) apart from desiccation, the tissues are typically not altered in any way, i.e. nothing has been added to the tissue and, except for volatile components (water), nothing has been removed, (d) the IR light is not damaging to the sample, thus

subsequent analyses are readily feasible, (e) with a bright synchrotron source or with focal plane array detectors, good spectra can be recorded at the diffraction limit. Recent papers on applications of IR to tissue visualization illustrate and emphasize these points [1–4]. Nevertheless, if the long-term goals are to be achieved, there must be further dialogue on how best to interpret and illustrate the spectral differences recorded in IR spectra.

IR spectral data maps present some interesting challenges for analysis, both from the viewpoint of the quantity of raw data and the choices for processing that data to obtain useful, meaningful information. Here, our focus is on the illustration of some of these options and a discussion of advantages and disadvantages, with specific reference to our work on Alzheimer's disease tissue. In addition, we consider some of the common practices in data acquisition that can increase the difficulty in extracting meaningful information.

\* Corresponding author. Tel.: +1 204 474 6262; fax: +1 204 474 7608.  
E-mail address: [kmgough@ms.umanitoba.ca](mailto:kmgough@ms.umanitoba.ca) (K. Gough).

Alzheimer's disease (AD) is the most common cause of dementia in the aging population. One of the characteristic hallmarks of this chronic neurodegenerative disorder is the accumulation of plaques in the brain [5]. The main component of plaques is the amyloid peptide (A $\beta$ ), which consists of 40 or 42 amino acids and adopts a highly aggregated  $\beta$ -pleated sheet fibril structure, known as an amyloid [6], in the plaque cores. We are studying brain tissue from TgCRND8 mice, a transgenic model of AD [7]. The animals express a double mutant form of human amyloid precursor protein 695 (KM670/671NL + V717F), under the control of the brain-specific PrP gene promoter. The characteristics include robust plaque pathology and cognitive impairment from an early age; small dense-cored plaques are detected by 3 months of age, concentrated in the cortex and hippocampus. As the animal ages, the plaques become larger and acquire a halo of diffuse amyloid. Purely diffuse plaques become more common in older animals and are most concentrated in the caudate and the molecular layer of the dentate gyrus. The dense-cored plaques are accompanied by dystrophic pathology and inflammation.

Classic histochemistry and immunohistochemistry are regularly used to illustrate different aspects of AD pathology [8–10]. However, these techniques are very specific, illustrating only one or two different targets, and require extensive tissue manipulation that can affect other features. Infrared (IR) spectroscopy has been used to study protein aggregation, since the position of the amide I band is dependent on the secondary structure of the protein [11–13]. Infrared microspectroscopy, which combines high spatial resolution with molecular-level information, provides a unique tool for studying plaques and associated changes in situ.

In this paper, we present a comparison of classic histology and immunohistochemistry with microspectroscopic visualization achieved through the analysis of a representative IR map that contains many features typical of TgCRND8 mouse brain. The data are processed in several ways, to reveal both the power and the potential pitfalls of this technique. Specific consideration is given to spectral acquisition parameters, pre-processing, some spectral artifacts, uni- and bi-variate spectral analysis contrasted to cluster methods with software such as CytoSpec, and some issues concerning the proper use of second derivatives of spectra.

## 2. Experimental

### 2.1. Animals

The animals in this study were treated in accordance with the guidelines of the Canadian Council on Animal Care and the protocols were approved by the University of Toronto and University of Manitoba. TgCRND8 mice and non-transgenic littermates were sacrificed by cervical dislocation. The brains were removed immediately and divided into left and right hemispheres. One half was fixed in 3%

buffered paraformaldehyde, dehydrated and embedded in paraffin; the other half was covered with OCT medium and snap frozen on dry ice. For the purposes of this paper, only one map, from an 11 month TgCRND8 mouse, will be considered in detail.

### 2.2. Sample preparation and tissue staining

For an initial examination of plaque morphology, the paraffin-embedded brains were cut into 6  $\mu$ m sections in the sagittal plane and placed on glass slides. Serial sections were stained with hematoxylin and eosin, Congo red, modified Bielschowsky silver stain and immunohistochemistry against ubiquitin (rabbit polyclonal, Dako, Carpinteria, CA). The frozen brains were cryotomed sagittally at 8  $\mu$ m thickness. For IR spectroscopic analysis, sections were thaw-mounted on reflective MirrIR<sup>TM</sup> slides (Kevley Technologies, Chesterland, OH). After IR analysis, they were stained with Congo red. Additional sections were placed on glass slides for immunofluorescence staining with anti-A $\beta$  peptide antibody 4G8 (Signet Laboratories, Dedham, MA), visualized with Alexa Fluor<sup>®</sup> 488 goat anti-mouse IgG secondary antibody (Molecular Probes, Eugene, OR), and counterstained with 4',6-diamidino-2-phenylindole. Immunofluorescence slides were imaged on a Zeiss Axiophot microscope equipped with a SenSys digital camera, and MacProbe v4.4 software (Applied Imaging, San Jose, CA).

### 2.3. Synchrotron IR mapping and data analysis

The map analyzed here was recorded with a Nicolet Magna 860 FTIR with a Spectra Tech Continuum IR microscope (U10B beamline, NSLS, BNL, Upton, NY). The spectra were recorded at 4  $\text{cm}^{-1}$  resolution, 12  $\mu$ m aperture, 10  $\mu$ m step size. Neither zero filling nor spectral smoothing was employed. A suitable region in the hippocampus was selected for spectral mapping by inspection, as the dense plaques are readily visible under bright light in the unfixed thaw-mounted sections. Uni- and bi-variate data analyses were performed with OMNIC/Atl $\mu$ s software (Thermo-Nicolet), while multivariate analyses were done with CytoSpec software (<http://www.cytospec.com>).

