Raman Spectroscopy Studies Of Multipotent Stromal Cells Differentiation For Bone Engineering Application

Raquel de Almeida Rocha Ponzoni

Doctor of Philosophy

University of York

Physics

March 2017

Abstract

Raman spectroscopy (RS) is a label-free method based in the inelastic scattering of laser-light and can be used non-destructively to provide a biomolecular fingerprint of cells. Mesenchymal stromal cells (MSCs) are known for their heterogeneity and, unlike other stem cells, lack a unique marker, thus compromising their application in regenerative medicine. In this work, RS was investigated for its capacity to discriminate MSCs subpopulations, whilst still providing markers of their function. The nuclei of four immortalised clonal MSC lines, which express the same surface proteins, have contrasting differentiation capacities and were not totally discriminated by global gene expression analyses, were investigated. Air-dried and live cell experiments were performed, as well as the induction of one cell line into ostegenesis and adipogenesis. Principal component, linear discriminant and peak intensity ratio (PIR) analyses were performed. Relative differences between proteins and DNA/RNA peaks provided the discrimination of the undifferentiated MSCs towards their differentiation competence. Raman maps linked this discrimination with differences in the nuclear morphology of the cells. Total discrimination of the four cell lines was also achieved by PIRs against DNA/RNA peaks. The discrimination of cell sub-types was mainly driven by differences in the relative intensities of DNA/RNA and protein peaks suggesting fundamental biomolecular differences in the nuclei of these cells. An opposing behaviour between phenylalanine and lipid peaks provided Raman markers that showed clear progression of the adipogenic differentiation. During osteogenesis, the MSCs formed mineralised nodules with an architecture similar to native human bone. The RS investigation of these MSCs showed that the obtained Raman markers could be used to predict differentiation competence in undifferentiated cells and then, followed throughout osteogenesis.

Dedication

To my parents, Wilmar and Marilei.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisors Yvette Hancock and Paul Genever for all their help and guidance throughout these years. Special thanks to Yvette for all her support, availability and who, since I first contacted, has always been excited about new research ideas! Thank you to Paul, whose attention helped me step into the world of biology.

I also would like to thank CNPQ for funding this research. Thanks to Julie Wilson, from the York Centre for Complex Systems Analysis for guiding me with R and with the statistical analyses. Thank you also to Peter O'Toole, head of Imaging and Cytometry from the Department of Biology for the great discussions about the approaches for the live cell analyses. My thanks to Martin Rumsby, also from the Department of Biology for donating the pure lipids used in this research.

I would also like to thank the staff of the Research Services in the Department of Physics, for all the help I have received, in particular to Dave Coulthard. Wayne Twigger, from HORIBA Scientific, thank you for always so promptly helping and fixing the issues with the Raman microscope.

I would like to acknowledge and also thank Kathrin Treacy for proof-reading this thesis, which was done in accordance with the University's Guidance on Proof-reading and Editing.

Thank you to everyone from Yvette's group for all the support in and outside the

office. I must thank James Geraets whose knowledge and advice was always so valuable. Thanks also to Sam Hollings who was always available for my questions about English grammar, biology, programming and British culture!

I thank everyone from Genever's lab, who were always available to help me. Special thanks to Amanda Barnes, Rebecca Pennock, James Fox, David Kuntin and Alice Carstairs.

From outside work, I would like to thank Artur Costrino, Thiago Galvao and Fabrizia Serafim for sharing coffees and laughs, and also for introducing me to board games (they kept me sane, mostly in the final year)! I would also like to thank Eliza and Tiago Ferreira for their special friendship. Thanks to all friends I made here in York, you made the rainy and gloomy days feel like sunny summer afternoons in Brazil. So, thank you Rebecca Otty, Glauber Pacelli, Antonio Pareja Sanchez, Spencer Cole, Kathrin Treacy, Thomas Jacobs, Alexandra Reyes, Rania Zein-Elabdin, Andre Colonese, Krista McGrath and Jenny Ferguson.

I cannot put into words how thankful I am for my husband, Andre Ponzoni, who supported me, and my idea of moving abroad and doing a PhD, from the application to the last words of this thesis. Andre, thank you for keeping me smiling and for putting this work into perspective. I also have to thank my dog Francobollo, who crossed the Atlantic to fill in my days with happy barks and a wagging tail. I also would like to thank my family, especially my parents, Marilei and Wilmar, to whom I dedicate this thesis. You have taught me that perseverance and hard work are the key for achieving your dreams.

Declaration

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references. This work has not previously been presented for an award at this, or any other, University. I hereby give consent for my thesis, if accepted, to be made available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations. Some of the material presented within this thesis forms part of the following papers/manuscripts:

S. James, J. Fox, F. Afsari, J. Lee, S. Clough, C. Knight, J. Ashmore, P. Ashton, O. Preham, M. Hoogduijn, **R. A. R. Ponzoni**, Y. Hancock, M. Coles and P. Genever. Multiparameter analysis of human bone marrow stromal cells identifies distinct immunomodulatory and differentiation-competent subtypes. *Stem Cell Reports*, 4:1004–1015, 2015 – Chapter 3. Joint work. Author carried out the Raman analysis.

R. A. R. Ponzoni, J. Fox, P. Genever and Y. Hancock. Raman spectroscopy assays for discriminatory assessment of heterogeneity and for cell-line comparisons, (in prep) – Chapters 3 and 4.

R. A. R. Ponzoni, P. Genever and Y. Hancock. Study of osteogenic and adiopogenic differentiation of mesenchymal stromal cells using quantitative Raman spectroscopy, (in prep) – Chapter 5.

Chapter 1

Introduction

This chapter explains the theory of the Raman effect, setting the basis for its application in biological and medical sciences. It also contains an introduction on mesenchymal stromal cells and reviews the applications of Raman spectroscopy on the analysis of cells. In this research, Raman spectroscopy is used for the characterisation and discrimination of mesenchymal stromal cells, which are already being used in regenerative medicine. However mesenchymal stromal cells form heterogeneous populations and lack unique markers of their differentiation potential.

1.1 The discovery of Raman

In 1928 Sir Chandrasekhara Venkata Raman revealed to the world a new radiation using nothing more than the sunlight from a heliostat, which had been concentrated into a powerful beam using a 7" telescope objective [1]. He discovered that when a beam of coloured light entered a liquid, a fraction of the light scattered by the liquid was of a different colour. After testing 60 different types of substances he also demonstrated that the nature of the scattered light was dependent on the type of sample [1].

In his early career C.V. Raman had specialised in the study of vibrations and sounds

of stringed instruments. However, in 1921 after a fifteen-day return trip from London to Bombay he became fascinated by the deep blue colour of the Mediterranean sea, and subsequently showed that this colour was the result of the scattering of sunlight by the water molecules – the same argument that Lord Rayleigh used to explain the colour of the sky [2]. C. V. Raman, obsessed with the light scattering phenomenon, started intensive studies measuring the light scattered by liquids and solids in the Indian Association for the Cultivation of Science [2]. His initial experiments were performed by visual observations rather than wavelength measurements. These measurements consisted of using a violet filter to isolate the violet light of the solar spectrum, which was then passed through a liquid sample (Figure 1.1). Most of the scattered light had the same colour as the incident violet beam – the Rayleigh scattered light. What C. V. Raman and his collaborator K. S. Krishnan showed was that if a green filter (which is complementary in colour to the violet filter) was positioned between the scattered light and the observer, a different colour was isolated [1, 2]. However, these initial visual observations were not enough to prove the effect. C. V Raman then used a pocket spectroscope and ultimately a quartz spectrograph (Figure 1.2) to photograph the scattered light and measure its wavelength. Finally, he published the quantitative results in the *Indian Journal of Physics* in 1928 [1]. For this work he was awarded the Nobel Prize in Physics in 1930.

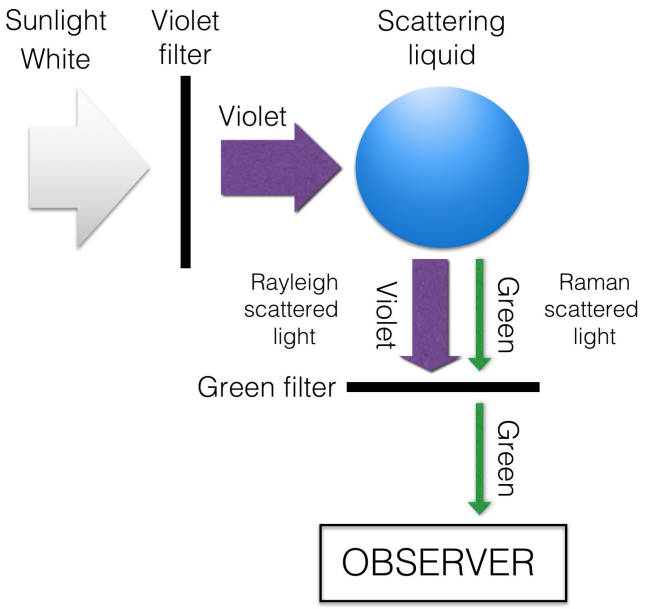


Figure 1.1: C. V. Raman's initial experimental set-up. Schematics of the first visual observations of the scattered light by Raman and his collaborator K. S. Krishnan, as per [2].

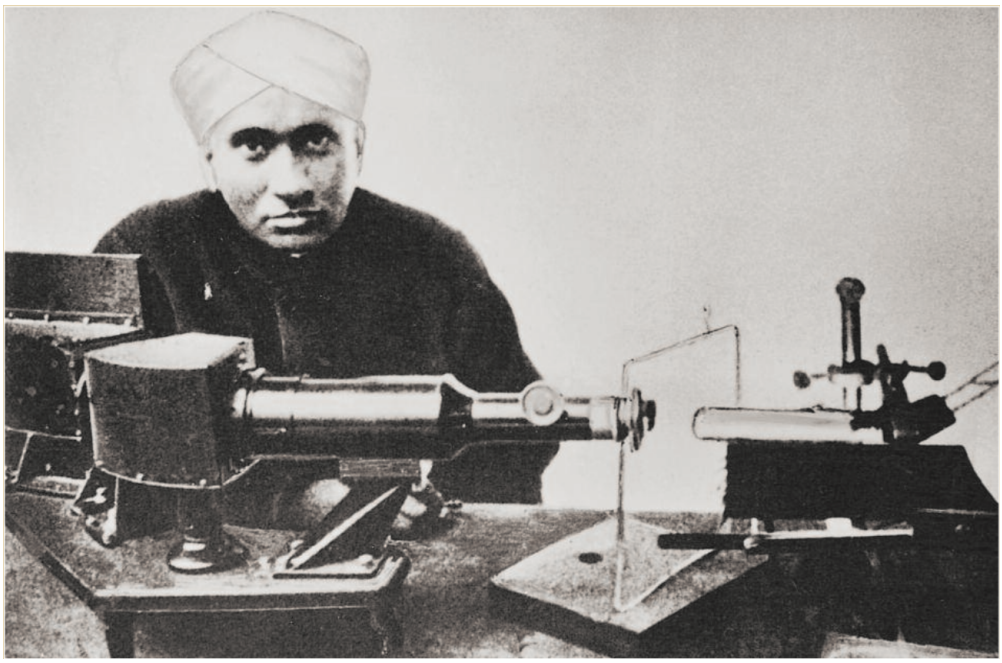


Figure 1.2: C. V. Raman and the quartz spectrograph used to measure the wavelength of the scattered light, as per [2].

1.2 The origin of the Raman spectra

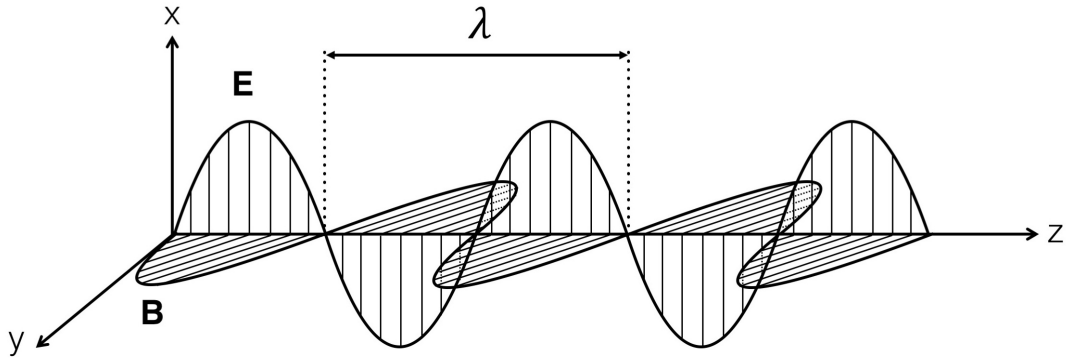


Figure 1.3: A plane-polarised electromagnetic wave, as per [3]. B and E are the magnetic and electric electric fields, respectively, and λ is the wavelength.

The Raman effect is a result of the interaction between light and matter, involving the inelastic scattering of light, which is usually observed in the direction perpendicular to the incident beam. Therefore, to describe the origin of the Raman spectra it is necessary to introduce concepts about electromagnetic radiation and how it interacts with matter.

Light is an electromagnetic wave consisting of an electric field (E) and a magnetic field (B), both perpendicular to each other (Figure 1.3). However, for the Raman effect, only the electric field component is involved.

In a wave, the distance between two points of the same phase is the wavelength (λ), which is commonly measured in units of nm (nanometer). The frequency (ν) of a wave is defined as the number of waves produced by the source in one second and it is measured in hertz (Hz) such that

$$\nu = \frac{c}{\lambda}, \tag{1.1}$$

where c is the speed of light (3×10^8 m/s) [3].

From the above expression, an important spectral parameter, the wavenumber $\tilde{\nu}$, can be obtained [3],

$$\tilde{\nu} = \frac{\nu}{c} = \frac{1}{\lambda}, \tag{1.2}$$

measured in cm^{-1} for spectroscopy measurements.

When a molecule interacts with an electromagnetic field, a transfer of energy to the molecule will only occur if Bohr's frequency condition is satisfied [3]. In other words,

$$\Delta E = h\nu, \tag{1.3}$$

where ΔE is the energy difference between two quantised states and h is Planck's constant (6.63×10^{-34} $\mathrm{m}^2\mathrm{kg/s}$). Therefore, the wavenumber $\tilde{\nu}$ is directly proportional to the transition energy [3].

The magnitude of ΔE depends on the origin of the transitions, which for the Raman effect originates from a vibrational change in state and accompanying change in polarisation of the molecule [3, 4]. According to classical theory, the strength of the electric field (E) of the incident laser beam (ν_0), ranging from the ultraviolet (UV) to the near-infrared region of the electromagnetic spectrum, can be written as

$$E = E_0 \cos(2\pi\nu_0 t), \tag{1.4}$$

where E_0 is the electric field amplitude and t is time. This electric field will perturb the electronic charge distribution in the molecule, inducing a dipole moment P, with strength determined by

$$P = \alpha E = \alpha E_0 \cos(2\pi\nu_0 t), \tag{1.5}$$

where α is the molecular polarisability [3, 4].

For a diatomic molecule vibrating at frequency ν_m, the atomic displacement q away from equilibrium is

$$q = q_0 \cos(2\pi\nu_m t), \tag{1.6}$$

where q_0 denotes the vibrational amplitude of the molecule. α can be approximated to the linear function for small displacements, q,

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial q}\right) q \tag{1.7}$$

where α_0 is the polarisability at the equilibrium position [3].

Combining Equation 1.5 and 1.7, the induced dipole moment becomes

$$P = \alpha_0 E_0 \cos(2\pi\nu_0 t) + \left(\frac{\partial\alpha}{\partial q}\right) q_0 E_0 \cos(2\pi\nu_0 t)\cos(2\pi\nu_m t). \tag{1.8}$$

By further use of the trigonometric formula for the product of two cosine functions (Appendix A, equation A.1.1) Equation 1.8 is re-written as

$$P = \alpha_0 E_0 \cos(2\pi\nu_0 t) + \frac{1}{2}\left(\frac{\partial\alpha}{\partial q}\right) q_0 E_0 \left[\cos\{2\pi(\nu_0+\nu_m)t\} + \cos\{2\pi(\nu_0-\nu_m)t\}\right]. \tag{1.9}$$

The first term in Equation 1.9 corresponds to Rayleigh (elastic) scattering processes resulting from an oscillating dipole that radiates light with the same incident frequency ν_0. The other two terms represent the inelastic light scattering that contains information about the molecular system due to its dependency on ν_m, or in other words, the Raman scattering [3, 4]. Specifically, the second term, which is the anti-Stokes scattering, is characterised by the frequency shift $\nu_0 + \nu_m$ and represents emitted radiation that is blue-shifted with respect to the incident beam [3,4]. The third term, results from the frequency shift $\nu_0 - \nu_m$, that occurs when the radiation emitted by the induced dipole moment is red-shifted with respect to the excitation frequency, *i.e.*, the Stokes scattering [3,4]. Thus, the peak positions in a Raman spectrum correspond fundamentally to a frequency shift from the incident laser beam frequency (ν_0) due to the molecule vibrational frequency (ν_m (Figure 1.4a) [3].

Rayleigh and Raman scattering events are illustrated in Figure 1.4, where the thickness of the arrows represent the intensity of the scattering. In Rayleigh (elastic) scattering the photon absorption causes a transition from the ground state (E_0) into a virtual excited state, the de-excitation of the molecule results in the emission of a photon with the same incident frequency (ν_0) [5]. In Raman (inelastic) scattering the emitted photon results from a transition from a virtual state to the ground state (anti-Stokes) or to the first

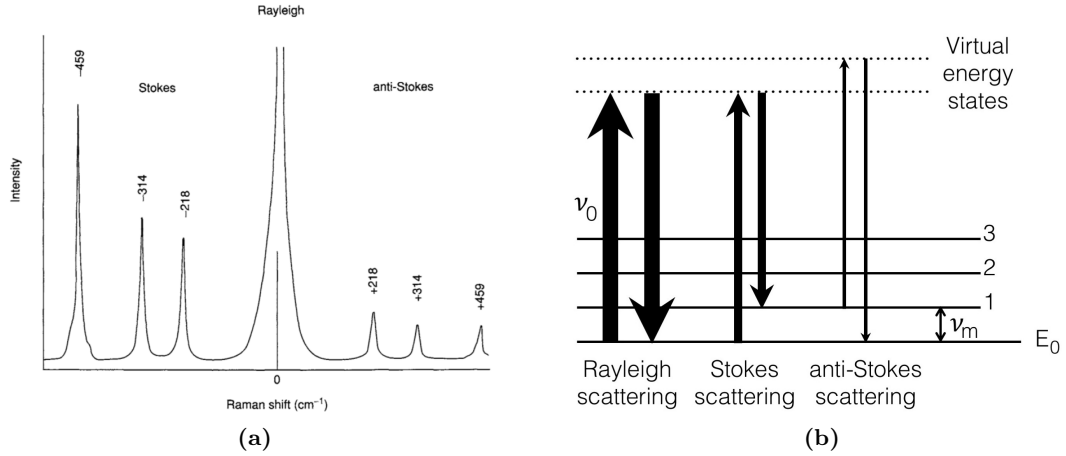


Figure 1.4: Rayleigh, Stokes and anti-Stokes scattering. (a) Raman spectrum of CCl_4 (using 488.0 nm excitation laser) showing the Rayleigh, Stokes and anti-Stokes scattering, as per [3]. (b) Energy level comparisons for the Rayleigh and Raman scattering. The width of the arrows represent the intensity of the scattering. Rayleigh scattering has the same frequency of the incident beam and is stronger, while Raman scattering are weaker and have the same vibrational frequency from the probed molecule.

excited vibration state (Stokes) of the molecule [5]. The virtual states are not stable and also not real states of the molecule. They are created from the polarisation caused by the interaction of the laser with the electrons in the molecule; thus, the energy of these states is determined by the frequency of the laser source used [6].

Anti-Stokes scattering has a reduced intensity relative to the Stokes scattering due to there being a much large population of molecules in the ground state E_0 than in the first excited state, E_1 via the Maxwell-Boltzman distribution [3]. This is further detailed in Subsection 1.2.1, where the classical description of the Raman effect is compared to the view of quantum mechanics. Because the Stokes lines are more intense than the anti-Stokes lines, it is common to measure only the Stokes side of the spectrum. Figure 1.4(a) shows the Raman spectrum of CCl_4 clearly illustrating the Rayleigh, Stokes and anti-Stokes scatterings.

Another very popular technique, that is not to be confused with Raman spectroscopy, is infrared (IR) spectroscopy. In IR spectroscopy, the molecule absorbs energy ΔE from the

IR source causing a change in vibrational state and dipole moment, as opposed to Raman spectroscopy where light ranging from UV to near IR causes a change in vibrational state [3, 5].

1.2.1 Classical versus quantum mechanics description of the Raman effect

Both classical and quantum mechanical interpretations of the Raman effect are based on the relation between the electric field E and the induced dipole moment P [7]. Thus, to compare classical and quantum mechanics interpretations, the vibration of a diatomic molecule will be considered for both treatments.

The harmonic oscillator model, derived from the classical mechanics, is a useful start point to better understand the molecular vibrations responsible for generating the characteristic bands observed in a Raman spectrum. Thus, a diatomic molecule with two masses m_1 and m_2 is considered. The chemical bond is classically regarded as a massless spring of force constant K, where x_1 and x_2 are the displacements of each mass from the equilibrium and r_1 and r_2 are the distances from the centre of gravity (C.G.) for each atom (Figure 1.5).

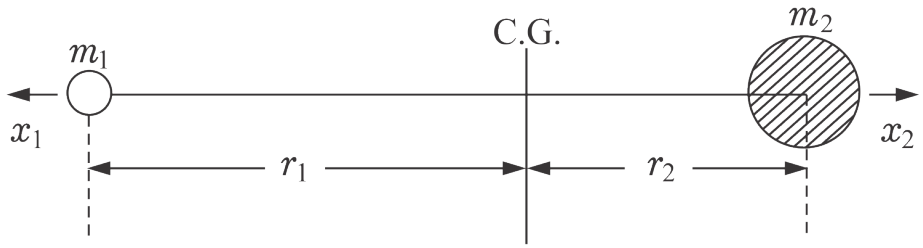


Figure 1.5: Schematic of a diatomic molecule. The masses m_1 and m_2 are positioned at distances r_1 and r_2 from the centre of gravity (C.G.), where x_1 and x_2 are the displacements of each mass from the equilibrium, as per [3].

In this system, although both atoms oscillate with different amplitudes, they share the same frequency, with both masses passing simultaneously through their equilibrium positions. Since the amplitudes are inversely proportional to the atom masses, the C.G. is kept stationary. Thus, the conservation of the centre of gravity requires the following relationships:

$$m_1 r_1 = m_2 r_2, \tag{1.10}$$

$$m_1(r_1 + x_1) = m_2(r_2 + x_2). \tag{1.11}$$

Since the chemical bond is considered as a massless spring obeying the Hooke's law, the restoring force is expressed as

$$f = -K(x_1 + x_2). \tag{1.12}$$

Finally, combing Equations 1.10, 1.11 and 1.12 and using Newton's equation of motion ($f = m\frac{d^2x}{dt^2}$), the following equation is obtained:

$$\frac{m_1 m_2}{m_1 + m_2}\left(\frac{d^2x_1}{dt^2} + \frac{d^2x_2}{dt^2}\right) = -K(x_1 + x_2), \tag{1.13}$$

which can be re-written as

$$\mu\frac{d^2q}{dt^2} = -Kq, \tag{1.14}$$

where μ is the reduced mass and q is the displacement [3]. The solution of this differential equation is Equation 1.6, which has been used to describe the classical origin of the Raman effect in the previous Section 1.2. Thus, for the harmonic oscillator, the displacement of the two atoms as a function of time varies periodically as a sine (or cosine) function [5]. Therefore, the classical vibrational frequency for the system can be described as [3, 5]

$$\nu = \frac{1}{2\pi}\sqrt{\frac{K}{\mu}}. \tag{1.15}$$

This shows that the frequency (or wavenumber) of the diatomic oscillator depends on the force constant K, which is a function of the bond energy between the two atoms, and on the atomic masses of the two atoms involved in the vibration. However, this diatomic harmonic oscillator is not an appropriate description for larger molecules, which have a more complex nature of vibration [5].

For the classical diatomic harmonic oscillator the potential energy (V) is given by

$$V = \frac{1}{2}Kq^2, \tag{1.16}$$

which determines a quadratic dependence between the potential V and the displacement q [3,5]. Therefore, a plot of V against q is a parabola that is symmetric to the equilibrium distance between the two atoms.

From the quantum mechanics perspective, the diatomic molecule can be treated as a motion of a single particle, with mass μ and potential energy expressed by Equation 1.16. For this system, the Schrödinger equation is written as [3]

$$\frac{d^2\psi}{dq^2} + \frac{8\pi^2\mu}{h^2}\left(E - \frac{1}{2}Kq^2\right)\psi = 0. \tag{1.17}$$

Considering that ψ is single-valued, finite and continuous, the eigenvalues for the Schrödringer equation are

$$E_i = h\nu\left(\upsilon_i + \frac{1}{2}\right),\ \upsilon_i = 0,\ 1,\ 2,\ \ldots \tag{1.18}$$

where ν is the classical vibrational frequency (Equation 1.15) and υ_i is a quantum number that can only have integer values [3,5]. Thus, from the quantum mechanics description, it is know that molecules can only exist in quantised energy states, meaning that the vibrational energy is not continuously variable and only assumes discrete values.

Figure 1.6a shows the vibrational levels for the quantum mechanical oscillator, along with the probability functions for the internuclear distance x_e within each energy level.

These distances (x_e) are expressed as probabilities due to the Heisenberg's uncertainty principle, which states that during a vibration it is not possible to know both the position and mass.

When both, classical and quantum treatments are compared, several marked differences arise. For the classical description, the total energy E is zero when q is zero. Conversely, for quantum mechanics, the lowest energy state (*i.e.*, $\upsilon = 0$) has the energy of $\frac{1}{2}\nu$, which results from Heisenberg's uncertainty principle [3]. While in classical mechanics the energy of this system can change continuously, in quantum mechanics it can change only in units of $h\nu$. Another difference, is that the vibration in confined within the parabola of the potential energy in classical mechanics, whilst in quantum mechanics, the probability of finding q outside the parabola is not zero (Figure 1.6a) [3].

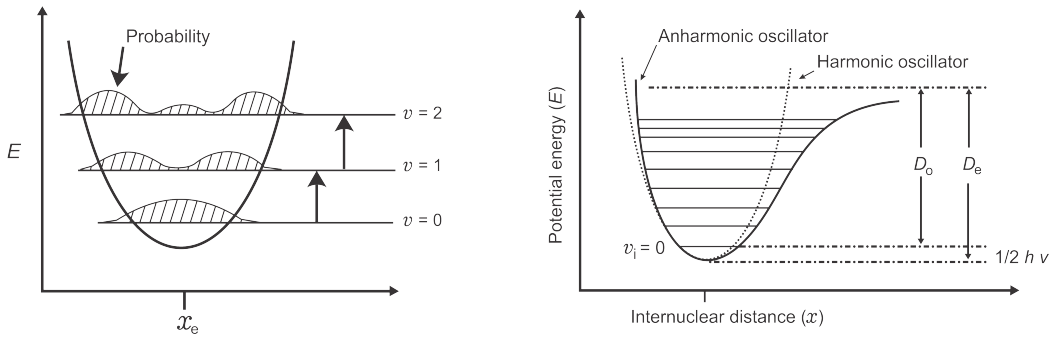
A more realistic description introduces the anharmonicity (Figure 1.6b), which is the result of a dipole moment that is not linearly proportional to the nuclear displacement coordinate. The energy for the anharmonic oscillator is given by

$$E_\nu = h\nu_e \left(\upsilon + \frac{1}{2}\right) - h\chi_e\nu_e \left(\upsilon + \frac{1}{2}\right)^2 \tag{1.19}$$

where ν_e is the frequency corrected for the anharmonicity and $\chi_e\nu_e$ indicates its magnitude [3, 5].

The deviation of the anharmonic oscillator from the harmonic oscillator becomes greater as the vibrational quantum number increases and the separation between adjacent levels becomes smaller for higher vibrational levels until a dissociation limit is reached [5]. For the harmonic oscillator only transitions to adjacent levels (fundamental transitions, i.e., $\Delta\upsilon = \pm 1$) are allowed, while for the anharmonic oscillator, overtones ($\Delta\upsilon = \pm 2, \pm 3, \dots$) and combination bands are also weakly allowed by selection rules [3,5].

Among the $\Delta\upsilon = \pm 1$ transitions allowed, the $\upsilon = 0$ to $\upsilon = 1$ is the most strong, as expected from the Boltzmann distribution law, for which the population ratio of both



(a) harmonic oscillator **(b)** anharmonic oscillator

Figure 1.6: (a) Vibrational levels for the quantum mechanical oscillator showing the probability functions for the internuclear distance x within each energy level. (b) Anharmonic quantum oscillator, which is a result of a dipole moment that is not linearly proportional to the nuclear displacement x_e. D_e and D_o are respectively the theoretical and spectroscopic dissociation energies. Both figures are reproductions as per [5].

levels is given by

$$\frac{P_{v=1}}{P_{v=0}} = e^{-\Delta E/kT}, \tag{1.20}$$

where ΔE is the energy difference between the two states, k is the Boltzmann's constant (1.3807×10^{-23} J/K), and T is the absolute temperature [3, 6, 8]. Therefore, at room temperature, and prior to the interaction with the laser, most molecules are likely to be in the ground vibrational state, thus explaining why the Stokes scattering is more intense than the anti-Stokes, as previously illustrated in Figure 1.4 [3, 6, 7].

Since the intensity of the anti-Stokes Raman scattering depends on population of the first excited state, the intensity ratio between anti-Stokes and Stokes scattering is given by the ratio of occupancy in both levels [6–8],

$$\frac{I_R(\nu_0 + \nu_m)}{I_R(\nu_0 - \nu_m)} = \frac{(\nu_0 + \nu_m)^4}{(\nu_0 - \nu_m)^4} e^{-h\nu_m/kT}. \tag{1.21}$$

This equation also allows the determination of the sample temperature. However, in order to do so, the Stokes and anti-Stokes intensities have to be transformed into absolute values and correct with respect to the spectral sensitivity of the spectrometer [8].

Finally, the intensity of the Raman scattered radiation I_R is given by

$$I_R \propto \nu_0^4 I_0 N \left(\frac{\partial \alpha}{\partial q}\right)^2, \tag{1.22}$$

where ν_0 is the frequency of the exciting laser, I_0 is the incident laser intensity, N is the number of scattering molecules in a given state, α is the polarisability of the molecules and q is the displacement (vibrational amplitude) [5]. This indicates that the Raman intensity depends on the laser frequency and varies with the fourth power of the observed frequency [5, 7]. The ν^4 factor is derived from the classical description of scattering from an oscillating induced dipole [7].

1.3 Applications of Raman spectroscopy

1.3.1 General applications

Raman spectroscopy (RS) can be used in a non-destructive manner to access chemical and structural information from a diverse range of materials and compounds. Initial applications of RS remained limited due to its weak signal and lack of sensitivity. However, with improvements in instrumentation, such as the development of better excitation sources to the advent of laser beams, spectrographs, holographic filters and charge-coupled devices (CCD), RS can now be applied to a wide range of materials and samples, being used to probe solids, liquids and gases [4, 9].

Due to minimal or no sample preparation requirements, RS has a wide range field of applications, covering areas from solid state physics and material science to biomedical and biological sciences [10]. In this vast range of applications, some examples include the structural characterisation of semiconductors and superconductors [11], identification of mineral phases and classification of rocks in mineralogy [12–14], analysis of fibres, drugs, inks, paints and biological fluids in forensic sciences [11, 15, 16]. As well as the characteri-

sation of archaeological artefacts (shell ornaments, pottery, ceramics) and bioarchaelogical samples (bone, teeth, etc.) in art and archaeology [17–21].

Recently, RS has been employed in the diagnosis of diseases, mainly in cancer. It has been used to compare noncancerous to cancerous human breast tissues [22], to identify and grade prostate cancer [23], to study the changes in the lipid content in hormone-treated breast and prostate cancer cells [24], and also to grade osteosarcoma cells [25]. RS was also used to monitor and analyse the formation of lipid droplets in hepatocytes exposed to lipolysis products [26]. RS has been coupled with optical micro-manipulations such as laser tweezers and microfluidic systems to allow the manipulation and biochemical analysis of biological particles in suspension [27]. Several probes have also been designed for fibre optic Raman spectroscopy and used for *in vivo* diagnostics of different types of cancer [28].

RS is also used in pharmaceutical research for molecular-based drug discovery, for designing drug delivery systems and also to assess the quality of finished products [29]. With such plasticity in applications and having weak scattering for water (in contrast to the strong absorption suffered by IR spectroscopy [3]), RS is an ideal technique to be applied in the fields of medicine and biology.

1.3.2 Probing cells with Raman spectroscopy

The Raman spectrum of a cell provides a biochemical fingerprint from which spectral features can be identified and then correlated to the cell biochemistry. Applications of RS applied to cells include cell phenotyping [23, 30–32], assessment of cell differentiation [33–37], identification of pathogens at the single cell level [38], monitoring population changes [39], probing different stages of the cell cycle [40] and evaluating cell apoptosis [41–44]. RS can be applied to live [31, 37, 41], fixed or air-dried cells [45, 46].

RS can be used for imaging, providing spatial distribution of the biochemical species

and cellular compartments [47,48]. It has been used to distinguish vesicles and organelles, such as the nucleus, the endoplasmic reticulum and the Golgi aparatus, without the use of labelling or staining [47, 49]. RS was used to show the subcellular organisation of the cytoplasm, in particular the distribution of mitochondria [50], the spatial distribution and chemical features of nucleoli [48] and to study apoptosis via mapping of the DNA and protein distribution in apoptotic HeLa cells [43]. It has also been effective in following the uptake of drugs in the studies of drug delivery systems [50, 51].

Due to the complexity of the spectra generated by cells multivariate techniques, such as principal component analysis (PCA), linear discriminant analysis (LDA) and clustering analysis, are used for data classification. RS along with PCA-LDA has been shown to discriminate between lung cancer cells, lung epithelial cells and lung fibroblasts, with the main spectral differences between these cells residing in their nuclei [52]. Crow *et al.* (2005) demonstrated that RS combined with PCA-LDA was able to discriminate between prostatic adenocarcinoma cell lines of varying degrees of biological aggressiveness [32]. Most of the variance observed between the cell lines by Crow et al. were accounted by differences in nucleic acid peaks (721 cm^{-1}, 783 cm^{-1}, 1305 cm^{-1}, 1450 cm^{-1} and 1577 cm^{-1}), DNA phosphate backbone band ($\sim$1096) and unordered protein vibrations (1250 cm^{-1} and 1658 cm^{-1}).

Another method used to discriminate between cells using RS involved peak ratios. Huang *et al.* (2015) achieved discrimination of nasopharyngeal cancerous from non-cancerous tissues by using either peak intensity ratios, or PCA-LDA analysis [53]. The use of ratios between specific Raman spectral peaks was also used to determine the lipid composition of lipid droplets in hepatocytes [26] and in hormone-treated breast and prostate cancer cells [24]. Furthermore, the intensity ratio between peaks related to unsaturated lipids and total lipids in breast and prostate cell spectra has been proposed as a molecu-

lar marker for normal and tumour cells, indicating that RS combined with a ratiometric analysis could also provide diagnostic markers [54].

RS is also able to discriminate between differentiated and non-differentiated embryonic stem cells (ESC). Chan *et al.* (2009) were able to discriminate the spectra of non-differentiated ESCs from their differentiated cardiomyocites by applying PCA-LDA to the spectra obtained [36]. A decrease in the magnitude of peaks correlated to nucleic acid vibrations at 813 cm^{-1} and 788 cm^{-1} of differentiated murine ESC in comparison to undifferentiated cells was noted by Notingher *et al.* (2004), who suggested the feasibility of using RS to monitor differentiation [55]. In another work, Notingher *et al.* (2004) used PCA and ratios between the peak areas of nucleic acid and phenylalanine vibrations, respectively at 813 cm^{-1} and 1005 cm^{-1}, to measure mRNA translation with these being used as markers for the differentiation state of murine ESCs [56]. Schulze *et al.* (2010) demonstrated that multivariate analyses were not required to discriminate undifferentiated from fully differentiated human ESC spectra, showing that protein to nucleic acid intensity ratios could be used as indicators of the differentiation state and type of differentiation [33]. Therefore, it can be seen that inherent to the Raman spectra of single cells are important information that can be used for detailed label-free cell discrimination and classification.

The use of RS for single cell identification is an attractive complement to established optical techniques. An increasing interest in using Raman-based applications was witnessed in the last two decades, in particular, due to the capability of RS to probe molecular signatures without the use of external labels [57]. Unlike techniques other methods that employ the use of X-rays, gamma rays or UV light which are harmful to life processes, spectroscopic methods employing visible or infrared light are deemed to be the least harmful [58].

Methods that make use of fluorescence probes and molecular stains, such as confocal fluorescence microscopy and two-photon microscopy, are frequently used in the biological sciences due to their powerful ability to reveal the inner workings of a cell [59]. Nowadays there are many thousands of fluorescent probes conceived to provide a means of labelling almost any aspect of a biological system, and the discovery of fluorescent gene products, in particular green fluorescent protein (GFP), has allowed the tagging of protein components of living systems, providing their visualisation [60]. However, considerable pitfalls for fluorescence techniques include the difficult delivery of labels, the issue that some labels work only for dead cells, that they perturb the very processes they are intending to study and also the barrier that exists with these methods suffering from a limited lifetime as a result of photobleaching [57]. Techniques that require a dye or chemical label are frequently questioned as to whether the dye interferes, or not, in the viability of the cells that are studied, or whether the diffusion of a receptor or a ligand is affected by the presence of the label [40].

The application of RS for the analysis of cells and tissues faces two major technical limitations: (i) the integration time per spectrum is usually long due to the low signal produced by biological samples; and (ii) autofluorescence can easily overwhelm the weak Raman signal [57]. Other techniques based on the Raman effect, such as surface enhanced Raman spectroscopy (SERS) and surface enhanced-resonance Raman spectroscopy (SERRS), try to overcome weak Raman scattering. However, these methods cannot be considered totally label-free since they make use of the adsorption of molecules by the substrate or dye molecules to enhance the Raman signal [4]. Other variations are the coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS), which emerged in the recent years, but are used only for rapid chemical imaging of biological samples [57]. Instruments that perform parallel acquisition from multiple points (line

laser) introduced a high-speed hyperspectral Raman imaging method that can be used to dynamically observe biological processes such as cell mitosis and apoptosis [61, 62] and the dynamical behaviour of specific molecules [63, 64]. Non-destructive Raman sampling depends on the laser power, wavelength and acquisition time, with the laser power and acquisition time being shown to cause drastic changes to the cell morphology [41].

Another difficulty concerning the application of RS for the analysis of cells is related to the methods used for data analysis. First, most analyses lack a parameter that determines whether enough spectra have been collected from the cell population to accurately represent it, with studies indicating that for point acquitisition anywhere between 5 to 50 spectra would be typical [65]. Some studies even combine live with air-dried datasets in order to augment the number of obtained spectra [33]. This is a bad practice because these spectra will reflect the biochemical differences between the live and the dried cell. RS studies have noticed a decrease in the intensity of DNA peaks in dead cells when compared to the spectra of live cells [41, 44].

In general, data analyses of Raman spectra range from only performing peak identifications and then correlating them with the literature databases [34, 41, 66], to multivariate methods as PCA, LDA and clustering analysis [31–33, 36, 67–69].

Peak identification may be inaccurate due to there being several biomolecule components that may contribute to the same peak [70]. For multivariate analyses to be effective on low signal-to-noise data, often a broad range of data pre-processing methods, such as baseline corrections, background removal, data smoothing and data normalisation (all of which can alter the spectra) need to be applied [3, 4].

Another important consideration for data analysis is that Raman spectra cannot have their peak intensities directly compared as this is dependent on several factors, such as the amount of material at the focal point, how well defined the focal plane is, the polarisation

direction of the laser relative to the sample, and also the conformational changes to the molecules [4]. The use of data normalisation allows for spectral comparison, however, normalising a cell's spectrum to any of its internal components which change due to cell type, cell lineage or even cell history, eliminates changes related to this constituent and enhances others [33]. Furthermore, while PCA is useful for data classification and illustrating the variation between samples, it does not reveal physically or chemically interpretable information of the sample that can quantify biomolecular changes [70, 71].

In ratiometric analyses, the comparison of spectra by means of peak intensity ratios, for example, circumvents the issue of directly comparing peak intensities from different spectra, as well as provides biomolecular markers [72]. Although simple, ratiometric analyses of Raman data carry limitations involving error-prone manual peak fitting, peak de-convolution and assumptions regarding peak position, width and shape that have to be made prior of the analysis [73, 74].

Thus, if experiments are executed avoiding cell autofluorescence or cell damage, and if the chosen data analysis is carefully performed, RS holds an advantage over other conventional methods due to label-free measurements reflecting the cell's components. RS can then be used to perform cell discrimination and classification or to monitor changes in the cell biochemistry during differentiation as justified by the vast amount of supportive literature reviewed here.

1.3.3 Assessment of cell populations

Cell populations that carry genetically identical cells can sometimes display some degree of heterogeneity by behaving differently to the same controlled environmental conditions. Identical cells might become different when they encounter different environment conditions, which then induces different cell responses [75]. However, phenotypic changes can

occur even in a controlled environment and these changes may be the result of asymmetric segregation of intrinsic determinants during cell division, which will lead to different gene expression patterns [75]. Even clonal populations under controlled culturing conditions have been shown to have some degree of heterogeneity, which emerges spontaneously [75] and can be noticed, for example, in different drug responses or transient drug tolerance [76, 77].

A population that shows significant heterogeneity is the one of mesenchymal stromal cells (MSC). These cells are isolated from adult and foetal tissues and can be induced to differentiate into linages of mesenchymal tissues, which include bone, cartilage and fat [78]. MSCs are most commonly derived from the bone marrow, forming heterogeneous populations that contain different levels of osteogenic (bone), adipogenic (fat) and chondrogenic (cartilage) competent cells, as well as other types of stromal cells [79–81].

Heterogeneity can be routinely depicted using several experimental techniques including microscopy, flow cytometry, electrophysiology, and global transcriptomic methods [82, 83]. Microscopy and flow-cytometry are nowadays essential tools for monitoring heterogeneity; however, they usually make use of labelling dyes, relying on fluorescence monitoring [83], and thus suffer from the same issues previously discussed, such as fading, photobleaching and chemical perturbation to the cell. Global genetic methods provide information about specific genes whose expression levels can be used to determine heterogeneity; however, these methods are usually time-consuming, expensive and destructive [83].

The identification of the MSCs is often done by using a panel of non-specific cell surface proteins which vary amongst research groups and may still result in heterogeneous populations containing colony-forming and different differentiation competent features, suggesting the existence of other MSCs subtypes [84]. All population-averaged assays

capture a population state and report how these states change in response to perturbations [82]. However, population distributions that rely on data being averaged across thousands or millions of cells in a sample, mask heterogeneity at the single-cell level by ignoring the presence of rare or small subpopulations of cells [82, 85].

Since spectroscopic methods provide fine-inspection of intracellular information, they can also be applied to characterise cell subpopulations and thus, provide information about heterogeneity. These fine-inspections, combined with single-cell interrogation, can show that for even seemingly identical cells phenotypic differences are always present in a population [82].

MSCs can be used as in vitro models for basic research and have been used in clinical-trials of cell-based therapies for the treatment of chronic orthopaedic disorders, such as osteoarthritis and osteoporosis [86,87]. However, their application into regenerative medicine is compromised by the inability to reliably characterise their subtypes using standard methods, such as cell surface markers or global genetic profiling [88]. Thus, they configure an important candidate to investigate the ability of RS to perform subpopulation discrimination, providing information about MSC heterogeneity. More details about these cells, their isolation and application are presented in the next section.

1.4 Mesenchymal Stromal Cells

Mesenchymal stromal cells (MSCs) were first identified by Friedenstein *et al.* (1968 and 1970) as a subpopulation of mouse and guinea pig bone marrow adherent cells capable of colony formation, and with fibroblast-like characteristics in that they possess the capability to differentiate into osteoblasts (bone) [89, 90]. Originally named as colony-forming unit-fibroblast (CFU-F), these cells were detected in human bone marrow in 1980 by Castro-Malaspina *et al.* [91]. Caplan, in 1991, proposed that these were mesenchymal *stem* cells

with the capability to differentiate into all cells of mesodermal lineage [92], stimulating their investigation for tissue regeneration. Their multilineage differentiation capability was proven by Pittinger *et al.* in 1999 who verified that expanded colonies derived from single cells could be induced to differentiate into osteoblasts (bone), adipocytes (fat) and chondrocytes (cartilage) [78].

MSCs display a heterogeneous morphology with several different terms been used to describe them, such as fibroblastoid cells, spindle shaped, flattened cells or very small round cells [93]. Apart from the diversified morphology, human MSCs express not only transcript characteristics of mesenchymal lineages, but also transcripts typical to endothelial, epithelial and neuronal cell lineages [94]. Given that some cells in these fibroblast-like plastic-adherent cultures have progenitor properties and display multipotent differentiation capacity, whereas others are totally devoid of this ability [93], it was clear that these were very heterogeneous populations and a convention in their nomenclature, as well as in the criteria adopted for their identification, was soon proposed.

Horwitz *et al.* (2005) suggested that these cells, regardless of the tissue from which they were isolated should be termed multipotent mesenchymal stromal cells, while the term mesenchymal stem cell should only be used for cells that met the specified stem cell criteria – the ability to produce daughter cells (self-renewal) as well as differentiate to multiple phenotypes [95].

One year later, the Mesenchymal and Tissue Stem Cell Committee of the International Society for Cellular Therapy proposed three criteria to define human multipotent mesenchymal stromal cells: firstly, the MSCs must be plastic-adherent when maintained in standard culture conditions; secondly, the MSCs must express CD105 (endoglin, SH2), CD73 (ecto-5'-nucleotidase) and CD90 (Thy1), and lack expression of haematopoietic markers CD45, CD34, CD14 or CD11b, CD79α or CD19 and HLA-DR surface molecules;

and thirdly, the MSCs must differentiate to osteoblasts, adipocytes and chondroblasts *in vitro* [96]. However, this list of conditions does not fully characterise the MSCs, which have been noticed to be metabolically active, secreting a vast array of cytokines along with the components of the extracellular matrix [97]. Other expression markers, complementary to those proposed by the committee, are used for MSC enrichment [93], such as the CD271 (low-affinity nerve growth factor receptor, LNGFR) [98] and the CD146 (melanoma cell adhesion molecule, MCAM), envisaging the selection of MSCs that are clonogenic and osteogenic *in vivo* [80,99]. Unlike haematopoietic stem cells (HSCs), which have relatively effective cell surface markers for their isolation [93], MSCs lack recognised and widely adopted markers.

Apart from bone marrow-derived cells, MSC-like populations have also been obtained from adipose tissue, placenta, skin, umbilical cord blood, umbilical cord perivascular cells, umbilical cord Wharton's jelly[1], dental pulp, amniotic fluid, synovial membrane[2] and breast milk [102]. Aside from MSCs, the bone marrow is also home to the HSCs, which give rise to blood and endothelial cells. However, MSCs represent only 0.001% to 0.01% of the population of cells in the bone marrow, being 10 fold less abundant than haematopoietic stem cells [78]. Although the MSCs have been thought of as being the same general population of cells regardless of the tissue of provenance, research has suggested that their gene expression reflects the tissue of origin, indicating that MSC characteristics and differentiation potential may vary according to the source of tissue [94, 102–104].

Nowadays the term 'mesenchymal stem cell', which was originally introduced referring

[1] A specialised tissue wrapped around the umbilical cord blood vessels, characterised by the presence of cellular elements and extracellular matrix, with its primary function being to protect the umbilical cord [100].

[2] The membrane that lines the ligament surrounding a freely movable joint. This membrane secretes the synovial fluid that lubricates the layers of cartilage forming the articulating surfaces of the joint [101].

to cultures of bone marrow stromal cells, was extended to non-skeletal tissues and is used to denote cultures of fibroblastic cells from virtually every tissue or organ [105]. Bianco *et al.* (2008) suggest that since CFU-Fs are present in other tissues beyond the bone marrow their nomenclature should denote the specific tissue from where they were isolated, *e.g.*, BM-CFU-F for those isolated from the bone marrow, AT-CFU-F for those isolated from adipose tissue, etc. [106]. Recently, Bianco and Robey (2015) have suggested that the term 'mesenchymal stem cell' should be replaced once there is no evidence that a stem cell with the characteristics of embryonic mesenchyme, or with multipotency beyond the skeletal lineages exists in postnatal bone marrow or in the postnatal organism [105]. They also suggest that these bone marrow progenitors of skeletal lineages should be called skeletal stem cells (SSC) [105]. This diversity of names therefore, clearly reflects the broad range of characteristics that make the MSCs important for medical applications. In this thesis, the studied cells are mesenchymal stromal cells derived from the bone marrow and the term MSC is used throughout the methods and results chapter to designate them. However, in the literature the term MSC is used regardless of the tissue of provenance, and when possible the cells' origin will be identified in the text.

1.4.1 Applications of MSCs in regenerative medicine

Regenerative medicine includes surgery, surgical implants, the use of biomaterial scaffolds, organ transplants and tissue engineering to manage chronic diseases or injuries in patients whose own regenerative mechanisms have failed to do so; the field is centrally focused on human cells, which may be somatic, adult stem cells or embryonic-derived cells [107]. MSCs are an attractive option for regenerative medicine as they remove the ethical issues associated with the use of embryonic stem cells because they can be used for autologous transplantation, which means that they can be isolated from the patient, expanded *in*

vitro, manipulated and then infused into the same patient [86].

Apart from their trilineage differentiation capacity, MSCs also have additional properties that aid tissue regeneration, such as the ability to secrete a wide range of growth factors that have trophic effects on surrounding cells, and the capacity to coordinate differentiation in tandem with differentiated and undifferentiated host cells [86]. MSCs also have immunomodulatory properties and can act on both adaptive and innate immune systems by suppressing T cells[3], suppressing dendritic cell[4] maturation, reducing B cell[5] activation and proliferation, inhibiting proliferation and cytotoxicity of NK cells,[6] and promoting the generation of regulatory T cells [102]. All of these properties have attracted the attention of clinical use for the MSCs, in particular for musculoskeletal diseases.

MSCs have been used for clinical trials targeting musculoskeletal diseases such as frac-

[3]Also known as T lymphocytes, these are cells from a population of lymphocytes that are the principal agents of cell-mediated immunity. T cells are derived from the bone marrow and migrate to the thymus to mature. Once matured they circulate between the lymph nodes and the bloodstream as naive T cells [101].

[4]One of the main types of antigen-presenting cell in the immune system, responsible for presenting antigen to naive T cells and inducing them to become effective components of the adaptive immune response [101].

[5]Also known as a B lymphocyte, this cell is derived from stem cells in the bone marrow, but unlike T cells do not mature in the thymus. In birds, this cell matures in the bursa of the cloaca – hence the name *B* cell. Each B cell has a unique set of receptor molecules on its surface, designed to recognise a specific antigen [101].

[6]Known as the natural killer (NK) cell, this cell type is a lymphoid cell that recognises and destroys tissue cells infected with pathogenic organisms. These cells are of important early line defence against infection, before more antigen-specific mechanisms of the adaptive immune response are mobilised by the T cells. NK cells become activated in response to interferons, or the release of cytokines and bind to target cells, releasing cytotoxic granules onto the surface of the target, which then penetrates the plasma membrane and induces programmed cell death (apoptosis) [101].

ture nonunions, avascular osteonecrosis[7], osteogenesis imperfecta[8], hypophosphatasia[9] and osteoarthritis [86, 87, 110], corresponding to 16% of the clinical studies started in the last decade [111]. Musculoskeletal diseases are considered the most common cause of severe long-term pain and physical disability and with an ageing population their incidence is projected to substantially increase [112]. MSCs have also been investigated as a therapeutic means towards multiple sclerosis[10], Crohn's disease[11], systemic lupus erythematosus[12] and renal transplantation [110].

Another application of MSCs derived from both adipose-tissue and bone marrow is on animal models for Graft-versus-host disease (GVHD) – a complication that occurs after stem cell or bone marrow transplantation, where the transplanted cells attack the

[7]A disease that results in the death of bony tissues, that are usually adjacent to a joint surface and are enclosed by avascular cartilage. The disease is usually caused by the loss of blood supply [108].

[8]From the latin *fragilitas ossium* this disease is a congenital disorder of the connective tissue formation that affects bone, teeth and soft tissues, causing brittle bones that easily fracture. There are four types of this disease with varying severity – the worst type being lethal at birth. Most types are inherited as autosomal dominant characteristics [108].

[9]A metabolic disorder characterised clinically by defective bone mineralisation, caused by mutation in the tissue-nonspecific isoenzyme of alkaline phosphatase [109].

[10]A chronic disease of the nervous system that affects young and middle-aged adults. In this disease, the myelin sheaths surrounding nerves in the brain and spinal cord are damaged, affecting the function of the nerves involved. The underlying cause of the nerve damage is still unknown, but an autoimmune process may be involved [108].

[11]This disease is also known as regional enteritis and is named after Burril B. Crohn (1884-1983) a US gastroenterologist. Crohn's disease is a chronic inflammatory disease that usually affects the ileum and colon causing inflamed lesions that are rich in inflammatory cytokines and tumour necrosis factor α [113].

[12]Also known by the acronym SLE, this disease is a chronic inflammatory autoimmune disease of the connective tissue, affecting skin and various internal organs. Typically, the disease forms the red butterfly-shaped rash on the face and causes arthritis and progressive damage to the kidneys. Heart, lungs and brain can also be affected by progressive attacks of inflammation followed by the formation of scar tissue (fibrosis) [108].

recipient's body. Application of MSCs resulted in an immunosuppressive effect leading to significant symptomatic improvement and prevention of GVHD [114, 115].

MSCs have also being investigated for the therapy of ischemic injuries such as myocardial infarction and cerebral ischemia. This application was driven by the fact that murine bone marrow-derived MSCs were identified as the progenitors of cells with phenotypic properties of cardiomyocites and neurons in *in vitro* studies [116, 117]. The potential for MSCs to repair damaged cardiovascular tissue and act as a therapeutic approach for cardiomyoplasty and infarction, together with their role in angiogenesis, has been reviewed by Pittenger and Martin (2004). In this review it was highlighted that one of the greatest attributes of MSCs is their potential to supply growth factors and cytokines for repairing tissue, along with the fact that they appear not to be rejected by the immune system [118].

One critical issue that impacts the use of MSCs in research and in clinical medicine is the paucity of information regarding how their different functional attributes are specified at the population level, a problem that is exacerbated by donor-to-donor and intra-population heterogeneity [119]. Variability, not only amongst single-cell derived colonies, but also within the same colony, have been observed with MSCs presenting wtih different potentials for differentiation and proliferation [120,121]. James *et al.* (2015) observed that bone-marrow derived immortalised MSC clones expressing typical cell-surface antigens exhibited contrasting trilineage differentiation capacities [84]. Since MSCs lack unique cell surface markers, clinical studies continue to use mixed stromal cells with unclear or unpredictable differentiation capacities [86].

Although there are established criteria for MSC selection amongst heterogeneous mixtures of primary cells, the idea that they will form a homogeneous population may be incorrect. Studies have reported that extensively cultured MSCs *in vitro* showed decreased capability of differentiation, increased cell size, changes in the global gene expression profile

and increased signs of cell age [93, 122–124].

The phenotypic and functional heterogeneity observed in primary cell cultures or clonal populations is reflected in MSC's cellular morphology, differential marker expression and variable differentiation potential; it could be a result of alterations caused by *ex vivo* culturing and *in vivo* heterogeneity and variable phenotypes that reflect the natural repertoire of these cells [93]. Even within a tissue such as bone marrow, there may be microenvironments wherein closely related or identical cells may express different surface molecules [118].

This biochemical heterogeneity may have some biological benefits related to the function of the MSCs, which are involved in tissue repair and inflammation and could play an important role in allowing the selection of the appropriate cell type for each extreme situation, reflecting the complexity of the stromal systems in the bone marrow [93]. However, if the principal aim is to use MSCs for regenerative medicine, then this application would rely on the expansion of progenitor cell populations whilst still preserving their capacity for multipotent differentiation. In this respect there is an importance placed on characterising the MSC subtypes since this characterisation impacts their therapeutic potency.

Studies cataloging the transcriptome[13] of human MSCs showed that they express transcripts of a diverse collection of proteins involved in the regulation of angiogenesis, hematopoiesis, cell motility, neural activities and immunity and defence [94, 119], reinforcing the concept that a single cell is unlikely to have all these properties and that they may be assigned to sub-populations.

MSCs need to be functionally defined and small differences in surface markers may not

[13]The full complement of RNA transcripts of the genes of a cell or organism. It provides a 'snapshot' of the expression pattern of the cell's genes [101].

be sufficient to distinguish types or subpopulations of cells [118]. Therefore, the question whether definitive markers for MSCs can be identified and more importantly, whether they can predict the MSC function, is still open. Therefore, it is necessary to identify markers that are related to their differentiation potency, and this may reside in the use of other methodologies, such as Raman spectroscopy.

1.5 Raman spectroscopy applied to MSCs

With its ability to obtain a biomolecular fingerprint of the analysed cell, RS is an emerging technique to characterise a diversity of cells, including the characterisation of stem cells and their differentiation. When applied to MSCs, RS is extensively used to assess *in vitro* osteogenic differentiation due to its capacity to straightforwardly detect the characteristic peak of hydroxyapatite (HA) [$Ca_5(PO_4)_3(OH)$] in the cell spectrum [68, 125].

Bone tissue consists of a carbonated hydroxyapatite-like mineral supported on a hydrated, collagen-rich protein matrix, thus RS uses the information from the phosphate and carbonate vibrations to characterise the mineral formed during osteogenic differentiation [126]. The main hydroxyapatite vibrations observed in the Raman spectrum are the PO_4^{3-} ν_1 symmetric stretch ranging from 950 cm^{-1} to 964 cm^{-1}, the PO_4^{3-} ν_3 asymmetric stretch between 1030 cm^{-1} and 1044 cm^{-1}, and the CO_3^{2-} ν_1 symmetric stretch ranging from 1065 cm^{-1} to 1070 cm^{-1} [126–129]. RS also captures the vibrations of the bone collagen matrix, with the major peaks being observed for proline (850–855 cm^{-1}), hydroxyproline (872–876 cm^{-1}), amide III (1243-1270 cm^{-1}), CH_2 scissoring vibration (1445-1450 cm^{-1}) and amide I (1595-1720 cm^{-1}) in the fingerprint region, and the CH stretch (2800–3100 cm^{-1}) in the high wavenumber region [126–129].

One of the first studies applying RS to analyse osteogenesis was performed by Pelled *et al.* (2007), who compared engineered bone tissue from genetically modified MSCs, derived

from bone tissue, and grown *in vivo* in the tight muscle of immunocompetent mice, with the femoral bone adjacent to the transplantation site [130]. Although they had observed very similar Raman spectra between implanted and native bones, their study did not focus on following the osteogenesis process per se.

The first RS study to definitively monitor the differentiation of human bone marrow-derived MSCs into osteoblasts was performed by Chiang *et al.* (2008) [131]. They followed the development of HA over the period of 7–21 days after the start of differentiation, when the MSC differentiation status was confirmed by alizarin red S staining for calcium. Their analyses were carried out with a 632.8 nm laser, focused through a water immersion lens, on cells placed onto a quartz substrate and kept in a phosphate buffered saline solution (PBS). The first sign of mineralised bone nodules in the spectra was noticed between the 14th and 21st day of their analyses, with the most dramatic changes occurring in the region around 960 cm^{-1}, which is dominated by the symmetric stretching of phosphate groups, including HA. However, other vibrations of the HA molecule, such as the phosphate ν_3 asymmetric stretch and the CO_3^{2-} ν_1 symmetric stretch were only observed at the 21st day. With this work, Chiang and collaborators showed that the RS analysis of HA was a clear and specific marker for monitoring the mineral deposition during osteogenic differentiation.

By means of RS, the quality of the mineral deposited by the cells can be evaluated and further compared to native bone via compositional measurements of the mineral-to-matrix ratio, carbonate-to-phosphate ratio, and the mineral crystallinity [128]. The mineral-to-matrix ratio (obtained from the division of the primary phosphate band, the PO_4^{3-} ν_1 symmetric stretch at ∼960 cm^{-1}, by a matrix band, such as the amide I band at ∼1654 cm^{-1} or the CH stretch at ∼2940 cm^{-1}) is indicative of the amount of mineralisation, while the carbonate-to-phosphate ratio (CO_3^{2-} ν_1 at ∼1070 cm^{-1} against PO_4^{3-} ν_1 at ∼960 cm^{-1}) provides insight into the biomolecular composition of bone, varying also with bone

architecture, carbonate substitution of the apatite, age and mineral crystallinity [128,132]. Finally, the full width at half maximum (FWHM) measurement of the primary phosphate band (PO_4^{3-} ν_1) indicates the strength of bone and its dependence on the amount of mineralisation, the degree of mineral crystallinity and changes in distribution of different crystal sizes; a thinner FWHM is correlated to higher crystallinity [128].

Another quality measurement is the position of the hydroxyapatite PO_4^{3-} ν_1 peak, which indicates the type of calcium phosphate species present, categorising them into three different band range frequencies. Disordered amorphous apatite from amorphous calcium phosphate or from A-type carbonate-substitution (where carbonate is substituted for hydroxide) has a band between 945 and 950 cm^{-1}. B-type carbonate-substituted apatite has a band from 955 to 959 cm^{-1}, while the band for crystalline non-substituted hydroxyapatite is between 962 and 964 cm^{-1} [129, 133].

Gentleman *et al.* (2009) used RS, along with other microscopy techniques such as transmission electron microscopy (TEM), scanning electron microscopy (SEM), together with nanoindentation mechanical testing and staining assays, to compare the *in vitro* mineralised nodules produced by cells (mouse embryonic stem cells – ESCs, neonatal calvarial osteoblasts and adult bone marrow-derived MSCs) to native bone [127]. However, the analyses were focused only at 28 days after osteogenic induction. From the Raman spectra of live cells, they observed that similarly to native bone, the nodules from the osteoblasts and MSCs contained three distinct mineral environments: B-type carbonate-substituted apatite, crystalline non-substituted hydroxyapatite and amorphous phosphate species, all identified by the position of the primary phosphate band that ranged from 945 to 964 cm^{-1}. All these mineral species together contribute to the bone's function. In contrast, the ESC nodules were dominated by a single mineral factor, similar to synthetic HA and failed to show the complex mineral interaction of the mineral species observed in the

nodules from the osteoblasts and MSCs. Mineral-to-matrix and carbonate-to-phosphate ratios contributed to these observations. However, they were obtained from band area ratios, which are more susceptible to background correction errors than peak intensity ratios [126].

Another work, although conducted on methanol fixed and air-dried human bone marrow-derived MSCs, also showed that RS was able to reveal the presence of the previous mentioned mineral species throughout the differentiation process, suggesting that the early stages of osteogenic differentiation of MSCs resembled endochondral ossification [126].

In general, the first sign of mineral deposition is observed after 14 days of the osteogenic induction in the form of crystalline, non-substituted apatite [126, 131, 134]. However, aggregated structures have been observed on day 12 [134], and the HA peak was detected rising above the spectrum background as early as day three on mouse MSC lines [34], day seven for human adipose-derived stem cells [68] and day nine for human bone marrow-derived MSCs [35].

RS mapping was also demonstrated to more accurately show the mineralised nodule distribution when compared to conventional staining assays, such as Alizarin red S [34], and more recently, high-speed Raman imaging was used to follow the mineralisation process of live murine osteoblasts for 24h at subcellular resolution [135]. Conventional Raman imaging has also been used to examine how mechanical deformation affects the bone ultrastructure [136], providing a correlation between biomolecular information and bone failure. These works display the efficacy of a non-destructive technique such as RS in the monitoring and quality evaluation of the mineral products of the MSC osteogenic derivatives.

RS has also been used to characterise adipogenic differentiation, however not to the same extent of osteogenic differentiation, with few works being published on this subject

to date, and mainly on adipose-derived stem cells (ADSC). Studies performed by Downes *et al.* (2011) and Mitchell *et al.* (2015) revealed significant differences in the spectra of adipo-induced ADSCs, mainly characterised by the presence of very sharp peaks at ∼1440 cm^{-1} and ∼1654 cm^{-1} and an increased intensity for the peak at ∼1305 cm^{-1}. All of these signatures related to vibrations of the CH_2 and C=C bonds in the lipids present in the lipid droplets formed during differentiation [137, 138]. Another investigation, also performed on ADSCs, but using CARS, revealed that the CH_2 peak at ∼2845 cm^{-1} could also be used as an indicator of adipogenesis [139].

The adipogenic differentiation of MSCs derived from the human bone marrow (or SSCs as denoted by the authors) has been investigated with RS and CARS by Smus *et al.* (2015) [140]. In their work, CARS was used to image adipogenesis from day 1 to day 14. The use of RS was restricted to the identification of the most suitable Raman band for imaging lipids with CARS, which also resulted in the CH_2 peak at 2845 cm^{-1} being chosen due to its strong presence in the spectra. Interestingly, other lipid peaks, such as those from the vibrations of the CH_2 (∼1440 cm^{-1}) and C=C (∼1654 cm^{-1}) bonds were only observed in the Raman spectra of the adipo-induced cells and not visualised in the spectra of cells kept in basal media.

Although RS has had very important applications in assessing mineral nodules or in indicating adipogenesis, it can be used to investigate other molecular changes throughout the differentiation process. There is however, a lack of studies on MSCs focusing on this type of discrimination, with the majority of the literature investigating these types of changes on embryonic stem cells. For example, Notingher *et al.* (2004) noticed that the area of peaks assigned to RNA (813 cm^{-1}) and DNA (786 cm^{-1}) decreased 75% and 50% respectively, over 16 days of differentiation for murine ESCs [55]. Another work that compared differentiated and non-differentiated human ESCs identified that the peak

intensity ratios of protein-related bands (757 cm^{-1}) to nucleic acid bands (784 cm^{-1}) were effective in discriminating their spectra [33]. A decrease in the intensity of DNA and RNA bands (785, 811, 1090 and 1320 cm^{-1}) was also noted in a live-cell RS study of human ESC during cardiomyocyte differentiation, providing discrimination between differentiated and non-differentiated cells [36]. Glycogen was also found to contribute to the discrimination between cardiomyocytes and non-cardiomyocytes [37, 69]. All of these differences between differentiated and non-differentiated cells establish RS to have a high molecular specificity that can be used towards a label-free method for cell characterisation and discrimination.

In order to ensure safe application of the MSCs in regenerative medicine, it is important to have these cells fully characterised and discriminated from other cell types such as fibroblasts or differentiation incompetent subpopulations. RS is a promising non-destructive and label-free technique that could be routinely incorporated into the study of these cells. Multivariate analyses of Raman spectra have provided discrimination between human MSCs and ESCs due to differences in DNA bands (780 and 1095 cm^{-1}) [141]. Pudlas *et al.* (2011) using multivariate analyses distinguished human bone marrow-derived MSCs from fibroblasts due to major spectral baseline differences attributed to specific endogenous autofluorescence patterns [31]. Such fluorescence resulted from higher mitochondrial activity and was correlated to the higher proliferative capacity of the MSCs. Fluorescence studies have also discriminated between normal and abnormal human bone marrow-derived MSCs and ESCs due to apparent differences in phenylalanine (1005 cm^{-1}) and protein (1158 and 1523 cm^{-1}) peaks, as well as a convolved band that includes vibrations of proteins, DNA/RNA and lipids (1450 cm^{-1}) [30]. However, there are no studies focusing on characterising the heterogeneity of MSCs, or identifying undifferentiated cells markers linked to their potential to undergo differentiation.

A better use of the MSCs in regenerative medicine requires the characterisation of their

heterogeneity, highlighting the need to establish the means to identify sub-populations that have contrasting differentiation capacities, as standard methods, such as the expression of surface proteins or global gene expression, are unable to provide this level of discrimination. The use of non-destructive techniques is also envisaged and it is here that Raman spectroscopy can be applied to its full potential, from cell-type characterisation, through to the assessment of the differentiation products, and up to label-free cell imaging. However, to be used for this intent, RS needs to be rigorously and carefully performed, from the execution of experiments to the data analysis.

1.6 Summary of research objectives

To address these limitations, this thesis proposes to use RS to characterise cells while investigating cell population heterogeneity and identifying biomarkers that perform cell discrimination MSCs. The research has been focused on the case of human bone marrow-derived mesenchymal stromal cells, known for their population heterogeneity and lack of unique cell surface protein expression markers for their identification.

Raman analyses of cells, in general, have lacked a means of determining and quantifying if the amount of collected spectra represents the investigated cell population. Therefore, one of the aims of this work is to develop a method for measuring the population heterogeneity thus, assessing the number of spectra needed to statistically represent such a cell population. This was done by means of obtaining the convergence of the standard error of the mean and standard deviation with an increasing number of collected spectra.

The use of RS for cell discrimination is limited becuase it is mainly used to compare different cell types, *e.g.*, differentiated from non-differentiated cells or normal from abnormal cells, without a deeper investigation on the potential of RS to discriminate sub-populations of cells. Thus, RS was applied to clonal lines of human bone marrow-derived

MSCs that express the same panel of surface proteins, but have contrasting differentiation capacities. The aim is to characterise these cells whilst obtaining spectral markers for their discrimination. The analysis of air-dried and live cells is assessed, followed by the RS investigations of osteogenesis and adipogenesis.

The thesis has a method chapter, three results chapters and final conclusions. Chapter two describes the cell culturing methods together with their preparation for the Raman experiments, which were performed on air-dried and live cells. This chapter also gives details of the micro-Raman system used, the acquisition parameters and experimental set-ups for the interrogation of cells. The data analyses by means of the convergence tests, multivariate and univariate analyses are also detailed in this chapter.

The aim of Chapter three is to verify the capability of RS to discriminate the homologue air-dried MSC lines whilst providing spectral markers linked to their differentiation competence. The identification of markers was done by means of the univariate ratiometric analysis, *i.e.* the calculation of peak intensity ratios (PIR), with minimal interference of pre-processing methods, since they can alter peak shape and intensity. These results were then compared to the multivariate analyses, PCA-LDA. This chapter also presents the convergence tests for the assessment of cell population heterogeneity and the examination of cell morphology via spatially-resolved Raman maps.

On the other hand, Chapter four verifies the use of RS for the characterisation and discrimination of the MSCs on live cell experiments. The viability of the interrogated cell is assessed after a Raman measurement, with the spectra and biomarkers being compared to those obtained from the air-dried cells. The analysis via PIRs was also checked against PCA-LDA.

Chapter five focuses on characterising the osteogenic and adipogenic differentiation. For the osteo-induced MSCs, the aim is to reveal if RS can identify markers other than

those related to the presence of hydroxyapatite. Parallel to it, the mineral content produced by the osteo-induced MSCs is assessed and compared to human and lamb bones. Finally, as per the adipo-induced cells, the study is focused on characterising this differentiation on MSCs, identifying the lipids stored in lipid droplets and revealing functional spectral markers related to adipogenesis.

Chapter 2

Methods

This chapter explains details of the micro-Raman system used, highlighting its main components, the parameters that were chosen for data acquisition and their implication in the obtained spectra. The experimental set-ups for air-dried and live-cell interrogation are described, as well as the culture of MSCs, their osteogenic and adipogenic inductions, along with their preparation for the Raman experiments. Staining assays performed on these cells are also described in the chapter. The data analysis section describes the uncertainties convergence tests that were used to evaluate if the averaged spectra represented the studied cell population. This section also explains how the biomolecular markers for cell characterisation were obtained from the ratiometric analyses and gives details on the multivariate analyses performed, such as principal component and linear discriminant analysis. Finally, the chapter also describes the methods for obtaining and generating the spatially-resolved Raman images.

2.1 Micro-Raman spectroscopy

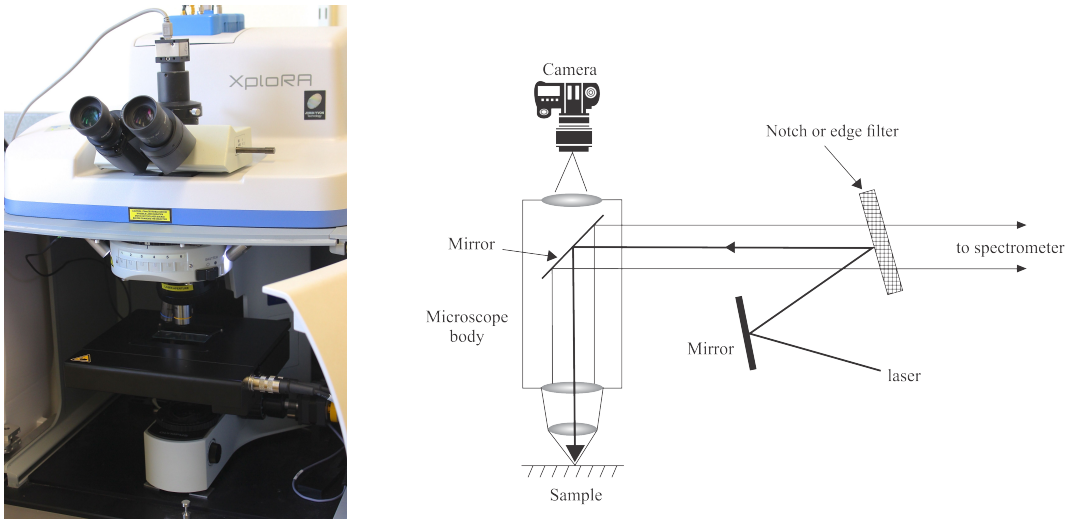
It was only in the 1990s that the Raman spectrometer was combined to an optical microscope, allowing laser light to be focused to a small spot of a few microns, and thus

establishing the micro-Raman spectroscopy [4]. The use of a microscope lens increased the collection efficiency of the system due to its high numerical aperture (NA); however, if used carelessly this assemble could cause heat damage to the sample as the light is focused into a microscopic spot [4].

All Raman data acquisitions were performed using a HORIBA XploRA micro-Raman system (Figure 2.1a). This equipment uses a 180° backscattering configuration, meaning that the laser is delivered through the microscope lens and the scattered light is collected back through it [6]. Figure 2.1b shows the typical 180° backscattered geometry, configuration on an upright microscope. In this geometry, it is necessary to use some device to separate the Raman-shifted backscattered light from the other wavelengths. This separation is done using an edge- or notch-filter that is designed to absorb all of the scattered light that has the same frequency as the incident beam [4,6]. The HORIBA system used in this research employs an edge-filter for separating the Raman from the Rayleigh scattering.

Micro-Raman systems often have adjustable components in the incident light pathway, such as the neutral density filter and in the scattered light pathway between the objective and the CCD detector, such as a pinhole (or hole), a slit and a diffraction grating. Figure 2.2 shows a schematic of the components of the HORIBA XploRA micro-Raman system that are present in the incident and scattered light pathways. All of these components are detailed in the next sections of this chapter and were configured to avoid the fluorescence caused by the choice of laser, whilst still performing a non-destructive sampling and providing good spectral resolution.

Raman spectroscopy has fluorescence as its natural opponent because the Raman scattering signal is typically very weak due to the non-resonant interaction of the laser photons with the sample molecules. Fluorescence is a resonant interaction of the light with the electronic states of the sample, which can be six orders of magnitude higher in intensity



(a) **(b)**

Figure 2.1: (a) HORIBA XploRA micro-Raman equipment and (b) system set up for an upright microscope using the 180° backscattering geometry as per [6]. The photographic camera on the schematics represents the eyepiece and optical camera present in the HORIBA XploRA equipment.

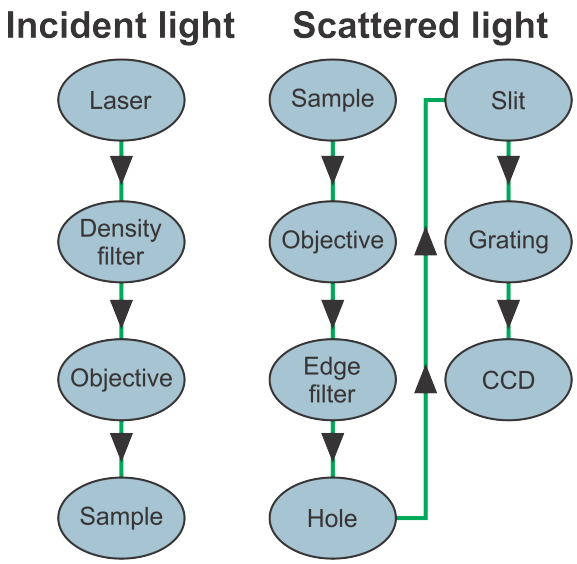


Figure 2.2: HORIBA XploRA micro-Raman system components present in the incident and scattered light pathways.

than the Raman interaction, therefore hindering the Raman signal [4, 6]. Many organic samples show fluorescence when excited at the UV, blue or green regions of the spectrum. Fluorescence can be avoided by selecting a wavelength that does not produce fluorescence or by reducing the probed volume [4].

The laser used for the experiments in this research was a green Nd:YAG (neodyniun yttrium aluminium garnet) solid state diode-pumped laser (532 nm, ∼7 mW). The Nd:YAG laser is a four-level laser that can be pumped continuously to generate continuous wave output at 1064 nm [7]. Nd:YAG lasers can also operate at 532 nm by using an intra-cavity crystal to double a significant fraction of the light exiting the cavity; by this reason they are also called frequency-doubled Nd:YAG lasers [3,7]. The efficiency of the diode-pumping allows the generation of several watts of output power for the 532 nm laser without the need for external cooling water or special electrical power, producing compact laser heads [7]

The Raman scattering intensity is proportional to ν^4 (or to $1/\lambda^4$), where ν is the frequency and λ the wavelength of the excitation laser source [7]. Thus, even though a green laser might cause some fluorescence when biological samples are probed, the 532 nm laser was chosen because it produces a more intense Raman scattering.

2.1.1 Microscope objective and neutral density filter

A good microscope objective will allow the highest possible collection of the scattered light together with the best spatial resolution (Δx), which is the smallest resolvable distance between two points [4]. This means that the ideal diffraction limited objective will image an Airy disk from an infinitely small object point (Figure 2.3a) [142]. The diameter of the first Airy ring is also the laser spot size (d_l) at the sample and it is generated by destructively interfering, diffracted wavefronts, and given by Equation 2.1 [142].

The resolvable distance between two points, or spatial resolution (Δx), is determined

by the Rayleigh criterion, which establishes that the centre of the Airy disk of the first object occurs at the first minimum of the Airy disk of the second object [8, 143], as exemplified by Figure 2.3b. The spatial resolution (Δx) is given by Equation 2.2 [142].

$$d_l = \frac{1.22\lambda}{NA}, \tag{2.1}$$

and

$$\Delta x = \frac{0.61\lambda}{NA}, \tag{2.2}$$

where λ is the wavelength of the excitation source and NA is the numerical aperture of the microscope objective [142, 144].

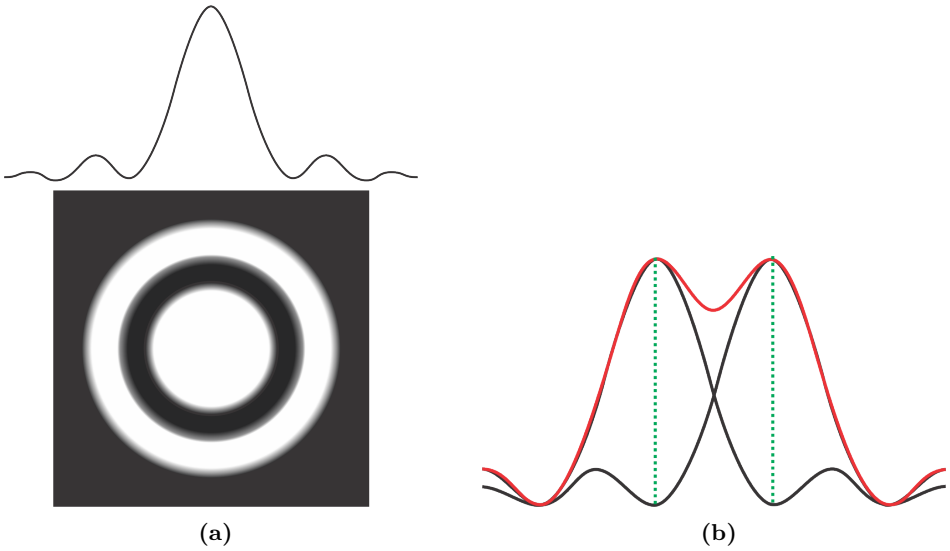


Figure 2.3: Schematics for the Airy disk and spatial resolution. (a) Airy disks created by an infinitely small object point and its intensity distribution. Image adapted from [142]. (b) The Rayleigh criterion for the spatial resolution establishes that the centre of the Airy disk of the first object occurs at the minimum of the Airy disk of the second object. This image was adapted from [8].

Equations 2.1 and 2.2 represent the theoretical limited values for both the spatial resolution and laser spot size. Thus, an objective with a high numerical aperture will produce a smaller laser spot size, reducing the sampled volume, which also helps to avoid fluorescence. The microscope objectives that were used in this research are listed in Table 2.1, along with their NAs and working distances.

The choice of microscope objective depends on the type of experiment being performed. For this research two types of experiments were performed: analyses of air-dried cells and analyses of live-cells; both experimental procedures are described later in this chapter, in Section 2.3. For the air-dried cell analyses, the 100x (0.9 NA) lens was chosen. The live-cell experiments were performed using the 63x (1.0 NA) dipping lens, which due to its longer working distance, caused minimal disturbance to the cell layer. For both lenses, the laser spot size generated by the 532 nm laser is $\sim$0.7 μm.

With such a small spot size the laser power has to be chosen well-below the point that leads to thermal decomposition of the sample. A useful measurement for the laser power is the laser power density at the sample, which is measured in W/cm^2. Although the laser power density allows different experimental set ups to be compared regarding the the laser power used, it is common that only the average laser power is disclosed. An acceptable laser power density depends on the absorption properties of the sample and, on its thermal and photochemical damage thresholds [7]. The laser power density at the sample can range from values below 10 W/cm^2 on macro samples to values higher than

Table 2.1: Microscope objectives for sample inspection used in this research. [1]Water dipping lens.

Objective	Numerical Aperture (NA)	Working distance (mm)
50x	0.75	0.38
63x dipping lens[1]	1.0	2.1
100x	0.90	0.21

10^6 W/cm^2 on microscopic samples [7].

Notingher *et al.* noticed morphological changes in cells irradiated for periods of 5 min with 488 nm and 514 nm lasers at 5 mW laser power, when using a 0.9 NA lens [41]. Thus, the laser power density at the sample, for their experiments, would be $\sim 1.3\times10^{10}$ W/cm^2. For transparent samples, the typical acceptable laser power for a green excitation is on the order of 10 mW for a 300 nm laser spot size [4], which results in a laser power density of 1.4×10^{11} W/cm^2. Therefore, the acquisition time, laser wavelength, laser power and, sampled area and volume have to be carefully chosen in biological experiments.

The use of a neutral density filter positioned between the laser and the objective lens reduce the laser power intensity at the sample. The micro-Raman spectrometer used in this research has the following filters available: 0.1%, 1%, 10%, 25%, 50% and 100%. A 100% neutral density filter allows full laser power at the sample. Thus, to reduce the laser power and avoid sample damage, a 50% neutral density filter was chosen, reducing the average laser power to $\sim$3.5 mW. Considering that the laser spot size is ~ 0.7 μm, the laser power density at the sample is $\sim 9.2\times10^9$ W/cm^2; therefore, smaller than the laser power density obtained by Notingher *et al* and that caused morphological alterations in the cells. The acquisition time used was 45 s over two repetitions, meaning that the cells were exposed to the laser for a much shorter period, thus avoiding laser induced damage. Spectra was acquired in real-time and the samples were visually inspected between every spectrum collection.

2.1.2 Pinhole

As previously explained, the two most straightforward ways to reduce fluorescence are to (i) select a laser wavelength in which the sample does not fluoresce, or (ii) reduce the detection volume. However, it is not always possible to use a different wavelength and

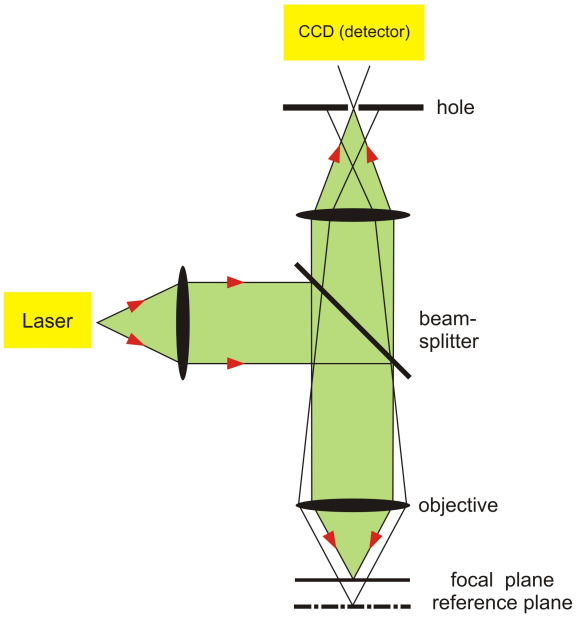


Figure 2.4: Schematic of a confocal pinhole as per [4], which shows that only the light from the focal plane is allowed to pass through the hole, thus allowing confocal Raman spectroscopy to be performed.

there are samples that will fluoresce regardless of the wavelength chosen.

A confocal detection setup limits the light collection to photons emitted only by the focal plane, consequently it reduces the fluorescence background as well as Raman photons [4]. This is done by introducing a pinhole (also referred to as hole) before the spectrometer and detector. The system used for this research has the option to select different pinhole sizes, thus by selecting smaller holes one can perform confocal Raman micro-spectroscopy.

The size of the hole is chosen so that only the central part of the focused laser can pass through it and reach the detector. Therefore, avoiding any other photons emitted by other focal planes (Figure 2.4). The hole size also affects the intensity of the Raman signal being collected as there is a physical barrier blocking the scattered light from reaching the detector, therefore, its size has to be appropriately chosen depending on the sample. In general, bulk sample analysis of non-fluorescent samples use bigger pinhole sizes than the analysis of transparent thin samples. If the sample is a thin layer on top of a supporting

substrate, the confocal setup allows the separation of the signal from the sample from the background signal [4].

The limit size for the pinhole is related to the size of the Airy disk (Figure 3.6) as projected onto the hole; thus, the smallest practical hole size is equal to the radius of the Airy disk [145, 146]. Even in diffraction-limited instruments, smaller holes do not improve the confocal setting and only reduce the amount of detected light [146].

The HORIBA XploRA system used has an adjustable pinhole located at the image plan of the microscope with setting choices of 100, 300 and 500 μm. Thus, in order to avoid photons emitted by other focal planes and perform confocal Raman spectroscopy, the 100 μm hole was used for all Raman acquisitions.

2.1.3 Slit and spectrometer

The spectrometer disperses light by means of a diffraction grating allowing the separation of wavelengths, hence Raman shifts [7]. However, a slit at the entrance of the spectrometer determines the angle and amount of light that will reach the optical bench (diffraction grating and CCD) and is an important factor when determining the spectral resolution [147]. A narrow slit results in loss of output signal, but conversely produces a higher divergence angle, completely illuminating the diffraction grating. Nonetheless, a broader slit may result in a small divergence angle partially illuminating the diffraction grating.

The HORIBA XploRA system has a single grating Czerny-Turner spectrograph (Figure 2.5), where the light entering the spectrometer after passing through the slit, is collimated to the diffraction grating by the collimating mirror and then further focused onto the CCD detector by a focusing mirror.

The spectral resolution depends on the groove density of the diffraction grating and is also a complex function of the slit-width, CCD pixel-width and spectrograph dispersion [7].

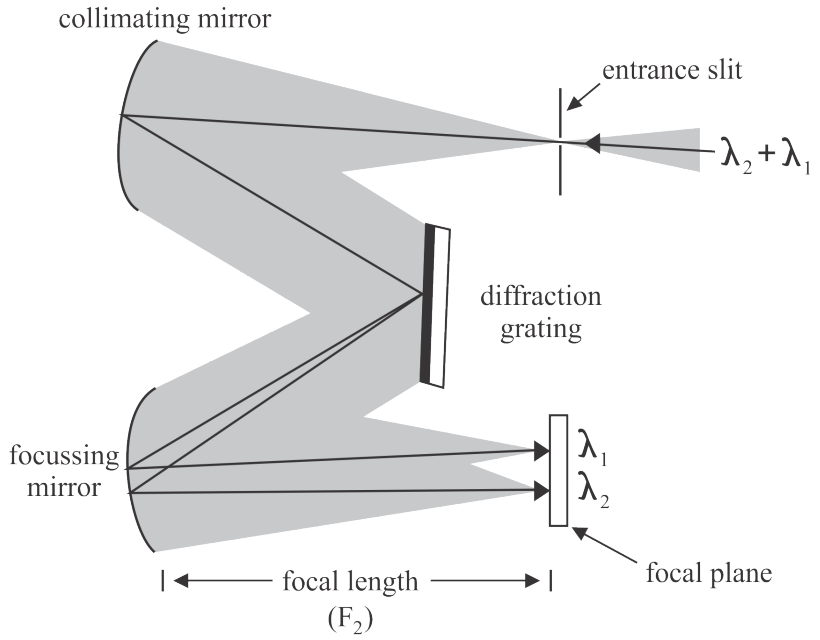


Figure 2.5: Schematics of a Czerny-Turner spectrograph. The scattered light enters through the slit, is collimated into the diffraction grating, and focused onto the detector. F_2 is the focal length, as per [7].

If the slit-width is bigger than the pixel-width, then the spectral resolution is influenced by the slit rather than by the CCD pixel. However, if the the slit is decreased to a value smaller than the pixel-width, then the slit-width will not improve resolution and will only decrease the signal [7].

The micro-Raman system used in this research has the following options for the slit size: 50, 100 and 200 μm. The available diffraction gratings are 600, 1200, 1800 and 2400 grooves/mm. In this system, the spectral resolution is 1 pixel, which corresponds to 1 cm^{-1} for a 2400 grooves/mm grating, or to 4 cm^{-1} when a 1200 grooves/mm grating is used. Therefore, to obtain the best spectral resolution, whilst maintaining good signal, the slit size chosen was 200 μm combined with the 2400 grooves/mm diffraction grating.

Another point of consideration is that spectral coverage and total acquisition time are linked. If a scanning spectrometer is used, then a high resolution over a wide Raman shift

requires that a large number of wavelengths increments be monitored, greatly increasing the acquisition time. The acquisition time is improved by using a multichannel (or multi-window) spectrometer, which will monitor many wavelengths at once, but still within a finite number of channels (or windows) [7]. The HORIBA system has both scanning and multi-window options, therefore the latter was chosen to reduce the total acquisition time for the experiments.

2.2 Cell culture

2.2.1 Summary of studied cells

All of the work performed on human samples was approved by the University of York, Department of Biology ethics committee and obtained by Prof. Paul G. Genever.

The four MSC lines (termed Y101, Y201, Y102 and Y202) employed in this research were developed by the research laboratory of Prof Paul G. Genever in the Biology Department at the University of York. The generation and phenotyping of the immortalised clonal cell lines used in this research are described by James *et al.* (2015) [84]. Thus, a brief summary of their work is described next, as per [84].

James *et al.* isolated a heterogeneous population of MSCs from the bone marrow of one donor and immortalised it by the overexpression of human telomerase reverse transcriptase (hTERT) followed by clonal isolation to generate a panel of cell lines. Four clonal cell lines (Y101, Y201, Y102 and Y202) were selected due to their strong clonal and stable growth characteristics, which were maintained over 400 days. The expression of surface antigens was determined by flow cytometry, with positive expression of CD29, CD44, CD73, CD90, CD105 and CD166 antigens and negative expression for CD34 and CD45 antigens, thus, matching the profile of their parental cells and typical of human primary

MSCs [96]. All clones were also negative for CD11b, CD14, CD169, CD1a, CD4, CD11c and CD83 antigens. James *et al.* also checked the clonal lines for tumourigenicity and no evidence was observed in *in vitro* colony transformation assays for non-adherent growth, or even following subcutaneous injection into immunocompromised mice. Characterisation of these four cell lines towards their tri-lineage differentiation – osteogenic, adipogenic and chondrogenic, followed up to 21 days after induction was also checked. Cell lines Y101 and Y201 showed tri-lineage differentiation capacity, whereas Y102 and Y202 had limited differentiation ability, exhibiting a non-haematopoietic stromal phenotype together with atypical morphology (flat and spread in comparison to the elongated fibroblastoid shape of the Y101 and Y201 cells). From the analysis of an array data for differential expression of genes encoding cell-surface proteins, it was also noticed that CD317 was highly expressed in the Y102 and Y202 while absent from the Y101 and Y201. The CD317+ fraction of cells sorted from primary MSCs had characteristics of colony-forming population, and appeared to have enhanced immunostimulatory capacity, thus, being closely related to the Y102 and Y202 cells.

Primary human MSCs studied with RS were isolated by Prof Paul G. Genever's research group from femoral heads during hip replacement surgery or as explant cultures from human tibial plateaus after routine knee replacement, and obtained with informed patient consent. For more details about their isolation, please refer to James *et al.* (2015) [84]. These primary human MSC populations are termed K72, K106, K148 and K149. The CD317+ primary fraction was analysed with RS, being previously sorted by Dr Charlotte Knight from Paul G. Genever's group using fluorescent activated cell sorting (FACS) from a primary population as described by James *et al.* (2015) [84]. Finally, the human dermal fibroblasts (HDF), used as a non-MSC stromal cell control for the RS investigations, were also maintained at the laboratory of Prof Paul G. Genever.

2.2.2 Culturing methods

All cells were kept frozen in liquid nitrogen, and initially thawed in a water bath. The cells were then grown in a T-25 flask (Corning, Sigma-Aldrich) with growth medium: Dulbecco's Moddified Eagle's Medium (DMEM) containing 4.5 g/L D-Glucose, L-Glutamine and Pyruvate (Gibco, Thermo Fisher), in which were added 1% of penicillin/streptomycin (100 U/mL - penicillin and 100 μg/mL - streptomycin; Gibco, Thermo Fisher) to prevent bacterial contamination. The growth medium was supplemented with 10% of foetal bovine serum (FBS) for the MSC lines and HDFs, whereas 15% FBS was used for the primary human MSCs. Once confluent, the cells were transferred to a T-75 flask (Corning, Sigma-Aldrich) and maintained in culture.

To promote cell growth, the culture flasks were kept in an incubator at 37°C and 5% CO_2 concentration. The growth medium was changed twice a week and the cells were passaged at a 4:1 ratio when 70%-80% of confluence was reached. This was first done by removing the growth medium and rinsing the flask with 1x phosphate buffered saline solution (PBS) (Gibco, Thermo Fisher) to remove dead cells, debris and any remaining of growth medium. A 0.05% trypsin-EDTA (Gibco, Thermo Fisher) was used to detach the cells from the plastic, which were then centrifuged for 5 min at 1200 rpm with 2 ml of growth medium. The trypsin supernatant was removed and the cells were re-suspended in fresh growth medium, split, re-seeded in the flask and re-incubated. Primary human MSCs were expanded and used between passages p1 and p7.

2.2.3 Preparation of cells for Raman spectroscopy

RS was performed on both air-dried and live cells and the experimental set-ups for both of them are described below. For both types of Raman measurements, the cells were detached from the flasks with trypsin following the same procedure described previously in Section

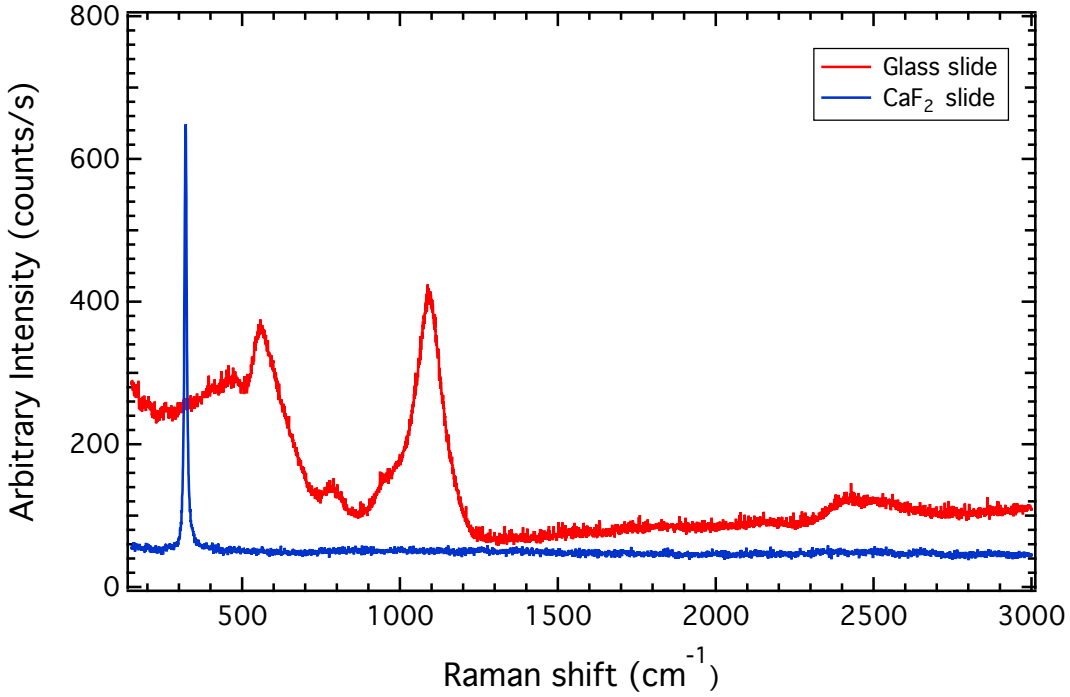


Figure 2.6: Glass versus CaF_2 spectra showing the different backgrounds created by both materials. For both spectra the acquisition time was 45 s and the laser power at the samples was $\sim 9.2 \times 10^9$ W/cm^2.