### 2.4. Second derivative analysis with simulated spectra

Model spectra with pure Lorentzian bands were created in Mathcad (MathSoft Engineering and Education, Inc., Cambridge, MA) to investigate some practical limits on the use of second derivatives in data analysis. We simulated each multicomponent spectrum as a sum of a preset number of Lorentzians with random numbers for the amplitude, wavenumber and bandwidth, to represent bands assigned as functional groups. Units are not relevant, so the "spectral" range was from 3 to 5. The first member of the set was centered at wavenumber 4 and had a constant bandwidth. The amplitudes were varied randomly each time

the Mathcad sheet was calculated. The input amplitude (intensity of the peak at wavenumber 4), the total intensity ( $F$ ) and the second derivative ( $F''$ ) of the summed peaks at wavenumber 4 were automatically captured. The graphs of input amplitude of the peak at wavenumber 4 versus  $F$  or  $F''$  could then be plotted. We considered sums of 5, 20 and 50 bands, to simulate spectra of increasing complexity.

In an ideal system of pure Lorentzians, Beer's Law must hold true for  $F''$ ; a plot of  $F''(4)$  versus input amplitude (4) must be linear; that is, the 'amplitude' of the second derivative is proportional to the amplitude of the original peak. Our test was designed to investigate how well the model would hold up in the presence of random interfering bands of varying intensity, as might be expected to occur in real tissue.

### 3. Results and discussion

#### 3.1. Classic methods of AD pathology visualization

AD plaques are classically imaged with histochemical stains and through immunohistochemistry with antibodies. Among the most widely used stains, both in standard pathology and research, are Bielschowsky silver and Congo red. Bielschowsky silver (Fig. 1A) stains plaques and dystrophic neurites, as well as normal neuronal processes,

but the amount of staining varies greatly with experimental conditions [14]. Congo red (Fig. 1B, E) stains any highly aggregated amyloid, regardless of its composition [15] and thus only reveals the core of dense-cored plaques. Although the staining process is easy and inexpensive, the chemical mechanisms are poorly understood and results can vary widely with protocol and reagent quality. A slightly more sophisticated alternative is immunohistochemistry, or immunostaining, which utilizes the binding of antibodies against specific epitopes in proteins, visualized with fluorochromes or through enzyme-substrate reactions with colored chromogens. Immunostaining is highly specific but each target requires its own antibody. For this study we used an anti-ubiquitin antibody which shows dystrophic pathology around plaque cores (Fig. 1C) and the monoclonal antibody 4G8, which binds to an epitope on the A $\beta$  peptide, regardless of its conformation (Fig. 1D). Fig. 1E shows a Congo red stained section serial to the section in Fig. 1D. These two sections illustrate different aspects of plaque morphology: the immunostain binds to the amyloid peptide in plaque cores, diffuse haloes of A $\beta$  peptide around the cores and to purely diffuse plaques, while Congo red staining shows only dense plaque cores. Fig. 1 shows that although classical methods can reveal a great deal of information, numerous techniques must be utilized to get a more complete picture, and multiple tissue sections are required.

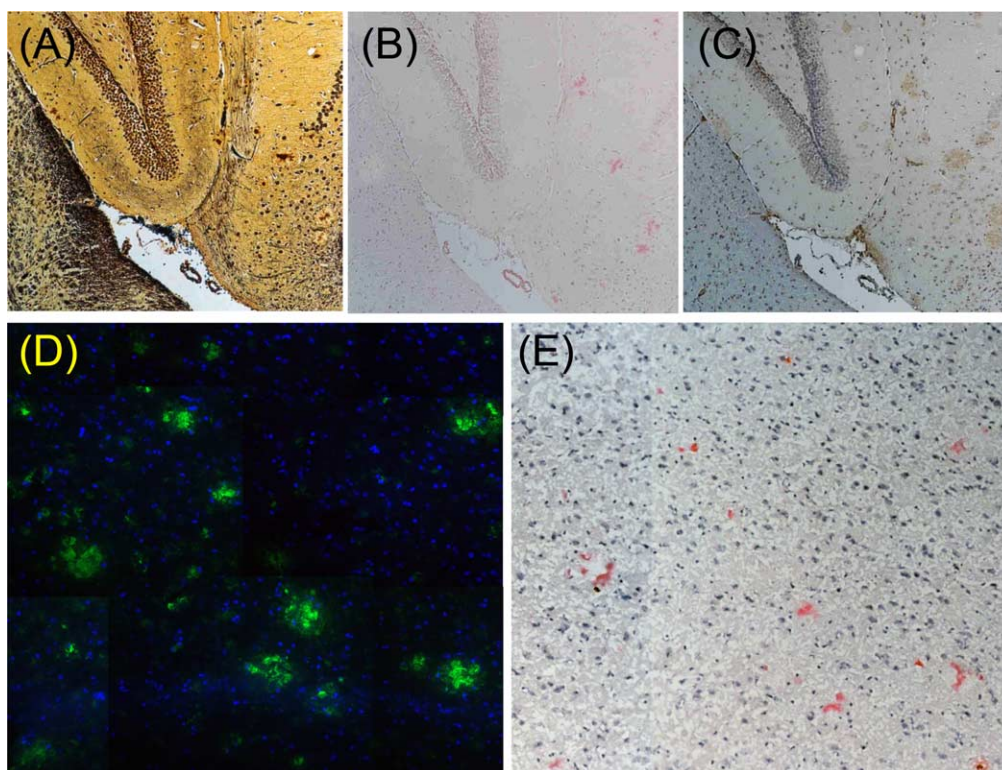


Fig. 1. Staining of paraffin-embedded serial slices from 11 month TgCRND8 mouse illustrates standard means of confirming plaques. Serial sections stained with (A) Bielschowsky silver, (B) Congo red, which shows the dense plaque cores and (C) anti-ubiquitin immunostain (brown). Nuclei stained blue with hematoxylin in B and C. (D) Immunofluorescent staining of A $\beta$  with antibody 4G8 and (E) Congo red stain of serial slice in cortex of 11 month TgCRND8 mouse.

### 3.2. Spectral visualization of plaques

In IR mapping, the goal is usually to create “false-color” images, where different colors in the image correspond to a specific feature in the tissue spectra. IR mapping data can be processed with a variety of options, including:

1. Uni-variate and bi-variate imaging, based on single peaks or ratios of two peaks, or integrated band areas thereof, often called functional group imaging.
2. Multi-variate analysis using a software package like CytoSpec, which can implement methods such as cluster analysis.

More sophisticated protocols include the development of neural networks and linear discriminant algorithms to improve robustness and accuracy for large-scale screening purposes [16,17]. We are not considering such protocols here.

A number of issues that affect spectral quality and inherent information must be considered during the IR data processing; these range from the initial data acquisition parameters (spectral resolution, S/N, zero filling), to baseline, normalization, scattering artifacts and pre-processing.

### 3.3. Example

A map in the hippocampus of an 11 month transgenic mouse is used to illustrate some of the issues that arise during analysis of IR map data. A photomicrograph of the unstained tissue (Fig. 2A) shows white matter, neuropil and neurons, as well as plaques (dark, dense spots), both in the hippocampus and in the arc of the corpus callosum (white matter) in the top left corner of the image. A photomicrograph of the same tissue stained with Congo red, following IR mapping, shows the location of dense plaque cores in both brightfield (Fig. 2B) and fluorescent mode (Fig. 2C).

In Fig. 2D, the plaques (red) are visualized on the basis of the amide I profile. The amide I band, a complex molecular vibration involving C=O stretch, CCN deformation and CN stretch [18], is sensitive to protein conformation. A maximum located between 1662 and 1650  $\text{cm}^{-1}$  is attributed to  $\alpha$ -helix; 1650–1640  $\text{cm}^{-1}$  to unordered structure; 1640–1620  $\text{cm}^{-1}$  to  $\beta$ -sheet; maxima above 1660  $\text{cm}^{-1}$  are associated with  $\gamma$ -turns and random coil [18,19]. Thus plaques, whose main constituent is the A $\beta$  peptide in a pleated  $\beta$ -sheet conformation, can be imaged by using the ratio of amide I peak areas attributed to proteins in a  $\beta$ -sheet conformation to those in an  $\alpha$ -helix conformation.

The spectral regions chosen for this map are shown by the shaded bars (Fig. 2E). Peak heights at specific wavenumbers could be used but are less advisable as the bands are broad, the location of the maximum shifts slightly, and exact locations of an “ $\alpha$ -helix” and a “ $\beta$ -sheet” maximum are not strictly definable. However, bi-variate spectral analysis of

amide I band gives false positives, with both the white matter in the corpus callosum and the CA1 neurons showing up as red and yellow, similar to the plaques. Examination of individual spectra (Fig. 2E) provides insight into the problem. The topmost spectrum shows the typical  $\alpha$ -helix maximum with a lower energy shoulder attributed to  $\beta$ -sheet, from a plaque region. The next lower shows a spectrum closer to the edge of the plaque, where the  $\beta$ -sheet component is not as prominent. The middle spectrum is that of normal tissue. The last two spectra are from white matter and from neurons (bottom). The full widths at half height (FWHH) in the latter two spectra are not altered greatly from the typical  $\alpha$ -helix bands, and there is no prominent  $\beta$ -sheet shoulder. There is a general shift of band centre, but it is spectroscopically distinct from the absorption around 1630  $\text{cm}^{-1}$  associated with  $\beta$ -sheet conformation. The ratio in the bi-variate analysis is sensitive to the band shift, and an erroneous result is reported. Ultimately, algorithms can be created to separate these many features, but the simple approaches clearly can lead to erroneous impressions.

Fig. 2F shows spectral artifacts created by (red spectrum) dense plaque core, which is almost opaque in IR, and by scattering off plaque edges (green spectrum). The light scattering off dense white matter creates multiple reflection artifacts and fringing, and hence a sinusoidal baseline (Fig. 2G, blue) that strongly perturbs the amide I profile.

Without inspection of individual spectra, these anomalies would be misinterpreted. The artifacts occur somewhat randomly in tissue samples, varying with quality, thickness, composition, etc. Data processing protocols must be carefully designed, as there is no simple way to predict or to identify and reject all such anomalies.

Another use of IR mapping is illustration of tissue morphology [20]. Uni- and bi-variate analyses of the CH<sub>2</sub> stretch region in our sample map illustrate white and grey matter, as well as neurons and plaque displacement of healthy tissue. Fig. 2H shows a univariate spectral map created on the basis of the height of the CH<sub>2</sub> stretch peak, shown as pixels, while Fig. 2I shows a bi-variate spectral map based on the CH<sub>2</sub>/CH<sub>3</sub> ratio, shown at fill level 10 for easier visualization. The spectra in Fig. 2J show that, as expected, white matter (blue) has the highest lipid content, followed by neuropil (green-yellow) and neurons (red), which have very low lipid content.