2.2. These cells were then seeded onto calcium fluoride slides (CaF_2) (Crystran Limited) placed into 10 cm Petri dishes. CaF_2 slides are preferred to glass slides because they do not show any background spectra in the spectral region of sampled cells. A comparison between the glass and CaF_2 spectra is shown in Figure 2.6. Glass creates a large spectral background that can jeopardise the weak Raman bands from the cells. In comparison, the only peak observed from CaF_2 is at 320 cm^{-1}, therefore outside of the acquisition range of any experiment performed in this study.

MSCs tend to form colonies, therefore the cell suspension was repeatedly pipetted prior seeding to allow for an even distribution of cells on top of the slide. The number of cells was counted using a hemocytometer with 2.5×10^4 cells/cm^2 seeded onto rectangular slides (75 mm x 25 mm, 1 mm thickness) for the air-dried experiments, or onto circular disks (13 mm diameter, 1 mm thickness) for the live cell experiments.

In order to avoid displacing the cells from the slides, the Petri dishes were carefully topped up with medium after 5 to 7 minutes from when the CaF_2 slides were seeded, being finally incubated. The CaF_2 slides were cleaned with the solvent NuSol rapide and rinsed with deionised water before each use. The first 24h of incubation were performed with 15 mL of growth medium, and a higher concentration of FBS (15%), with 0.1% of Amphotericin B (0.25 μg/mL - HyClone, GE Healthcare Life Sciences) to prevent fungal incidence. These first 24h of incubation at a higher serum concentration in the medium allowed the cells to attach to the surface of the CaF_2 slides. Raman cell spectra are known to be dependent on the cell cycle [40], thus after the initial incubation, the medium was replaced by a low serum concentration medium with 0.5% FBS and re-incubated for other 24h to synchronise the cell cycle at G_0 via nutritional deprivation [148]. Therefore, before any data collection could occur, the samples had to be prepared 48h in advance.

For the air-dried experiments, the growth medium was removed and the slide was then rinsed three times with 1x PBS and air dried before Raman data collection. No fixative solutions were used as these can influence the Raman signal of the cells; air drying and sample desiccation provide good signal intensity and reasonable preservation of cell components visible in the Raman spectrum [45].

For live cell experiments, the medium was removed and the disks were then rinsed three times with 1x Hank's balanced salt solution (HBSS) (Gibco, Thermo Fischer). The Petri dish was then topped up with 17 mL of 1x HBSS and the cells were then analysed using a x63 water dipping lens. As the cells had to be kept outside their ideal growth conditions, the RS analyses were performed with 4h from the point of removing the samples from the incubator.

The RS acquisition parameters used for both air-dried and live cell data collection are summarised in Section 2.3.

2.2.4 Osteogenic and adipogenic differentiation

The differentiation assays were performed on the Y201 cell line due to its tri-lineage (osteogenic, adipogenic and chondrogenic) differentiation capacity [84]. Cells were seeded near confluence on 24 well plates and also onto CaF_2 disks (13mm diameter x 1mm thickness) that were placed inside 10 cm Petri dishes. The well plates were used to perform histological staining assays to confirm differentiation, while the CaF_2 disks were used for the Raman data collection. In both cases, the cells were incubated with DMEM containing 1% penicillin/streptomycin (100 units/mL - penicillin and 100 μg/mL - streptomycin; Gibco, Thermo Fischer), 0.1% Amphotericin B (0.25 μg/mL - HyClone, GE Healthcare Life Sciences) and supplemented with 15% FBS. This culture medium is termed here as basal medium as it does not contain any of the supplements needed to promote cell differentiation. After the cells were confluent, they were then induced into osteogenic or adipogenic differentiation, with this being day 0 of the differentiation process.

Osteogenic differentiation was induced with basal medium supplemented with L-ascorbic acid-2-phosphate (50 μg/mL), β-glycerophosphate (5 mM) and dexamethasone (10 nM). In order to induce adipogenic differentiation, the basal medium was supplemented with dexamethasone (1 μM), isobutyl-methylxanthine (500 μM), insulin (1 μg/mL) and indomethacin (100 μM). Along with the osteogenic and adipogenic differentiation experiments, a control experiment was prepared by keeping the cells in basal medium and incubating them for the same period of time.

The osteogenic, adipogenic and basal media were changed twice a week. The differentiation and control experiments were stopped at days 0, 7, 14 and 21 for Raman analyses and staining assays. The Raman analyses were performed on live cells with the data acquired using a x63 dipping lens immersed in 1x HBSS and the acquisition parameters described in Section 2.3. Since Raman cell spectra are dependent on the cell cycle, the

cells were also synchronised to G0 via nutritional deprivation by replacing the medium, either basal, osteogenic or adipogenic, with a low-serum (0.5% FBS) medium 24h prior to the data collection.

2.2.5 Alkaline phosphatase and von Kossa staining

Mineralisation was confirmed using alkaline phosphatase (ALP) and von Kossa staining [149]. These assays were performed on the cells seeded onto the 24 well plates, and conducted at days 0, 7, 14 and 21 for samples kept in both basal and osteogenic conditions. First, the cells were washed twice with 1x PBS. A filtrated (Millex GP 0.22 μm sterile filter) alkaline phosphatase (ALP) reagent mix was then added and left for 1 min. This reagent mix consists of 0.2 mg/mL naphtol AS-MX (Sigma-Aldrich) in 1% N,N-dimethylformamide diluted in 0.1 M Tris (base) pH 9.2 plus 1 mg/mL of Fast Red TR. The cells were then washed twice with 1x PBS and fixated with 4% paraformaldehyde for 5 min. This procedure was followed by a wash with 1x PBS and a wash with deionised water. A 1% silver nitrate solution was then added and the plates were left on a light box for 60 min. Plates were again washed three times with deionised water, followed by 2.5% sodium thiosulphate for 5 min. Finally, the plates were washed twice with deionised water and left in a 20% glycerol in PBS for imaging. Bright field images were saved on a Zeiss AxioZoom stereomicroscope using a planApo S 0.63x lens and on a Leica DM IRB microscope with x5 and x10 objective lenses.

2.2.6 Oil Red O staining

Lipid accumulation was examined by Oil Red O staining of the cells seeded onto the 24 well plates and performed at days 0, 7, 14 and 21 for samples kept in both basal and adipogenic conditions. At each time point analysed, the culture medium was removed

and the wells were rinsed with 1x PBS and then fixed with 4% paraformaldehyde for 10 min. The plates were then rinsed once with deionised water and left for 5 min in 60% isopropanol. Following this, isopropanol was then removed and Oil Red O staining solution (0.3% Oil Red O in 60% isopropanol) was added for 10 min. One final rinse with 60% isopropanol was then performed to remove the excess of Oil Red O stain together with three rinses with deionised water. The samples were stored in 20% glycerol in PBS with bright field images saved using a Leica DM IRB microscope with x5 and x10 objective lenses.

2.3 Summary of RS acquisition parameters for cell analyses

Many of the selected settings for the adjustable components of the micro-Raman system were described in Section 2.1, and are summarised below.

For this thesis, all spectra collection were performed using an HORIBA XploRA micro-Raman system, equipped with an automated X-Y-Z motorised stage and CCD detector kept at -50°C. The system allows visual optical inspection and Raman mode spectral data collection. The acquisition of cell spectra was performed using a x100 (0.9 NA) lens, for air-dried cell experiments, and a x63 (1.0 NA) dipping lens, for the live cell analyses.

The XploRA system is operated using Labspec software, which allows the selection of settings for all adjustable components. During this research, Labspec 5 was initially used, and then later upgraded to Labspec 6. The samples were irradiated for 45 s (per spectral window) using a 532 nm laser. The Raman acquisition consisted of an average of two accumulations of irradiation time per spectral window to allow the software to correct for cosmic rays that may reach the CCD, and also to decrease noise. All spectra were obtained over the nucleus of the cells, unless they had been induced to differentiate, in which case the data were obtained from their differentiation products, *i.e.* mineralised

nodules for osteo-induced cells and lipid droplets for the adipo-induced cells.

The laser power was reduced to 3.5 mW (laser power density of $\sim 9.2 \times 10^9$ W/cm^2) by selecting a 50% neutral density filter. A pinhole of 100 μm was used to allow for confocal Raman acquisition. A slit of 200 μm combined with a 2400 grooves/mm diffraction grating provided a spectral resolution of ~ 1 cm^{-1}.

The spectral range was dependent on the type of experiment: for air-dried cell analyses, spectra were collected from 600 cm^{-1} to 1800 cm^{-1} (fingerprint region), while for live-cell analyses, data were obtained from the fingerprint region plus high-wavenumber region (2600 cm^{-1} to 3200 cm^{-1}).

Air-dried experiments were performed on the undifferentiated MSC lines (Y101, Y201, Y102 and Y202), primary human MSCs (K72), the sorted primary fraction (CD317+) and HDFs. For these experiments, five spectra were obtained from the nucleus of each cell. The second order standard deviation and the standard error of the mean were checked for their convergence to verify that the quantity of spectra was sufficiently representative of the cell line analysed. These convergence tests are explained further on Subsection 2.6.2.

For the air-dried experiments, in total, 560 spectra were collected for Y101 cell line (over six experimental repeats), 200 spectra were collected for Y102 cell line and 180 spectra were collected for Y201 and Y202 cell lines (over two experimental repeats). The HDFs and primary human MSCs were analysed once with 100 spectra obtained for each sample. An additional experiment was performed for the primary human MSC K72, where three cells were randomly selected and had 36 spectra collected from the nucleus of each one of them, in order to individually compare them to the MSC lines.

Live cell experiments were performed on the undifferentiated MSC lines (Y101, Y201, Y102 and Y202) and for MSC line Y201 during osteogenic and adipogenic differentiation. In these experiments only a single spectrum was collected per cell to avoid cell death caused

by increased laser exposure. In total, 48 spectra were obtained for each undifferentiated MSC line over three experimental repeats (16 spectra obtained for each experiment). These quantities of spectra obtained were also checked for their convergence of the standard error of the mean and second order standard deviation. During the differentiation process, either osteogenic or adipogenic, cells were analysed at days 0, 7, 14 and 21 with 16 spectra obtained for each time-point.

In order to prevent laser-induced damage, all cells were monitored during real-time acquisition and with optical inspection performed after each Raman measurement. In addition, a trypan blue assay was developed to assess if the laser exposure could cause cell death during the live cell experiments. This assay is described in Subsection 2.3.1.

Before each experiment, the Raman system was auto-calibrated with respect to the wavenumber position using a Si slab. This procedure is automated in Labspec and only requires the Si slab to be focused under the objective.

2.3.1 Trypan blue assay for live cell experiments

A trypan blue staining assay was performed to check if the laser exposure could induce cell death during the live cell data collection. The MSCs were seeded onto CaF_2 disks and incubated, as described in Subsection 2.2.3. Using the x63 dipping lens immersed in 1x HBSS, a random cell was selected and optical images were saved. The cell was then irradiated using the 532 nm laser and under the same acquisition conditions described in Section 2.3. After both the fingerprint and the high-wavenumber regions were analysed, the total amount of time during which the cells were laser irradiated was determined as 7.5 min. Following the laser irradiation, the HBSS was drained and 200 μL of 0.2% trypan blue solution (Thermo Fisher) was added for 1 min [150]. The CaF_2 disk was then rinsed three times using 1x HBSS and the Petri dish was then further topped up, so that new

images of the chosen cell could be saved. Finally, as a control, the HBSS was again drained and, in order to kill the cells, 300 μL of 70% ethanol was added for 5 min. The trypan blue staining was then repeated and more optical images were saved.

This protocol was repeated three times for three randomly chosen cells. It is important to highlight that the cell specificity, where the same cell was evaluated before and after laser exposure, was only possible due to the use of an automated X-Y-Z stage.

2.4 Bone fragment analyses

Bone fragments were investigated using RS in order to compare their mineral properties to the mineral produced by the MSCs during osteogenic differentiation. This included the analyses of human and lamb bone fragments.

The human bone originated from the femoral heads or tibial plateaus obtained during hip or knee replacement surgeries and was donated to Prof Paul G. Genever's research group. Small fragments, from four different donors (here termed FH1, FH2, FH3 and K1) were obtained during the isolation of the primary human MSCs and were frozen at -80°C. They were thawed in HBSS after 24 h from being first frozen and analysed using Raman micro-spectroscopy with the x63 (1.0 NA) dipping lens. Spectra were obtained from various parts of the fragments using the same acquisition parameters described for the live cell analyses in Section 2.3.

The lamb bone was prepared for RS by Jake Hodgetts from Dr Yvette Hancock's research group with the methods of preparation described in his Master's dissertation [151]. This method included using hot water maceration to prepare a section of the rib, which was boiled for 20-30 min to remove the flesh and fat, cooled down and air-dried. A thin section of the bone was then cut and polished until a thickness of 5-30 μm was obtained. The RS acquisition parameters for this lamb bone fragment were the same used for the

human bone fragments, however, it was analysed in air using the x100 (0.9 NA) lens.

2.5 Lipid analyses

Fatty acids, lipids and phospholipids were analysed using RS in an attempt to the identify the lipids stored in the lipid droplets formed during adipogenesis of the MSCs. All samples in this study were provided by Dr Martin G. Rumsby from the Department of Biology, University of York. The lipid samples were analysed either as pure lipids or dissolved into chloroform/methanol-2:1 (CM) (Table 2.5). A small volume (5 μL) of the liquid samples was pipetted onto a CaF_2 slide and immediately analysed. For solid samples, a small quantity of the powder was deposited onto the slide and sampled. To avoid thermal decomposition of the samples, a x50 lens was used so that the laser spot size and working distance were larger. The acquisition parameters were again the same as those used for live cell analysis as described in Section 2.3.

Table 2.2: List of fatty acids, lipids and phospholipids analysed using RS.

Classification	State	Dissolution	Sample
Sterol	Solid	Pure	Cholesterol
Fatty acid	Liquid	Pure	Eicosapentaeonic acid (EPA)
Fatty acid	Liquid	Pure	Linoleic acid
Fatty acid	Liquid	Pure	Oleic acid
Fatty acid	Solid	Pure	Palmitic acid
Phospholipid	Liquid	CM (50 mM)	Phosphatidic acid
Phospholipid	Liquid	CM (50 mM)	Phosphatidylcholine
Phospholipid	Liquid	CM (50 mM)	Phosphatidylethanolamine
Phospholipid	Liquid	CM (50 mM)	Phosphatidylserine
Fatty acid	Liquid	Pure	Stearidonic acid
Triacylglycerol	Liquid	CM (50 mM)	Triolein
Fatty acid	Liquid	Pure	γ-linolenic acid

2.6 Data analyses

2.6.1 Ratiometric analysis

The ratiometric analyses of the Raman data were performed using minimal pre-processing and, thus each individual spectrum per sample was not smoothed or normalised. Spectra obtained for the air-dried experiments were initially baseline corrected using a linear interval baseline, extending from the first to the last spectral point using the *Raman tool set* package [152]. Further data analyses, such as spectra averaging and peak fitting, were carried out using IGOR Pro 6.32 software.

The spectra obtained for each sample were first averaged thus, obtaining a representative spectrum for each sample. Statistical quantities such as the standard deviation (SD) and standard error of mean (SE) were calculated directly on IGOR Pro, over the whole spectrum, and are represented by equations 2.3 and 2.4, respectively [153].

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} \tag{2.3}$$

and

$$SE = \frac{SD}{\sqrt{n}}, \tag{2.4}$$

where n is the total number of measurements, or spectra obtained, y_i is the intensity of each point in the spectra and $\bar{y}$ is the mean of the intensity for each point. The SD represents how scattered the measurements are, while the SE indicates the uncertainty around the mean. It can be seen from equation 2.4 that the SE depends on the sample size n and it decreases with it.

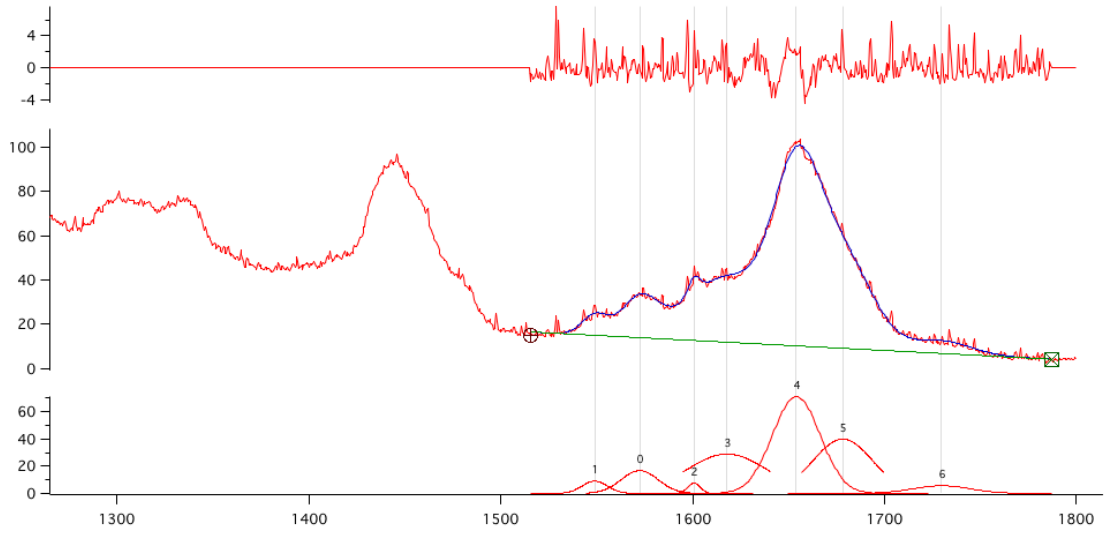
More comprehensive examination of the spectra included Gaussian peak fitting along with linear baseline subtraction for each band envelope fitted. Where necessary, peak

deconvolution was performed during the fitting process.

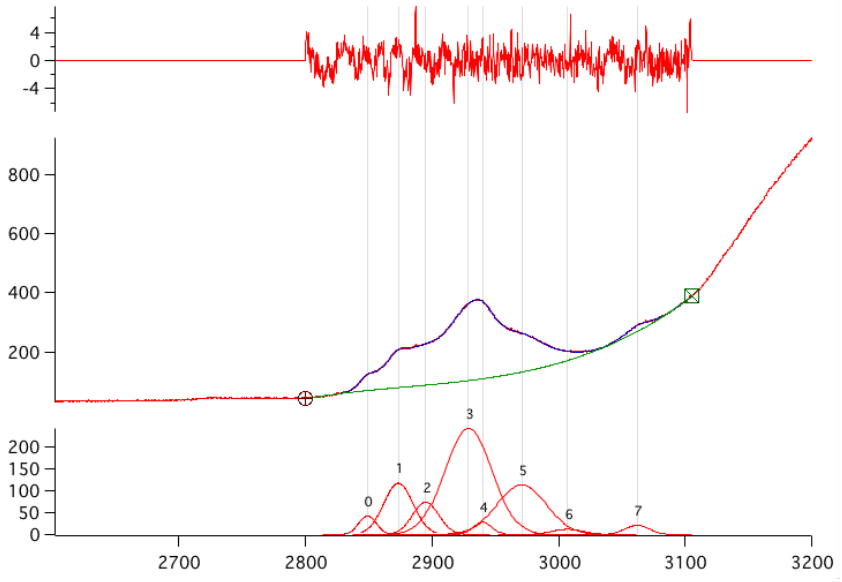
The fitting of the averaged spectra was performed for regions containing single peaks or groups of convolved peaks. When a fitting is performed on IGOR Pro, one can inform the software to automatically locate peaks in a selected region. Thus, the software estimates the number of peaks and their position for that range. These peak positions are then used as initial guesses during the fitting process and may change when the fitting is finished and optimised. Other initial guesses that are automatically calculated by the software include the peak width and amplitude. The peak fitting is concluded when a minimum value is obtained for the χ^2 and the final values for peak position, width, full width at half maximum (FWHM), area and amplitude are reported.

Examples of the IGOR Pro multi-peak fitting procedures are shown for the 1530 cm^{-1} to 1780 cm^{-1} range in the fingerprint region, which contain seven peaks (Figure 2.7a) and in high-wavenumber region from 2800 cm^{-1} to 3100 cm^{-1}, which contains eight peaks (Figure 2.7b). Subfigure 2.7a is the average of 180 spectra obtained for the air-dried Y201 cell line, while subfigure 2.7b is the average of 48 spectra obtained for the live analyses of the same cell line. As these figures show, linear baseline subtraction was performed for ranges fitted in the fingerprint region, whilst for the high-wavenumber region, a cubic function was baseline subtracted from the experimental data.

The process of fitting peaks to the experimental data generates a fitting error that only informs how well a peak was fitted, reporting nothing about the biological variance within that sample. For example, in the air-dried samples, the fitting error was typically of the order of 1.3% of the peak intensity, while the SE of the mean for the same peak was about 4.1% of its intensity. Therefore, the measure of uncertainty used for the peak intensities, in all of the Raman datasets studied was the SE of the mean obtained for each peak.



(a)



(b)

Figure 2.7: Example of multi-peak fitting using IGOR Pro. Both fitting examples show the experimental data in red, the fitted curve in blue and the individual peaks that generated the fit. Also shown are the baseline corrections in green and the fitting residuals (top). (a) The fingerprint region fitted with seven peaks ranging from 1530 cm^{-1} to 1780 cm^{-1} where a linear baseline was used and is the average of 180 spectra obtained for the air-dried Y201 cell line and (b) high-wavenumber region fitted with eight peaks ranging from 2800 cm^{-1} to 3100 cm^{-1}, which was corrected using a cubic baseline subtraction, and is the average of 48 spectra obtained for the analyses of live cells from the same cell line.

The ratiometric analyses are based on the calculation of peak intensity ratios (PIRs) derived from the peak intensities obtained during the fitting process. PIRs were calculated for all peaks relative to all others in order to identify possible biomolecular markers and also between specific peaks to calculate certain parameters, *e.g.* the hydroxyapatite crystallinity in the mineral nodules observed during osteogenesis and the unsaturation levels of the fatty acids stored in the lipid droplets formed during adipogenesis. Since each peak has its uncertainty measured by the SE, the uncertainty associated with the PIRs is therefore the propagation of the SE for the quotient expressed by Equation 2.5:

$$\frac{A \pm \delta A}{B \pm \delta B} = \frac{A}{B} \pm \frac{A}{B}\sqrt{(\frac{\delta A}{A})^2 + (\frac{\delta B}{B})^2} \tag{2.5}$$

where A and B are the intensities of the two peaks used for calculating the PIR and δA and δB are the SE associated with them. The first term on the right is the calculated PIR and the second is its uncertainty.

2.6.2 Convergence tests

Convergence tests were developed to ensure that the statistics are converged and that the averaged spectra represents the population average, thus assessing the quantity of spectra obtained on each experiment. These convergence tests are represented by three different approaches: graphs of (i) the convergence of the SE and (ii) the convergence of two times the standard deviation (2x SD), over the full averaged spectrum; and (iii) plots of the percent SE (%SE) of a certain PIR against the number of collected spectra. The SE and 2x SD convergence tests correspond to spectral averages up to the total amount of spectra collected per experiment. The %SE convergence test consists on checking how the SE of a certain PIR decreases as the number of obtained spectra increases. The %SE convergence tests attained to particularities of the different experiments performed in this study and

are described below.

In the air-dried experiments five spectra were obtained from random positions in each cell nucleus. These spectra were randomly averaged in groups of five, and from these averages the PIRs and respective percent SE were determined. For the live cell analyses, one spectrum was randomly obtained per cell, and therefore these spectra were randomly included one-by-one in the convergence calculations. The PIRs used on the convergence tests of the undifferentiated cells, either air-dried or live cells, were the 970/1085 and 1656/1085. For the osteo-induced MSCs, it was used the PIR 1654/1088, and for the adipo-induced MSCs, the PIR 1654/1448. The 970 cm^{-1} peak is related to vibrations of proteins and DNA/RNA, the 1085 cm^{-1} peak represents vibrations from lipids, carbohydrates, phospholipids and DNA/RNA, and the 1444 cm^{-1} and 1656 cm^{-1} peaks are assigned to lipids and proteins [36, 42, 56, 154–159]. Most of the spectral differences noticed for the MSCs, undifferentiated, osteo-induced and adipo-induced cells, were observed for these peaks, thus their inclusion for this test.

2.6.3 Principal component and linear discriminant analysis

In conjunction with the peak intensity ratio analyses, principal component analysis (PCA) and linear discriminant analysis (LDA) were performed on the MSC lines using the statistical package R Version 3.3.0, which was compiled through RStudio Version 0.99.902. For these analyses, each spectrum was linear baselined in the fingerprint region and cubic baselined in the high-wavenumber region. Other pre-processing methods included normalisation to the total spectrum area and smoothing by a cubic spline process, all of these were done using the *Raman tool set* package [152]. The spectra obtained for the air-dried experiments were then averaged to represent each analysed cell, once five spectra were obtained per cell. The averaging was not performed for the spectra obtained for the live

cell experiments as only one spectra was obtained per cell.

PCA is an unsupervised multivariate technique used for data classification based on their similarity. The PCA method finds the best low-dimensional representation of the dataset by generating orthogonal principal components (PCs) accounting for the major variance in the dataset [32]. For the PCA, each point in the Raman spectrum is considered a variable and each analysed cell is an observation. Thus, the number of observations for the air-dried experiments is 112 for the Y101, 40 for the Y102 and, 36 for Y102 and Y202. For the live cell analyses one spectrum was obtained per cell; therefore, the 48 spectra obtained for each undifferentiated MSC line, and the 16 spectra obtained for the osteogenesis and adipogenesis analyses, correspond to the amount of observations. The data used for the PCA was also standardised, meaning that it has mean zero and standard variation equals to one; thus, all variables are equally weighted.

The PCA components, or PCs, are uncorrelated and the first few retain most of the variation present in all of the original variables used [160]. Thus, scatter plots of the scores of the first three PCs were generated. The loadings of each one of these PCs were also plotted against the wavenumbers. The loadings are the coefficients of the linear combination that defined each of the PCs and describe the greatest variance in the spectra.

The LDA is a supervised classification model that optimises data classification by maximising the between-group variance whilst decreasing the within-group variance. By applying PCA prior to the LDA, the data is reduced and the LDA does not incorporate as variables the same number of observations [36, 161]. This type of PCA-LDA analyses have been used before in Raman data analyses demonstrating good group separation and classification [32, 36, 161, 162]. In order to prevent overfit of the data with the LDA, the criteria used to select the number of PCs incorporated into the LDA was to use half of the number of observations obtained for the smallest group [36]. For the air-dried cell analyses,

the smallest group pertained to the Y201 and Y202 MSC lines with 180 spectra obtained, thus generating 36 observations (from the average of the 5 spectra obtained in each cell). Therefore, to avoid overfitting, the half of this number of observations is 18, totalling the amount of PCs incorporated into the LDA. For the live cell analyses of the undifferentiated MSCs, all groups (Y101, Y201, Y102 and Y202) had 48 spectra collected, one per cell; thus, 24 PCs were used. Finally, for the osteogenesis and adipogenesis experiments, 16 spectra were obtained for each time point (day 0, 7, 14 and 21); therefore, eight PCs were used in the LDA.

Additionally, a classification model was built where two thirds of the data were used to train the PCA-LDA algorithm while the other third was tested for their classification. Finally, the model predictions were presented in a confusion matrix which described the performance of the classification model.

2.7 Raman mapping

Due to the long amount of time required for their collection, Raman maps were only performed on the air-dried samples using the same acquisition parameters described in Section 2.3. The sample culturing and preparation for Raman map analysis followed the same procedures previously explained in Sections 2.2.2 and 2.2.3. A cell was randomly selected in the population and had its nucleus scanned. The number of spectra obtained in each cell nucleus varied with its size and ranged from 361 to 484 spectra per cell. Mappings were only collected for the fingerprint region (600 cm^{-1} to 1800 cm^{-1}) consuming approximately 19h per map acquisition.

The spatially-resolved RS map images were processed using the IGOR Pro 6.32 software. All spectra were corrected using a linear baseline subtraction from the first to the last spectral point, in the 600 cm^{-1} to 1800 cm^{-1} range. As IGOR Pro allows the user

to write procedures to automate tasks, a procedure was then written to perform peak fitting for each spectrum obtained from the cell nucleus. The procedure fitted each peak, or group of convolved peaks, by performing a linear baseline subtraction for the fitted range as explained in Section 2.6.1. These fittings were then followed by the calculation of the PIRs, and finally, based on their intensities, a colour profile image (heat map) was created.

2.8 Additional analyses

Two additional experiments were performed on the four MSC lines by Dr James Fox, from Prof Paul G. Genever research group, and are described bellow. The cell ploidy experiment was used to determine in which stage of the cell cycle the MSC lines were after being synchronised. The nuclear counterstaining and Image J analyses were used to verify the nuclear area and perimeter of the MSC lines.

2.8.1 Cell ploidy

The cell ploidy experiment was used to confirm that the cells had been synchronised to the G0 after nutritional starvation. Cells in normal growth media at ∼70% confluence were briefly trypsinised; 125,000 cells were transferred to 6 well plates and allowed to adhere for 5h. Cells were washed 3x with PBS then incubated for 16h in DMEM containing only 0.5% serum. For flow cytometry, cells were washed and trypsinised to a single cell suspension then fixed in ice-cold methanol for 10 mins and further washed 2x with PBS. Hoechst (Life Technologies) nuclear acid stain was applied diluted in PBS to a final concentration of 5 μg/ml for a period of 30 mins followed by a PBS wash. Cells were immediately analysed on a Beckman Coulter CyAn ADP flow cytometer (Beckman Coulter, High Wycombe, UK) respecting a gating strategy to remove debris and doublets to record at least 10,000

events. Data were analysed using Summit v4.3 software.

2.8.2 Nuclear counterstaining and Image J analysis

Cells at ∼70% confluence were briefly trypsinised, 20,000 cells were transferred in triplicate per cell line and allowed to adhere in 24 well plates for 5h. Cells were washed x3 with PBS then incubated for 16h in DMEM containing only 0.5% serum. Cells were fixed with 3% paraformaldehyde (Park Scientific) for 20 mins and gently washed 3x with PBS. DAPI (2 μg/ml) was applied for 10 mins before an additional 3x washes in PBS. Each replicate was imaged at 10x magnification on a Leica DM IRB microscope (Leica) coupled with a Leica DC500 camera. Images were processed in Image J v1.49 (National Institutes of Health, USA). Briefly, the RGB images were threshold adjusted and particles were analysed with area and perimeter measurements being recorded. Images were screened for validity with exclusion of: all cells at the edge of the image; small objects that were not cells and large objects that were two or more cells that Image J could not discriminate, each image taken included at least 95 cells that met the inclusion criteria. Data from the three Image J replicates were transformed to μm or μm^2 and averaged in GraphPad Prism v6.07.

Chapter 3

Undifferentiated mesenchymal stromal cell line discrimination and heterogeneity assessment

MSCs are known to form heterogeneous populations, which together with a lack of unique cell surface expression markers, compromises their full application in regenerative medicine. To circumvent these limitations, this chapter proposes to use RS biomolecular fingerprinting to characterise air-dried samples of four immortalised clonal MSC lines. Thus, the chapter evaluates the ability of RS to discriminate subpopulations of cells, where all four clonal MSC lines express the same panel of surface proteins, but have contrasting differentiation capacities. The MSC lines were also compared to HDFs and to a primary MSC fraction sorted by the positive expression of the CD317 protein. PCA, a multivariate method for data analysis, was used to reveal the spectral bands corresponding to most of the variance across the MSC lines and PCA-LDA was used for data classification. Univariate analysis of the spectra, by means of peak intensity ratio (PIR) analyses, were used

to identify biomarkers linked to cell fate, *i.e.* RS markers that discriminate differentiation competent from differentiation incompetent cell lines. Spatially-resolved Raman maps performed on these air-dried MSCs allowed for localisation of the discriminating markers on the sampled nuclei, revealing the differences in morphology between the studied cell lines. The population heterogeneity was also investigated, by means of the %SE, SE and 2x SD convergence tests. Finally, through the use of PIR markers, single cells from a heterogeneous primary MSC population were characterised with RS aiming to identify any similarities with the clonal MSC lines.

3.1 Averaged spectra and peak assignments

The laser spot of $\sim$0.7 μm allowed all spectra to be collected exclusively over the nucleus of the air-dried single cells, as described in Section 2.3. The analysed samples were the four hTERT immortalised MSC clonal lines (Y101, Y201, Y102 and Y202), a primary MSC population sorted for the positive expression of the CD317 protein (CD317+), HDFs and a heterogeneous primary population (K72). The culturing methods and sample preparation for the Raman analyses were described in Section 2.2.

The averaged Raman spectra and the standard error of the mean envelopes of the studied cells are shown in Figure 3.1 along with the identification of the peaks included in this study. The peak positions in Figure 3.1 were achieved by averaging the positions obtained from the peak fittings performed on the cell lines, as per Section 2.6.

As described in Chapter 2, these spectra represent the average of 560 spectra obtained for the Y101 cell line, 200 spectra for the Y102 cell line and 180 spectra for the Y201 and Y202 cell lines. The Y101 cell line underwent six experimental repeats, whereas the other cell lines included two experimental repeats. HDFs and primary human MSC populations (CD317+ and K72) were analysed using 100 spectra each that were obtained from a single

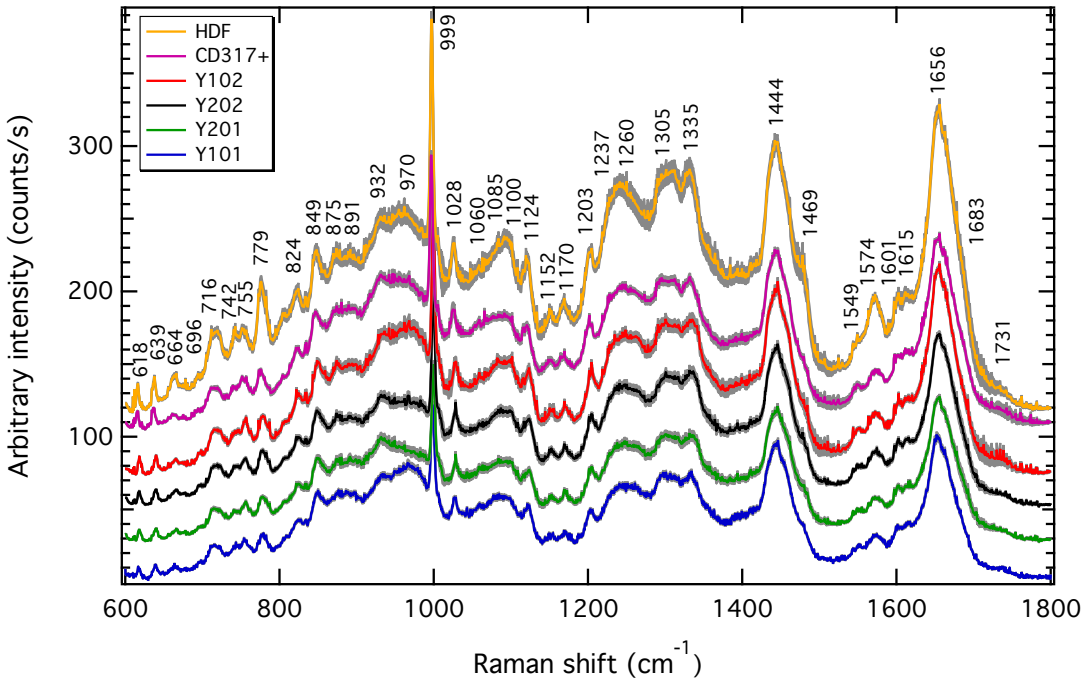


Figure 3.1: Averaged Raman spectra of air-dried undifferentiated MSC lines, HDFs and CD317+ cells. The spectra were obtained at random positions in the nucleus of the cells comprising each of these populations. The standard error of the mean envelopes (grey curves) are also shown. The peak positions are the averages obtained from peak fitting routines performed on each cell line. The spectra were offset for clarity.

experiment. These averaged spectra are a statistically converged representation of each population, as determined by the %SE convergence test, together with the SE and 2x SD full spectral convergences. These convergence tests were described in Section 2.6.2 of Chapter 2 and their results are discussed in Section 3.4.1 of this chapter. The convergence tests are displayed in Appendix B, in Sections B.2 and B.3. It is important to highlight that although the convergence tests determined the minimum quantity of spectra needed to describe each MSC line, the full amount of spectra obtained for each of the lines was used. This was chosen so that the variability between the different experimental repeats could be assessed. The convergence tests were established first to determine if the averaged spectra is representative of the population of cells and, secondly, to assess the cell population heterogeneity.

The four MSC lines have contrasting differentiation capacities, as described in Subsection 2.2.1. Specifically, the Y101 and Y201 cell lines possess trilineage differentiation capacity, whereas the Y102 and Y202 cell lines are differentiation incompetent. The CD317+ primary population fraction was sorted using fluorescence activated cell sorting (FACS) from a donor different than that of the MSC lines. The CD317 protein is positively expressed only by the Y102 and Y202 MSC lines [84], thus this primary fraction was used as a control to confirm the use of RS for subpopulation discrimination, while the HDFs were used as a non-MSC stromal cell control.

The general and detailed peak assignments in Table 3.1, were obtained for each measured peak according to the literature. Some of the peak positions are assigned to more than one biomolecule constituent due to the overlapping vibrations of different molecules. For example, the peak at 1085 cm^{-1} is assigned to lipids, phospholipids, carbohydrates and DNA/RNA [40, 42, 56, 154, 163]. The peak positions displayed in Table 3.1 correspond to the average of peak positions obtained for all collected spectra. Since the SE for most of these peak positions was below 1 cm^{-1}, which is the equipment's precision, this value was used as a measure of their uncertainty; however, peaks at 1469 cm^{-1} (proteins, lipids and DNA/RNA) and 1683 cm^{-1} (lipids and proteins) showed greater difference in their mean positions across the cell lines, with a standard error of the mean (SE) of $\pm$ 3 cm^{-1}, and this was the measure of their uncertainty. This increase in peak position uncertainties may be due to both peaks being fitted as shoulders of other more intense peaks at 1444 cm^{-1} and at 1656 cm^{-1}.

Table 3.1: General and detailed Raman peak assignments. The assignments correspond to the measured peak positions, averaged across the spectra shown in Figure 3.1, as defined in the literature. The measured peak position uncertainty is the equipment's precision (± 1 cm^{-1}) except for the 1469 cm^{-1} and 1683 cm^{-1} peaks, which showed greater differences across the cell lines. For these peaks the SE of ± 3 cm^{-1} was used as the uncertainty measure.

Peak position (cm^{-1})	Literature range (cm^{-1})	General biomolecular assignments	Detailed biomolecular assignments
618	618-623	Proteins	C-C twist of proteins [154]. Skeletal phenylalanine [41, 66, 156]. Phenylalanine C-C twist [42, 56].
639	640-645	Proteins	C-S stretch & C-C twist of tyrosine [154]. Tyrosine [41, 66, 156]. C-C twist of tyrosine [42, 56].
664	666-678	DNA/RNA	Guanine and thymine ring breathing mode [41, 42, 56, 154]. Tyrosine-G backbone in RNA [154].
696	700	Proteins	Aminoacid methionime ν(C-S) *trans* [154].
716	717-719	Lipids	$CN^{+}(CH_3)_3$ stretch in lipids [36, 42, 56, 154]. C-N membrane phospholipid head stretch [48, 154]. Phospholipids [32].
742	746	DNA	Thymine ring breathing mode [154, 163].
755	755-760	Proteins	Tryptophan ring breathing mode [33, 36, 42, 48, 56, 141, 154, 164].
779	774-788	DNA/RNA	Uracil ring breathing mode [154]. Thymine and uracil ring breathing modes [43]. Cytosine and thymine modes [66]. Uracil, cytosine and thymine ring breathing modes [32, 42, 56]. DNA/RNA O-P-O symmetric stretch [30, 141, 163]. Uracil, cytosine, and thymine overlapping with O-P-O stretch in DNA [33, 48, 156, 164].
824	823-830	Proteins & DNA/RNA	DNA O-P-O asymmetric stretch [32, 163]. Tyrosine out-of-plane ring breathing and O-P-O stretch in DNA [36, 42, 56, 154]. Proline and hydroxyproline, out-of-plane tyrosine ring breathing, DNA/RNA asymmetric O-P-O stretch [48].
849	850-853	Proteins	Tyrosine ring breathing [36, 41, 42, 56, 66, 154, 156, 164]. C-C in proline and tyrosine ring breathing mode [69]. Protein related peak [33].

Continued on next page

Table 3.1 – *Continued from previous page*

			C-C stretch in collagen, tyrosine ring breathing, glycogen, polysaccharides C-O-C stretch [48].
875	874-877	Proteins, lipids & carbohydrates	Proteins C-C stretch, antisymmetric vibrations of choline $N(CH_3)_3$ [154]. C-C-N+ symmetric stretch (lipids) and C-O-C ring (carbohydrates) [42, 56]. Hydroxyproline C-C stretch in collagen, tryptophan ring deformation, choline asymmetric stretch in phospholipids and C-O-C stretch in carbohydrates [48].
891	889-891	Carbohydrates	Saccharide band, methylene and C-C skeletal [154].
932	915-938	Proteins	Skeletal C-C stretch α-helix [32, 33, 41, 69, 126, 154, 156, 164]. C-C backbone stretch α-helix and C-O-C glycogen [36, 42, 48, 56].
970	968-975	Proteins & DNA/RNA	C-O-P phosphodiester residue [40]. Ribose (DNA/RNA) [159]. Phosphate monoester groups of phosphorylated proteins and DNA/RNA cellular nucleic acids [154, 157, 158].
999	1000-1005	Proteins	Phenylalanine ring breathing [30, 33, 34, 36, 40–42, 48, 56, 66, 126, 154, 156, 159, 163]. Phenylalanine ring breathing or histone proteins if collected in nucleus [164].
1028	1029-1033	Proteins	Phenylalanine ring breating [33, 36, 41, 42, 56, 154]. Collagen, keratin, C-N stretch in proteins, Phe C-H in plane bending, phospholipids and polysaccharides [48].
1060	1060-1083	Lipids, carbohydrates, phospholipids & DNA/RNA	Peak indicative of phospholipids [126]. Lipid peaks [66]. DNA PO^{-2} backbone stretch, lipids chain C-C stretch, carbohydrates C-O and C-C stretch [42, 56, 154].
1085	1083-1095	Lipids, carbohydrates, phospholipids & DNA/RNA	DNA/RNA PO^{-2} symmetric stretch [40]. DNA PO^{-2} backbone stretch and C-C gauche in lipids and phospholipids [154, 163]. DNA PO^{-2} stretch, lipids chain C-C stretch, carbohydrates C-O and C-C stretch [42, 56].
1100	1095-1100	Lipids, proteins, carbohydrates & DNA/RNA	DNA PO^{-2} stretch [32, 41, 43, 66, 141, 154, 164]. DNA PO^{-2} stretch and protein C-N vibration [155]. C-C vibration of gauche bonded chain [154]. Peak indicative of lipids [126]. DNA PO^{-2} stretch, lipids chain C-C stretch, carbohydrates C-O and C-C stretch [36, 42, 56].
1124	1123-1128	Lipids &	Protein C-N stretch [36, 41, 42, 56, 66, 154, 155].

Continued on next page

Table 3.1 – *Continued from previous page*

		proteins	ν(C-C) skeletal backbone in lipid (transconformation) [154, 163]. C-C stretch in proteins and C-O stretch in carbohydrates [41, 42, 154]. Peak indicative of phospholipids [126]. Lipids [159].
1152	1152-1158	Proteins	C-C and C-N stretch in proteins [36, 42, 56, 154]. Carotenoids [154].
1170	1170-1176	Proteins	C-H in plane bend of tyrosine [42, 56, 154]. Tyrosine and phenylalanine [41]. Protein related peak [33].
1203	1200-1209	Proteins	Tyrosine and phenylalanine [41, 42, 56, 154, 163]. Phenylalanine and tryptophan [36]. CH_2 wagging of glycine and proline [154]. Amide III [154, 164].
1237	1237-1240	Proteins, lipids & DNA/RNA	Amide III and CH_2 wagging of glycine and proline [154]. PO^{-2} antisymmetric stretch [40]. Amide III and thymine [41]. Amide III, thymine, adenine and =CH bend of lipids [42, 56].
1260	1257-1263	Proteins, lipids & DNA/RNA	Amide III, thymine, adenine and =CH bend of lipids [36, 42, 56, 154]. α-helix proteins [32]. Amide bond [159].
1305	1300-1305	Proteins, lipids & DNA/RNA	CH_2 deformation in lipids, adenine and cytosine [154]. Adenine and CH deformation [41]. =CH deformation in lipids [66, 156]. Lipids, proteins (Amide III), adenine and cytosine [163]. Nuclear acids and lipids [32]. CH_2 twist in lipids [42, 56, 69]. Lipid band [30]. Peak indicative of phospholipids [126].
1335	1335-1339	Proteins & DNA/RNA	Guanine and CH_3CH_2 wagging of collagen [154]. DNA and RNA [66]. Adenine, phenylalanine and CH deformation [163]. Adenine, guanine and CH deformation [155]. Adenine and CH deformation [41].
1444	1440-1447	Lipids & proteins	CH_2 deformation in lipids and proteins [154,163,164]. CH deformation in proteins [41, 43, 155]. CH_2 deformation in lipids [30, 69]. CH in carbohydrates [34].
1469	1450-1485	Proteins , lipids	Nuclear acids [32].

Continued on next page

Table 3.1 – *Continued from previous page*

		& DNA/RNA	Guanine, adenine, CH deformation of proteins and lipids [36]. Guanine, adenine, CH deformation of proteins, lipids and carbohydrates [42, 56]. C=N stretch [154]. Guanine, adenine, CH deformation of proteins, lipids and carbohydrates [42, 56, 154].
1549	1548-1554	Proteins	Tryptophan [154]. Amide III [34].
1574	1573-1578	DNA/RNA	Guanine and adenine [36, 41, 42, 56, 154, 155, 164]. DNA/RNA [66]. Peak indicative of lipids [126].
1601	1600-1607	Proteins	C=C of Phenylalanine and tyrosine [41, 42, 56, 155]. Amide I C=O stretch and phenylalanine [154]. Protein band [66].
1615	1610-1617	Proteins	C=C of tyrosine and tryptophan [42, 56, 154, 155]. C=C of tyrosine and phenylalanine [36].
1656	1651-1660	Lipids & proteins	Amide I [30, 34, 41, 126, 159, 164]. Amide I and C=C stretch of lipids [36, 42, 56, 154–156]. C=C stretch of lipids [66, 69, 163]. α-helix proteins [32]. C=C stretch of phospholipids [43].
1683	1680-1685	Lipids & proteins	Amide I [154]. Amide I and C=C stretch of lipids [42, 56].
1731	1729-1743	Lipids	C=O ester group in lipids [42, 43, 56, 154, 156].

The spectra of the four MSC lines and primary CD317+ in Figure 3.1 show similar features, with the main visual differences being observed for the protein peak at 932 cm^{-1}, the protein & DNA/RNA vibrations at 970 cm^{-1} and for the lipids, phospholipids, carbohydrates and DNA/RNA band at 1085 cm^{-1}. The peaks at 779 cm^{-1} and at 1574 cm^{-1}, both assigned to DNA/RNA nucleic acid vibrations, are also more distinct in the HDF population compared to the other cell lines.

3.2 Multivariate analysis: PCA-LDA of the air-dried MSC lines

Multivariate analysis, comprised of PCA followed by LDA, was performed on the spectra obtained from the air-dried MSC lines Y101, Y102, Y201 and Y202. PCA is an unsupervised technique used for data classification that generates a low-dimensional representation of the data by orthogonal principal components (PCs) accounting for the major variance in the dataset. LDA is a supervised method that is used to maximise the variance in the dataset, and in this work it was carried out using the PCs obtained from the PCA.

For these analyses, the five spectra obtained for each single cell were averaged, baseline corrected, area-normalised and smoothed, as explained in Section 2.6.3. Once different amounts of spectra were obtained for each of the MSC lines, the number of observations for each group is also different, such as that it corresponds to 112 observations for the Y101 line, 40 observations for the Y102 line and 36 observations for the Y201 and Y202 lines. Such uneven sampling size affects the interpretation of the classification with the LDA. However, regardless of the total sample size, the ability to discriminate the cell lines is governed by the smaller of the groups [165]. Although an uneven sample size may bias the interpretation of multivariate analyses, it is common that different amount of spectra per group is used in the Raman analyses of cells [23, 31, 36, 162].

Figure 3.2 shows the scores obtained for the PCA analysis performed on the MSC lines. The first three PCs accounted for 67% of the total variance in the dataset, with variances of 34%, 23% and 10%, respectively, for PC1, PC2 and PC3. PCA alone is not able to fully discriminate the four MSC lines; however, cell line Y101 starts to be separated from the others when PC1, PC2 and PC3 are plotted together (Figure 3.2).

The loadings of these three principal components are shown in Figure 3.3. PC1 load-

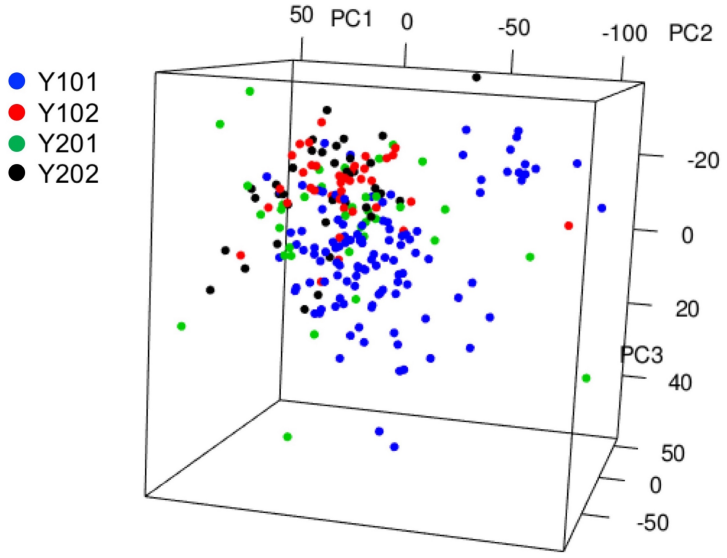
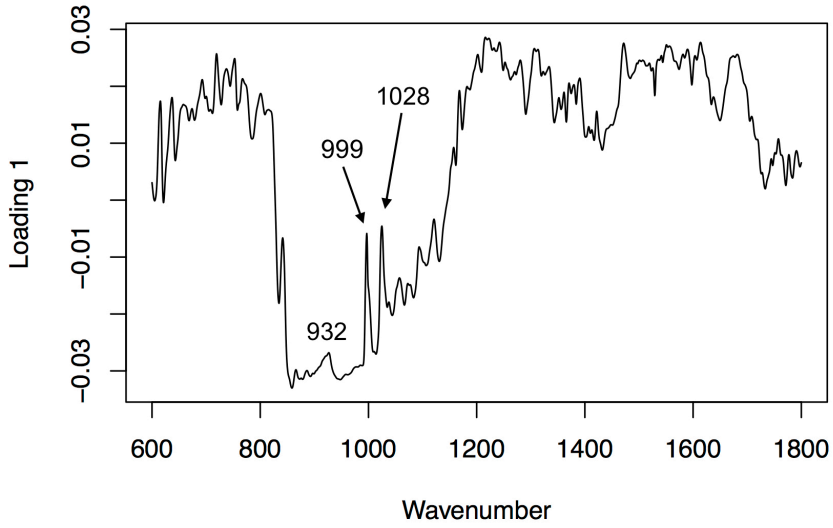


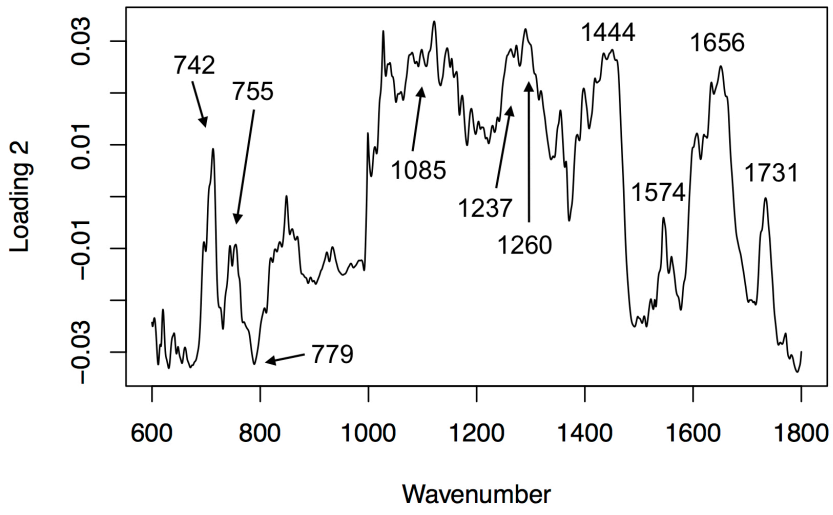
Figure 3.2: PCA scores for the MSC lines. PCA is not able to fully discriminate the four cell lines, however, the Y101 MSC line is shown nearly separate from the other lines. PC1, PC2 and PC3 accounted for 34%, 23% and 10% of the total variance in the datasets.

ing indicates a significant contribution of the background common to these cell lines, specifically between 800 cm^{-1} and 1200 cm^{-1} (Figure 3.3a). This reflects the pronounced background, observed across the cell lines, in the region of the 970 cm^{-1} peak (proteins and DNA/RNA). These background contributions were captured by the PCA due to the baseline correction used, which consisted on a linear function being subtracted from the first to the last spectral point. Raman background generated by cells were used in the literature to discriminate MSCs from HDFs [31], and also tumours from non-tumours cells [166]. These different backgrounds are generated by different cell phenotypes, which have specific endogenous autofluorescence patterns [31]. Therefore, the background contribution revealed by PC1 loading might be due to phenotypic differences between the analysed cells. PC1 loading also indicates the protein peaks at 932 cm^{-1}, 999 cm^{-1} and 1028 cm^{-1} as responsible for the variation in the dataset.

The PC2 loading (Figure 3.3b) is less affected by the background and provides a better understanding of the statistical variance in the dataset. It shows that the DNA/RNA

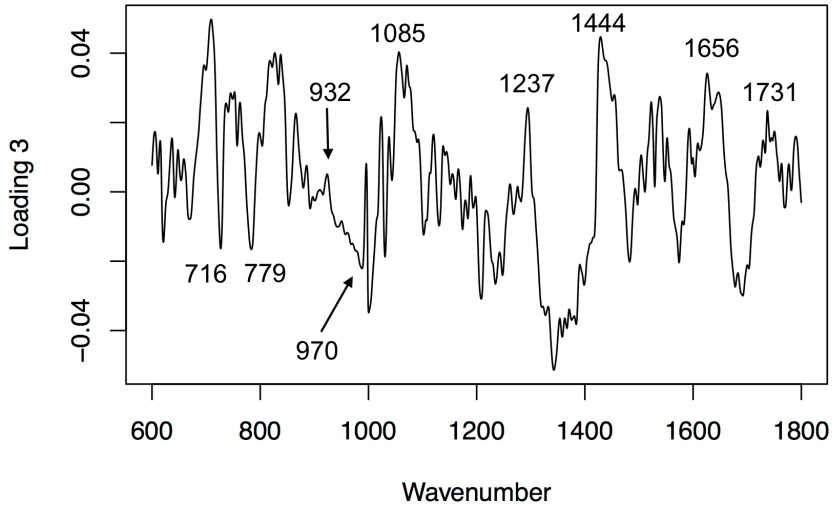


(a)



(b)

Figure 3.3: PCA of the MSC line datasets showing the loadings of (a) PC1, (b) PC2 and (c) PC3, which accounted, respectively, for 34%, 23% and 10% of the total variance. The peaks indicated in each loading were the major contributors of each PC score - *continued on the next page.*



(c)

Figure 3.3: *Continued from the previous page.* PCA of the MSC line datasets showing the loadings of (a) PC1, (b) PC2 and (c) PC3, which accounted, respectively, for 34%, 23% and 10% of the total variance. The peaks indicated in each loading were the major contributors of each PC score.

vibrations, such as 742 cm^{-1}, 779 cm^{-1}, 1085 cm^{-1}, 1237 cm^{-1}, 1260 cm^{-1} and 1574 cm^{-1}, and that the lipids and proteins peaks at 1444 cm^{-1}, 1656 cm^{-1} and 1731 cm^{-1} were the major contributors for this PC scores.

Apart from the previously peaks, PC3 loading (Figure 3.3c) reveals that the lipids vibration at 716 cm^{-1}, the DNA/RNA peaks at 779 cm^{-1} and 970 cm^{-1}, and the protein peak at 932 cm^{-1} also contributed to the variation between the MSC lines.

Although PCA was not able to discriminate the four MSC lines, it revealed that most of the variation in the dataset was represented by DNA/RNA and protein and lipid bands. From global gene expression analyses, it is known that these cell lines segregate into two distinct groups that share similar characteristics: Y101 and Y201 (01s) MSC lines versus Y102 and Y202 (02s) MSC lines [84]. Significant differences between these two groups were observed in the expression of genes involved in the cell cycle, DNA replication and other processes associated to cell replication, cell adhesion, endochondral ossification and adipogenesis. In addition, these two groups also have distinct differentiation capacities. Thus, their spectra might be reflecting these intrinsic different characteristics associated with DNA replication, cell proliferation and differentiation competence.

LDA optimises data classification by maximising the between-group variance whilst decreasing the within-group variance. Thus, LDA was applied to the dataset using the principal components obtained from the PCA. This ensures that the data is reduced and does not incorporate as variables the same number of observations [36]. However, not all PCs can be used in the LDA as this causes an overfitting of the data. Thus, as proposed by Chan *et al.* (2009), the number of PCs incorporated into the LDA should be half of the number of spectra obtained for the smallest group, *i.e.*, in the case of the MSCs this corresponds to 18 PCs accounting for 95% of the variance in the dataset [36]. The combined method of PCA-LDA has been used in the literature to demonstrate good separation and

classification of healthy and tumorogenic bone cells [161], different bone cells such as osteoblasts, alveolar bone cells and osteosarcoma [162], non-differentiated human ESC and their cardiac derivatives [36], and also between different prostatic adenocarcinoma cell lines [32].

Figure 3.4 illustrates the separation that PCA-LDA achieved between the spectra of the four air-dried MSC lines, with the group clustering easily identified. The two figures show different rotations of the same scatterplot. The PCA-LDA was cross-validated with a prediction algorithm. Two thirds of data (150 spectra obtained from the average of five spectra/cell) were used for training the algorithm, while the other third (74 spectra, also comprising the average of five spectra/cell) was used for the predictions. The spectra was randomly chosen for both training and prediction testing, with the amount of selected spectra being proportional to the size of each group.

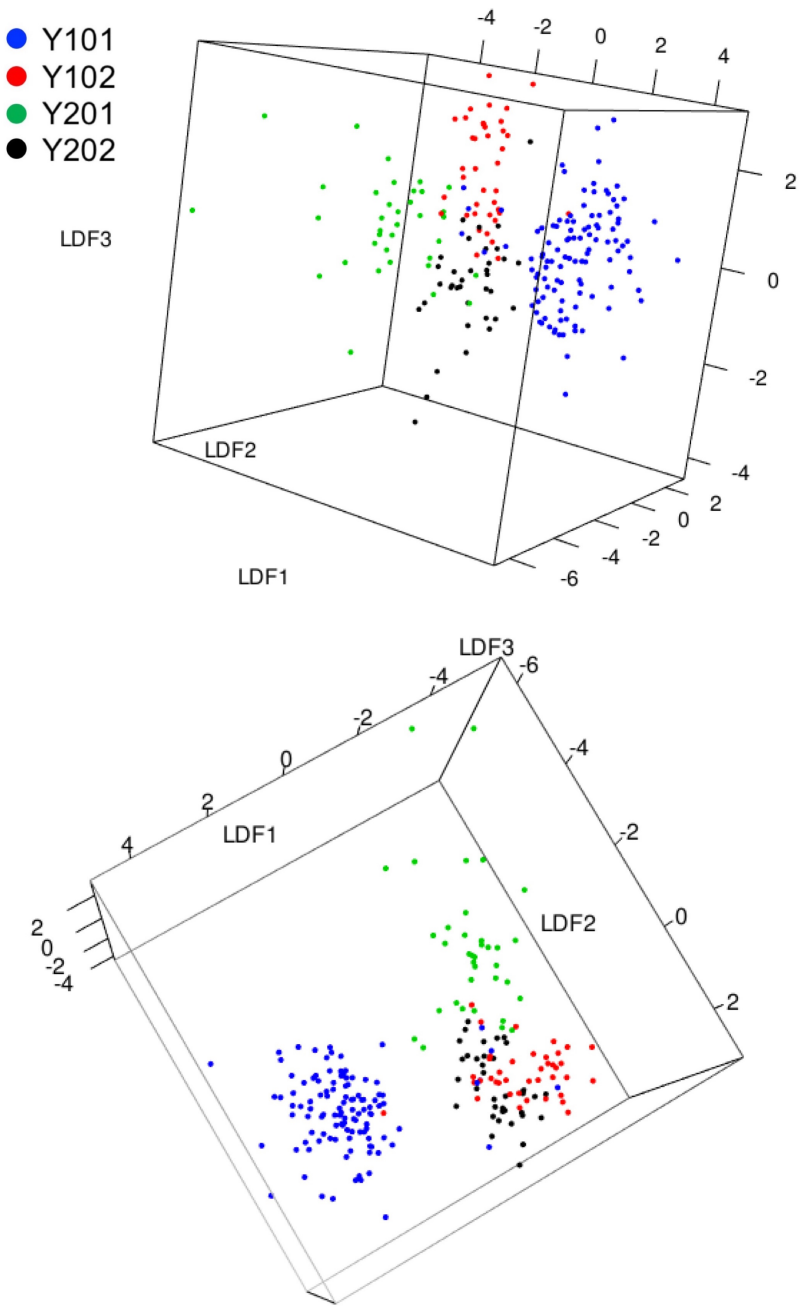


Figure 3.4: PCA-LDA three-dimensional score plots. The score plots show the discrimination of the four MSC lines achieved with PCA-LDA. The two images show different rotations for the same score plot.

The prediction results are shown in the confusion matrix presented in Table 3.2. A total of six cells were misclassified, with all of them predicted as belonging to the Y102 cell line. From these wrongly predicted cells, one was from Y101 cell line, two from

Y201 cell line and three from Y202 cell line. The percentage of wrong predictions was 8%, demonstrating this as a suitable method for their classification. If the groups of differentiation competent (Y101 and Y201) and differentiation incompetent (Y102 and Y202) cell lines are compared, 94% of the differentiation competent cells are correctly classified. This evidences the accuracy of the method in classifying the MSCs towards their differentiation capacity.

Table 3.2: Confusion matrix of the PCA-LDA classification algorithm results.

	Predicted cell line			
Cell line	Y101	Y201	Y102	Y202
Y101	39	0	1	0
Y201	0	10	2	0
Y102	0	0	10	0
Y202	0	0	3	9

The results obtained with the PCA-LDA indicate that the spectra of the four MSC lines are distinct. Although these four cell lines have not been previously fully discriminated by conventional methods, such as global gene expression [84], their Raman spectra suggests that the four of them might be different. The PCA-LDA also reveals a full separation of the Y101 cell line from the others (Figure 3.4). Interestingly, this cell line is capable of spontaneous osteogenic differentiation and shows poor adipogenic competence [84]. Furthermore, as indicated by the PCA loadings, the major contributors of the variation between spectra of the four MSC lines were the DNA/RNA, lipids and protein peaks.

3.3 PIR analyses of the air-dried MSC lines

The PCA-LDA analysis confirmed the ability of RS to discriminate the four MSC lines studied. PCA also suggested that the protein, lipids and DNA/RNA peaks contributed to

the variation between the spectra of these lines. Although these four MSC lines express the same panel of surface proteins, the RS results obtained so far indicate that they have distinct characteristics. Thus, the use of ratiometric analyses, by means of peak intensity ratios (PIRs), allows the underlying source of differences to be identified.

These PIRs were calculated from the peak intensities obtained from the fitting of the averaged spectra shown in Figure 3.1. The processes of peak fitting and PIR calculation were described in Section 2.6.1.

Discriminatory Raman markers were observed for PIRs against peaks at 932 cm^{-1}, 970 cm^{-1}, 1060 cm^{-1}, 1085 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} peaks, and are shown from Figure 3.5 to Figure 3.10. The peaks at 932 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} are assigned to proteins, while the peaks at 970 cm^{-1}, 1060 cm^{-1} and 1085 cm^{-1} are associated to DNA/RNA vibrations (Table 3.1). Peaks, such as the 932 cm^{-1}, 970 cm^{-1} and 1085 cm^{-1}, had already been associated to the PCA loadings (Section 3.2).

PIRs against the peaks at 932 cm^{-1}, 1060 cm^{-1}, 1085 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} discriminated the MSC lines into two groups, the 01s and the 02s. The PIRs against the peak at 970 cm^{-1} were found to discriminate the four cell lines. The set of markers obtained for each one of these PIRs are discussed next, initially for the PIRs obtained against peaks related to DNA/RNA vibrations, and later for the PIRs calculated against protein peaks.

At first, a reduced set of PIR markers obtained against the 1085 cm^{-1} peak is shown. These Raman markers are in agreement with the previously published result obtained for these MSCs and that comprised only one dataset for each cell line [84]. The PIRs obtained against the 1085 cm^{-1} peak separate the 01 group (Y101 and Y201) from the 02 group (Y102 and Y202), as shown in Figure 3.5. Although the group separation is maintained, higher values were obtained for the 1656/1085 PIR in comparison to the first published

dataset. This was due to a change in the fitting range for the 1656 cm^{-1} peak, thus causing a change in its subtracted baseline. For the published dataset, the peak fitting ranged from ∼1600 cm^{-1} to ∼1800 cm^{-1}, while now the band was fitted from ∼1500 cm^{-1} to ∼1800 cm^{-1}. This change was made to accommodate, in the same baseline, the small peaks neighbouring the 1656 cm^{-1} peak (Figure 2.7).

The group separation obtained by the PIRs reflects the contrasting differentiation capacity of these cell lines and can be used for their discrimination. The biggest separation between 01s and 02s is noted for the PIR of peaks assigned to proteins and lipids (1444 cm^{-1} and 1656 cm^{-1}). The PC2 and PC3 loadings also showed that these two peaks contributed to the dataset variance (Figure 3.3b and c). Thus, the 1444/1085 and 1656/1085 PIRs confirm a difference in the proteins and lipids relative to DNA/RNA between the two groups. Additionally, the 970/1085 PIR was able to discriminate the four cell lines (Figure 3.5). The variation in the relative intensity of the 970 cm^{-1} peak is discussed later when the PIRs against it are presented.

As explained before, the CD317 protein, which has anti-viral properties, is highly expressed in the Y102 and Y202 cell lines while absent for the Y101 and Y201 [84]. Thus, this population was used as control for the subpopulation discrimination. The PIRs obtained for the MSCs were compared to those obtained from a primary sorted CD317+ population, from which 100 spectra were obtained across 20 cells (5 spectra/cell). The intensity ratios obtained against the 1085 cm^{-1} peak confirmed the relatedness of this sorted population fraction (CD317+) with the 02s group, indicating that they have the same protein and lipid to DNA characteristics (Figure 3.6).

MSCs have similar morphologies and properties to non-differentiating stromal cells (fibroblasts), such as the expression of matrix proteins and the expression myofibroblastic cells markers (α-smooth muscle actin) [167]. Therefore, it is important to be able to

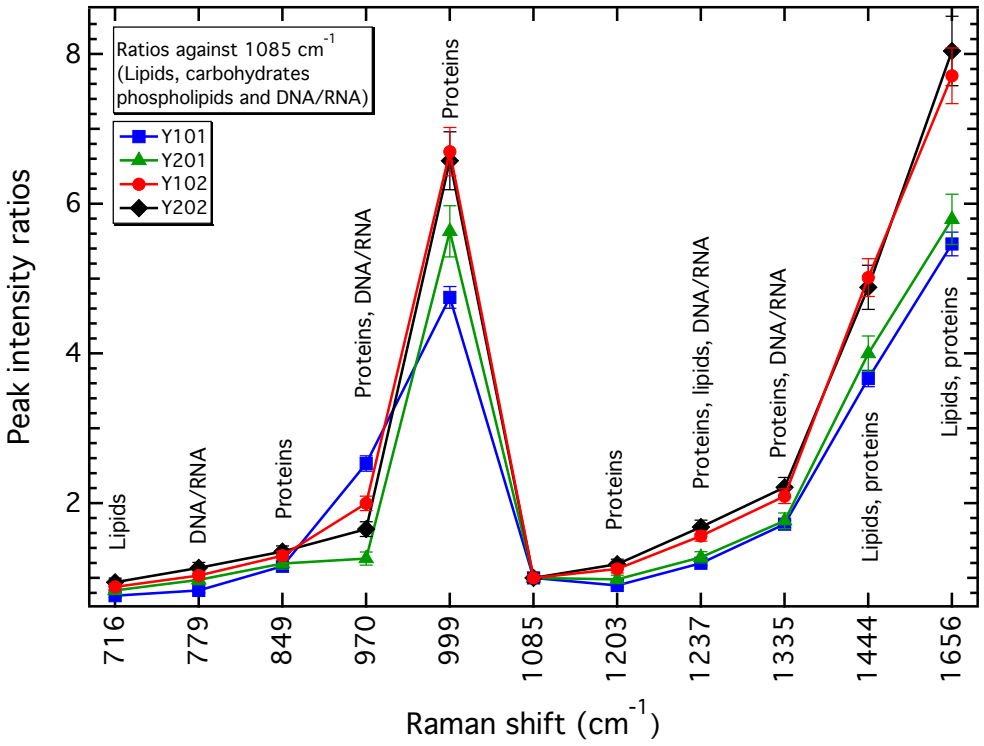


Figure 3.5: PIR comparison for the MSCs against the 1085 cm^{-1} peak. This peak is assigned to lipids, carbohydrates, phospholipids and DNA/RNA (PO_2^- DNA/RNA stretching vibration). The results show the separation of for the four MSC lines into two groups: the 01 group (Y101 and Y201) and the 02 group (Y102 and Y202). The error bars are the propagation of the fitted intensity SEs.

identify and discriminate one from the other. RS combined with PCA has already been used to discriminate HDFs from MSCs [31]. Thus, human dermal fibroblasts (HDFs) were used as non-MSC control and their PIRs markers were compared to those obtained from the MSC lines. The PIRs against the 1085 cm^{-1} peak revealed that the HDFs are very similar to the 02 group (Figure 3.7). However, two very distinct features, highlighted by the 780/1085 and 1237/1085 PIRs, discriminated the HDFs from the MSCs. Both 780 cm^{-1} and 1237 cm^{-1} peaks have in common vibrations of the DNA/RNA molecules. These results show the relatedness between 02s and HDFs, both differentiating incompetent, whilst still providing Raman markers for the discrimination of the MSCs from the HDFs.

The discrimination of cells by means of protein-to-DNA intensity ratios was used by

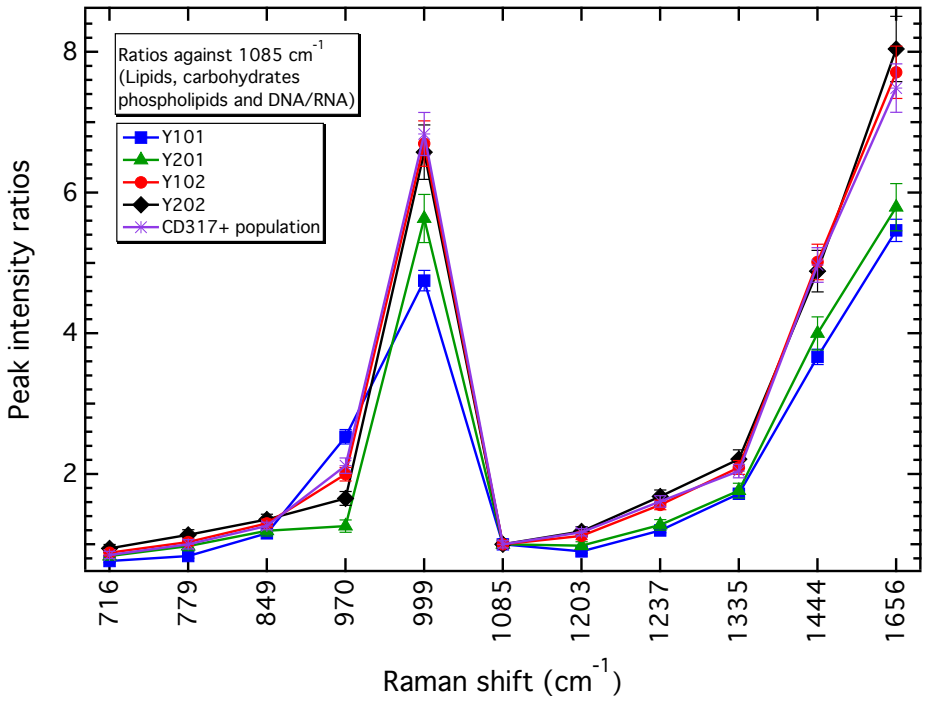


Figure 3.6: PIR comparison against the 1085 cm^{-1} peak between the MSC lines and the CD317+ primary MSC fraction. The error bars are the propagation of the fitted intensity SEs.

Schulze *et al.* (2010), for discriminating between differentiated and non-differentiated human embryonic stem cells (ESCs) [33]. Their discriminatory intensity ratios were calculated between the intensity of the protein-related peak at 757 cm^{-1} and the nucleic acid band at 784 cm^{-1} [33]. Other Raman markers used to discriminate between differentiated and non-differentiated live murine ESC, were obtained by Notingher *et al.* (2004), who noted that the relative intensity of peaks assigned to RNA (813 cm^{-1}) and DNA (786 cm^{-1}) decreased 75% and 50% respectively, over 16 days of differentiation. Similar relative decreases in the RNA and DNA bands that were also noted in a live-cell study of human ESC during cardiomyocyte differentiation provided discrimination between differentiated and non-differentiated cells [36]. Thus, aiming to identify if there were any other markers with the ability to separate the two groups, an expanded set of PIRs against the 1085 cm^{-1} peak was calculated. These PIRs were obtained from intensity ratios containing all peaks

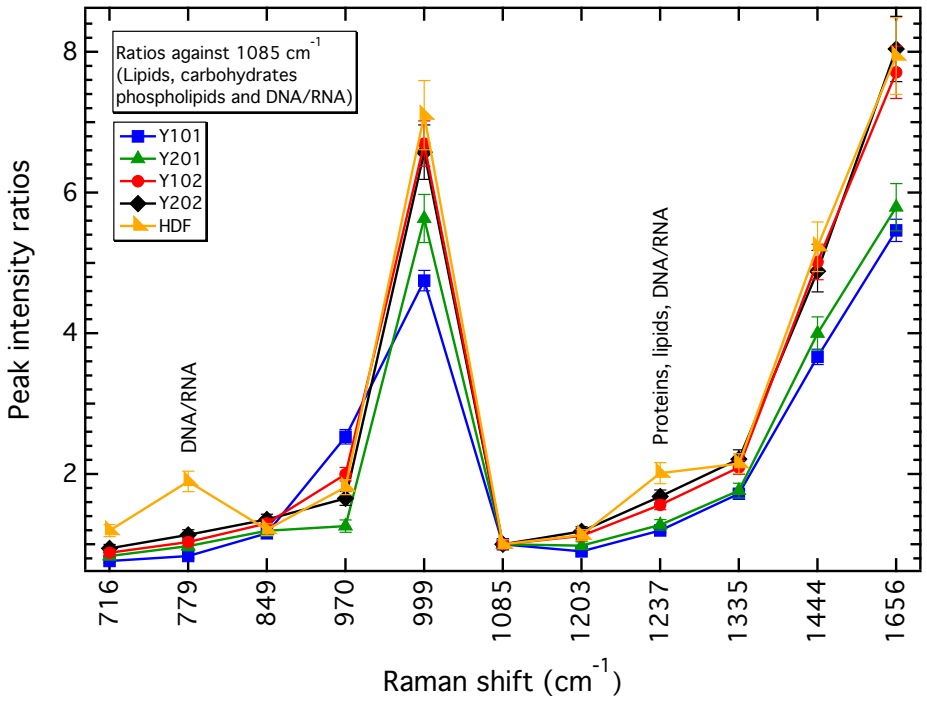
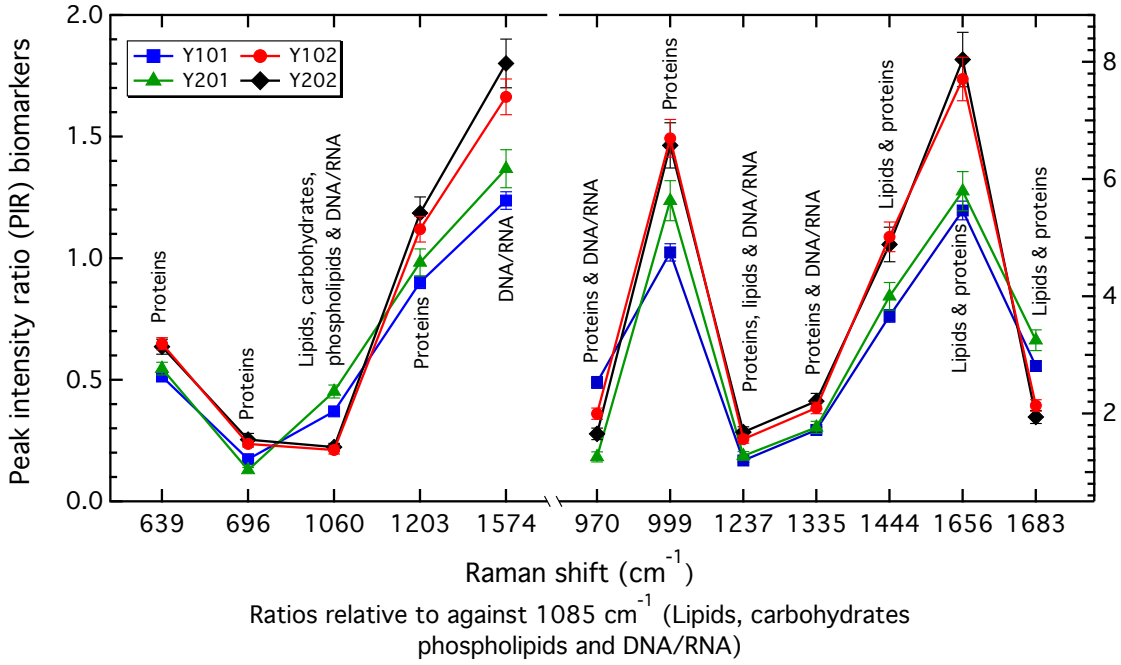


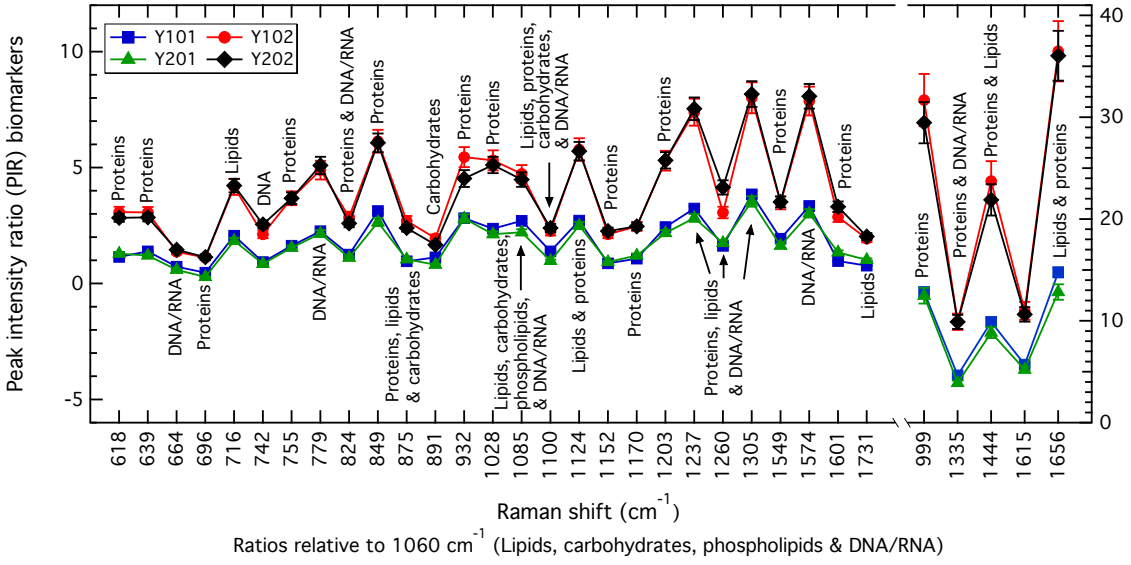
Figure 3.7: PIR comparison against the 1085 cm^{-1} peak between the MSC lines and the HDFs. The error bars are the propagation of the fitted intensity SEs.

and only the discriminatory Raman markers are shown in Figure 3.8a. The expanded set confirms that the separation of the 01s and 02s is pertained to a difference in proteins and lipids relative to DNA/RNA.

Similarly, the PIRs obtained against the 1060 cm^{-1} peak are also presented as an expanded set of markers (Figure 3.8b). These Raman markers made even more evident the separation between the 01 group and the 02 group. However, the 1060 cm^{-1} peak has a very low relative intensity, which results in the much higher values obtained for the markers. Although it gives similar group separation, this is a difficult peak to be fitted as it is convolved with the 1085 cm^{-1} peak. The 1060 cm^{-1} peak is also assigned to lipids, carbohydrates, phospholipids and DNA/RNA (Table 3.1). For the band ranging 1060–1100 cm^{-1} the DNA/RNA vibration observed is the symmetric phosphate backbone stretch [32, 40, 41, 43, 66, 141, 154, 164]. Thus, the discrimination of 01 and 02 groups, obtained by



(a)



(b)

Figure 3.8: Expanded set of discriminatory PIRs for the MSCs against the (a) 1085 cm^{-1} and (b) 1060 cm^{-1} peaks, showing the separation of the lines into two groups: 01 group (Y101 and Y201) and 02 group (Y102 and Y202). The error bars are the propagation of the fitted intensity SEs.

these ratios, reinforces the differences in lipids and proteins relative to DNA/RNA.

Such group separation against DNA/RNA peaks may be the effect of a difference in

the nucleus size of the 01s and 02s MSCs, causing a variation in the density distribution of such molecules in the nucleus. Different molecular densities in cell lines with different nuclear size have been noticed with RS in a study conducted by Chan *et. al* (2006) that compared normal human lymphocytes with transformed lymphocyte cell lines [168]. The larger nucleus of the transformed cells was found to affect the compactness of chromatin and the concentration of DNA in the probed volume, concluding that the DNA density decreased for larger nuclei, while the overall DNA content increased [168]. Although not significantly different, the Y102 and Y202 cells have slightly greater area and cell perimeter than the Y101 and Y201 (Appendix E, Figure E.2), indicating the likelihood of different molecular densities in the probed volume. Therefore, since the PIRs for the 02s group calculated against DNA/RNA (1060 cm^{-1} and 1085 cm^{-1}) peaks had higher values than those obtained from the 01s group, it could be assumed that they have smaller DNA/RNA densities within the probed volume, in agreement with their slightly larger nuclei sizes.

Discriminatory PIRs of the four MSC lines were identified against the 970 cm^{-1} peak (Figure 3.9), which is assigned to vibrations of proteins and DNA/RNA (Table 3.1). In the image, the 875/970 PIR seems overlapped between the cell lines, as the choice of scale is unable to show it to be a distinct marker. The 970 cm^{-1} peak and other protein and lipid peaks were revealed as major contributors of the PC3 loading of the principal component analysis (Figure 3.3c).

As the PIRs against the 970 cm^{-1} peak were fully discriminatory these markers were also calculated for the CD317+ primary fraction and HDFs to further compare against the MSCs. While previous PIRs against the 1085 cm^{-1} peak mainly showed a relatedness of the CD317+ population to the 02 group (Figure 3.6), the PIRs against the 970 cm^{-1} peak present a match between the CD317+ and Y102 cell line (Figure 3.10a). These markers also show that the HDF results oscillate between the Y201 and Y102 PIRs (Figure 3.10b).

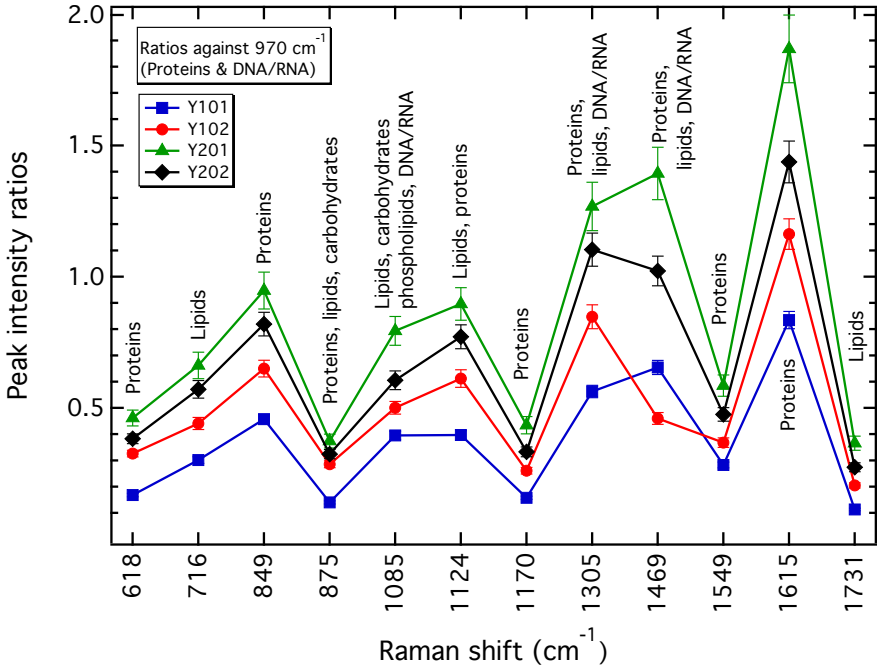
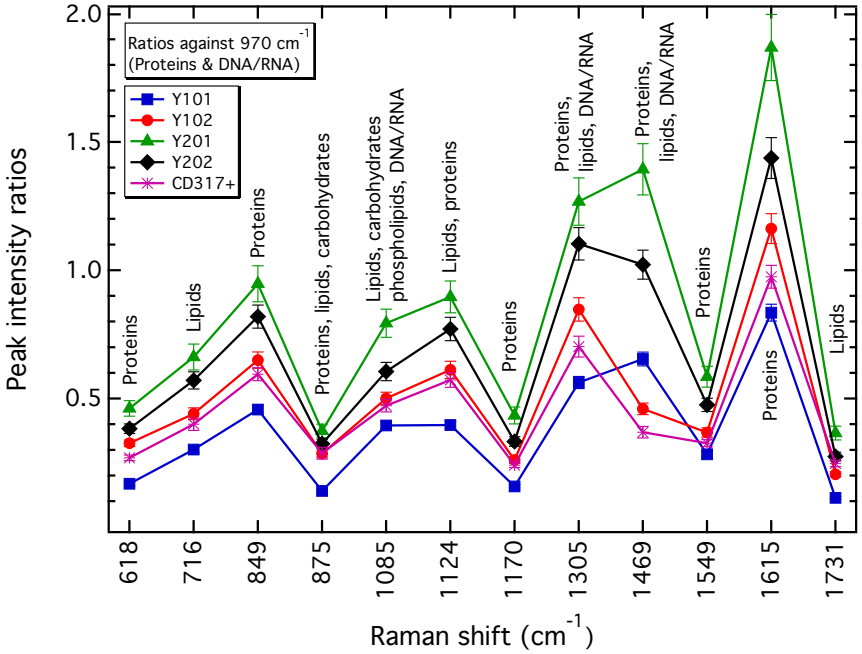


Figure 3.9: Discriminatory PIRs for the MSCs against the 970 cm^{-1} peak (assigned to proteins and DNA/RNA vibrations). The graph shows only PIRs that fully discriminate the four undifferentiated MSC lines: Y101, Y201, Y102 and Y202. The error bars are the propagation of the fitted intensity SEs.

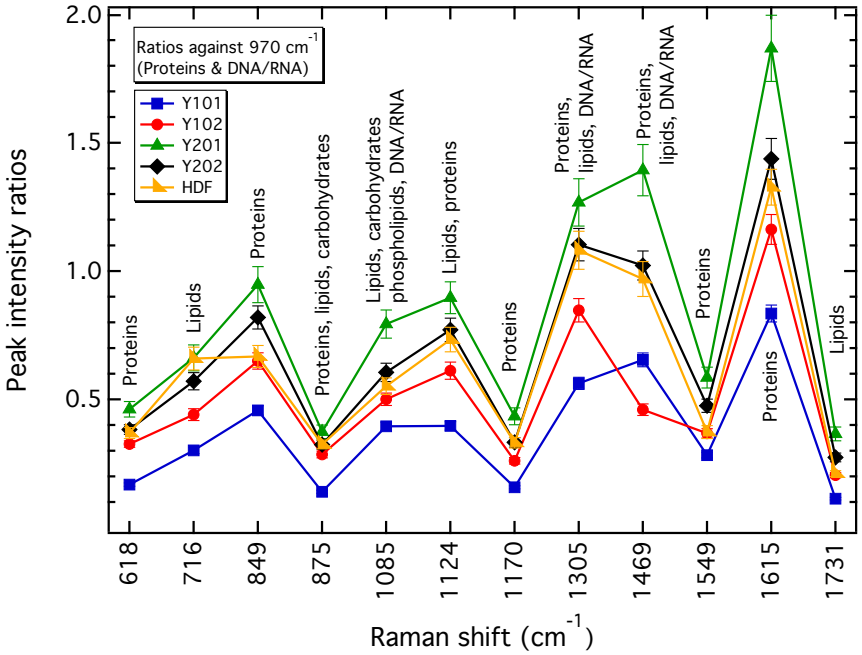
They exhibit evidence of close similarities between the PIRs of the HDFs with the Y202 line that were not evident from the PIRs against the 1085 cm^{-1} peak .

Discriminatory PIRs against the 970 cm^{-1} peak confirm that the greatest differences between the four cell lines is found for ratios of proteins and lipids against DNA/RNA vibrations. Other PIRs of peaks assigned only to DNA/RNA nucleic acids, such as the 664 cm^{-1}, 742 cm^{-1}, 779 cm^{-1} or 1574 cm^{-1} peaks, against the 970 cm^{-1} peak did not discriminate between the four cell lines. It is also noticed, from Figure 3.9, that a variation of the 970 cm^{-1} peak intensity relative to other cellular components, across the cell lines, allowed the discrimination of the four MSC lines. Thus, it is important to investigate the possible causes of such variation.

Although assigned to vibrations of proteins and DNA/RNA, the 970 cm^{-1} band is very closely positioned to the hydroxyapatite (HAP) band at 960 cm^{-1} (957–970 cm^{-1}), which



(a)



(b)

Figure 3.10: PIR comparison against the 970 cm^{-1} peak (assigned to proteins and DNA/RNA vibrations) between MSCs and (b) the primary sorted CD317+, and (b) the HDFs. The error bars are the propagation of the fitted intensity SEs.

is an indication of osteogenic differentiation [35, 127, 131]. The mineralisation process during cell differentiation can be assessed by detecting hydroxyapatite or its precursors, such as amorphous calcium carbonate (ACC), octacalcium phosphate (OCP), β-tricalcium phosphate (β-TCP) and dicalcium phosphate dehydrate (DCPD) [35, 131, 134]. However, the spectra obtained from the MSCs reflect their undifferentiated state. Thus, although the Y101 cell line has the capability of spontaneously differentiate into osteoblasts, HAP or its precursors were not observed for this cell line.

The 970 cm^{-1} peak is assigned to phosphorylated proteins and cellular nucleic acids by Movasaghi *et. al* (2007) [154], Meurens *et. al* (1996) [157] and Wong *et. al* (1991) [158]. Both works of Meurens *et. al* (1996) and Wong *et. al* (1991) credited the appearance of this peak to cancerous cells and malignant tissues. However, no evidence of tumourigenicity in the MSC lines was found using *in vitro* colony-transformation assays for non-adherent growth [84]. Therefore, the relative variations in the intensity observed for this peak cannot be credited to the presence of cancerous cells.

The appearance, in all spectra, of the 1731 cm^{-1} peak, which becomes apparent after a width shrinkage of the 1656 cm^{-1} band, indicates the presence of carbonyl (C=O) stretching vibrations associated with oxidative damage occurring during cell death [43,44]. A study performed on dying HeLa cells noticed an increase in the relative intensity of peaks assigned to DNA and RNA, which was then associated to DNA condensation [43]. The 970 cm^{-1} peak is also assigned to a C-O-P phosphodiester residue by Matthaus *et. al* (2008) [40] and to vibrations of the deoxyribose in B-DNA by Prescott *et. al* (1984) [169]. Thus, as a consequence of the cell death process, the variations in the relative intensity of this peak could be related to the DNA condensation. However, peaks assigned only to nucleic acid vibrations showed constant relative intensities and did not provide any cell line discrimination when PIRs were calculated against them.

It is important to highlight that the results presented here correspond to experiments performed on air-dried samples. Along with desiccation, air-drying is considered one of the most effective methods of preserving cells for Raman experiments, as cell responses to fixatives are dependent on their internal structure and membrane components [45]. However, dehydration causes loss of structural integrity and leads to cell death [170]. Although during the desiccation process there is a shrinkage of capsular layers, a crowding of macromolecules, changes in volume of cell compartments, changes in biophysical properties and damage to external layers, some human cells might still tolerate desiccation for short periods [170,171]. An increase in the phosphorylation of proteins was noted as early as 30 min in human embryonic kidney cell lines, showing that human cells react rapidly to desiccation by mitogen-activated protein kinase (MAPK) activation [172]. Thus, since the 970 cm^{-1} peak is also assigned to vibrations of phosphate monoester groups of phosphorylated proteins and DNA/RNA nucleic acids [154, 157, 158], this might indicate that the variations in its relative intensity could be related to the air-drying process. Protein phosphorylation also has a regulatory role in cell death mechanisms, for which two different cell pathways may occur: apoptosis or necrosis [173]. The apoptosis pathways are very complicated and involve a large number of proteins, however they also imply an increase in the phosphorylation of proteins [173].

Other PIRs against the 932 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} protein peaks, reveal some more details about the possible changes caused by the air-drying process. Although they all discriminate the MSCs into 01s and 02s (Figures 3.11 and 3.12), discriminatory PIRs of the four cell lines are observed for the 970/1549 and 970/1615 PIRs (Figure 3.12a and b). Additionally, the 970/932 PIR discriminates the two 01 cell lines, whilst grouping the 02 together (Figure 3.11). The relative intensity of the 970 cm^{-1} peak against the 932 cm^{-1} peak is similar for the 02 MSCs, higher for the Y101 cell line and lower for the

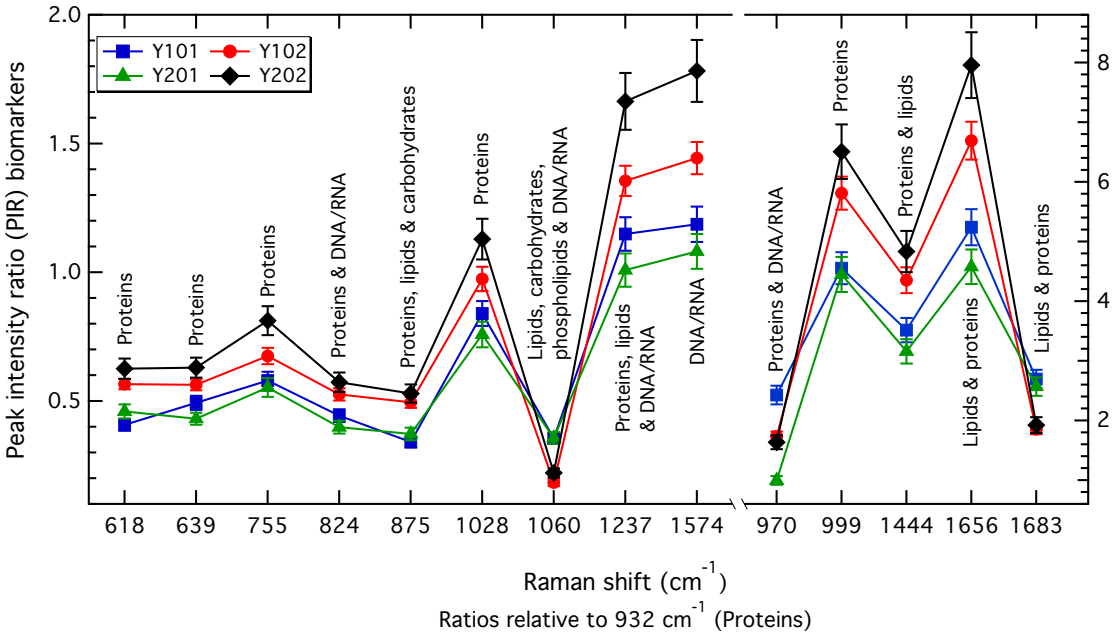
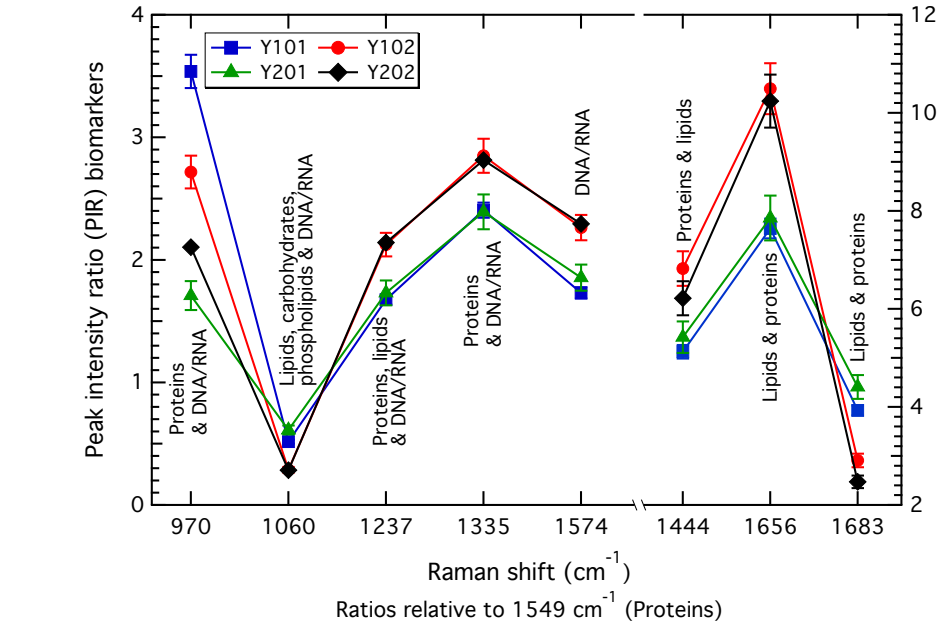


Figure 3.11: Expanded set of MSCs discriminatory PIRs against the 932 cm^{-1} peak, showing the separation of the cell lines into two groups: 01 group (Y101 and Y201) and 02 group (Y102 and Y202). The error bars are the propagation of the fitted intensity SEs.

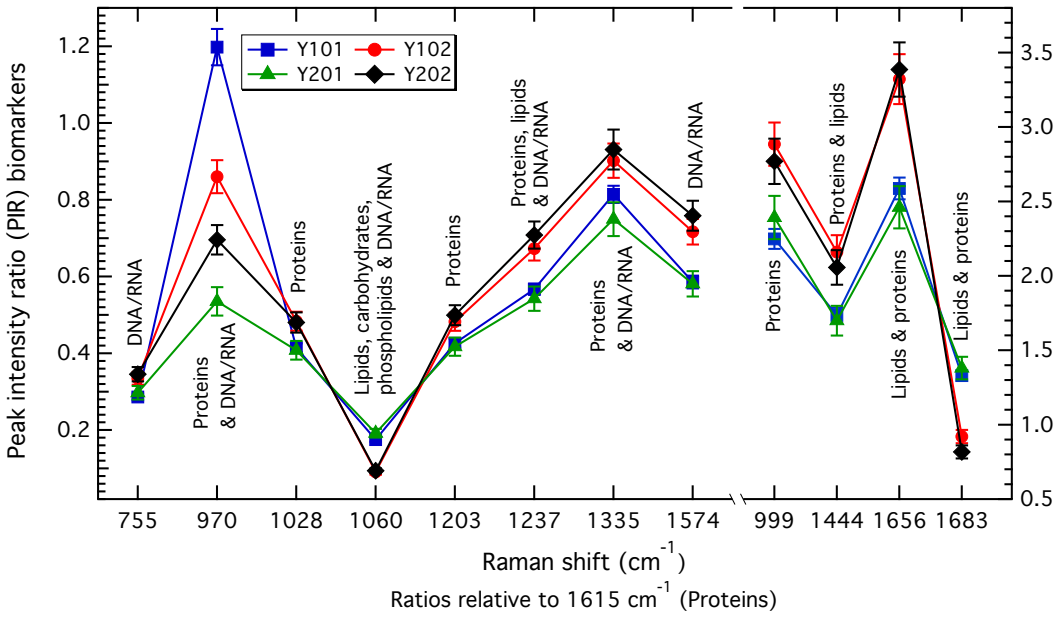
Y201 cell line. Both peaks had been previously indicated by the PCA loading 3 (Figure 3.3c). Thus, the variations in the relative intensity of the 970 cm^{-1} peak could be the reflex of different protein phosphorylation between the MSC lines as result of air-drying or cell death.

The 1549 cm^{-1} and 1615 cm^{-1} peaks correspond to tryptophan and to tyrosine vibrations (Table 3.1). Both tyrosine and tryptophan are common proteins involved in intrinsic and extrinsic pathways of apoptosis [174], where tyrosine phosphorylation has been assumed as a protection mechanism against apoptosis [173]. Thus, the MSC group separation provided by the PIRs obtained against the 1549 cm^{-1} and 1615 cm^{-1} peaks might also indicate that 01s and 02s respond differently to the air-drying process, a characteristic that still made possible their discrimination.

In summary, from all PIR makers obtained, it was noticed that those calculated against protein (932 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1}) and DNA/RNA (1060 cm^{-1} and 1085



(a)



(b)

Figure 3.12: Expanded set of MSCs discriminatory PIRs against the (a) 1549 cm^{-1} and the (b) 1615 cm^{-1} peaks, showing the separation of the cell lines into two groups: 01 group (Y101 and Y201) and 02 group (Y102 and Y202). The error bars are the propagation of the fitted intensity SEs.

cm^{-1}) vibrations allowed the discrimination between differentiation competent and incompetent MSC lines. One of the aims of this chapter was to investigate the MSCs prior being induced into differentiation, thus verifying if RS would be able to discriminate them. The results presented so far attest that RS depicted differences between these cell lines can be used for their discrimination.

The MSC heterogeneity is one of the major drawbacks in their application into regenerative medicine. So far the discrimination of differentiation competent cells and their progenitors lacks of an unique marker and is done only by a panel of surface expressed proteins [96] that may still not result in homogeneous populations [84, 120, 121]. Thus, from the results obtained, RS can be used to assess MSC heterogeneity, identifying subpopulations and providing markers for the selection of cells prone to undergo differentiation.

However, the results of this chapter were drawn from experiments performed on air-dried cells. Thus, it is important to verify the consistency of the PIR markers obtained across different experimental repeats and between a single-cell and the population average. These investigations are discussed in the next sections of this chapter. Nonetheless, in order to validate that the variations, across the cell lines, for the relative intensities of the 970 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} peaks are related to the air-drying process, it is necessary to conduct live cell analyses of the MSC lines. These live-cell analyses are discussed in Chapter 4.

3.4 Population heterogeneity

3.4.1 Convergence tests

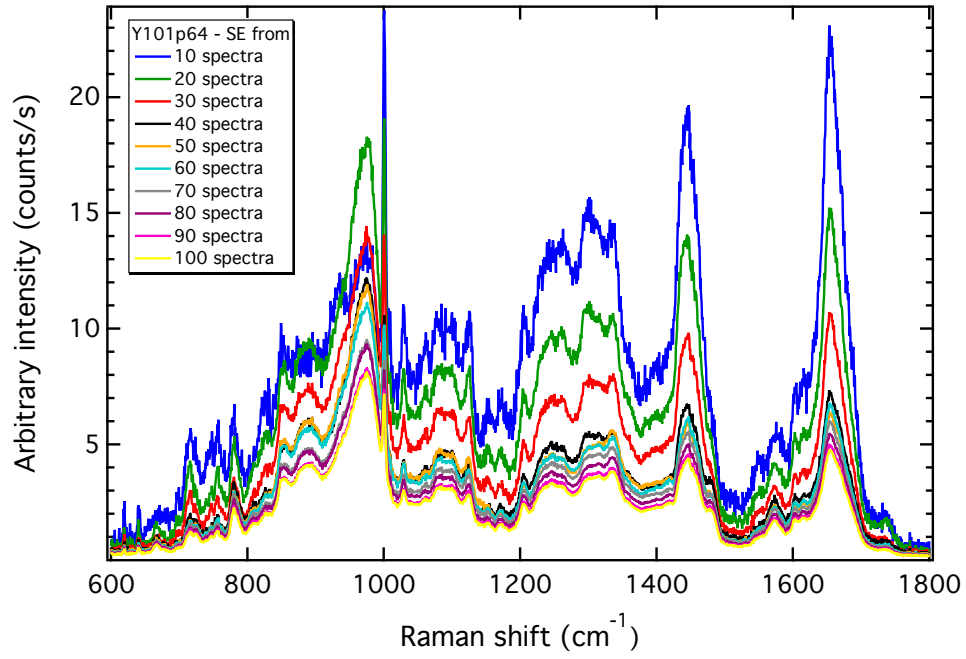
One method to evaluate the heterogeneity of the cell populations included in this study, is the examination of the %SE associated to PIRs, the SE and 2x SD over the full av-

eraged spectrum against the number of collected spectra. These tests are here termed 'convergence tests' and were described in Section 2.6.2. These convergence tests assess the quantity of spectra needed for each sample, so that the lowest SE of the peak intensities is achieved for that dataset. The convergence tests ensure that the statistics are converged and that the averaged spectra represents the population average.

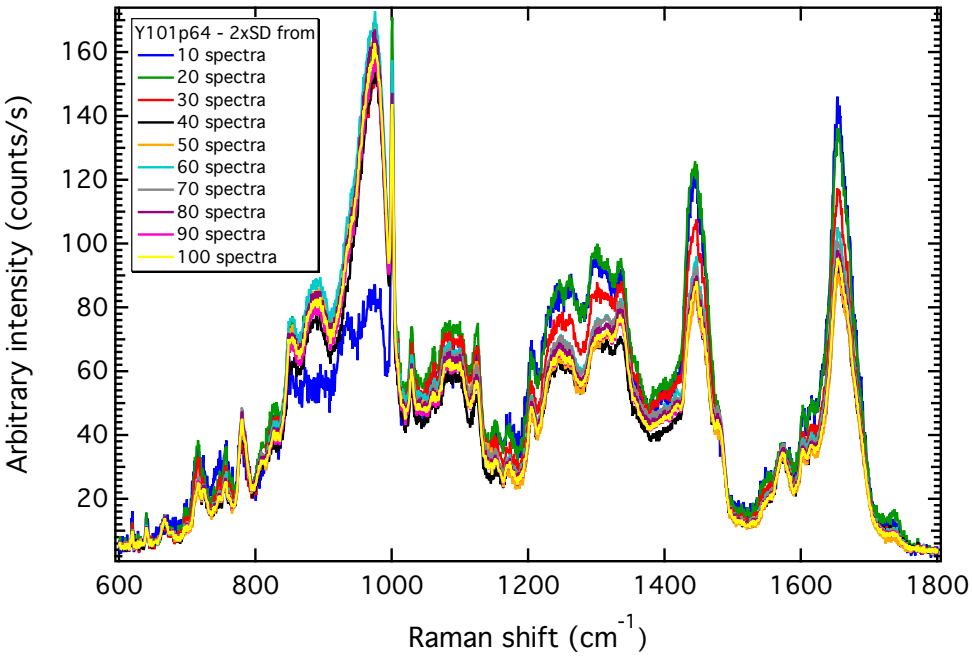
The SE and 2x SD graphs obtained for each experiment performed for the air-dried MSCs are displayed in Appendix B, Sections B.2 and B.3. Examples of these graphs are shown in Figure 3.13. These graphs show how the SE and 2x SD change with the increasing number of spectra in the spectral average. The SE decreases as the number of averaged spectra increases, converging to a minimum error value (Figure 3.13a).

On the other hand, the 2x SD curves oscillate between a mid point, with its amplitude decreasing as the number of spectra increased (Figure Figure 3.13b). The 2x SD graphs show, in general, that most of the dispersion in the peaks relative intensities are observed for the 970 cm^{-1} (proteins and DNA/RNA), 1085 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA), 1237 cm^{-1} (proteins, lipids and DNA/RNA), 1305 cm^{-1} (proteins, lipids and DNA/RNA), 1335 cm^{-1} (proteins and DNA/RNA), 1444 cm^{-1} (proteins and lipids) and 1656 cm^{-1} (proteins and lipids) bands. Apart from the latter two bands, all other bands comprise vibrations of DNA and RNA molecules.

The third test used was the %SE convergence associated with the PIRs. Two ratios were considered for this test: the 970/1085 PIR, which discriminated the four MSC lines, and the 1656/1085 PIR, which showed the largest separation between groups of differentiation competent (Y101 and Y201) and incompetent (Y102 and Y202) cell lines. For both PIRs the %SE was calculated and plotted against the number of averaged spectra. The graphs obtained for each experiment are presented in Appendix B, Section B.1. An example of this %SE convergence test is shown in Figure 3.14. Each graph was fitted with



(a)



(b)

Figure 3.13: Example of the (a) SE and (b) 2x SD convergence tests. These examples show the convergence of the SE and 2x SD, associated to the peak intensities, with the increasing number of spectra in the spectral average.

an exponential curve (Equation A.2.2 from Appendix A), from which the decay constant (τ) and the converged %SE value (y_0) were used to measure the heterogeneity associated with each population.

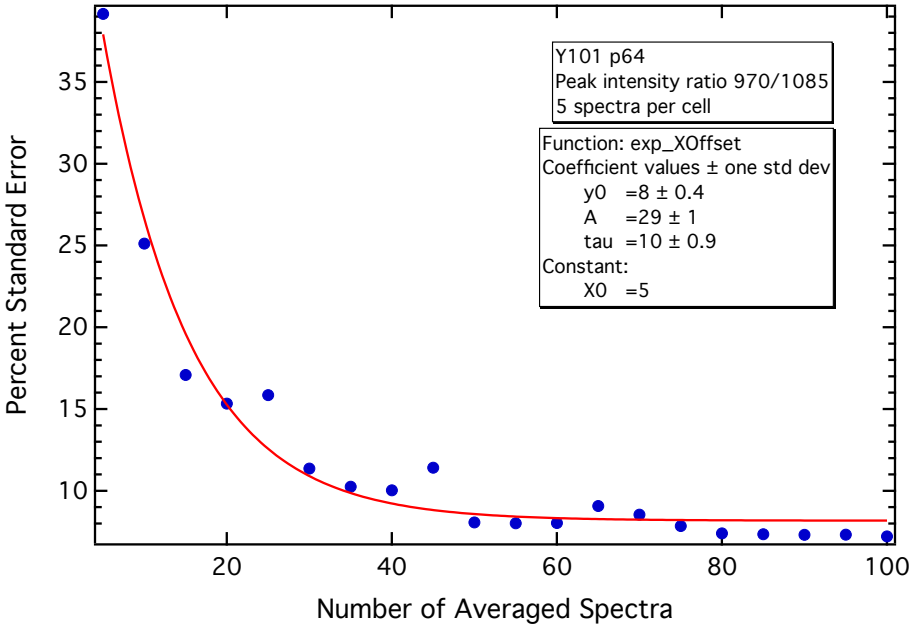


Figure 3.14: Example of %SE convergence test. It shows the decay of the %SE associated with the 970/1085 PIR as the number of averaged spectra increases. The red line represents the exponential fitted function.

Table 3.3 presents the average PIR decay constant of cell populations and the minimum %SE achieved. A smaller τ is related to a more rapid decay of the exponential curve, meaning that a sample converges to a minimum value of the standard error with a smaller quantity of data across the population. In other words, the decay constant may be used to measure the heterogeneity of the population, as a smaller constant would compare to a more homogeneous sample.

From the decay constant obtained for the 970/1085 PIR (protein and DNA/DNA ratio) it can be observed that all MSC lines have decay rates smaller than the HDF and K72 populations, indicating that they are more homogeneous than the others. Amongst the four MSC lines, Y201 is the most homogeneous of them with a mean decay rate of

Table 3.3: Average of the decay constants and converged %SE obtained for 970/1085 and 1656/1085 PIRs for the different cell populations analysed, along with the number of experiments performed. A smaller decay constant is related to a rapid decay of the %SE. For the MSC lines (Y101, Y201, Y102 and Y202) the maximum and minimum decay constants obtained are shown in the brackets. For the other cell types, since the experiments were performed once, the uncertainty corresponds to the standard deviation of the exponential fitting.

		PIR 970/1085		PIR 1656/1085	
Sample	Exp. repeats	Averaged decay constant (τ)	Converged %SE	Average decay constant (τ)	Converged %SE
Y101	6	10.0 (6.2–14.0)	7.0 (4.5–8.6)	7.8 (4.4–11.9)	7.3 (6.6–8.6)
Y201	2	8.5 (5.7–11.6)	9.5 (7.6–11.6)	12.5 (9.1–16)	8.0 (5.8–10.6)
Y102	2	9.0 (5.5–12.0)	8.5 (7.3–9.5)	8.5 (7.2–9.9)	9.0 (7.4–10.5)
Y202	2	11.0 (10.0–12.7)	8.5 (5.8–12)	6.5 (5.4–8)	10.5 (7.5–13.8)
CD317+	1	7 ± 1	7.0 ± 0.6	7 ± 1	7.0 ± 0.5
HDF	1	14 ± 1	6.0 ± 0.3	15 ± 1	7.0 ± 0.4
K72	1	17 ± 2	5.0 ± 0.3	15 ± 3	4.0 ± 0.7

8.5, ranging from 5.7 to 11.6 for the two repeat experiments performed. In comparison, the decay constant for Y202 shows it is the most heterogeneous of them, with a mean decay rate of 11 and ranging from 10 to 12.7 for the two experimental repeats. The CD317+ sorted primary fraction is the most homogeneous with a decay constant of 7 ± 1, however only one experiment was performed for this population. With decay constants of 14 ± 1 and 17 ± 1, the HDFs and primary K72, respectively, are depicted as the most heterogeneous of all samples. The average converged %SE was below 10% for all samples.

Similar results were obtained for the 1656/1085 PIR (proteins and lipids/DNA ratio). Its decay constant is also smaller for the MSC lines than for the primary K72 and HDF cells. The CD317+ sorted primary presented a decay constant in the same range observed for the MSCs. However, when only the MSC lines are considered, the Y201 is indicated as the most heterogeneous sample, whilst Y202 as the most homogeneous. The converged %SE was again below 10% for all samples, apart from Y202, which presented a value of

10.5% for the 1656/1085 PIR. Thus, although this cell line had the smallest decay rate, which can indicate its homogeneity, the SE converged to higher values than the other lines.

The studied MSCs are clonal lines derived from single-cell colonies obtained from a single donor and selected for their strong clonal and stable growth characteristics [84]. Thus, it was expected that they would be more homogeneous than the primaries. However, as seen for the minimum and maximum decay constant and converged %SE values, populations with different heterogeneous characteristics were analysed. Clonal heterogeneity in the differentiation potential had already been noticed on immortalised MSCs, suggesting that MSCs had subgroups of cells [81]. However, the heterogeneity observed for the same MSC subpopulations is consistent with differences in the expression levels of cell-cycle and extracellular matrix genes observed for single-cells within the same colony, which then displayed different potential for differentiation and proliferation [120].

Bone marrow cells form colonies of adherent cells with fibroblast-like characteristics [93]. These primary populations contain cells that have progenitor properties and multi-potent differentiation capacity (the MSCs), and others unable to differentiate, which are called fibroblasts [80, 93]. Thus, the higher values obtained for the decay constants associated to these two populations (K72 and HDF) reflect their heterogeneity. The primary sorted CD317+ fraction represents ∼1%–3% of a primary MSC population [84], thus the lower decay constant achieved for the CD317+ agrees that they had been sorted. However, although these results agree with literature observations about these cells, they constituted only one experimental repeat. Thus, since a broader range was observed for the MSC's decay constant, the values obtained for the HDFs, K72 and CD317+ could be representing minima and maxima observations for these cells if they were repeated.

The convergence test results show that the uncertainties had converged and that the averaged spectra is a statistically converged representation of the population averages used

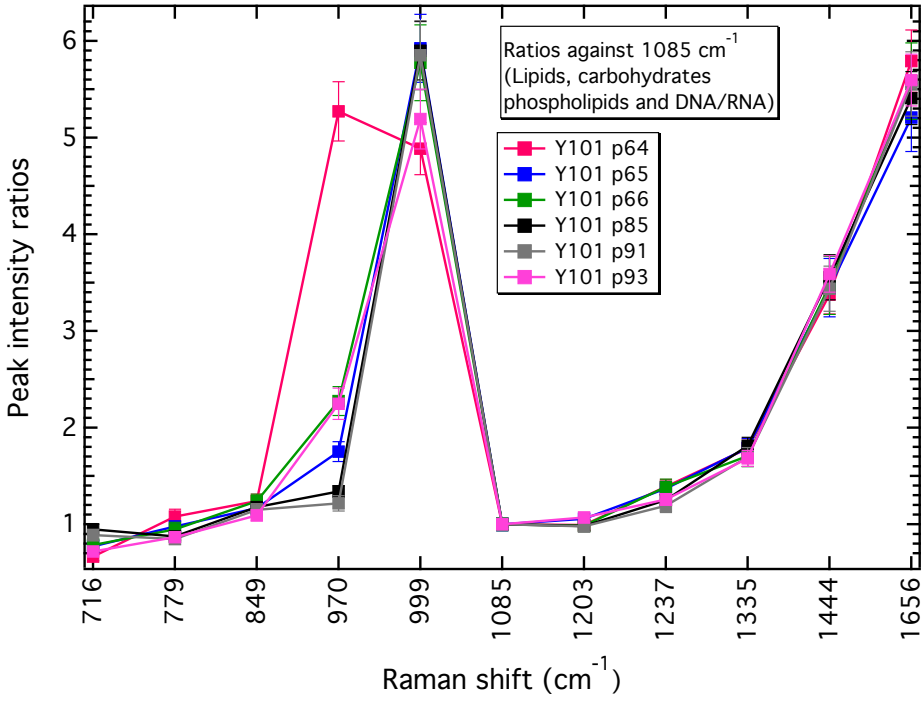
for the PIR analyses. However, to be used as a method for assessing the heterogeneity degree of a population, many more experimental repeats would be needed.

3.4.2 PIR analyses and population heterogeneity

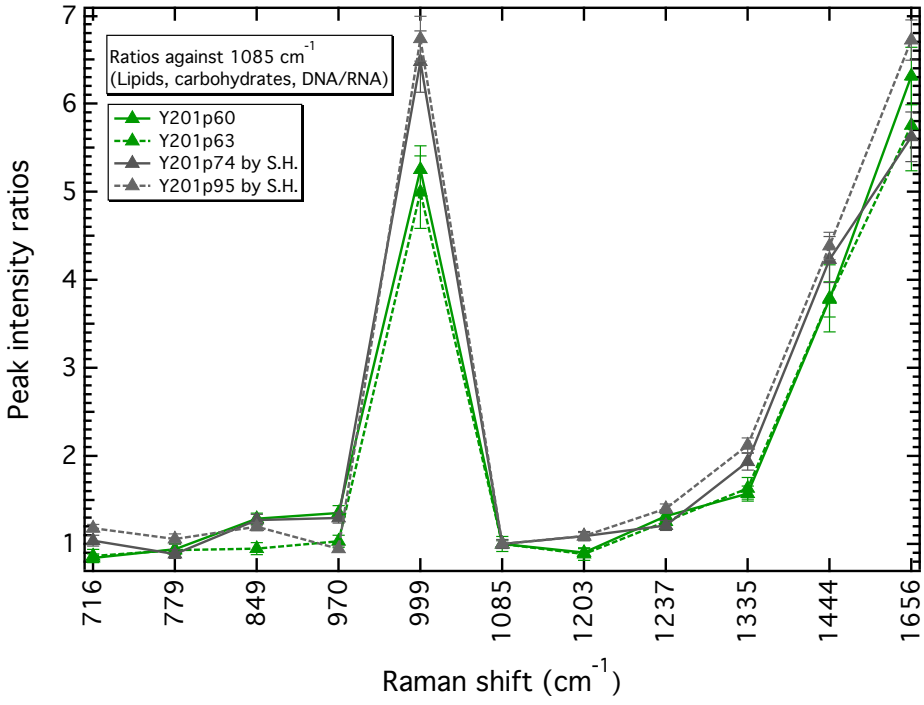
The proposed convergence tests indicated a degree for the heterogeneity associated with each cell line or cell type. This section explores how the heterogeneity affects the ratiometric analysis by comparing the different experiment repeats through the PIRs against the 1085 cm^{-1} peak (lipids, carbohydrates, phospholipids and DNA/RNA), which provided the discrimination of the 01 and 02 groups, or in other words, between the differentiation competent and incompetent cell lines.

Figures 3.15 show these PIR comparisons for Y101 and Y201 cell lines, while Figure 3.16 presents the same comparisons for the Y102 and Y202 cell lines. The experiments were repeated six times for the Y101 cell line. The four experimental repeats shown for Y201 cell line include two datasets (Y201p74 and Y201p95) of experiments and data analyses performed independently by Sam Hollings. For the other cell lines, Y102 and Y202, the experiments were repeated twice.

The 970 cm^{-1} peak (proteins and DNA/RNA) is one of the most affected spectral bands throughout the experimental repeats performed for the MSC lines (Figures 3.15 and 3.16). The 970/1085 PIR varied mostly for the Y101 cell line, with small variations observed for the other cell lines. Although the Y101 cell line can spontaneously start osteogenic differentiation, as explained in the previous section, no evidence for the hydroxyapatite or its precursors was found. Other minor differences are observed for the 999/1085 and 1656/1085 PIRs, but mainly for Y101 and Y201 cell lines (Figures 3.15). The 999 cm^{-1} peak is assigned to the phenylalanine, while the 1656 cm^{-1} peak is related to the C=C stretch in proteins and lipids (Table 3.1).

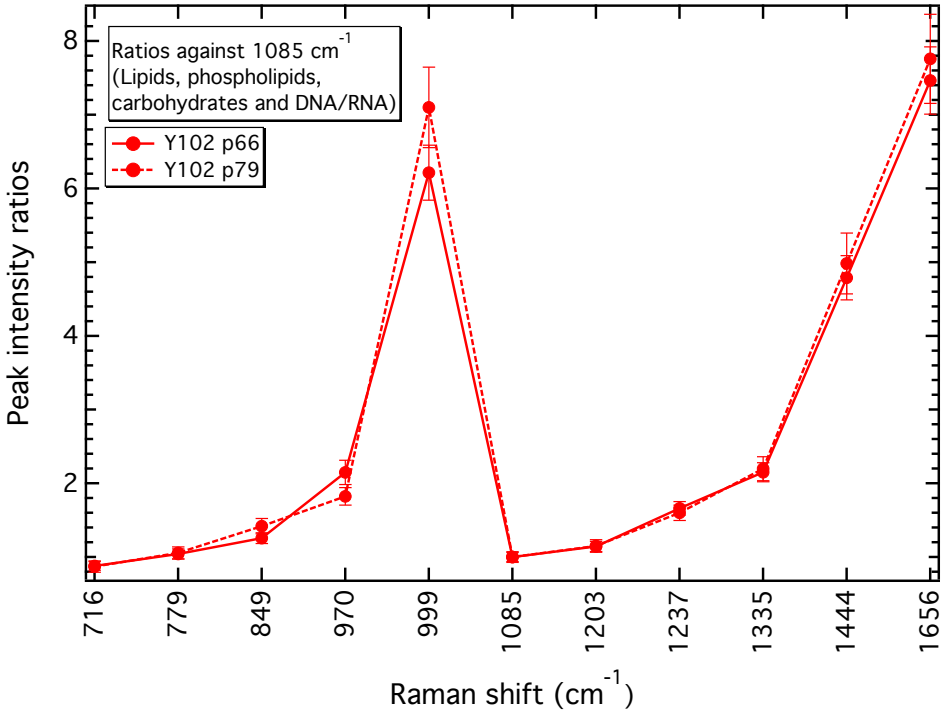


(a)

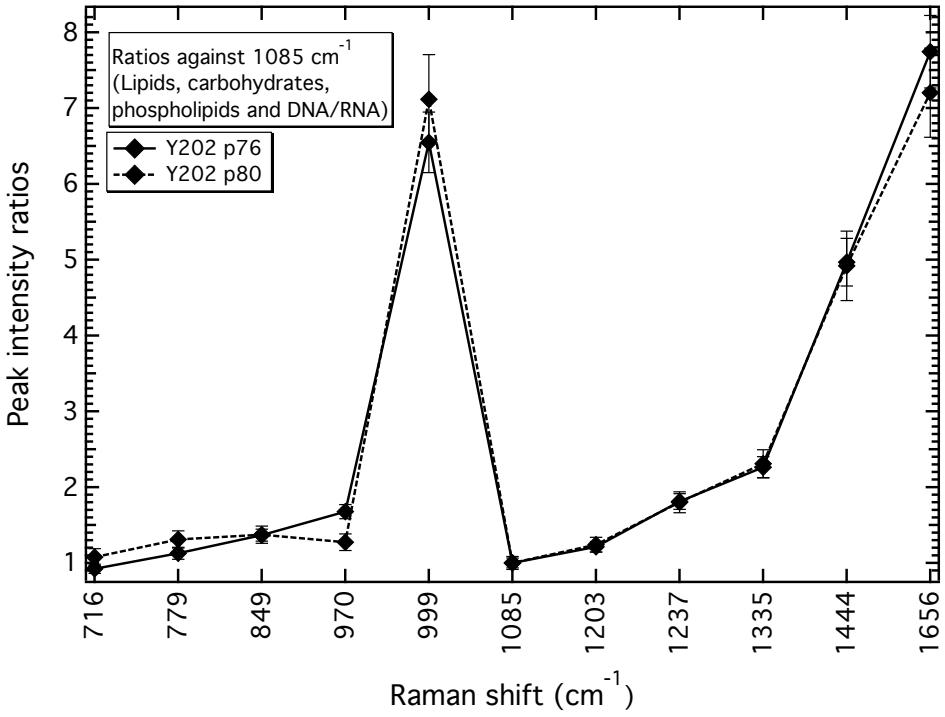


(b)

Figure 3.15: Comparison of PIRs against the 1085 cm^{-1} peak between different experiment repeats performed for the (a) Y101 and (b) Y201 MSC lines. The Y201p74 and Y201p95 repeats (in grey) were performed independently by Sam Hollings. The error bars are the propagation of the fitted intensity SEs.



(a)



(b)

Figure 3.16: Comparison of PIRs against the 1085 cm^{-1} peak between different experiment repeats performed for the Y102 and Y202 MSC lines. The error bars are the propagation of the fitted intensity SEs.

The 970 cm^{-1} and 999 cm^{-1} peaks are fitted together in the same band envelope, while the 1656 cm^{-1} peak is fitted in the same band envelope of other seven peaks. Thus, these variations could also be affected by fitting differences in these complicated regions. Another possibility is that factors such as the cell cycle and the use of cells harvested after sub-culturing are causing variations in the Raman spectra obtained from single-cells [40, 175]. This has not been the case, as the investigated MSCs were cell cycle synchronised and the cells were harvested from cultures when ∼70–80% of confluence had been reached, as described in Sections 2.2.2 and 2.2.3.

Bai *et al.* (2015) showed that Raman spectra of senescent and young human umbilical cord MSCs presented differences in the relative intensity of the 934 cm^{-1}, 980 cm^{-1}, 1003 cm^{-1}, 1097 cm^{-1}, 1127 cm^{-1}, 1157 cm^{-1}, 1174 cm^{-1} and 1521 cm^{-1} peaks [176]. Among these differences there is the inclusion of the 980 cm^{-1} peak for the spectra of senescent cells. However, although the differences were visible in the spectra obtained by them, their results pertaining the 980 cm^{-1} band were not statically significant due to the large standard deviations they obtained for this peak [176]. It is important to highlight that Bai *et al.* investigated primary MSCs, which lose multipotent differentiation capacity and have their replicative potential reduced with increasing cell doublings [79, 93, 124]. The MSC lines investigated in this thesis were immortalised using human telomerase reverse transcriptase (hTERT) and, therefore, do not suffer from a limited lifespan due to senescence [84].

In the previous sections, the variations in the relative intensity of the 970 cm^{-1} peak were linked to the air-drying process. However, this still has to be validated with the analysis of live-cells. In general, major differences, across the cell lines, for the 970/1085 PIR were only observed for the Y101 cell line, which indicates that the variations observed for the other cell lines and PIRs would not compromise their discrimination if this method

is used.

3.4.3 Spatially-resolved Raman images

The analyses of air-dried cells allowed maps of the cell nucleus to be performed. Raman maps provide the identification of differences in the morphology of the MSCs that could be linked to their discrimination. The number of spectra obtained in each map varied with each nucleus size. The spatially-resolved Raman images were constructed based on the values of the 1656/1085 PIR. This was one of the PIRs responsible for the grouping of the 01s (Y101 and Y201) and the 02s (Y102 and Y202) and presented the biggest separation between the two groups (Figures 3.5 and 3.8a). This PIR is related to ratios of proteins and lipids to DNA/RNA (Table 3.1). Other peaks that also provided discrimination, such as the 1549 cm^{-1} and 1615 cm^{-1} peaks, or the 970 cm^{-1} peak, which separated the four cell lines, were not used as they might be associated with biomolecular changes due to the air-drying process.

Raman maps of the nuclei of the MSCs were constructed for the 1656/1085 intensity ratio, as explained in Section 2.7 and are shown in Figure 3.17. From these spatially-resolved Raman images, it can be noticed that the Y101 and Y201 sampled cells have slightly smaller nuclei compared to the Y102 and Y202, which is in agreement with their measured nuclei perimeter and area (Appendix E, Figure E.2). The Y101 and Y201 lines have elongated fibroblastoid morphology, which is typical of MSCs, whereas the Y102 and Y202 cell lines were flattened and spread [84]. However, their nuclei shapes may have been distorted by the manual selection of the mapped area.

The maps showed that both the Y101 and Y201 cell lines have similar values for the PIR, which is evenly distributed in the nucleus area (Figure 3.17 (a) and (b)). In contrast, the mappings of the Y102 and Y202 MSC lines revealed areas that concentrated very

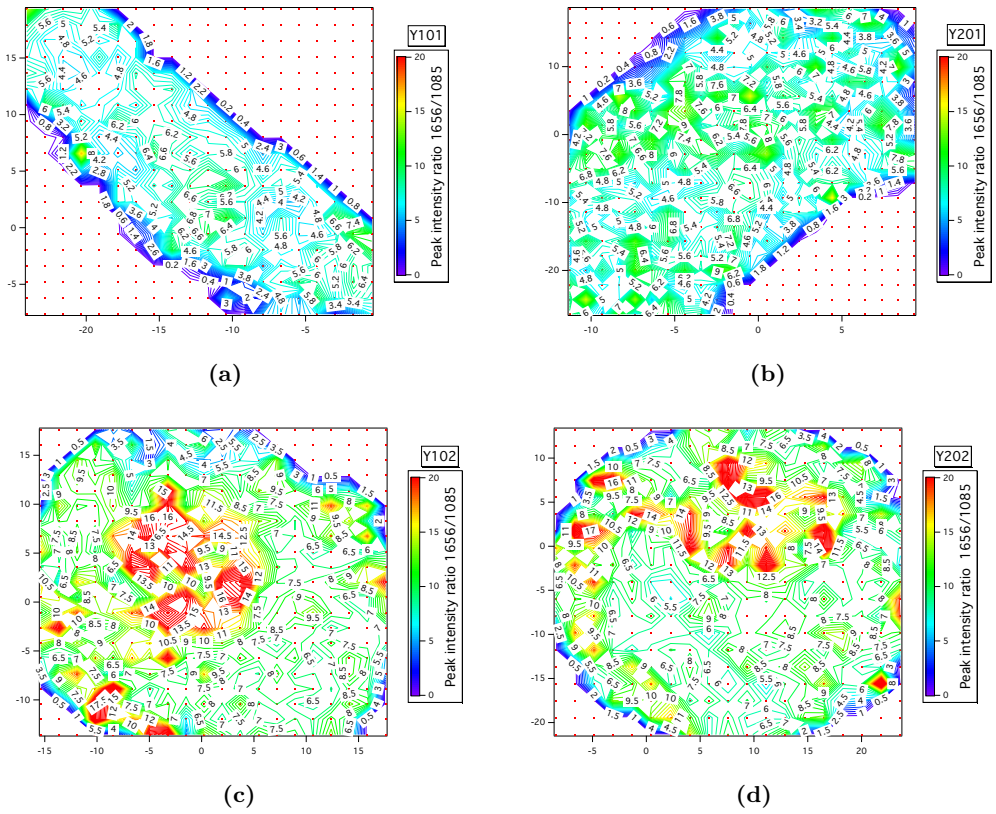


Figure 3.17: Spatially-resolved Raman maps from the nucleus of single cells representing the 1656/1085 PIR for the MSC lines Y101, Y201, Y102 and Y202. The x and y axes are the spatial coordinates (in μm) of each spectrum.

intense PIR values, ranging from value of 11 to 20, located near the centre and towards the periphery of their nuclei (Figure 3.17 (c) and (d)).

These areas with higher PIRs values can be identified as the nucleoli, which are nuclear subcompartiments of ribosome formation and consist of a network of RNAs and proteins [48]. The nucleolus is a huge aggregate of macromolecules and unlike other cell organelles it is not bound by a membrane [177]. Its structure is formed by granular components (GCs), fibrilar centres (FCs) and dense fibrilar components (DFCs) [178]. RS has already been used to identify the biomolecular composition of these structures, showing that the DFCs had high RNA content, the FCs were DNA-rich regions, while the GCs were protein-rich regions [48]. The red areas in the maps could, therefore, be explained as being GC regions

of the nucleolus since the mapped ratio (1656/1085) is associated with protein vibrations.

As the nucleolus is a very dynamic organelle, its number, size and shape can vary greatly depending on the cell type, cell-cycle and culture conditions. It is self-organised and also involved in the cell cycle regulation [178]. The nucleolus has its own cycle, disassembling each time the cell goes through mitosis and re-assembling during G1 [177]. The ploidy experiment shows that the serum starvation, through which the cells were synchronised, has pushed them towards G0/G1 and S phase, with particularly more 02 cells found in the S phase (Appendix E, Figure E.3). The fact that more cells from Y102 and Y202 populations were in the S phase suggests that they had their nucleoli already re-assembled, explaining the morphological differences captured by the Raman maps.

A relative frequency distribution of the 1656/1085 PIR is presented in Figure 3.18. The relative distribution means that the number of PIR occurrences had been normalised to the total number of spectra obtained from the different nuclei sizes.

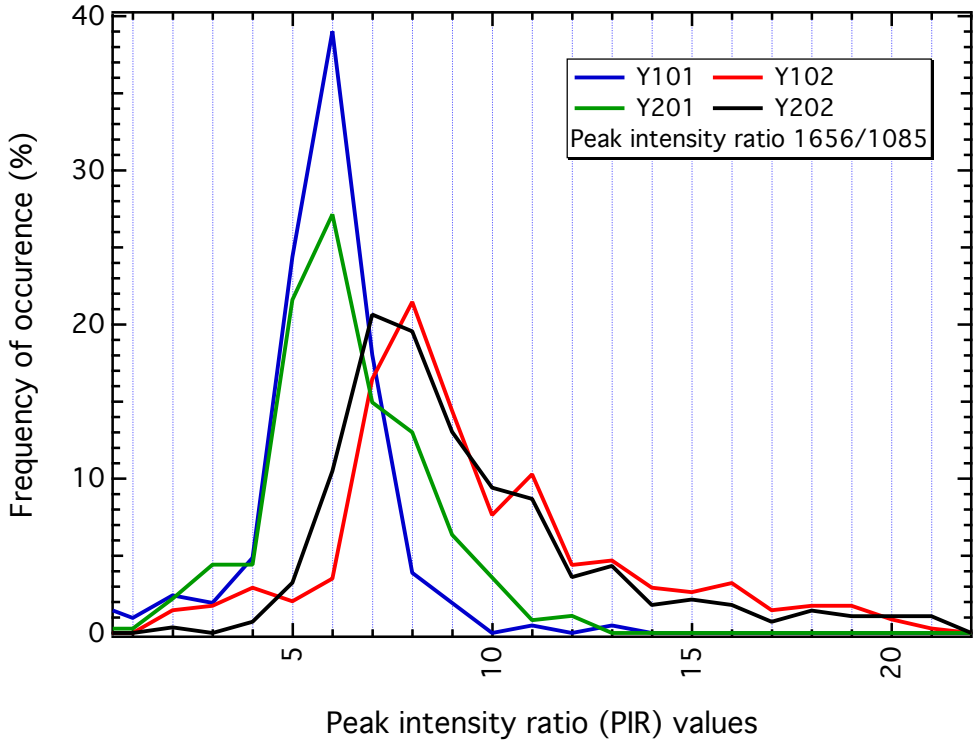


Figure 3.18: Relative frequency distribution of the 1656/1085 PIR from the spatially-resolved Raman maps of the MSC lines Y101, Y201, Y102 and Y202. The relative frequency distribution means that the number of occurrences was normalised to the mapped nucleus area as the cells have different nuclear sizes.

From Figure 3.18 it can be seen that, similar to the markers obtained for the averaged spectra (Figures 3.5 and 3.8a), the distribution shows group separation of the MSCs. Specifically, Y101 and Y201 have similar distributions of low values for the 1656/1085 PIR, with the most frequent value being six, while Y102 and Y202 has the value eight as the most frequent 1656/1085 PIR. The different distributions of the 1656/1085 PIR for 01s and 02s are, therefore, related to the phenotypic differences revealed by the Raman images and are correlated to the previously obtained Raman markers.

3.5 PIR analyses of primary MSCs

Primary MSC populations are known for being heterogeneous and containing cells with contrasting differentiation capacities, as well as containing cells that may or may not have progenitor characteristics [80, 93]. Thus, RS was applied to single cells from a heterogeneous primary MSC population (K72), with the aim to characterise them against the clonal hTERT MSC lines. For these comparisons, three random cells had their nuclei mapped just for collecting a greater quantity of spectra. This map data collection was described in Section 2.3. The average of the 36 spectra obtained from the nucleus of each of the three different single cells investigated along with their optical images is shown in Figure 3.19.

Two different sets of peak intensity ratios were calculated for these cells: against 1085 cm^{-1} and against 970 cm^{-1}, since these bands were used to discriminate between the 01 and 02 group as well as to perform full cell line discrimination. The comparison of the PIRs from these three K72 primary cells against those obtained for the hTERT MSC clonal lines are presented in Figures 3.20, 3.21 and 3.22. K72 cell 1 showed similar markers to the 02 group for the PIRs against 1085 cm^{-1} (Figure 3.20a). However, its markers oscillated between the Y102 and Y201 cell lines, specifically for the 1305/970 and 1549/970 PIRs, which identified it as Y201 (Figure 3.20b). This shows that while this cell has closer molecular features to the 02s group it also has a Y201-like constitution.

K72 cells 2 and 3 have their markers positioned alongside the 01 group for most of the 1085 cm^{-1} PIRs and all their PIRs against the 970 cm^{-1} peak are siting bellow the Y101 PIRs (Figures 3.21 and 3.22). This was due to the higher relative intensity of the 970 cm^{-1} peak against peaks of other cellular components for these two cells, as seen in Figure 3.19. Such higher relative intensity for the 970 cm^{-1} peak is a characteristic of the Y101 cell line, being not present in the other cell lines.

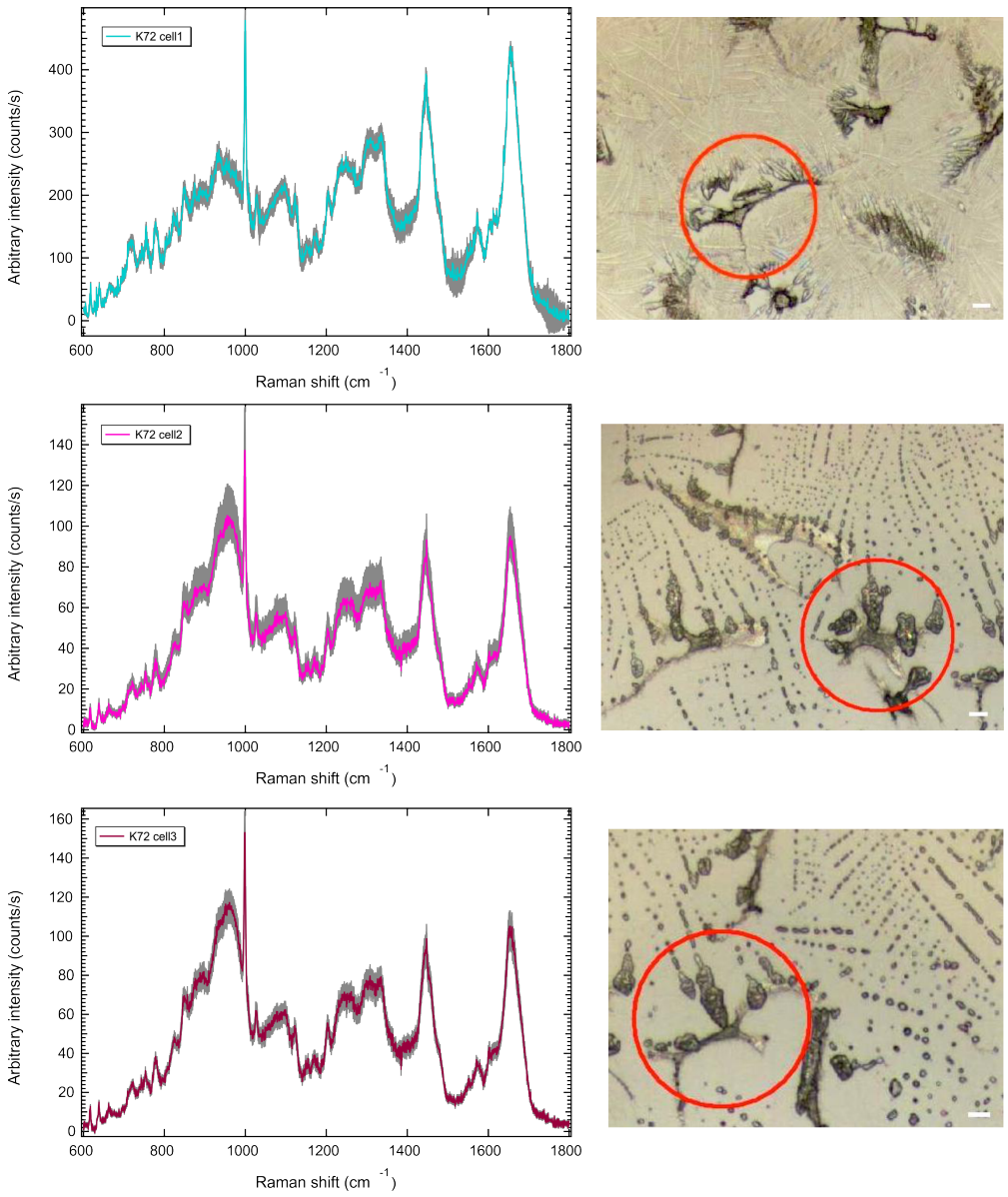
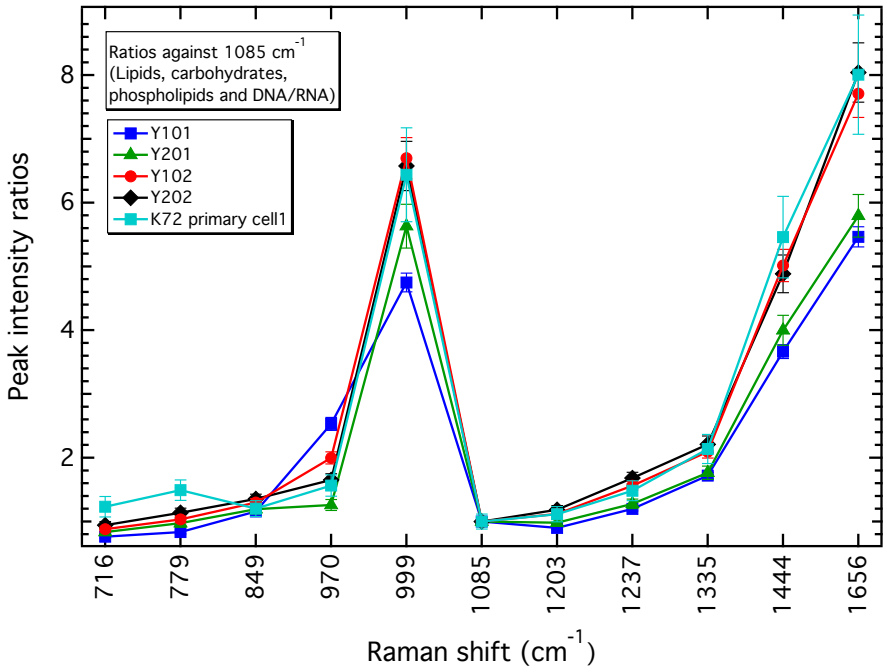
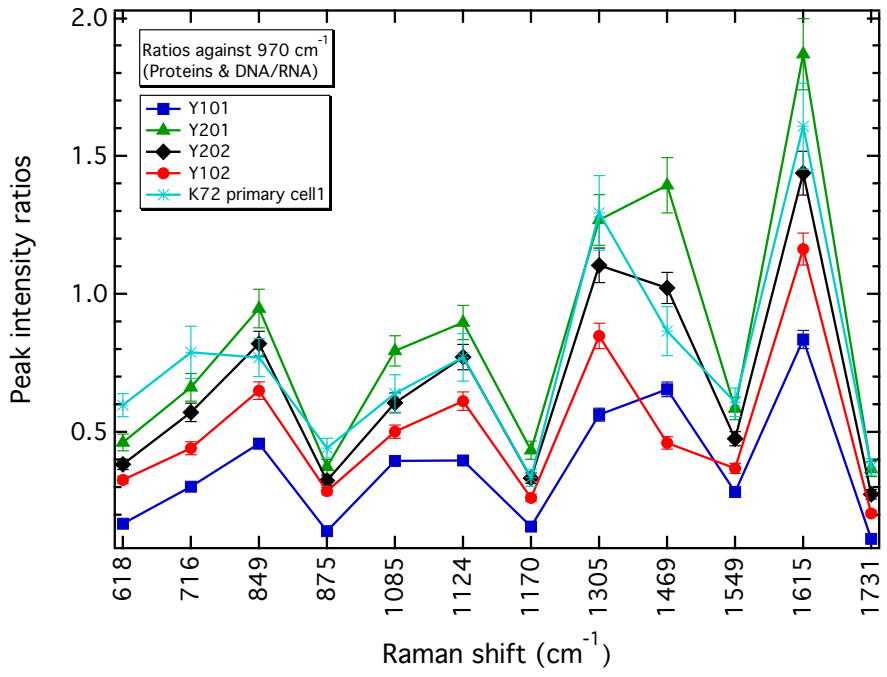


Figure 3.19: Raman spectra of single cells from the K72 primary population alongside their optical images (the scale bar is 20 μm). The spectra represent the mean of 36 spectra collected in the nucleus of each cell and are shown with their standard error envelopes.

Although only three cells were analysed for this primary MSC population, the variable profile on the PIRs obtained from cell to cell correlates with the heterogeneity of such populations. However, the results obtained against the 970 cm^{-1} peak have to be carefully considered as this peak seems to be affected by the air-drying process. While this is still an exploratory attempt, the results showed that cells with the same biomolecular profile

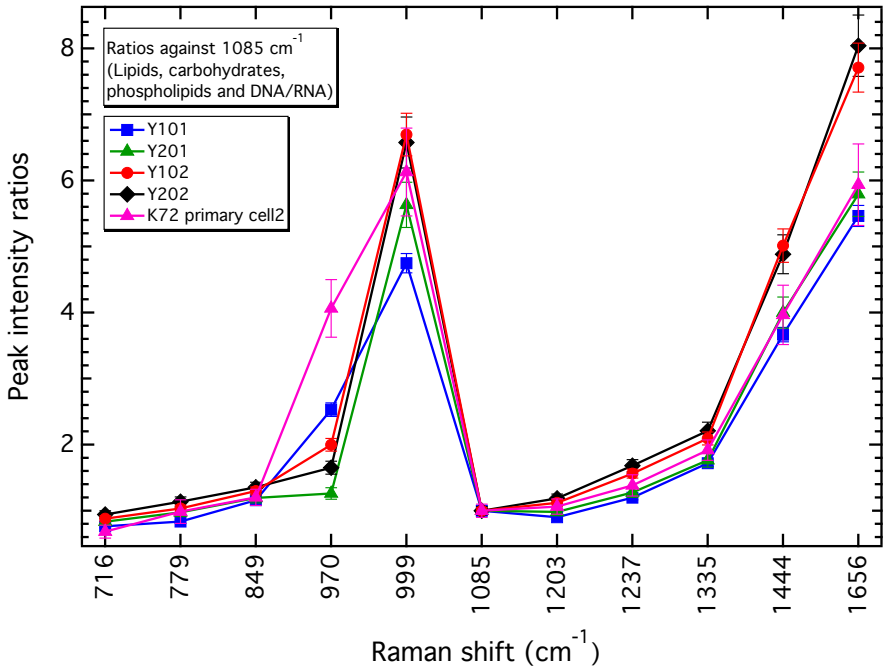


(a)

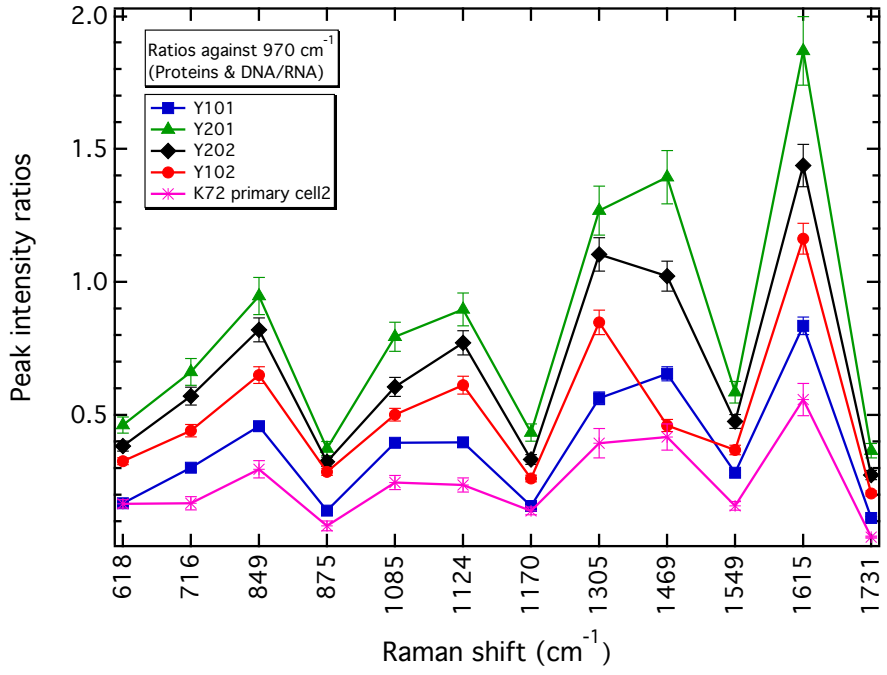


(b)

Figure 3.20: Raman markers from PIRs against peaks at (a) 1085 cm^{-1} and (b) 970 cm^{-1} for K72 cell 1. The PIRs are compared to those obtained for the Y101, Y201, Y102 and Y202 MSC lines.

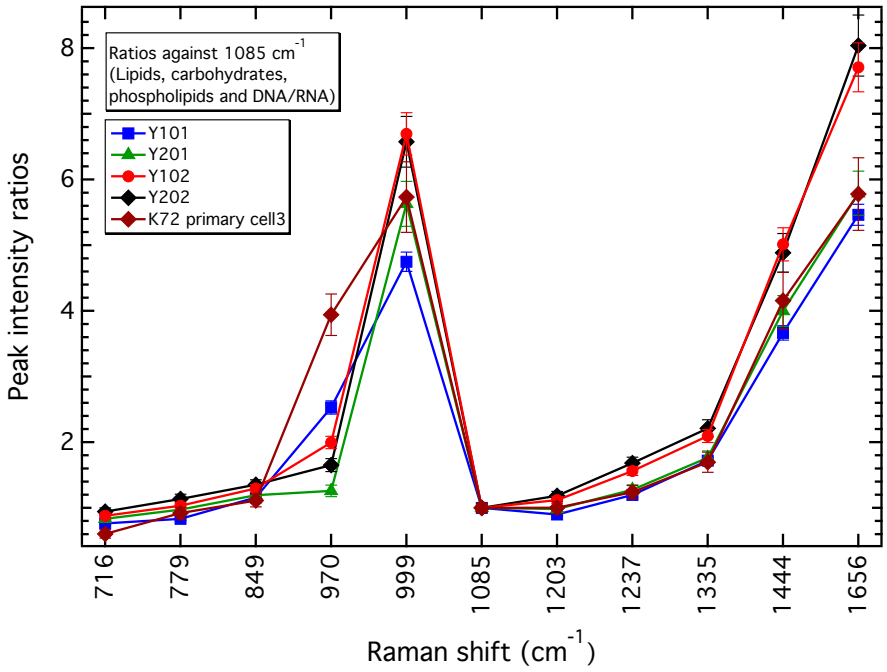


(a)

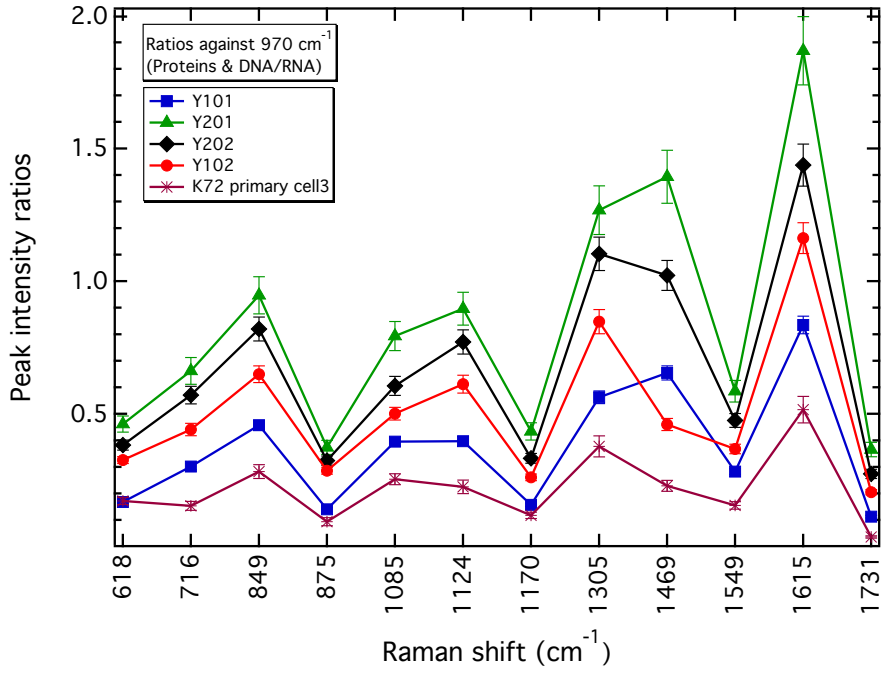


(b)

Figure 3.21: Raman markers from PIRs against peaks at (a) 1085 cm^{-1} and (b) 970 cm^{-1} for K72 cell 2. The PIRs are compared to those obtained for the Y101, Y201, Y102 and Y202 MSC lines.



(a)



(b)

Figure 3.22: Raman markers from PIRs against peaks at (a) 1085 cm^{-1} and (b) 970 cm^{-1} for K72 cell 3. The PIRs are compared to those obtained for the Y101, Y201, Y102 and Y202 MSC lines.

of the human immortalised MSC lines can be identified, using RS, in a primary MSC population.

3.6 Conclusions

The four immortalised clonal MSC lines studied have contrasting differentiation capacities, while the Y101 and Y201 are trilineage (osteogenic, adipogenic and chondrogenic) competent, the Y102 and Y202 are not. However, both groups (01s and 02s) express the cell-surface antigens commonly used to identify cells with progenitor ability and differentiation capacity. Additionally, the four cell lines were not discriminated by standard methods such as global gene expression. Unlike other cell types, such as haemotopoietic stem cells, MSCs lack recognised and widely adopted markers thus representing an excellent population for investigating the discriminating capabilities of RS.

Multivariate analysis (PCA-LDA) discriminated the spectra of the four air-dried MSC lines. The percentage of wrong predictions achieved by the PCA-LDA algorithm was 8%, demonstrating it as a suitable method for the MSC lines classification. The percentage achieved for correct predictions of differentiation competent cells was 94% when the groups of differentiation competent (Y101 and Y201) and incompetent (Y102 and Y202) cell lines were compared. PCA loadings revealed that a Raman background feature between the MSCs could be due to phenotypic differences across the cells. The loadings also indicated that the peaks contributing to the variation between the MSC lines were related to DNA/RNA vibrations (970 cm^{-1}, 1060 cm^{-1}, 1085 cm^{-1}, 1574 cm^{-1}), and protein and lipid vibrations (932 cm^{-1}, 999 cm^{-1}, 1444 cm^{-1}, 1656 cm^{-1}, 1731 cm^{-1}).

In agreement with the results obtained from the multivariate analysis, discriminatory markers for the cell lines were observed for PIRs against peaks at 932 cm^{-1} (proteins), 970 cm^{-1} (proteins and DNA/RNA), 1060 cm^{-1} (lipids, carbohydrates, phospholipids

and DNA/RNA), 1085 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA), 1549 cm^{-1} (proteins) and 1615 cm^{-1} (proteins) peaks. PIRs that provided discrimination between the four MSC lines were obtained against the 970 cm^{-1} peak, while PIRs against the other peaks separated 01s (Y101 and Y201) from 02s (Y102 and Y202). These PIRs allowed the discrimination of undifferentiated cell lines towards their differentiation competence, and are related to differences in lipids and proteins relative to DNA/RNA.

The primary sorted CD317+ cells were used as a control for the subpopulation discrimination as this protein is highly expressed in the Y102 and Y202 cell lines while absent for the Y101 and Y201. The PIRs obtained against the 1085 cm^{-1} peak confirmed the relatedness of this fraction (CD317+) with the 02 group, indicating that they have the same protein and lipid to DNA/RNA characteristics. Human dermal fibroblasts (HDFs) were used as a non-MSC control, with the PIRs against the 1085 cm^{-1} peak revealing that they were very similar to the 02s. However, the HDFs could be discriminated from the MSCs by the 780/1085 and 1237/1085 PIRs; both 780 cm^{-1} and 1237 cm^{-1} peaks have in common vibrations of the DNA/RNA molecule.

As indicated by PIRs calculated against peaks related to DNA/RNA vibrations, the MSC group separation obtained may be the effect of a difference in the nucleus size of the 01 and 02 MSCs, causing a variation in the density distribution of DNA/RNA molecules in the nucleus. Nuclear counterstaining and image J analysis revealed that, although not significant different, the 02s have slightly greater area and cell perimeter than the 01s.

Although the PIRs obtained from the 970 cm^{-1} peak discriminated the four cell lines, its relative intensity greatly varied among the MSCs and between different experimental repeats. It mostly varied for the Y101 cell line, which can spontaneously differentiate into osteoblasts. However, no evidence for the hydroxyapatite or its precursors was found. Cell-cycle and sub-culturing factors were also not found to be the case for this peak relative

intensity variation across the cell lines. The investigated MSCs were cell cycle synchronised and the cells were harvested from cultures when ∼70–80% of confluence was reached. The inclusion of a peak at 980 cm^{-1} was observed, in the literature, for senescent primary MSCs. However, the investigated MSC lines were immortalised and, therefore, do not suffer from a limited lifespan.

The 970 cm^{-1} peak is assigned to vibrations of DNA/RNA, but also to phosphorylated proteins. Protein phosphorylation increases during cell desiccation and has a regulatory role in cell death mechanisms, which could reflect the variations observed in the relative intensity of this peak. The 01s and 02s discriminating PIRs obtained against 1549 cm^{-1} and 1615 cm^{-1} peaks could also be reflecting such changes associated to the air-drying. They are assigned to tyrosine and tryptophan vibrations, which are common proteins involved in apoptosis. Thus, the PIRs against the 970 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} peaks might indicate that these MSC lines react differently to the air-drying process in such a fashion that their discrimination was optimised. However, in order to understand if these differences are intrinsic to the cells or an effect of the air-drying process it is necessary to analyse live-cells.

Population heterogeneity was also investigated by means of %SE associated to PIRs, and SE and 2x SD convergences over the full averaged spectrum. A measure of the heterogeneity degree was attempted by obtaining the decay rate of a PIR %SE, which is inversely proportional to the amount of averaged spectra. The %SE convergence tests indicated that the HDF and K72 primary populations were less homogeneous than the MSC lines; it also indicated that the CD317+ sorted primary population was the most homogeneous. This is in agreement with the knowledge that primaries and HDFs are more heterogeneous, and also agrees with the fact that the primary CD317+ had been sorted, thus representing a small fraction of MSCs. The convergence test also showed that the

uncertainties, associated to each experimental repeat, had converged and that the averaged spectra represented the population average used for the PIR analyses. However, in order to use these convergence tests to measure the degree of heterogeneity of a population, many more experimental repeats are needed, along with the analyses of different cell types.

The spatially resolved Raman images, constructed over the 1656/1085 PIR, indicated the presence of nucleoli in the Y102 and Y202 cells. A cell ploidy experiment revealed that, after synchronisation, more cells from these populations were in the S phase. This suggests that they already had their nucleoli re-assembled, which explains the morphological differences captured by the Raman maps. The mapping of these MSCs also revealed different distributions for the 1656/1085 PIR of the 01s and 02s which are related to the phenotypic differences revealed by the Raman images, and therefore to their discrimination.

Finally, the recognition of single-cells with a similar biomolecular profile to the MSC lines allows the identification of subpopulations in primary MSCs, as demonstrated by the results obtained from the three investigated cells from the K72 population. Although the whole nucleus had to be scanned, the results suggest that RS could complement standard protocols for selecting cells with progenitor ability and differentiation competence in such heterogeneous primary populations.

In summary, this chapter demonstrated that RS is able to provide subpopulation discrimination and also to evaluate heterogeneity. The recognition of these MSCs subpopulations is important once primary populations are being used in clinical-trials. The subpopulation discrimination was obtained with both univariate (PIRs) and multivariate (PCA-LDA) analyses. The PIR analysis provided sufficient information to reveal what underlies the cell line differences.

Chapter 4

Live cell analyses of undifferentiated mesenchymal stromal cell lines

Chapter 3 showed that RS was able discriminate subpopulations of MSCs. This was demonstrated by multivariate (PCA-LDA) and univariate (PIR) analyses. The identified PIR markers were able to discriminate these MSC lines into two groups according to their trilineage differentiation capacities. Fully discriminatory markers of the four subpopulations were identified for PIRs against the protein and DNA/RNA peak at 970 cm^{-1}. However, it was noticed that the relative intensity of this band was substantially altered from one experimental repeat to the other. Thus, since these fluctuations could have been enhanced by changes occurring in dried samples, this chapter focuses on the analyses of live undifferentiated MSCs. It aims to verify if the subpopulation discrimination is attainable in live cells. Although the air-dried experiments allowed a prompt characterisation of these cell lines, the discrimination of live cells is envisaged for their future use in clinical

trials.

Initially, the cells were checked via a trypan blue staining assay to verify if the laser could induce cell death. Then, the spectra obtained from the four MSC lines (Y101, Y201, Y102 and Y202) were analysed by PCA to reveal the bands that contributed to the spectral variance, classified by a PCA-LDA algorithm and by means of PIRs Raman discriminatory markers were identified. Finally, as done previously in Chapter 3, the heterogeneity amongst these cell lines was investigated by using the uncertainties convergence tests.

4.1 Trypan blue viability assay

A trypan blue dye assay was carried out to verify if the laser exposure could damage the cell membrane and induce cell death. Viable cells have an intact cell membrane that is capable of excluding certain dyes, such as trypan blue, which is one of the most common dyes for exclusion assays [150]. Trypan blue has already been used to evaluate cell membrane integrity of cardiomyocytes derived from human embryonic stem cells after exposure to a 785 nm laser during Raman acquisition [36].

As explained in Subsection 2.2.3, for the live cell experiments the cells were seeded onto CaF_2 discs, incubated with growth medium, synchronised and then kept in HBSS during the Raman acquisitions. The trypan blue assay used in this study was tailored so that the same cell was tested before and after laser irradiation (Subsection 2.3.1).

Figure 4.1 shows the three repeats for the trypan blue assay. A cell was randomly chosen in the population and the nucleus was irradiated using the same spectral acquisition conditions of the Raman data collection (Section 2.3). The column A, in Figure 4.1, shows the cells prior to laser irradiation, while column B shows the same cells after being irradiated and stained with trypan blue. The dye is not seen present in the irradiated cells,

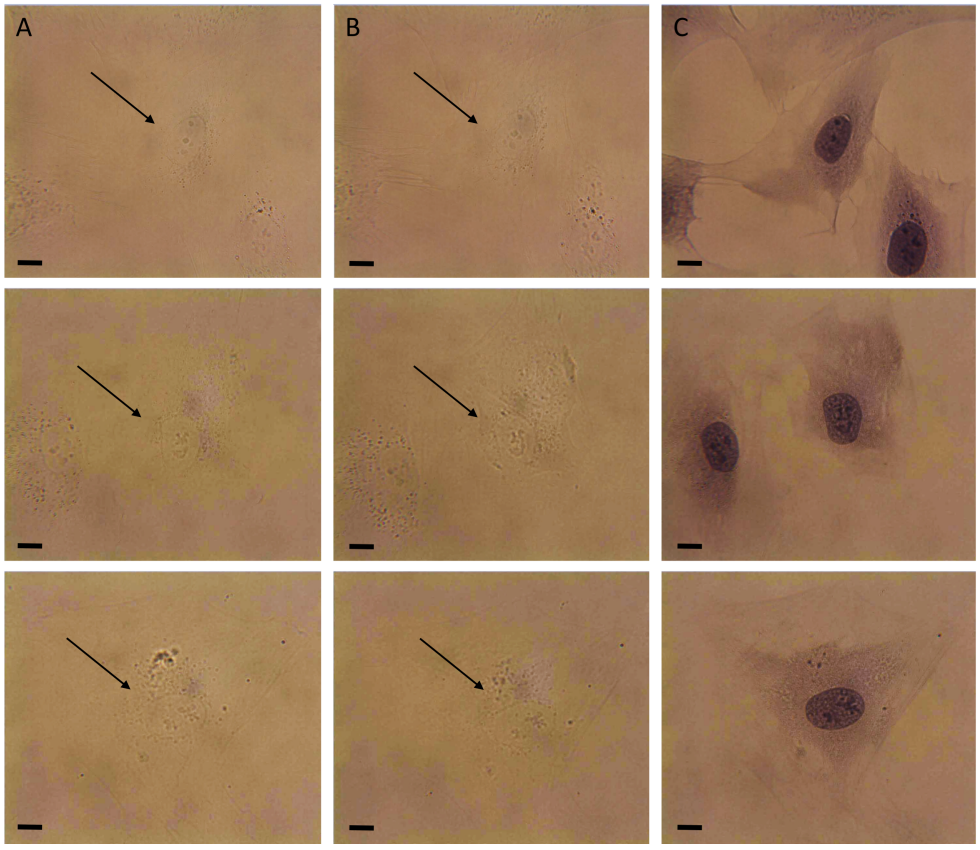


Figure 4.1: Cell specific trypan blue staining assay. The trypan blue dye assay was performed on the MSCs to check for cell membrane damage caused by the laser exposure. Each row corresponds to one experimental repeat. A cell was randomly chosen (column A) and laser irradiated with the same acquisition parameters used for the Raman data collection. The arrows point to the selected cell. The staining assay was performed right after laser exposure (column B) showing no damage to the cell membrane and after the cells were immersed in 70% ethanol for 5min, revealing the dye uptake by non-viable cells (column C). Scale bar is 10 μm.

confirming that the laser exposure did not damage the cell membrane. Following this and as a control, the whole population was immersed in 70% ethanol and again stained with trypan blue. The dye uptake by the cells (column C) is now visible, revealing that they are no longer viable. These results indicate that the laser exposure did not cause any drastic biological change to the cells, and that the spectra further obtained from them represent the biomolecular composition of viable living cells.

4.2 Averaged live cell spectra

A total of 48 spectra, obtained over three experimental repeats for each MSC line, were averaged and then peak fitted (Section 2.6). The convergence tests, described in Section 2.6.2 ensure that the averaged spectra is a statistically converged representation of the population average. The results of each test, %SE, SE and 2x SD convergences, for each sample and experimental repeat correspond to 48 graphs and are shown in Appendix B.4. These results are further discussed in the Section 4.5 of this chapter.

Figure 4.2 shows the averaged spectra and the standard error of the mean (SE) envelopes of the live undifferentiated MSC lines – Y101, Y201, Y102 and Y202. All studied peaks are identified in the figure and their peak positions were calculated by averaging the positions obtained from the peak fittings performed on each cell line, as per Section 2.6. The uncertainty of the measured peak positions is the equipment's wavenumber precision of ± 1 cm^{-1}. For the live cell experiments, the Raman data was collected from the fingerprint (600 – 1800 cm^{-1}) and high-wavenumber (2700 – 3100 cm^{-1}) regions. The HBSS in which the cells were kept during Raman data collection does not produce any peak in the fingerprint region and causes only the curved background in the high-wavenumber region of the spectrum (Figure 4.3).

Two peaks that were not visible in the air-dried spectra (Figure 3.1) were observed in the fingerprint region of the live cells at positions 1403 cm^{-1} and 1422 cm^{-1}. The peak 1403 cm^{-1} is assigned to vibrations of proteins and lipids, while the peak 1422 cm^{-1} is assigned to protein, lipids and DNA/RNA vibrations [41, 42, 56, 154, 168]. Their detailed assignments and the assignments of the peaks observed in the high-wavenumber region are listed in Table 4.1. The assignments of all other peaks have already been listed in Table 3.1 in Chapter 3.

The four undifferentiated live MSC lines show similar spectral features, with the finger-

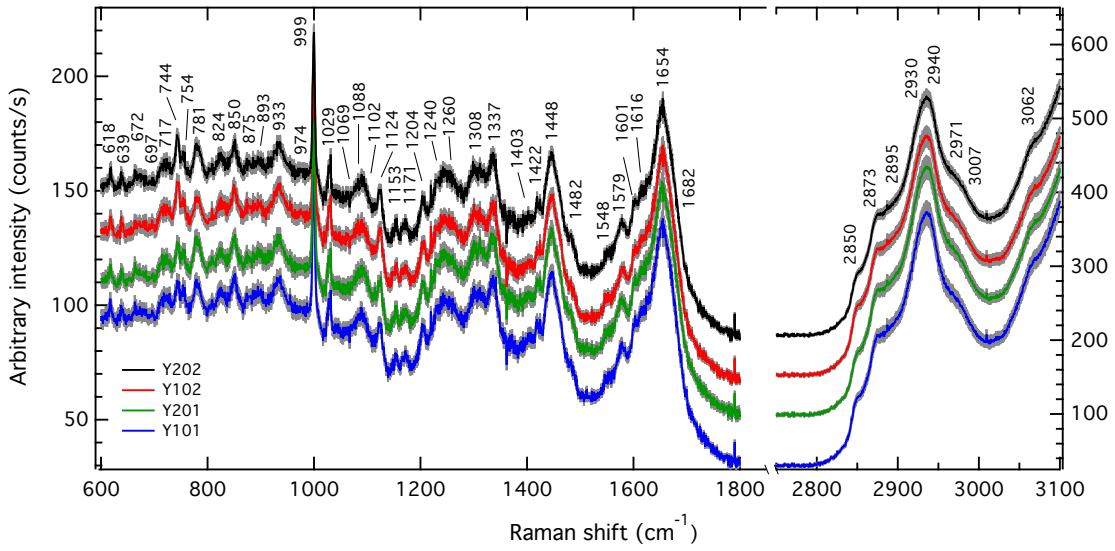


Figure 4.2: Averaged Raman spectra of live undifferentiated MSC lines. Each population mean contains three experimental repeats from the data obtained at random positions of the cells nuclei. The undifferentiated MSC spectra are shown with their respective standard error of the mean envelopes (grey curves) and were offset for clarity. All identified peaks are assigned in Tables 3.1 and 4.1.

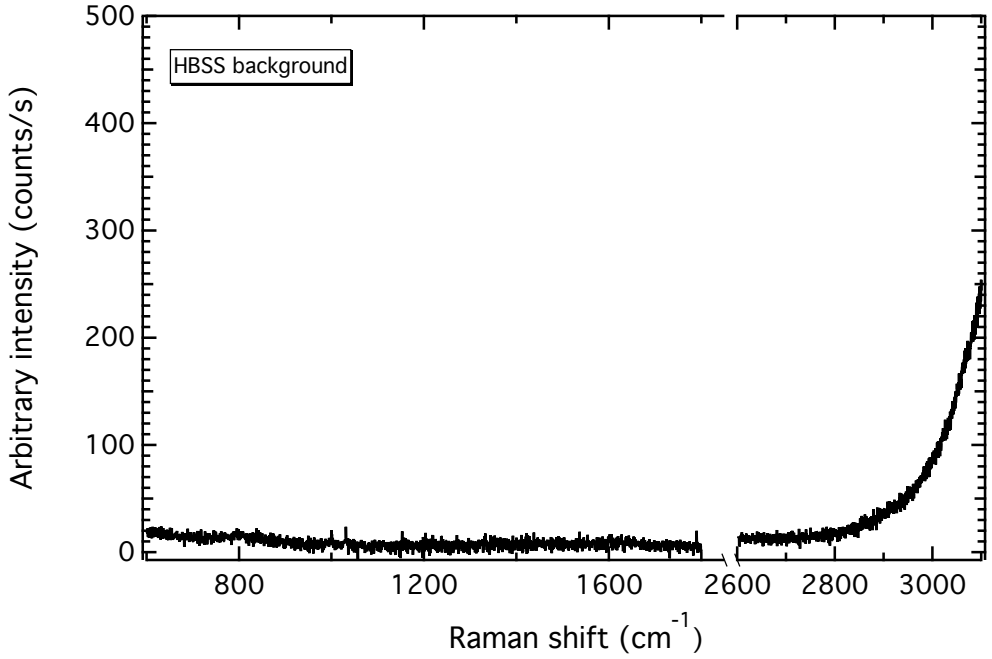


Figure 4.3: HBSS spectrum obtained with the same Raman acquisition parameters used for the cell spectra collection.

Table 4.1: Assignments of new peaks observed in the live cells Raman spectra. Each assignment corresponds to the measured peak positions, averaged across the spectra of the live cells (Figure 4.2). The uncertainty of the measured peak positions is the equipment's wavenumber precision of ± 1 cm^{-1}.

Peak position (cm^{-1})	Literature range (cm^{-1})	General biomolecular assignments	Detailed biomolecular assignments
1403	1400-1404	Proteins & lipids	NH in-plane deformation, C(C=O)O^{-1} for amino acids and glutamic acid [154]. CH deformation in proteins and lipids [179].
1422	1420-1424	Protein, lipids & DNA/RNA	Deoxyribose as marker for B and Z DNA [154, 179]. Guanine and adenine [168]. CH deformation in proteins and lipids [41,42,56, 154]
2850	2850-2851	Lipids	CH_2 symmetric stretching of lipids [24, 30, 72, 154, 179, 180].
2873	2873-2876	Lipids & proteins	CH_3 symmetric stretch of lipids and proteins [24, 72, 154, 179].
2895	2893-2895	Lipids & proteins	CH_3 symmetric stretch of lipids and proteins [72, 154].
2930	2929-2930	Lipids	CH_2 asymmetric stretch of lipids [24, 72, 154, 179].
2940	2935-2940	Lipids & proteins	CH vibration in lipids and proteins [72,154,179].
2971	2959-2970	Lipids & proteins	CH_3 vibration in lipids, fatty acids cholesterol and cholesterol esters [72, 154]. CH_3 asymmetric stretching of lipids and proteins [24]. Asymmetric stretching of methoxy groups [154, 179].
3007	3008-3015	Lipids	=CH vibration in lipids and fatty acids [72, 154, 179]. Unsaturated lipid =CH stretch [154]. Olefinic =CH stretching from unsaturated lipids and cholesterol esters [24].
3062	2800-3050	Lipids	Acyl chain vibrations [181]. OH, NH and CH stretching vibrations [179].

print region of the spectra dominated by protein, lipid and DNA/RNA vibrations, while the high-wavenumber region comprises vibrations assigned only to proteins and lipids (Figure 4.2). Visible differences across the four live undifferentiated MSC lines are noted for the 744 cm^{-1} and 781 cm^{-1} peaks, both assigned only to DNA/RNA vibrations.

In the literature, the comparison of Raman spectra obtained from live and dead cells, revealed that the magnitude of DNA/RNA peaks (in spectra normalised to the area of peak 1444 cm^{-1}) decreased considerably for non-viable cells [41, 44]. The two new peaks observed in the fingerprint region of live cell spectra (1403 cm^{-1} and 1422 cm^{-1}) were indistinguishable from the background in the air-dried MSC spectra (Figure 3.1). If the spectra of the air-dried (Figure 3.1) and live MSCs (Figure 4.2) is further compared, it reveals that the DNA/RNA peaks at 744 cm^{-1}, 781 cm^{-1} and 1579 cm^{-1}, are more evident in the live cell spectra. Such decrease in the relative intensity of DNA/RNA peaks has been attributed, in the literature, to changes in the DNA structure and integrity, for the case of apoptotic cells [156, 182].

Notingher *et al.* (2002) noticed that a small peak at 1743 cm^{-1}, correlated to the carbonyl C=O stretching vibration in phospholipids, became apparent in the spectra of dead cells due to the shrinkage of the amide I peak width (at $\sim$1660 cm^{-1}) [44]. This peak was observed in the spectra obtained from the air-dried MSCs (Figure 3.1); however, positioned at 1731 cm^{-1}. In the literature, the position for this vibration ranged from 1729 cm^{-1} to 1743 cm^{-1} [42, 43, 56, 154, 156]. Thus, the FWHM of the amide I band was compared between the spectra obtained from live and air-dried MSCs. The averaged FWHM, across the four MSC lines, reduces from 50 $\pm$1 cm^{-1} for the live cells, to 33 $\pm$ 2 cm^{-1} for the air-dried cells, indicating modification of the protein conformation in the air-dried cells and thus, in agreement with Notingher's observations. The uncertainties shown for the FWHM averages are the SE.

Another difference noticed between the live and air-dried cell spectra is pertained to the positions of some of the fitted peaks. Table 4.2 compares the averaged peak positions of the air-dried and live MSCs for which the difference was bigger than 1 cm^{-1}, which is the equipment's wavenumber precision. High-wavenumber peaks were not compared because this region was only investigated for the live cells. Apart from the 1656 cm^{-1} peak, a shift to lower wavenumbers was noticed, in the air-dried cell spectra, for all peaks listed in Table 4.2. These shifts might be associated to changes in DNA/RNA structure, which is also reflected by a decrease in their relative intensities [44, 156, 182], and to a change in the overall protein conformation in the air-dried cells [182, 183]. Low wavenumber shifts of protein peaks have already been noticed in the spectra of dead cells with Fourier transformed IR spectroscopy [183]. Although these shifts have also been described in literature for the 1656 cm^{-1} peak [182, 183], the same was not observed for the air-dried cells. This may have been influenced by an increase in the relative intensity of the lipid peaks. An increase in the relative contribution of lipids in the air-dried cells is observed by the presence of the 1743 cm^{-1} peak and the decrease in the FWHM of the 1656 cm^{-1} band [41, 156], and was also correlated to reduced protein signals in apoptotic cells [182].

The most striking difference between the live (Figure 4.2) and air-dried (Figure 3.1) MSC spectra lies for the proteins and DNA/RNA peak at $\sim$970 cm^{-1}. This band showed great variation in its relative intensity between the four air-dried MSC lines, being responsible for the full cell line discrimination obtained by the PIRs (as shown on Figure 3.9). However, contradicting all other DNA/RNA assigned peaks, which were more evident in the live MSC spectra, this peak is almost unnoticeable. The 970 cm^{-1} peak is originated from the vibrations of phosphate monoester groups of phosphorylated proteins and DNA/RNA nucleic acids [154, 157, 158]. As explained in Chapter 3, protein phosphorylation has a regulatory role in cell death mechanisms [173] and is also associated

Table 4.2: Comparison of the peak positions for the averaged spectra obtained from air-dried and live cell experiments. The peak position uncertainty is ±1 cm^{-1}, which is the instrument's wavenumber precision. However, the position of the 1469 cm^{-1} peak showed varied greatly across the air-dried cell lines, for which the SE of ±3 cm^{-1}. The general peak assignments were defined from the Table 3.1 in Chapter 3.

Peak position (cm^{-1})		General peak assignment
Air-dried cell	Live cell	
664	672	DNA/RNA
742	744	DNA
779	781	DNA/RNA
891	893	Carbohydrates
970	974	Proteins and DNA/RNA
1060	1069	Lipids, carbohydrates, phospholipids and DNA/RNA
1085	1088	Lipids, carbohydrates, phospholipids and DNA/RNA
1100	1102	Lipids, proteins, carbohydrates and DNA/RNA
1237	1240	Proteins, lipids and DNA/RNA
1305	1308	Proteins, lipids and DNA/RNA
1335	1337	Proteins and DNA/RNA
1469	1482	Proteins, lipids and DNA/RNA
1574	1579	DNA/RNA
1656	1654	Proteins and lipids

to desiccation [172]. Therefore, the differences in the relative intensity of the 970 cm^{-1} band, observed in the air-dried spectra, could be the reflection of different protein phosphorylation between the MSC lines as result of the air-drying process. This indicates that the four MSC lines may modulate cell death differently and explains why such an intense feature was not observed for this band in the spectra of the live MSCs.

4.3 Multivariate analysis: PCA-LDA of the live MSC lines

Multivariate analyses were used to investigate which spectral bands accounted for most of the variance between the four live MSC spectra. The spectra obtained in each cell was baseline corrected, area normalised and smoothed as per Section 2.6.3. Since the

fingerprint and high-wavenumber regions were collected for each Raman spectrum, both were included in the multivariate analysis.

Figure 4.4 shows the PCA scores for the live MSC spectra. The first three PCs correspond respectively to 23%, 14% and 7% of the variance in the dataset. Similarly to the PCA performed on the spectra of the air-dried cells (Figure 3.2), the unsupervised method alone does not discriminate the four MSC lines.

The loadings of these three PCs are shown in Figure 4.5. The PC1 loading (Figure 4.5a) is dominated by a background feature, similarly to what was pointed by the PC1 loading from the PCA performed in the air-dried MSC lines (Figure 3.3). As explained in Chapter 3, Raman backgrounds have been used in the literature to discriminate cells, being associated to phenotypical differences between them [31,166]. Apart from the background feature, loading 1 also shows the peaks at 999 cm^{-1} (phenylalanine), 1422 cm^{-1} (proteins, lipids and DNA/RNA), 1482 cm^{-1} (proteins, lipids and DNA/RNA) and 1654 cm^{-1} (lipids and proteins) as major contributors, in the fingerprint region, for the PC1 scores. For the

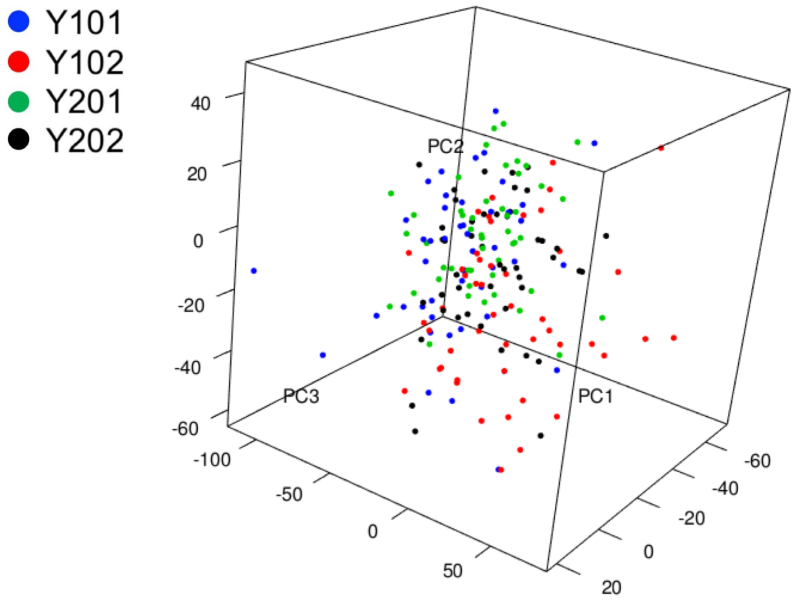


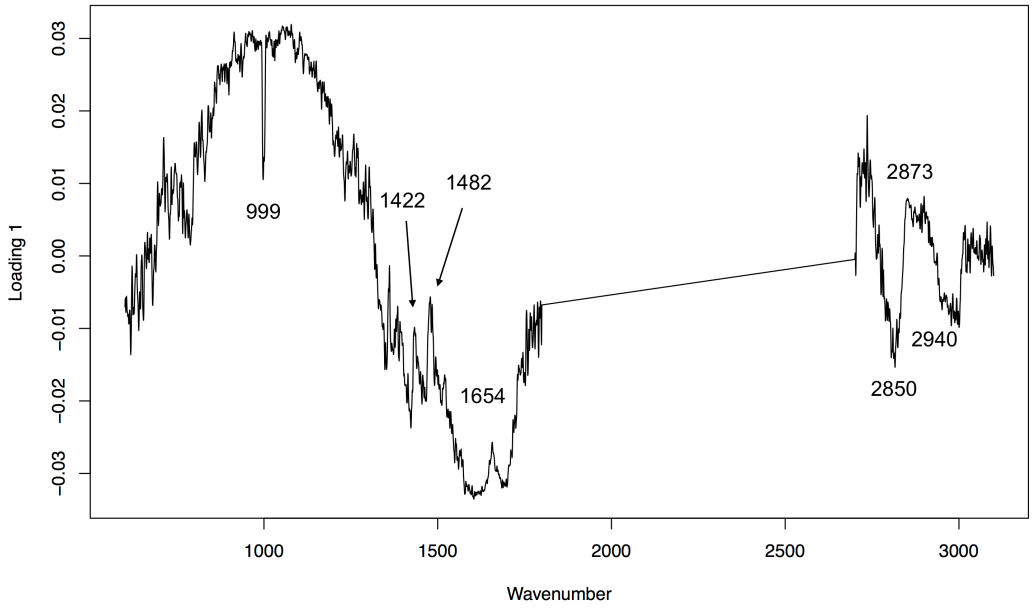
Figure 4.4: PCA scores for the live MSC spectra. The PCA components PC1, PC2 and PC3 accounted respectively for 23%, 14% and 7% of the variance.

high-wavenumber region, it shows that the 2873 cm^{-1} (lipids and proteins), 2850 cm^{-1} (lipids) and 2940 cm^{-1} (lipids and proteins) peaks also contributed to the variance between the cell lines.

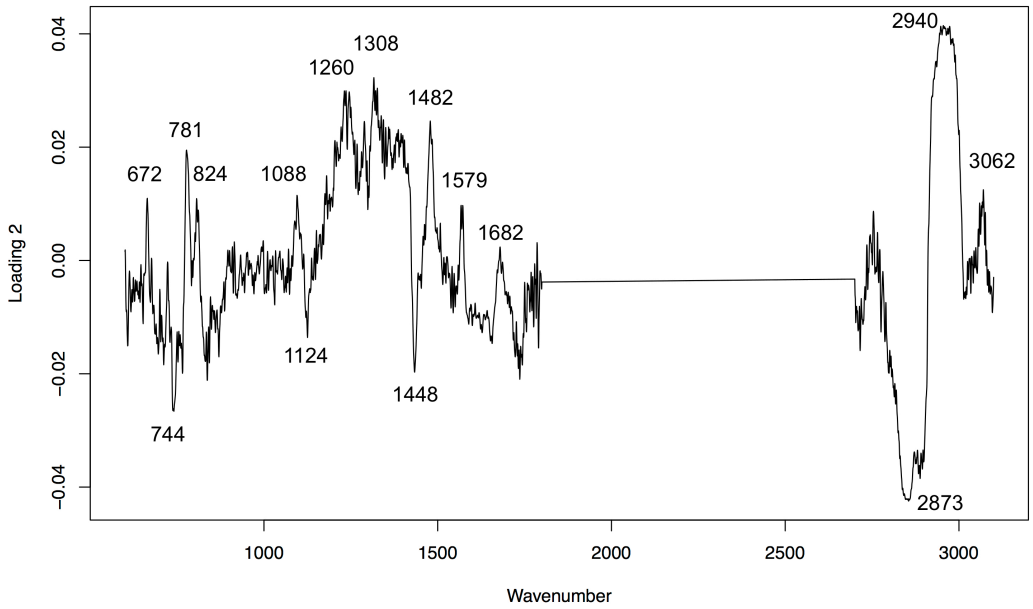
The PC2 loading (Figure 4.5b) reveals many other peaks responsible for the variance in the dataset, such as the fingerprint region peaks assigned only to DNA/RNA vibrations (672 cm^{-1}, 744 cm^{-1}, 781 cm^{-1} and 1579 cm^{-1}), other peaks assigned to DNA/RNA, but also to proteins and lipids (824 cm^{-1}, 1088 cm^{-1}, 1260 cm^{-1}, 1308 cm^{-1} and 1482 cm^{-1}) and those associated to lipids and proteins (1124 cm^{-1}, 1448 cm^{-1} and 1682 cm^{-1}). The peaks associated to DNA/RNA vibrations at 781 cm^{-1}, 1579 cm^{-1} and 1088 cm^{-1} also contributed to the PC2 loading of the air-dried cells (Figure 4.5b). For the high-wavenumber region, the peaks at 2873 cm^{-1} (lipids and proteins), 2940 cm^{-1} (lipids and proteins) and 3062 cm^{-1} (lipids) contributed to the variance in the dataset.

PC3 loading (Figure 4.5c), which accounted for 7% of the variance in the dataset, highlights the peaks at peaks at 850 cm^{-1} (proteins), 933 cm^{-1} (proteins), 1337 cm^{-1} (proteins and DNA/RNA) and 1654 cm^{-1} (lipids and proteins), and also the aforementioned peaks observed in PC2 loading. It also highlights, for the high-wavenumber the peaks at 2873 cm^{-1} (lipids and proteins) and 2930 cm^{-1} (lipids). Although PCA did not discriminate the spectra of the live cells, the PCA loadings revealed the peaks associated to DNA/RNA, proteins and lipids vibrations that largely contributed to the PCA scores.