Figs. 3 and 4 show the results obtained from cluster analyses on the original, unprocessed IR map data, using the CytoSpec software. Cluster analysis is based on finding the similarity between objects (in this case, between IR spectra) by calculating the distance matrix, then creating groups of similar objects with a hierarchical clustering algorithm [21]. The software allows the user many choices, including selection of spectral region(s) and a choice of algorithms, not to mention a wide array of spectral pre-processing options. In our experience, the best parameters for illustrating features are not necessarily intuitive, yet the obvious potential for obtaining clear images is demon-

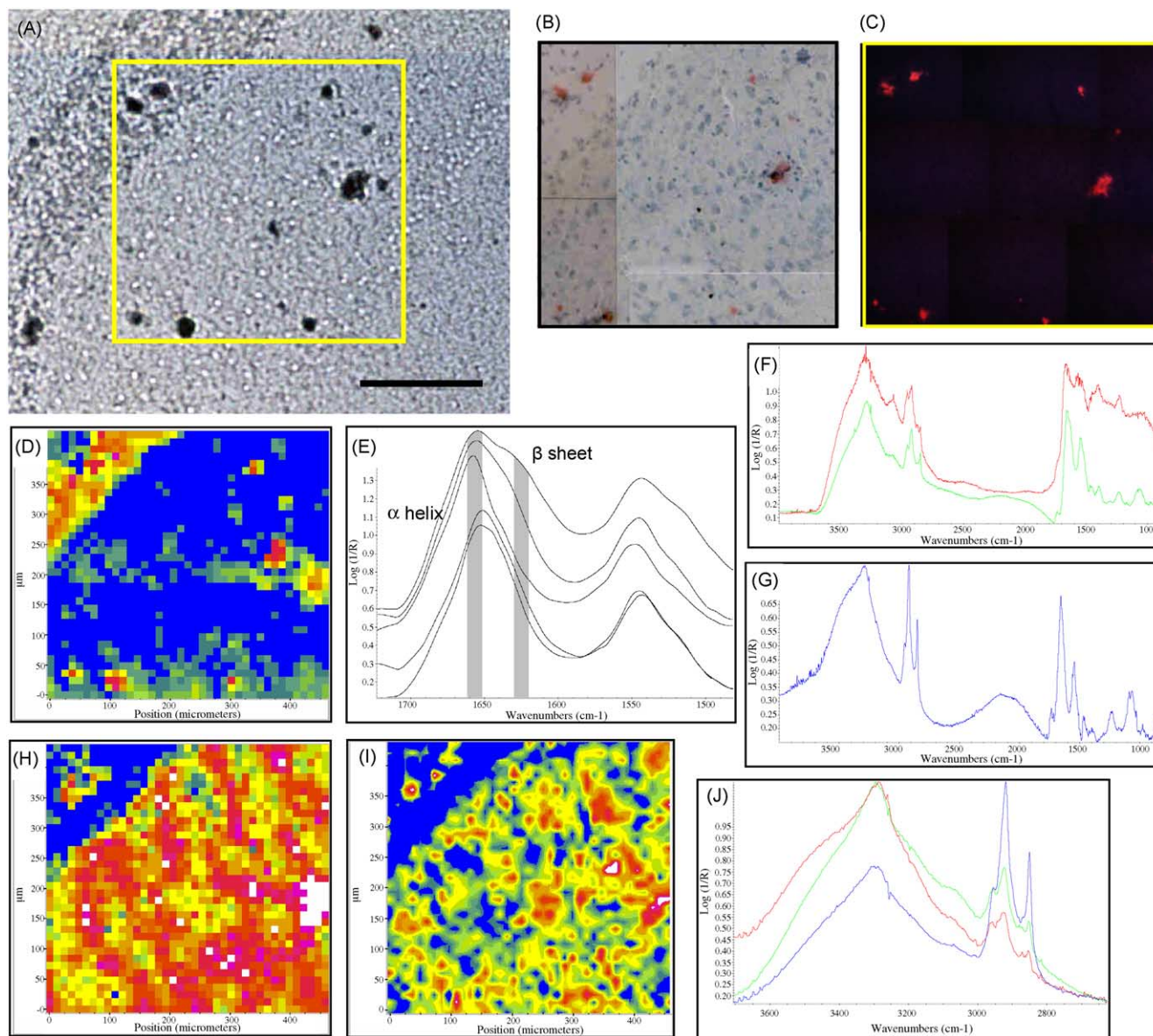


Fig. 2. (A) Photomicrograph of unstained tissue in the hippocampus of an 11 month old TgCRND8 mouse. The yellow box outlines map area. Same tissue after Congo red staining (done after IR mapping) in (B) brightfield and (C) fluorescent mode. (D) A bi-variate spectra map created on the basis of the amide I band profile. Plaques, as well as the white matter of the corpus callosum show up in red. (E) The amide region showing spectra of plaque core, plaque edge, normal neuropil, white matter and CA1 neurons (from top to bottom). The grey regions indicate the peak areas, due to  $\alpha$ -helix and  $\beta$ -sheet used for secondary structure analysis in Fig. 2E. (F) Spectral artifacts encountered with plaques: the dense plaque cores are almost opaque to IR light (red) and scattering off plaque edges creates spectral artifacts (green). (G) Spectral artifacts encountered with white matter: light scattering off dense white matter creates sinusoidal baseline (blue), which can strongly perturb the amide I profile. (H) Uni-variate spectral map created on the basis of the height of the  $\text{CH}_2$  stretch peak. (I) Bi-variate spectral map bases on the  $\text{CH}_2/\text{CH}_3$  ratio, (pixels blended for ease of visualization). (J) The CH stretch profiles are characteristic of white matter (blue), neuropil (green) and neuron (red). (Scale bar = 200  $\mu\text{m}$ ).

strated. Fig. 3 shows the results of cluster analysis performed on the amide I and II spectral regions. This method successfully separates the plaques from the remaining tissue morphology, where the bi-variate analysis (Fig. 2D) was unclear in the corpus callosum region. However, the clustering diminishes the detection range (plaque extent) in the tissue. The color scheme (Fig. 3A) in the cluster analysis image corresponds to different components, including plaques, white and grey matter. The average spectra (Fig. 3B) are also color coded to match the processed

map colors. In this simple cluster, there is little significant difference in the grey matter designated by the light and dark blues. The accompanying dendrogram shows that the division into the selected four components was a reasonable choice.

In Fig. 4, the cluster analysis on the CH stretch region may be compared to the results from the uni- and bi-variate analyses in Fig. 2H–J. The map (Fig. 4A) shows slight differences in the edge and centre of the corpus callosum, and dramatically demonstrates the location of most of the neurons (red), as compared to the tissue (Fig. 4B).

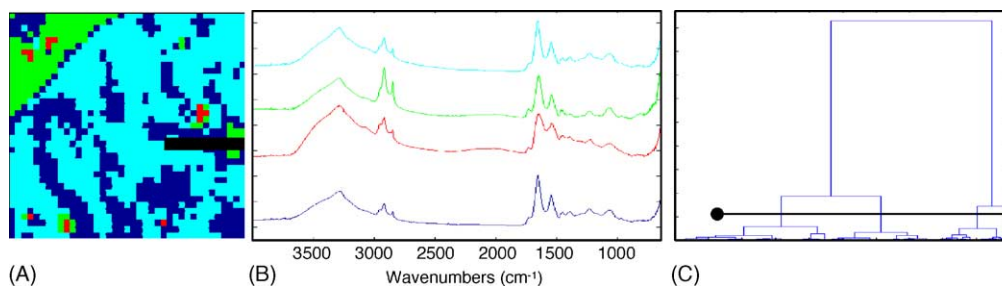


Fig. 3. Cluster analysis based on Amide I and II bands successfully separates the AD plaques from the remaining tissue morphology, where the bi-variate analysis Fig. 2 was unclear in the corpus callosum region. (A) Cluster analysis map shows plaques (red), corpus callosum, CA1 (green) and neurons and neuropil (blue). Some pixels were omitted due to artefacts (black). (B) Average spectra color-coded to processed map and (C) the corresponding dendrogram.

The compact group of neurons along the right side (arrow) is the beginning of the cornu ammonis (CA1) region. The average spectra (Fig. 4C) again support the assignment and the dendrogram (Fig. 4D) validates the choice of four distinct groups. The two plaques located inside the curve of the corpus callosum (upper left) cluster with the neuron group, demonstrating the absence of lipid in the aggregated protein core.

### 3.4. Comparison of real and model spectra

Data analysis is often performed on derivative (most often second order) IR spectra. The justification sometimes offered is that this removes baseline effects and increases the “resolution” of the data. The behavior of derivative spectra has been explored in detail [22–27], however, there are several points that may get overlooked. Since the signal-to-

noise ratio decreases as higher order derivatives are taken, the data are often smoothed prior to derivatization. The Savitzky–Golay method of calculating derivatives from spectra (the most widely used method) incorporates a smoothing procedure [28]. This lowers spectral resolution and may obscure important details, particularly small peaks and shoulders, while creating false maxima through smoothing of noise. In addition, derivative calculations lead to the creation of artifactual peaks on both sides of real peaks. Zero-filling, a common practice to reduce noise, introduces non-data points into the interferogram, creating an illusion of smoother data without providing any new information. Ultimately, the derivatives of spectra are only approximations that vary depending on the method employed, the number of points used and the pre-processing. In this section, we consider some issues concerning spectral manipulation and interpretation.

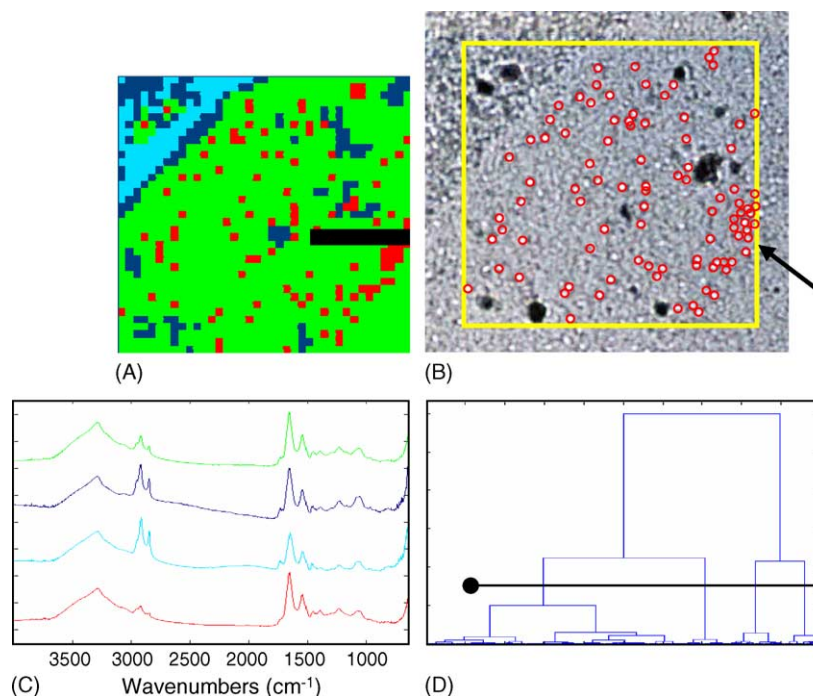


Fig. 4. Cluster analysis based on the CH stretch region is an excellent tool for the location of most neurons. (A) Cluster map showing neurons (red), corpus callosum (blue), neuropil (green) and plaque and corpus callosum (yellow). Some pixels were omitted due to artefacts (black). (B) Neurons (O) marked on a micrograph of unstained tissue, arrow points to cluster of CA1 neurons. (C) The average spectra and (D) corresponding dendrogram.

### 3.4.1. Real and model spectra: band position and assignment

Real spectra are collections of discrete points, dependent on the chosen spectral resolution. Our model spectra are true functions, thus we are testing an idealized system. The summed Lorentzian function appeared as normal tissue samples tend to, with many peaks blending in with other peaks (Fig. 5A:  $F(x)$ ). The second derivative (Fig. 5B:  $F''(x)$ ) effectively resolved the location of the test peak better than could be done in the original tissue simulation. However, this was only possible when we did not create an excessive number of Lorentzians; with  $n > 20$ , the second derivative contained too many overlapping interferences to correctly determine the original Lorentzian band position.