In order to maximise the between group variance and verify if the MSC lines discrimination could be obtained, a LDA was performed to the dataset, as described in Section 2.6.3. Again, the LDA was applied to the PCs and, to prevent overfitting of the data, the number of PCs input into the LDA algorithm were half the number of spectra obtained for the smallest group [36]. Thus, since a total of 48 spectra were obtained for each of the live MSC lines, 24 PCs were used, totalling 70% of the variance in the dataset. For

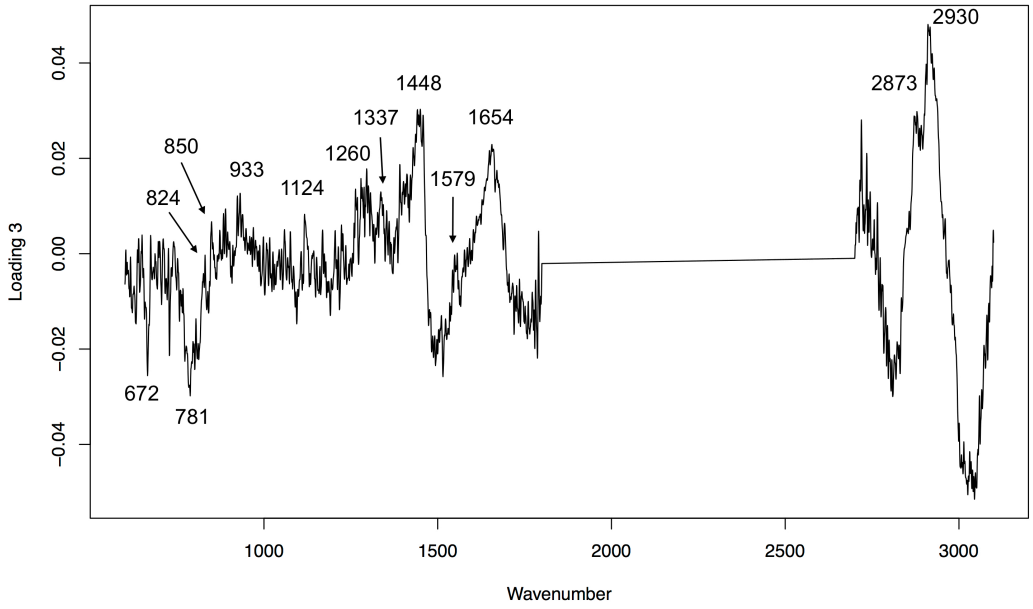


(a)



(b)

Figure 4.5: Principal component analysis of the live MSC lines datasets showing the loadings of (a) PC1, (b) PC2 and (c) PC3. The peaks indicated in each loading were the major contributors of each PC scores - *continued on the next page.*



(c)

Figure 4.5: *Continued from the previous page.* Principal component analysis of the live MSC lines dataset showing the loadings of (a) PC1, (b) PC2 and (c) PC3. The peaks indicated in each loading were the major contributors of each PC scores.

the air-dried dataset, the number of PCs used totalled 95% of the variance; however, if this was applied to the live cell data, 127 PCs would need to be incorporated into the algorithm, thus causing overfitting.

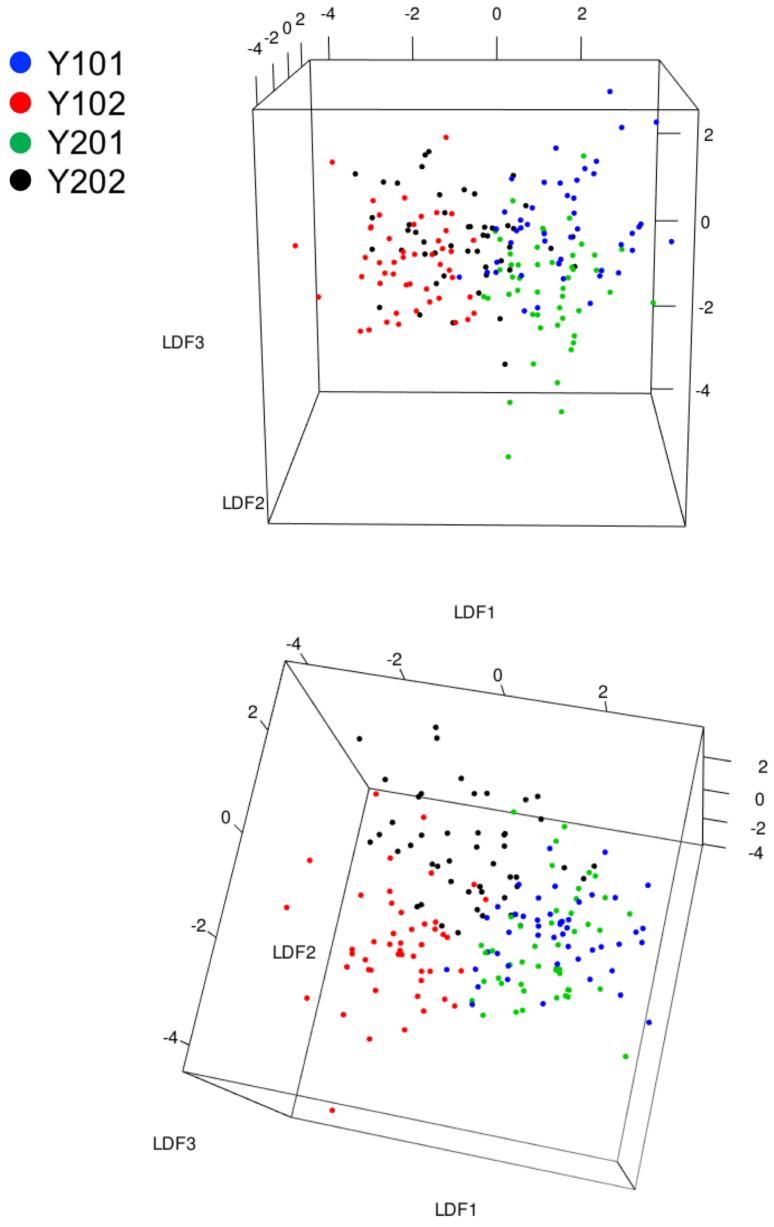


Figure 4.6: PCA-LDA scores of the live MSC lines. Both three dimensional score plots represent the same results, but were rotated for better visualisation. LDF1 separates the 01s (Y101 and Y201) from the 02s (Y102 and Y202) groups. LDF2 is responsible for separating Y102 from Y202 and LDF3 shows the start of separation between the Y101 and Y201.

It can be seen from Figure 4.6 that the PCA-LDA clearly separates the four live MSCs into two distinct groups, one for the 01s (Y101 and Y201) and other for the 02s (Y102 and Y202). These two groups of cells have distinct differentiation capacities, the 01s are able of trilineage differentiation, whereas the 02s are differentiation incompetent. This 01s and 02s discrimination was also achieved by hierarchical clustering analysis and PCA performed on the global gene expression analyses of these MSCs [84]. From Figure 4.6, it can be seen that LDF3 shows a start of separation between the 01s (top image), while the LDF2 nearly separates Y102 from Y202 (bottom image). This suggests that the number of collected spectra might not have been sufficient for the multivariate analyses, once the spectra is individually compared. Although the convergence tests indicated that the amount of spectra obtained was enough the represent the samples (Appendix B.4, in Sections B.5 and B.6), these tests were conceived for the analyses of averaged spectra and not individual spectrum.

A cross-validation of the PCA-LDA method was performed with two thirds of the data (128 spectra) used to train the prediction algorithm while the other third (64 spectra) was used to test it. The result of the data classification obtained by the PCA-LDA is presented in Table 4.3, in the form of a confusion matrix describing the performance of the algorithm. The total error of the method was 34% (percentage of wrong predictions), a value much higher than the 8% obtained for the air-dried cells. This increase is explained by the fact that the predictions were tested for the four MSC lines, whilst the algorithm was able to only discriminate these cell lines into two groups. However, it can be seen from Table 4.3 that most of the cells were predicted to the correct group with only four 01s cells being classified as 02s and two cells from the 02s classified as 01s. In this case, 91% of the cells were correctly classified as 01s or 02s. This indicates that this method, together with the number of acquired spectra, is suitable for classifying the cells towards their differentiation

Table 4.3: Confusion matrix of the PCA-LDA classification algorithm results obtained for the live MSCs.

	Predicted cell line			
Cell line	Y101	Y201	Y102	Y202
Y101	9	7	0	2
Y201	2	10	0	2
Y102	0	0	10	3
Y202	1	1	4	12

competence, therefore, into the 01 or 02 groups.

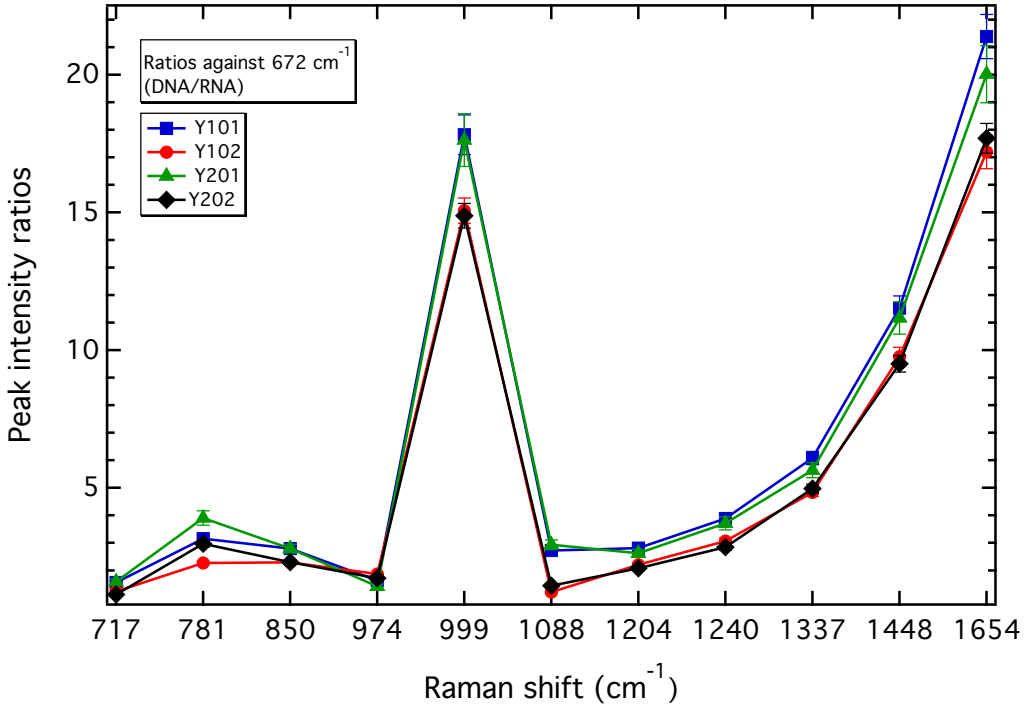
4.4 PIR analyses of the live MSC lines

The multivariate analysis was not able to fully discriminate the spectra of the four live MSC lines. However, it indicated that DNA/RNA vibrations and also to a background feature contributed to the majority of the variance between the cell lines. Thus, the univariate analysis by means of PIRs was applied to the fingerprint and high-wavenumber spectral regions of the dataset, as described in Section 2.6.1.

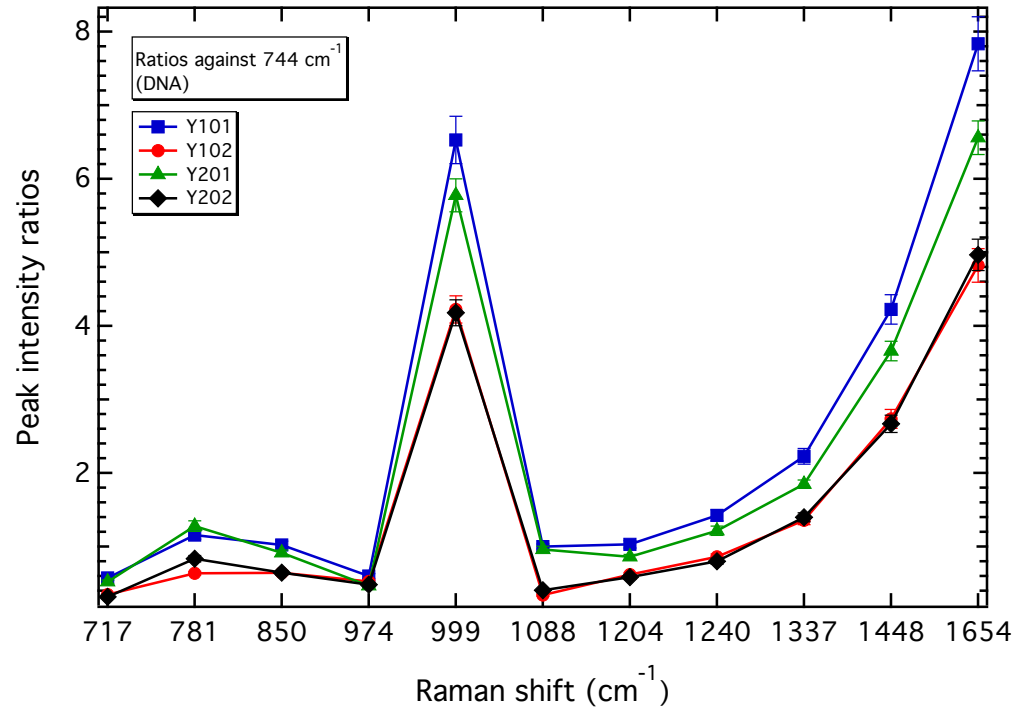
Although the PCA loadings revealed that the high-wavenumber vibrations also contributed to the dataset variance, the PIRs obtained against peaks assigned only to lipids, such as the 2850 cm^{-1}, 2930 cm^{-1}, 3007 cm^{-1} and 3062 cm^{-1} peaks, did not show discrimination between the cell lines (Appendix C, Figures C.1 and C.2). However, intensity ratios against peaks assigned to both lipids and proteins vibrations, such as the 2873 cm^{-1}, 2895 cm^{-1}, 2940 cm^{-1} and 2971 cm^{-1} peaks, provided discriminatory markers for 01s and 02s (identified by arrows in Figures C.3 and C.4, Appendix C). The high-wavenumber results indicate that the four MSC lines have a difference in proteins relative to other cellular components, and also that they display similar lipid characteristics.

In accordance to the PCA loadings of the fingerprint region, discriminatory markers were obtained against peaks related to DNA/RNA. Amongst them are intensity ratios

against the 672 cm^{-1}, 744 cm^{-1} and 781 cm^{-1} peaks, which are assigned only to vibrations of DNA/RNA (Figures 4.7 and 4.9a). Other discriminatory markers were obtained for PIRs calculated against the 1069 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA), 1088 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA), 1422 cm^{-1} (protein, lipids and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA) peaks, and shown from Figure 4.8 to Figure 4.10. For detailed peak assignments, please refer to Tables 3.1 and 4.1. These discriminatory PIRs are explained below and were organised between those that provided markers for differentiation competence, separating 01s from 02s, and between those that provided the discrimination of the four cell lines.

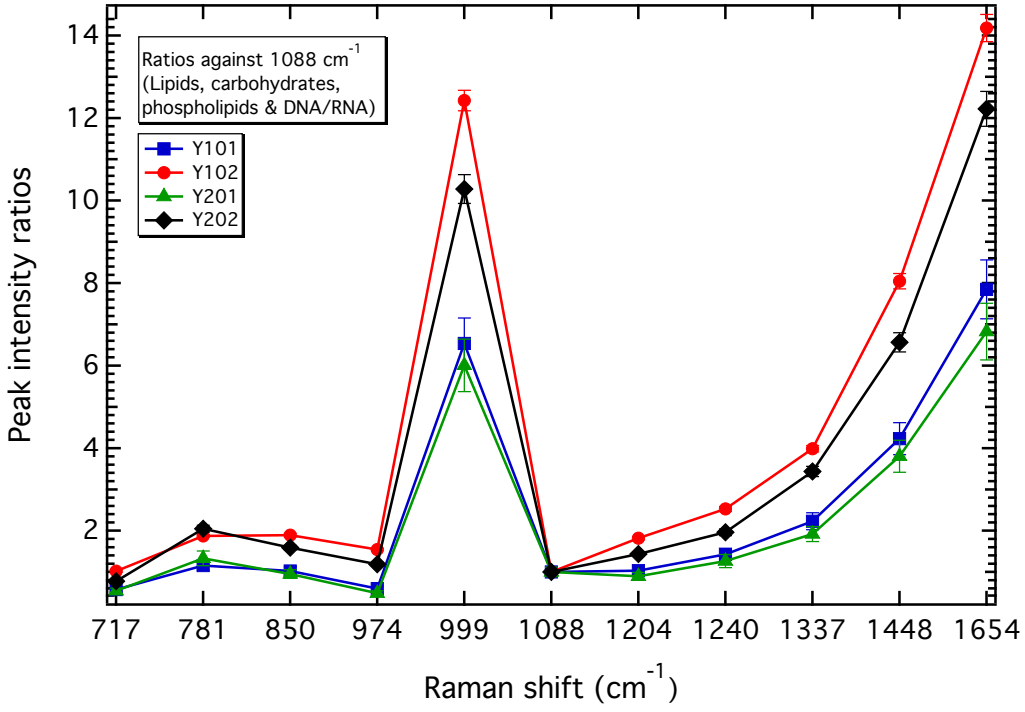


(a)

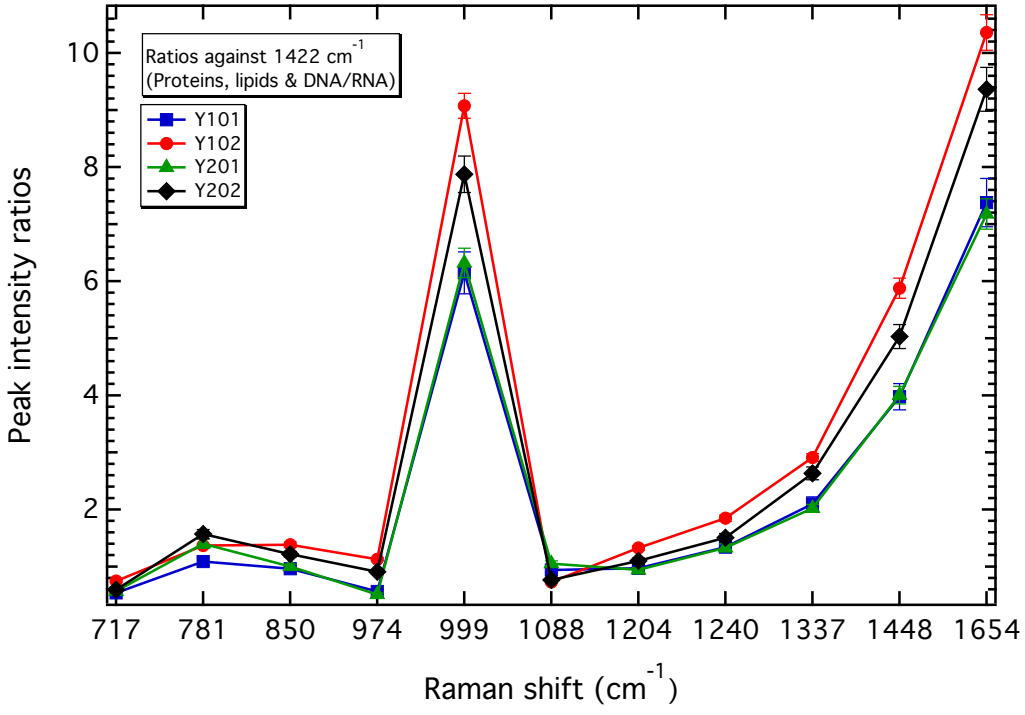


(b)

Figure 4.7: PIR comparison for the MSCs against the (a) 672 cm^{-1} and (b) 744 cm^{-1} peaks, showing the group separation between 01s and 02s cell lines. The error bars are the propagation of the fitted intensity SE.

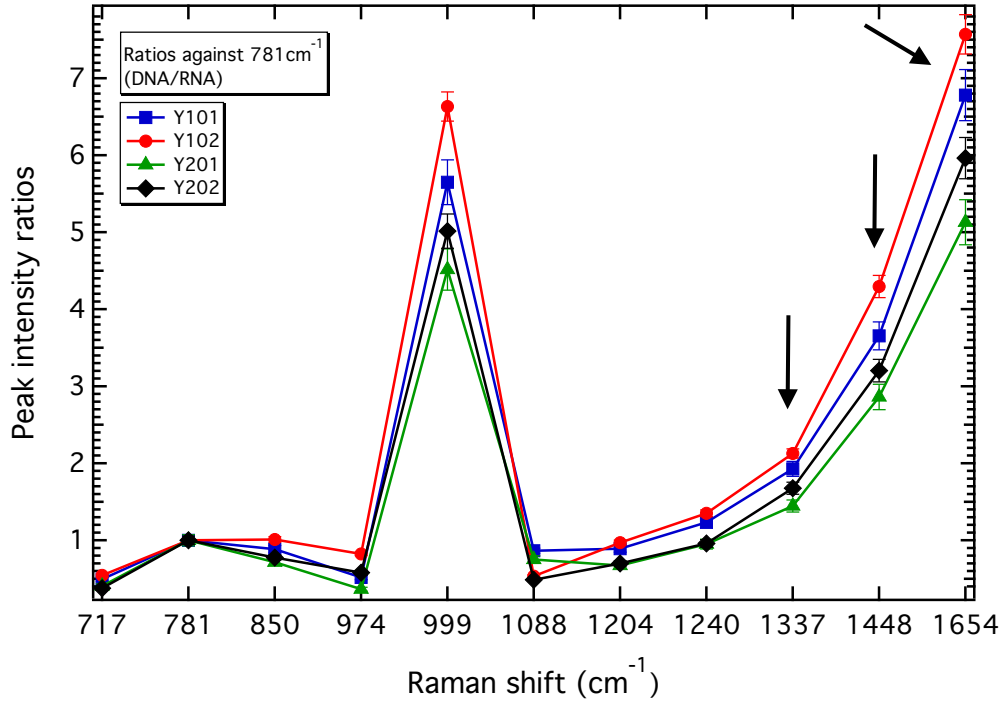


(a)

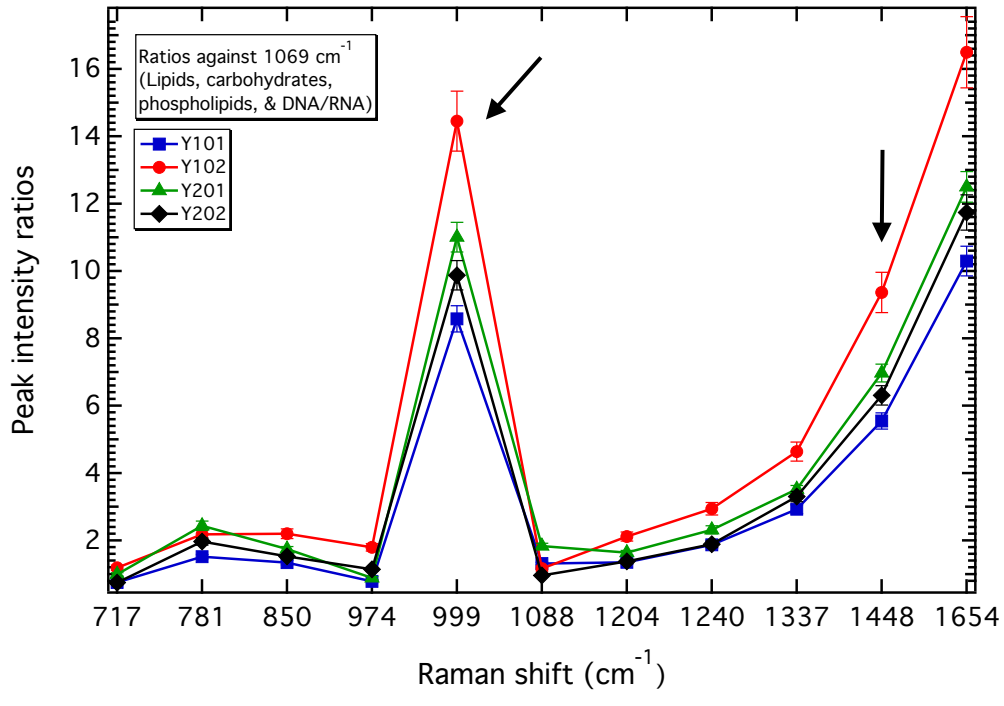


(b)

Figure 4.8: PIR comparison for the MSCs againts the (a) 1088 cm^{-1} and (b) 1422 cm^{-1} peaks, showing the group separation between 01s and 02s cell lines. The error bars are the propagation of the fitted intensity SE.



(a)



(b)

Figure 4.9: PIR comparison for the MSCs against the (a) 781 cm^{-1} and (b) 1069 cm^{-1} peaks, showing the presence of discriminating markers (indicated by the arrows) between all cell lines.

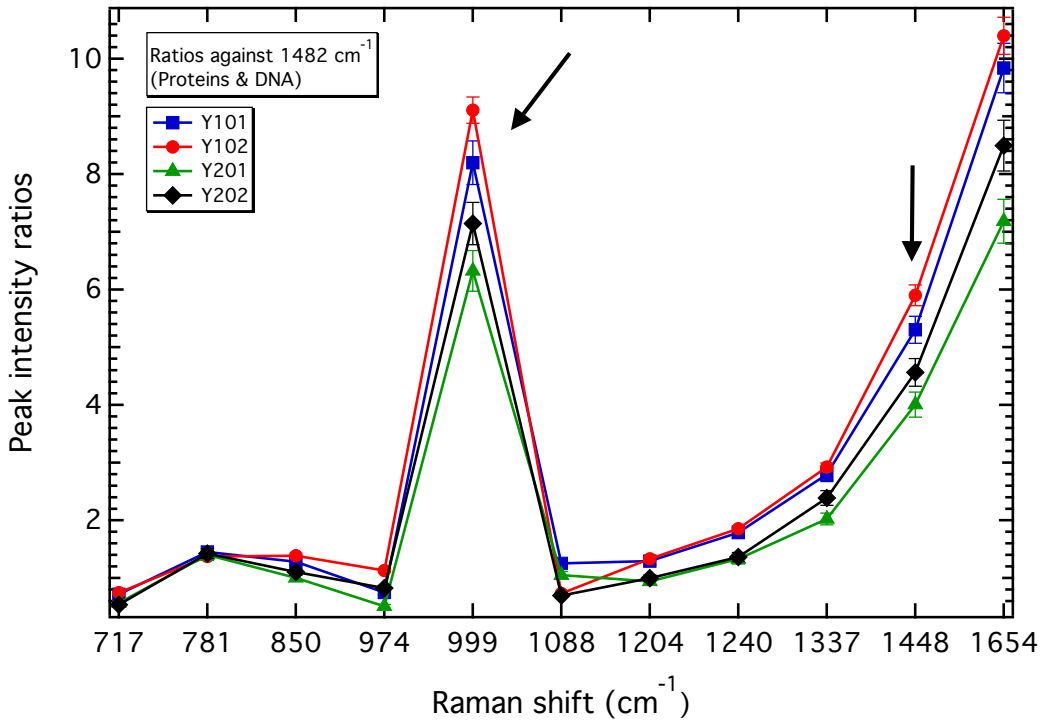


Figure 4.10: PIR comparison for the MSCs against the 1482 cm^{-1} peak, showing the presence of discriminating markers (indicated by the arrows) between all cell lines. The error bars are the propagation of the fitted intensity SE.

The discrimination of cells in relation to differences in DNA and RNA relative to other cellular components, such as lipids and proteins, has been verified by Chan *et al.* (2009) for human ESCs and the cardiomyocytes derived from them [36]. In their work, the lower signals of the DNA bands, at 785 cm^{-1} and 1090 cm^{-1} (normalised to the intensity of the protein and lipid peak at 1450 cm^{-1}), of the differentiated cells were used as an indicative of lower proliferation after the cells have developed into a more matured phenotype [36]. A decrease in the relative intensity of nucleic acids peaks during the process of differentiation was also used to discriminate differentiated from non-differentiated murine ESCs [55, 56]. Notingher *et al.* calculated a ratio between RNA and protein contributions, in the Raman spectrum, and related it to the degree of RNA translation in individual cells [56]. From this ratio, they concluded that the differences in the relative intensity between DNA

and proteins were indicating a low transcription of mRNA in the undifferentiated cells. Schulze *et al.* (2010), also on a study with human ESCs, proposed that intensity ratios of protein-related bands (757 cm^{-1}) against nucleic acid bands (784 cm^{-1}) were effective for discriminating non-differentiated from differentiated cells [33]. They also verified that the relative intensity of proteins increased with differentiation, whilst the relative intensity of nucleic acids decreased, reflecting the higher proliferation rate and also the bigger nucleus of the undifferentiated human ESCs. However, their results might have been jeopardised by combining live and air-dried cell spectra to obtain a larger dataset. Although all these studies were focused in non-differentiated and differentiated ESCs, they showed that the discrimination between these cells was obtained from differences in the relative intensities of peaks assigned to DNA, RNA and proteins. The discriminatory PIRs of 01s and 02s were obtained against peaks associated to DNA/RNA vibrations, such as the 672 cm^{-1} (DNA/RNA), 744 cm^{-1} (DNA/RNA), 1088 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1422 cm^{-1} (protein, lipids and DNA/RNA) peaks. These differentiation competence markers, or in other words, markers that segregated 01s from 02s are shown in Figures 4.7 and 4.8. The greatest separation between these two groups is observed for the 999 cm^{-1} (proteins) and 1654 cm^{-1} (proteins and lipids) peaks. These peaks have in common vibrations related to proteins; phenylalanine in the case of the 999 cm^{-1} and amide I for the 1654 cm^{-1}. Since there was no major difference in the relative intensity of the lipid peaks, demonstrated by the high-wavenumber PIRs, the separation obtained by the 1654 cm^{-1} peak was mainly driven by protein vibrations other than lipids. Thus, indicating that a difference in the DNA/RNA relative to proteins caused the separation between these two groups.

The PIRs calculated against the peaks at 781 cm^{-1} (DNA/RNA), 1069 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA)

(Figures 4.9 and 4.10) did not separate the MSC lines into two groups; however, they provided full cell line discriminatory Raman markers, which are indicated by arrows in the images. These four cell line discriminatory PIRs were obtained for the 999 cm^{-1} (phenylalanine), 1337 cm^{-1} (proteins and DNA/RNA), 1448 cm^{-1} (lipids and proteins) and 1654 cm^{-1} (lipids and proteins) peaks, which also have in common vibrations of protein molecules.

Once full cell line discrimination was observed in the PIRs, a deeper investigation including all spectral peaks was conducted against the previous ratios to verify if there were any other full cell line discriminatory Raman markers. Total cell line discrimination was obtained for PIRs against the 781 cm^{-1} (DNA/RNA), 1069 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA) peaks, and are shown from Figures 4.11 to 4.13. The ratios against the 781 cm^{-1} peak comprised the biggest set of discriminating PIRs. From these three sets of fully discriminatory Raman markers, it can be seen that the majority of them were obtained between intensity ratios of proteins and DNA/RNA peaks. These results further indicate that the discrimination of the four MSC lines is also pertained to differences in the DNA/RNA relative to proteins, as previously suggested.

Downes *et al.* (2011) suggests that the observed drop of the DNA/RNA levels relative to other cellular components during differentiation could mean either that their relative concentration is reducing in the nucleus, or that the nucleus-to-cytoplasm size ratio is decreasing during differentiation [68]. Taking this into account, it could be suggested that the 01s and 02s have different nuclei sizes relative to their cytoplasm. 02s have significantly greater area and perimeter compared to both Y101 and Y201 [84]. However, it is important to remember that the Raman data was collected solely over the nucleus, with a laser spot size of $\sim$0.7 μm. The measurement of the nucleus area through nuclear counterstaining

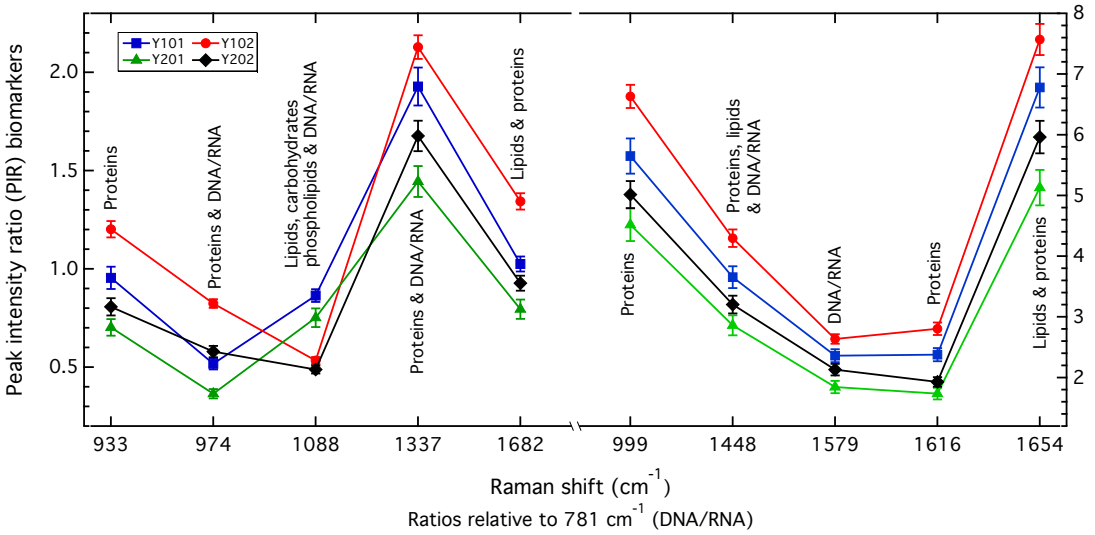


Figure 4.11: Expanded set of discriminatory PIRs for the MSCs against the 781cm^{-1} peak. Note that the graph is split into two sides to accommodate the different Y-axis ranges and allow the visualisation of the Raman markers of each MSC line. The error bars are the propagation of the fitted intensity SE.

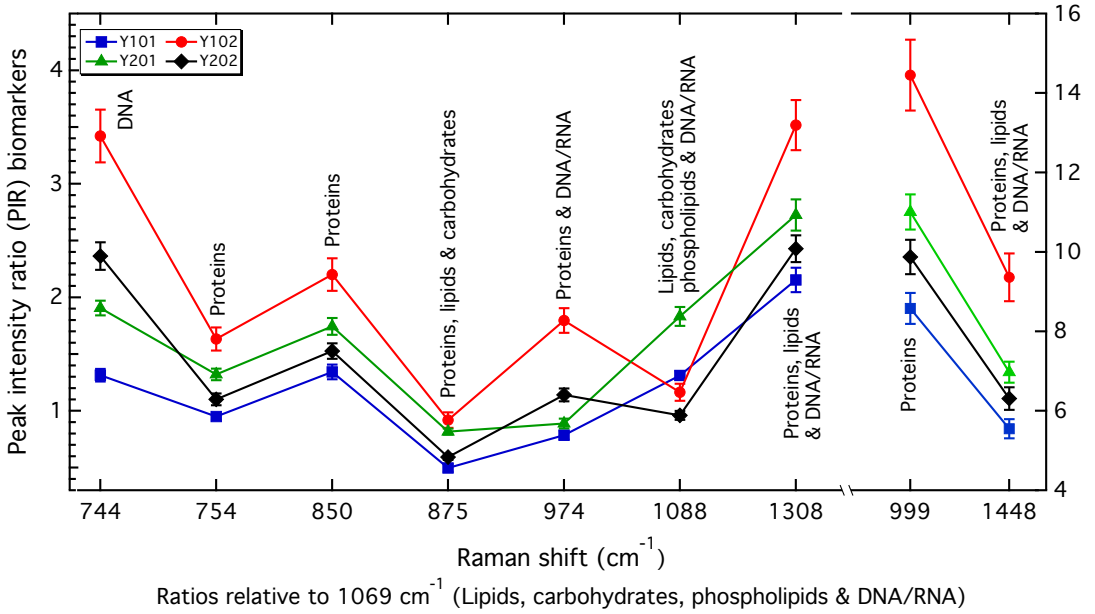


Figure 4.12: Expanded set of discriminatory PIRs for the MSCs against the 1069 cm^{-1} peak. Note that the graph is split into two sides to accommodate the different Y-axis ranges and allow the visualisation of the Raman markers of each MSC line. The error bars are the propagation of the fitted intensity SE.

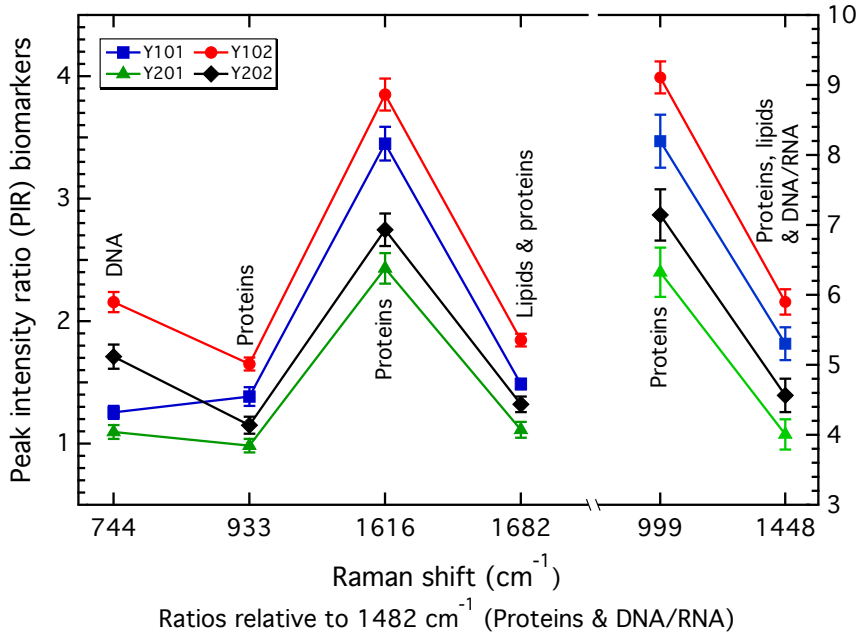


Figure 4.13: Expanded set of discriminatory PIRs for the MSCs agains the 1482 cm^{-1} peak. Note that the graph is split into two sides to accommodate the different Y-axis ranges and allow the visualisation of the Raman markers of each MSC line. The error bars are the propagation of the fitted intensity SE.

and image analyses (details of these methods in Section 2.8), did not show a significant difference in the nuclear size of the MSC lines, however, it revealed a consistent trend for larger nuclei in the 02's compared to the 01's (Appendix E, Figures E.1 and E.2). This suggests that these morphological differences between the two MSC groups could have also influenced their discrimination.

The biological properties of cells are encoded in the transcriptome, while the phenotype and function are determined by the proteins translated [184], for which the protein localisation is also associated with function [185]. Nuclear proteomics investigate the collective actions and interactions of proteins in the nucleus [186]. Usually these studies compare the nuclear proteins in undifferentiated and differentiated cells to understand their dynamics during cellular phenotype commitment [187]. Nuclear proteomics has identified differences in the nuclear concentration of several proteins, such as those involved in tran-

scription regulation, proliferation, pro-apoptosis and chromatin remodelling as function of ESC differentiation [186]. An increase in the expression levels of 81 proteins, together with the decreasing levels of other 21 proteins, has been noted before and after four days of osteogenic induction in hTERT MSCs by the expression profiling of membrane proteins [188]. Similarly to these proteomics studies that pointed to a different concentration of proteins in the nucleus of differentiated and non-differentiated ESCs, the RS studies performed in such cells also directed the discrimination of them towards a difference in proteins relative to DNA [33, 36, 55, 56].

Interestingly, a proteomics study of two undifferentiated MSC clones that express the same cell surface antigens, but that have different differentiation capacities and different proliferation rates, also demonstrated differences in the nuclear proteome of these two groups of cells [189]. In this work, Mareddy *et al.* (2009) found differences in the expression of proteins related to cell cycle, cell morphology, metabolism and proliferation [189]. They also showed that the trilineage competent cells over-expressed calcium-binding and actin-binding proteins, such as CAM and TM; the first regulates apoptosis and osteoclastogenesis, while the second is associated to the acceleration of cell division [189]. For the MSC lines here studied, the results obtained with RS, combined with PIR analyses, indicated a difference in the relative intensity of the DNA/RNA peaks to the other cellular components, mainly to proteins. Such difference allowed not only the discrimination between differentiation competent and incompetent cell lines (Figure 4.7 and 4.8), it also provided the separation of the four cell lines (Figure 4.11 to 4.13). Thus, it is probable that these cells lines have different protein characteristics. As discussed previously, the lower relative intensity of the DNA/RNA peaks has been attributed to lower proliferation rates of cells. Therefore, the relative difference in the DNA/RNA to proteins, obtained with the PIRs, may indicate that these cell lines have different proliferation rates, which has

engendered their discrimination. However, without further experiments, it is not possible to correlate these different proliferation rates to the contrasting differentiation capacities of the two discriminated groups or to differences between the cell lines.

Comparing the set of markers obtained for live the cells against those obtained for air-dried MSCs (Section 3.3), it is clear that the discrimination in the live cells was driven by differences in DNA/RNA relative to proteins. Amongst the air-dried set of markers, only three discriminatory ratios were obtained against peaks that have vibrations of the DNA/RNA molecules. Such ratios were against the 970 cm^{-1}, 1060 cm^{-1} and 1085 cm^{-1} peaks. The 970 cm^{-1} peak was the main actor in the full cell line discrimination obtained for the air-dried MSCs, while the other two peaks provided group separation. However, for the live MSCs spectra, the 970 cm^{-1} peak (observed at the position 974 cm^{-1}) is hardly distinguished from the background, and was also not evident in the PCA loadings (Figure 4.5). The intense feature of the 970 cm^{-1} peak in the air-dried samples could be the result of increased phosphorylation of proteins as result of the air-drying process and possibly cell death, as it is known that protein phosphorylation has a regulatory role in cell death mechanisms [173] and it is increased during desiccation [172]. Other ratios that separated 01s and 02s, in the air-dried experiments, were obtained against the 1549 cm^{-1} and 1615 cm^{-1} peaks, assigned to tyrosine and tryptophan, which are also common proteins involved in cell death pathways [174]. Thus, the PIRs obtained in Chapter 3 obtained against peaks at 970 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1}, and that provided discrimination between the cell lines, could have been directly affected by air-drying process.

The discriminatory markers obtained against DNA/RNA peaks, for the air-dried cells, indicate that although they might reflect the changes associated to the air-drying process, they still retained some of the differences in DNA/RNA relative to proteins that were observed for the live cells. This explains the similar group discrimination obtained by PIRs

against peaks ∼1060 cm^{-1} and ∼1085 cm^{-1} in both experiments. Other ratios obtained for the air-dried cells, and not related to DNA/RNA vibrations, suggest that these cells might modulate cell death differently, or that the air-drying process enhanced their differences. Nonetheless, it is clear that the set of PIRs obtained for the live cells, once not affected by biomolecular changes driven by air-drying process, allowed the identification of many more markers for full cell line discrimination, evidencing the robustness of RS for subpopulation discrimination.

4.5 Population heterogeneity: convergence tests and PIRs of the live MSC lines

The convergence tests used to determine if the number of obtained spectra represented the population average, considering a convergence of the statistical quantities, can also be used to evaluate the heterogeneity of such a cell population. Thus, similarly to what was done in Sections 3.4.1 and 3.4.2 of Chapter 3, in this section the convergence tests are also used for this purpose and are applied to each MSC line, throughout the three experimental repeats.

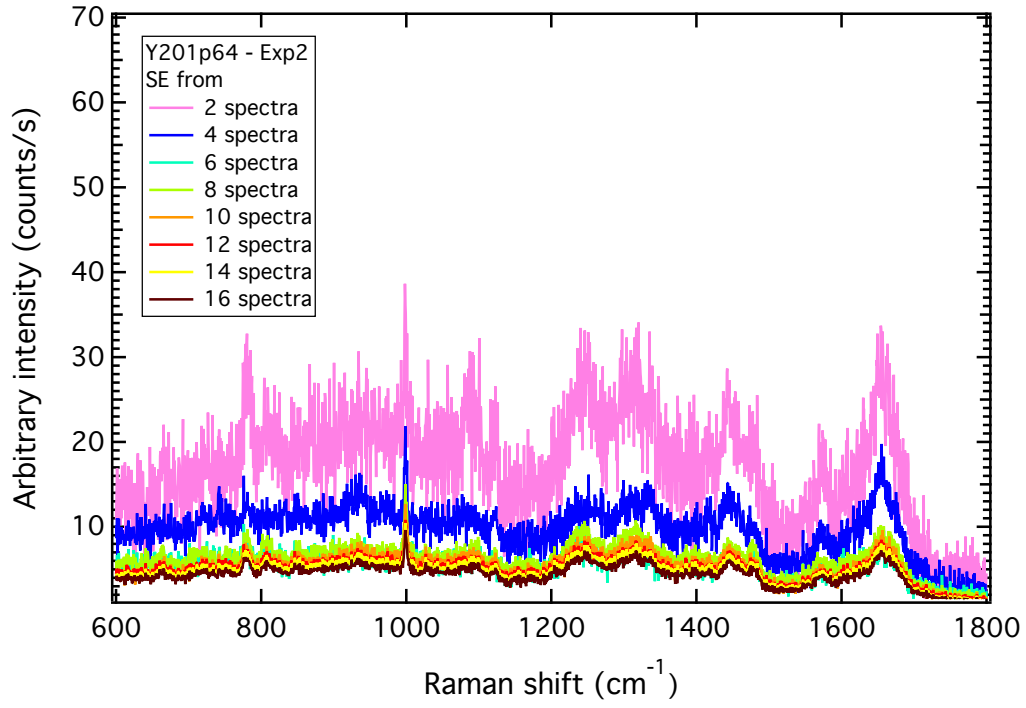
Full spectral convergences of the SE and 2x SD were obtained for averages containing an increasing number of spectra, as described in Section 2.6.2. These results are shown in Appendix B, in Sections B.5 and B.6. Examples of the SE and 2x SD full spectral convergences are shown in Figure 4.14. From the Figure 4.14a it can be seen that the SE decreases to a minimum error value with the increasing in the number of averaged spectra. The 2x SD (Figure 4.14b) oscillates between a mid point, with its amplitude decreasing as the number of spectra increases. When compared to the 2x SD convergence tests obtained from the air-dried experiments (Figure 3.13), the 2x SD obtained for the

live cells displayed less dispersion from cell to cell spectrum (Figure 4.14b).

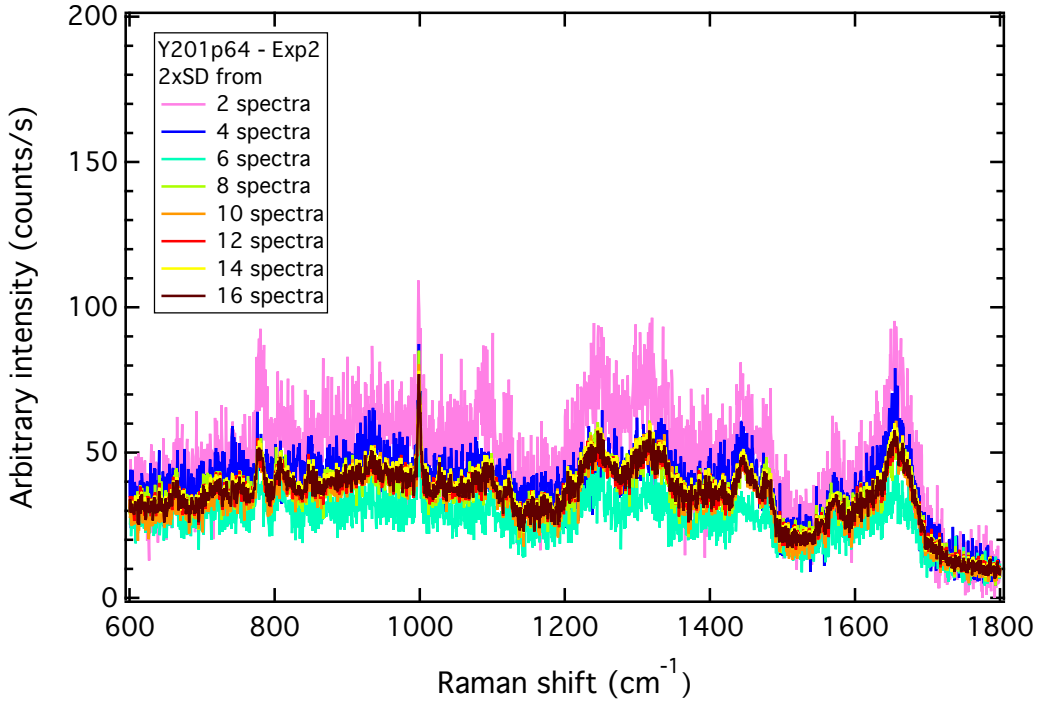
The method to obtain the convergence of the %SE associated with the 970/1085 and 1656/1085 PIRs was described in Section 2.6.2. Both PIR convergences are used in this analysis, so that they can be compared to the air-dried results. The relative intensity of the 970 cm^{-1} peak, assigned to phosphorylated proteins and DNA/RNA, varied considerably between the air-dried samples, possibly reflecting the changes associated to the air-drying process, while it displayed a constant low relative intensity for the live cells. As per the 1656/1085 PIR, it discriminated differentiation competent from incompetent MSC lines on both air-dried and live cell analyses. It is important to remember that for the live cell analysis, there was a small shift in the frequency of these peaks as described on Table 4.2. Thus, for the live cell results they are represented by the positions at 974cm^{-1}, 1088 cm^{-1} and 1654 cm^{-1}.

For both ratios (974/1088 and 1654/1088) the %SE was calculated and plotted against the number of averaged spectra. The individual results for each cell line and experimental repeat are shown on Appendix B, Section B.4. Each graph was fitted with an exponential curve (Equation A.2.2 from Appendix A), from which the decay constant (τ) and the converged %SE value (y_0) were used to represent the heterogeneity associated to each population/experiment. Examples of both PIRs %SE convergence tests are shown in Figure 4.15. While for the air-dried cells the %SE convergence tests indicated a convergence of the %SE from ∼50 spectra onwards (Appendix B.1), for the live cells, the %SE convergence was obtained from ∼10 spectra (Appendix B.4).

Table 4.4 shows the averaged decay constant of each cell line, obtained for each PIR, across the different experimental repeats. A smaller decay constant is associated to a rapid decay of the exponential curve, which means that the statistical quantity measured (%SE) converges with a smaller quantity of spectra. It can be seen that independently

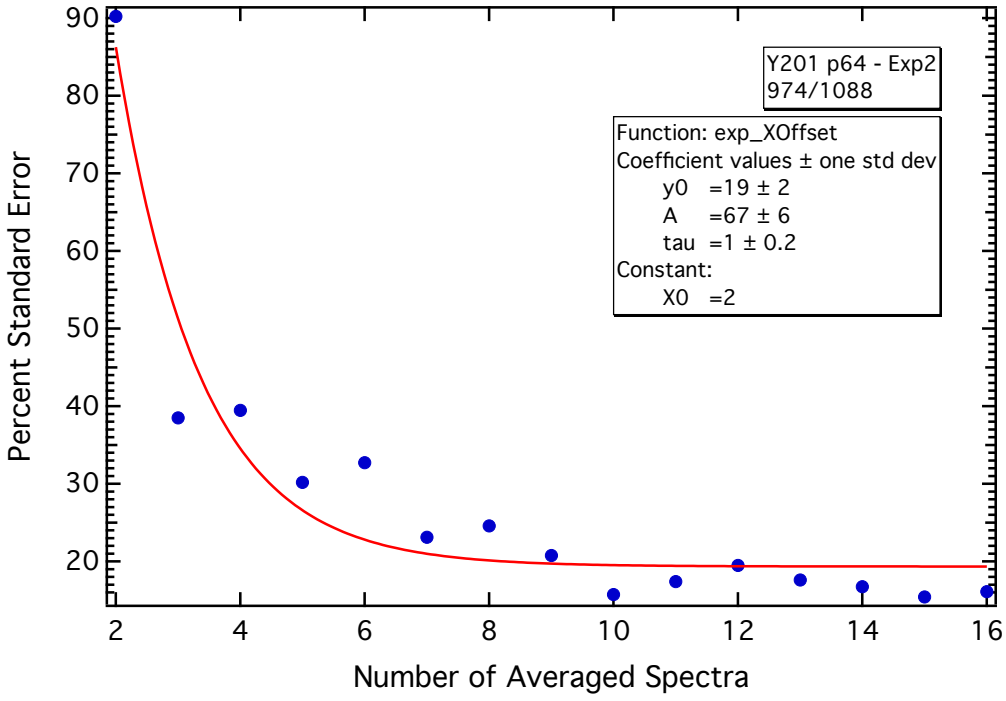


(a)

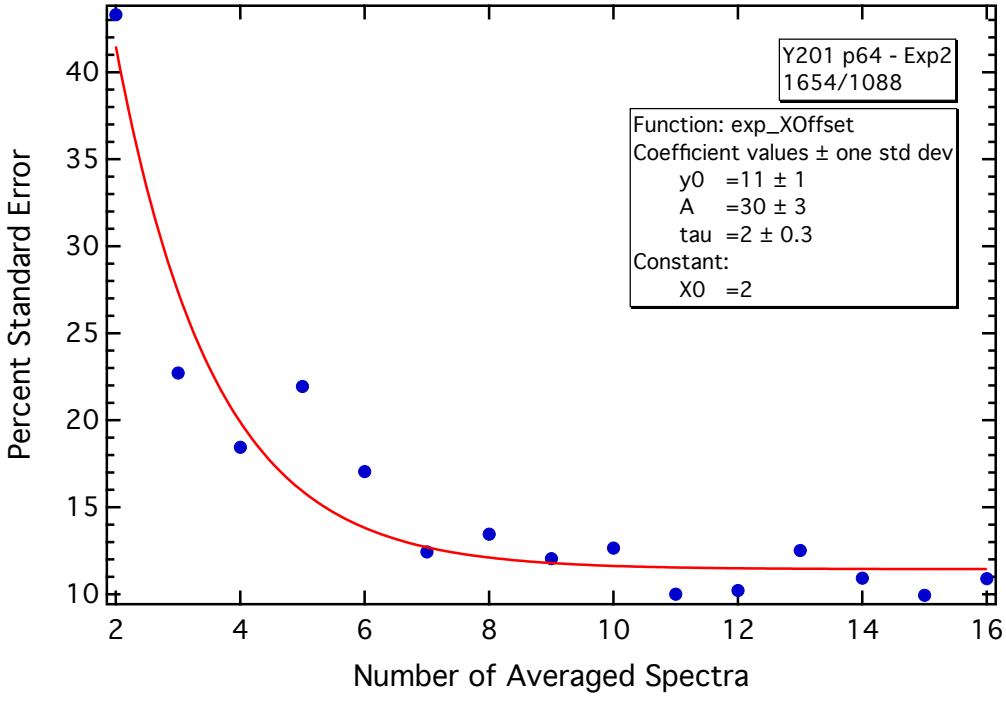


(b)

Figure 4.14: Example of the (a) SE and (b) 2x SD full spectral convergence test. These examples show the convergence of both SE and 2x SD, associated to the peak intensities, with the increasing number of spectra in the spectral average.



(a)



(b)

Figure 4.15: Examples of %SE convergence tests for the (a) 974/1088 and (b) 1654/1088 PIRs. They show the decay of the %SE associated with the PIR as the number of averaged spectra increases. The red line represents the exponential fitting.

of the PIR analysed all four MSCs displayed similar decay constants and converged %SE, which reflects the greater homogeneity of these live cell populations.

When the τ obtained for the air-dried cells (Table 3.3) is compared to the τ from the live cells, it can be noted that they are much lower for the live cells. The averaged decay constants of the air-dried cells ranged from 6.5 to 12.5 for both PIRs, while those obtained for the live MSCs ranged from 0.9-2.0, for both ratios. This suggests that most of the variability observed in the air-dried populations was linked to changes associated to the air-drying process. Interestingly, the average of the converged %SE was similar for both datasets, ranging from 7.0 to 12.5 for the air-dried cells and from 6 to 13.7 for the live cells. This indicates that many more spectra per experiment were necessary, in the air-dried experiments, to achieve the same level of uncertainty of the live cell experiments. These results also show the robustness of sampling live cells and are in agreement with the variance obtained from the full spectral 2x SD graphs (Appendix B, Section B.6).

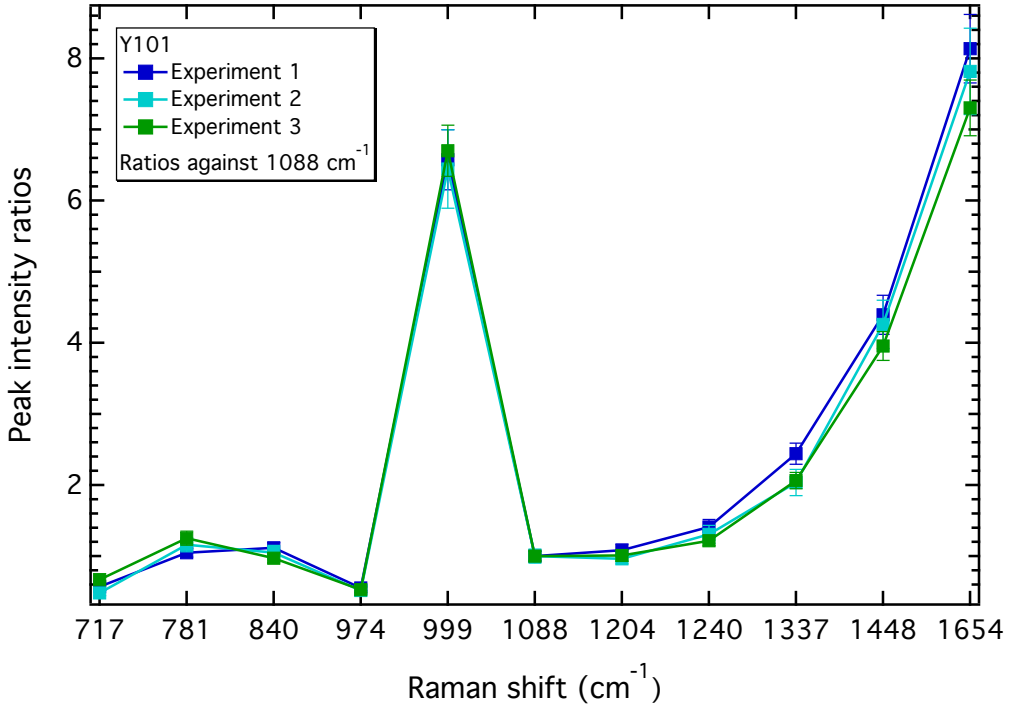
In order to assess if there were pronounced biochemical differences captured by RS between the experimental repeats performed for the live cell analyses, the intensity ratios against the 1088 cm^{-1} peak (lipids, carbohydrates, phospholipids and DNA/RNA) were calculated for each experiment. For this, the spectra obtained in each repeat was averaged

Table 4.4: Average of decay constants and converged %SE obtained for the 974/1088 and 1654 PIRs for the live cell populations analysed. Three experimental repeats were performed for each cell line. A smaller decay constant is related to a rapid decay of the %SE. The maximum and minimum values obtained are shown in the brackets.

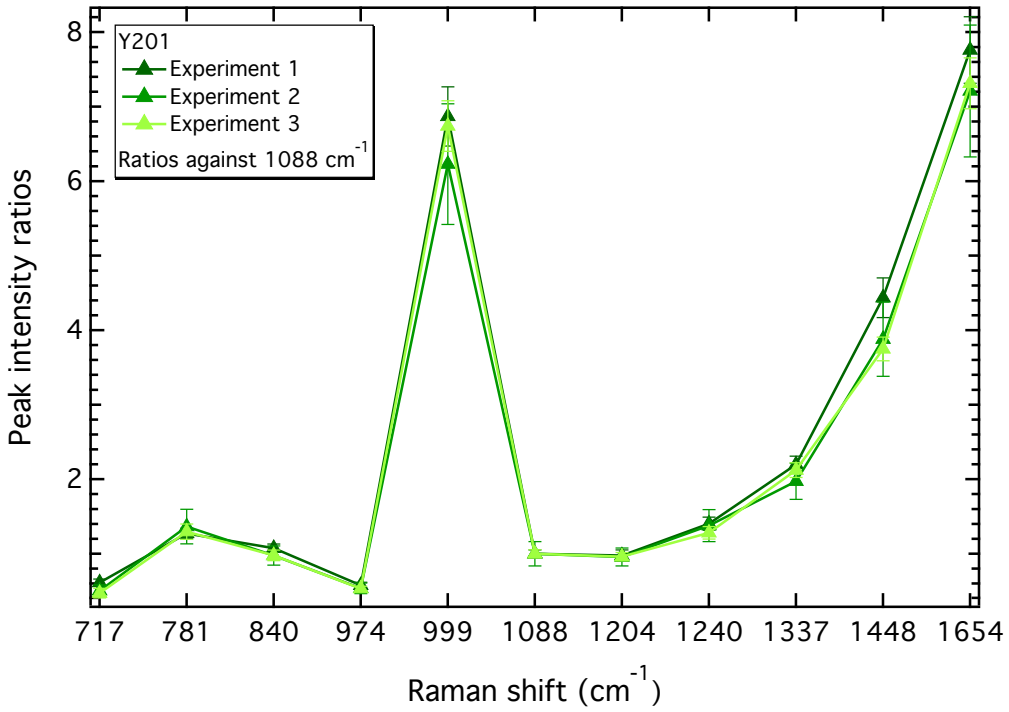
	PIR 974/1088		PIR 1654/1088	
Sample	Averaged decay constant	Converged %SE	Average decay constant	Converged %SE
Y101	1.3 (0.8–2.2)	8 (6.1–10)	1.3 (0.7–2.1)	7.7 (6.4–9.7)
Y201	1.3 (0.8–2.3)	13.7 (9–21)	1.7 (0.9–2.3)	8 (4.4–12)
Y102	1.3 (0.8–2.2)	7.3 (5.2–9)	2.0 (0.9–3.3)	6 (1.9–9.7)
Y202	1 (0.8–1.2)	9.7 (7.3–12.4)	0.9 (0.8–1.2)	8.3 (6.5–10.8)

and fitted, such as described in Section 2.6. As explained earlier, this PIR demonstrated group discrimination between differentiation competent and incompetent MSC lines on both air-dried and live cell analyses.

These results are shown in Figures 4.16 and 4.17, where the error bars are the propagation of the SE over the calculated PIRs. The images show that apart from a small variation in the 781/1088 PIR for Y202 all other markers remained nearly constant for all experimental repeats, showing only fluctuations no bigger than their uncertainty bars. These results contrast with those obtained for the air-dried cells (Figures 3.15 and 3.16), which showed an increased variation for the relative intensity of the 970 cm^{-1} peak. Therefore, these results attest the reproducibility of the obtained markers and the robustness of live cell analyses.

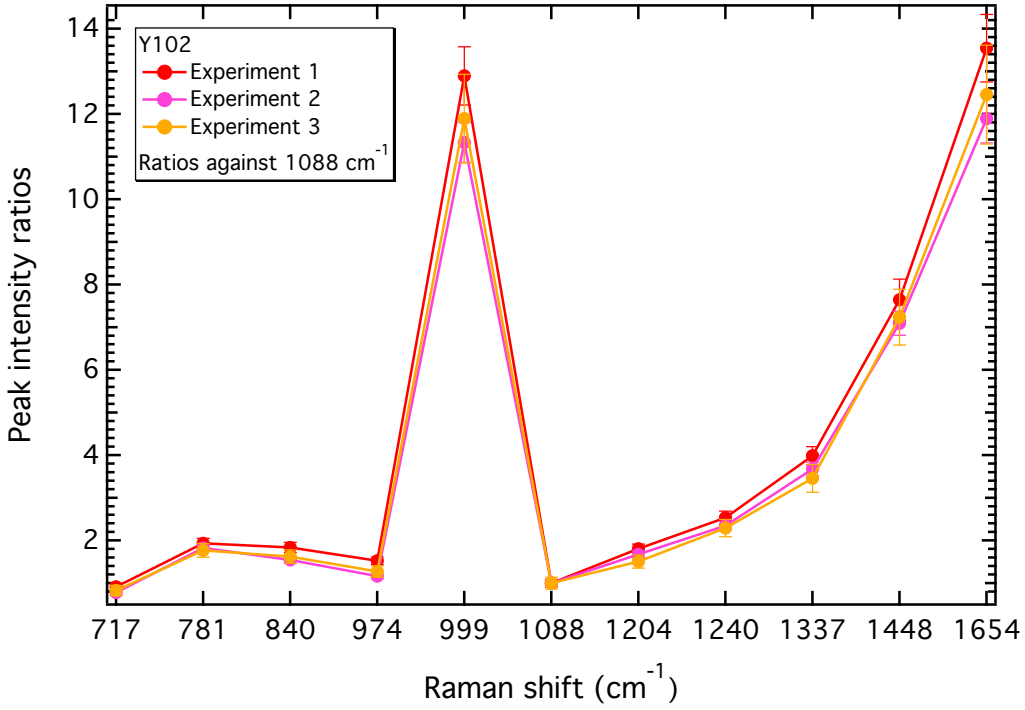


(a)

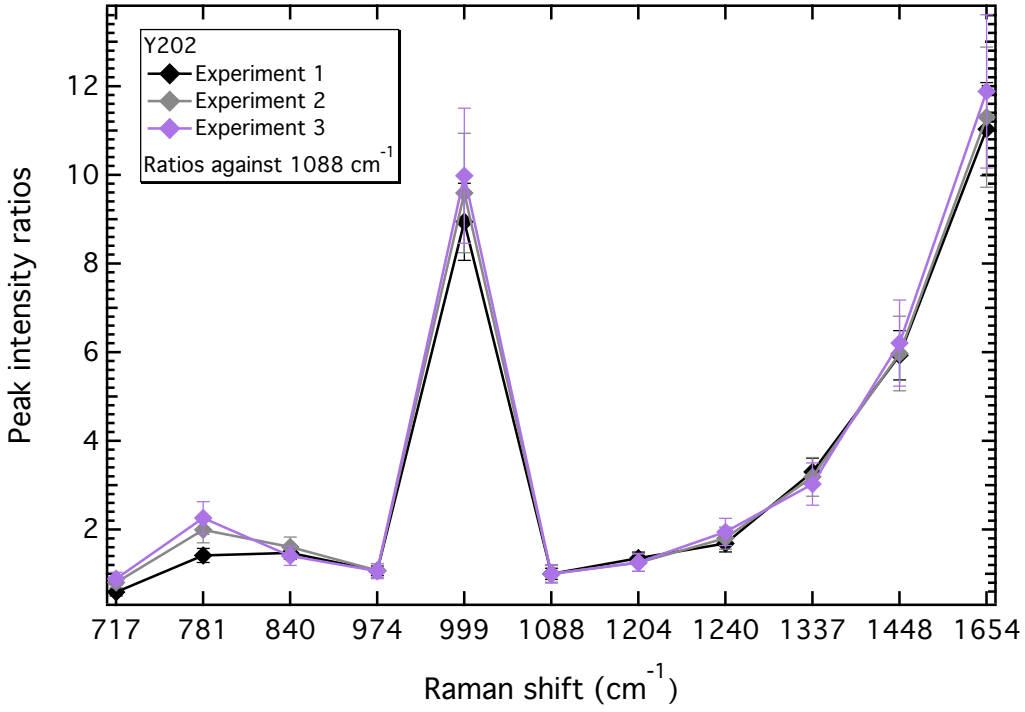


(b)

Figure 4.16: Comparison between the PIRs against the 1088 cm^{-1} peak for the experiment repeats performed for the live (a) Y101 and (b) Y201 MSC lines. The error bars are the propagation of the fitted intensity SE.



(a)



(b)

Figure 4.17: Comparison between the PIRs against the 1088 cm^{-1} peak for the experiment repeats performed for the live Y102 and Y202 MSC lines. The error bars are the propagation of the fitted intensity SE.

4.6 Conclusions

The main objective of this chapter was to verify if discriminatory Raman markers could be obtained from the analyses of live MSCs. The chapter also aimed to understand if the PIRs obtained for the air-dried cells reflected the differences of DNA/RNA relative to protein in the MSC lines and how much these results were affected by the air-drying process. A trypan blue assay assured that the laser exposure did not cause any drastic change to cell membrane and that the spectra obtained from them represented the biomolecular composition of living viable cells.

The analyses of live cells rendered not only the discrimination of the MSCs into groups related to their trilineage differentiation potential, as it also provided, by means of the PIR analyses, Raman markers for the discrimination of the four cell lines. All discriminatory Raman markers found by the PIR analyses, for the live cells, were obtained from intensity ratios against DNA/RNA peaks.

Sets of group discriminatory markers, related to trilineage differentiation capacity, were obtained against the peaks at 672 cm^{-1} (DNA/RNA), 744 cm^{-1} (DNA/RNA), 1088 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1422 cm^{-1} (proteins, lipids and DNA/RNA). Sets of discriminatory Raman markers of the four cell lines were obtained against the 781 cm^{-1} (DNA/RNA), 1069 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA) peaks, and show the potential of RS to discriminate these subpopulations. Both sets of markers indicated that differences in the DNA/RNA relative to protein could be driving the cell line discrimination. The multivariate analyses corroborated these findings, indicating that most of the variance in the live cell dataset was produced by the DNA/RNA and protein peaks. The PCA also associated the variation in the dataset with the Raman background of the analysed cells. Raman background has been associated, in the literature, to phenotypical differences

among the cells.

Differences in the intensity of DNA/RNA peaks relative to protein peaks have been associated, in the literature, to different proliferation rates and therefore used to discriminate cells. However, differences in DNA/RNA relative to proteins could also be associated to the slight different nuclear sizes of the MCS lines. In the literature, differences in the nuclear proteome of differentiation competent and incompetent cells were linked to the expression of proteins related to cell cycle, cell morphology, metabolism and proliferation. Thus, it is clear that RS discriminated the MSC lines based in the differences in DNA/RNA relative to proteins. However, it is necessary to conduct further experiments to confirm if these differences are correlated to a different proliferation rate of the MSC lines or to differences in their nuclear proteome and how they translate into the contrasting differentiation capacities.

In contrast to the air-dried spectra, the data obtained from the live cells represented more homogeneous populations as shown by the convergence tests. Live cell analyses of different experimental repeats did not suffer from major fluctuations in the relative intensity of the PIRs calculated, resulted in more constant datasets and reinforced the robustness of live cell analyses. The 01s and 02s discriminatory PIRs against the $\sim$1069 cm^{-1} and $\sim$1088 cm^{-1} peaks were obtained for both air-dried and live cell experiments. This indicates that, although the air-drying process affected the Raman spectra, the air-dried cells still retained some of the relative differences between DNA/RNA and proteins that were observed for the live cells.

In summary, RS is a label-free method able to identify markers linked to MSC function, thus providing the discrimination between trilineage differentiation competent and incompetent cell lines. The identification of discriminatory Raman markers of the four undifferentiated live MSC lines reinforces the capacity of RS for characterising and dis-

criminating subpopulations, as these cells lines were not fully distinguished by standard methods, such as global gene expression. Therefore, RS is an effective technique for cell characterisation that can be used in tandem with standard methods for understanding the biomolecular differences across cell populations, and to identify and select different, but closely related, cell subtypes.

Chapter 5

Raman spectroscopy assessment of osteogenic and adipogenic differentiation of mesenchymal stromal cells

The previous two chapters showed that RS can distinguish the subpopulations of differentiation competent and incompetent MSCs. It also provided an insight into the biomolecular differences between these cell lines, providing markers for their full discrimination. In this chapter, the osteogenic and adipogenic differentiations of the Y201 cell line are characterised with RS. This cell line was chosen because it shows both osteogenic and adipogenic competence, whereas Y101 spontaneously differentiate into osteoblasts and has poor adipogenic capacity [84].

For the osteogenesis, the aim is to verify if RS can identify markers other than the presence of hydroxyapatite. Although, parallel to it, the mineral produced by the cells was

also assessed and compared to human and lamb bones. For the study of adipogenesis, the chapter focuses on characterising the adipo-induced MSCs since, to date, this was mainly done for adipose-derived stem cells. Finally, it also attempts to identify the lipid species stored in the lipid droplets from the adipocytes.

5.1 Population heterogeneity: convergence tests during osteogenesis and adipogenesis

RS was used to assess the Y201 MSC line cultured in basal, osteogenic and adipogenic media at days 0, 7, 14 and 21, as explained in Section 2.2.4 of Chapter 2. As indicated by the convergence tests presented in Section 4.5, the Raman analyses of live cells allowed that a lower number of spectra were collected in comparison to air-dried analyses. Those convergence tests also indicated that the air-dried cell populations were less homogenous, possibly due to biomolecular changes associated to the air-drying process. Thus, Raman data acquisition was performed on live cells kept in HBSS, with one spectrum collected per cell. Details of the acquisition parameters are described in Section 2.3. The convergence tests were applied to the spectra obtained from the basal, osteogenic and adipogenic samples, in order to check if the averaged spectra were representative of the population average.

Expressive changes are expected in the spectra during osteogenesis and adipogenesis as confirmed by literature [126, 138]. Therefore, peaks that would not be hindered by other prominent peaks appearing during the differentiation were chosen to compose the %SE convergence test associated with the PIRs. For example, the 974/1088 PIR was not used because, during osteogenesis, the presence of the hydroxyapatite peak at ∼960 cm^{-1} could obscure the peak at 974 cm^{-1}. Therefore, for the osteo-induced samples,

the 1654/1088 PIR allows the comparison of non-differentiated cells at day 0 with the cells at day 21, when mineralisation is expected to have happened [84]. This PIR is associated to vibrations of lipids and proteins (1654 cm^{-1}) and the peak 1088 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA). For the adipo-induced samples, the chosen PIR was the 1654/1448 once it contains the vibrations of the CH_2 (1448 cm^{-1}) and C=C (1656 cm^{-1}) groups present in lipids and would also allow the comparison of the non-differentiated MSCs with the adipocytes. For the samples kept in basal condition, both PIRs were used to allow the comparison of the basal cells with the osteo- and adipo-induced MSCs. The graphs of each %SE convergence test, SE and 2SD stacks are displayed in Appendix B; Sections B.7, B.8 and B.9 contain the tests for the cells kept in basal medium; Sections B.10, B.11 and B.12 display the tests for the osteo-induced cells; and Sections B.13, B.14 and B.15 show the convergences of the adipo-induced cells. Examples of these convergence tests are shown in Figures 5.1 and 5.2.

Full spectral convergences of the SE and 2x SD were obtained for the analysed samples as exemplified by Figure 5.1. It can be seen that the SE associated to the peak intensities converges to a minimum value as the number of obtained spectra increases (Figure 5.1, images a, c and e). As per the 2x SD, it oscilates between a mid point and stabilises (Figure 5.1, images b, d, and f).

Table 5.1 summarises the %SE convergence test results obtained for each experiment. Examples of these tests are shown in Figure 5.2 for basal, osteo- and adipo-induced cells at day 21. Each %SE convergence test was fitted with an exponential curve (Equation A.2.2 in Appendix A), from which the decay constant (τ) and the converged %SE value (y_0) values were used to represent the heterogeneity associated to each experiment. Although the cells kept in basal, osteogenic and adipogenic media were going through different processes during the time-points measured, they have similar decay constants, which indicates that

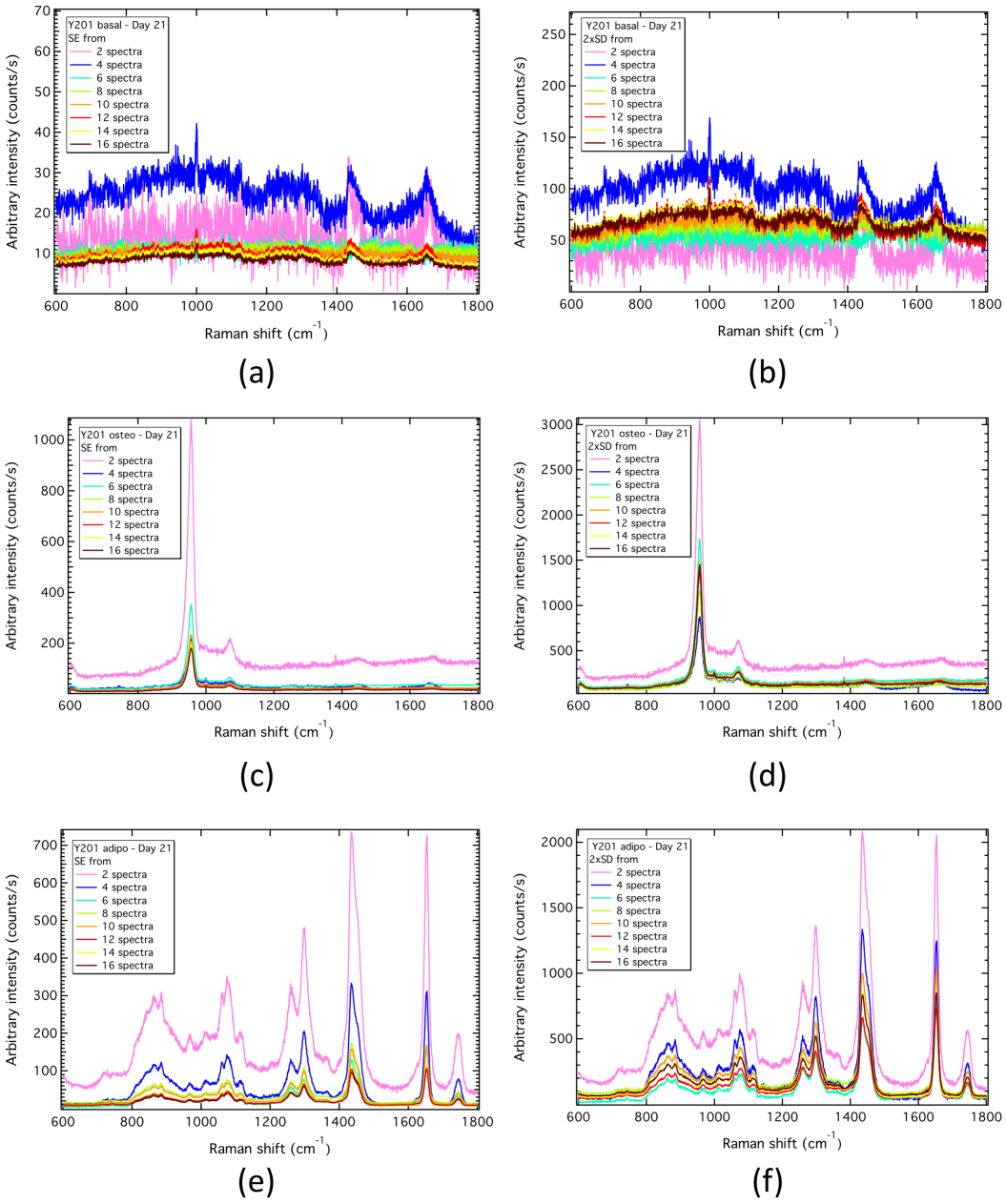


Figure 5.1: Examples of the SE (left column) and 2x SD (right column) convergence tests for the basal (a, b), osteo- (c, d) and adipo-induced (e, f) cells. All graphs show the results obtained at day 21 for the fingerprint region. For the complete set of results please see Appendix B, Sections B.11 and B.12 for the osteogenic convergence tests, and Section B.14 and B.15 for the adipogenic convergence tests.

their uncertainties converged at the same rate. However, as indicated by the %SE the cells that underwent differentiation process, either osteogenic or adipogenic, showed higher converged %SE values for days 7, 14 and 21 in comparison to the cells kept in basal

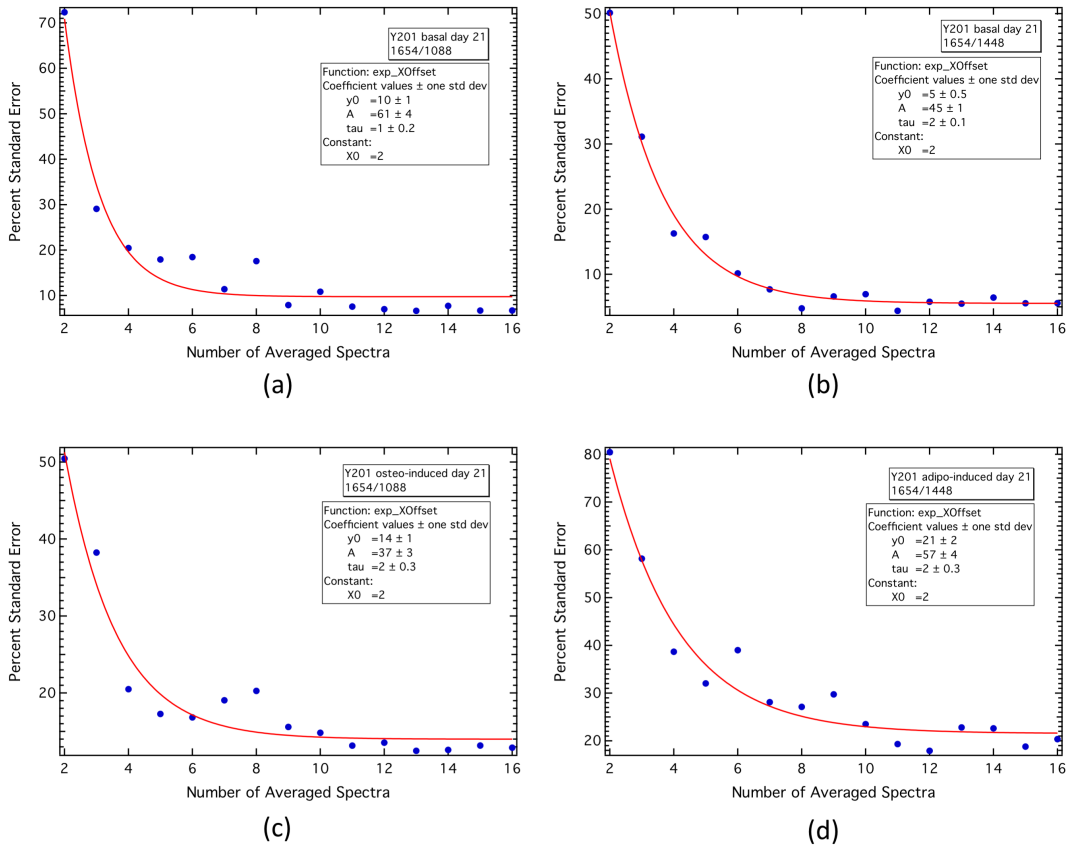


Figure 5.2: Examples of the %SE convergence tests for the basal, osteo- and adipo-induced cells. For the cells kept in basal medium, the %SE convergence test was obtained for the (a) 1654/1088 and (b) 1654/1448 PIRs. Graphs (c) and (d) show the %SE convergence tests for the osteo-induced cells (1654/1088 PIR) and adipo-induced cells (1654/1448 PIR), respectively. The red line represents the exponential fitting. All graphs show the results obtained at day 21 for the fingerprint region. For the complete set of results please see Appendix B.

medium. This higher %SE values reflect the diversity of such populations. The osteogenic samples, for example, presented mineralised nodules with different sizes, shapes, thickness, and crystallinity that varied across the time-points. The lipid droplets were also observed in varied sizes. This is further discussed in Sections 5.2 and 5.3.

Although these converged %SE values are higher if compared to those of the cells cultured in basal medium, the %SE has converged for the experiments, as seen from their graphs (Appendix B, Sections B.7, B.10 and B.13) and confirmed by the full spectral

SE convergences (Appendix B, Sections B.11 and B.14). Therefore, the averaged spectra reflects the population average. The 2x SD graphs (Appendix B, Sections B.12 and B.15) also show that the dispersion of the spectra obtained from these populations reflected the varied mineral content in the osteogenic samples, and the relative increase of the lipid content during adipogenesis.

Table 5.1: Convergence tests decay constants for basal, osteo- and adipo-induced MSCs. The decay constant and the converged %SE obtained for the PIRs 1654/1088 and 1654/1448 are presented for the different experiments performed. A smaller decay constant is related to a rapid decay of the error.

	PIR 1654/1088		PIR 1654/1448	
Time-point	Decay constant	Converged %SE	Decay constant	Converged %SE
Basal cells				
Day 0	2 ± 0.3	10 ± 2	2 ± 0.4	12 ± 2
Day 7	2 ± 0.3	12 ± 1	1 ± 0.2	11 ± 1
Day 14	1 ± 0.2	10 ± 1	1 ± 0.2	9 ± 2
Day 21	1 ± 0.2	10 ± 1	2 ± 0.1	5 ± 0.5
Osteo-induced cells				
Day 0	3 ± 0.3	15 ± 2	–	–
Day 7	2 ± 0.2	25 ± 1	–	–
Day 14	2 ± 0.2	21 ± 0.6	–	–
Day 21	2 ± 0.3	14 ± 1	–	–
Adipo-induced cells				
Day 0	–	–	2 ± 0.3	12 ± 1
Day 7	–	–	2 ± 0.2	29 ± 1
Day 14	–	–	2 ± 0.5	17 ± 2
Day 21	–	–	2 ± 0.3	21 ± 2

5.2 Osteo-induced MSCs

When cultured in osteogenic medium, the MSCs differentiate into osteoblasts, responsible for bone matrix synthesis and subsequent mineralisation, which is done by the accumulation of calcium deposits in the extracellular matrix [190]. Eventually, some of the

osteoblasts get trapped in the bone matrix, giving rise to osteocytes, which are the most abundant cells in bone [190].

The ossification process, or osteogenesis, is carried out by two processes: (i) the intramembranous ossification and the (ii) endochondral ossification [190, 191]. The intramembranous ossification occurs during the formation of flat bones and is an essential process during natural healing of bone fractures [190]. In this process, the MSCs (osteoprogenitor cells) proliferate, condense and differentiate into osteoblasts, which then start the production of bone matrix and become osteocytes when eventually surrounded by the collagen fibres; therefore, forming a rudimentary bone that later gradually thickens [190]. The endochondral ossification occurs in the formation of long bones and most of the rest of the bones in the body [190]. It is a two step process initially characterised by the formation of a cartilaginous region that gets vascularised and promotes the migration of osteoblasts to it [190, 191]. The second step is the mineralisation of the cartilage template, which is initially deposited in a disorganised structure, containing a high proportion of osteocytes. Such structure is called woven bone and is gradually replaced by lamellar bone, which is stronger and filled with cylindrical arrangements of mineralised collagen fibres [190, 191].

Octacalcium phosphate and amorphous calcium phosphate are known precursors of the hydroxyapatite [191, 192]. The initial mineral formation is believed to be under cellular control, while the mineral propagation is mediated by the type I collagen in the extracellular matrix [191]. The basic nanoscale structure of bone consists in mineralised collagen fibres, with the mineral concentrated in holes zones of collagen fibrils [191, 193]. The assemblage and organisation of collagen creates intramollecular spaces assumed to be the site of mineralisation, from which the crystals grow and proliferate [191,193,194]. The collagen fibrils are formed by the self-assemblage of triple helices of collagens, with the hydroxyapatite crystals formed in the hole zones between collagens [193, 194], as shown

in Figure 5.3 (a). The hydroxyapatite crystals have a sheet-like shape and grow over and around the surface of the collagen fibrils, in a fashion that the crystal c-axis is orientated along the longitudinal axes of the fibrils, suggesting that the crystal nucleation and growth process is controlled by the collagen as shown in Figure 5.3 (b) [193, 194]. Finally, these mineralised fibrils organise parallel to one another, forming the mineralised collagen fibres, as in Figure 5.3 (c) [193, 194].

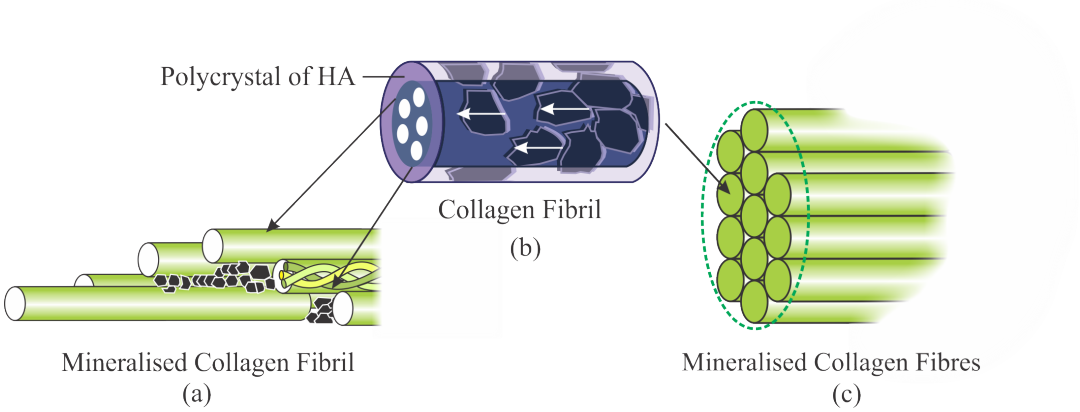


Figure 5.3: Collagen mineralisation schematic adapted from [193]. The bone nanostructure consists in mineralised collagen fibres. The process starts when (a) hydroxyapatite crystals are formed in the hole zones between collagens. (b) These crystals grow over and around the collagen fibrils. (c) The mineralised fibrils organise parallel to one another, forming the mineralised collagen fibres.

RS can easily observe the initial mineral deposited by the osteoblasts during osteogenic differenation [125,127,131]. In this Section, osteogenesis was induced with dexamethasone, L-ascorbic-acid-2-phosphate and β-glycerophosphate, as described in Section 2.2.4, and followed up to 21 days for the MSC line Y201. To confirm osteogenesis, histological ALP and von Kossa staining were performed at days 0, 7, 14 and 21 in samples prepared in parallel to those that were analysed with RS. The experimental procedures for the staining were described in Section 2.2.5.

Figure 5.4 shows the optical images for the histological alkaline ALP and von Kossa staining. This is a traditional staining assay that detects an increase in the extracellular

alkaline phosphatase enzyme activity and PO_4^{3-} deposits [149]. The images obtained compare cells kept in the basal condition with those induced into osteogenic differentiation. The top image on Figure 5.4 shows an overview of the well plates, while the bottom image shows the same wells at a higher magnification (using a x5 lens), where day 0 corresponds to the day when the differentiation was induced. No changes are seen on the histological staining images of the cells kept in basal medium, whereas osteo-induced cells showed, from day 14 onwards, the dark brown colour characteristic of the von Kossa staining and originated from the reaction of the phosphate deposits with the silver nitrate used.

The ALP staining, which used naphthol-AS-MX and Fast Red TR salt, preceded the von Kossa staining. The ALP hydrolyses the naphthol-AS-MX into a phenol component that combines with the Fast Red TR salt producing a pink/red colour [195]. This pink/red colour was observed during the staining process, but it was obscured by the dark brown from the metallic silver originated from the von Kossa staining and it is not visible in the saved images.

The Raman spectra obtained from the cells cultured in the basal and in the osteogenic medium, at days 0, 7, 14 and 21, are shown in Figures 5.5a and 5.6, respectively. The images display their averaged Raman spectra and respective SE envelopes. Small changes are observed in the spectra obtained from the cells kept in basal condition. These changes are consistent to cytoplasm contribution into the nucleus signal, which may have been caused by an increase of the cell layer thickness. Figure 5.5b shows the comparison between the averaged spectra (16 per sample) obtained from the nucleus and from the cytoplasm of the live Y201 cells. A decrease in the relative intensity of peaks assigned to nucleic acids (marked with the arrows) is seen in the cytoplasm spectrum. Other changes are observed in the high-wavenumber region, mainly for the relative intensity of the lipid peak at 2850 cm^{-1}. Assignments for these peaks are listed on Tables 3.1 and 4.1.

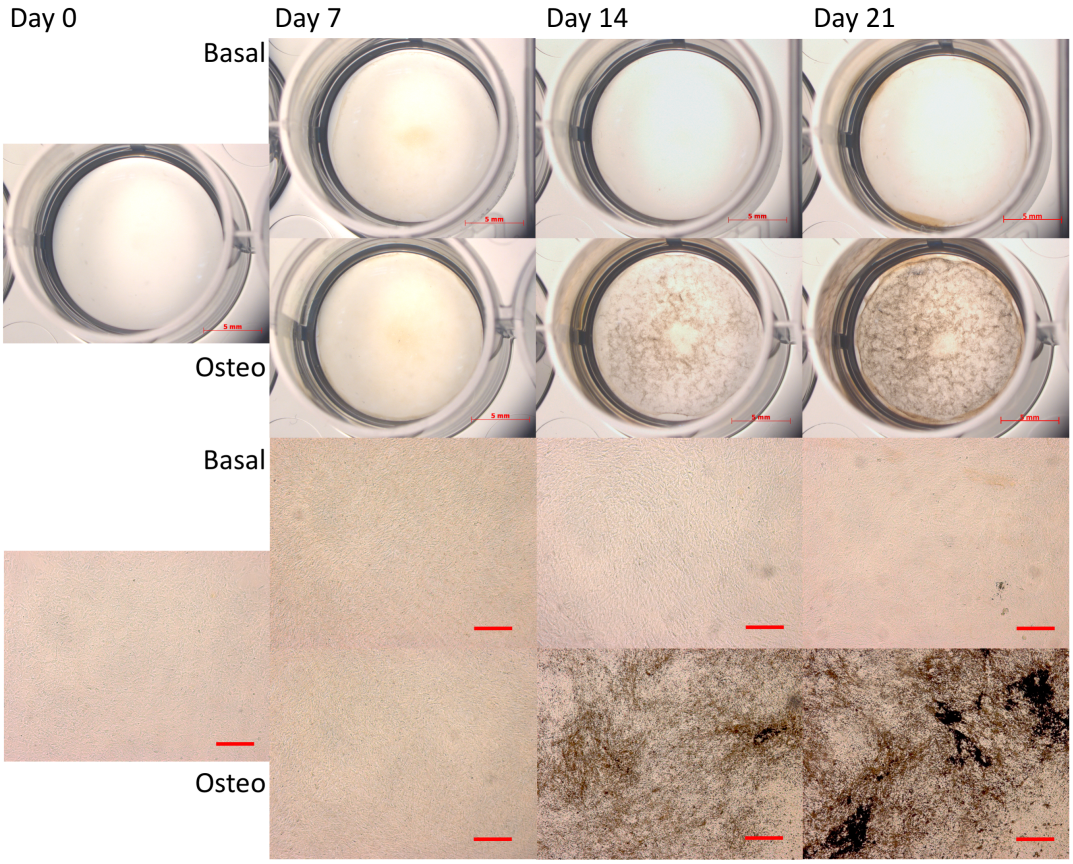
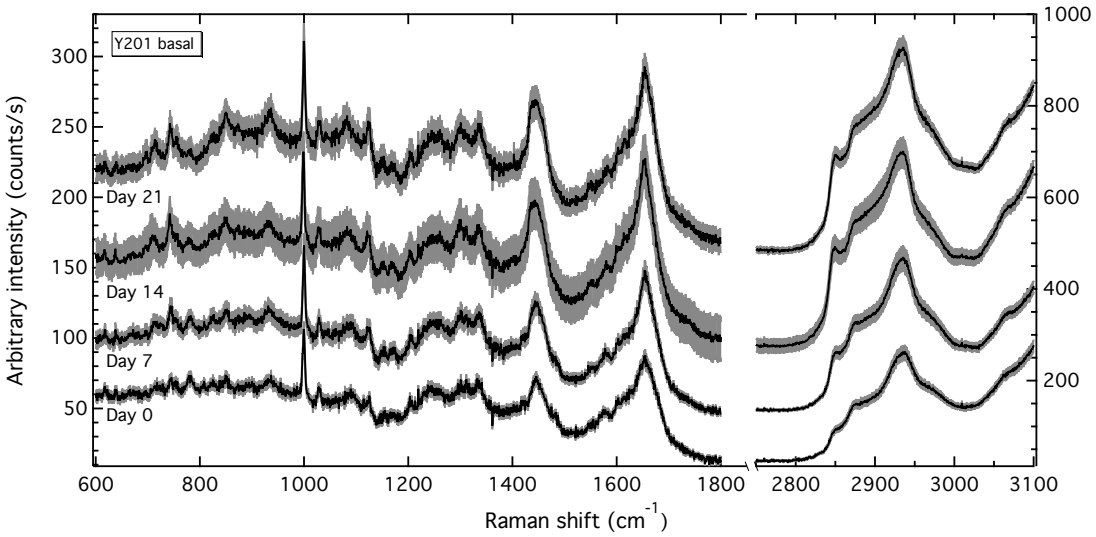
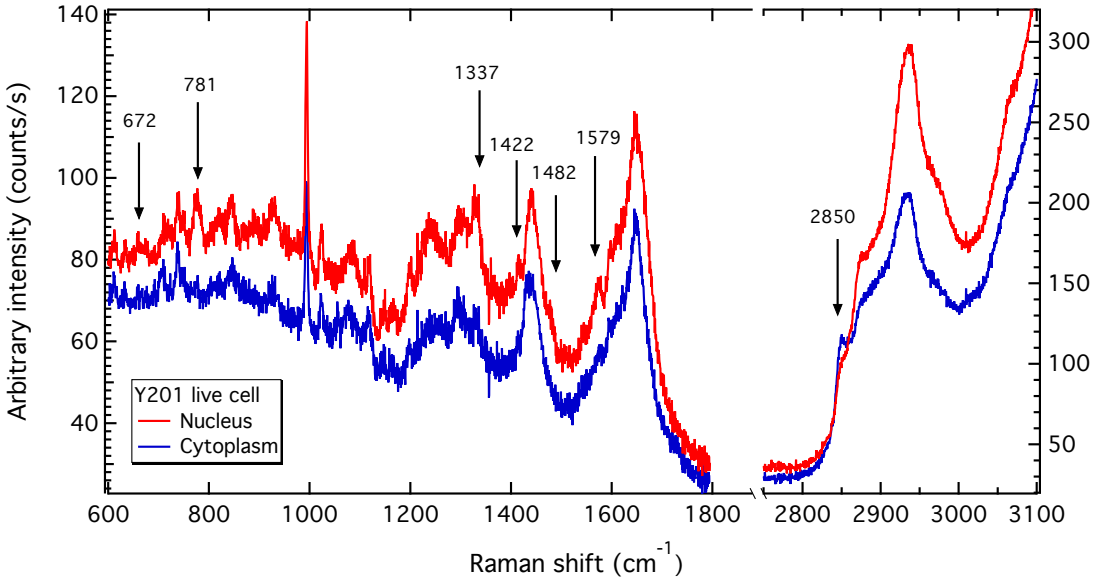


Figure 5.4: ALP and von Kossa staining of osteo-induced cells. The staining was performed at days 0, 7, 14 and 21 on cells kept in the basal and in the osteogenic medium. Top images: stereomicroscope images showing the entire wells. Bottom images: more details of the mineral content were captured with a 5x objective on a conventional microscope. Scale bars are 5 mm (top images) and 250 μm (bottom images).