The assignment of vibrational bands or regions to particular modes, such as the amide I mode of proteins, is based on empirical evidence and the intrinsic intensity of the different modes is unknown. There have been several attempts to simulate spectra of proteins in different conformations [29–31], but the computational problems are massive compared to the available computational power today, and the problem remains unsolved. Hydrogen bonding or other differences in environment are responsible for the variations in the amide I position, but could alter band shape and intensity considerably. In practice, amide I modes and other functionally or

conformationally distinct bands should not be assumed to be represented by one peak or even one region.

### 3.4.2. Real and model spectra: amplitude

When a small number of Lorentzians (e.g.: 5) are used to simulate the spectrum, Beer's Law holds as expected for both  $F$  and  $F''$  (Fig. 6A and B). However, with increasing numbers of bands present ( $n = 20$ , Fig. 6B and E;  $n = 50$ , Fig. 6C and F) problems arise due to interference from overlapping peaks. This happens for both the original spectrum and the second derivative; thus, using the second derivative spectrum does not provide any intrinsic advantage over the original spectra, even with this simple model when only one pure Lorentzian with a constant band width is considered. In real tissue, millions of modes exist, each producing a Lorentzian band, so if 50 modes cannot be handled, certainly 50 million would likewise fail.

This conclusion is somewhat misleading. In our model for  $n = 50$ , many of the peaks are at effectively identical positions (i.e. one Lorentzian might be centered at 4.000, while another is at 3.998). The test only took into account the amplitude of the Lorentzian at 4.000, despite the fact that the almost-degenerate Lorentzian may (and in real-life situations likely will) have been from the same functional group. If most of the contributions in the region arise from the same functional group, then the model works fairly well.

Since the second derivative depends on both original peak height and FWHH [24,25], there is a risk in choosing to ignore bandwidth variation. If the bands are inhomogeneously broadened, there will be additional problems with applying the second derivative. In Fig. 7, the upper trace shows a portion of the IR spectrum of polystyrene film, the lower trace, the second derivative. This standard reference spectrum was selected as it is a convenient example of the advantages and disadvantages of the derivatization process.

The magnitude of a peak in the second derivative spectrum will still bear a relationship to number of functional groups present in the sample, however, the relative intensity of two bands with different intrinsic absorption coefficients and FWHH will necessarily change. In the polystyrene example, the strong partially resolved triplet, around  $700\text{ cm}^{-1}$ , dominates the second derivative spectrum by virtue of the narrow bandwidth in the original absorbance spectrum. The broad bands around  $750$  and  $540\text{ cm}^{-1}$  are significantly diminished relative to it. In fact, the reduction of broad bands and enhancement of small features is often cited as a reason for choosing the second derivative. However, this process is just cosmetic and can introduce artifacts. In the polystyrene film example, the bands all originate from the same compound, in the same homogeneous sample, but the alteration in relative magnitudes will be the same for all the bands from the different components in a mixture. To measure relative amounts, it is necessary to establish separate calibrations for the individual components, and then determine the ratio from the separate calculations, just as for the original spectral data.

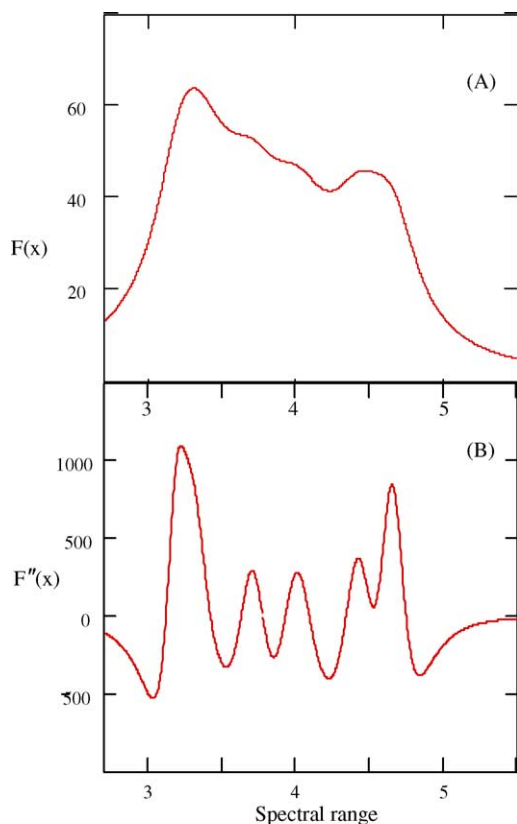


Fig. 5. Simulated spectra created from pure Lorentzian peaks with random position, linewidth and intensity. (A) Typical example of a model spectrum formed as the sum of 10 pure Lorentzian bands. (B) Second derivative of the model spectrum in (A), multiplied by  $-1$ .