(a)



(b)

Figure 5.5: (a) Averaged Raman spectra of basal Y201 MSC line for the time-points at days 0, 7, 14 and 21, along with their respective SE envelopes. (b) Comparison between cytoplasm and nucleus Raman spectra for the same cell line. Note the decrease in the relative intensity of peaks assigned to nucleic acids (marked with arrows) in the fingerprint region and the changes in the high-wavenumber region, mainly for the lipid peak at 2850 cm^{-1}. Such changes are noted in the spectra of MSCs in basal medium at days 14 and 21. In both graphs the spectra were offset for clarity.

A dramatic change is observed in the spectra of the osteo-induced cells (Figure 5.6), during the experiment time-course. This is characterised by the prominence of peaks at 956 cm^{-1} and 1069 cm^{-1}, respectively assigned to the phosphate (PO_4^{3-}) ν_1 symmetric stretching of hydroxyapatite and to the carbonate (CO_3^{2-}) ν_1 symmetric stretching [126, 127, 134]. The obtained spectra has similar characteristics to that of other studies on osteogenesis differentiation of MSCs and the presence of phosphate and carbonate indicates mineral deposition by the cells [41, 126, 127, 131].

The acquired Raman spectra revealed the presence of mineralised nodules as soon as day 7 (insert graph in Figure 5.6), which were not detected by the ALP and von Kossa staining method used. However, these nodules comprised only two spectra out of sixteen obtained during acquisition, indicating their small, but not inconsequential effect. Optical images of the nodules found at day 7, and their respective spectra, are shown in Figure

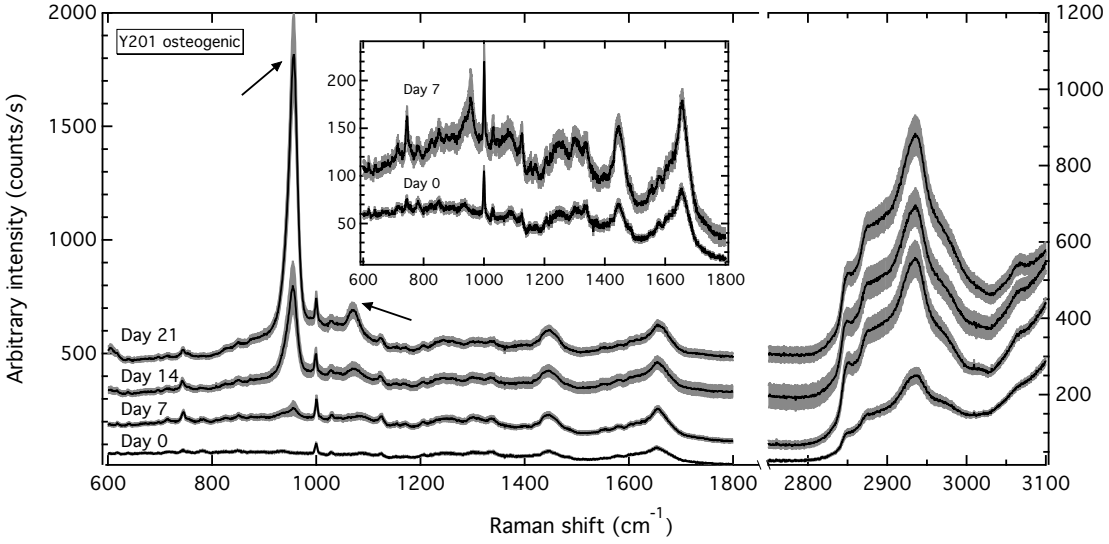
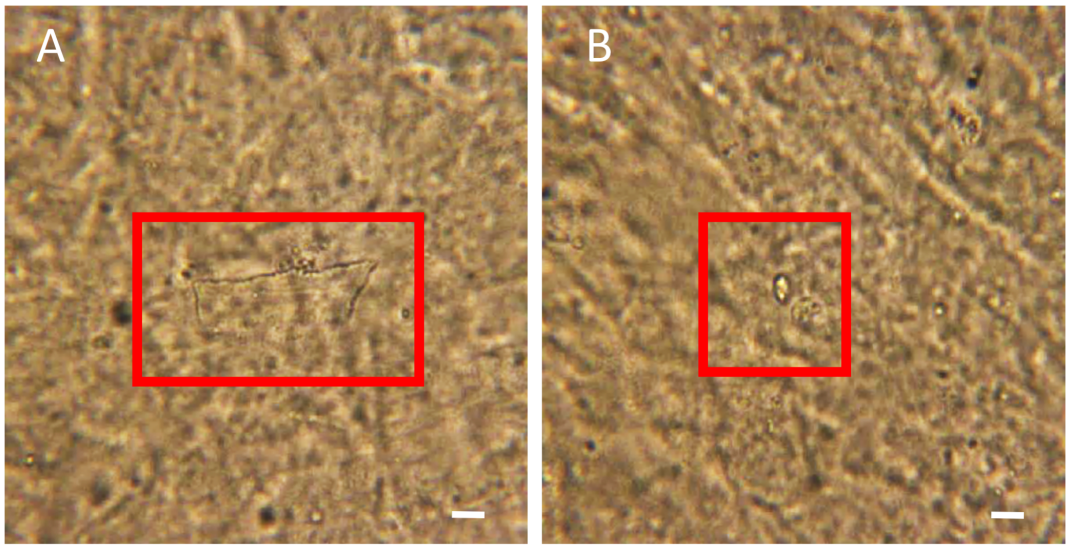


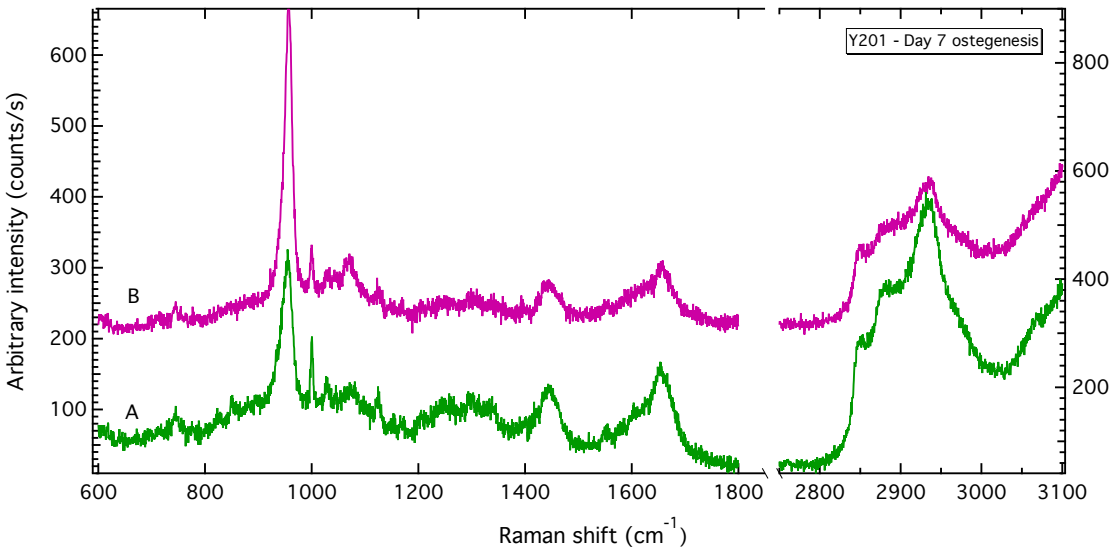
Figure 5.6: Averaged Raman spectra of the osteo-induced Y201 MSC line. The spectra are shown for the time-points at days 0, 7, 14 and 21 along with their respective SE envelopes. Dramatic changes are observed from day 7 onwards in the 956 cm^{-1} and 1069 cm^{-1} bands (marked with arrows), respectively assigned to the mineral apatite and mineral carbonate, which indicates mineralisation. The inset shows details of the fingerprint region for the spectra of both days 0 and 7. The spectra were offset for clarity.

5.7. In general, the hydroxyapatite signature in the Raman spectra of human MSCs is observed after 14 days of osteogenic induction [126, 131], but some aggregated nodule structures have been noticed at day 12 [134]. The hydroxyapatite peak has also been detected above the background as early as day 3 on mouse MSC lines [34], day 9 for human MSCs [35] and day 7 for human adipose-derived stem cells [68].

At day 21 the mineralised nodules were easily observed throughout the sample, covering large extensions over the cell surface layer. Figure 5.8 shows, as an example, the optical images and respective Raman spectra obtained in four different regions. The images reveal that these structures have different sizes and shapes and, that during the stages of the differentiation process they render different spectral signatures. Such heterogeneity has been also noticed by Ghita *et. al* (2014) during the osteogenic differentiation of human MSCs. In their work they observed that the central part of the nodules accumulated more crystalline hydroxyapatite, whereas the disordered amorphous apatite was predominantly found at their edges [134].

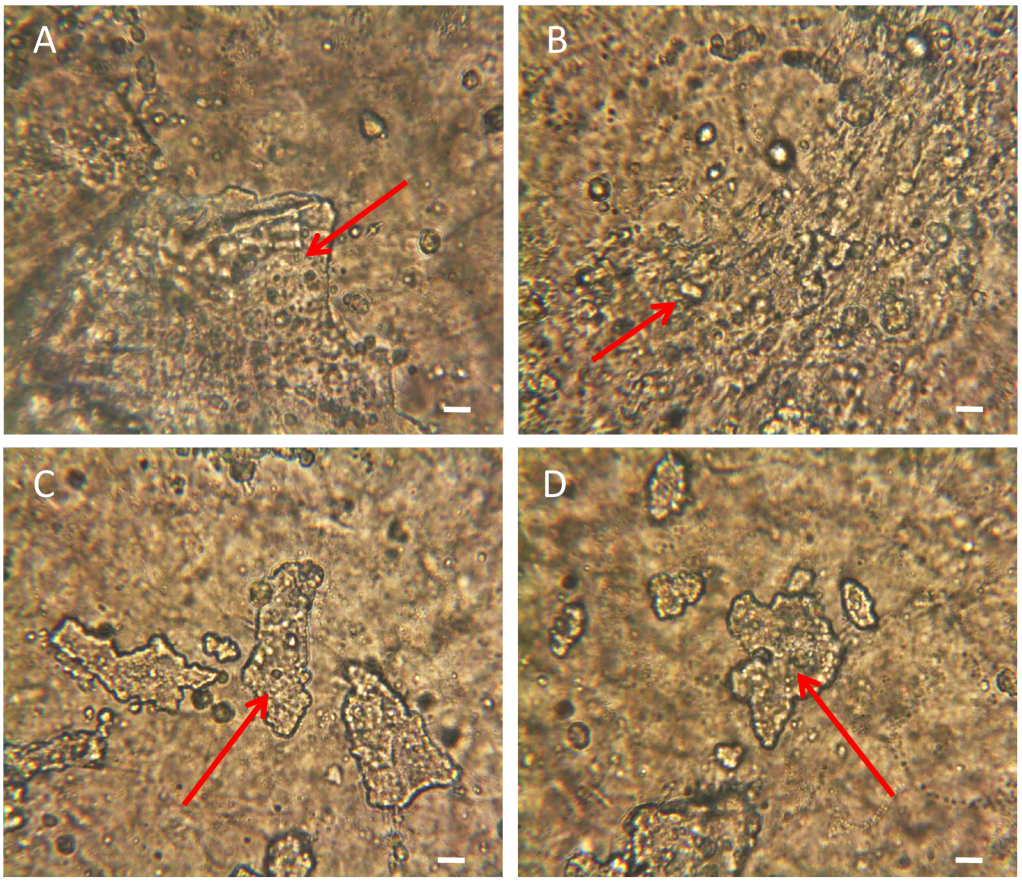


(a)

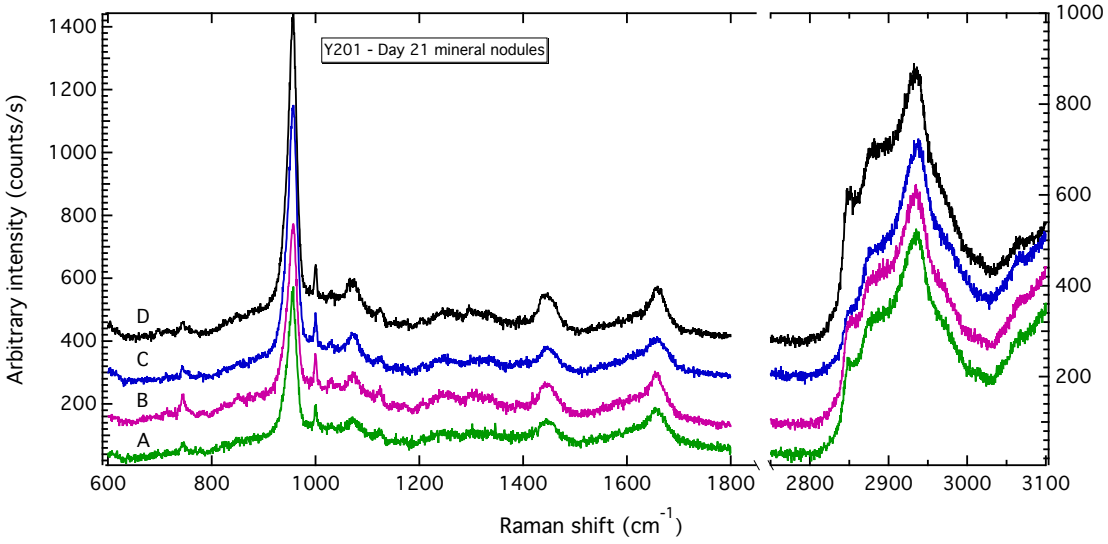


(b)

Figure 5.7: Mineralised nodules observed at day 7 on the osteo-induced cells. (a) Optical images of the two mineralised nodules found, marked in A and B, and their respective (b) Raman spectra, which were offset for clarity. Scale bar is 10 μm.



(a)



(b)

Figure 5.8: Mineralised nodules observed at day 21 on the osteo-induced cells. (a) Optical images of different regions showing the diversity of nodule sizes and shapes, A to D, and (b) their respective Raman spectra, which were offset for clarity. The arrows point to where the spectra were collected. Scale bar is 10 μm.

The mineralised nodules produced by the osteo-induced MSC had their spectra compared across the different time-points of the differentiation process and also compared to human and lamb bone. Four different fragments of human bone, obtained from femoral heads (FH1, FH2, FH3) and knee (K1), were analysed while immersed in HBSS. The lamb bone (L), sectioned from the rib, was macerated with hot water and air-dried before Raman spectra were collected. The preparation of these samples and details about the Raman acquisition are found in Section 2.4. Figure 5.9 shows a comparison of the averaged spectra obtained for the MSCs at day 21 with the averaged spectra obtained from the human and lamb bones. Table 5.2 lists the main peaks observed for the bones, their measured positions, literature range and peak assignments, in accordance to the literature [41, 126–129, 154].

It is known that the position of the peak assigned to the hydroxyapatite PO_4^{3-} ν_1 vibration indicates the type of calcium phosphate species present in the mineral content of

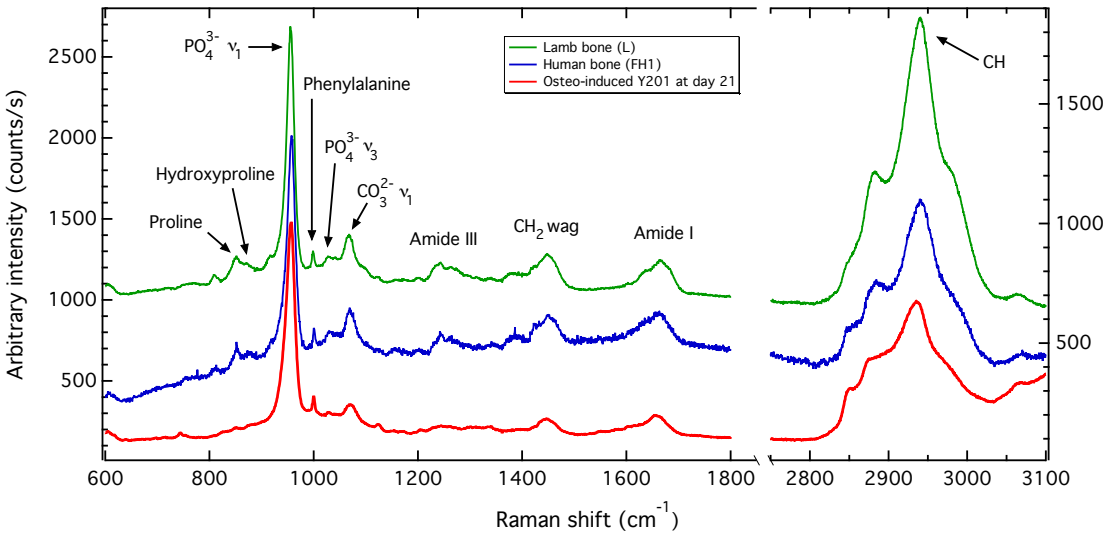


Figure 5.9: Raman spectra of lamb bone, human bone and mineralised nodules at day 21. The image shows a comparison between the averaged Raman spectra obtained for the lamb bone, human bone sample FH1 and mineralised nodules produced by the osteo-induced cells after 21 days. The spectra were offset for clarity.

Table 5.2: Assignments of the main peaks observed in the mineralised nodules and bones. The measured peak positions are compared to a position range found in literature along with general and detailed descriptions of their assignments. The measured peak position uncertainty is the equipment's wavenumber precision of 1 cm^{-1}.

Measured position (cm^{-1})	Literature range (cm^{-1})	General assignments	Detailed assignments
850	850-855	Proteins	C-C in proline, C-C stretch in collagen [127,128, 154] and hydroxyproline ring [126]
876	872-876	Proteins	Hydroxyproline [127, 128]
956	950-964	Mineral apatite	PO_4^{3-} ν_1 symmetric stretch [126–129]
1000	999-1001	Proteins	Phenylalanine [126–129, 154]
1029	1030-1044	Mineral apatite	PO_4^{3-} ν_3 asymmetric stretch [126–129, 154]
1069	1065-1070	Mineral carbonate	CO_3^{2-} ν_1 symmetric stretch [126–129, 154]
1260	1243-1270	Proteins	Amide III [126–129, 154]
1448	1445-1450	Lipids & proteins	CH_2 wag [126–129, 154]
1654	1595-1720	Proteins	Amide I [126–129, 154]
2940	2935-2940	Lipids & proteins	CH vibration in lipids and proteins [128,154,179]

bones and mineralised nodules [129, 133]. Therefore, the position of this peak can be used to categorise these species into three different band ranged frequencies. Disordered amorphous apatite from amorphous calcium phosphate or from A-type carbonate-substitution (where the carbonate is substituted for hydroxide) has a band between 945 and 950 cm^{-1}, B-type carbonate-substituted apatite has band from 955 to 959 cm^{-1}, while for crystalline non-substituted hydroxyapatite the band is between 962 and 964 cm^{-1} [129, 133]. Thus, it is important to deconvolve the hydroxyapatite region during the peak fitting process to allow the identification of the mineral species.

The Raman spectra of the mineralised nodules is also able to provide information about the quality of the mineral deposited by the cells and assess how similar it is to

native bone. This is obtained by the univariate analysis through the mineral/matrix ratio (956/850 PIR), the carbonate/phosphate ratio (1069/956 PIR), the CH/phosphate ratio (2940/956 PIR) and the relative presence of crystalline to disordered apatite (956/945 PIR) [41,126–128]. The FWHM of the main PO_4^{3-} ν_1 peak also indicates the crystallinity of the mineral species [128]. In this work the main PO_4^{3-} ν_1 peak corresponded to the position of the B-type carbonate-substituted apatite peak at 956 cm^{-1} .

The PO_4^{3-} ν_1 band was best fitted with two peaks, one at ∼956 cm^{-1} indicating the presence of B-type carbonate-substituted apatite and the other at ∼945 m^{-1}, which indicated the presence of disordered amorphous apatite [129, 133]. Similar fitting for the mineralised nodules of osteogenic induced human MSCs has also been performed in literature [134]. Such fitting, therefore, separates the contributions of the disordered and crystalline phases in the Raman spectra. From Table 5.3 it can be seen that the PO_4^{3-} ν_1 position is between 955 cm^{-1} and 958 cm^{-1} for all samples indicating the predominance of B-type carbonate-substitute apatite in the mineralised nodules as well as in the bone samples. The crystallinity of the mineral produced by the cells increased as indicated by the decrease of the peak FWHM during the 21 days of differentiation.

Table 5.4 shows other mineral properties obtained from the Raman spectra of the mineralised nodules, human and lamb bone. The 956/945 PIR easily displays the mineral phase of the analysed data, indicating that the presence of crystalline apatite relative to the disordered amorphous phase increased, in the nodules, throughout the differentiation process. It achieved, at day 21, a value close to that of the human bone, showing that the nodules were evolving towards a more crystalline phase. Other indication for an increased crystallinity of the nodules is the carbonate/phosphate ratio (PIR 1069/956), which slightly decreased reaching values similar to those of human bone, suggesting that the carbonate substitution into the mineral lattice has decreased.

Table 5.3: Fitted parameters of the PO_4^{3-} ν_1 peak for osteo-induced MSCs, human and lamb bones. The peak position and FWHM uncertainties are 1 cm^{-1}, which is the equipment's precision.

	Peak position (cm^{-1})	Peak FWHM (cm^{-1})
Osteo-induced MSCs		
Day 0	–	–
Day 7	957	17
Day 14	955	17
Day 21	956	16
Human bone samples		
FH1	958	16
FH2	958	15
FH3	957	15
K1	957	15
Lamb bone sample		
L	956	14

Table 5.4: Mineral properties of osteo-induced MSCs, human bone and lamb bone. The properties were obtained via PIRs obtained from the curve fitting of the Raman spectra.

	Crystallinity (956/945)	Mineral/matrix (956/850)	Carbonate/phosphate (1069/956)	CH/phosphate (2940/956)
Osteo-induced MSCs				
Day 0	–	–	–	–
Day 7	2 ± 1	3 ± 2	0.2 ± 0.1	0.7 ± 0.3
Day 14	4 ± 1	26 ± 6	0.10 ± 0.03	0.10 ± 0.03
Day 21	4.4 ± 0.8	76 ± 11	0.09 ± 0.01	0.050 ± 0.001
Human bone samples				
FH1	5.1 ± 0.4	13 ± 1	0.10 ± 0.01	0.08 ± 0.01
FH2	5 ± 1	17 ± 5	0.20 ± 0.04	0.10 ± 0.02
FH3	5.1 ± 0.5	18 ± 2	0.10 ± 0.01	0.05 ± 0.06
K1	5.5 ± 0.9	13 ± 2	0.20 ± 0.03	0.40 ± 0.06
Lamb bone sample				
L	3.4 ± 0.6	12 ± 2	0.10 ± 0.02	0.20 ± 0.03

Although the lamb bone showed similar crystallinity to the human bone samples, demonstrated by the FWHM of the main phosphate peak (Table 5.3), the crystallinity ratio 956/945 indicated a larger presence of the disordered amorphous phase relative to

crystalline apatite, suggesting that the lamb bone is less crystalline than the human bone samples. This is further confirmed by the small mineral/matrix ratio (PIR 956/850), which indicates a smaller presence of mineral relative to the bone matrix.

The mineral/matrix ratio (PIR 956/850) of the mineralised nodules increased considerably, reaching values higher than those obtained for the human and lamb bones (Table 5.4). This could suggest that there is an increased amount of mineral relative to bone matrix at day 21. This is corroborated by the decrease of the CH/phosphate ratio (PIR 2940/956). However, this intense mineral/matrix ratio can also be an indication that the collagen-rich protein matrix of the bone is not fully formed in this samples.

Such intense mineral/matrix ratio values were also obtained by McManus *et al.* (2011) and Chiang *et al.* (2008), who compared intensity ratios of the hydproxyapatite phosphate peak against proline (850 cm^{-1}) and phenylalanine (1000 cm^{-1}), respectively [126, 131]. The use of intensity ratios is preferred over area ratios once they are less susceptible to baseline or background correction errors [126]. However, the choice of matrix band to be used in the ratio calculation affects its value. For example, ratios using the amide I band at 1654 cm^{-1} can only be calculated from the area, but are affected by the baseline corrections [126]. Ratios involving the phenylalanine peak at 999 cm^{-1} do not provide robust data since non-collagenous matrix proteins contribute to this peak intensity [126]. Thus, since only collagenous matrix proteins contribute to the proline peak, the mineral/matrix ratio 956/850 is measuring only the bone matrix content.

From the spectra obtained in the mineralised nodules it was noted that although the mineral matrix was nearly formed and had characteristics similar to those of bone (indicated by the crystallinity ratio), the same cannot be assumed by the collagen-rich protein matrix (Figure 5.9). Peaks assigned to proline and hydroxyproline have merely been distinguished from the background, while peaks related to amide III, amide I and CH

vibrations are not as well resolved in the spectra obtained at day 21 of the osteo-induced cells as they are in the human and lamb bone spectra. However, these features are in accordance to other works where the osteogenesis of the MSCs has been studied [126,127]. In these studies, the spectra obtained from the nodules reflected the predominance of type-I collagen; however, their spectra also displayed a low relative intensity for the matrix peaks when compared to the spectra of native bone [126,127].

The results presented in Table 5.4 show that the mineral found in the mineralised nodules, which were produced by the Y201 MSC line after 21 days of osteogenic induction, have similar architecture to adult human bone, being composed by different mineral species. During the analysed time-points the reduction in the amount of disordered amorphous phosphate species, indicated by the crystallinity ratio 956/945, contributed to an increase in the mineral crystallinity of the nodules. The increase in the mineral crystallinity during ostegeonesis has been reported by several other studies on osteo-induced MSCs [126, 127, 131, 134]. The presence of only non-substituted crystalline hydroxyapatite, although increases the crystallinity, does not resemble the complex mineral interaction of the mineral species present in native bone [126–128]. Therefore, the fact that amorphous disordered and B-type carbonate substituted apatite were found in the osteo-induced MSCs and that the carbonate substitution decreased during osteogenesis suggests that the MSCs are producing mineral that resembles bone.

5.2.1 Multivariate analysis: PCA-LDA of the osteo-induced MSCs

Multivariate analysis by means of PCA and PCA-LDA were used to inspect the data in order to verify their discrimination, as described in Section 2.6.3. They were applied to the spectra obtained from the osteo-induced MSCs and compared from day 0 to day 21. At day 0 the cells have not started differentiation and their spectra resembles that obtained

for the undifferentiated Y201 cells in Chapter 4.

The unsupervised PCA applied to the MSCs cultured in osteogenic medium was able to capture the separation between undifferentiated and differentiated cells, for which the components PC1, PC2 and PC3 accounted for 39%, 18% and 10% of the total variance of the dataset (Figure 5.10). It can be seen from the scoreplot that the PCA separated the cells into two groups, clearly discriminating spectra from days 0 and 7 from those obtained at days 14 and 21. Although mineralisation was observed at day 7, this corresponded only to two spectra out of the 16 obtained for that time-point. Similar separation of the time-points has been observed in literature, where the overlapping of spectra of adjacent time-points was attributed to the continuous phenotypic changes occurring due to differentiation and to the cells' many functional aspects [126, 159].

Figure 5.11 shows the loadings corresponding to PCs 1, 2 and 3. PC1 loading is mainly characterised by the background of the regions of the mineral apatite at 956 cm^{-1} and of the mineral carbonate at 1069 cm^{-1}. Peaks were identified for the phenylalanine at 999

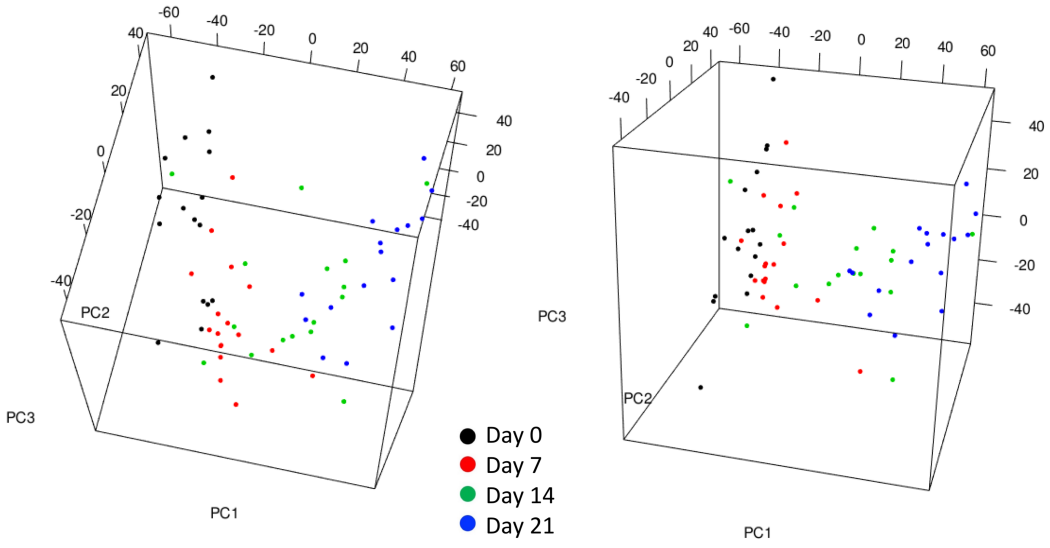
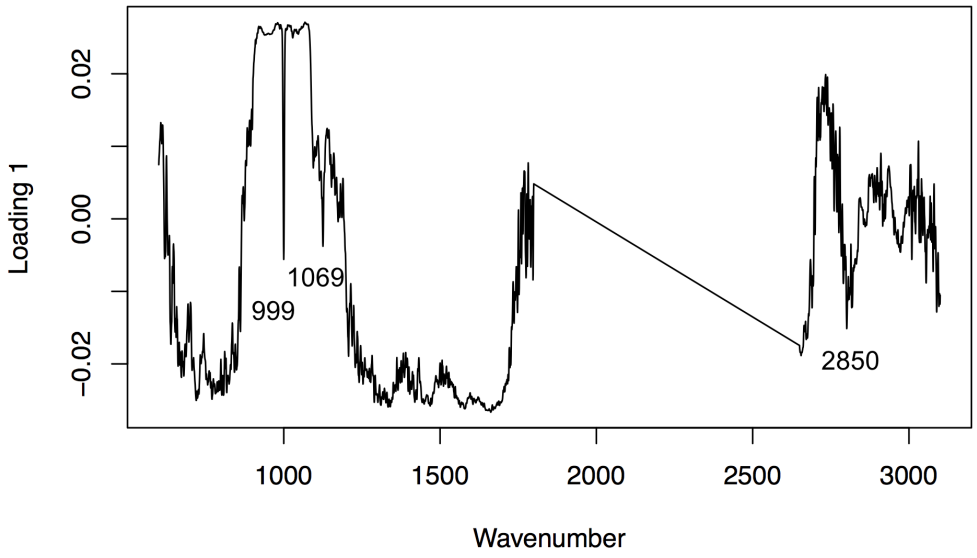


Figure 5.10: PCA scores for the osteo-induced MSCs. The PCA components PC1, PC2 and PC3 accounted for 39%, 18% and 10% of the total variance of the data. The images show different rotations of the same scoreplot.

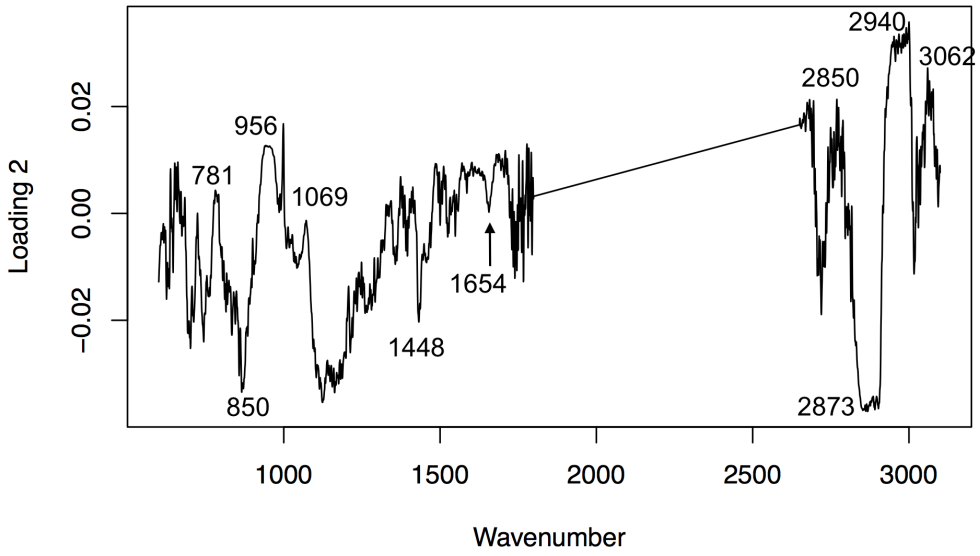
cm^{-1}, for the mineral carbonate at 1069 cm^{-1} and for the lipids at 2850 cm^{-1}.

Loading 2 highlights the bands related to the vibrations of the hydroxyapatite at 956 cm^{-1} and at 1069 cm^{-1}, showing that both peaks, allied to the different backgrounds, contributed to these PC scores. Other peaks contributing to the variance in this PC were the lipids and protein peaks at 1448 cm^{-1} and 1654 cm^{-1}. Apart from lipids vibrations, the bands corresponding to these two peaks are also related to the CH_2 and to the amide I vibrations of the proteinaceous matrix formed during osteogenesis, which could be indicative of the formation of the collagen-rich protein matrix in the differentiated cells. Another indication for it is the presence of lipid peaks in the high-wavenumber region (2850 cm^{-1}, 2873 cm^{-1}, 2940 cm^{-1} and 3062 cm^{-1}).

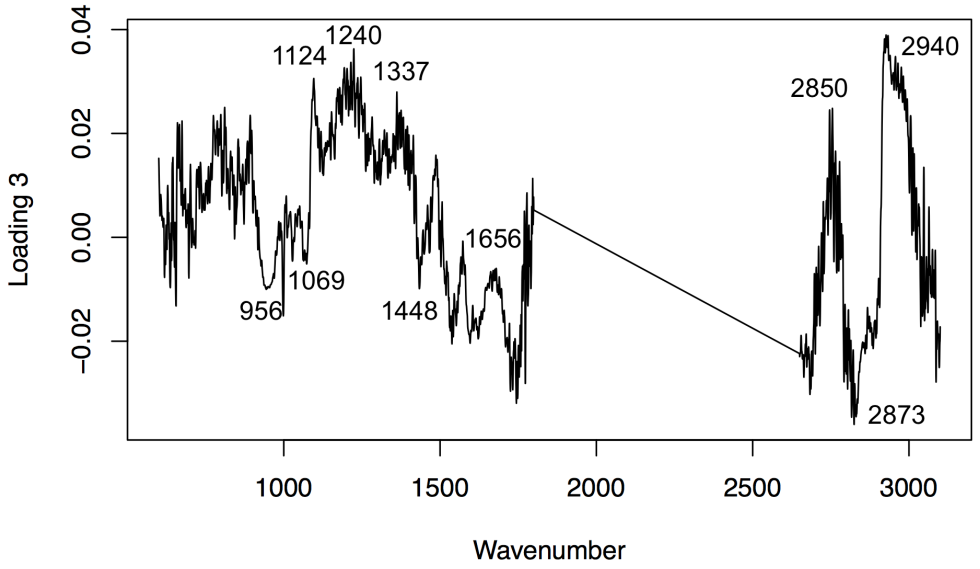


(a)

Figure 5.11: PCA of the osteo-induced MSCs showing loadings of (a) PC1, (b) PC2 and (c) PC3. The peaks indicated in each loading were the major contributors of each PC scores.- *continued on next page.*



(b)



(c)

Figure 5.11: *Continued from previous page.* PCA of the osteo-induced MSCs showing loadings of (a) PC1, (b) PC2 and (c) PC3. The peaks indicated in each loading were the major contributors of each PC scores.

Loading 2 also shows that the DNA/RNA peak at 781 cm^{-1} and the proline peak at 850 cm^{-1} also contributed to the PC scores. This can be atributed to the relative decrease of the DNA/RNA peaks against the proteins, as observed from the spectra of the osteo-induced cells (Figure 5.6). Such contribution has also been observed in the osteogenic differentiation of dental pulp stromal cells [159].

Loading 3 shows some of the same contributions previously determined, but it also highlights the contribution of other protein peaks related to the collagen-rich matrix formation, such as the 1124 cm^{-1} (phospholipids), 1240 cm^{-1} (amide III) and 1337 cm^{-1} (CH deformation) peaks.

In order to improve the separation between the time-points, LDA was applied over the PCA results, since it maximises the variance between groups. For this, only 8 PCs accounting for 79% of the variance were incorporated into the LDA algorithm. This was chosen so that the number of PCs input into the algorithm corresponds to half of the number of spectra obtained for the smallest group (16 spectra), thus preventing overfitting [36]. Figure 5.12 presents the PCA-LDA scores which show a continuous and clear phenotypic transition from undifferentiated MSCs (at day 0) to osteoblasts (at day 21). The two red points from day 7 sitting with groups belonging to days 14 and 21 correspond to the two mineralised nodules observed at this time point.

The PCA-LDA algorithm was cross-validated with a training and testing method described in Section 2.6.3. For this cross-validation, two thirds of the dataset were used for training the algorithm, while the rest was used for testing the predictions. The predictions are summarised in the confusion matrix shown on Table 5.5. It can be seen that none of the undifferentiated cells were classified as belonging to days 14 and 21. However, four cells from day 14 were wrongly predicted, two of them for day 7 and the other two for day 21. This indicated the continuous phenotypic changes suffered by these cells during

the process of differentiation. Finally, only one cell from day 21 was misclassified at day 14. If the four time-points are taken into account, the percentage of cells correctly classified is 70%. If the classification of differentiated (days 14 and 21) and undifferentiated (days 0 and 7) cells is compared, for which only two cells were misclassified at day 7, the percentage of correctly classified cells is 91%. These results suggest that the PCA-LDA model could be used for tracking and classifying osteogenic differentiation.

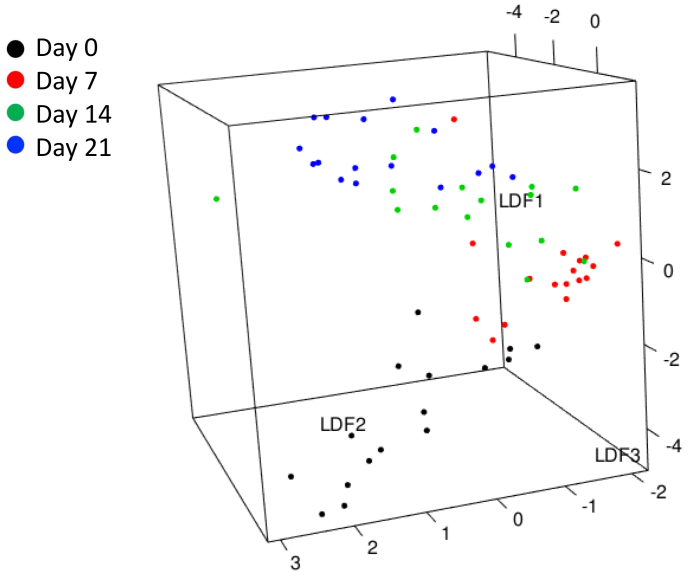


Figure 5.12: PCA-LDA scores for the osteo-induced MSCs. It shows a continuous and clear phenotypic transition from undifferentiated MSCs (at day 0) to osteoblasts (at day 21).

Table 5.5: Confusion matrix of the PCA-LDA classification algorithm results for the osteo-induced MSCs.

	Predicted time-point			
Time-point	Day 0	Day 7	Day 14	Day 21
Day 0	4	2	0	0
Day 7	0	4	0	0
Day 14	0	2	2	2
Day 21	0	0	1	6

5.2.2 PIR analyses of the osteo-induced MSCs

The univariate method was here used with the aim to identify markers of the osteogenic differentiation in the spectra obtained from the osteo-induced MSCs. The emergence of the hydroxyapatite peak in the spectra is a common osteogenesis marker; however, as indicated by the multivariate analysis, peaks related to DNA/RNA and to protein vibrations could also be used. For such, PIRs were calculated from the spectra obtained at days 0, 7, 14 and 21 as per Section 2.6.1.

These PIRs were obtained from ratios of all peaks. The discriminatory markers are shown from Figure 5.13 to 5.20. Such markers were observed in the ratios against peaks 717 cm^{-1} (lipids), 781 cm^{-1} (DNA/RNA), 933 cm^{-1} (proteins), 945 cm^{-1} (disordered amorphous phosphate), 956 cm^{-1} (carbonate substituted apatite - hydroxyapatite), 1069 cm^{-1} (lipids, phospholipids, DNA/RNA and mineral carbonate), 1102 cm^{-1} (lipids, proteins, carbohydrates and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA).

In general, the relative intensity of peaks 781 cm^{-1} (DNA/RNA), 1102 cm^{-1} (lipids, proteins, carbohydrates and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA) decreased during osteogenesis (Figures 5.14, 5.16 and 5.15, respectively). These peaks have in common their assignment to the vibrations of the DNA molecule. Other works investigating ESCs also noticed a decrease in the relative intensity of DNA and RNA bands during differentiation [33, 36, 56]. However, for these MSCs, this can be explained by the spectral changes produced by the minerals and collagen matrix present in the mineralised nodules formed during osteogenesis. These vibrations could have hindered such DNA peaks, producing the well separated markers obtained for day 21. The PCA had also shown the importance of the DNA/RNA peak at 781 cm^{-1} and the proline peak at 850 cm^{-1}, for the variance in the dataset. Adittionaly, the PIR method was able to evidence it by demonstrating that other peaks also assigned to DNA/RNA vibrations suffered the

same changes in their relative intensities. Similar trend for the markers was also observed for the lipid peak at 717 cm^{-1} (Figure 5.13).

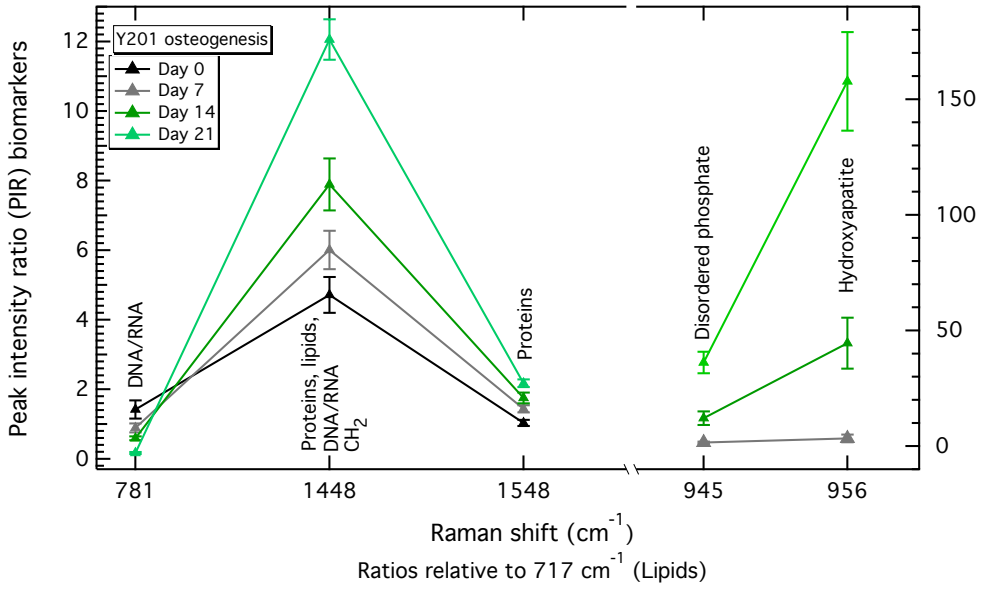
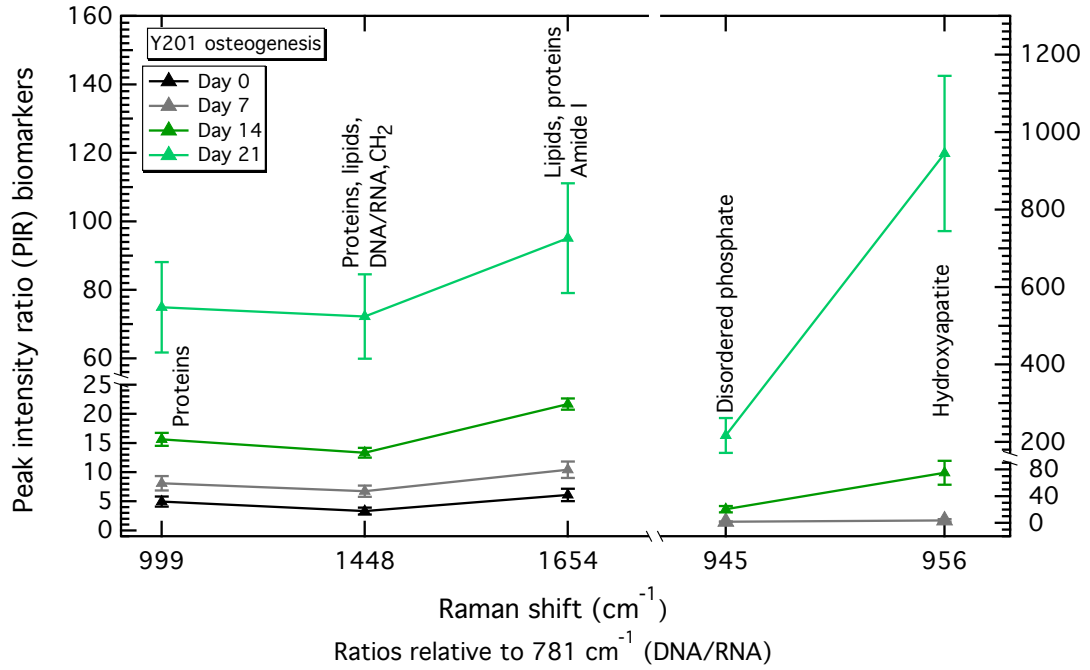
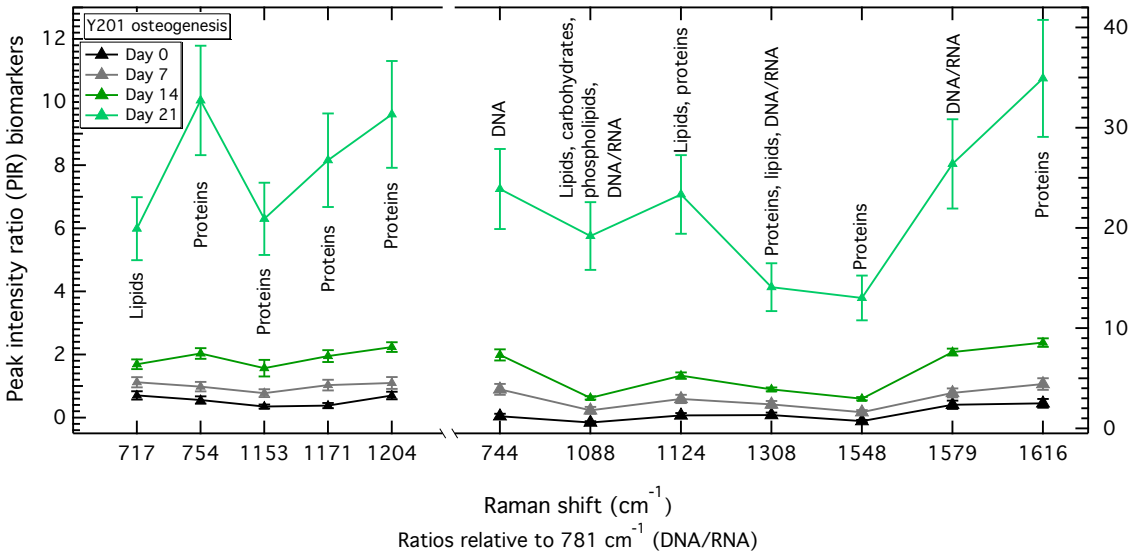


Figure 5.13: PIRs against peak 717 cm^{-1} (lipids) for osteo-induced MSCs showing the discrimination between different time-points of the osteogenesis. Attempt for the different scales of the right and left Y-axes. The error bars are the propagation of the fitted intensity SE.

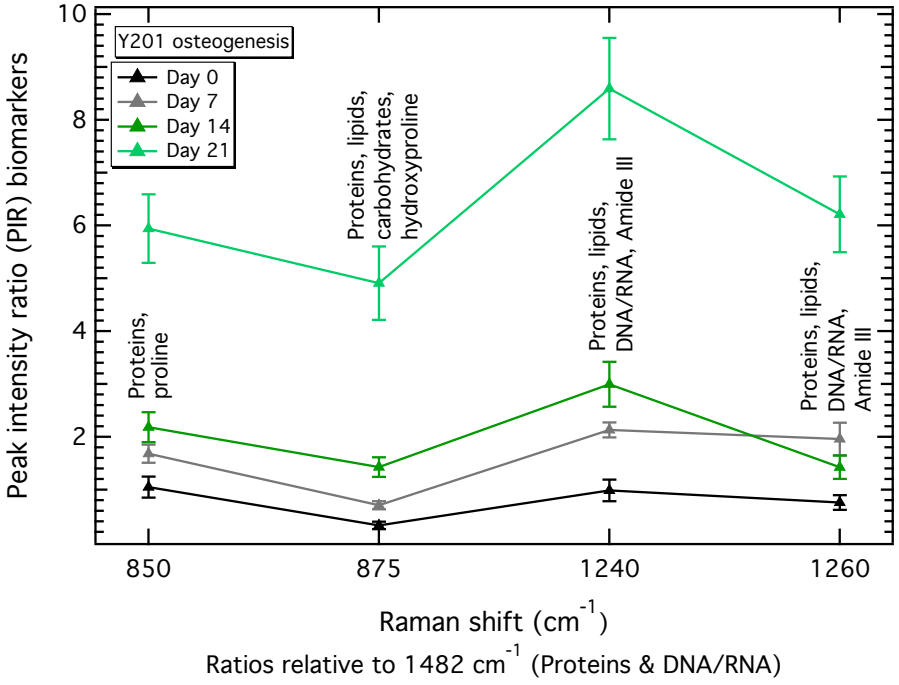


(a)

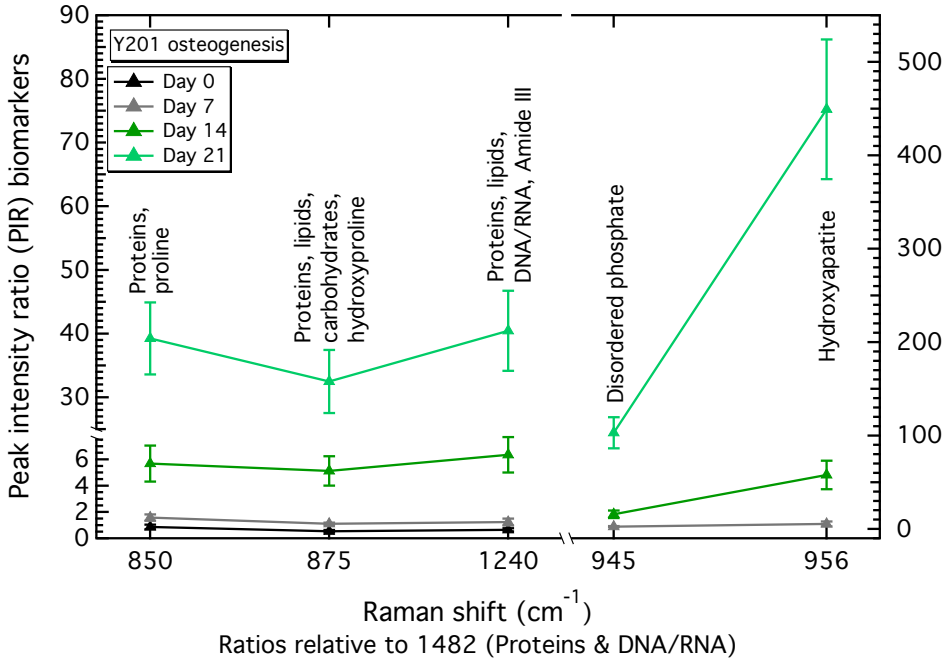


(b)

Figure 5.14: PIRs against peak 781 cm^{-1} (DNA/RNA) for osteo-induced MSCs. Due to very contrasting scales, these ratios were split into two graphs (a) and (b), which show the decrease in the relative intensity of this peak against others. Also, note the different scale for the right and left Y-axes. The error bars are the propagation of the fitted intensity SE.



(a)



(b)

Figure 5.15: PIRs against peak 1482 cm^{-1} (lipids, proteins, carbohydrates and DNA/RNA) for osteo-induced MSCs. Due to very contrasting scales, these ratios were split into two graphs (a) and (b). The markers show the decrease in the relative intensity of this peak during osteogenesis. The error bars are the propagation of the fitted intensity SE.

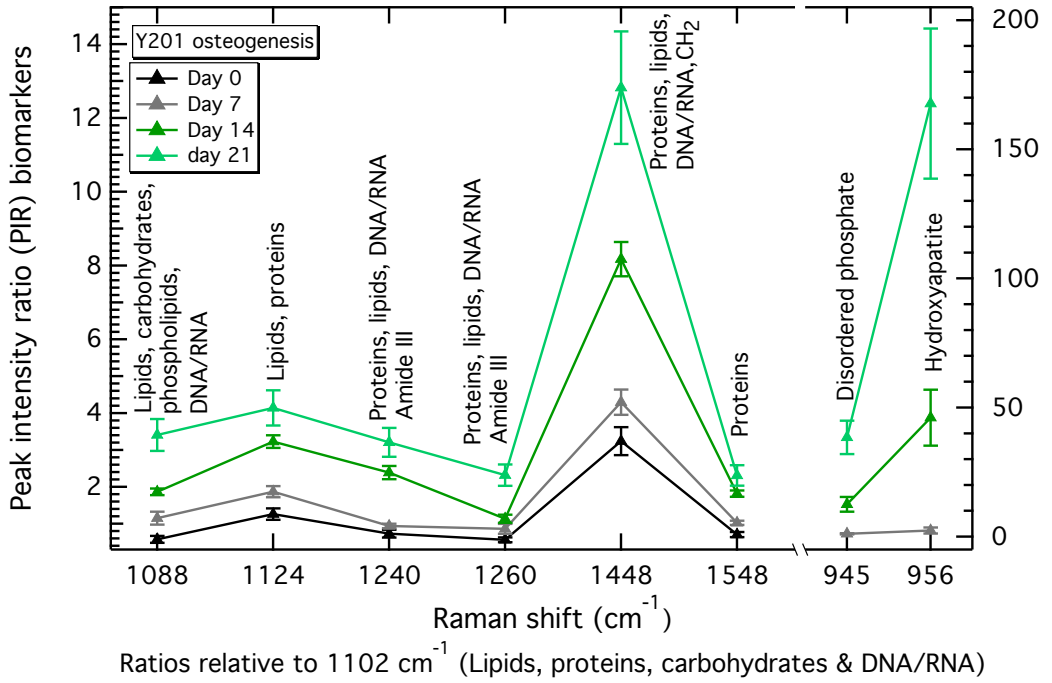


Figure 5.16: PIRs against peak 1102 cm^{-1} (lipids, proteins, carbohydrates and DNA/RNA) for the osteo-induced MSCs. Note the different scale for the right and left Y-axes. The error bars are the propagation of the fitted intensity SE.

Markers obtained against peak 933 cm^{-1} (Figure 5.17) showed an oscillation of this peak intensity relative to other cellular components. It initially decreased, from day 0 to day 7, when it started to increase relatively to other biomolecules. The peak at 933 cm^{-1} is assigned to vibrations of α-helix proteins and to the skeletal C-C of collagen backbone [127, 196, 197]. Thus, the variations in this peak relative intensity could be due to the start on the production of the collagen-rich protein matrix during differentiation. Type-II collagen has been noticed in the Raman spectra of pre-mineralised MSCs, while type-I collagen was observed on mineralised cells from day 21 onwards [126, 127]. In literature, this was correlated to an increased expression of chondrogenic markers, such as the genes *Sox9* (gene essential for cartilage formation) and *Col2a1* (type-II collagen gene), during osteogenesis [127]. However, in this work, type-II collagen peaks were not observed

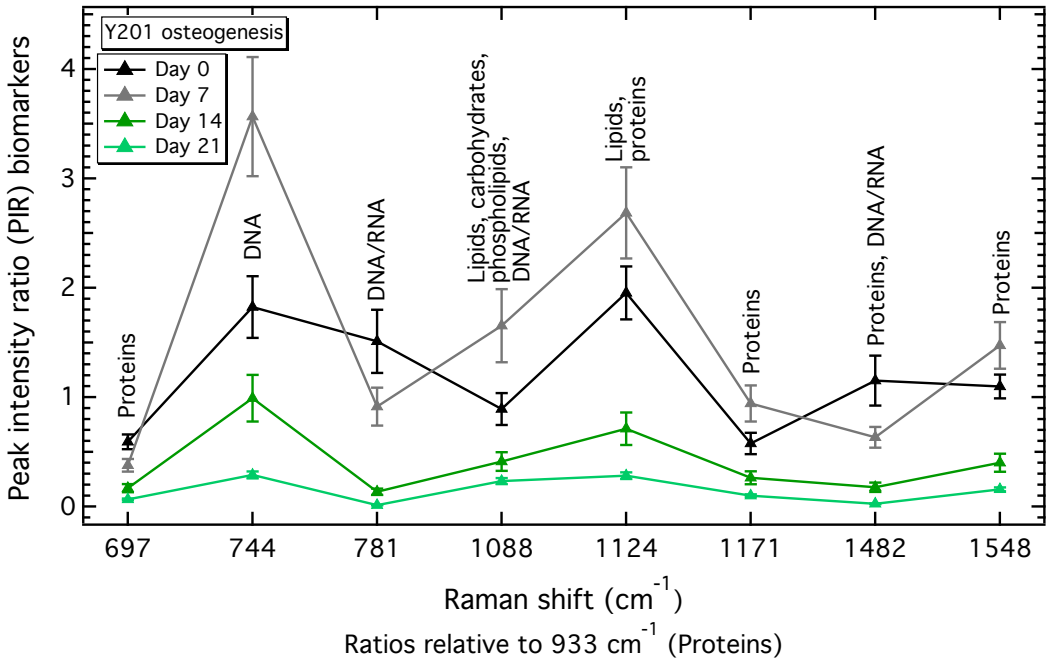
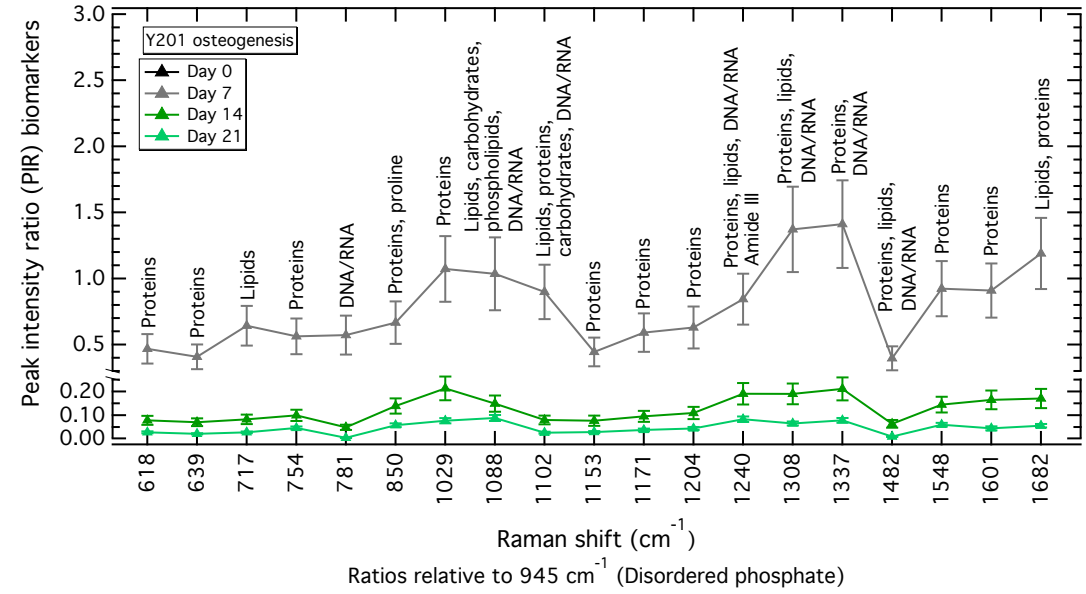


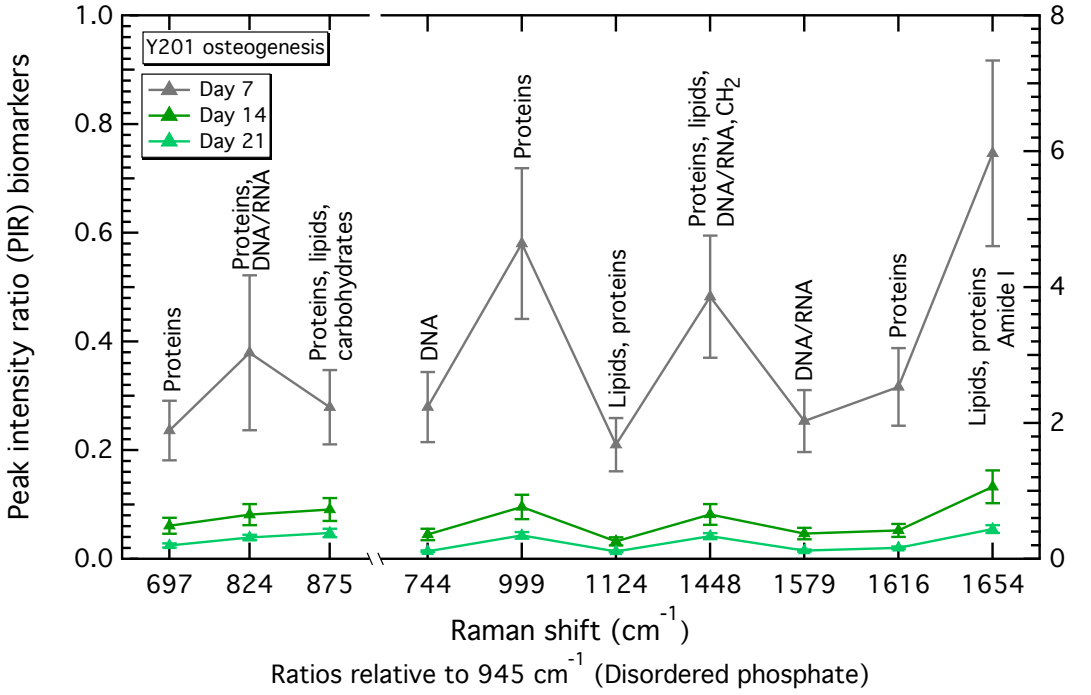
Figure 5.17: PIRs against peak 933 cm^{-1} (proteins) for osteo-induced MSCs, showing full discrimination between undifferentiated MSCs, early (day 7) and late stages (days 14 and 21) of the *in vitro* osteogenesis. The error bars are the propagation of the fitted intensity SE.

in the pre-mineralised MSC's averaged spectra (at day 7), nor were the Y201 MSCs tested for the expression of these genes in previous work. It could be that type-II collagen would be noted in the spectra between days 0 and 7; however, these other time-points were not evaluated. Thus, the oscillation observed in the relative intensity of the 933 cm^{-1} peak could be reflecting the transition from type-II to type-I collagen, since during the mineralisation process the cells deposit the minerals over an extra-cellular matrix mainly composed of type-I collagen [127]. The spectra obtained from the mineralised nodules (at days 14 and 21) reflected this protein content by the prominent bands of the amide I (1654 cm^{-1}) and amide III (1240 cm^{-1}).

The ratios against the phosphate vibrations 945 cm^{-1} and 956 cm^{-1} expose the clear difference between initial (at day 7) and late (at day 21) time-points during osteogenesis (Figures 5.18 and 5.19).

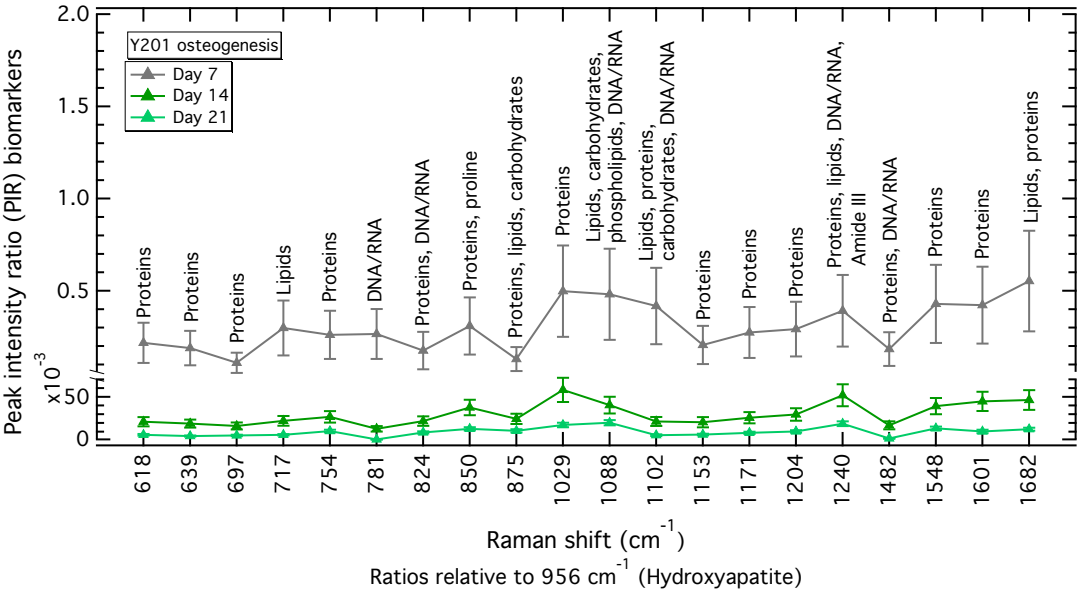


(a)

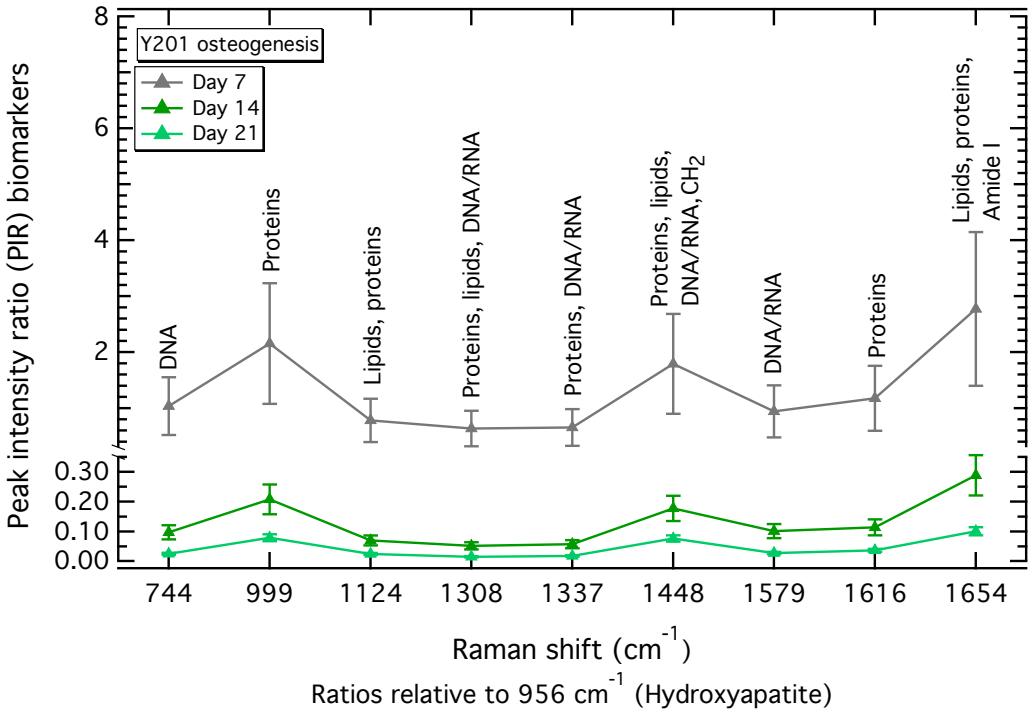


(b)

Figure 5.18: PIRs against peak 945 cm^{-1} (disordered amorphous phosphate) for osteo-induced MSCs. Due to very contrasting scales, these ratios were split into two graphs (a) and (b), which show the increase in the relative intensity of this peak against others. In order to better visualise the differences, the Y-axis was split on graph (a), while (b) has different scales for the right and left Y-axis. The error bars are the propagation of the fitted intensity SE.



(a)



(b)

Figure 5.19: PIRs against peak 956 cm^{-1} (B-type carbonate substituted hydroxyapatite) for osteo-induced MSCs. Due to very contrasting scales, these ratios were split into two graphs (a) and (b), which show the increase of this peak relative to others. In order to better visualise the differences the Y-axis was also split on both graphs. The error bars are the propagation of the fitted intensity SE.

Both 945 cm^{-1} and 956 cm^{-1} peaks were only detected in the spectra from day 7 onwards, when the mineralisation process was first detected. Thus, the markers of Figures 5.18 and 5.19 do not show PIRs for day 0. It is important to note that the carbonate/phosphate (1069/956) and the crystallinity (956/945) ratios are also not shown in Figures 5.18 amd 5.19 because although they have changed during osteogenesis, indicating a reduction in the carbonate substitution into the mineral lattice and an increase in the mineral crystallinity, they were not fully discriminated for all time-points. The PIRs obtained against disordered apatite (945 cm^{-1}) and hydroxyapatite (956 cm^{-1}) are in accordance to the other works that assign the hydroxyapatite peak as the main marker for osteogenesis [35, 126, 131, 134, 137].

Finally, the ratios against peak 1069 cm^{-1} (lipids, phospholipids and DNA/RNA on the undifferentiated cells and mineral carbonate for the differentiated MSCs) also showed a good separation between initial (at day 7) and late (at day 21) time-points of osteogenesis (Figure 5.20), being an additional mineral marker. This band is initially assigned to lipids, phospholipids and DNA/RNA vibrations in the undifferentiated MSCs [42,56,66,126,154]. However, with the start of the mineralisation process, its intensity increases due to the carbonate vibrations, which occur in the same frequency (Table 5.4), resulting in such separation. The discrimination achieved by the PIRs against the mineral apatite and mineral carbonate vibrations agree with the contributions of both peaks in the PCA loadings (Figure 5.11).

Thus, as achieved by the multivariate analyses, the use of an univariate method by means of PIRs is able to not only discriminating between undifferentiated and differentiated MSCs, but of also of providing markers that can be used as indicators of the different stages of the osteogenic differentiation process. The PIR analyses also provided markers for osteogenesis other than the presence of peaks related to the mineral apatite

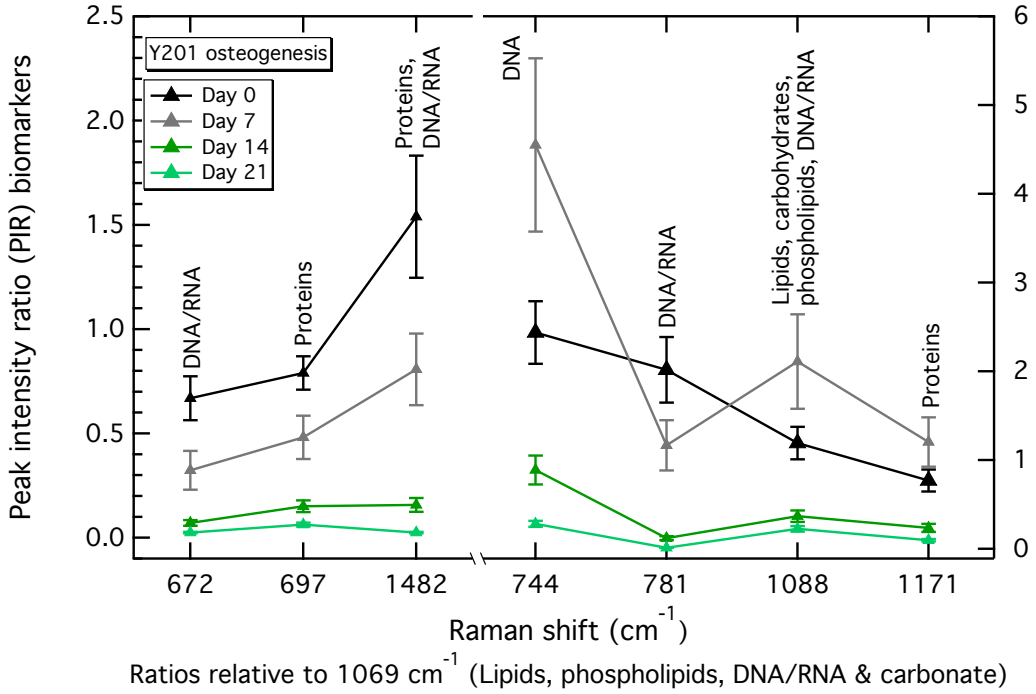


Figure 5.20: PIRs against peak 1069 cm^{-1} (lipids, phospholipids and DNA/RNA on the undifferentiated cells and mineral carbonate for the differentiated MSCs) for the osteo-induced MSCs. Note the different scale for the right and left Y-axes. The error bars are the propagation of the fitted intensity SE.

and mineral carbonate. Interestingly, three of the PIRs found as markers of osteogenesis have already been identified as full cell line discrimination markers for the four undifferentiated MSCs (Section 4.4). Such peaks were the 781 cm^{-1} (DNA/RNA), 1069 cm^{-1} (lipids, phospholipids and DNA/RNA on the undifferentiated cells and mineral carbonate for the differentiated MSCs) and 1482 cm^{-1} (proteins and DNA/RNA). These could be used to predict osteogenesis competence in undifferentiated cells and then tracked during differentiation.

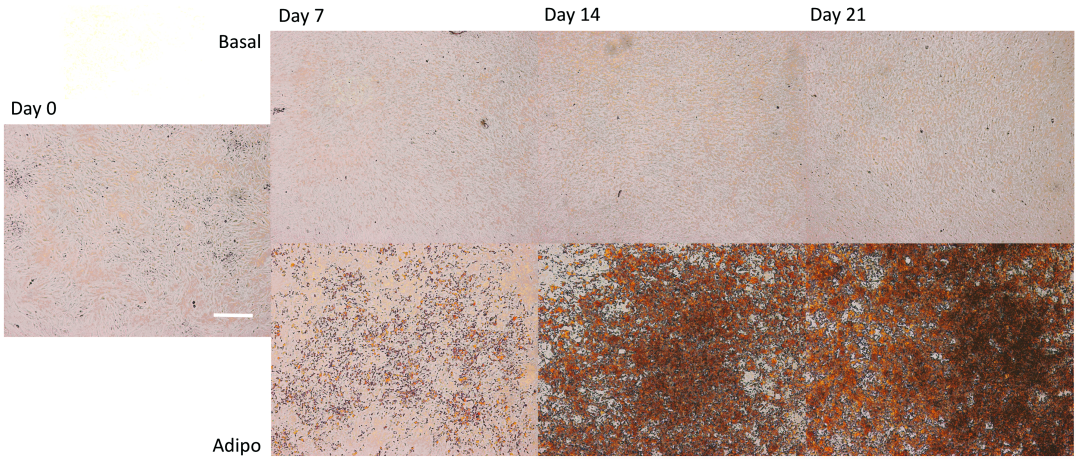
5.3 Adipo-induced MSCs

Adipogenesis is characterised by the formation of adipocytes, which were identified from day 7 onwards on the adipo-induced Y201 MSC line. The adipogenesis was induced and

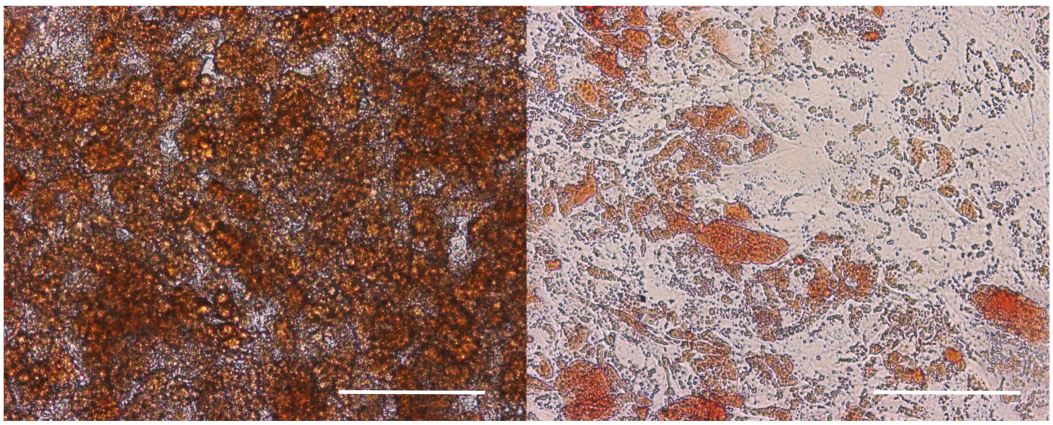
followed up for 21 days as described in Section 2.2.4. It can be verified by the presence of lipid droplets, validated by Oil Red O staining on the basal and adipo-induced cell cultures (Section 2.2.6). The red colour of the staining was perceived, initially, at day 7 and radically increased at days 14 and 21 for the adipo-induced cells, while no changes were observed for the basal cultures (Figure 5.21a). Figure 5.21b shows the staining at day 21 using a higher magnification lens (x10), in contrast to the x5 lens used on Figure 5.21a. This image highlights the denser area in the centre of the well, and also shows more details of the lipid droplets in the well's edge, where fewer cells were present.

Raman data was obtained on cells kept in HBSS at days 0, 7, 14 and 21 once adipogenesis had been induced. The preparation of the samples for RS, as well as the acquisition parameters used, are respectively detailed in Sections 2.2.3 and 2.3. The convergence tests results, shown in Section 5.1 of this chapter, confirmed that the spectra averaged at each time-point is sufficient to reflect that population average.

Figure 5.22 shows the averaged Raman spectra and the SE envelope of the data acquired on each time-point for the cells kept in basal medium (Figure 5.22a) and in adipogenic medium (Figure 5.22b). The spectra obtained for this basal culture is similar to that obtained for the basal culture of the osteogenic experiment (Figure 5.5a), also showing an increase of the cytoplasm signal at days 14 and 21. Figure 5.23 presents the spectra obtained at day 21, highlighting the positions of the identified peaks. The peak position uncertainty is ± 1 cm^{-1}, which is the equipment's wavenumber precision. The assignments for these peaks are described in Table 5.6, together with the range found in the literature for their positions.

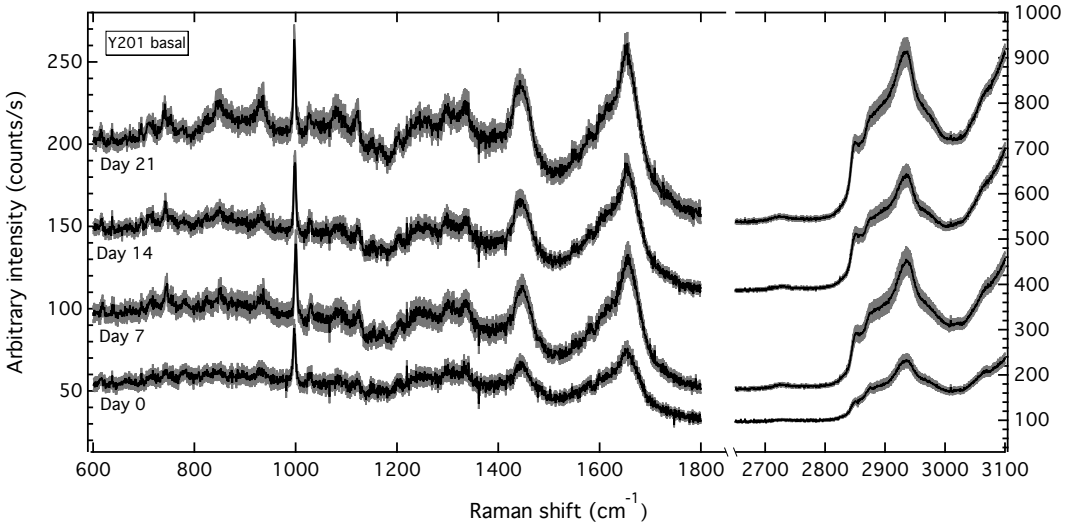


(a)

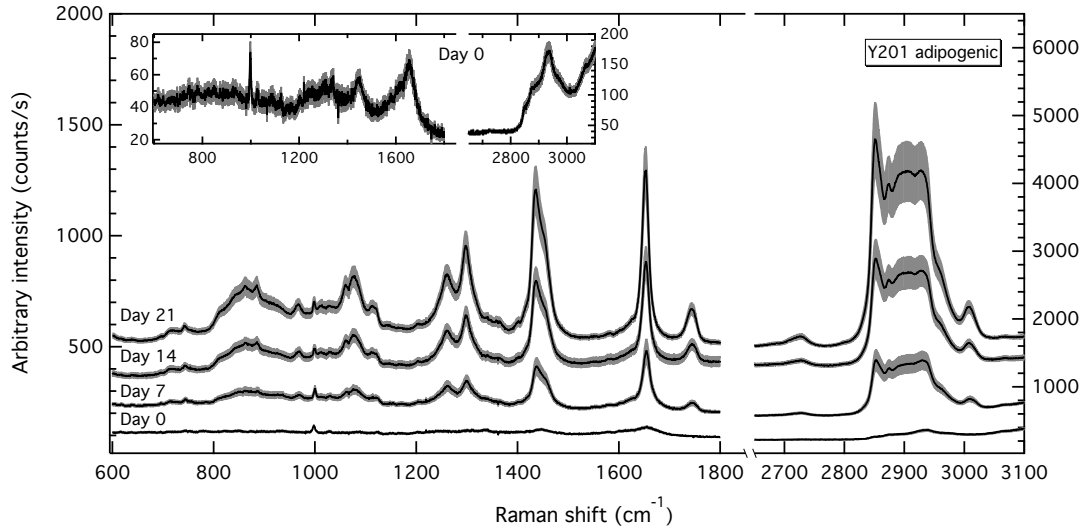


(b)

Figure 5.21: Oil Red O staining of adipo-induced MSCs. The staining was performed at days 0, 7, 14 and 21 on basal and adipo-induced cells cultured in 24 well plates. Image (a) shows the red staining of the lipid droplets present on the adipo-induced MSCs, which were absent in those kept in basal conditions. (b) Adipo-induced cells at day 21 of the adipogenesis. These images were captured in two different regions of the well, one at the centre showing a dense concentration of adipocytes and the other at the edge. Scale bar is 250 μm for both images.



(a)



(b)

Figure 5.22: Averaged Raman spectra of (a) basal and (b) adipo-induced Y201 MSC line. Both images display the averaged spectra and the SE envelope of the data obtained at days 0, 7, 14 and 21 on live MSCs. The inset in image (b) shows details of the day 0 averaged spectra. In both graphs the spectra were offset for clarity.

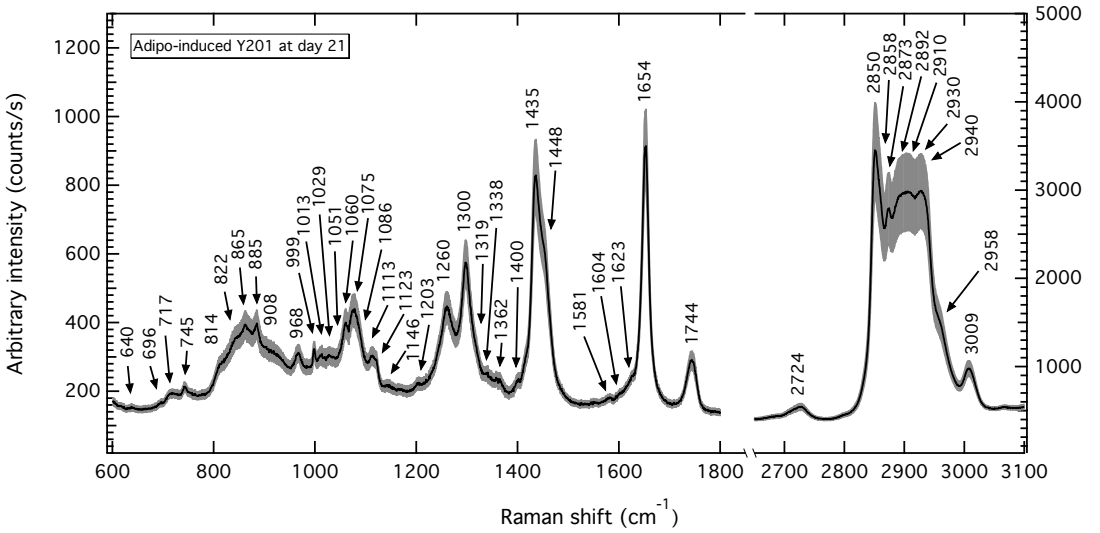


Figure 5.23: Day 21 averaged spectra of the adipo-induced Y201 MSC line. The image shows the averaged spectra with the SE envelope and highlights the identified peaks.

A radical change in the spectral profile of the adipo-induced MSCs is seen from day 7 onwards with the prominence of lipid peaks (Figure 5.22b). To allow the comparison with the undifferentiated MSCs, the image inset in this figure shows details of the averaged spectra obtained at day 0. The differences are marked by the rising of very sharp bands, markedly one band containing peaks 1435 cm^{-1} and 1448 cm^{-1} related to the CH_2 vibration in lipids, and the other with the 1654 cm^{-1} peak, which characterises the C=C stretch in lipids [137, 138, 179, 198]. In the non-differentiated cells, these two bands are dominated by lipid and protein vibrations; for example, the peak at 1654 cm^{-1} is assigned to vibrations of the C=C bond in both proteins and lipids. However, in proteins this is affected by the low energy of the α-helix conformation and the high energy of the β-sheet conformation, causing its broadening, while lipids, this C=C bond vibration gives rise to sharper peaks [137]. Thus, it can be assumed that the major peaks observed in the spectra obtained from the adipocytes are mainly originated from lipids.

Although there are few publications using RS to characterise adipogenesis, these changes are in accordance to works published by Downes *et al.* (2011) and Mitchell

et al. (2015) on adipose-derived stem cells (ADSCs) [137, 138] and to work conducted by Smus *et al.* (2015) on MSCs [140]. In the high-wavenumber region, it can be noted that the relative intensity of peak 2850 cm^{-1}, related to the CH_2 symmetric stretch in lipids has also increased. This observation is also in agreement with CARS analyses of adipo-induced ADSCs and MSCs, which revealed that this peak could be used as an indicator of adipogenesis [139, 140]. Although the lipid droplets varied in size, Figure 5.22b shows that from day 7 onwards the spectra only suffered a proportional increase of the peaks relative intensities, which was also observed by Mitchell *et al.* (2015) for the ADSCs [138].

Table 5.6: Peak assignments and literature reference of the adipo-induced MSCs. The assignments in bold correspond to the main lipid peaks observed in the spectra of the adipo-induced MSCs. The measured peak positions and their literature range are also shown. The uncertainty of the measured peak positions is 1 cm^{-1}, corresponding to the equipment's wavenumber uncertainty.

Measured position (cm^{-1})	Literature range (cm^{-1})	General biomolecular assignments	Detailed biomolecular assignments
640	640-646	Proteins	C-C twist of proteins, tyrosine [179]
696	696-700	Proteins	ν(C-S) trans (amino acid methionine) [179]
717	717-719	Phospholipids	C-N (membrane phospholipids head) [179]
745	745-748	Proteins	Symmetric breathing of tryptophan [179]
814	813-817	Proteins	C-C stretch of collagen [179]
822	823-826	Saccharides	Out-of-plane ring breathing of tyrosine [179]
865	865	Proteins	C-C stretch of collagen [179]
885	**870-890**	**Lipids**, & proteins	**C-O-C skeletal vibration in lipids** [198] and C-C in collagen and protein bands [179]
908	904	Proteins	C-C skeletal stretch [179]
968	968-971	Lipids & proteins	C-C bonds in collagen [138], Phosphorylated proteins and ν(C-C) wagging [179]
999	999-1008	Proteins	Phenylalanine [138, 179]
1013	1011-1013	Proteins	Tryptophan ring breathing [179]
1029	1029 -1030	Proteins	O-CH_3 stretching of methoxy groups and phenylalanine of collagen [179]
1051	1048-1053	Proteins & saccharides	C-O stretching and C-N stretching in protein and glycogen [179]
1060	**1057-1069**	**Lipids**	**C-C stretch in lipids** [24, 179, 198]
1075	**1073-1082**	**Lipids**	**C-C stretch in lipids** [24, 179, 198]

Continued on next page

Table 5.6 – *Continued from previous page*

1086	1082-1086	Lipids & proteins	C-N bonds in proteins and lipids [138, 179]
1113	**1114-1119**	**Lipids**	**C-C stretch in lipids** [24, 179, 198]
1123	**1125**	**Lipids**	**C-C stretch in lipids** [24, 179]
1146	1149	Carbohydrates	Carbohydrates [179]
1203	1204	Proteins	Amide III and CH_2 wagging in proteins and collagen [179]
1260	**1258-1268**	**Lipids** & proteins	**=CH deformation in lipids** [24, 198] and Amide III [138, 179]
1300	**1300-1304**	**Lipids**	**CH_2 twisting in lipids** [24,179,198], textbfCH_3 and CH_2 in lipids [138]
1319	1319	Proteins	CH_3CH_2 twisting in collagen [179]
1338	1337-1339	Lipids & proteins	CH_2CH_3 wagging, twisting and/or bending mode of collagen and lipids [179]
1362	1359-1364	Proteins	Tryptophan [179]
1400	1400	Proteins	NH in plane deformation and ν(C=O)O^- of amino acids [179]
1435	**1420-1450**	**Lipids**	**CH_2 scissoring vibration in lipids** [24, 179]
1448	**1440-1450**	**Lipids**	**CH_2 and CH_3 scissoring** [198], **CH_2 in fatty acids** [138, 179]
1581	1580-1583	Proteins	C=C bending of phenylalanine [179]
1604	1602-1607	Proteins	C=C in-plane bending mode of phenylalanine and tyrosine [179]
1623	1623	Proteins	Tryptophan [179]
1654	**1654-1656**	**Lipids**	**C=C stretch in lipids** [24, 138, 198]
1744	**1746-1749**	**Lipids**	**C=O stretch in lipids** [198]
2724	2724	Lipids & proteins	CH stretch vibration [179]
2850	**2850-2875**	**Lipids**	**CH_2 symmetric stretch in lipids** [24,179,198]
2858	**2850-2875**	**Lipids**	**CH_2 symmetric stretch in lipids** [24,179,198]
2873	**2850-2875**	**Lipids**	**CH_2 symmetric stretch in lipids** [24,179,198]
2892	**2889-2908**	**Lipids** & proteins	**CH_2 asymmetric stretching in lipids** [198], CH_3 symmetric stretch of lipids and proteins [72, 154]
2910	**2912-2930**	**Lipids**	**CH_2 asymmetric stretch in lipids** [24,72,154, 179]
2930	2930-2940	Lipids & proteins	CH vibration in lipids and proteins [24, 179]
2940	2930-2940	Lipids & proteins	CH vibration in lipids and proteins [24, 179]
2958	2958	Lipids & proteins	CH_3 asymmetric stretch [24, 179]
3009	**3007-3015**	**Lipids**	**Unsaturated =CH stretching in lipids** [24, 179, 198]

5.3.1 Multivariate analyses: PCA-LDA of the adipo-induced MSCs

A PCA was applied to the adipo-induced MSC spectra from day 0 to day 21 in order to unveil the most important Raman bands responsible for the variance between the time-points, as described in Section 2.6.3. The unsupervised PCA results (Figure 5.24) show the clear separation between the spectra obtained at day 0 and the spectra from the lipid droplets at days 14 and 21. The spectra from day 7 appears scattered across the two groups (Figure 5.24). The overlapping of adjacent time-points is expected if continuous phenotypic changes are occurring [159]. The PCA scoreplot indicates that these changes are continuous but, from day 14 onwards, many of the cells share the same phenotype and functional aspects. These results agree with the similar findings obtained for ADSCs [138]. PC1, PC2 and PC3 accounted, respectively, for 51%, 11% and 6% of the total variance and their loadings are shown in Figure 5.25. PC1 is the responsible for most of the separation between the spectra at day 0 and days 14 and 21.

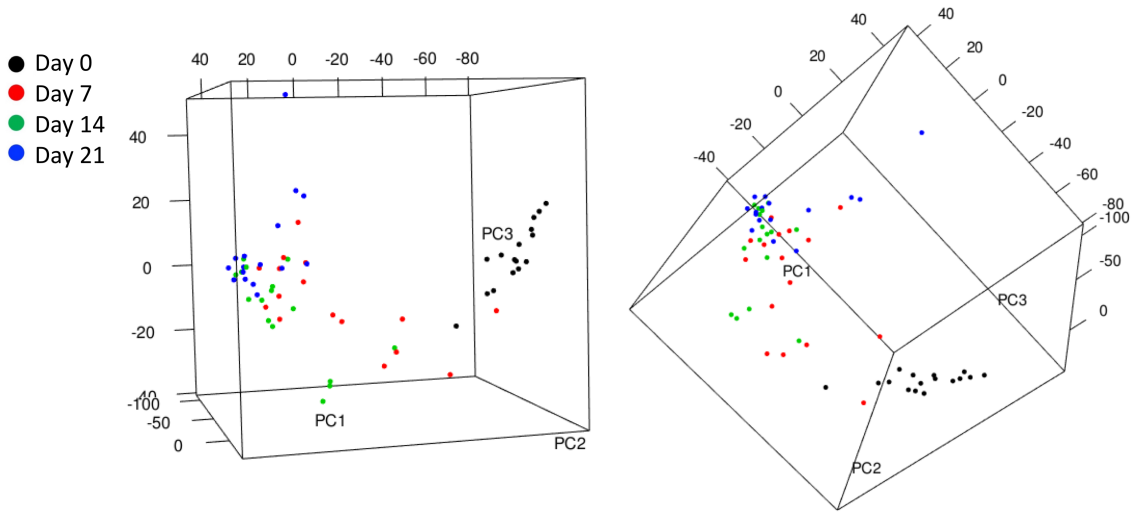
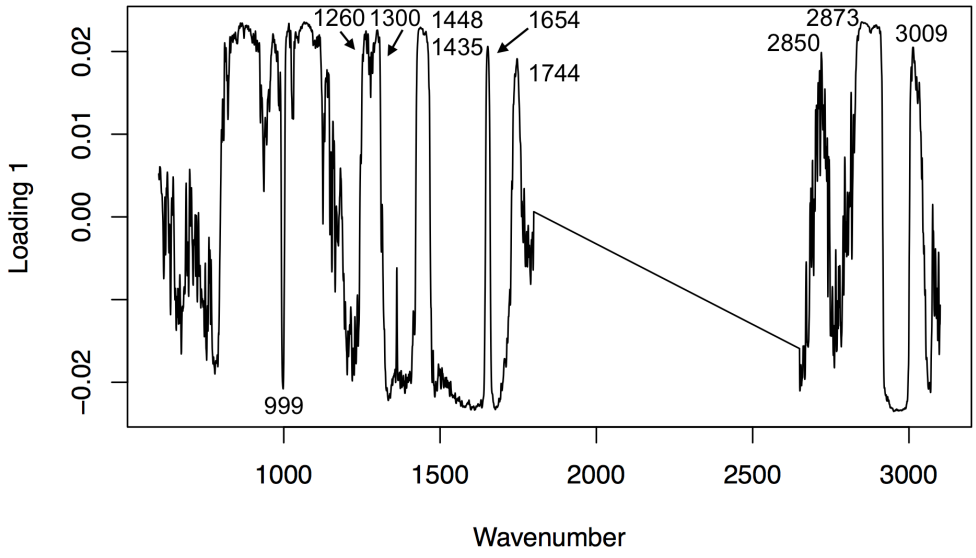
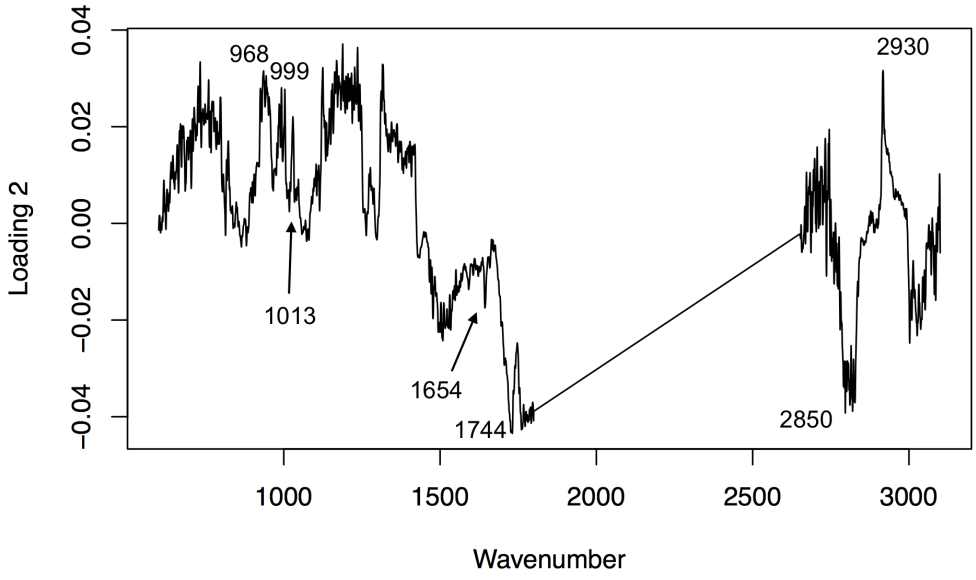


Figure 5.24: PCA scores of the adipo-induced Y201 MSC line. PCA components PC1, PC2 and PC3 accounted, respectively, for 51%, 11% and 6% of the total variance. The two images show different rotations of the same scoreplot.

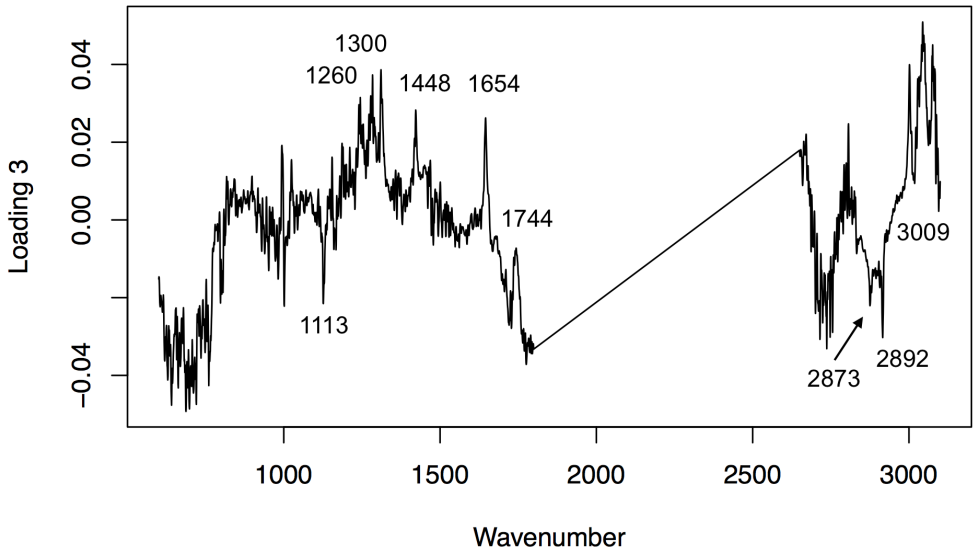


(a)



(b)

Figure 5.25: PCA of the adipo-induced MSCs showing loadings of (a) PC1, (b) PC2 and (c) PC3. The peaks indicated in each loading were the major contributors of each PC scores. - *continued on next page.*



(c)

Figure 5.25: *Continued from previous page.* PCA of the adipo-induced MSCs showing loadings of (a) PC1, (b) PC2 and (c) PC3. The peaks indicated in each loading were the major contributors of each PC scores.

The analysis of the PC loadings allows a better understanding of Raman bands contributing to the variance in the data. PC1 loading (Figure 5.25a) shows that the phenylalanine peak (999 cm^{-1}) and all other peaks associated to lipids (1260 cm^{-1}, 1300 cm^{-1}, 1435 cm^{-1}, 1448 cm^{-1}, 1654 cm^{-1}, 1744 cm^{-1}, 2850 cm^{-1}, 2873 cm^{-1} and 3009 cm^{-1}) are major contributors for the PC1 scores. The PC2 loading (Figure 5.25b) highlights, in the fingerprint region, the peaks associated to proteins (968, 999 and 1013 cm^{-1}) and the lipid peaks at 1654 cm^{-1} and 1744 cm^{-1}. Other lipid and protein peaks are observed in the high wavenumber region, such as the peaks at 2850 cm^{-1} (CH_2 stretch in lipids) and at 2930 cm^{-1} (CH vibration in proteins and lipids).

From the PCA scoreplot, it can be seen that PC3 separated some of the cells at day 14 from those at day 21 (Figure 5.24). The PC3 loading (Figure 5.25c), responsible for

6% of the variance, similarly highlights the previously mentioned lipids peaks. However, it shows that the peak at 1113 cm^{-1} (C-C stretch in lipids) also contributed to the PC3 scores.

The results obtained with the PCA show that adipogenesis clearly affects the cell biomolecular signature as demonstrated by the PC1 loadings. It was this PC component, which accounted for 51% of the variance in the dataset, that separated undifferentiated MSCs from the adipocytes. In the study performed on ADSCs by Mitchell *et. al* (2015), the combination of PC1 and PC2 resulted in the separation between differentiated and undifferentiated cells; however, in their work the protein peaks appeared only in the second principal component [138]. Although in the MSCs the protein contribution was verified since PC1, the obtained results are in accordance with literature findings and show that the relation between proteins and lipids was responsible for the discrimination of the adipocytes from the undifferentiated MSCs.

Similarly to the PCA-LDA model used for the osteo-induced MSCs, a model was built for the adipo-induced cells. This model also used 8 PCs, accounting for 82% of the variance in the dataset, in order to prevent the overfitting of the data. Details about the PCA-LDA are found in Section 2.6.3. The PCA-LDA algorithm did not notably improve the results obtained by the PCA itself, only better grouping the spectra obtained at day 0 (Figure 5.26). The cross-validation of the model was performed with a training and testing algorithm, with the prediction results summarised in Table 5.7. As explained in Section 2.6.3, two thirds of the data were used to train the model and the other third was used for the predictions. The model correctly classified all spectra from day 0, but spectra from day 7 were wrongly predicted to days 14 and 21. The spectra from days 14 and 21, for which the characteristic lipid peaks were well distinguished, were not incorrectly classified as belonging to days 0 or 7. However, they were wrongly predicted among days

14 an 21. Although the algorithm correctly discriminated all undifferentiated cells, the percentage of cells correctly classified is 73% if all four time-points are considered. This percentage shows that the model is not able to discriminate between the adipocytes at different time-points, such as days 7, 14 and 21. The model showed a very clear progression of the differentiation and it was good at discriminating the undifferentiated cells (day 0) from the adipocytes (days 7, 14 and 21), which is indicated by the absence of incorrect predictions to day 0.

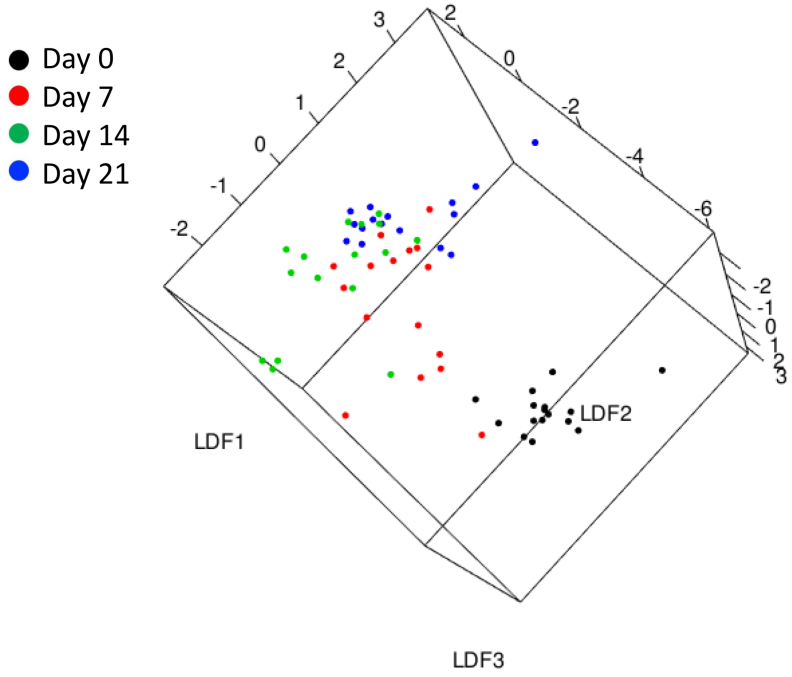


Figure 5.26: PCA-LDA scores for the adipo-induced MSCs, showing the phenotypic transition from undifferentiated cells (day 0) to adipocytes (days 7, 14 and 21).

Table 5.7: Confusion matrix of the PCA-LDA classification algorithm results for the adipo-induced MSCs.

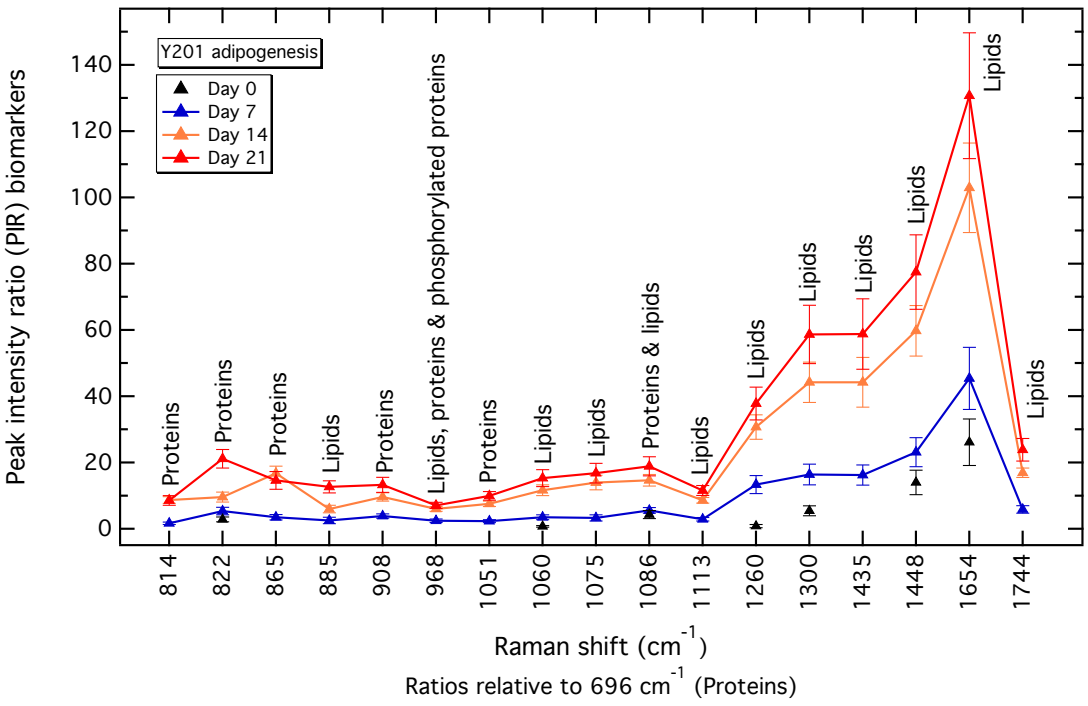
	Predicted time-point			
Time-point	Day 0	Day 7	Day 14	Day 21
Day 0	6	0	0	0
Day 7	0	4	1	1
Day 14	0	0	2	3
Day 21	0	0	1	4

5.3.2 PIR analysis of the adipo-induced MSCs

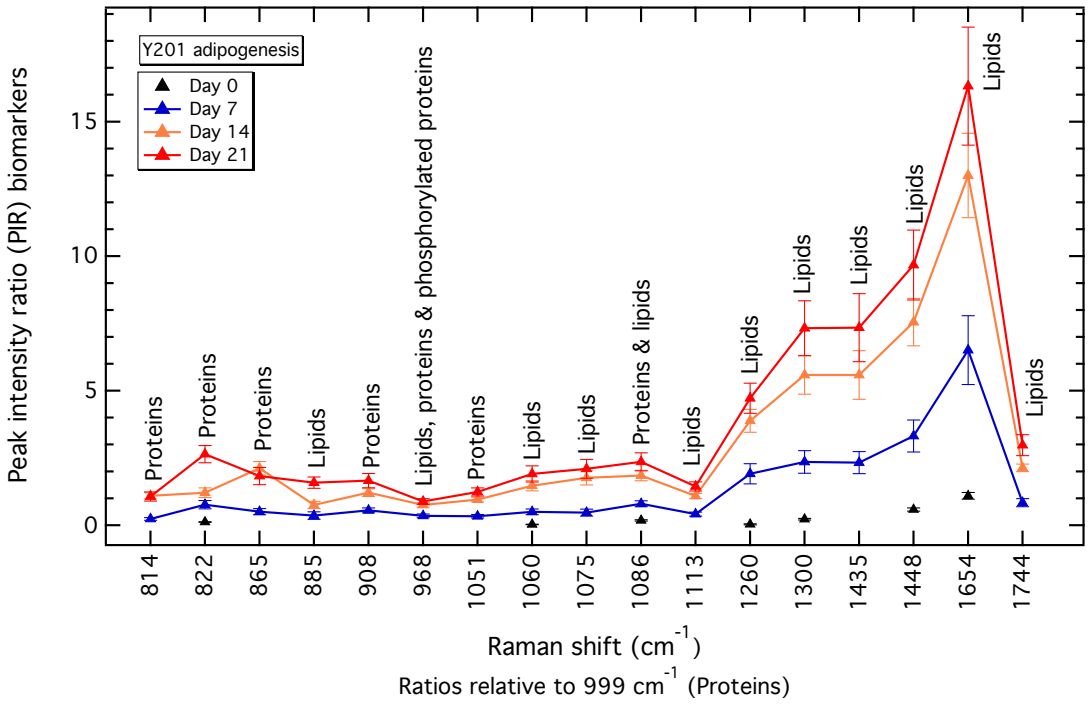
From the results of the multivariate analyses it was clear that presence of the sharp lipid peaks in the spectra and their relation to the protein peaks could be used as markers for adipogenesis. PCA and PCA-LDA clearly separated the spectra of the undifferentiated cells (at day 0) from the spectra of the adipocytes (at days 7, 14 and 21). However, it only separates to a certain degree the adipo-induced MSCs at days 7, 14 and 21. Thus, the ratiometric analysis (Section 2.6.1) was applied to the data obtained at days 0, 7, 14 and 21, in order to identify if there are markers that could be related to the stage of the differentiation process.

The PC1 loading indicated the phenylalanine peak at 999 cm^{-1} and the sharp lipid peaks as major contributors of the scores obtained in PC1 (Figure 5.25a); nonetheless, ratios were calculated between all identified peaks. From the ratiometric analysis, it was noted that PIRs of lipid peaks against other lipid peaks did not discriminate the time points. This suggests that the lipid type was the same throughout the time-points and that the relative intensity of the lipid peaks increased equally during adipogenesis. The increase in the relative intensity of lipid peaks has been suggested to be caused by the increase of the thickness of the cells [138].

Discriminating markers were only found against the protein peaks at 696 cm^{-1} and 999 cm^{-1} assigned to the C-S vibration in methionine and to phenylalanine, respectively (Figure 5.27). Many of the peaks observed in the spectra of the adipo-induced MSCs were not present in the spectra of the undifferentiated cells (day 0). Thus, in Figure 5.27, the PIRs obtained at day 0 are not connected by a line.



(a)



(b)

Figure 5.27: PIRs against the (a) 696 and (b) 999 m^{-1} peaks for the adipo-induced MSCs, showing markers for the time-point discrimination. The error bars are the propagation of the fitted intensity SE.

The 696 cm^{-1} peak was not identified in the PCA loadings (Figure 5.25) and this could be related to the small contribution of this peak to the PC scores. The markers obtained against the phenylalanine peak (999 cm^{-1}) for the ratios of peaks at 1260 cm^{-1}, 1300 cm^{-1}, 1435 cm^{-1}, 1448 cm^{-1}, 1654 cm^{-1} and 1744 cm^{-1} agree with the PC1 loading where they were highlighted as major contributors of this PC scores (Figure 5.25a).

For both ratios, the adipo-induced MSCs at day 7 have markers well separated from those at days 14 and 21. Notably, the relative intensity of lipid peaks against proteins has substantially increased from day 7 to days 14 and 21, as seen on the PIRs from 1260 cm^{-1} to 1654 cm^{-1}. However, the majority of PIRs were not able to resolve the spectra obtained at days 14 and 21, resembling the results obtained by the PCA and PCA-LDA. The only discriminatory markers found for all time-points are the PIRs of the 822 cm^{-1}, 885 cm^{-1} and 1744 cm^{-1} peaks, for both ratios. The 822 cm^{-1} peak is assigned to proteins and DNA/RNA vibrations, while the 885 cm^{-1} and at 1744 cm^{-1} peaks are assigned, respectively, to the C-O-C skeletal and C=O stretch vibrations in lipids. These PIRs could, therefore, be used for the full discrimination of the adipogenic stage, from day 7 to day 21 of the differentiation process. However, due to the difficulty in fitting the very convoluted region from 800 cm^{-1} to 950 cm^{-1}, the single peak at 1744 cm^{-1} is a better candidate. In addition, the phenylalanine peak (999 cm^{-1}) is well defined at all time-points, thus ratios against this peak are favourable over those calculated against the 696 cm^{-1} peak. In summary, ratios against the phenylalanine peak can be used as markers of the stage of the adipogenic differentiation and the 1744/999 PIR for its full discrimination.

5.3.3 Identification of the lipid species

The spectra obtained from the adipo-induced MSCs (from day 7 onwards) resulted mainly from the signal captured from lipids stored in the lipid droplets. Thus, they were compared to the spectra of other pure fatty acids and lipids in order to attempt the identification of the stored lipids. From a literature search on databases of Raman spectra obtained from lipids and fatty acids, it was inferred that the adipo-induced cell spectrum was similar to that of oleic acid, triolein or tripalmitolein [196, 198].

It is known that lipid droplets are an important part of the energy storage system of cells and that they are formed by a core of neutral lipids, such as triacylglycerols (TAG) and sterol esters, enclosed by a monolayer of phospholipids and proteins [199]. Thus, other pure fatty acids and phospholipids commonly found in membranes were also included in the analysis. All lipids and fatty acids analysed are listed on Table 2.5 from Section 2.5 together with the methods for sample preparation and the Raman acquisition parameters. Although it was indicated by the literature search that tripalmitolein has a similar spectra to that obtained from the cells, a sample of tripalmitolein was not available for the comparisons.

From the data collected, it was noted that the spectra of oleic acid and triolein (Figure 5.28) were very similar with those obtained from the adipo-induced MSCs (Figure 5.23). The spectra obtained from all other listed lipids are presented in Appendix D, from Figure D.1 to D.5.

Triolein is a triacylglyceride that is composed of unsaturated oleic acid and is derived from glycerol [198]. It can be seen that the major spectral difference between triolein and oleic acid is the absence, on the latter, of the 1744 cm^{-1} peak, which corresponds to the C=O stretch [198]. Therefore, among the lipids analysed in this study, triolein is the closest candidate for the signature observed in the adipo-induced MSCs. However,

Table 5.8: Raman characteristics of the adipo-induced MSCs intracellular lipid and triolein.

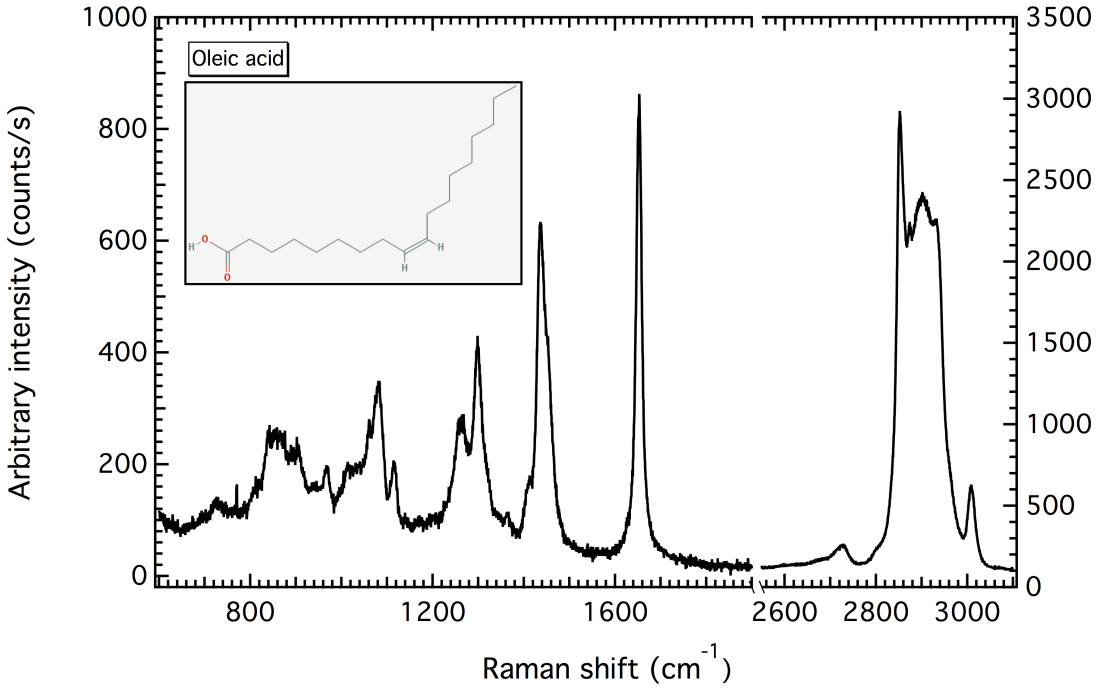
	Unsaturation Degree		TUFA/TFA	Chain length	
	(1654/1448)	(1654/1308)	(3007/2850)	(2930/2958)	(1448/1744)
Day 0	–	–	–	–	–
Day 7	2.8 ± 0.7	2.8 ± 0.7	0.4 ± 0.1	1.1 ± 0.2	2.8 ± 0.7
Day 14	2.3 ± 0.4	2.3 ± 0.4	0.30 ± 0.07	1.2 ± 0.2	2.6 ± 0.4
Day 21	2.2 ± 0.5	2.2 ± 0.4	0.30 ± 0.07	1.3 ± 0.2	2.5 ± 0.5
Triolein	2.3 ± 0.1	2.5 ± 0.1	0.30 ± 0.01	1.70 ± 0.08	2.6 ± 0.1

triolein and tripalmitolein are TAGs with very similar Raman signatures [198] and thus, also similar to the spectra obtained from the adipo-induced MSCs.

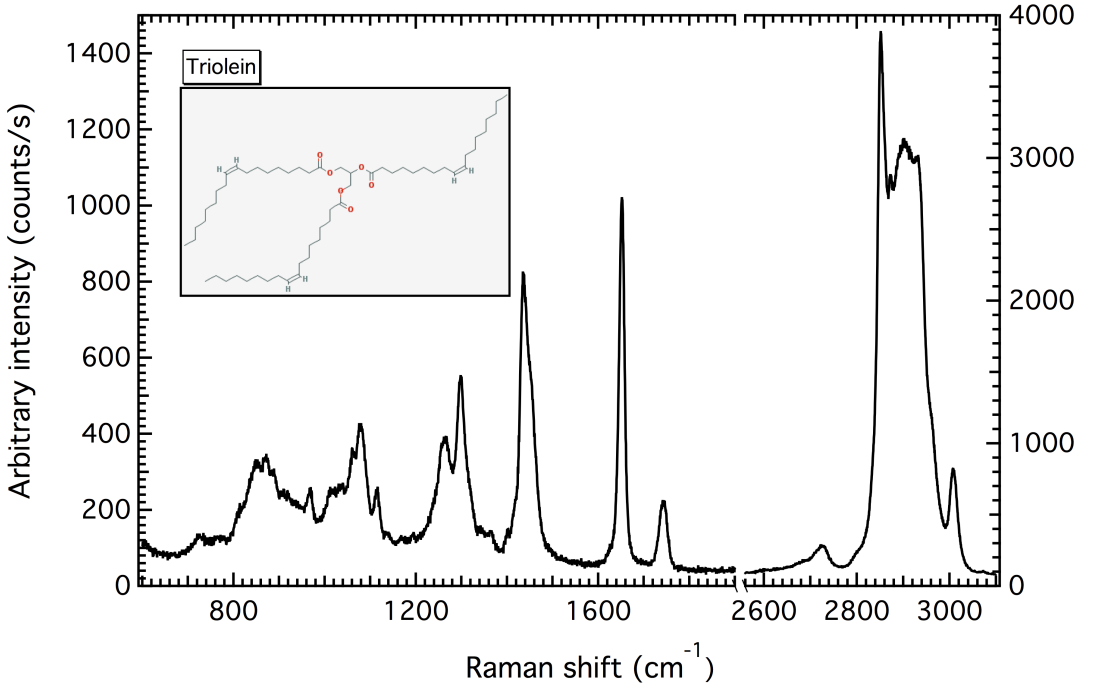
In order to have a better understanding of the lipids found in the adipo-induced MSCs, a lipidomic analysis was performed. This consists of obtaining the unsaturation degree of the fatty acid, an indication of its chain length and the relative quantities of unsaturated fatty acids to the total amount of fatty acids (TUFA/TFA) by calculating PIRs from certain Raman peaks [24, 54, 72]. The unsaturation degree is obtained by the 1654/1448 (C=C stretch/CH_2 scissoring) and 1654/1300 (C=C stretch/CH_2 twisting) PIRs, while the chain length is calculated from the 2930/2958 (CH_2 stretch/CH_3 stretch) PIR and the TUFA/TFA from the 3010/2850 (=CH stretch/CH_2 stretch) PIR [24, 54].

These ratios were compared between triolein and the time-points at days 7, 14 and 21 from the adipo-induced MSCs (Table 5.8). The lipidomics show no difference between the ratios obtained from the cells, indicating that the same lipid type was present at all time-points, in agreement with the results found so far via univariate and multivariate analyses. Triolein displayed an unsaturation degree, chain length and TUFA/TFA that are similar to those obtained for the adipo-induced MSCs.

The chain length ratio is not the value of the lipid chain length itself, but an indication if its size has changed from one sample to the other. It can be seen that for both chain



(a)



(b)

Figure 5.28: Raman spectra obtained from (a) oleic acid and (b) triolein. Note the different range on the left and right Y-axes. The insert window shows the chemical structure of each lipid as per [200].

length ratios, the lipid found in the adipo-induced MSCs had comparable values to those of triolein. Based on these results it can be assumed that the lipid droplets formed during adipogenesis contain a high level of triolein. However, since tripalmitolein has a very similar Raman signature to triolein, it could also lead to analogous results.

Adipocytes from the white adipose tissue (WAT) contain a single large lipid droplet (10-100 μm) in its cytoplasm, thus being called unilocular adipocytes; in contrast, the multilocular adipocytes from the brown adipose tissue (BAT) contain several small lipid droplets (2-10 μm) [201, 202]. The size of the adipocyte is related to the cell lipid content [201] and its volume is a determinant of cell's functionality; larger adipocytes exhibit higher metabolic activity and secrete more chemoattractants to immune cells [203]. WAT are a site for fuel storage and act as a regulatory centre of energy; in contrast, the BAT protect the organism from hypothermia once they are specialised in thermogenesis [201].

From the Oil red O staining (Figure 5.21) and optical images saved during Raman inspection at day 21 (Figure 5.29), it can be observed that the lipid droplets formed on the adipo-induced MSCs seem to be multilocular. Their multilocular architecture could indicate that these cells may form brown adipose tissue; however, at this point and without further experiments it is not possible to further evaluate their functionality.

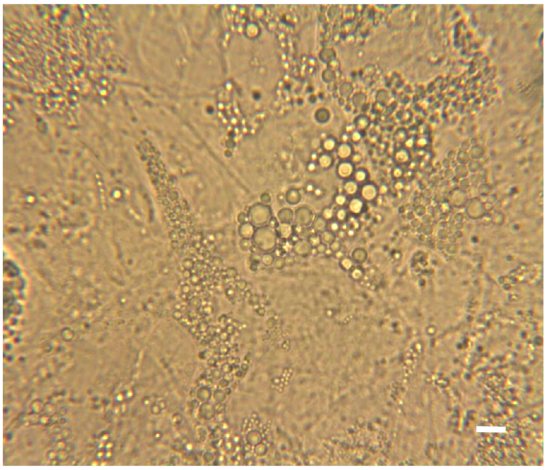


Figure 5.29: Optical image of the adipo-induced MSCs at day 21 shows multilocular lipid droplet structure. Scale bar is 10 μm.

The indication by RS for a major presence of triolein, or possibly tripalmitolein, in the lipid droplets is in accordance to the fact that most lipid bodies store mainly TAGs [199]. The fact that MSCs can differentiate into adipocytes may be linked to the observation that the bone marrow is partially replaced by adipose tissue with ageing; however, some forms of osteoporosis may be linked to an increased tendency of osteoblasts precursors to differentiate into adipocytes [204]. These initial results therefore shine some light on knowing which lipid species are stored in the lipid droplets and help to understand the functionality of the adipocytes differentiated from the MSCs.

5.4 Chapter conclusions

Both osteogenesis and adipogenesis were easily detected by RS due to the contrasting spectra between undifferentiated and differentiated cells. Osteogenesis was observed from day 7 onwards in the form of mineralised nodules deposited by the cells. Adipocytes were also noticed from day 7 onwards and their spectra contained the intense signature of lipids.

The convergence tests verified that the averaged spectra used in the analyses repre-

sented the population average for each experiment. However, the higher converged %SE obtained for the cells under differentiation, in comparison to undifferentiated cells, reflected the diversity of such populations. Mineralised nodules of different sizes and shapes were observed during osteogenesis and the relative intensity of the lipids peaks increased during adipogenesis.

The use of ratiometric analysis allowed the assessment of the mineral deposited by the cells during osteogenesis. The mineralised nodules produced by the Y201 MSCs after 21 days of osteogenic induction were found to have similar architecture to adult human bone, composed of different mineral species. The position of the phosphate group (PO_4^{3-} ν_1) indicated that B-type carbonate substituted apatite and disordered amorphous phosphate were present in both mineralised nodules and native human bone samples. After 21 days, a reduction of the disordered amorphous phosphate species contributed to an increase of the mineral crystallinity of the nodules, and values comparable to those of native bone were obtained. Therefore, the fact that amorphous disordered and B-type carbonate substituted apatite were found in the osteo-induced MSCs suggests that they are producing a mineral content resembling that of native bone.

On the other hand, although the mineral matrix was nearly formed and had characteristics similar to those of bone, the same could not be assumed for the collagen-rich protein matrix. Peaks assigned to proline and hydroxyproline had merely been distinguished from the spectrum background, while peaks related to amide III, amide I and CH vibrations were not well resolved in the spectra obtained at day 21 of the osteo-induced cells.

The multivariate analyses and PIRs indicated that markers of osteogenesis comprised not only peaks assigned to the mineral apatite, but also to DNA/RNA and proteins. The PIR analysis revealed that ratios against lipids (717 cm^{-1}), proteins (933 cm^{-1}), DNA/RNA (781 cm^{-1}) and bands with overlapping vibrations of lipids and proteins with

DNA/RNA (1102 cm^{-1} and 1482 cm^{-1}) also provided discriminating markers for osteogenesis and its evolution from day 0 to day 21.

In addition, ratios against the peak at 933 cm^{-1} could indicate the start of the formation of the collagen-rich protein matrix during osteogensis, since this band is observed in the spectra of type-II and type-I collagen. However, to securely use this peak as a marker for the development of the protein matrix, the MSC line would need to be tested for their expression of the genes *Sox9* (gene for cartilage formation) and *Col2a1* (type-II collagen gene).

As expected, the PIRs obtained against disordered apatite (945 cm^{-1}) and hydroxyapatite (956 cm^{-1}) provided discriminating markers of osteogenesis. PIRs obtained from the mineral carbonate (1069 cm^{-1}) also showed a good separation between initial (day 7) and late time-points (day 21) of osteogenesis and can be additionally used.

PIRs against the 781 cm^{-1} (DNA/RNA), 1069 cm^{-1} (lipids, phospholipids and DNA/RNA on the undifferentiated cells and mineral carbonate for the differentiated MSCs) and 1482 cm^{-1} (proteins and DNA/RNA) peaks also provided discriminating markers for the four undifferentiated MSC lines (Y101, Y201, Y102 and Y202). Thus, these PIRs can be used as predictive markers of osteogenesis competence in undifferentiated cells and then tracked during osteogenic differentiation.

Adipogenesis revealed a dramatic change in the spectral signature of the MSCs once the spectra of the adipo-induced MSCs was dominated by lipid peaks. The main differences between the non-differentiated and the adipocytes was noticed by the rising of very sharp lipid peaks on regions of 1448 cm^{-1} and 1654 cm^{-1}. T

The PIRs revealed that the lipid species remained constant throughout the time-points, and only its relative intensity increased. In accordance to the multivariate analysis, PIR markers were identified against the phenylalanine peak (999 cm^{-1}); thus, corroborating

the PCA suggestion about the contributions of protein and lipid peaks to the variation observed in the spectra during the adipogenesis. The identification of adipogenesis markers based on the intensity of lipid peaks relative to phenylalanine is a straightforward method once the phenylalanine peak remained very well defined after 21 days of differentiation and it was not hindered by the lipid peaks.

Finally, RS by means of lipidomic analysis indicated the presence of triacylglycerols, either triolein or tripalmitolein as the major component of the lipid signature obtained from the lipid droplets. Optical images also showed a multilocular lipid droplet structure. However, further experiments are necessary to elucidate both adipocyte composition and functionality.

In summary, RS was demonstrated as an important method to characterise differentiation and also to identify markers for osteogenesis and adipogenesis. The use of PIRs to obtain biomolecular markers proved to be a robust method once it was able to indicate the stage of differentiation for either osteogenesis and adipogenesis. RS, coupled with a careful data analysis, provided an insight into the cell's chemical features, allowing trackable markers of the cell biomolecular changes to be unlocked.

Chapter 6

Conclusions

In this thesis the potential of RS to characterise cells, evaluate cell population heterogeneity, discriminate subpopulations of cells and assess differentiation was investigated. Thus, RS was applied to the analysis of mesenchymal stromal cells, which form heterogeneous populations containing cells that may or may not have progenitor characteristics and have dissimilar potential for differentiation. Although these cells are being used in clinical trials, there is a lack of unique markers for their identification.

This thesis showed that RS was able to discriminate homologous MSCs subpopulations. The discriminated subpopulations were four hTERT immortalised clonal MSC lines (Y101, Y201, Y102 and Y202) that have not been totally distinguished by global gene expression analyses. These four cell lines express the same surface proteins from a panel widely used to identify MSCs; however, the 02 cell lines are incapable of trilineage differentiation. RS, combined with PCA-LDA and peak intensity ratio analyses, discriminated between the differentiation competent (01s) and incompetent (02s) cell lines, in experiments performed on air-dried and live cells.

In the air-dried experiments, the discrimination of 01s and 02s was obtained for peak intensity ratios against the 932 cm^{-1} (proteins), 1060 cm^{-1} (lipids, carbohydrates, phospho-

lipids and DNA/RNA), 1085 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA), 1549 cm^{-1} (proteins) and 1615 cm^{-1} (proteins) peaks. The discrimination of the four cell lines was obtained, for these same experiments, against the 970 cm^{-1} (proteins and DNA/RNA) peak. The results from Chapter 3 revealed that the discrimination of the MSC lines obtained against the 970 cm^{-1}, 1549 cm^{-1} and 1615 cm^{-1} peaks may point to a different reaction of these cells to the air-drying process. The discriminatory Raman markers obtained against these peaks were related to different relative intensities of peaks assigned to phosphorylated proteins (970 cm^{-1}), which are known to increase during desiccation, and to tyrosine and tryptophan (1549 cm^{-1} and 1615 cm^{-1}), known to be linked to cell death mechanisms.

For the live cell experiments, all discriminatory Raman markers were obtained from intensity ratios against peaks related to DNA/RNA vibrations. Thus, Chapter 4 indicates that although the air-dried cells were affected by the air-drying process, they retained some of the relative differences between DNA/RNA and proteins that were observed for the live cells. In the live cells, many other Raman markers for the total discrimination of the four cell lines were obtained, such as the intensity ratios against the 781 cm^{-1} (DNA/RNA), 1069 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1482 cm^{-1} (proteins and DNA/RNA) peaks. Trilineage differentiation competence discrimination of the MSC lines was obtained against the intensity ratios for the 672 cm^{-1} (DNA/RNA), 744 cm^{-1} (DNA/RNA), 1088 cm^{-1} (lipids, carbohydrates, phospholipids and DNA/RNA) and 1422 cm^{-1} (proteins, lipids and DNA/RNA) peaks.

It is evident that the discrimination of the MSC lines was pertained to relative differences between DNA/RNA and proteins. Although these findings are consistent with previous RS research performed on differentiated and non-differentiated embryonic stem cells, which discriminated them based on the relative intensities between DNA/RNA and

proteins, in this thesis these relative differences were verified for undifferentiated subpopulations of MSCs. Therefore, they could be used to discriminate cells that are capable of trilineage differentiation, providing new markers for their characterisation. This discrimination, based in DNA/RNA and protein relative differences, is in line with a nuclear proteomic study, also performed in homologous undifferentiated MSCs, that disclosed differences in the expression of proteins between differentiation competent and incompetent cells.

The Raman maps, although performed on air-dried samples, revealed morphological differences in the nucleus of the analysed MSCs. Although the four cell lines were cell-cycle synchronised, the 02s (differentiation incompetent lines) presented regions with high relative intensity of proteins. These regions were attributed to re-assembled nucleoli as many more of these cells were found in the S phase after synchronisation. Despite the fact that only one Raman map was performed per cell line, the peak intensity ratio markers obtained for these cells agreed with those obtained for their population. This suggests that this morphological difference has also driven their discrimination.

This thesis also proposed to study the heterogeneity of cell populations. The use of tests associated to the standard error of the mean and standard deviation allowed a measure of the heterogeneity degree of such a cell population. The tests showed that the air-dried cells were more heterogeneous than the live cells, and also, that differentiated MSCs were more heterogeneous than undifferentiated cells. This pointed to an understanding of how the experimental set-up influences the results, providing a tool to evaluate how many spectra are needed to characterise a cell population. Unfortunately, due to the nature of the tests, which were developed based on the uncertainties associated to the averaged spectra, they may not be suitable to evaluate the quantity of spectra needed if multivariate analyses are being performed.

In Chapter 5, the MSCs were investigated during osteogenesis and adipogenesis; however, due to the nature of the differentiation experiments, the RS results reflect one experimental repeat. The results showed that the osteo-induced Y201 MSC line formed mineralised nodules with similar architecture to native human bone. Although the collagen-rich protein matrix of the nodules did not seem completely formed after 21 days of osteogenic induction, the obtained results are compatible with other studies also performed on osteo-induced MSCs. The peak intensity ratio analyses provided Raman markers other than those usually associated to the presence of hydroxyapatite and mineral carbonate. They were related to proteins, lipids and DNA/RNA vibrations, such as those obtained against the 717 cm^{-1}, 933 cm^{-1}, 781 cm^{-1}, 1102 cm^{-1} and 1482 cm^{-1} peaks. Uniquely, these markers show a clear progression of the osteogenic differentiation. This thesis also provided markers that could be used to predict differentiation competence in undifferentiated cells and then, followed throughout osteogenesis as the peak intensity ratios obtained against the 781 cm^{-1}, 1069 cm^{-1} and 1482 cm^{-1} peaks. The markers obtained during adipogenesis revealed that proteins and lipids have opposing behaviours, which correlates with studies already performed in adipose-derived stem cells. The increase in the relative intensity of lipids can be measured against the phenylalanine peak (999 cm^{-1}), thus providing not only Raman markers of the adipogenesis but also of its progression. Unlike other research performed during adipogenesis, this thesis also attempted the identification of the lipids species stored in the lipid droplets. The results suggested high levels of triolein and/or tripalmitolein in the adipo-induced MSCs.

The PCA-LDA proved capable of classifying the undifferentiated MSCs and predicting their differentiation competence on both air-dried and live cell analyses. During the osteogenesis and adipogenesis, the multivariate analyses indicated the continuos phenotypic changes suffered by the MSCs during the differentiation process.

Considering the results obtained from this thesis, possible future research avenues are described next.

This thesis offered suggestive evidence that differences in the DNA/RNA relative to proteins made possible the discrimination of the MSC lines. These findings suggest that it is useful to focus on the study of the MSCs nuclear proteome, particularly prior to differentiation, in order to verify if these cell lines over-express different proteins.

Another avenue for further study would be research into the specific biomolecular changes caused by the air-drying process. The findings suggested that these cell lines were differently affected by it. Thus, this different modulation might uncover new markers for their discrimination.

Further research with primary MSCs should be expanded. Although the clonal hTERT MSC lines can be used as models for understanding the biomolecular differences between differentiation competent and incompetent cells, the hTERT over-expression can affect other biological functions. Thus, it is important to investigate such discrimination in primary cells.

The Raman mappings revealed morphological differences between these cell lines that could be linked to their discrimination; however, they were performed on air-dried cells. Although these mappings usually take around 19h to be completed, ultra-fast Raman mapping could be applied to investigate the nucleus morphology of live cells.

Osteogenesis revealed that the mineralised nodules produced by the osteo-induced MSCs had similar architecture to native human bone. Thus, it would be interesting to investigate the similarities of the adipo-induced MSCs with adipocytes from brown and white adipose tissues.

This research showed that RS is capable of performing subpopulation discrimination and of identifying markers linked to MSC function. Therefore, it can be used non-

destructively and as a label-free method for sorting these cell sub-types in heterogeneous primary MSC populations. This thesis also highlights that RS is an effective tool for cell characterisation and that it can complement standard methods for understanding the biomolecular diversity of cell populations.

Appendix A

General equations and formulas

A.1 Trigonometric identity for the product of two cosine functions

$$\cos\alpha\cos\beta = \frac{\cos(\alpha+\beta)+\cos(\alpha-\beta)}{2} \tag{A.1.1}$$

A.2 Exponential function used for the fitting of the %SE convergence tests

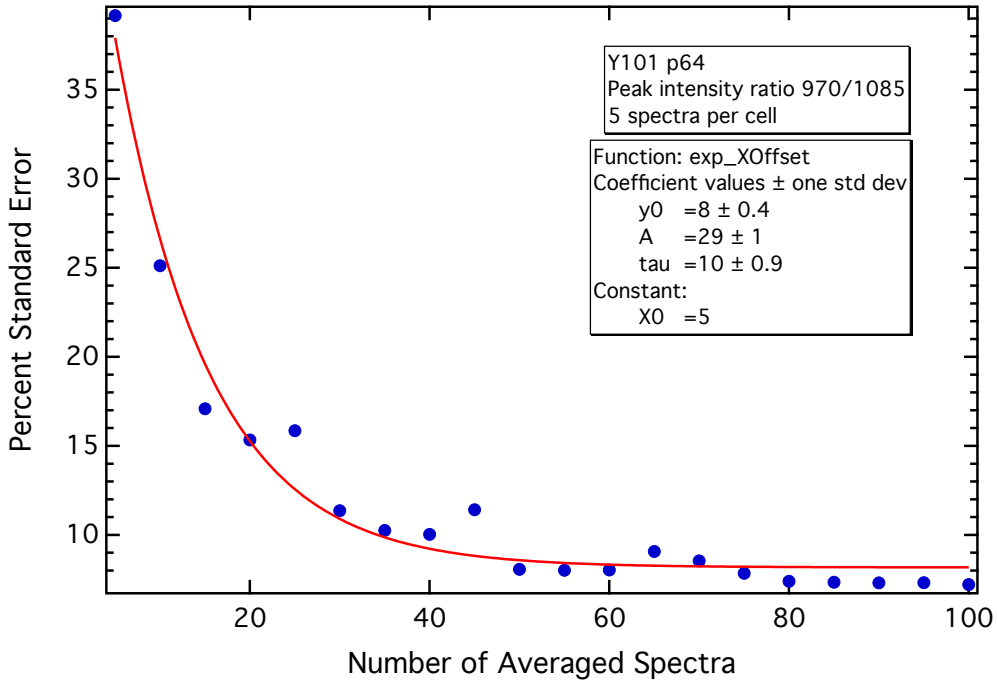
$$y = y_0 + Aexp\left(\frac{x-x_0}{\tau}\right), \tag{A.2.2}$$

where y_0 is the converged %SE value, A is the amplitude, x_0 is a constant representing the initial amount of averaged spectra and τ is the decay constant. This is a predefined exponential function from IGOR Pro version 6.32.

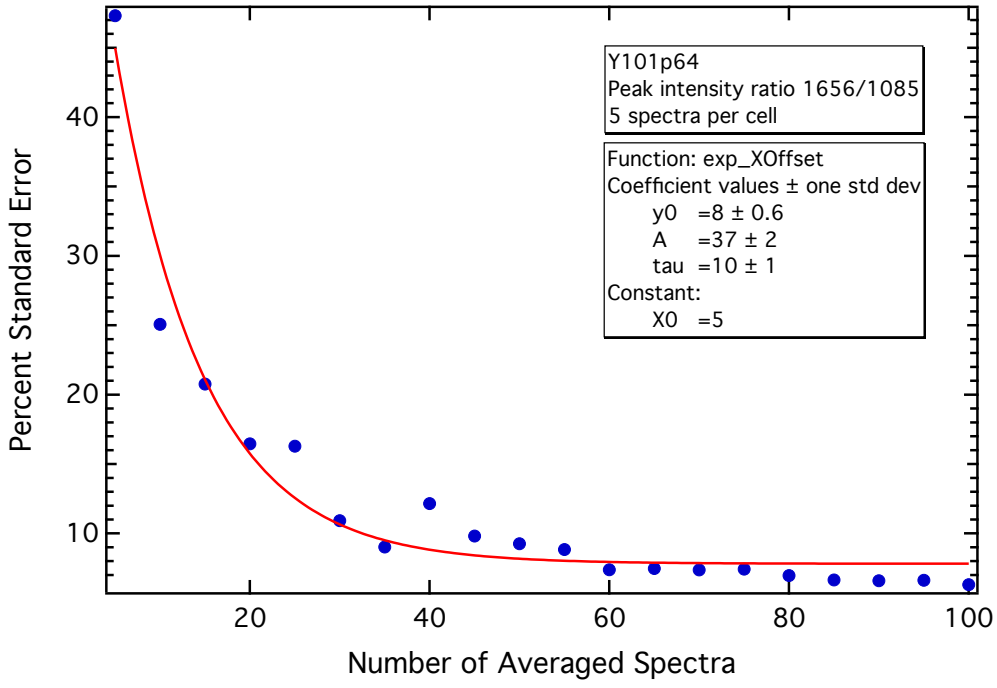
Appendix B

Convergence tests

B.1 Air-dried MSCs, HDF, CD317+ and K72 – %SE convergence tests

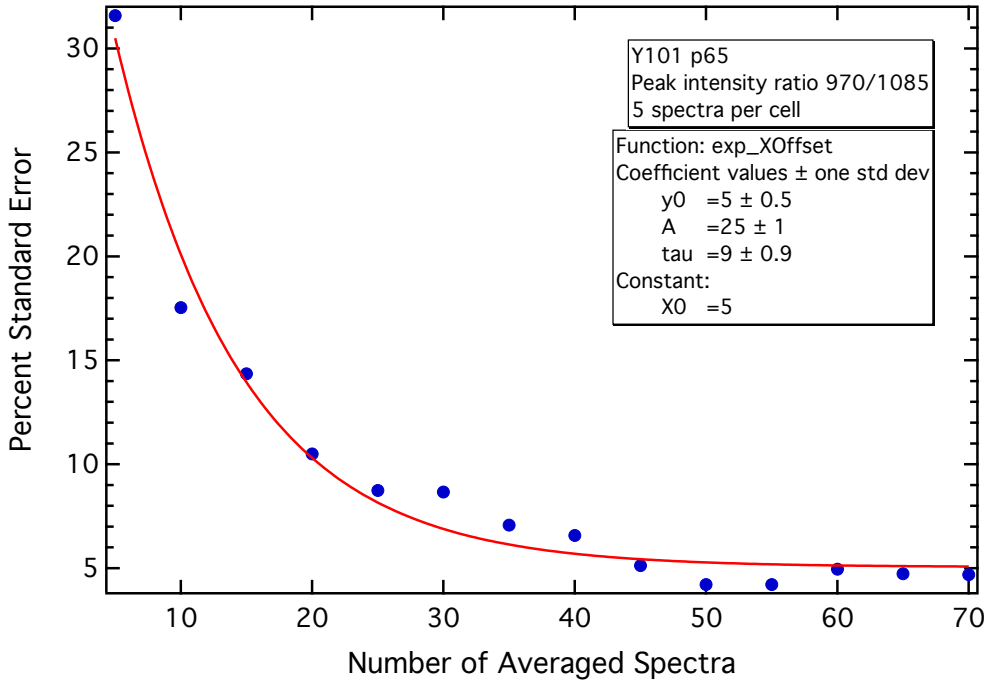


(a)

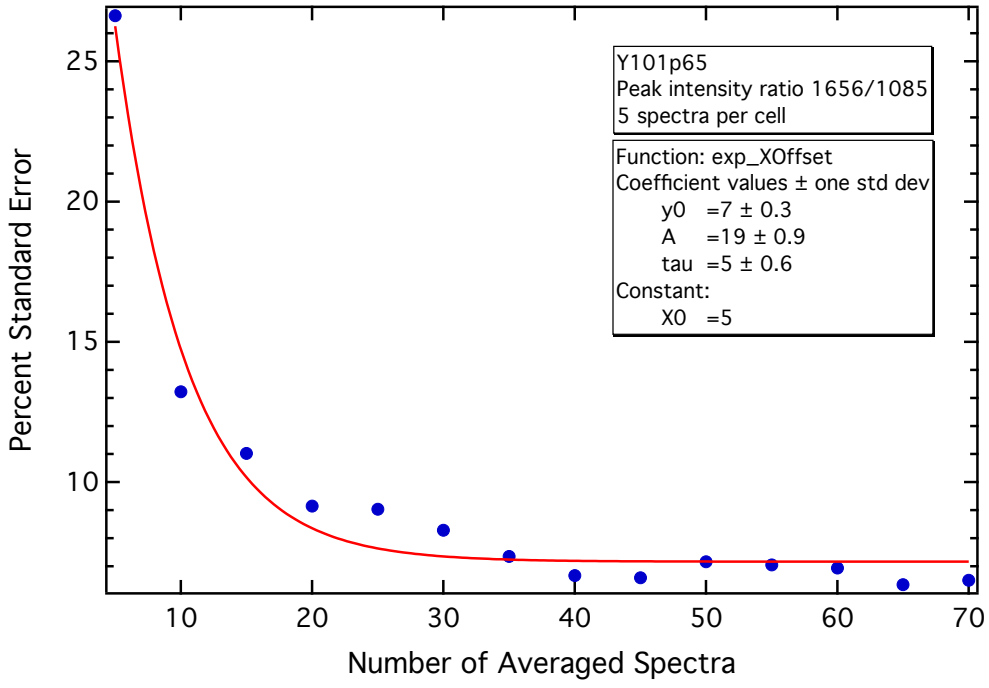


(b)

Figure B.1: %SE convergence tests of the Y101 MSC line for the air-dried analyses – experiment 1. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

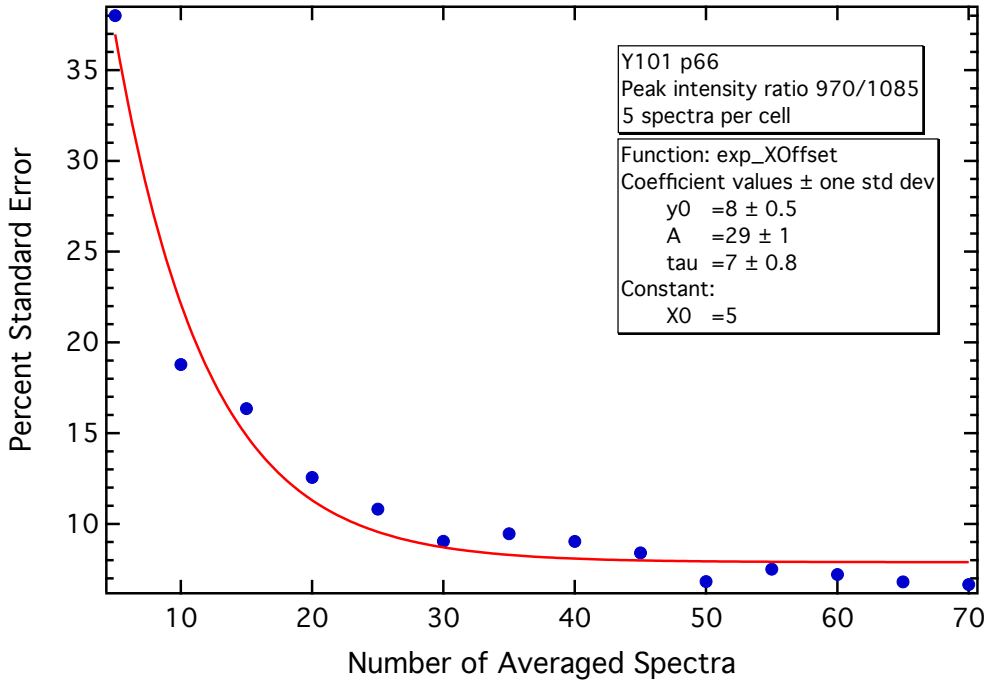


(a)

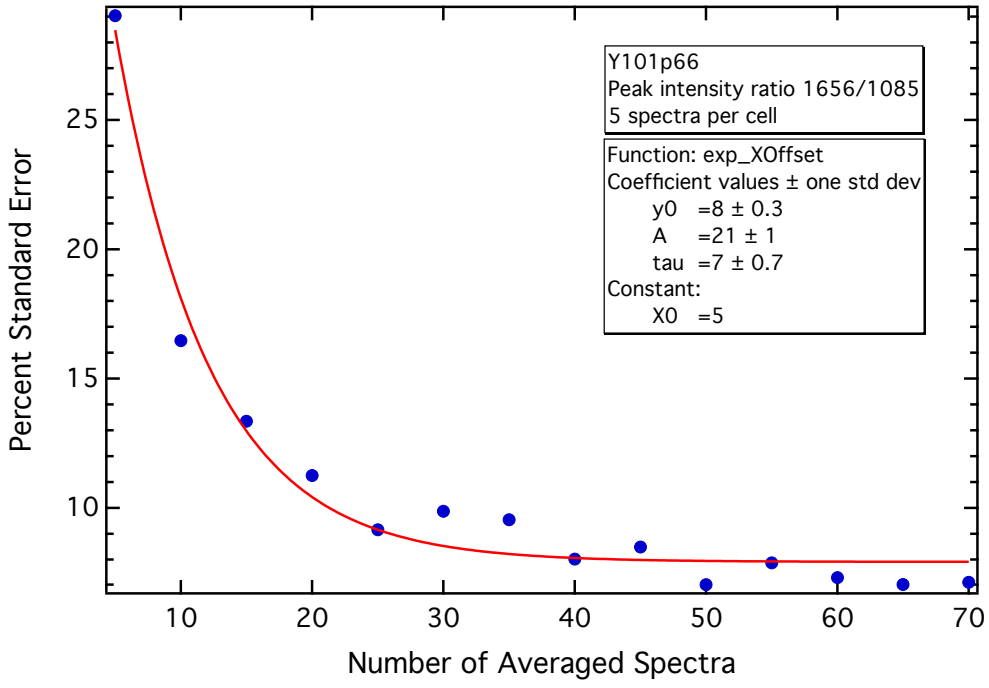


(b)

Figure B.2: %SE convergence tests of the Y101 MSC line for the air-dried analyses – experiment 2. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

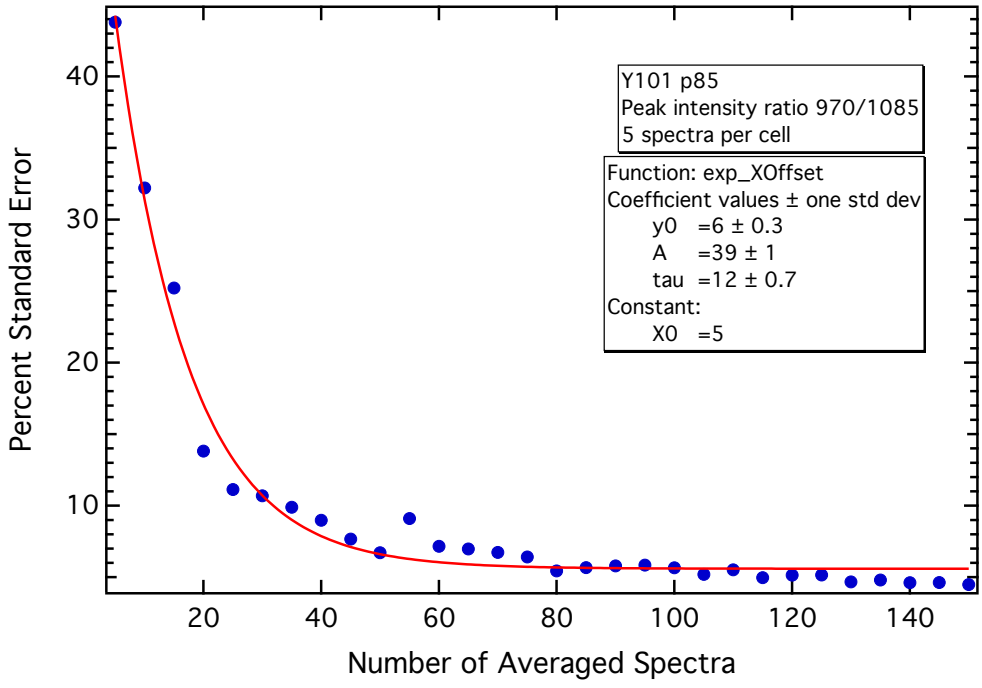


(a)

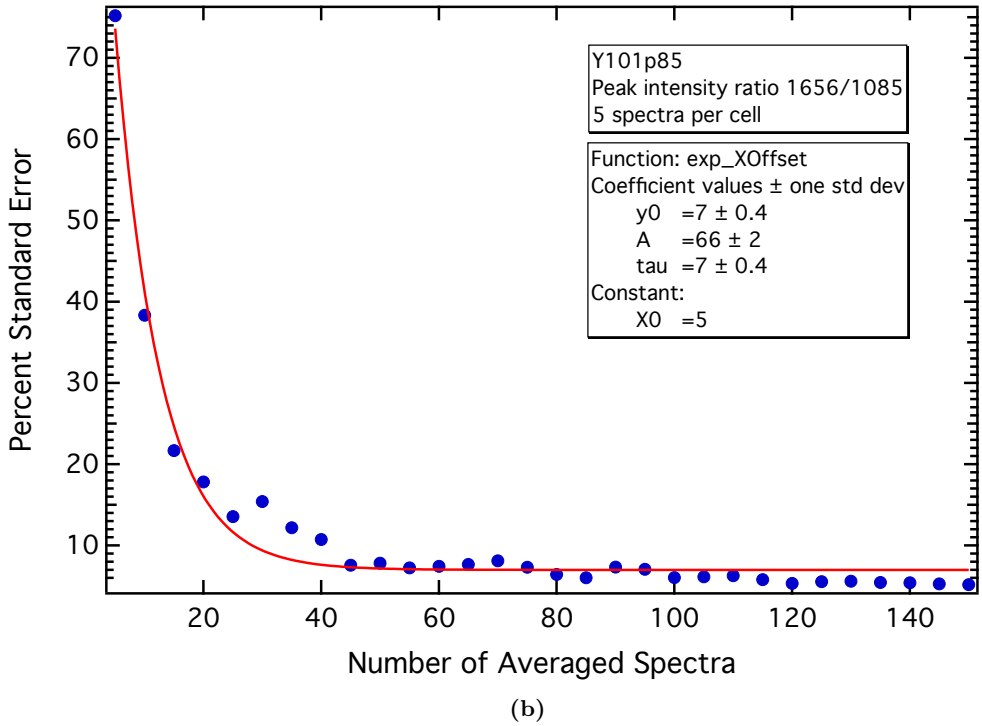


(b)

Figure B.3: %SE convergence tests of the Y101 MSC line for the air-dried analyses – experiment 3. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

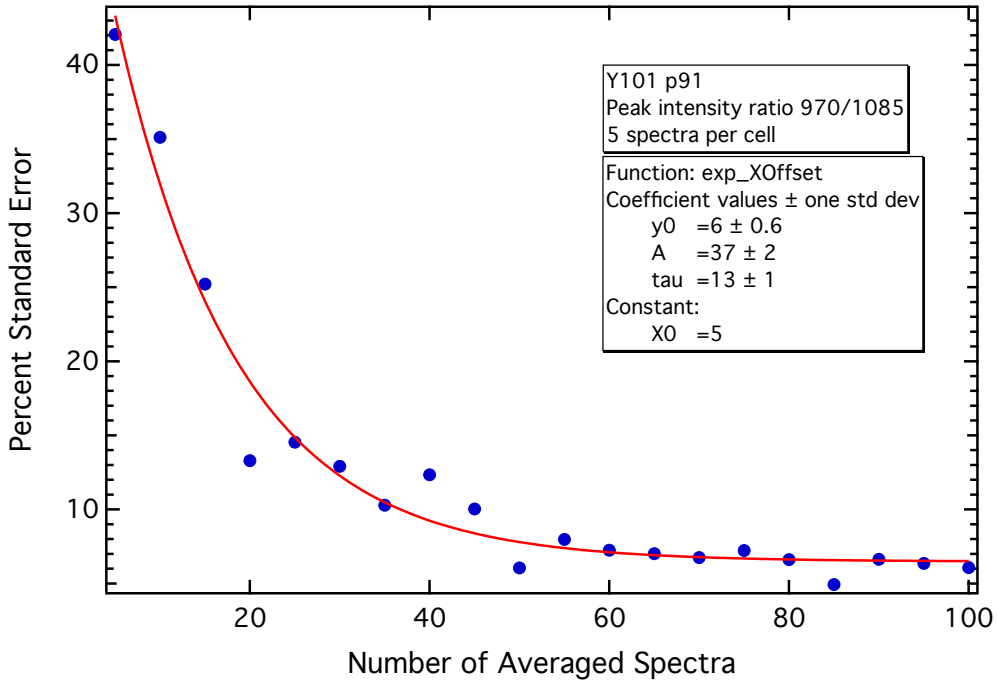


(a)

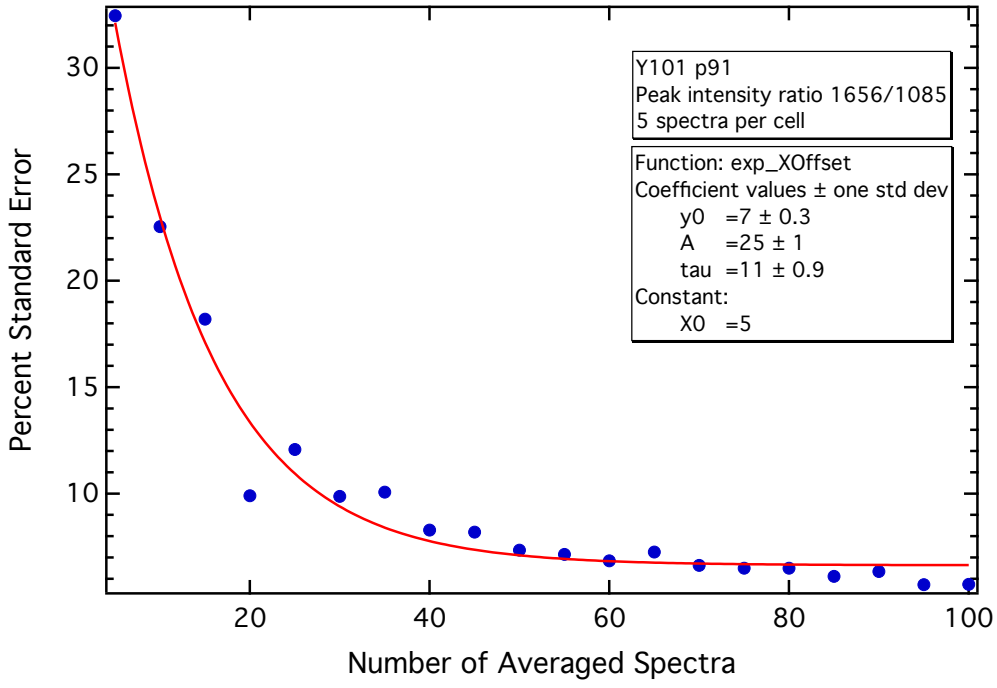


(b)

Figure B.4: %SE convergence tests of the Y101 MSC line for the air-dried analyses – experiment 4. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

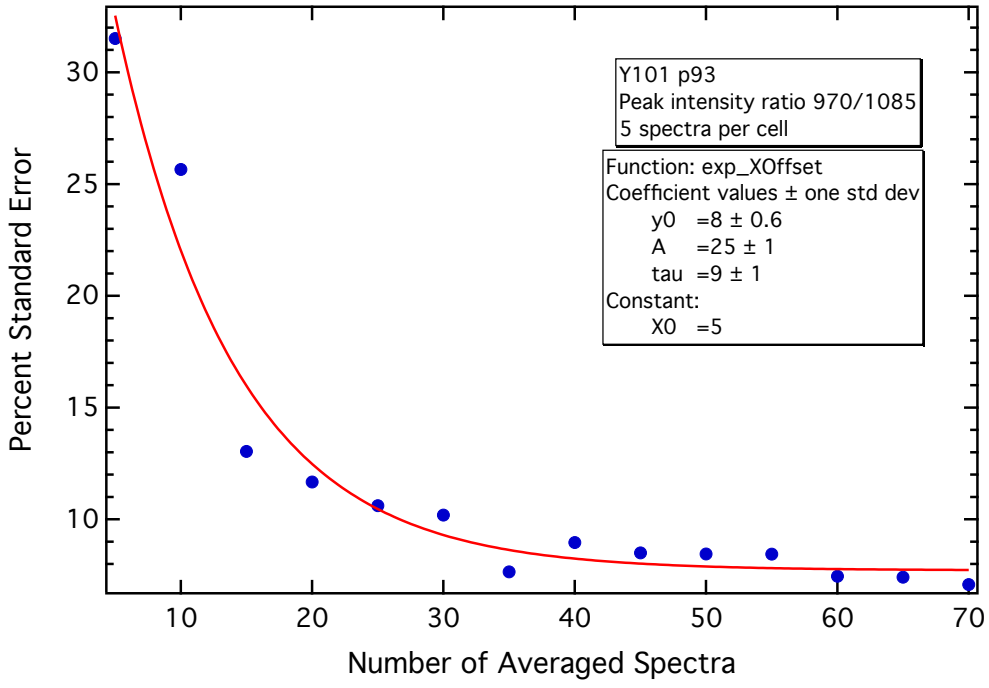


(a)

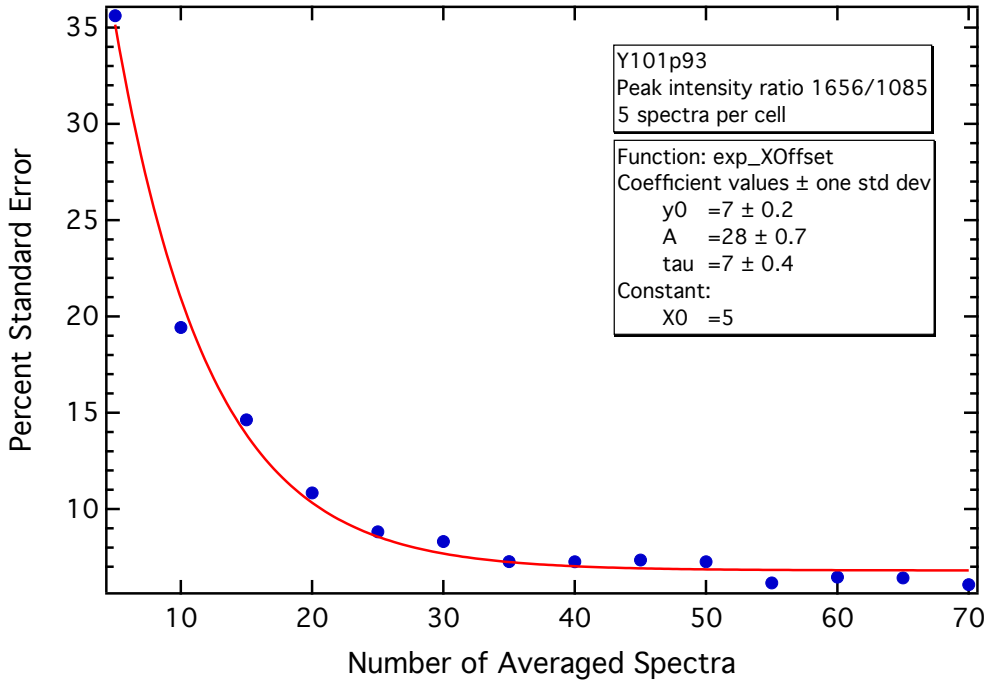


(b)

Figure B.5: %SE convergence tests of the Y101 MSC line for the air-dried analyses – experiment 5. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

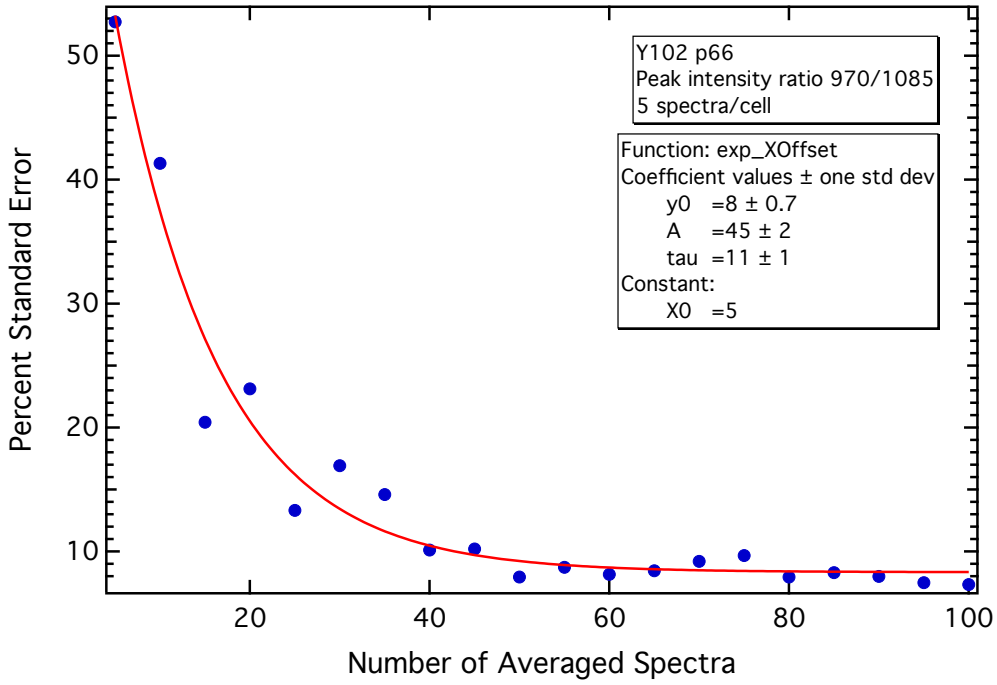


(a)

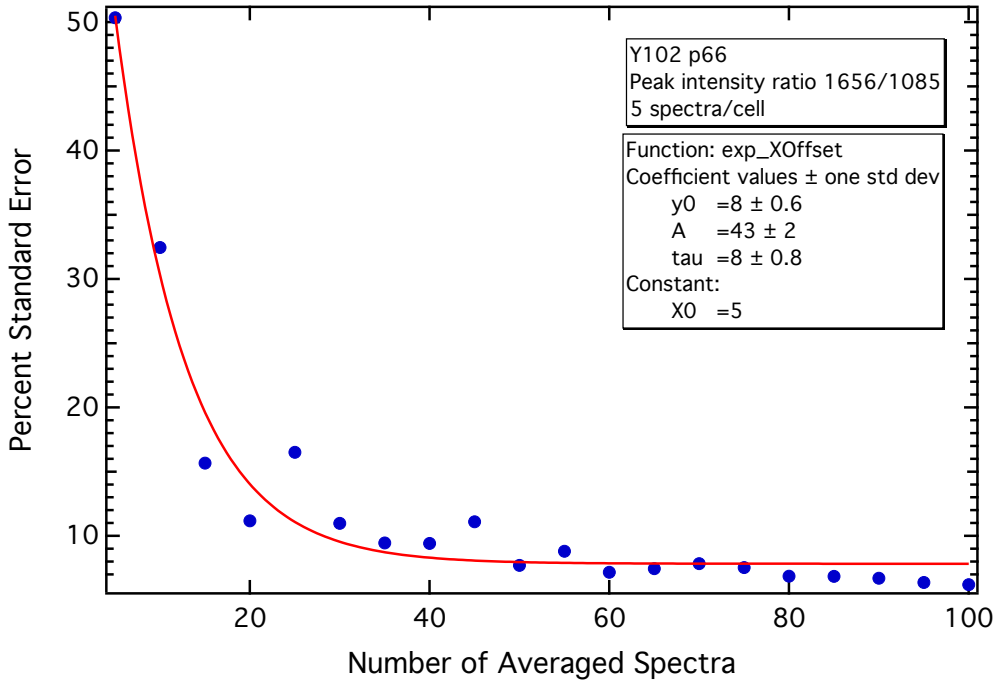


(b)

Figure B.6: %SE convergence tests of the Y101 MSC line for the air-dried analyses – experiment 6. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

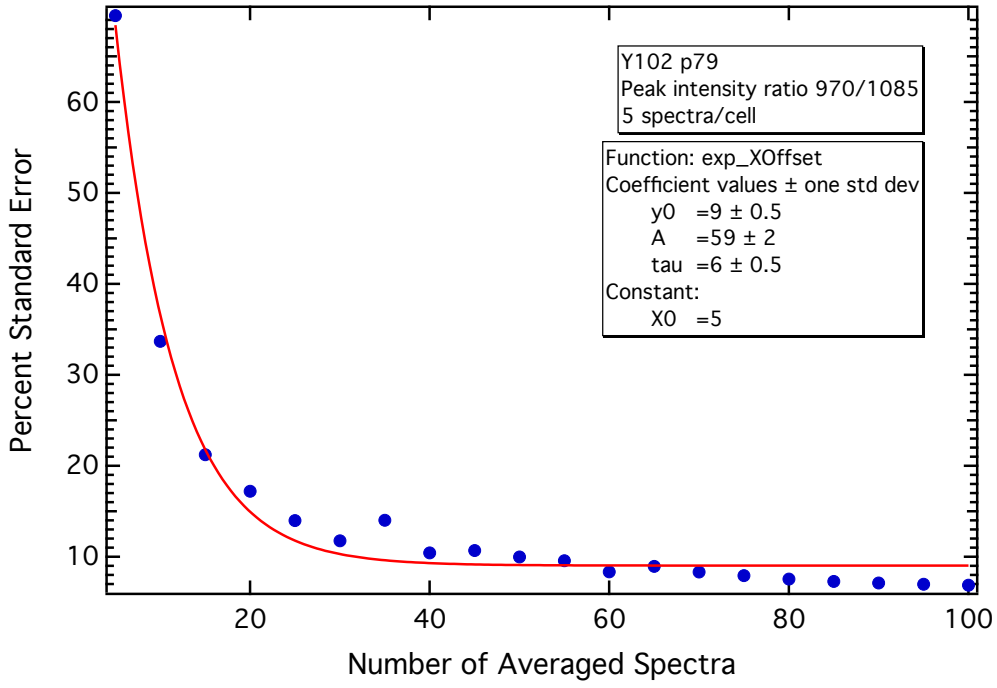


(a)

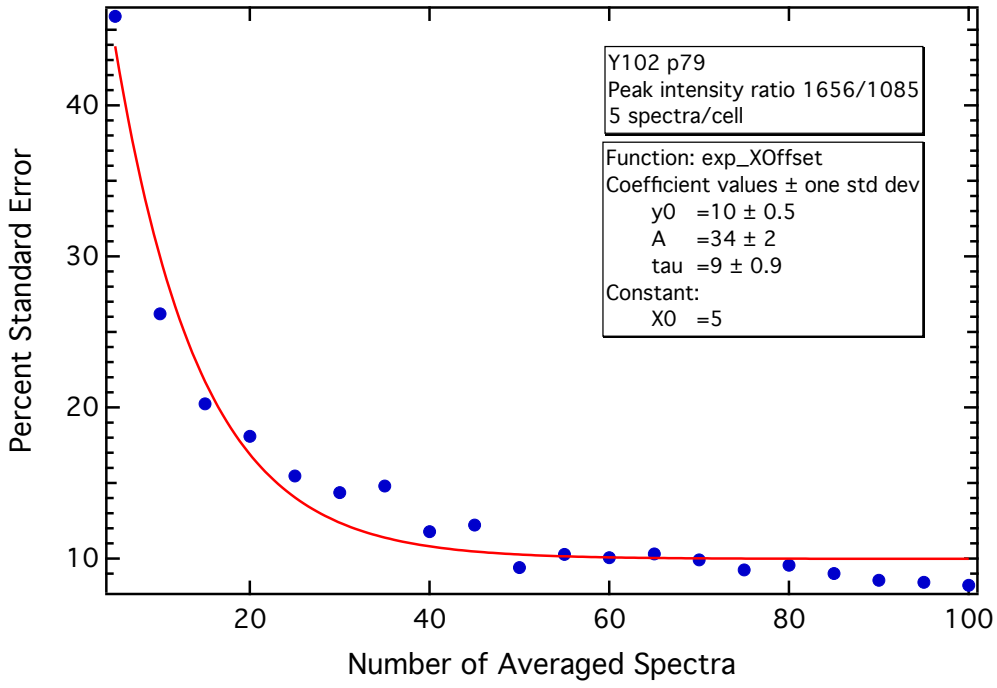


(b)

Figure B.7: %SE convergence tests of the Y102 MSC line for the air-dried analyses – experiment 1. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

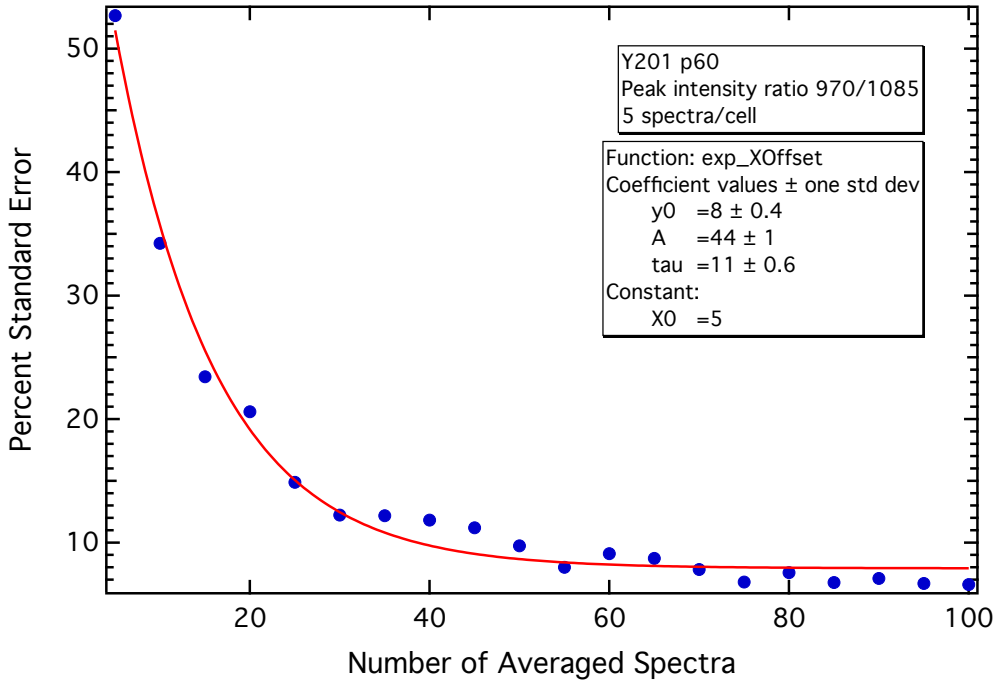


(a)

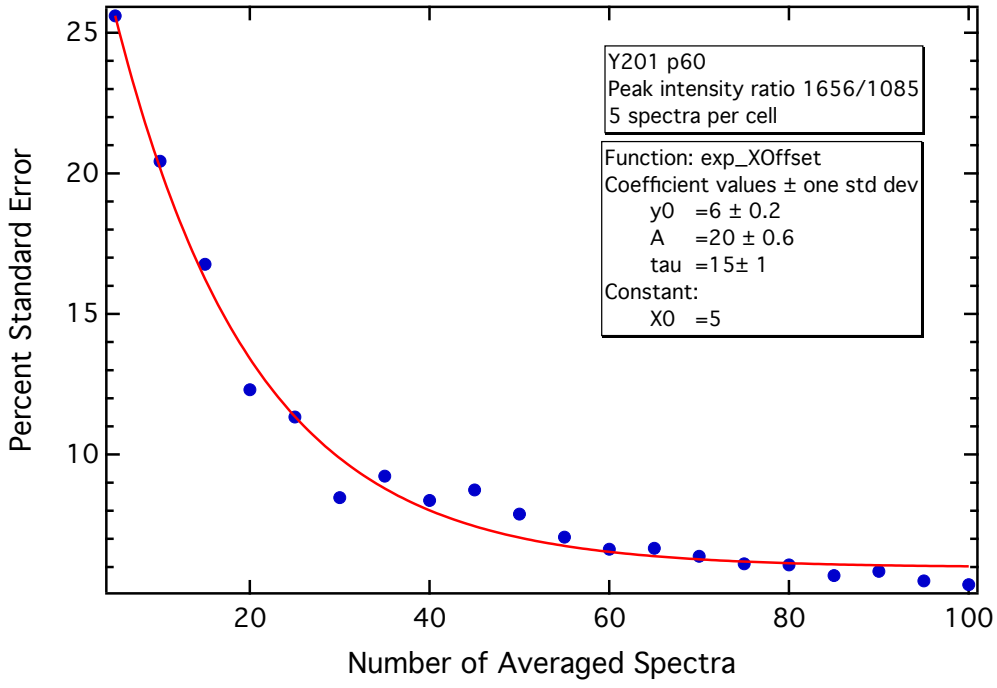


(b)

Figure B.8: %SE convergence tests of the Y102 MSC line for the air-dried analyses – experiment 2. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

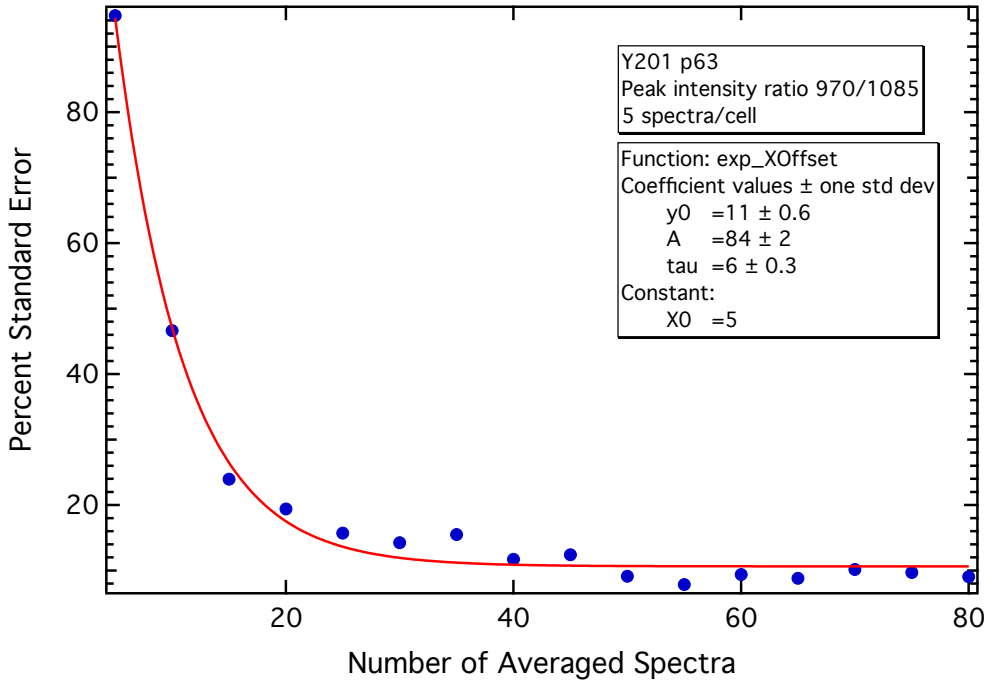


(a)

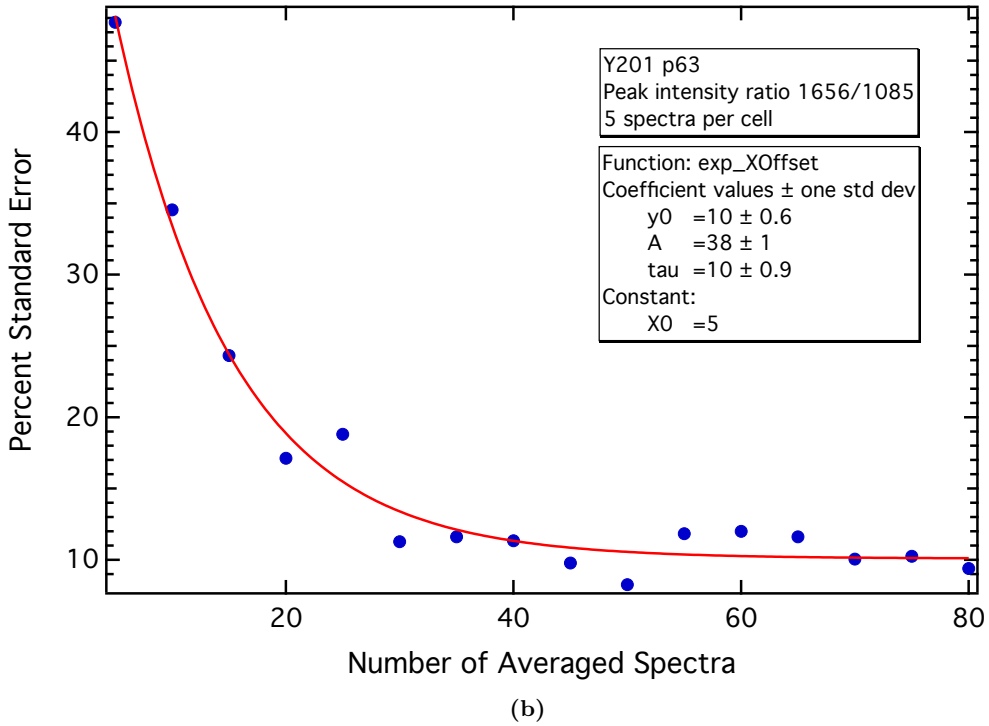


(b)

Figure B.9: %SE convergence tests of the Y201 MSC line for the air-dried analyses – experiment 1. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

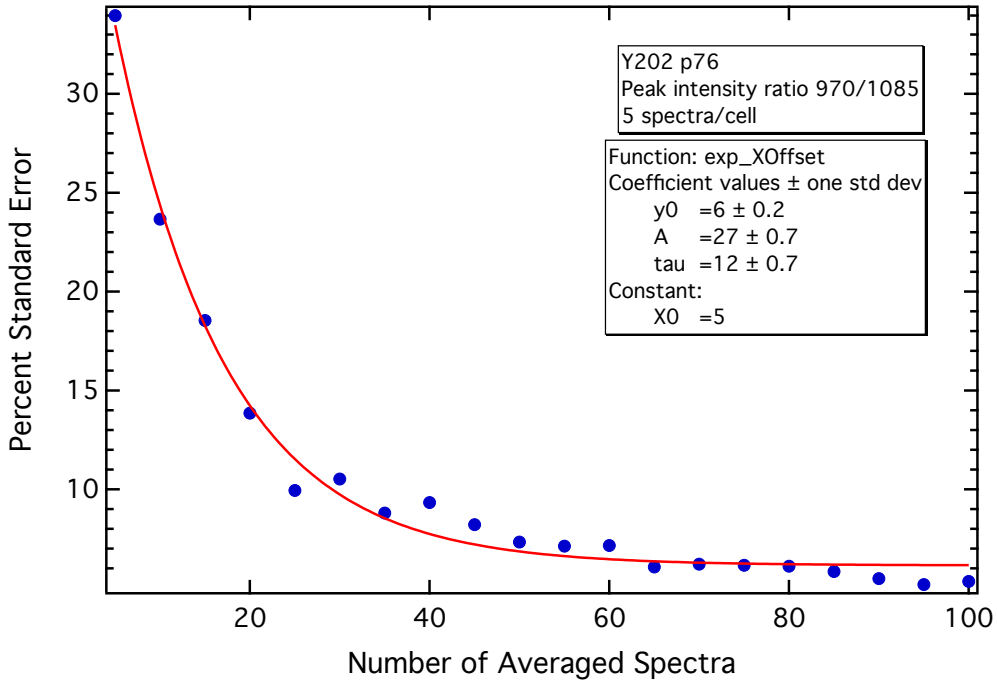


(a)

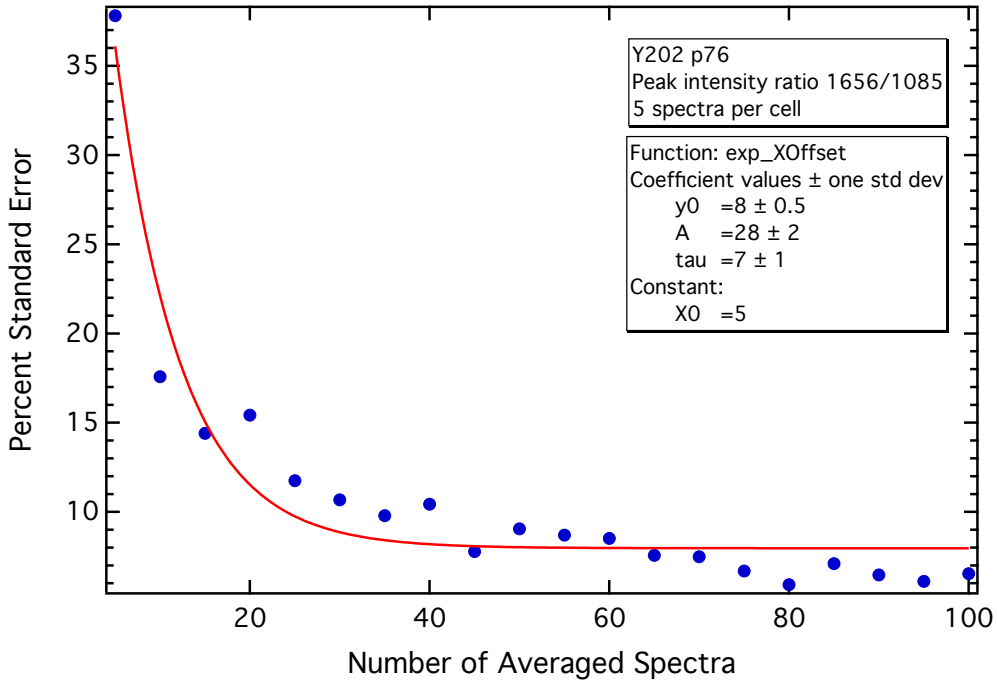


(b)

Figure B.10: %SE convergence tests of the Y201 MSC line for the air-dried analyses – experiment 2. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

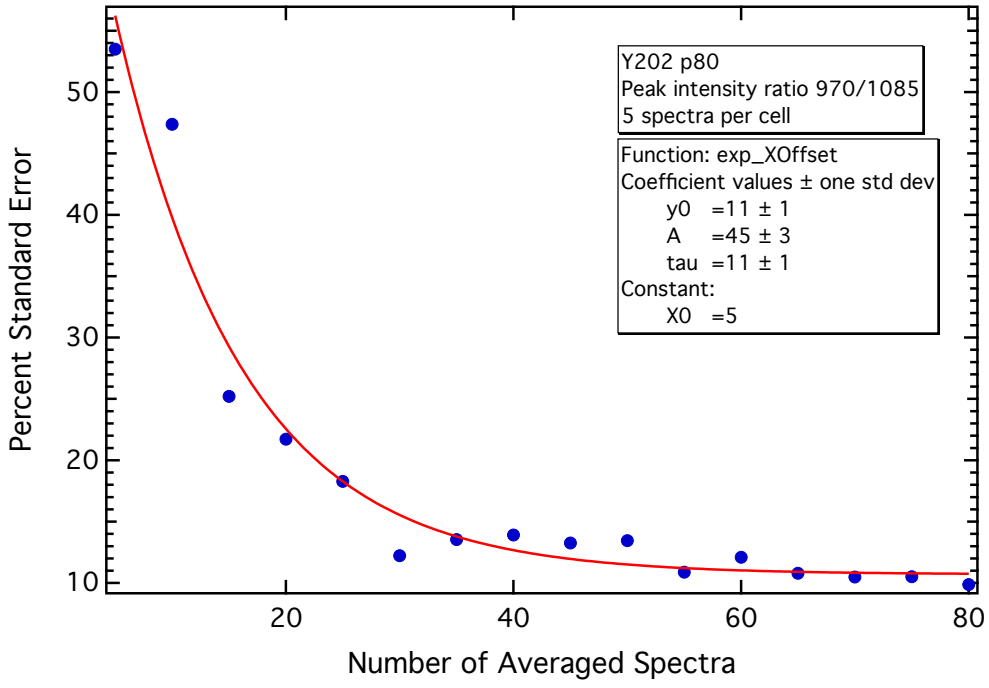


(a)

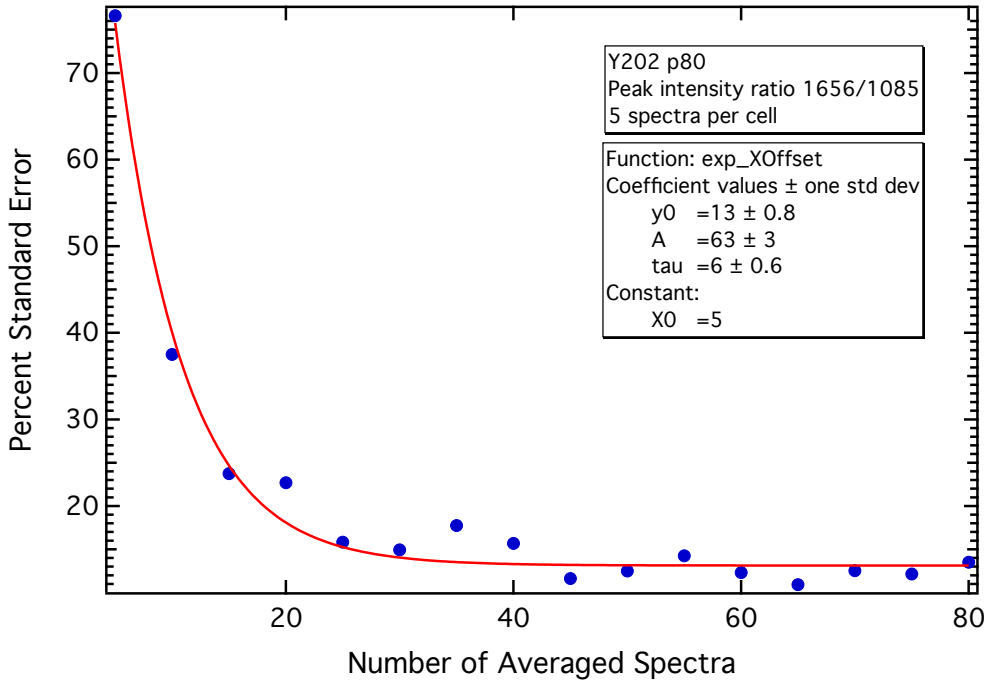


(b)

Figure B.11: %SE convergence tests of the Y202 MSC line for the air-dried analyses – experiment 1. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

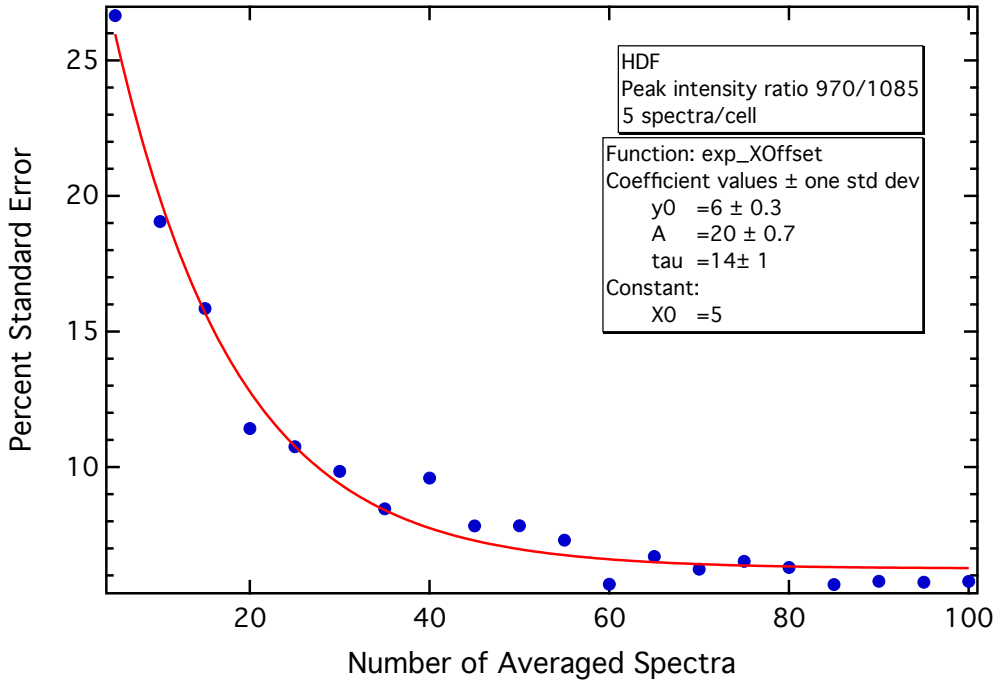


(a)

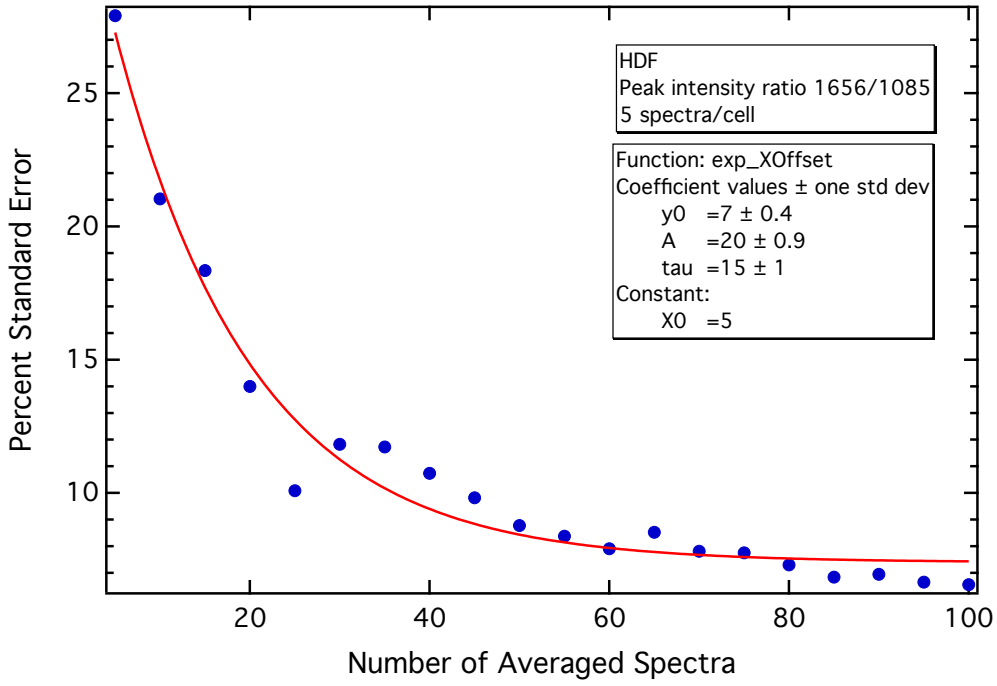


(b)

Figure B.12: %SE convergence tests of the Y202 MSC line for the air-dried analyses – experiment 2. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

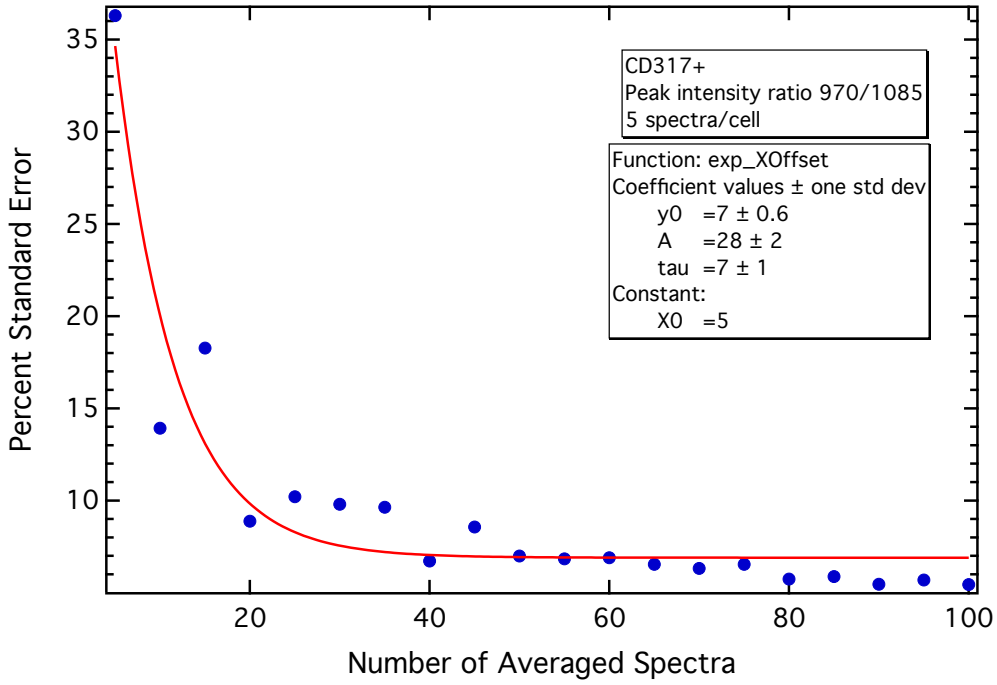


(a)

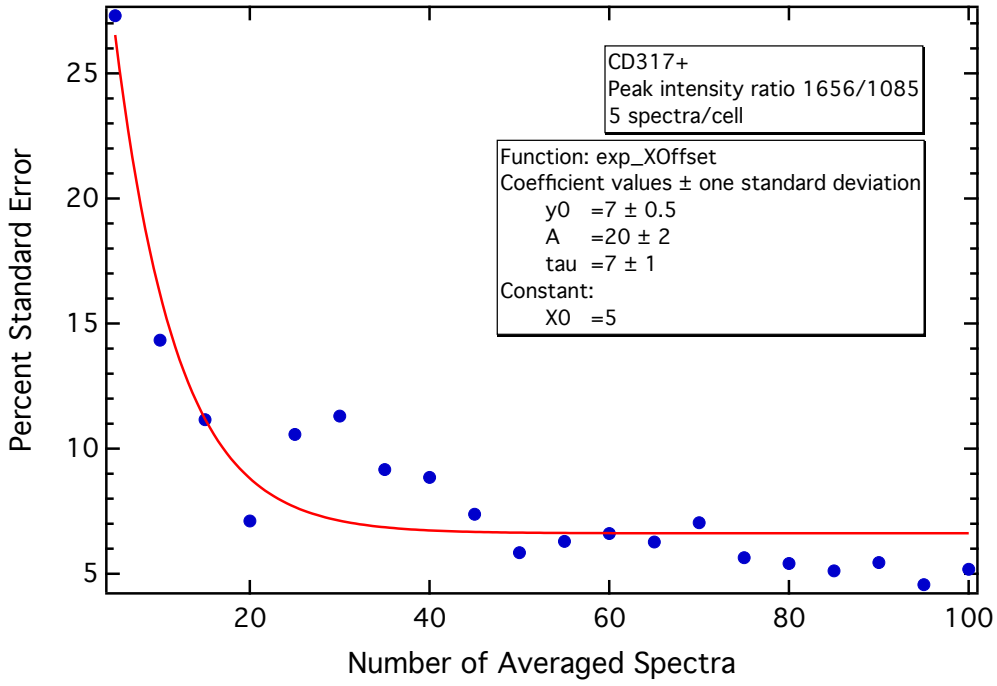


(b)

Figure B.13: %SE convergence tests of the HDF for air-dried analysis. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

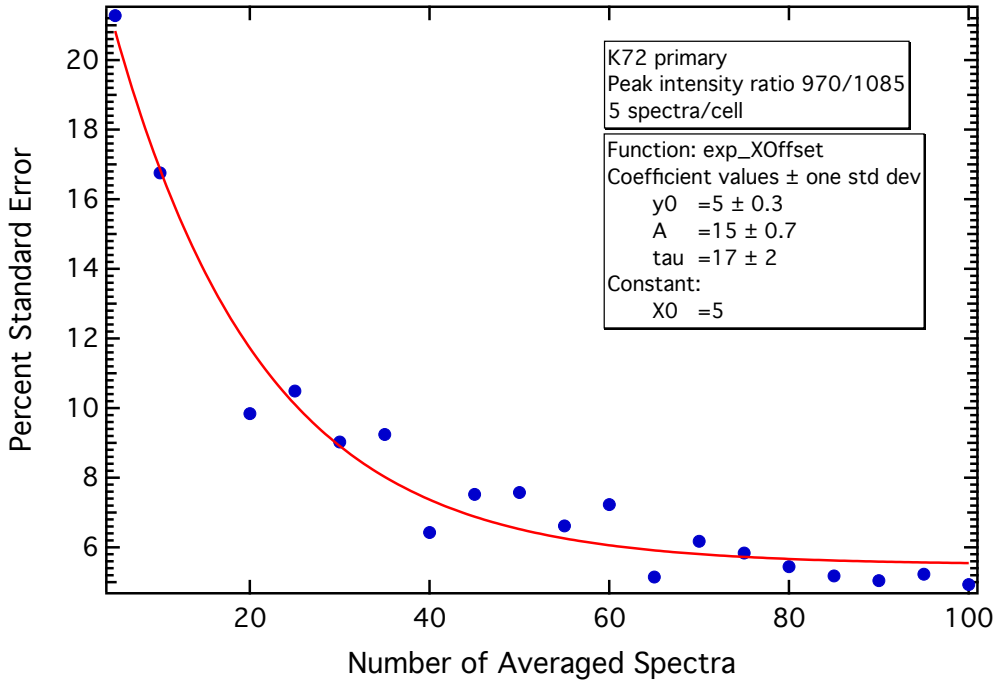


(a)

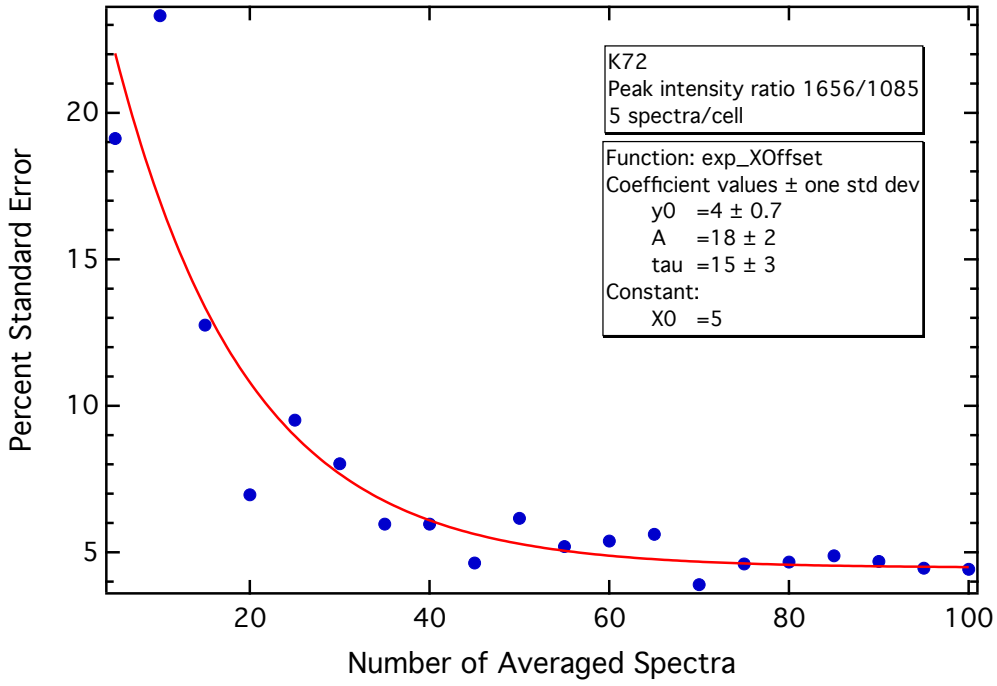


(b)

Figure B.14: %SE convergence tests of the CD317+ sorted fraction for air-dried analysis. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.