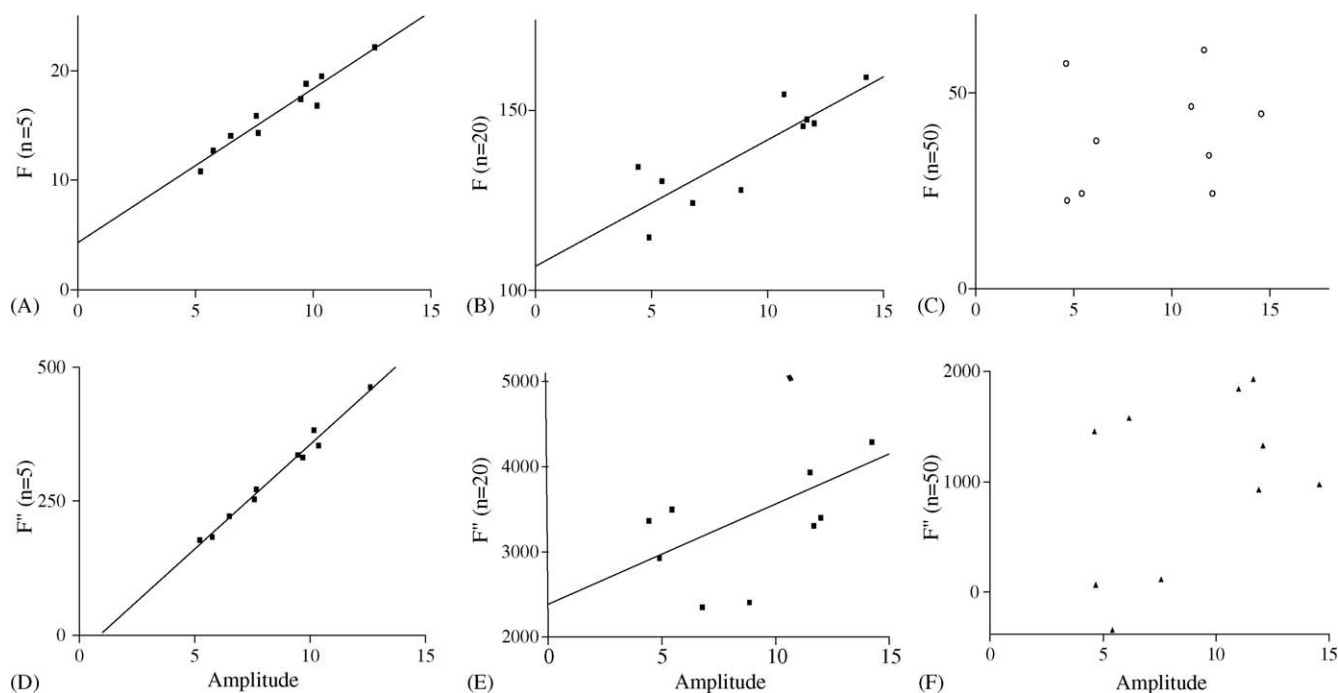


Fig. 6. Correlation between input amplitude value for the unique band at 4 and  $F(4)$ , the total intensity at 4 in the simulated spectrum or at  $F''(4)$ , the corresponding value in the second derivative spectrum. (A, B, C) Correlations for  $F(4)$  vs. amplitude, number of pure Lorentzians in simulated spectrum,  $n = 5, 20$  and  $50$ , respectively. (D, E, F) Similar correlations for  $F''(4)$  and amplitude.

Attempting quantitative analysis with IR spectra on tissue sections poses several problems. In standard analytical chemistry, one identifies and removes interferences, or accounts for them in the model. However, as tissues are extremely complex, standards are difficult if not impossible to manufacture. The exact tissue thickness is also not usually known. Only semi-quantitative analysis is possible (i.e. relative amounts of a component across samples). Attempting quantitation on derivative spectra complicates the problem further, rather than simplifying it.

Even when working with the absorbance spectrum, quantitation relies on the assumption that there are no accidentally degenerate bands from other species. The accidental degeneracy of phosphate and sulfate bands at about  $1235\text{ cm}^{-1}$ , along with the strongest feature of the

amide III series of bands in collagen, is a good example. When the complex, heterogeneous mixture of components in tissue is being studied, the use of either the original or the second derivative requires that the summed bands truly represent the functional groups, that tissue thickness is constant throughout and that the intensity calibration of each band has been properly established. At best, one can typically report only relative composition, or a change in composition.

The second derivative is often used to eliminate baseline anomalies and to reduce the appearance of broad bands. We believe that the not-insignificant amount of smoothing that is required to achieve decent-looking second derivative spectra and the ensuing loss of resolution far outweighs any benefits. This is fairly rudimentary spectroscopy, but it is sometimes possible to lose sight of such simple factors when working with complicated tissue samples. We note that it is still perfectly valid to use the second derivative spectra in the development of diagnostic markers, where no attempt is being made to interpret individual bands, though care with signal to noise ratios and smoothing is still required.

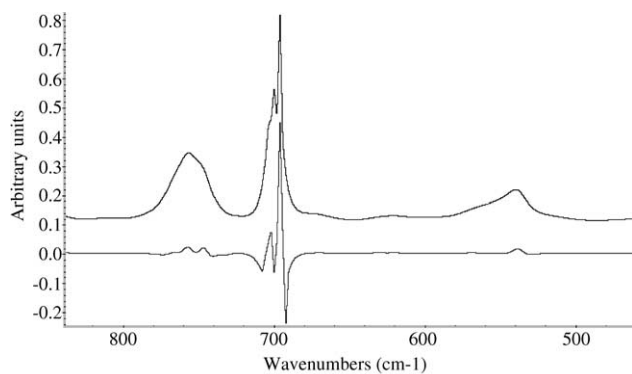


Fig. 7. Portion of the IR spectrum of polystyrene film (top). The second derivative of the spectrum, multiplied by  $(-1)$  (bottom).

#### 4. Summary

Intrinsic differences in tissue composition can produce observable differences in IR spectra. This is the fundamental concept behind IR imaging. The guidelines for sample preparation, data collection and analysis are still being explored and refined. In this paper, we have raised only a few

of the issues that are encountered. Spectral artifacts may arise because certain tissue types are more reflective or denser; therefore, before attempting pre-processing, such as smoothing, baseline correction or normalization, one must check carefully whether it is appropriate, and what may be learned from the presence of the artifacts.

Observed bands are inhomogeneously broadened collections of accidentally degenerate peaks. Under idealized conditions, the second derivative can be used to resolve the location and amplitude of a pure Lorentzian quite well, provided that Lorentzians are not excessively numerous or excessively close. However, there is insufficient evidence to support or refute the use of second derivatives to resolve the locations and concentration of types of molecules in tissue, where the peaks are not perfectly modeled by Lorentzians and other factors may be present.