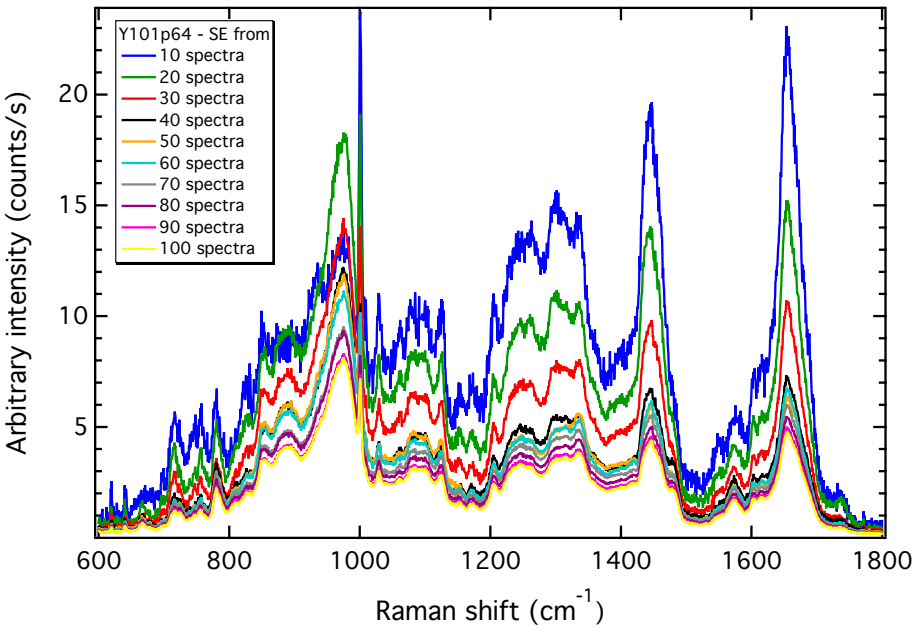
(a)



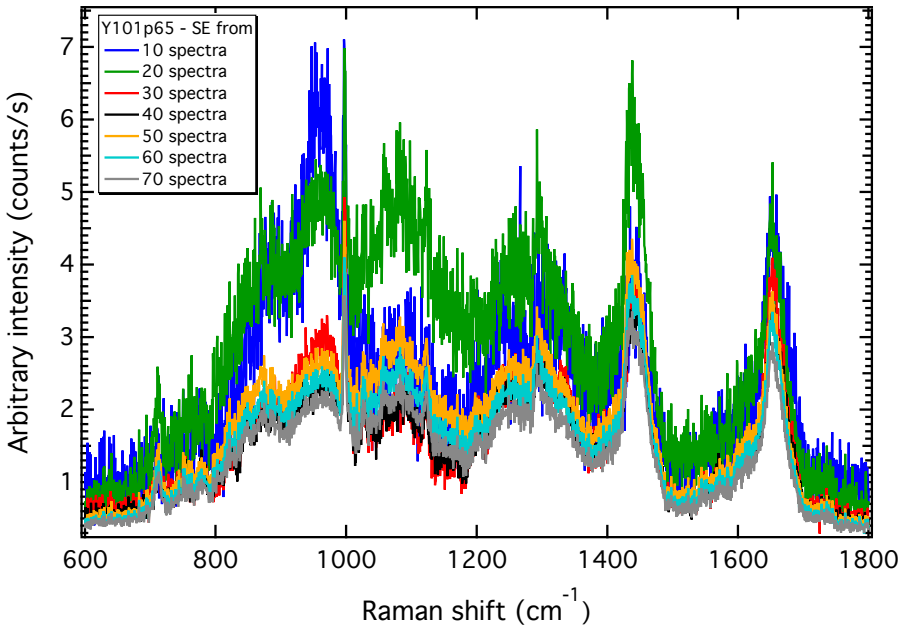
(b)

Figure B.15: %SE convergence tests of the K72 primary for air-dried analysis. Convergence for PIRs (a) 970/1085 and (b) 1656/1085.

B.2 Air-dried MSCs, HDF, CD317+ and K72 – SE convergence tests

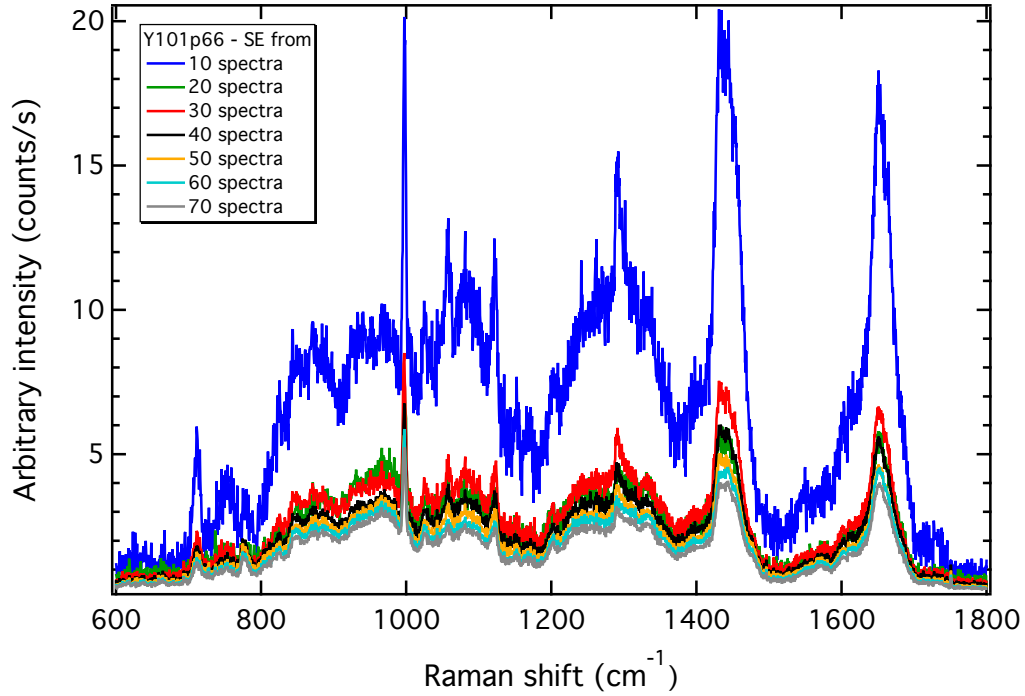


(a)

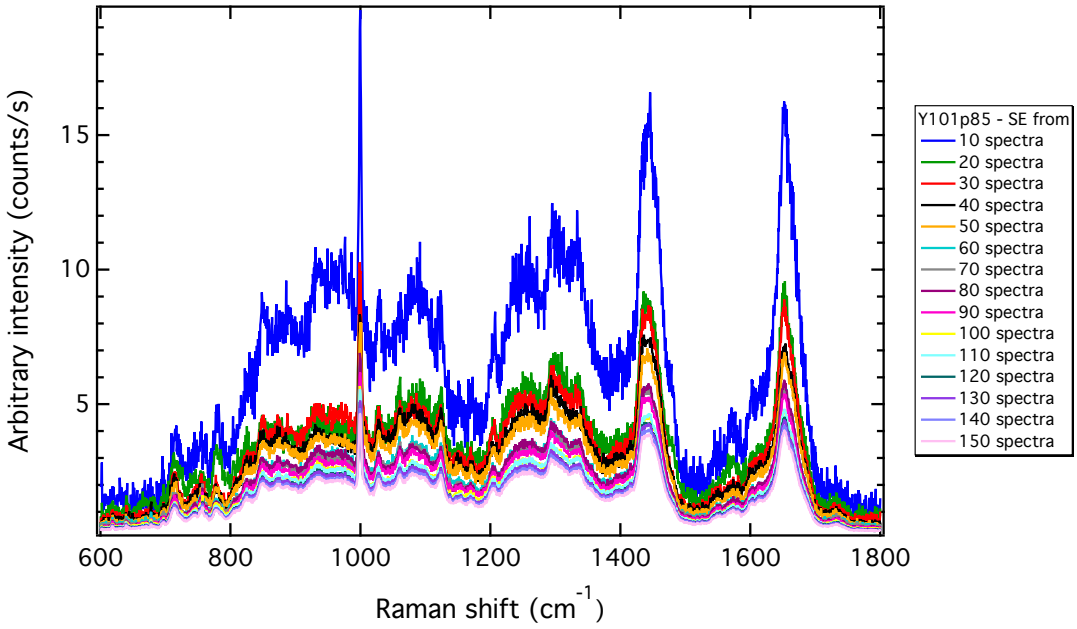


(b)

Figure B.16: SE stacks of MSC line Y101 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.

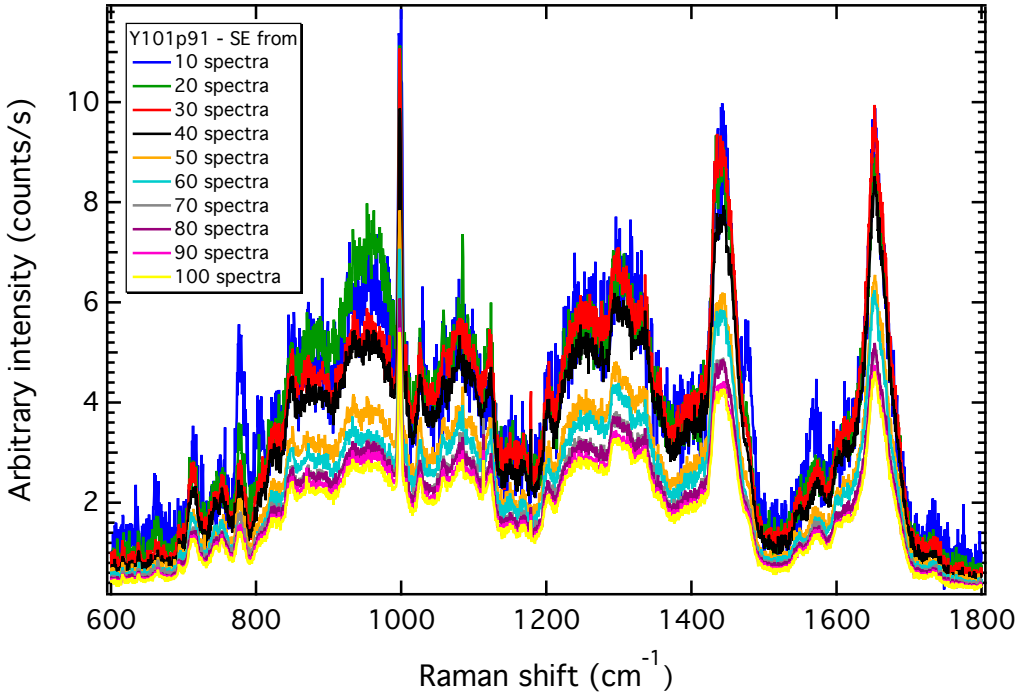


(a)

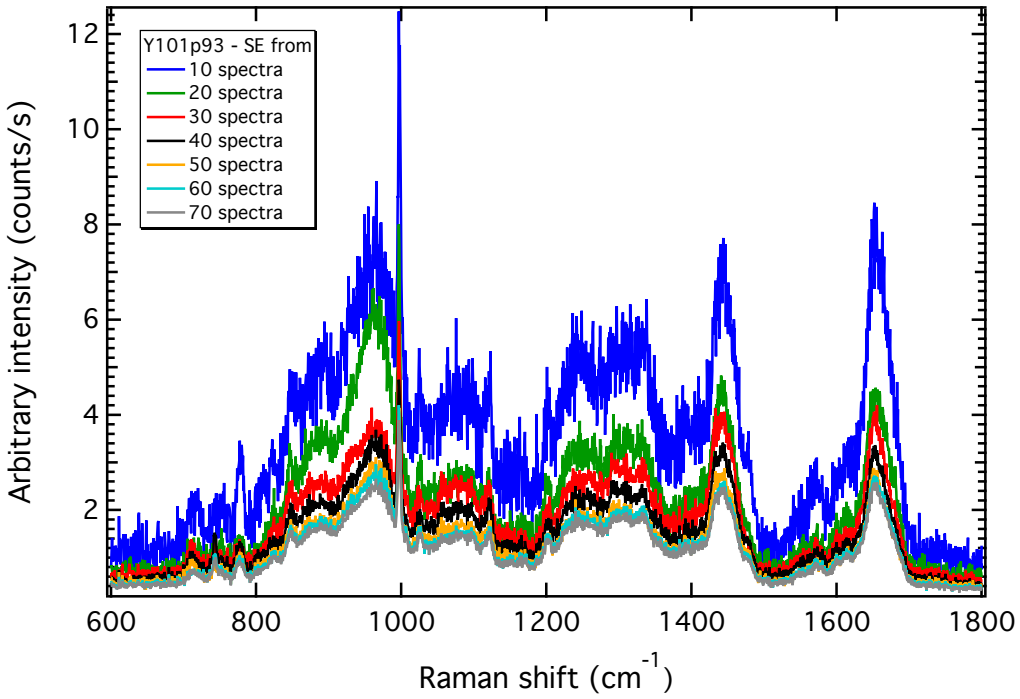


(b)

Figure B.17: SE stacks of MSC line Y101 for air-dried analyses. (a) Experiment 3 and (b) experiment 4.

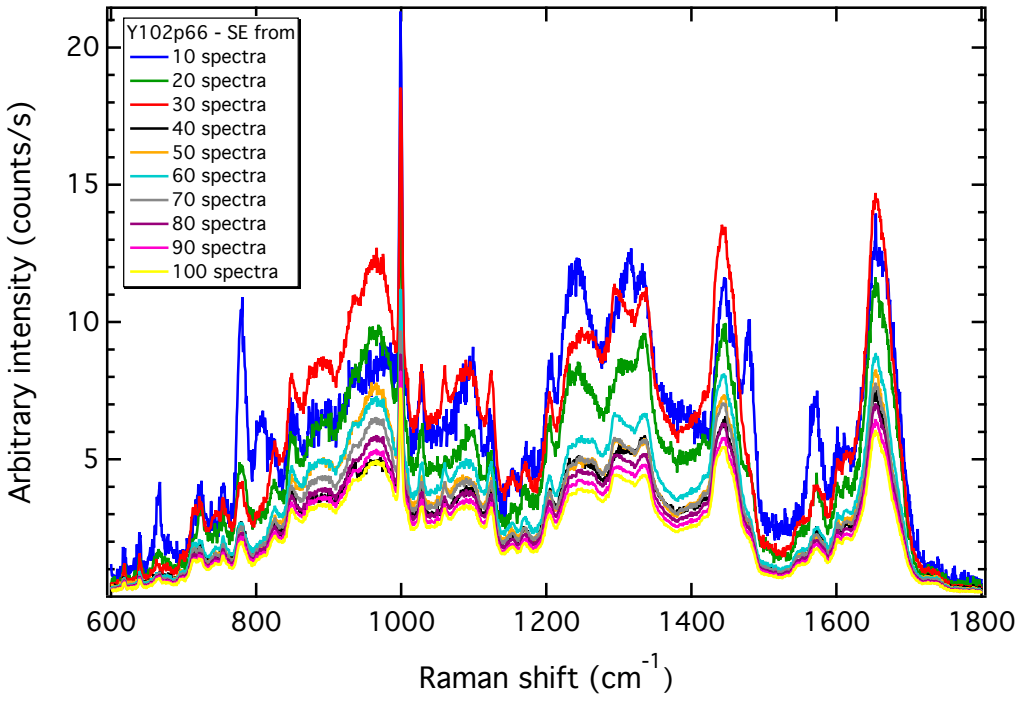


(a)

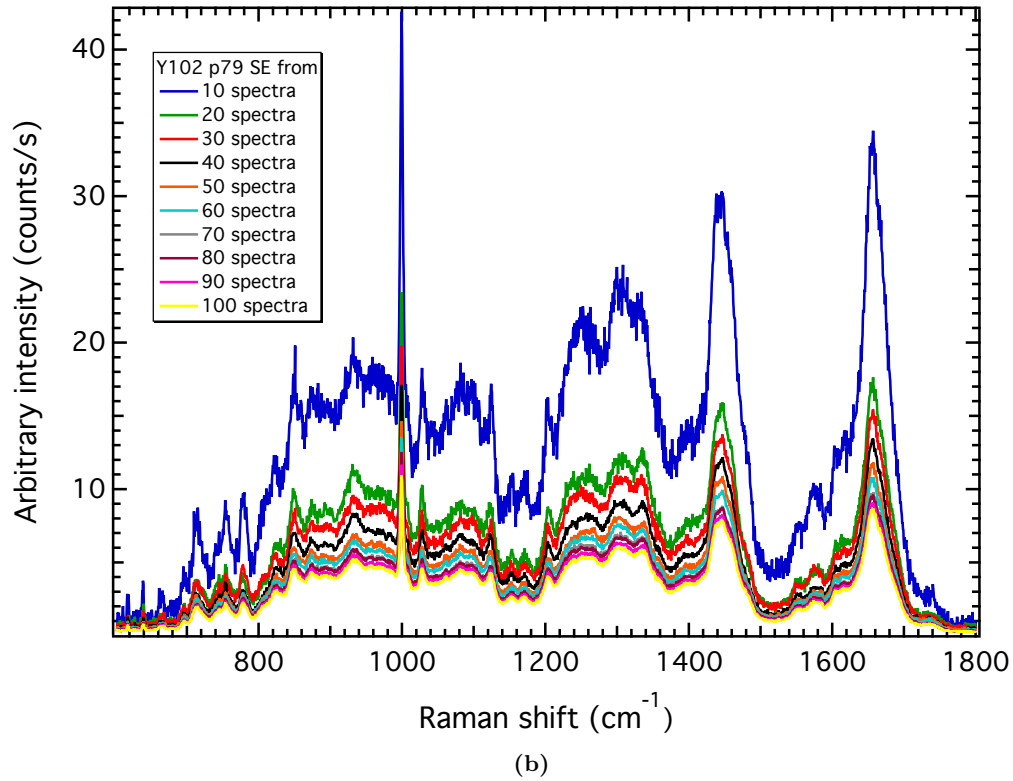


(b)

Figure B.18: SE stacks of MSC line Y101 for air-dried analyses. (a) Experiment 5 and (b) experiment 6.

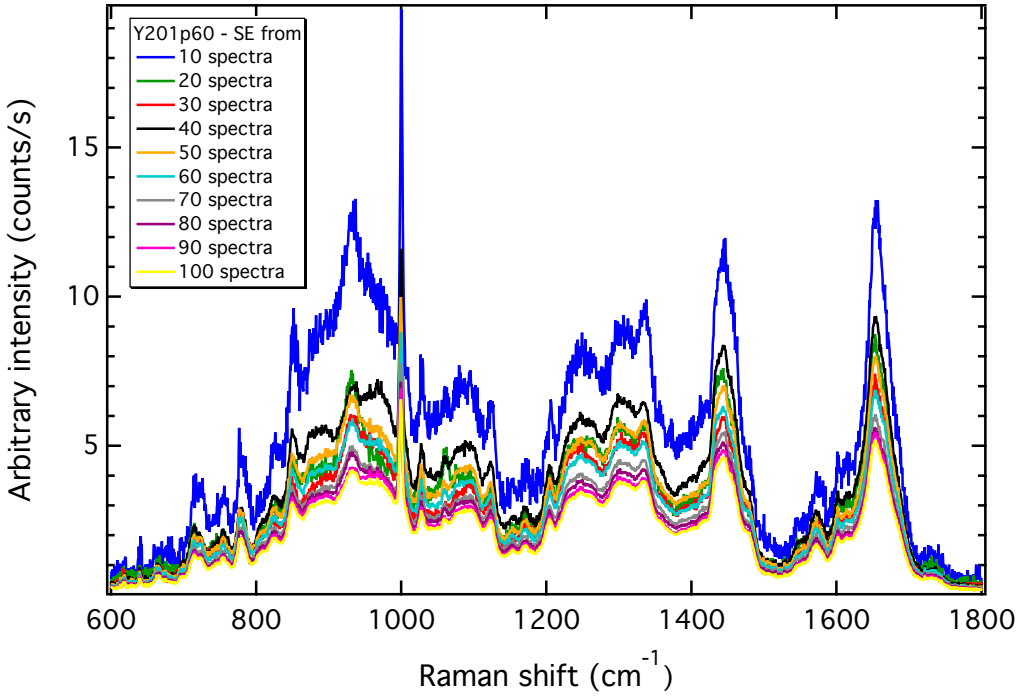


(a)

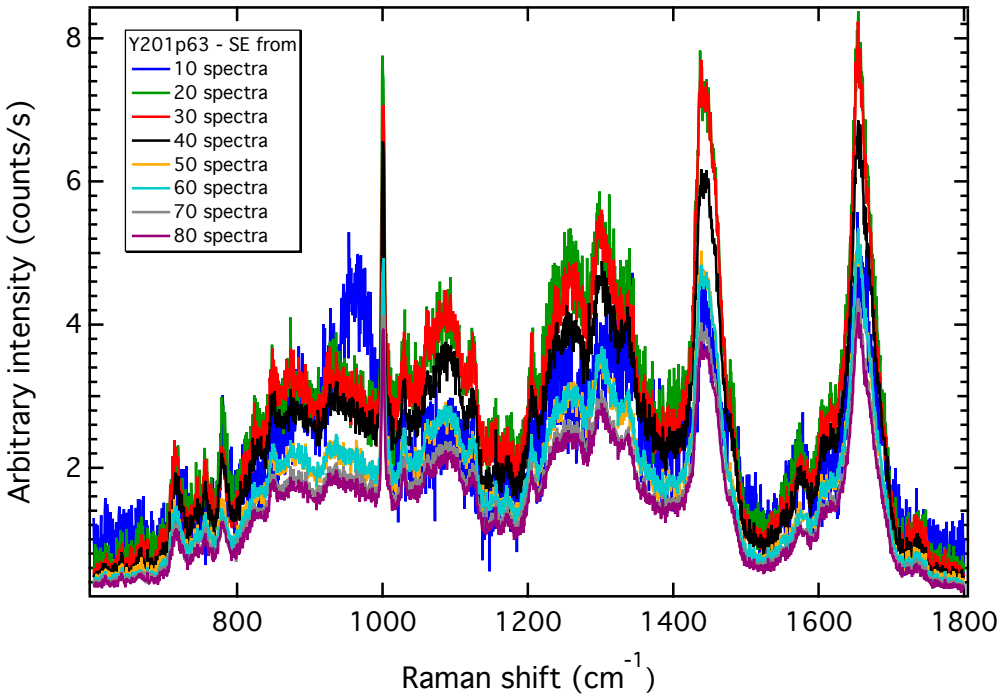


(b)

Figure B.19: SE stacks of MSC line Y102 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.

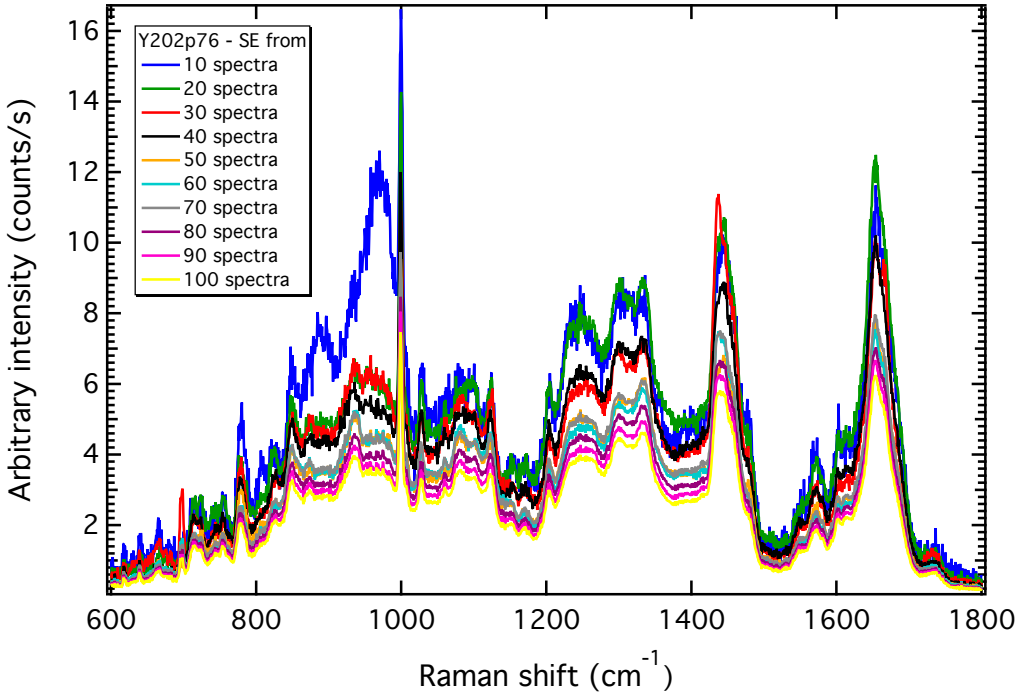


(a)

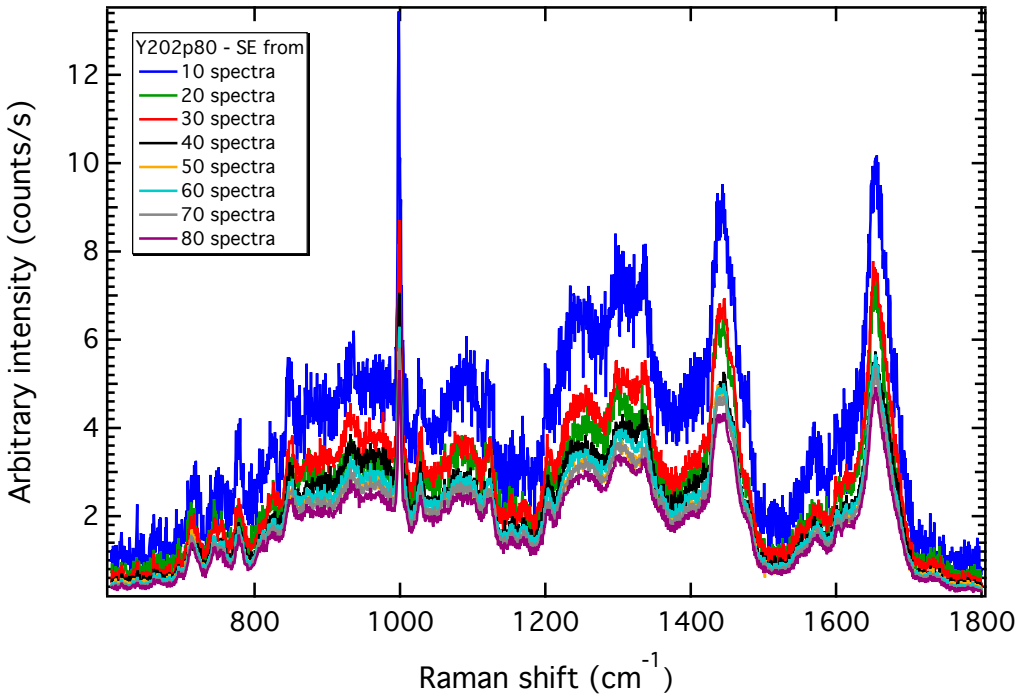


(b)

Figure B.20: SE stacks of MSC line Y201 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.

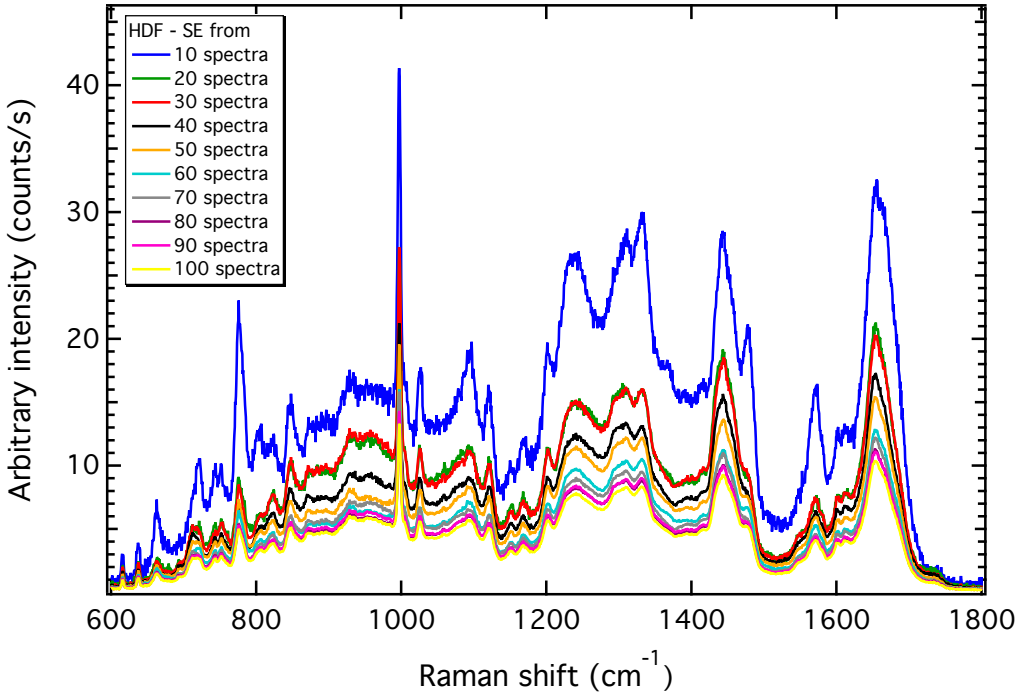


(a)

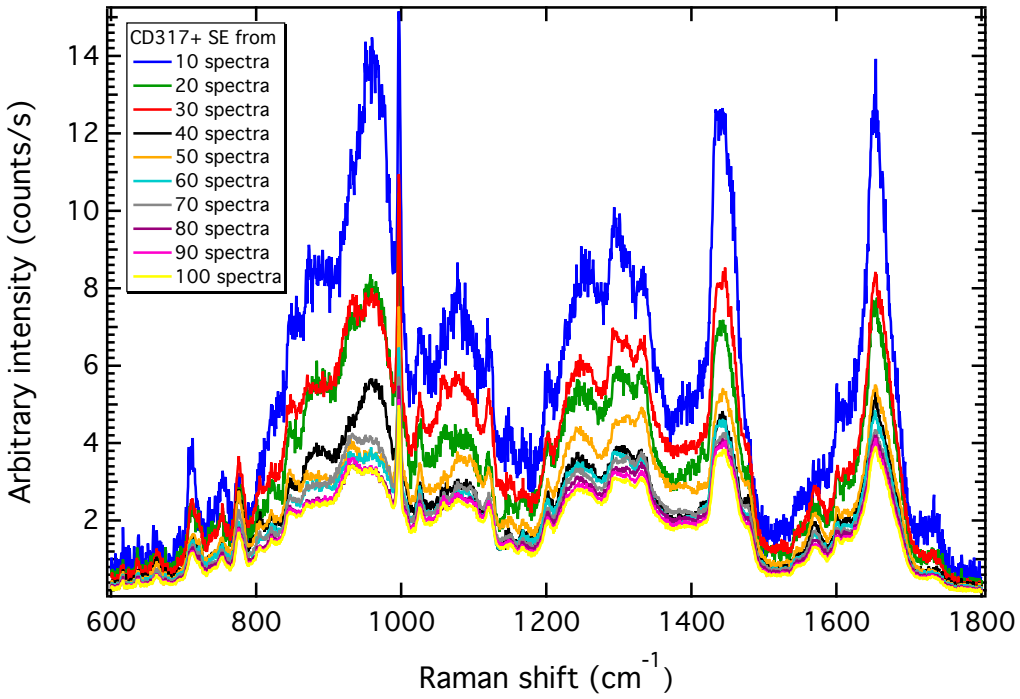


(b)

Figure B.21: SE stacks of MSC line Y202 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.



(a)



(b)

Figure B.22: SE stacks of of (a) HDF and (b) CD317+ for air-dried analyses.

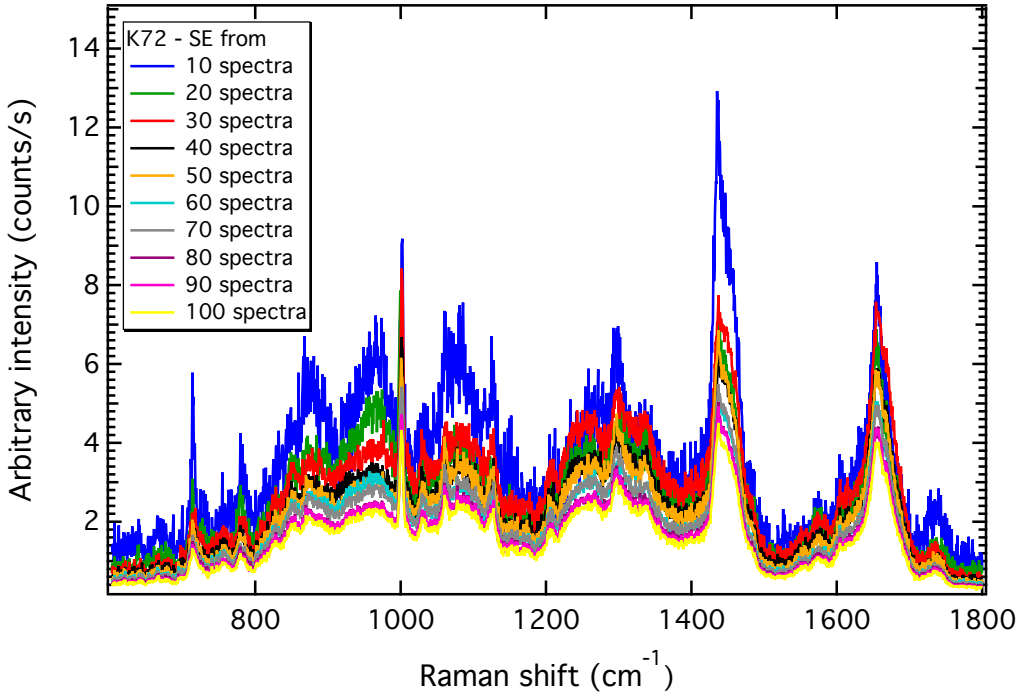
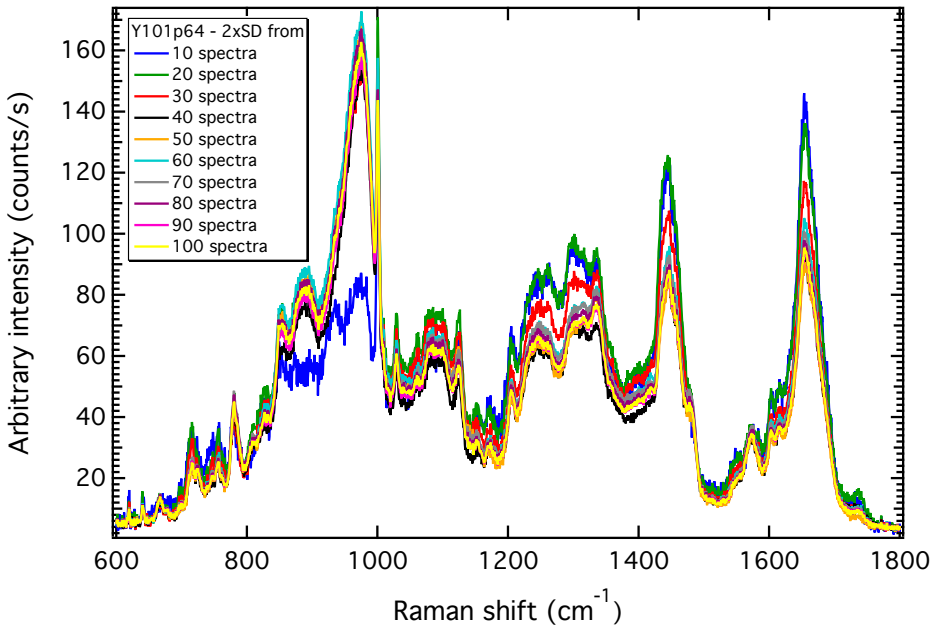
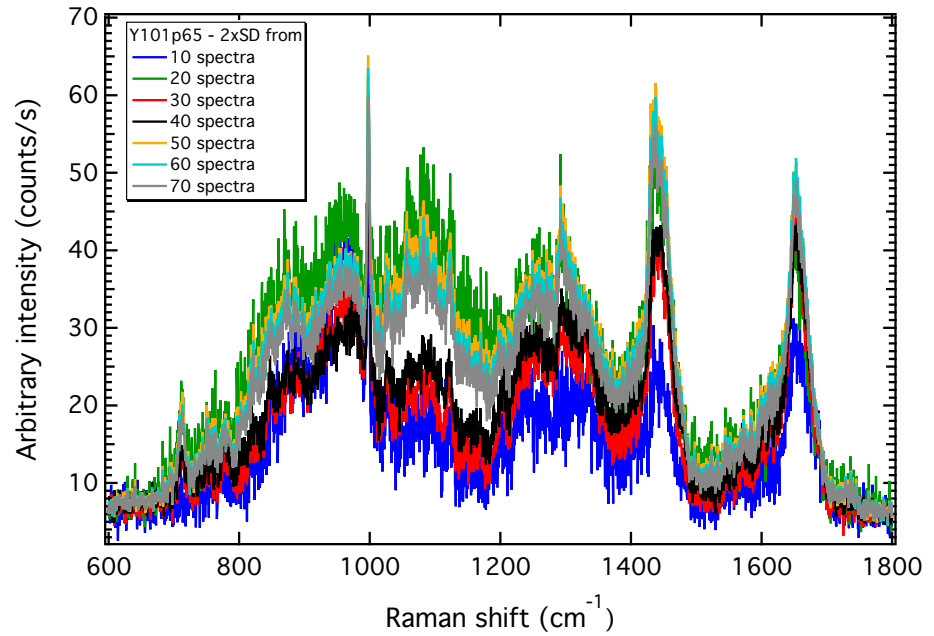


Figure B.23: SE stacks of of K72 primary for air-dried analyses.

B.3 Air-dried MSCs, HDF, CD317+ and K72 – 2x SD convergence tests

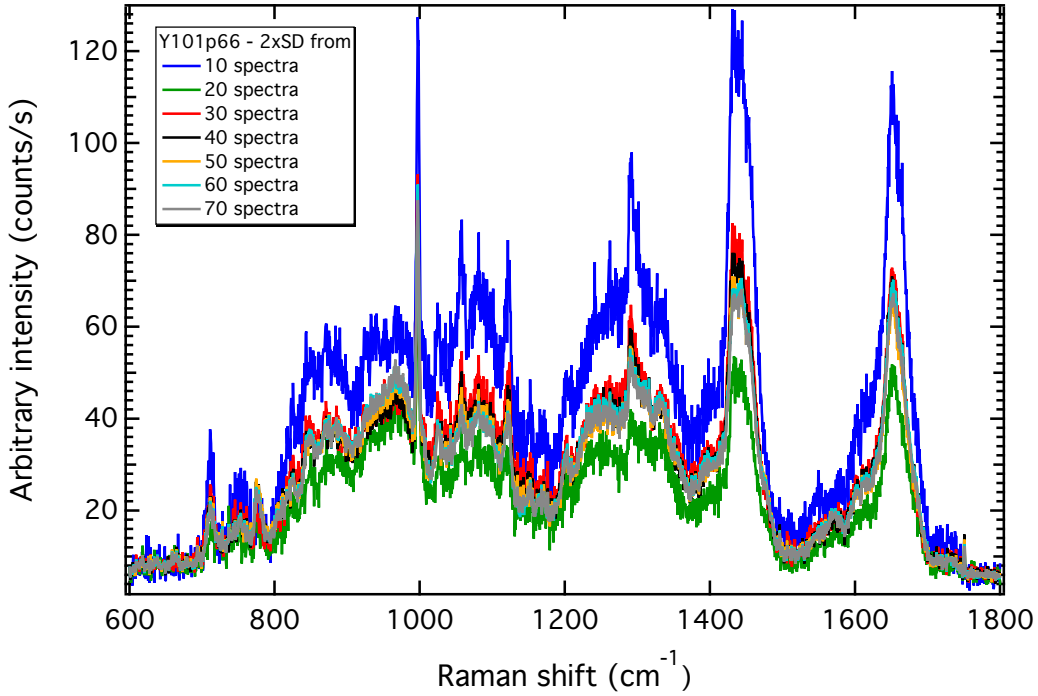


(a)

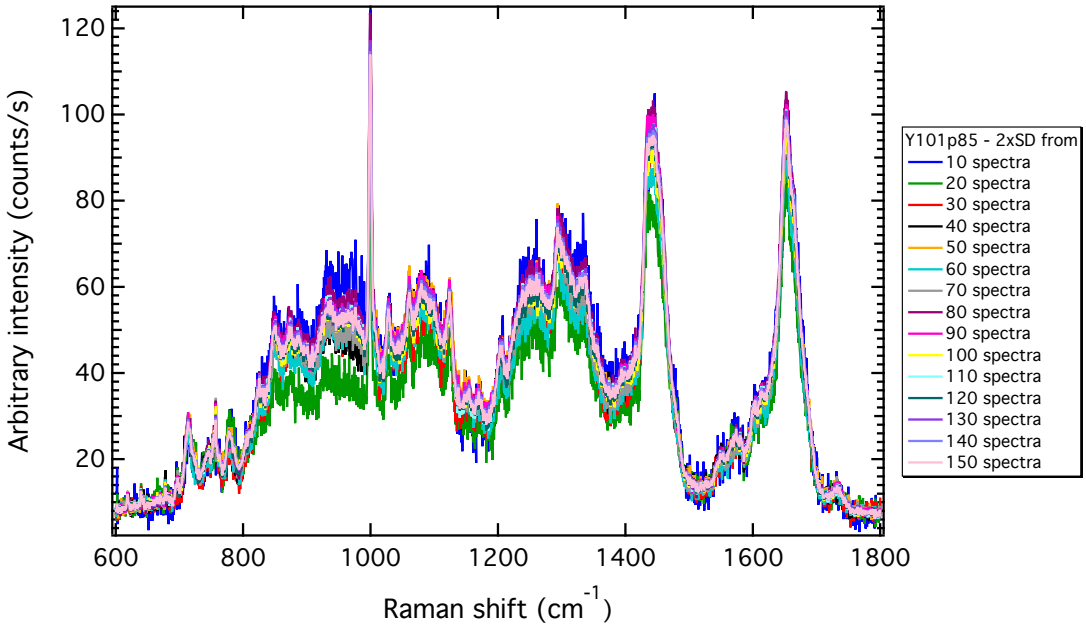


(b)

Figure B.24: 2x SD stacks of MSC line Y101 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.

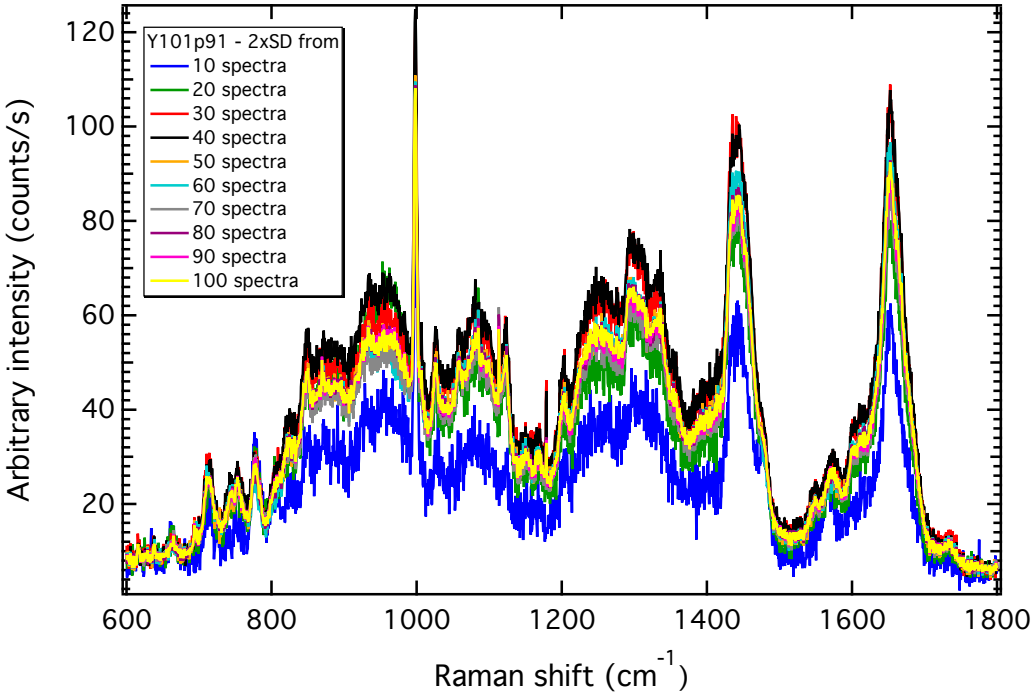


(a)

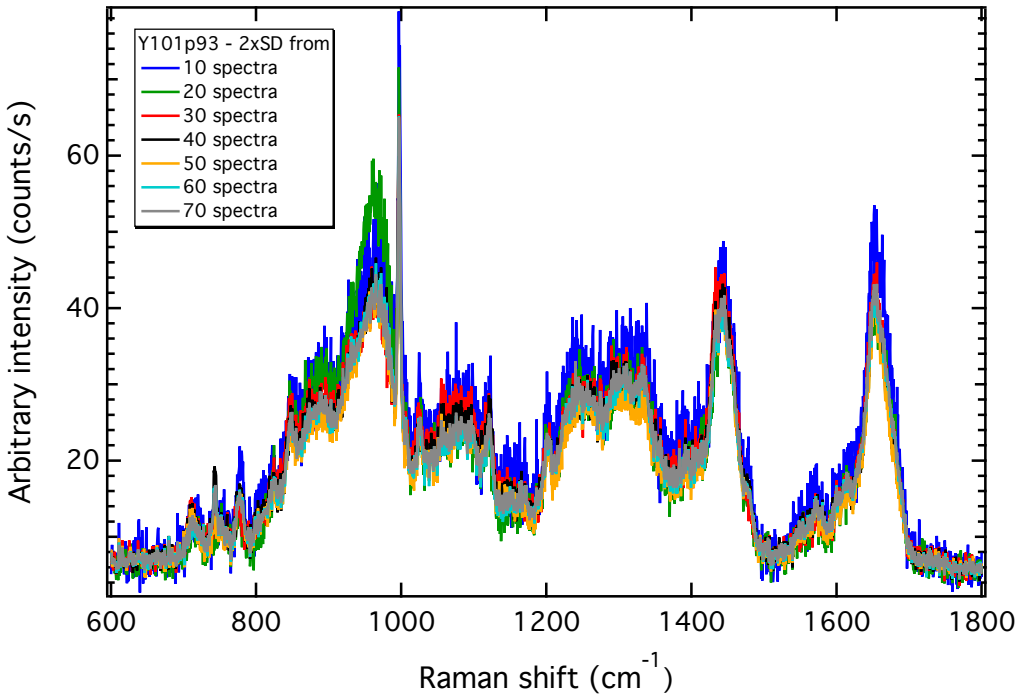


(b)

Figure B.25: 2x SD stacks of MSC line Y101 for air-dried analyses. (a) Experiment 3 and (b) experiment 4.

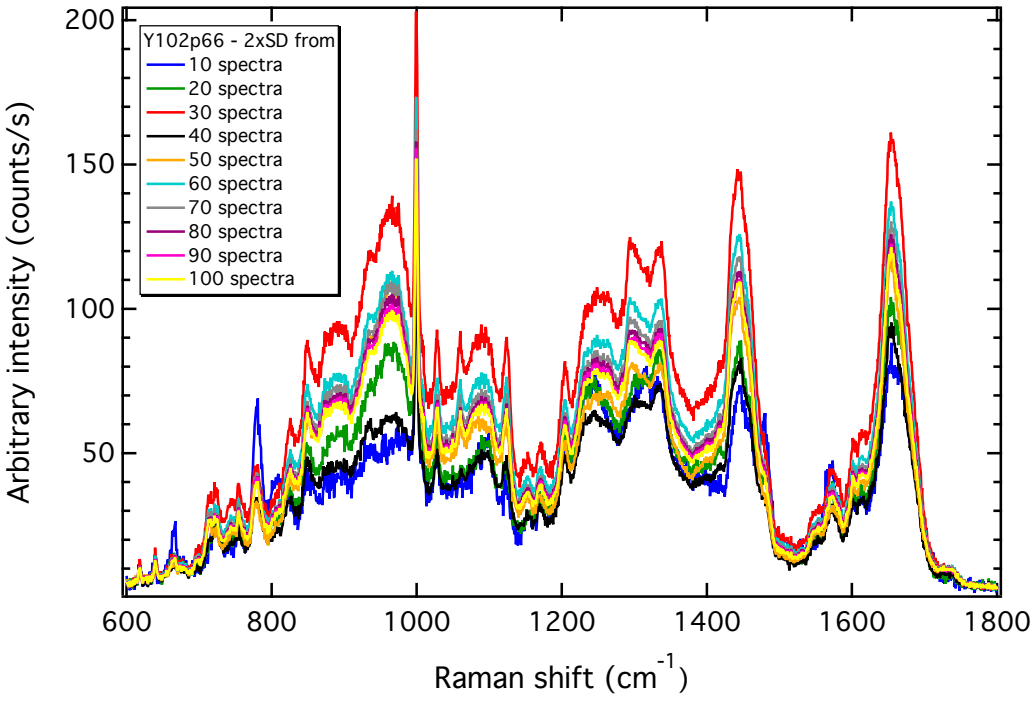


(a)

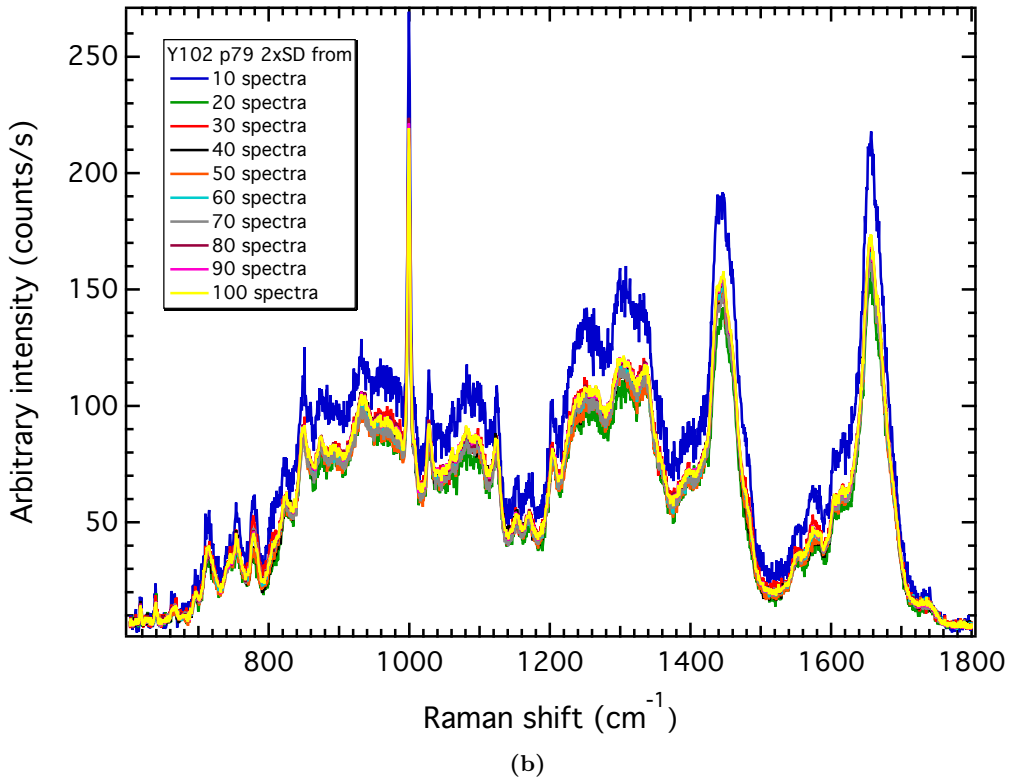


(b)

Figure B.26: 2x SD stacks of MSC line Y101 for air-dried analyses. (a) Experiment 5 and (b) experiment 6.

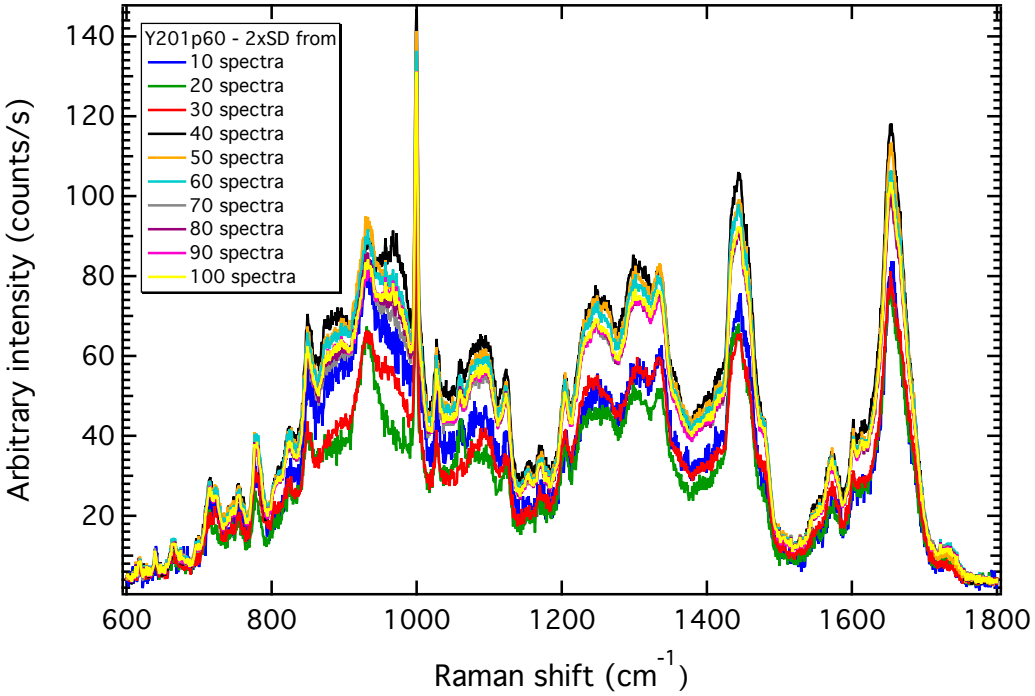


(a)

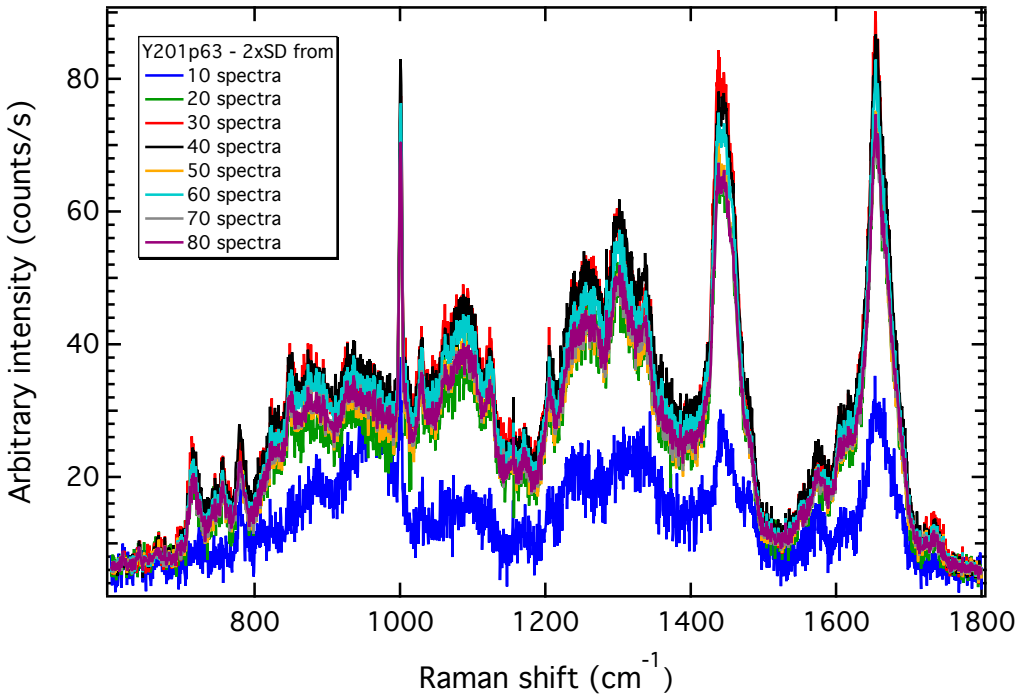


(b)

Figure B.27: 2x SD stacks of MSC line Y102 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.

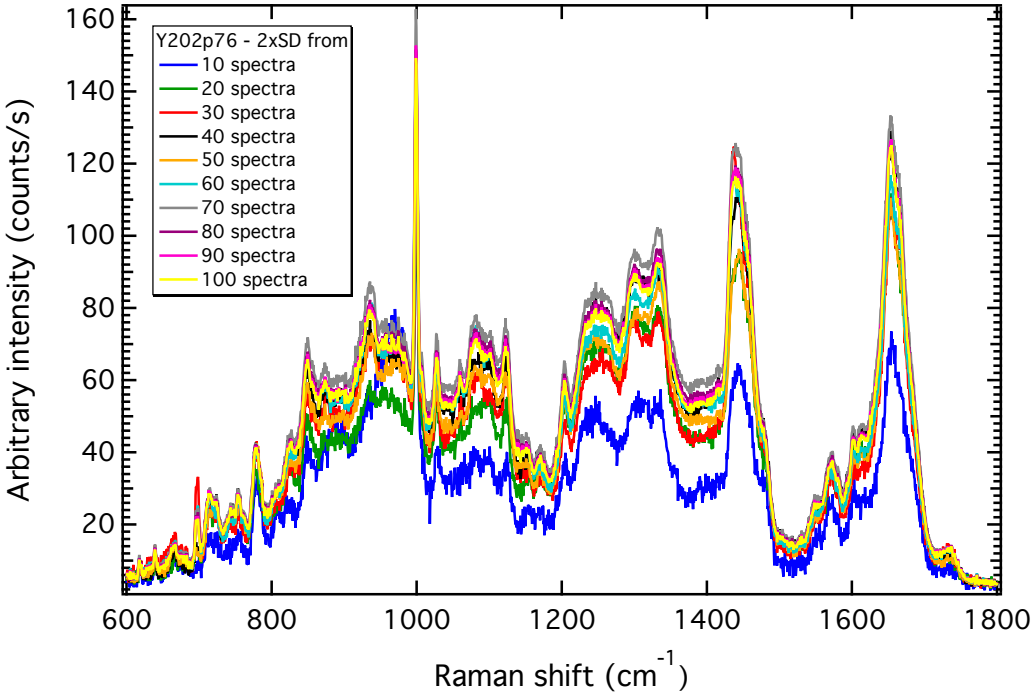


(a)

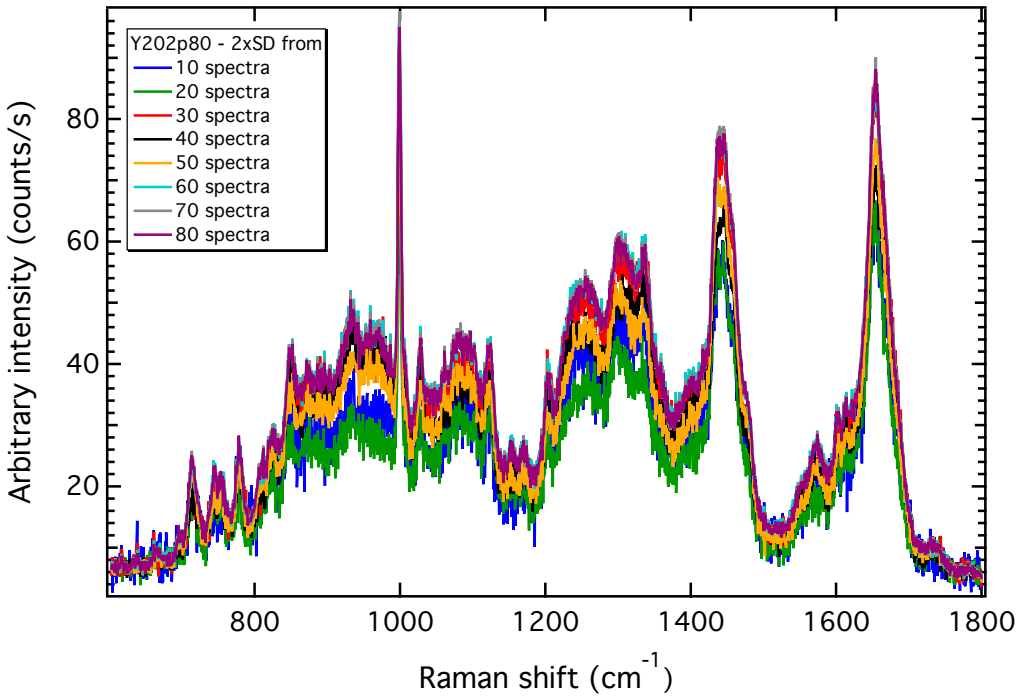


(b)

Figure B.28: 2x SD stacks of MSC line Y201 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.

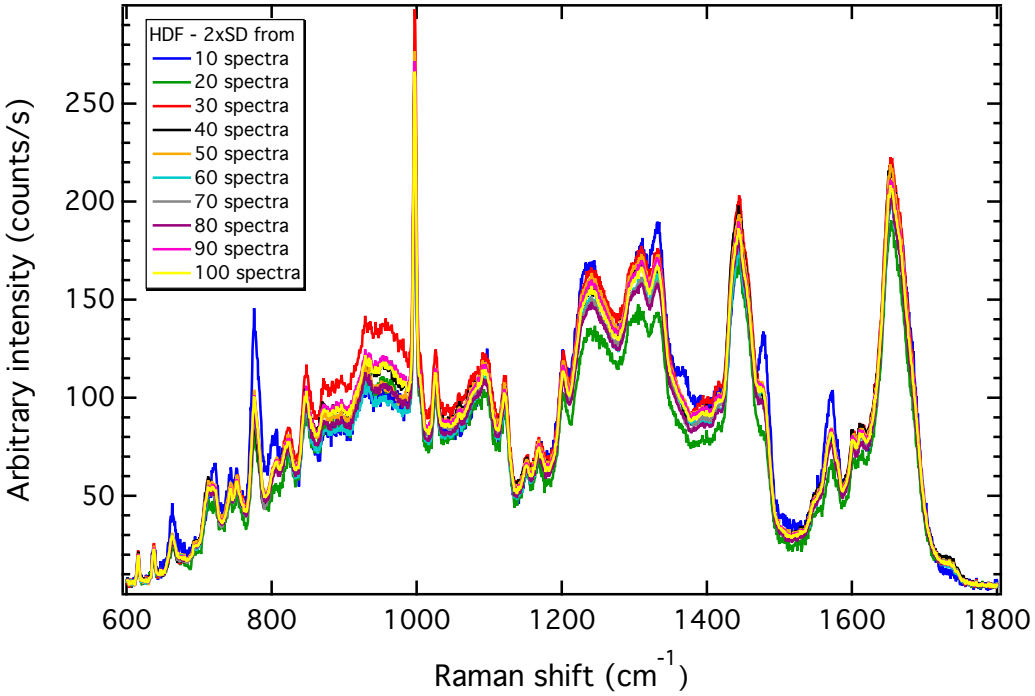


(a)

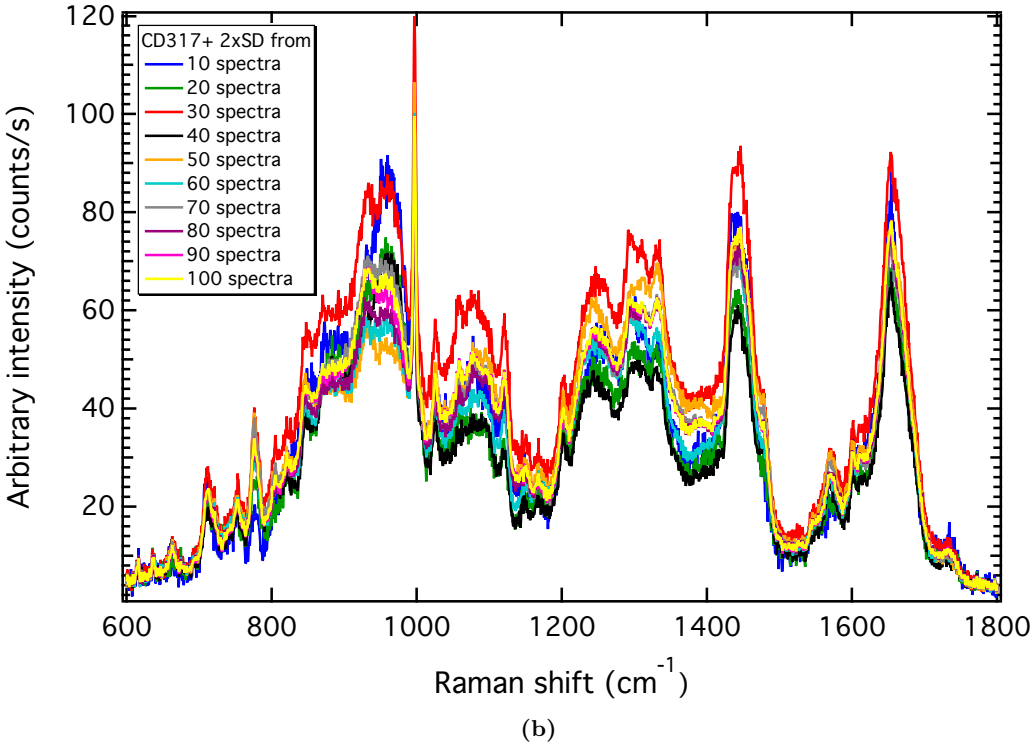


(b)

Figure B.29: 2x SD stacks of MSC line Y202 for air-dried analyses. (a) Experiment 1 and (b) experiment 2.



(a)



(b)

Figure B.30: 2x SD stacks of (a) HDF and (b) CD317+ sorted fraction for air-dried analysis.

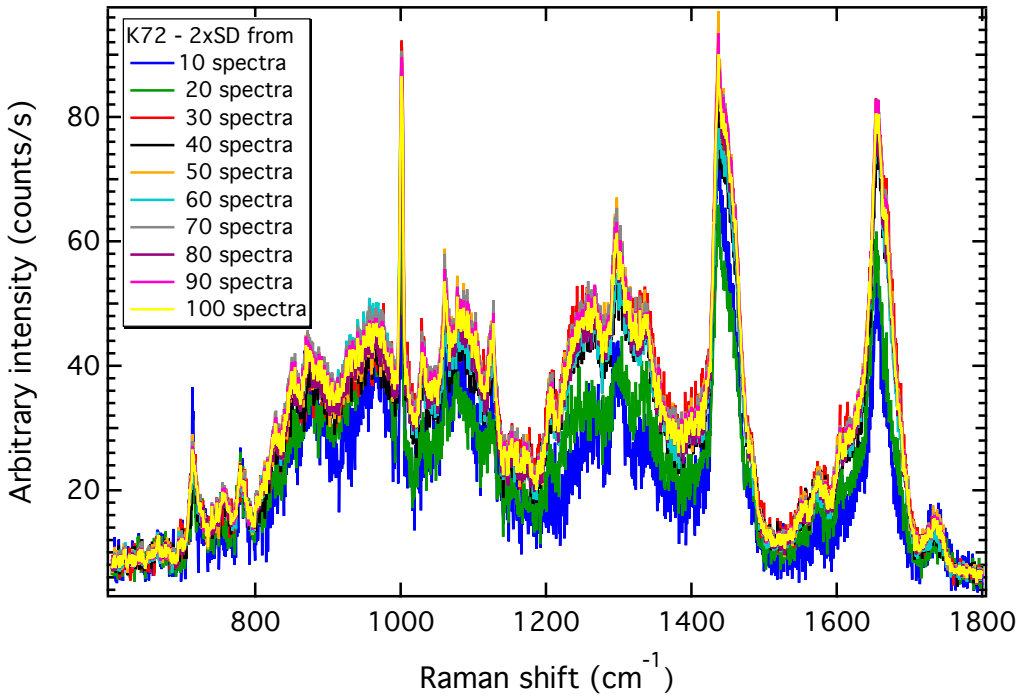
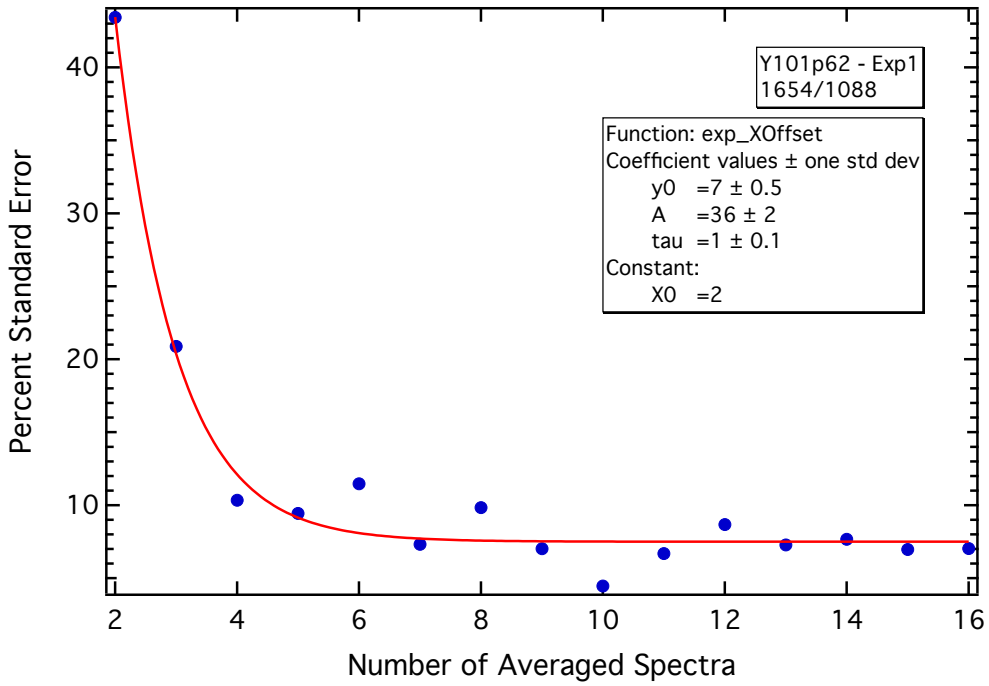
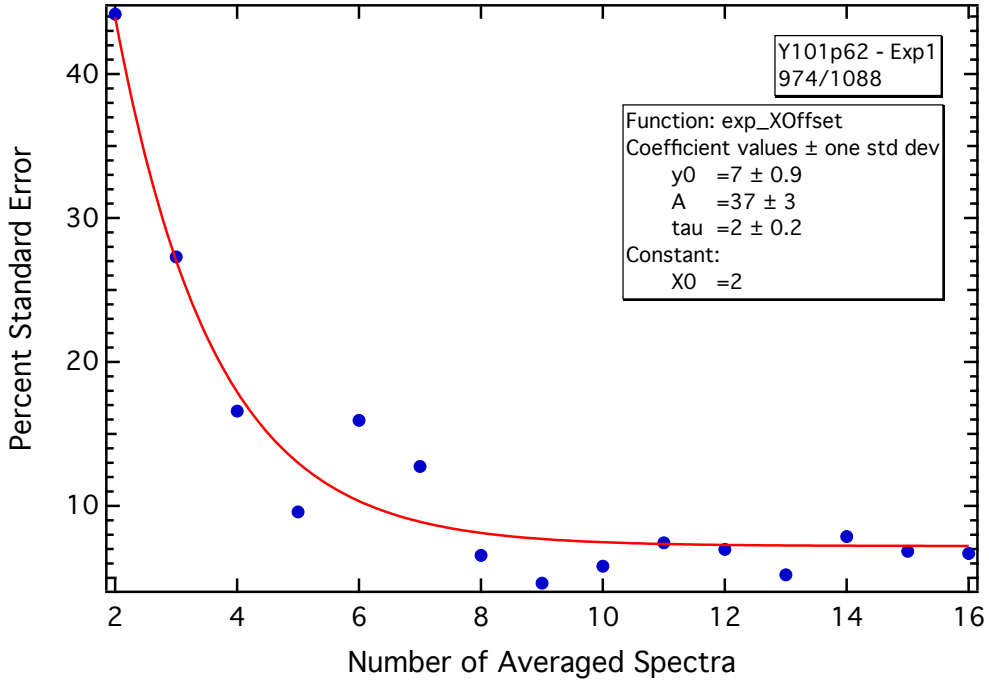


Figure B.31: 2x SD stacks of K72 primary for air-dried analyses.