Data processing software allows a tremendous variety of pre-processing, uni-, bi- and multi-variate analysis methods. If we are to successfully develop this technique and see its implementation become widely accepted, we must be rigorous in our self-assessment and critique of our choices. It is vitally important to be familiar with the sample, to have at least a basic understanding of the theory behind data processing and spectroscopy, to know physical limitations of the technique, such as spectral and spatial resolution and to look at the spectra before drawing conclusions.

### Acknowledgements

Funding was provided by grants to KMG, MRDB and DW from the Canadian Institutes of Health Research and NSERC Canada. MR has been supported through an NSERC PGS-A scholarship, a graduate fellowship from the University of Manitoba and the McCrorie-West Family Fellowship for Alzheimer research. AB is supported in part by an NSERC USRA award. We thank Dr. Michael Mayne for assistance in managing the animals. Technical support was provided by Sharon Allen, Susan Janeczko and Cheryl Taylor-Kashton. Fluorescent immunohistochemistry was performed at the Genomic Centre for Cancer Research and Diagnosis. The authors are grateful to Dr. R. Julian (SRC) and Drs. L. Miller, L. Carr and N. Marinkovic (NSLS) for technical assistance. The SRC is funded by NSF (Award No.DMR-08442). Beamtime at NSLS was made available in part through the Canadian Light Source Participating Research Team at U10B.

### References

- [1] N. Jamin, P. Dumas, J. Moncuit, W.H. Fridman, J.L. Teillaud, G.L. Carr, G.P. Williams, *Proc. Natl. Acad. Sci. U.S.A.* 95 (1998) 4837.
- [2] N.P. Camacho, P. West, P.A. Torzilli, R. Mendelsohn, *Biopolymers* 62 (2001) 1.
- [3] P. Lasch, W. Haensch, N. Lewis, L.H. Kidder, D. Naumann, *Appl. Spec.* 56 (2002) 1.
- [4] K.M. Gough, D. Zelinski, R. Wiens, M. Rak, I.M.C. Dixon, *Anal. Biochem.* 316 (2003) 232.
- [5] D.J. Selkoe, *Physiol. Rev.* 81 (2001) 741.
- [6] G.G. Glenner, E.D. Eanes, H.A. Bladen, R.P. Linke, J.D. Termine, *J. Histochem. Cytochem.* 22 (1974) 1141.
- [7] M.A. Chishti, D. Yang, C. Janus, A.L. Phinney, P. Horne, J. Pearson, R. Strome, N. Zuker, J. Loukides, J. French, S. Turner, G. Lozza, M. Grilli, S. Kunicki, C. Morissette, J. Paquette, F. Gervais, C. Bergeron, P.E. Fraser, G.A. Carlson, P.St. George-Hyslop, D. Westaway, *J. Biol. Chem.* 276 (2001) 21562.
- [8] D.W. Dickson, J. Farlo, P. Davies, H. Crystal, P. Fuld, S.C. Yen, *Am. J. Path.* 132 (1988) 86.
- [9] S.M. Gentleman, C. Bruton, D. Allsop, S.J. Lewis, J.M. Polak, G.W. Roberts, *Histochemistry* 92 (1989) 355.
- [10] M.L. Schmidt, K.A. Robinson, V.M.Y. Lee, J.Q. Trojanowski, *Am. J. Path.* 147 (1995) 503.
- [11] H. Fabian, L.P. Choo, G.I. Szendrei, M. Jackson, W.C. Halliday, L. Otvos Jr., H.H. Mantsch, *Appl. Spectrosc.* 47 (1993) 1513.
- [12] L.P. Choo, D.L. Wetzel, W.C. Halliday, M. Jackson, S.M. LeVine, H.H. Mantsch, *Biophys. J.* 71 (1996) 1672.
- [13] S. Seshardi, R. Khurana, A.L. Fink, *Meth. Enzymol.* 309 (1999) 559.
- [14] S. Litchfield, Z. Nagy, *Acta Neuropathol.* 1001 (2001) 17.
- [15] G.G. Glenner, E.D. Eanes, D.L. Page, *J. Histochem. Cytochem.* 20 (1972) 821.
- [16] P. Lasch, D. Naumann, *Cell. Mol. Biol.* 44 (1998) 189.
- [17] P. Lasch, E.N. Lewis, L.H. Kidder, D. Naumann, *Proc. SPIE-INT Soc. Opt. Eng.* 3920 (2000) 129.
- [18] J. Bandekar, *Biochim. Biophys. Acta.* 1120 (1992) 123.
- [19] W.K. Surewicz, H.H. Mantsch, D. Chapman, *Biochemistry* 32 (1993) 389.
- [20] D.L. Wetzel, S.M. LeVine, *Spectroscopy* 8 (1993) 40.
- [21] P. Lasch, W. Haensch, D. Naumann, M. Diem, *Biochim. Biophys. Acta.* 1688 (2004) 176.
- [22] W.F. Maddams, W.L. Mead, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 38A (1982) 437.
- [23] S. Hawkes, W.F. Maddams, W.L. Mead, M.J. Southon, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 38A (1982) 445.
- [24] W.F. Maddams, M.J. Southon, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 38A (1982) 459.
- [25] H. Mark, J. Workman Jr., *Spectroscopy* 18 (4) (2003) 32.
- [26] H. Mark, J. Workman Jr., *Spectroscopy* 18 (9) (2003) 25.
- [27] H. Mark, J. Workman Jr., *Spectroscopy* 18 (12) (2003) 106.
- [28] A. Savitzky, M.J.E. Golay, *Anal. Chem.* 36 (1964) 1627.
- [29] H. Torii, M. Tasumi, in: H.H. Mantsch, D. Chapman (Eds.), *Infrared Spectroscopy of Biomolecules*, Wiley-Liss, New York, 1996, p. 1.
- [30] J.W. Brauner, C. Dugan, R. Mendelsohn, *J. Am. Chem.* 122 (2000) 677.
- [31] J. Kubelka, T.A. Keiderling, *Am. Chem.* 123 (2001) 12048.