B.4 Live MSCs – %SE convergence tests

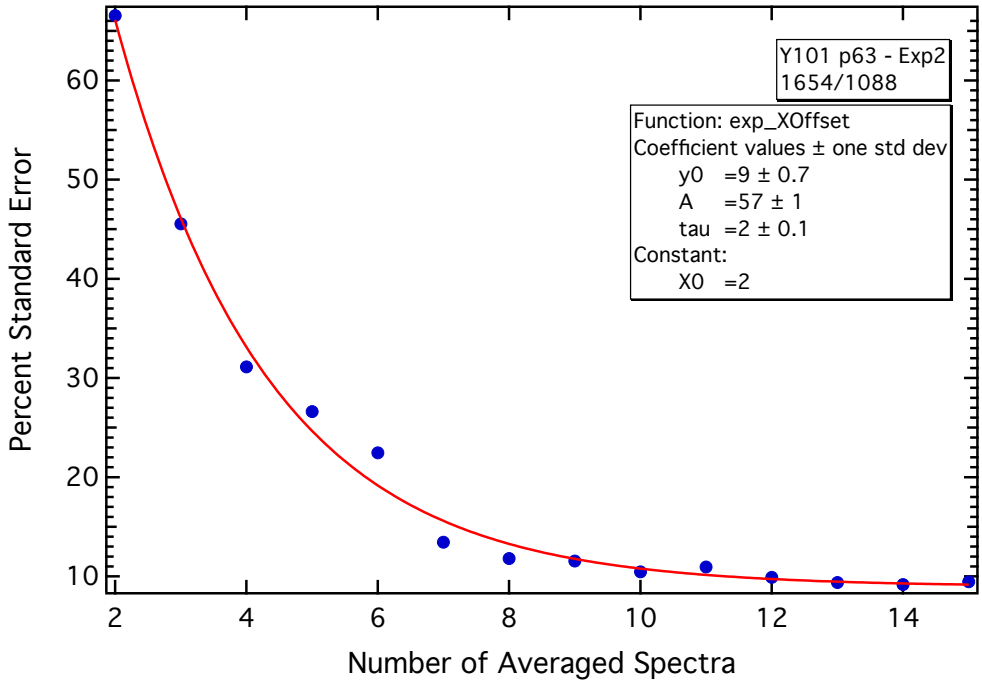


(a)

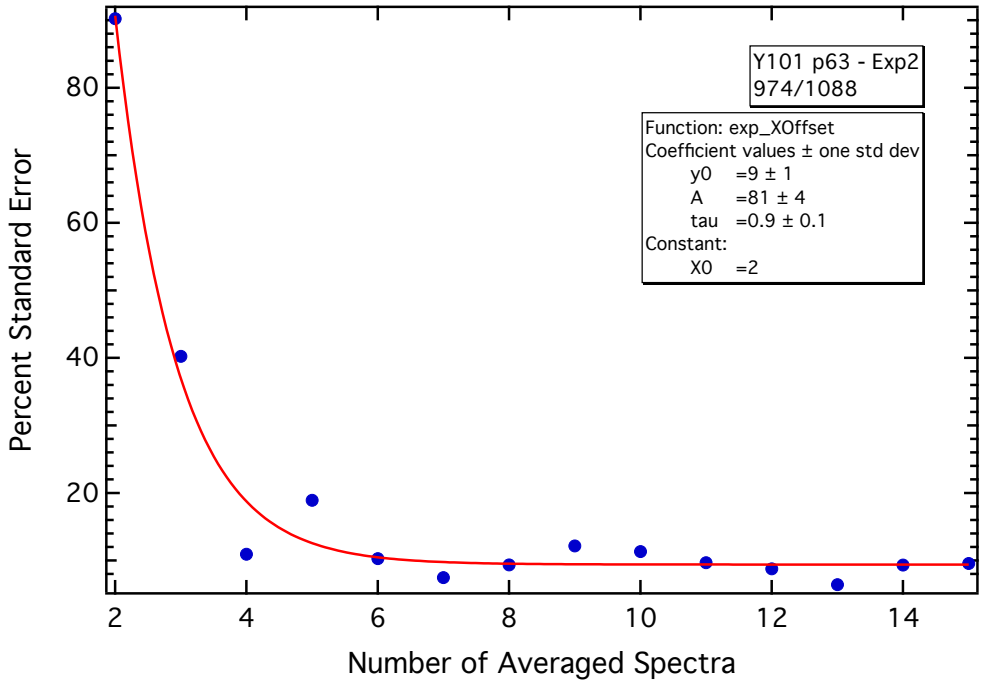


(b)

Figure B.32: %SE convergence tests for live cell analyses of MSC line Y101 experiment 1. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

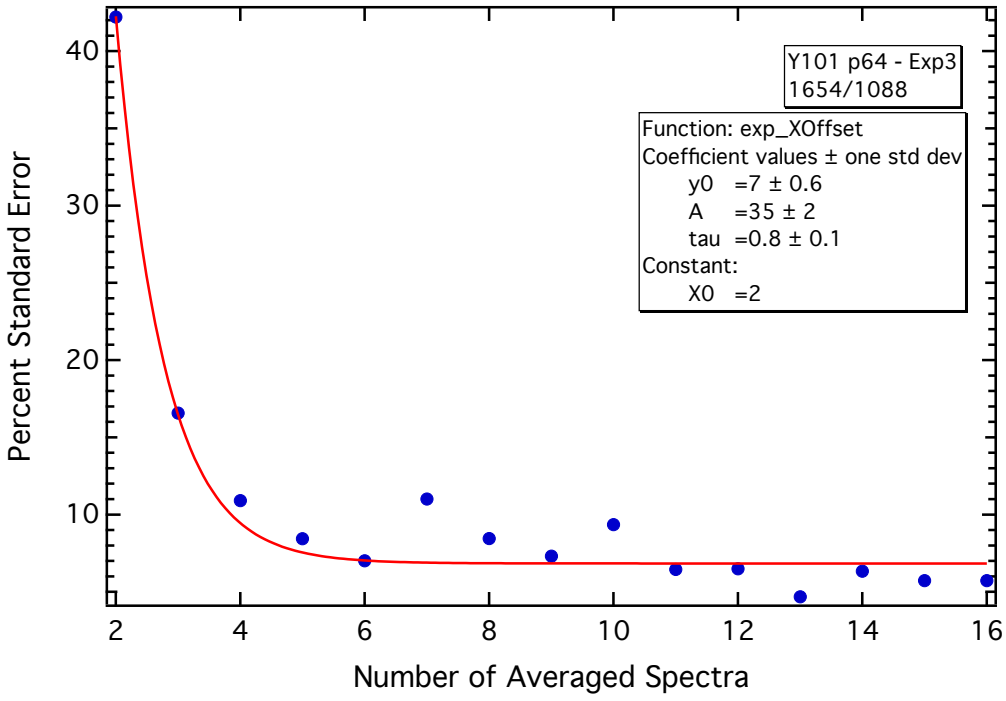


(a)

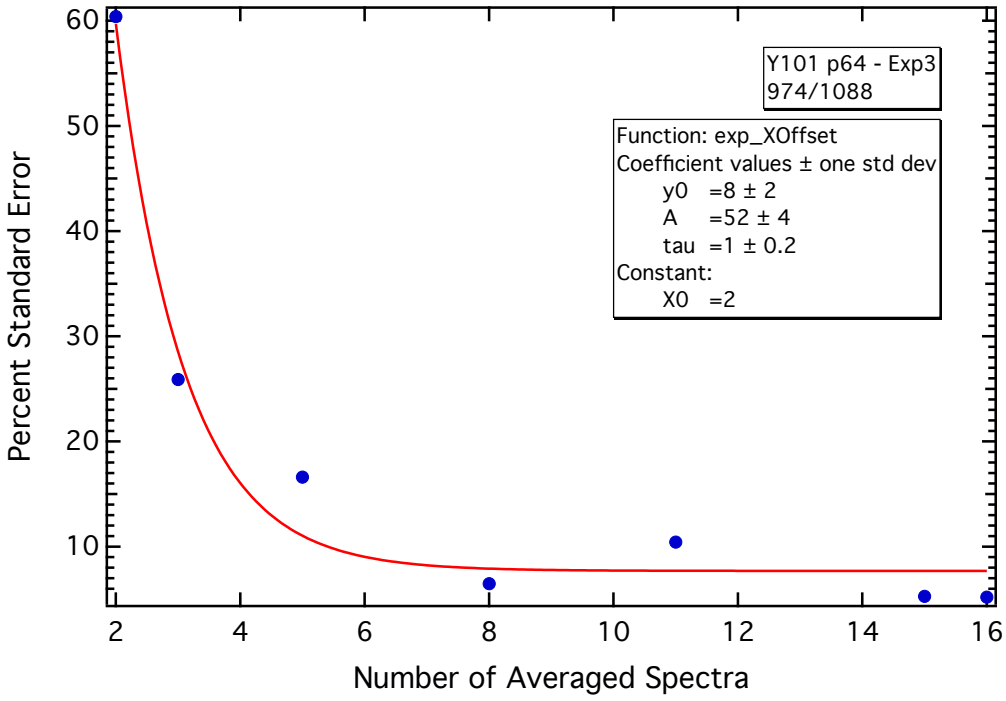


(b)

Figure B.33: %SE convergence tests for live cell analyses of MSC line Y101 experiment 2. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

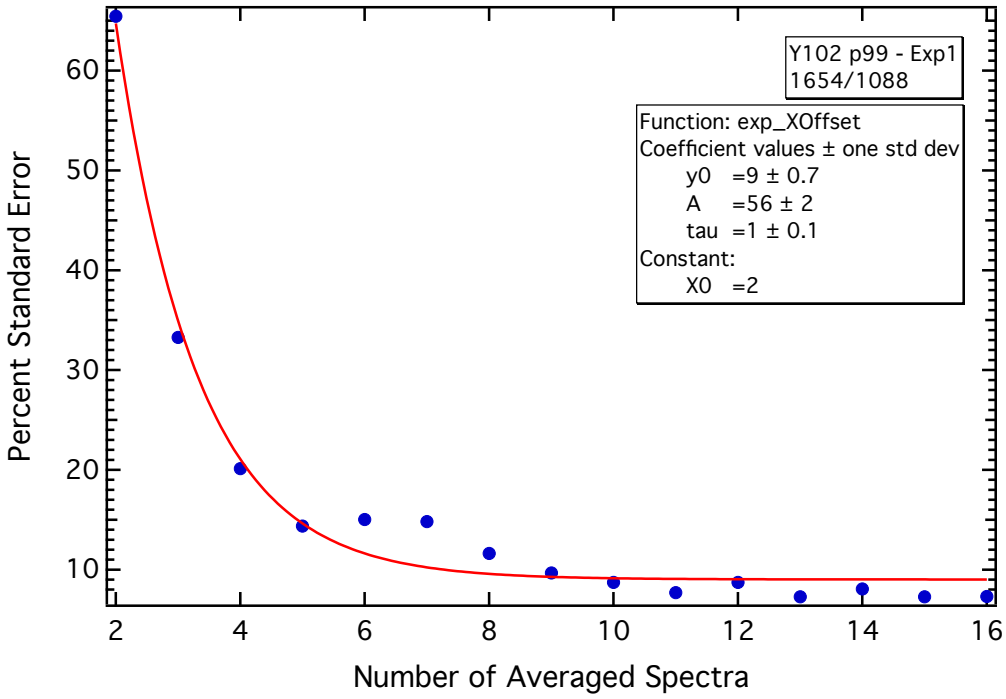


(a)

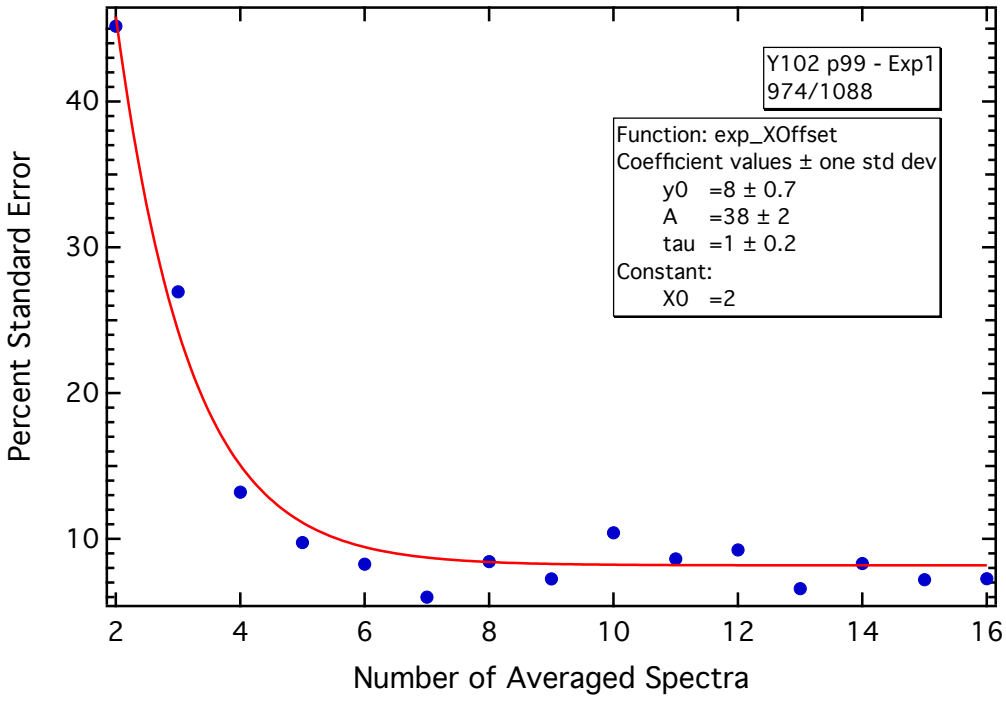


(b)

Figure B.34: %SE convergence tests for live cell analyses of MSC line Y101 experiment 3. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

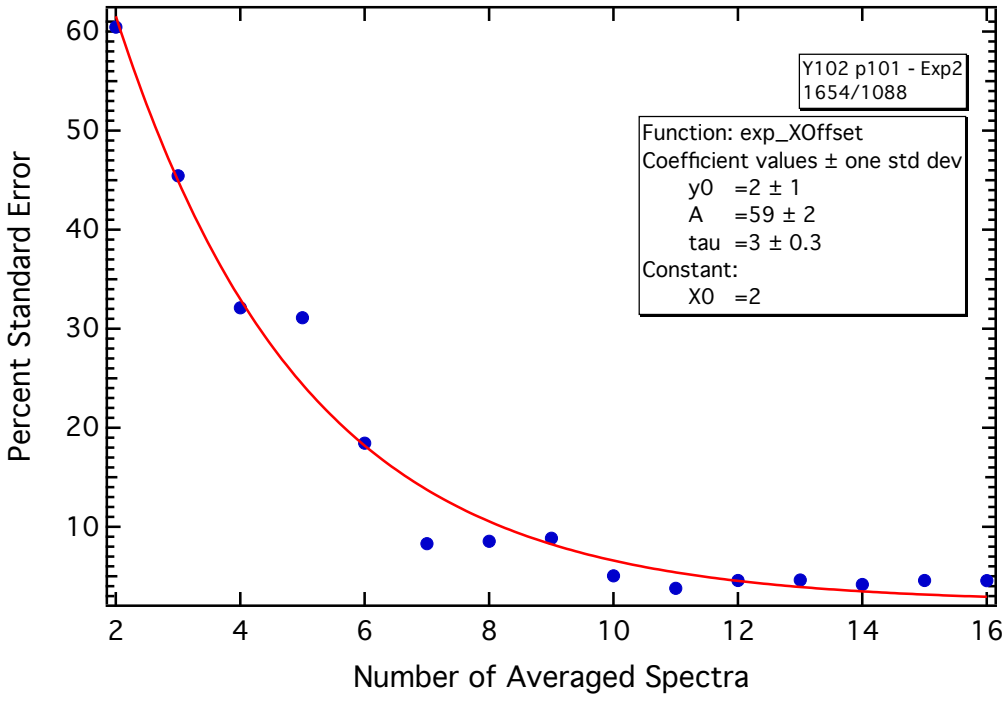


(a)

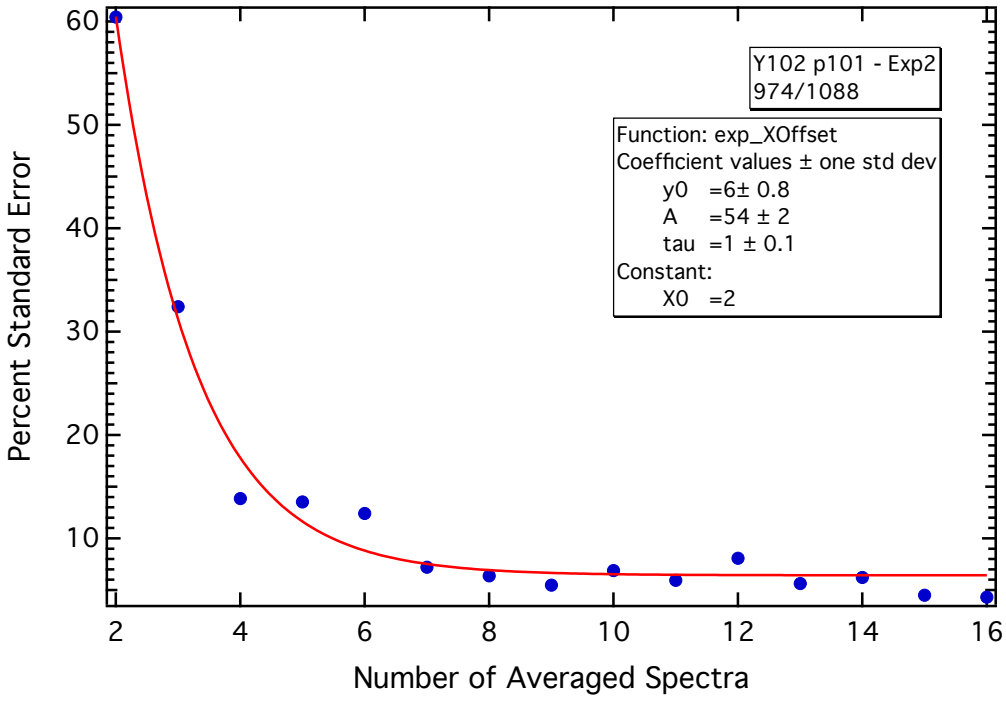


(b)

Figure B.35: %SE convergence tests for live cell analyses of MSC lines Y102 experiment 1. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

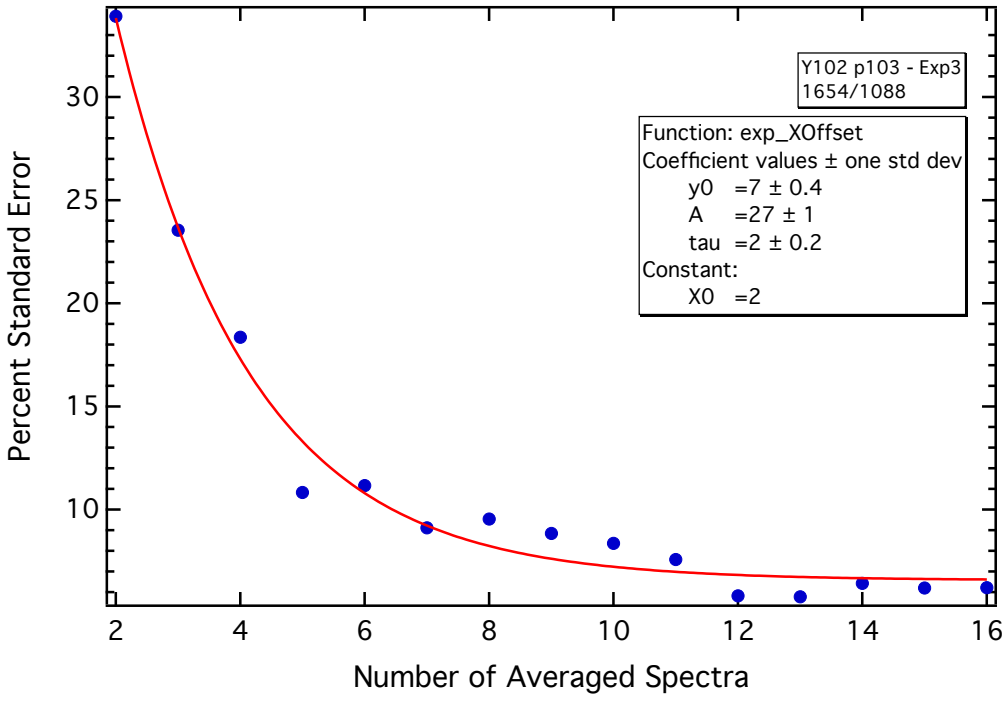


(a)

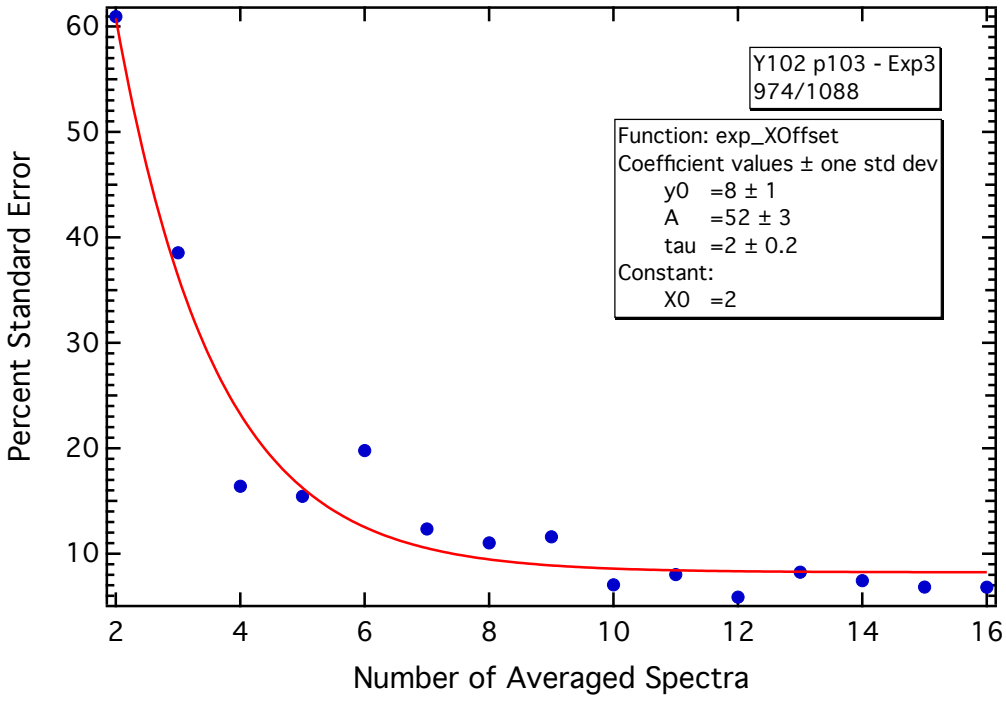


(b)

Figure B.36: %SE convergence tests for live cell analyses of MSC line Y102 experiment 2. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

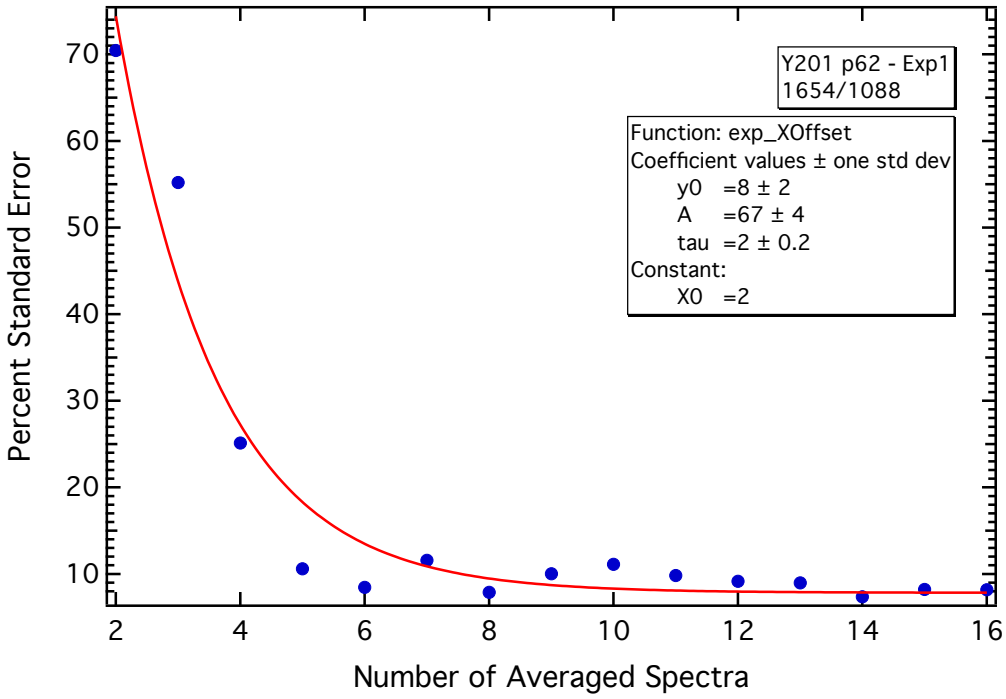


(a)

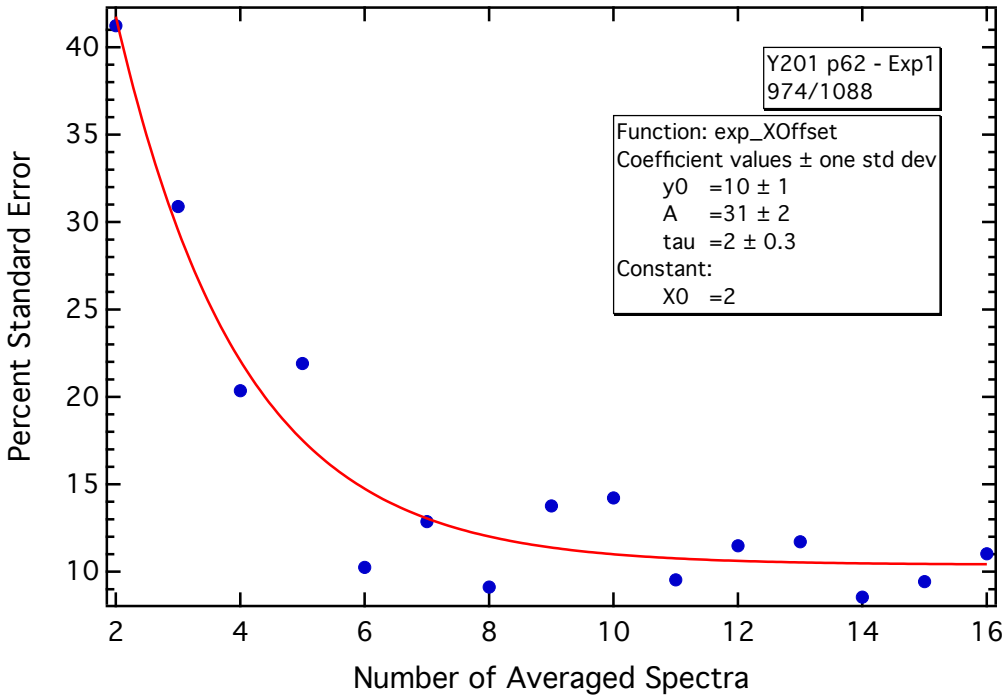


(b)

Figure B.37: %SE convergence tests for live cell analyses of MSC line Y102 experiment 3. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

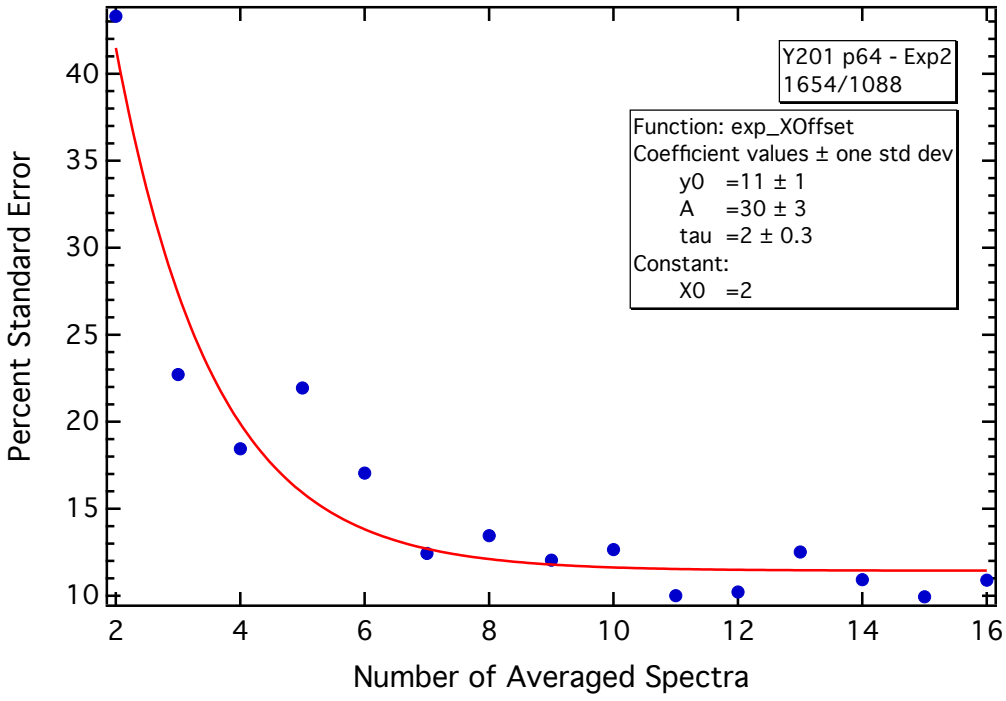


(a)

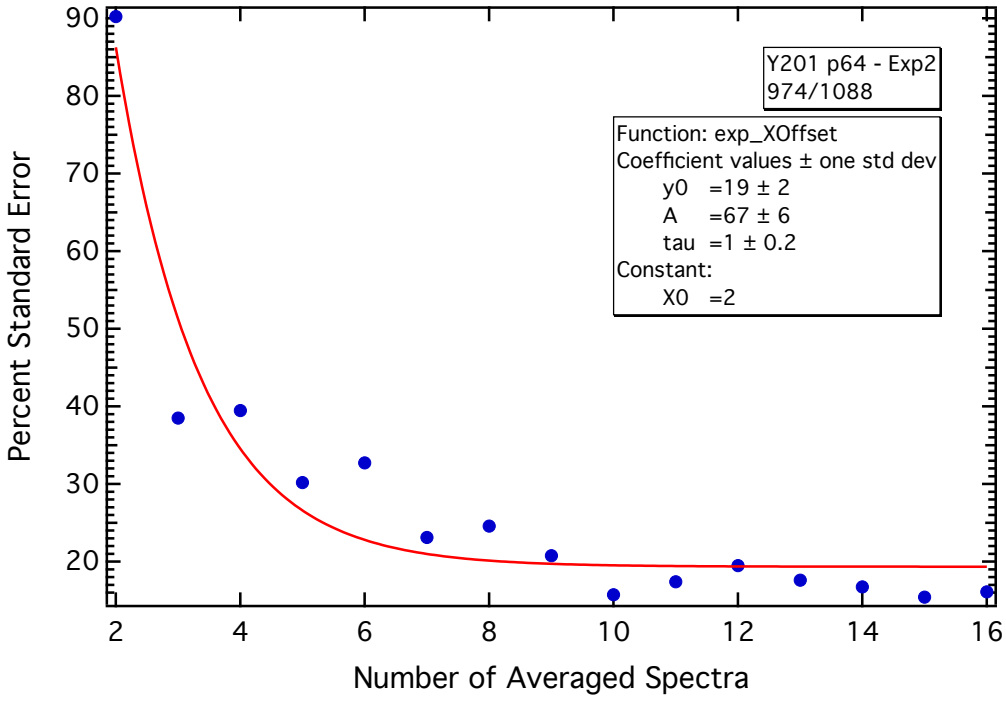


(b)

Figure B.38: %SE convergence tests for live cell analyses of MSC line Y201experiment 1. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

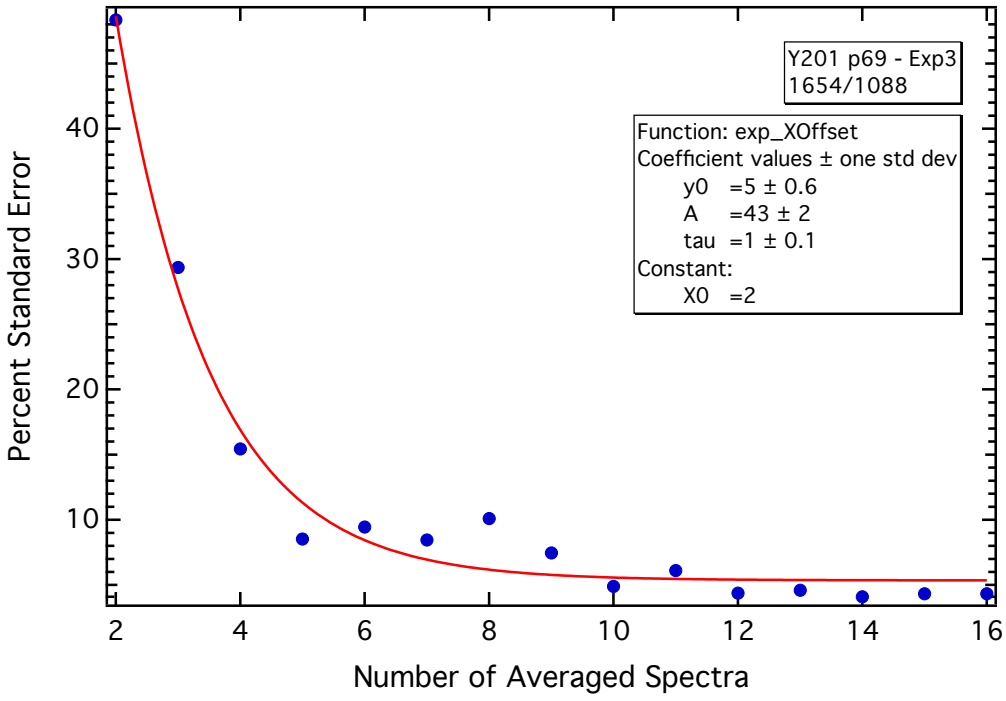


(a)

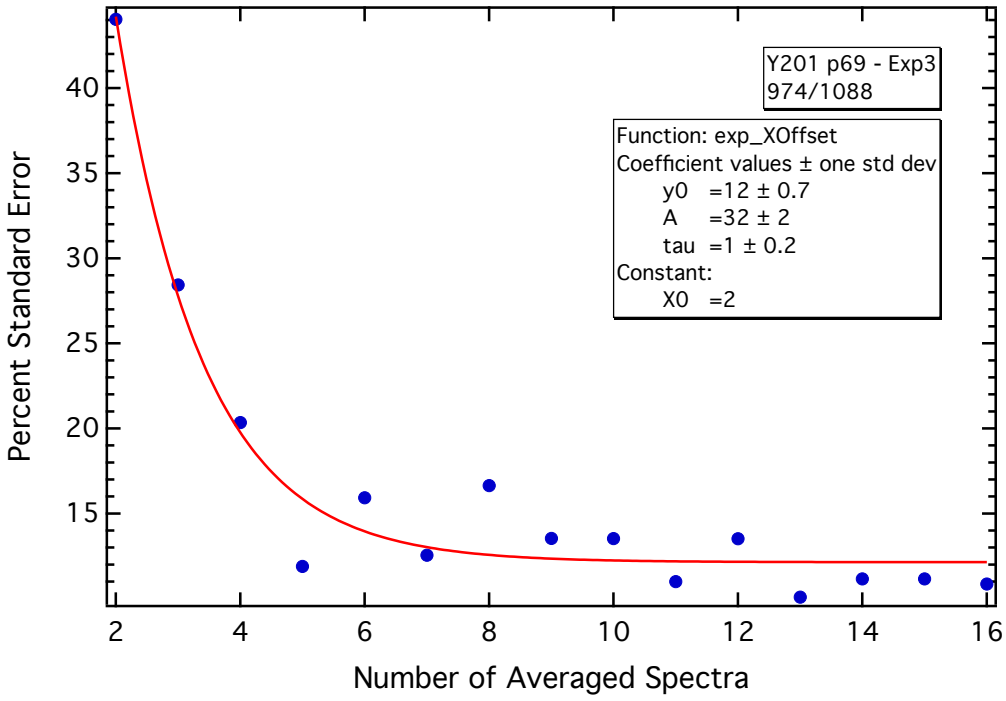


(b)

Figure B.39: %SE convergence tests for live cell analyses of MSC line Y201 experiment 2. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

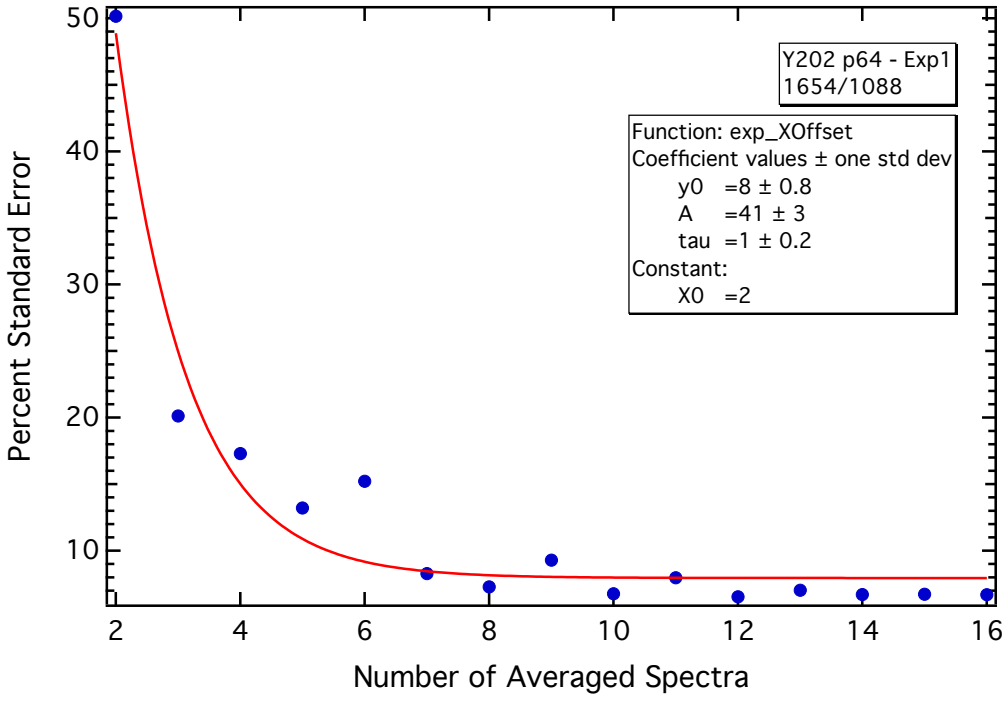


(a)

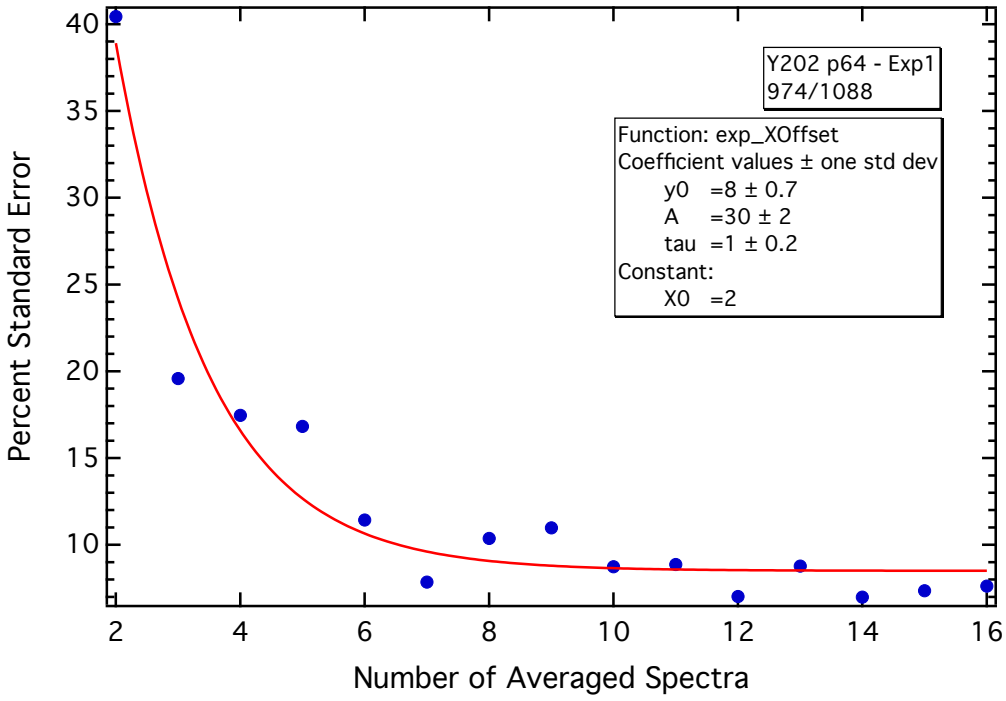


(b)

Figure B.40: %SE convergence tests for live cell analyses of MSC line Y201 experiment 3. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

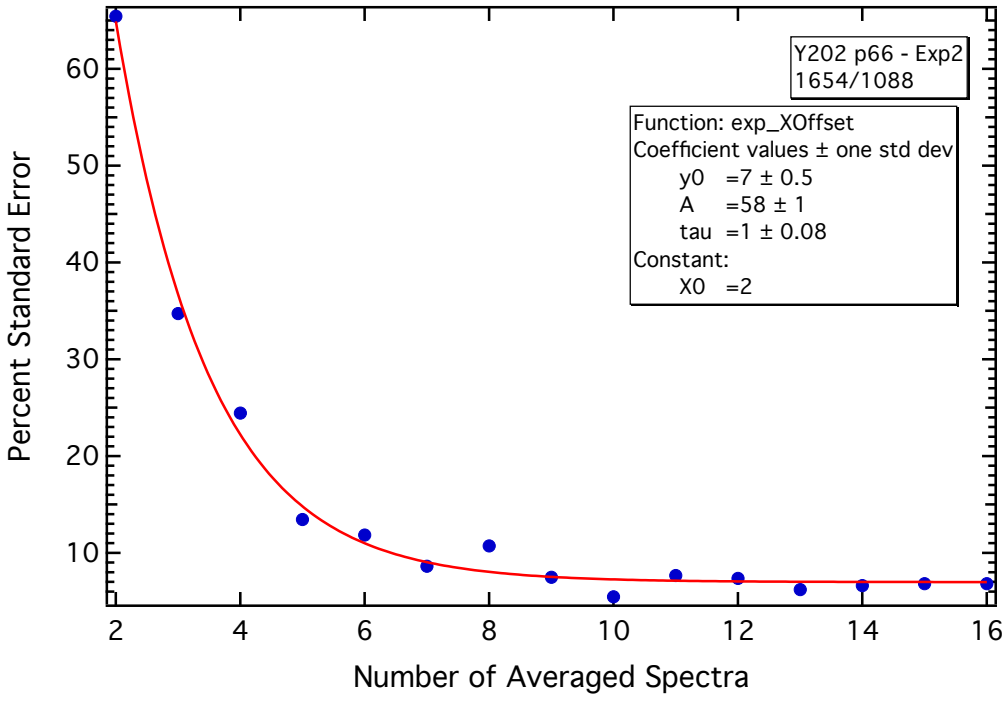


(a)

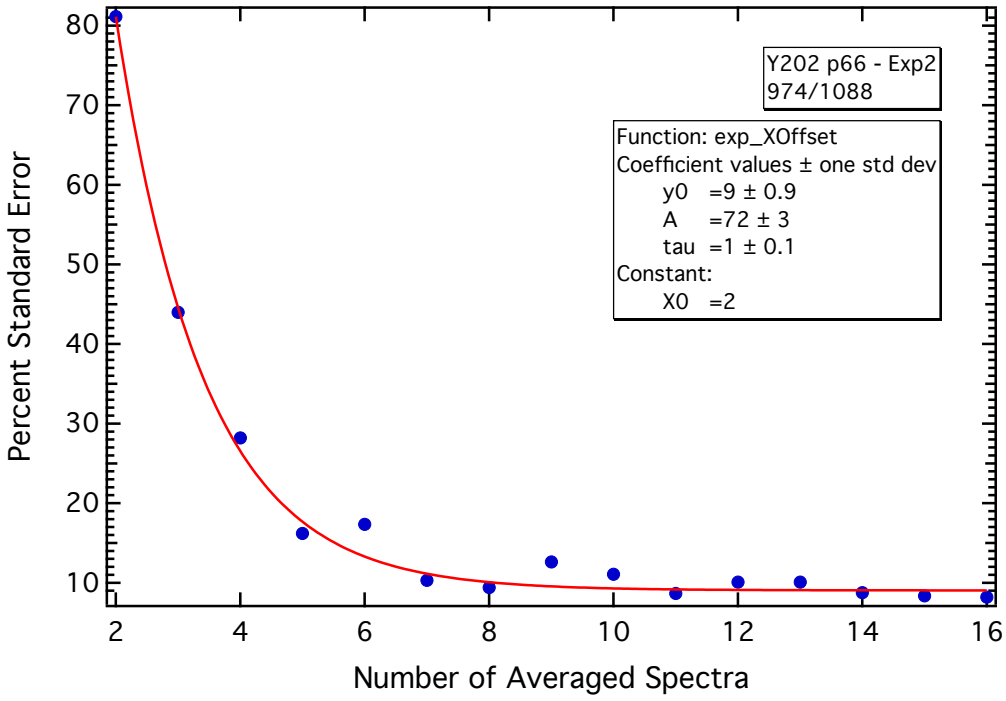


(b)

Figure B.41: %SE convergence tests for live cell analyses of MSC line Y202 experiment 1. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

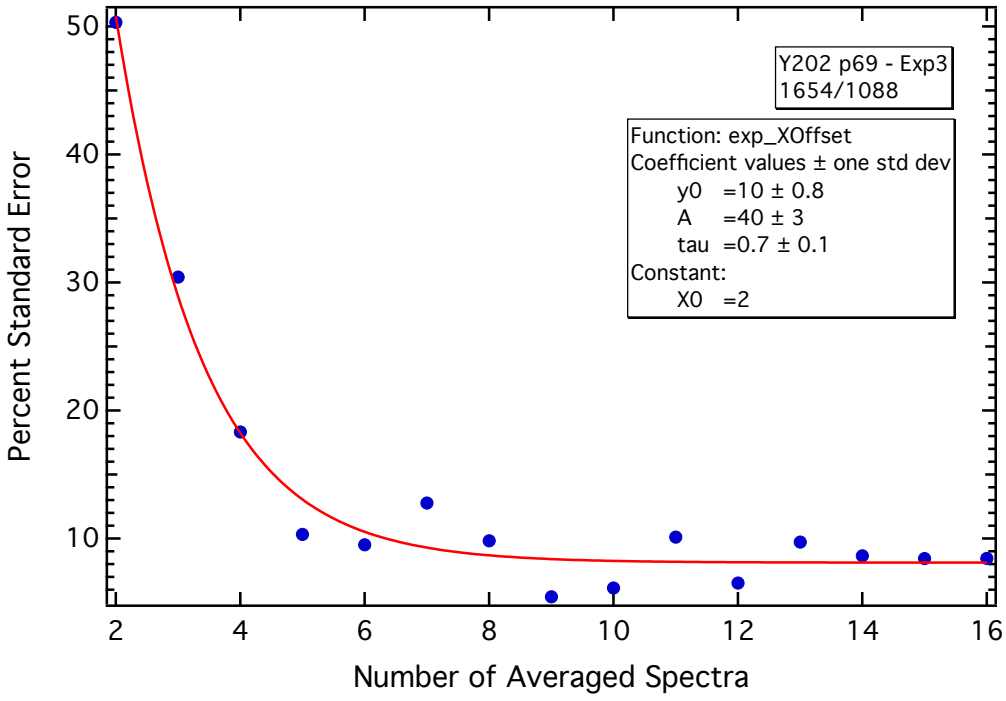


(a)

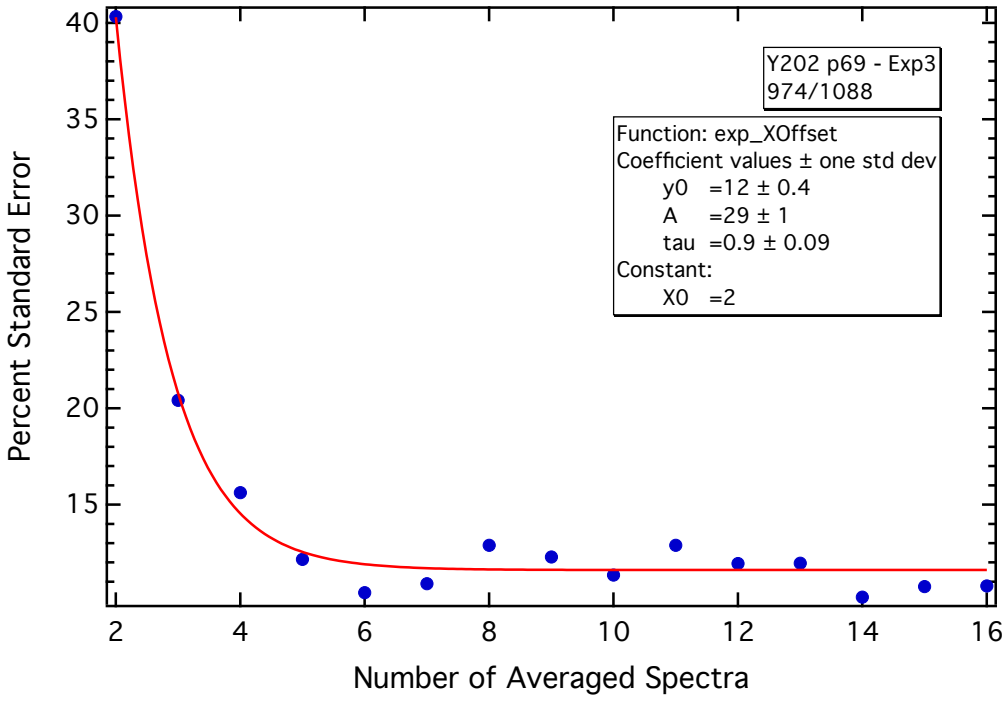


(b)

Figure B.42: %SE convergence tests for live cell analyses of MSC line Y202 experiment 2. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.



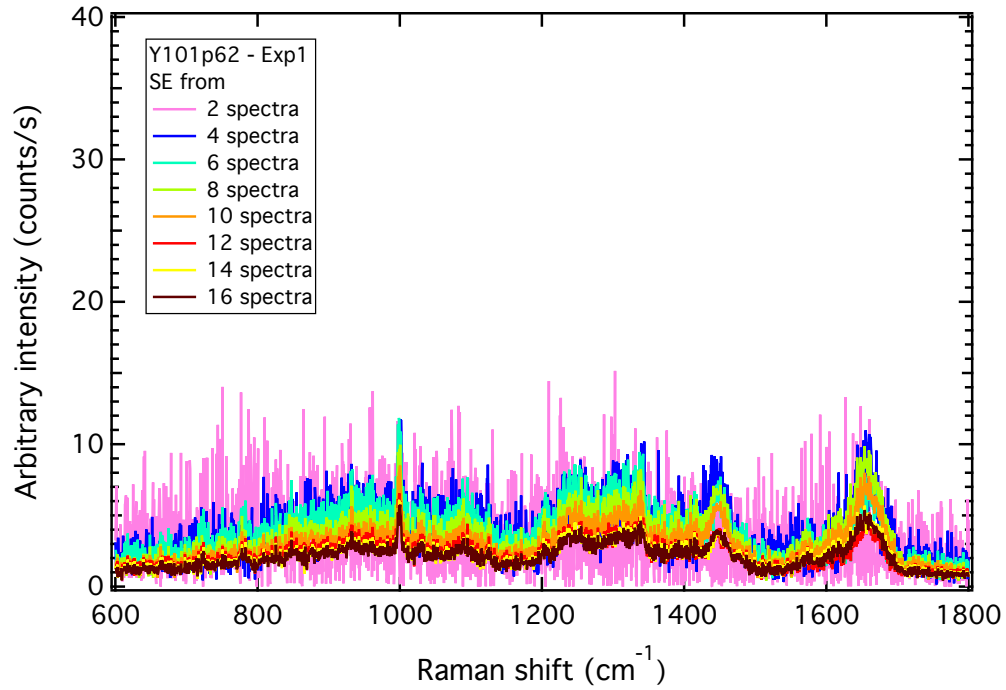
(a)



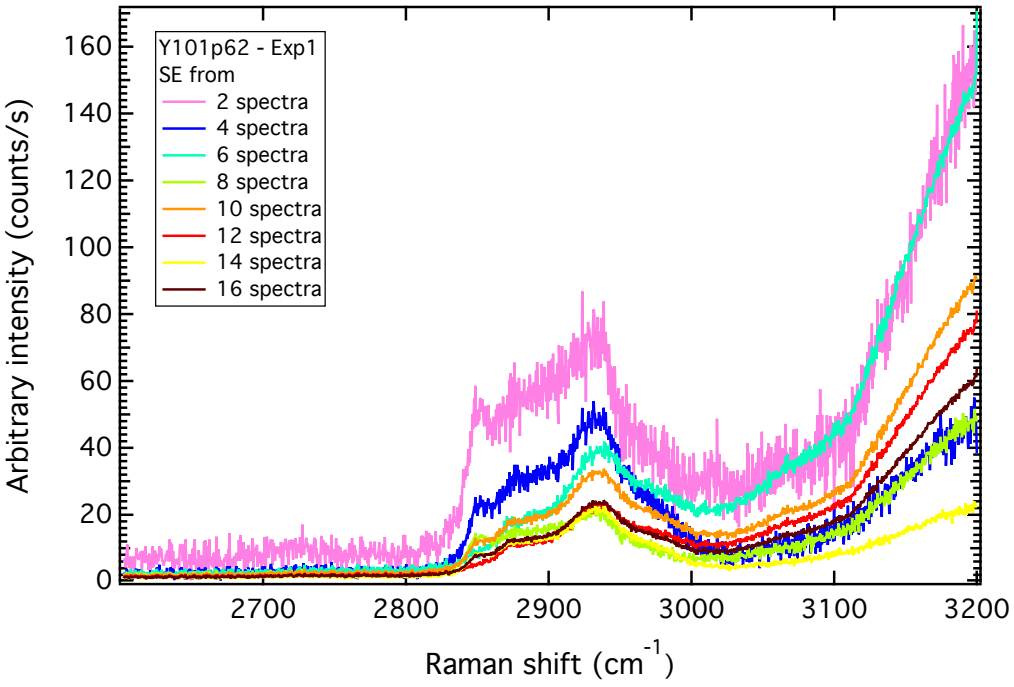
(b)

Figure B.43: %SE convergence tests for live cell analyses of MSC line Y202 experiment 3. Convergence for PIRs (a) 1654/1088 and (b) 974/1088. Convergence for PIRs (a) 1654/1088 and (b) 974/1088.

B.5 Live MSCs – SE convergence tests

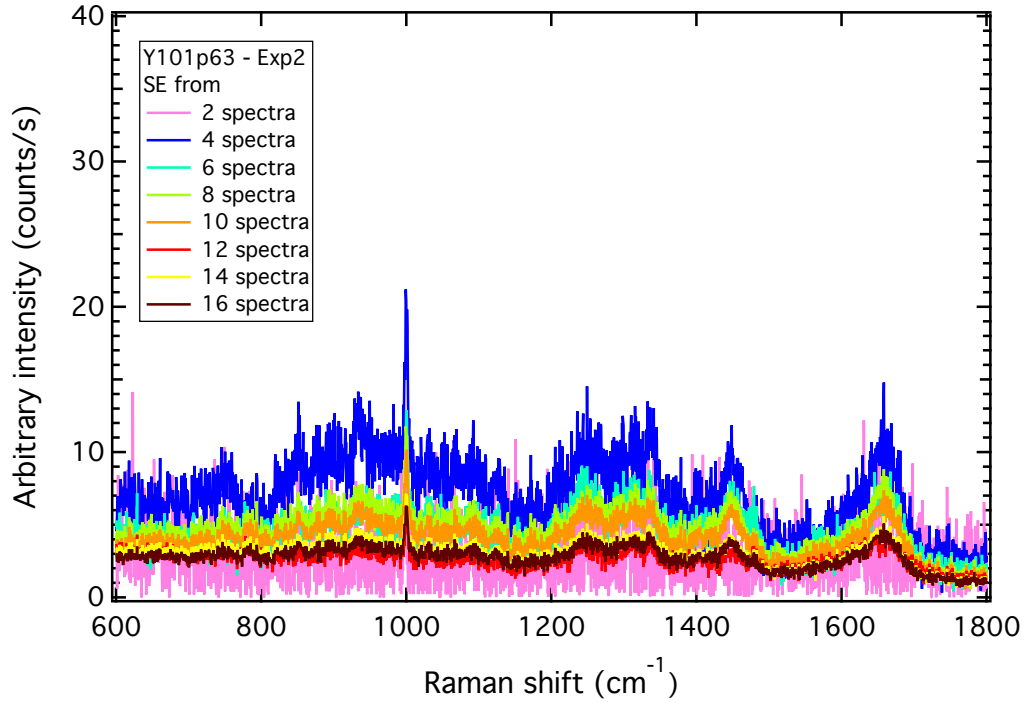


(a)

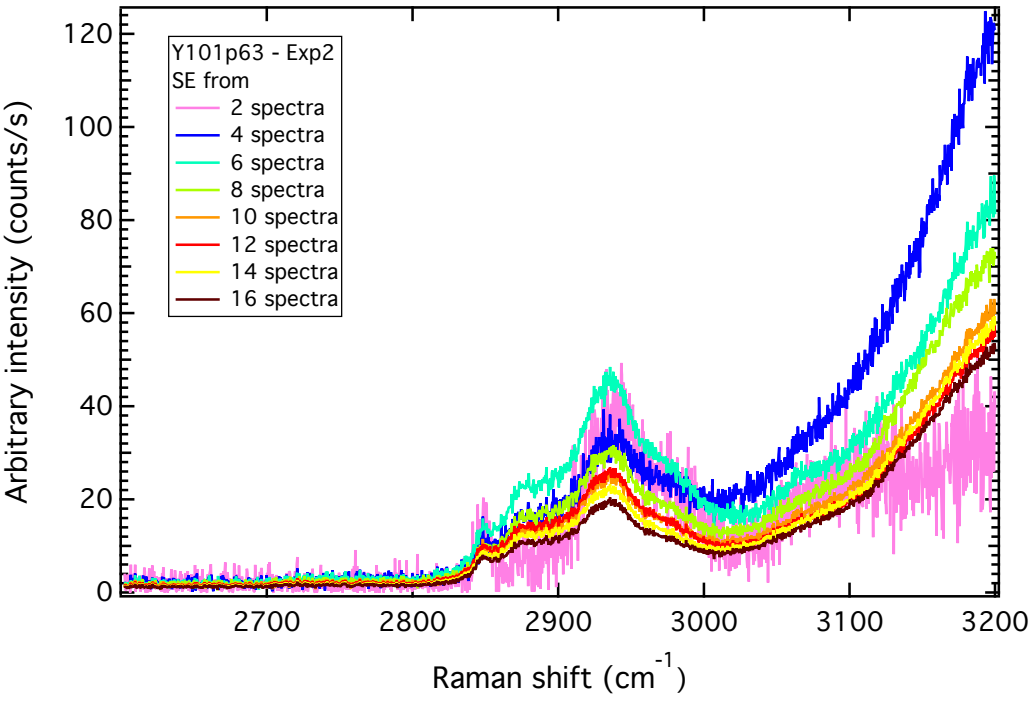


(b)

Figure B.44: SE stacks for live cell analyses of MSC line Y101 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

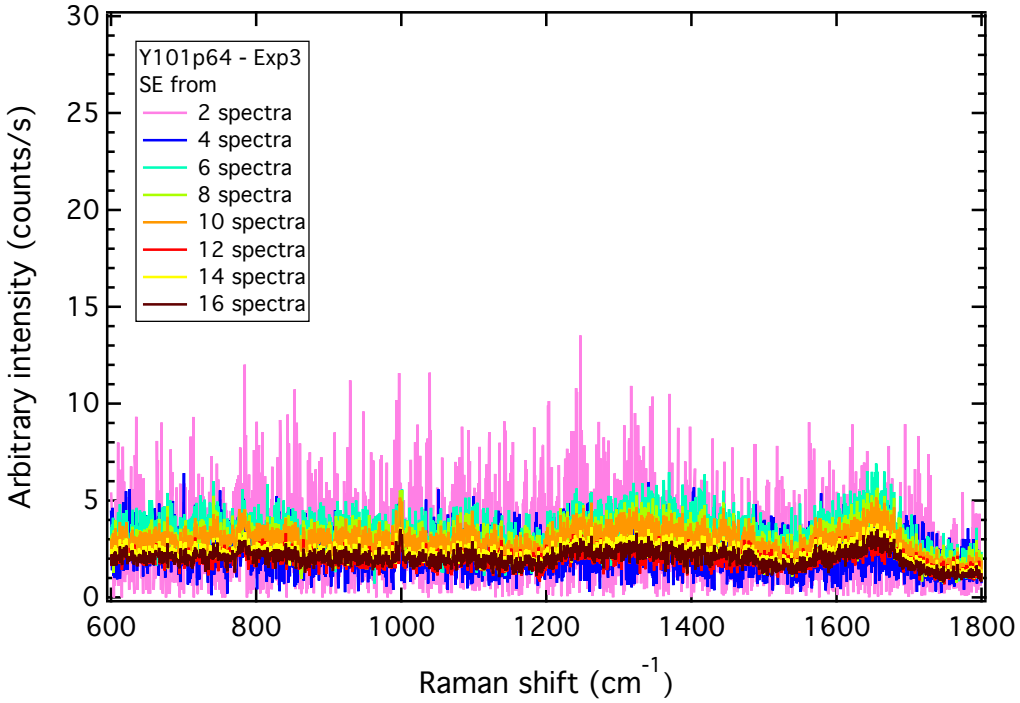


(a)

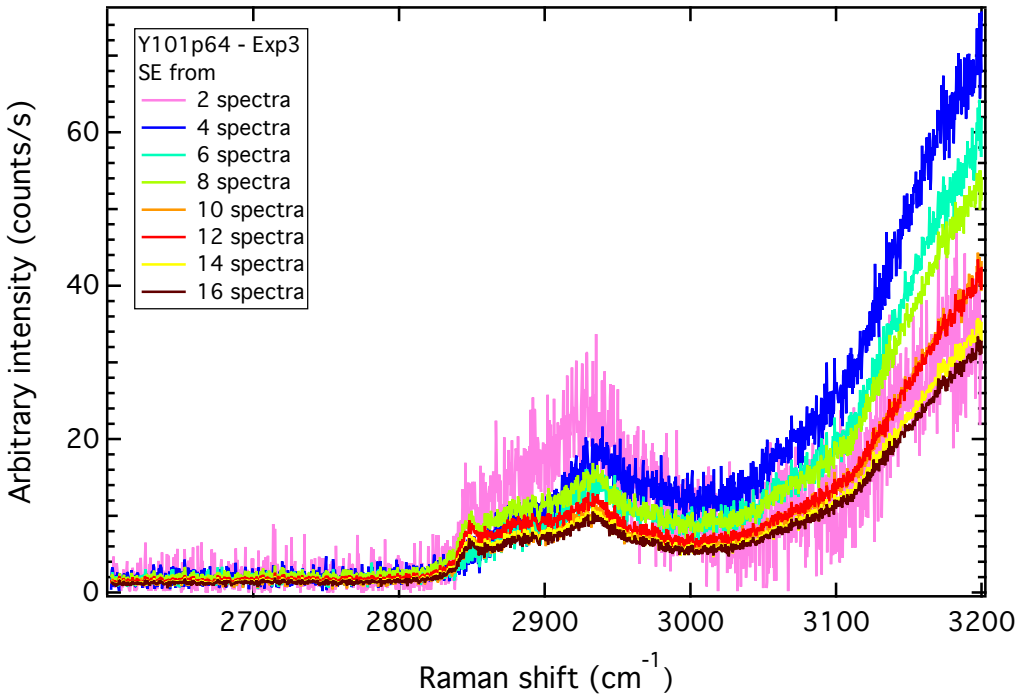


(b)

Figure B.45: SE stacks for live cell analyses of MSC line Y101 experiment 2. (a) Fingerprint and (b) high wavenumber regions.

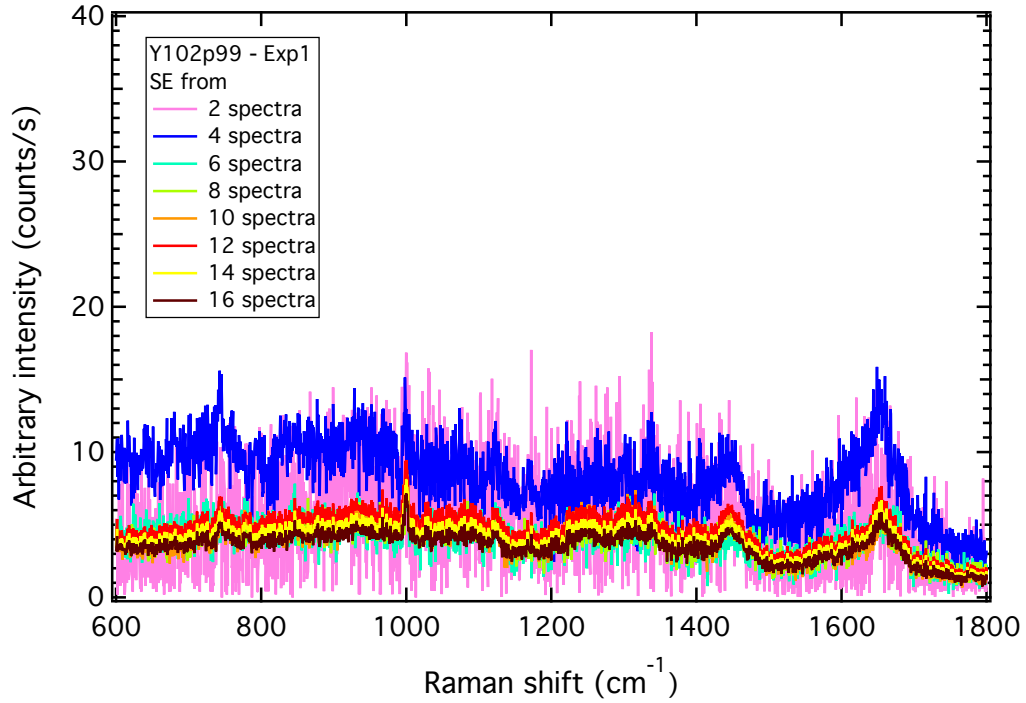


(a)

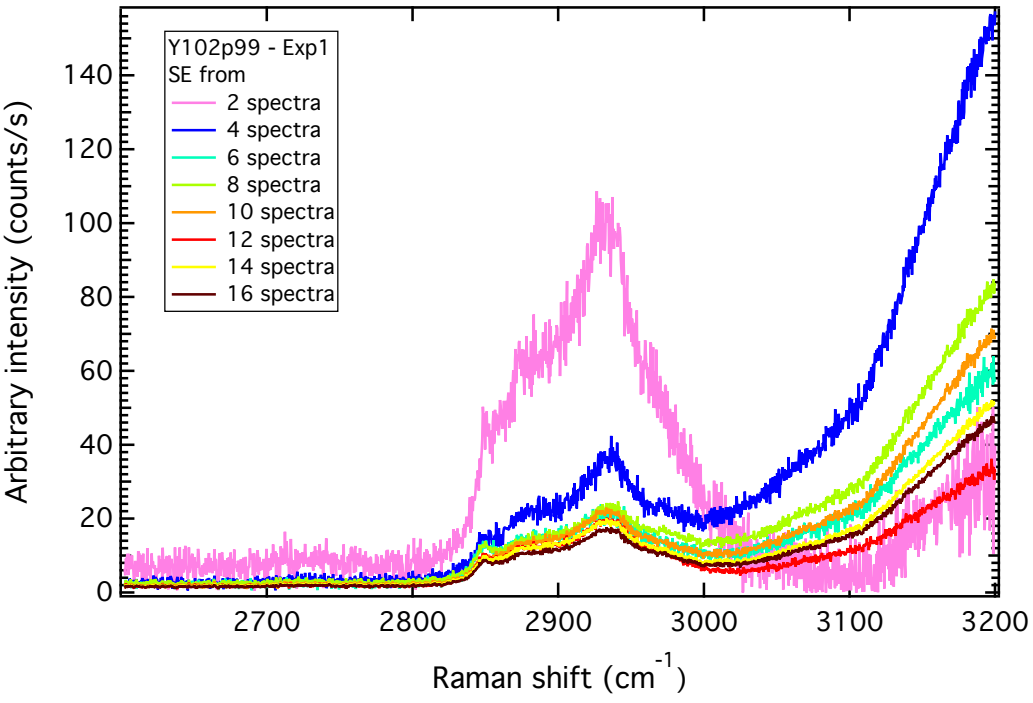


(b)

Figure B.46: SE stacks for live cell analyses of MSC line Y101 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

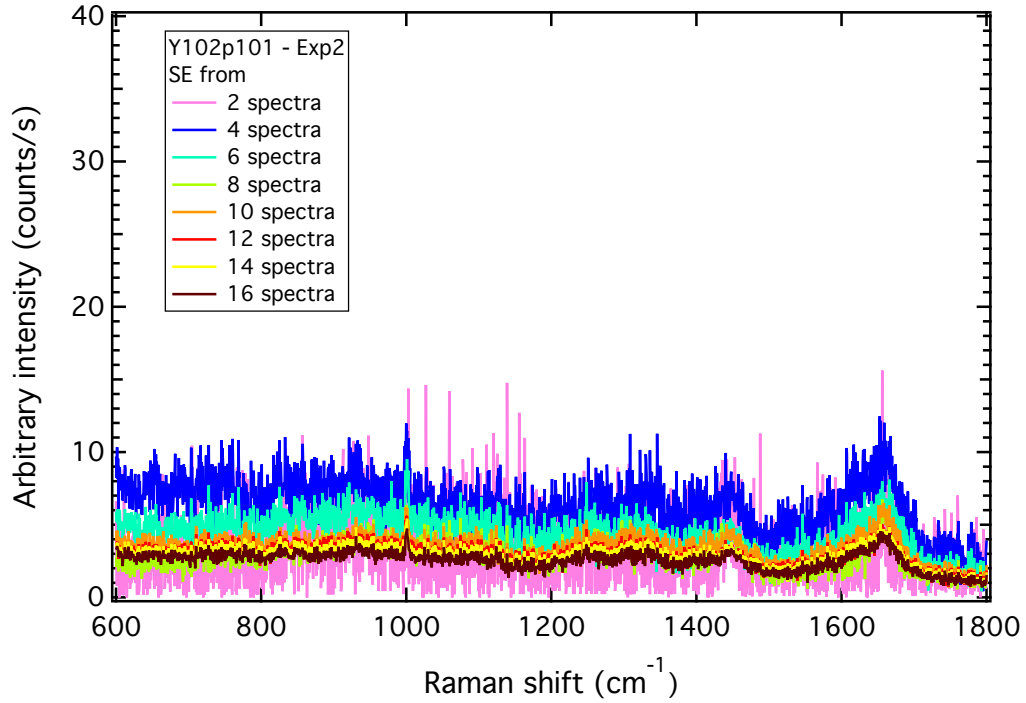


(a)

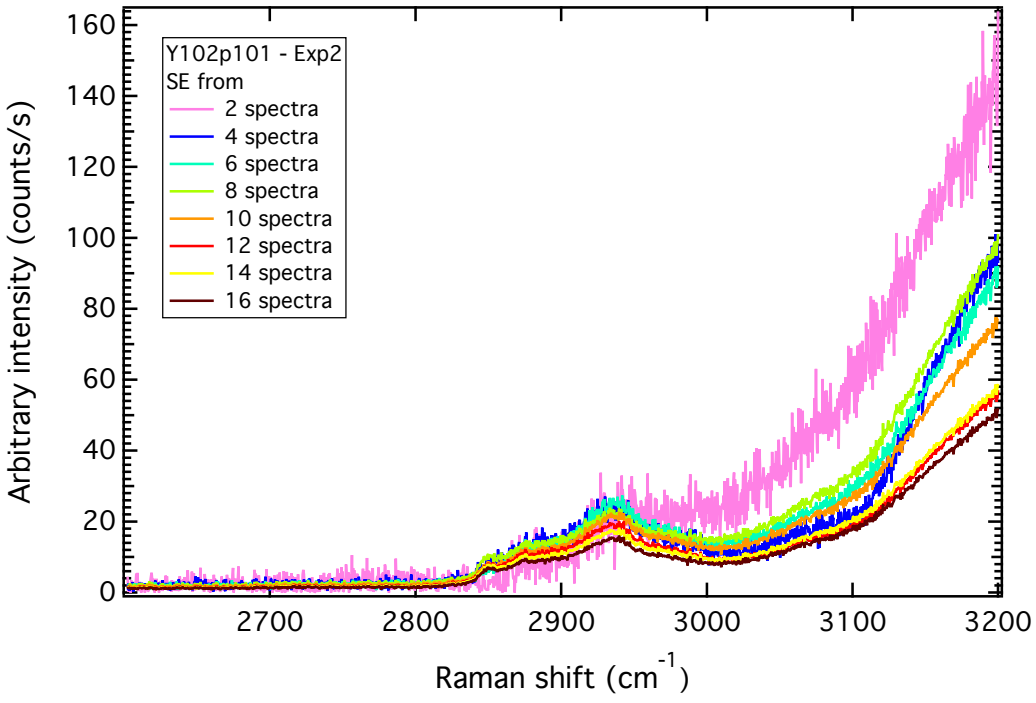


(b)

Figure B.47: SE stacks for live cell analyses of MSC line Y102 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

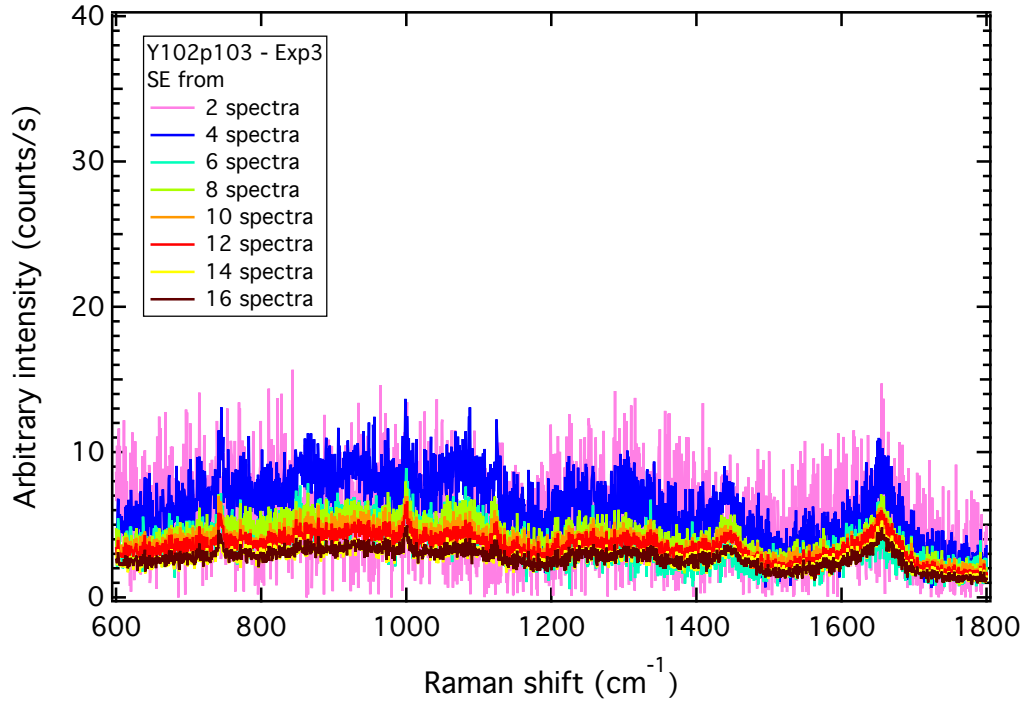


(a)

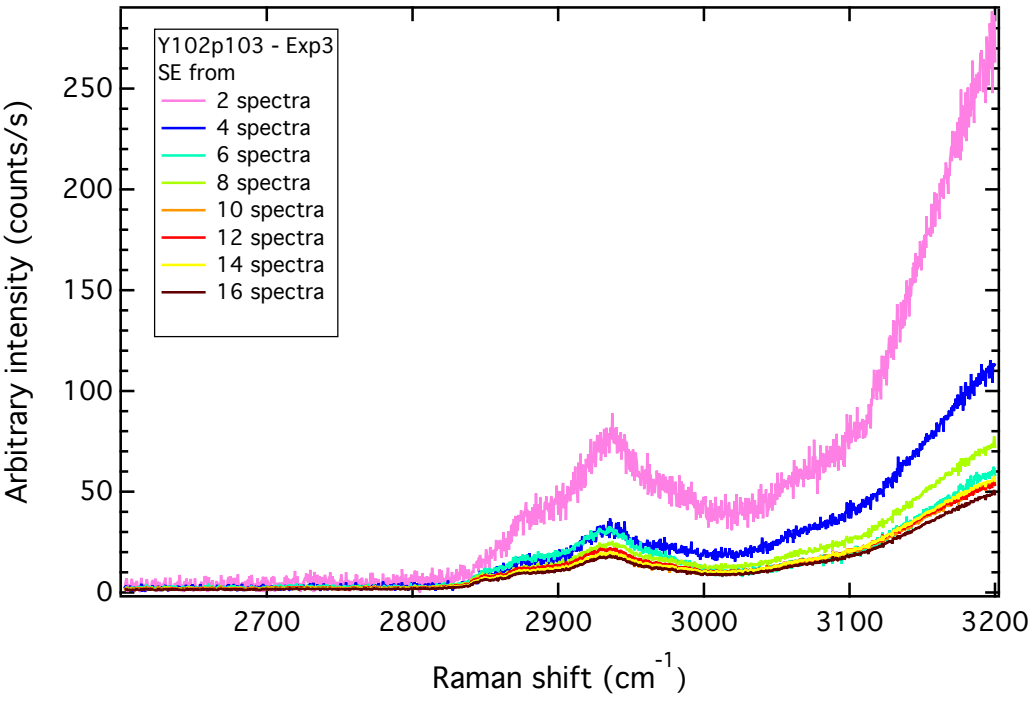


(b)

Figure B.48: SE stacks for live cell analyses of MSC line Y102 experiment 2. (a) Fingerprint and (b) high wavenumber regions.

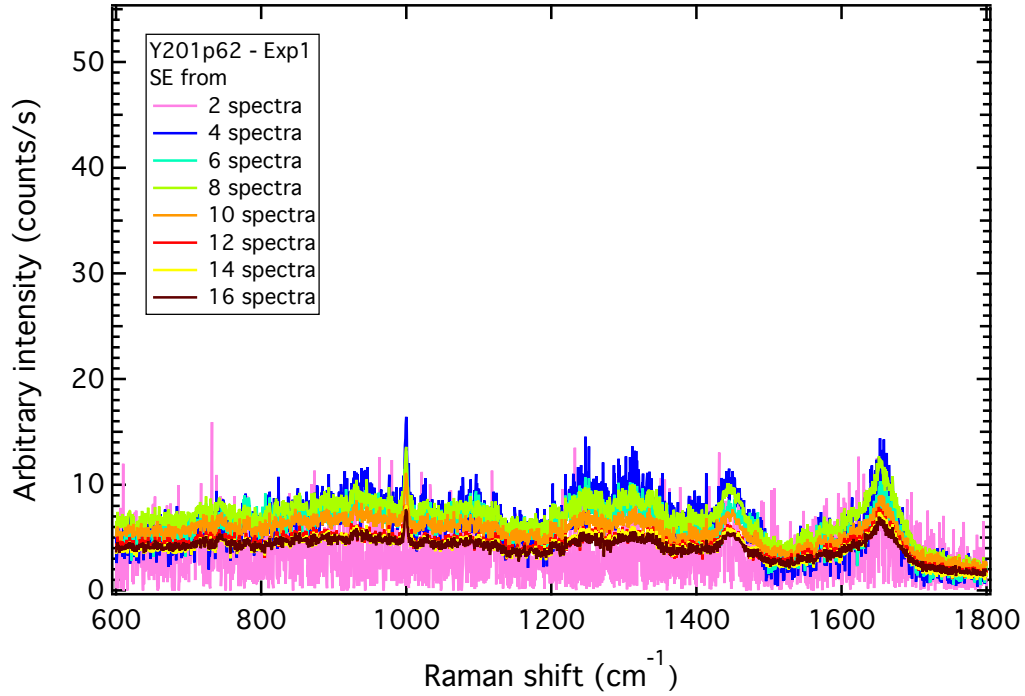


(a)

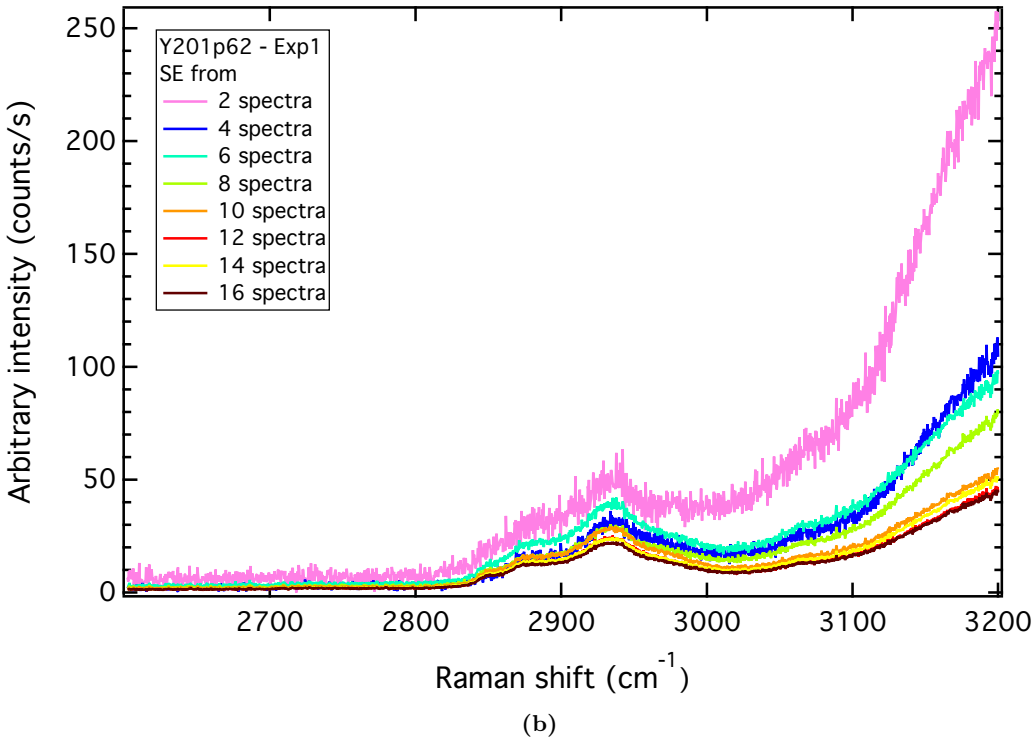


(b)

Figure B.49: SE stacks for live cell analyses of MSC line Y102 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

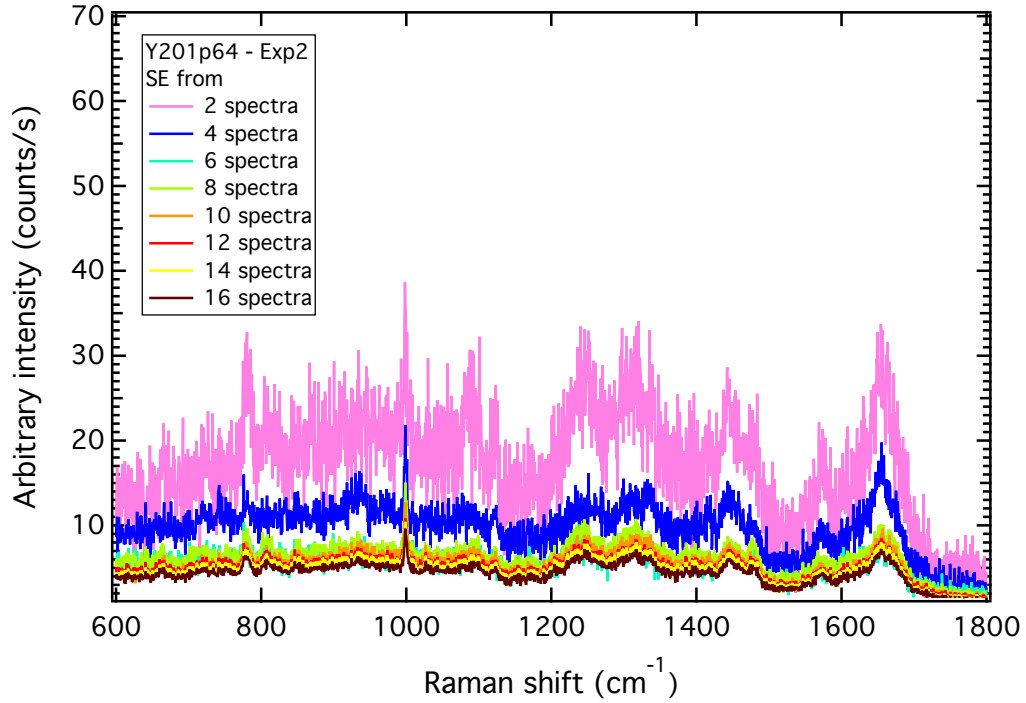


(a)

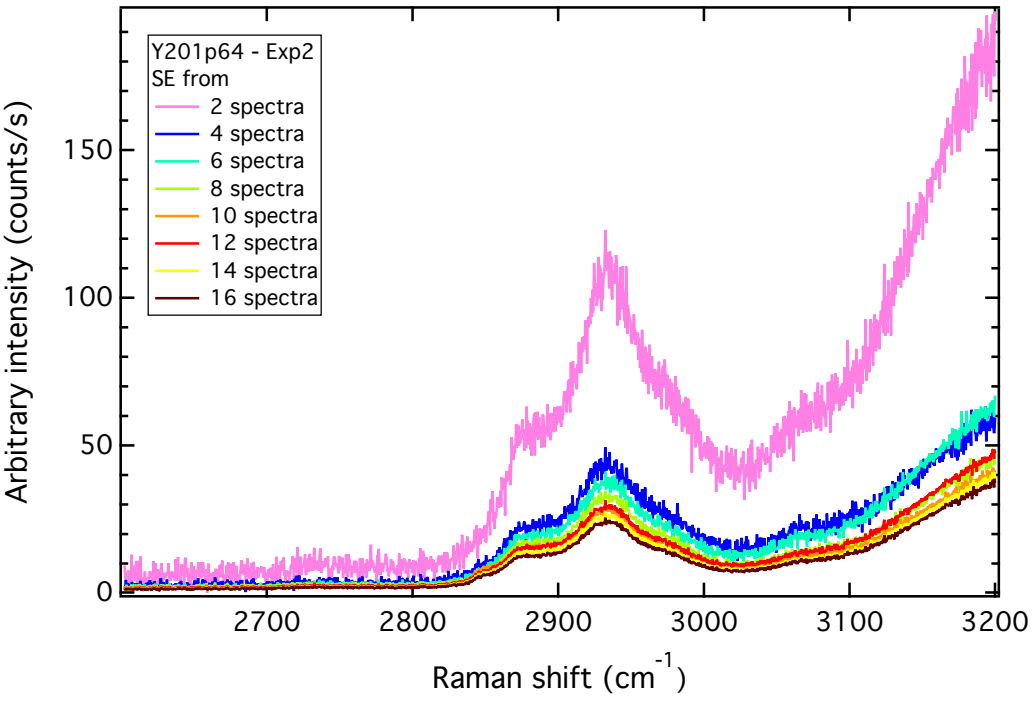


(b)

Figure B.50: SE stacks for live cell analyses of MSC line Y201 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

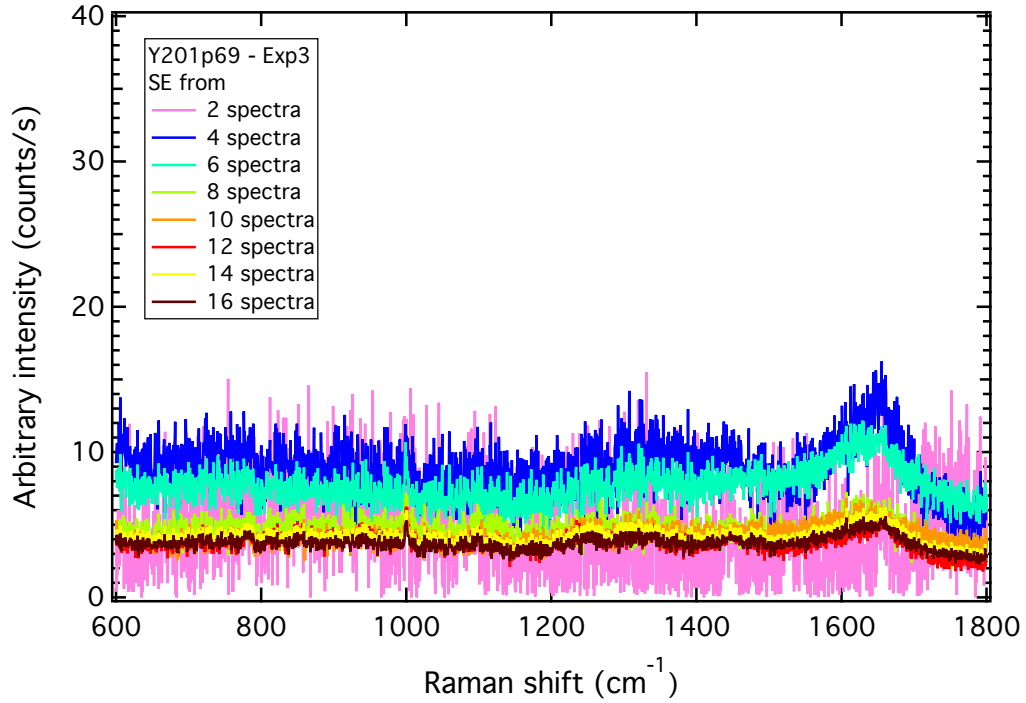


(a)

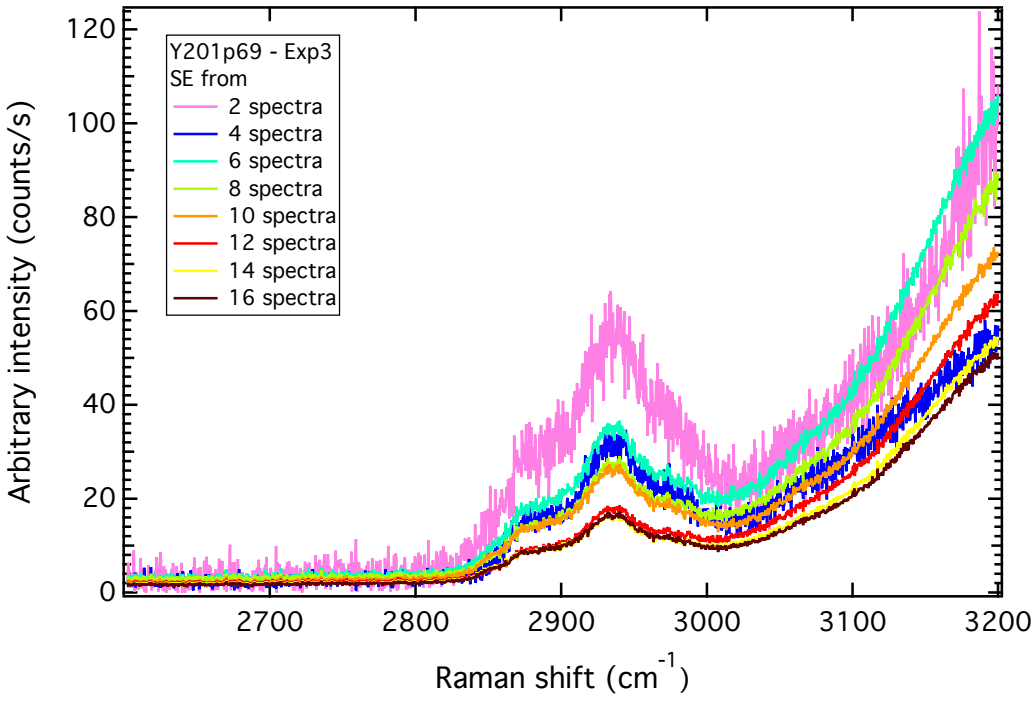


(b)

Figure B.51: SE stacks for live cell analyses of MSC line Y201 experiment 2. (a) Fingerprint and (b) high wavenumber regions.

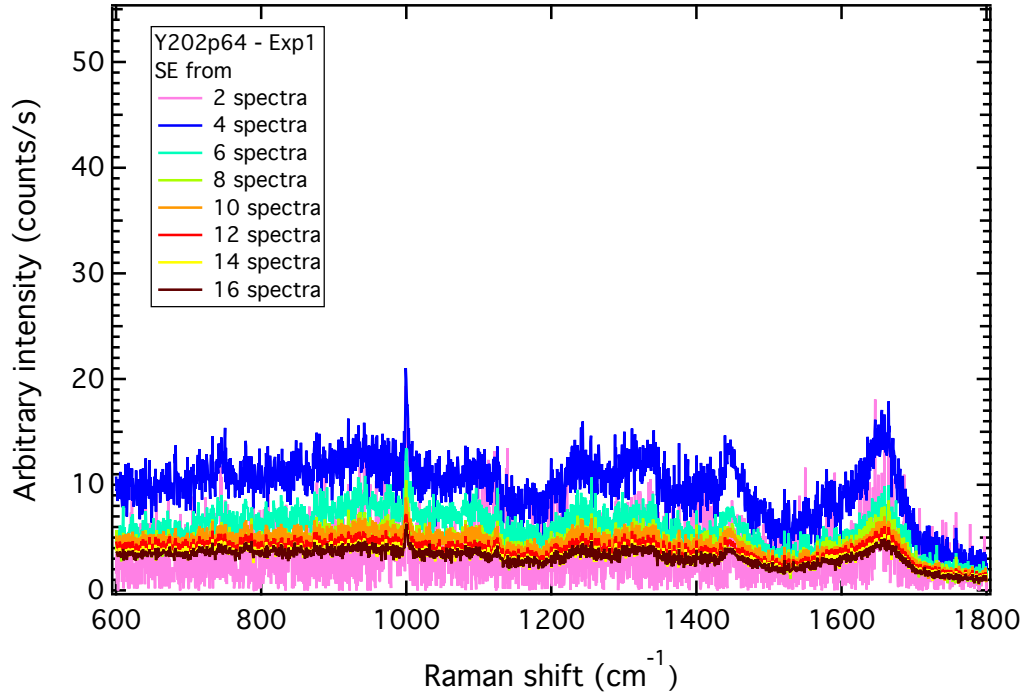


(a)

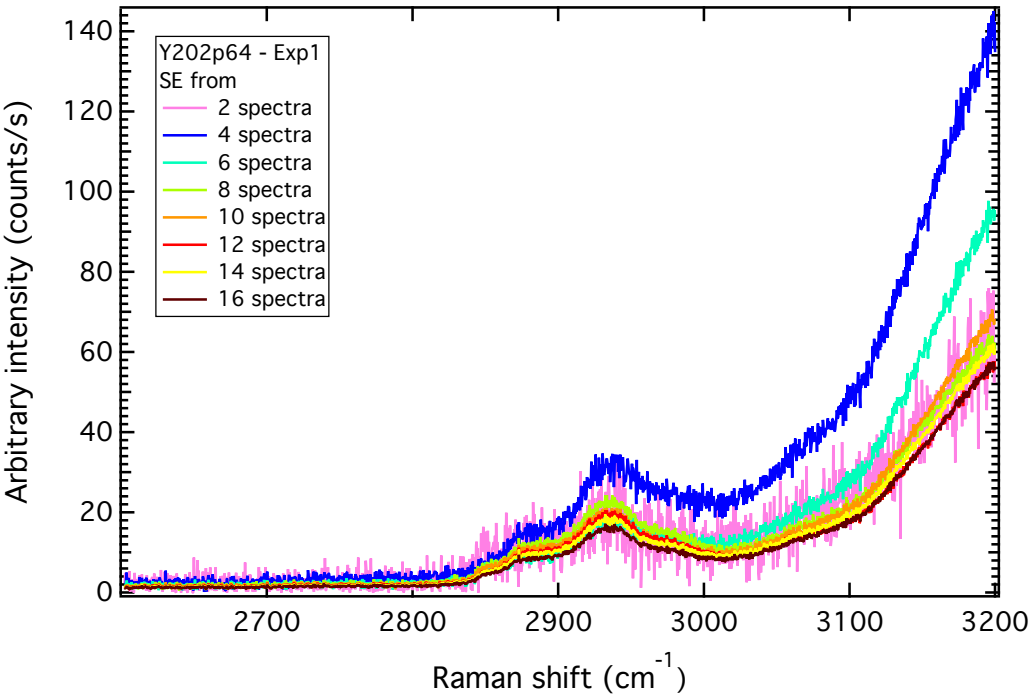


(b)

Figure B.52: SE stacks for live cell analyses of MSC line Y201 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

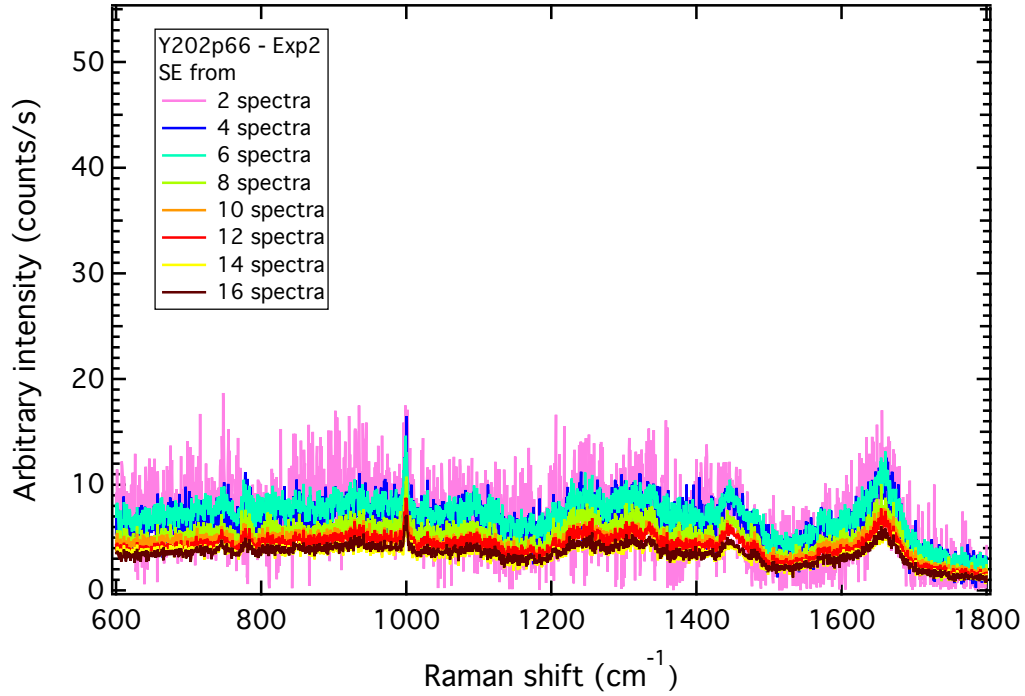


(a)

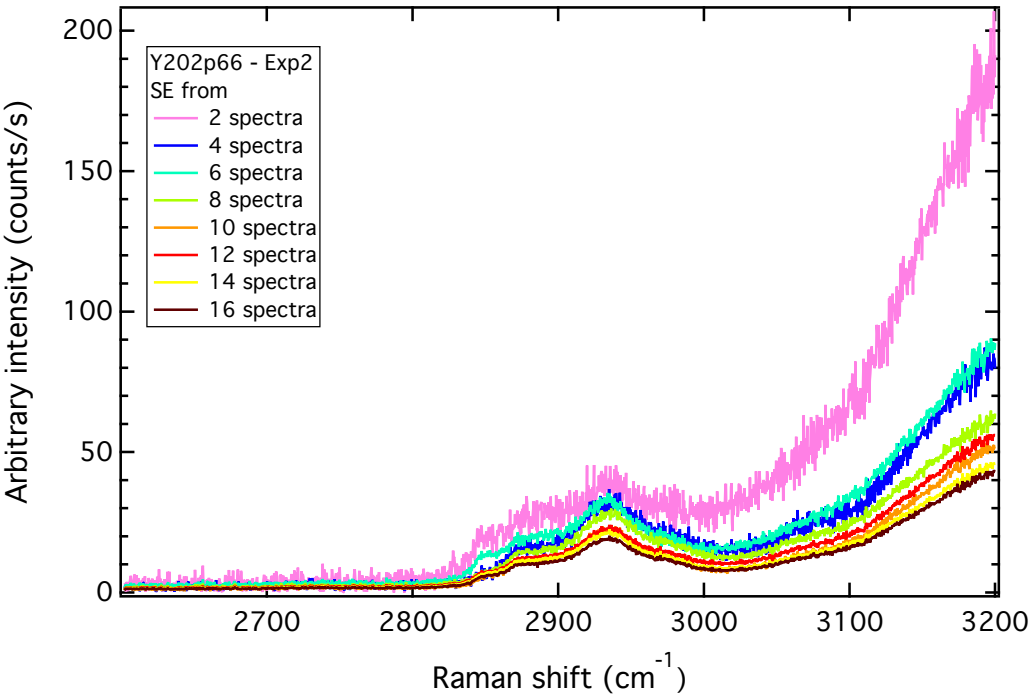


(b)

Figure B.53: SE stacks for live cell analyses of MSC line Y202 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

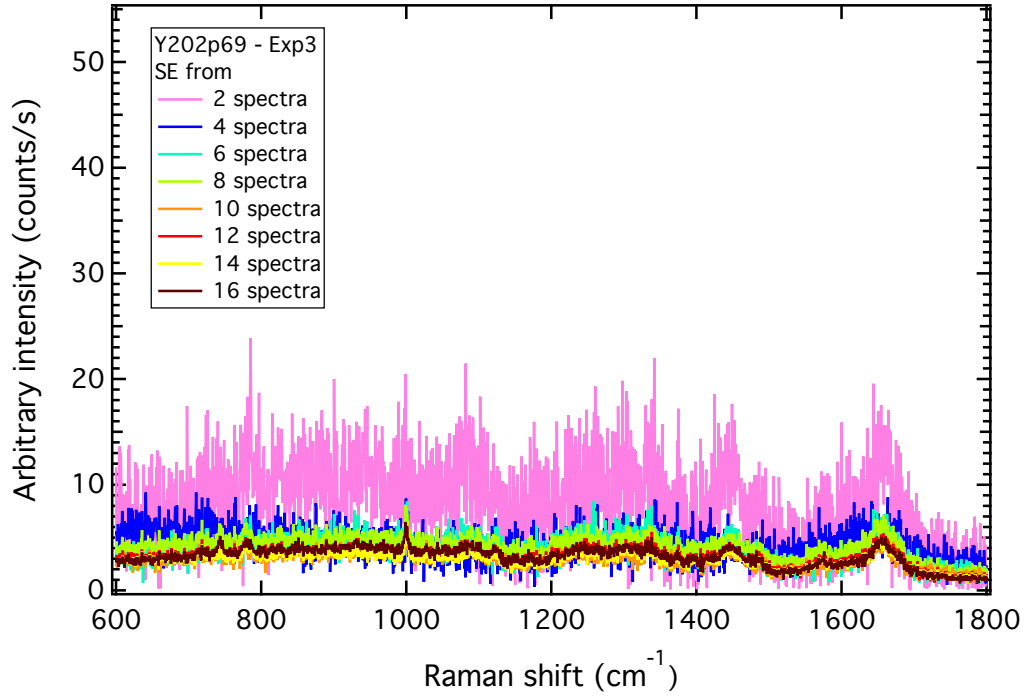


(a)

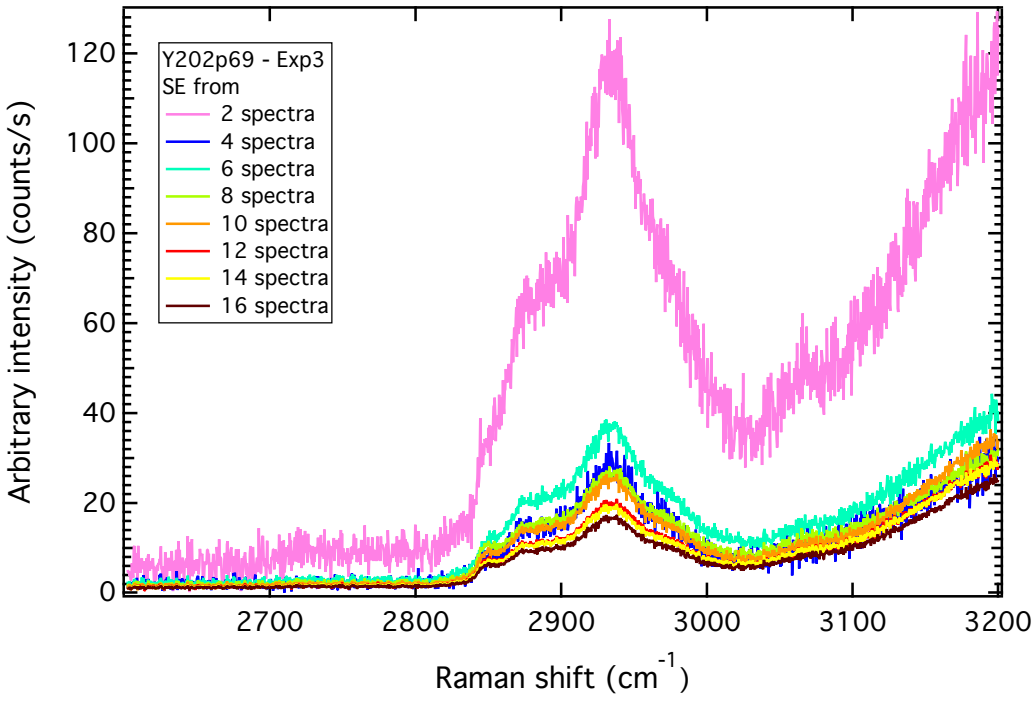


(b)

Figure B.54: SE stacks for live cell analyses of MSC line Y202 experiment 2. (a) Fingerprint and (b) high wavenumber regions.



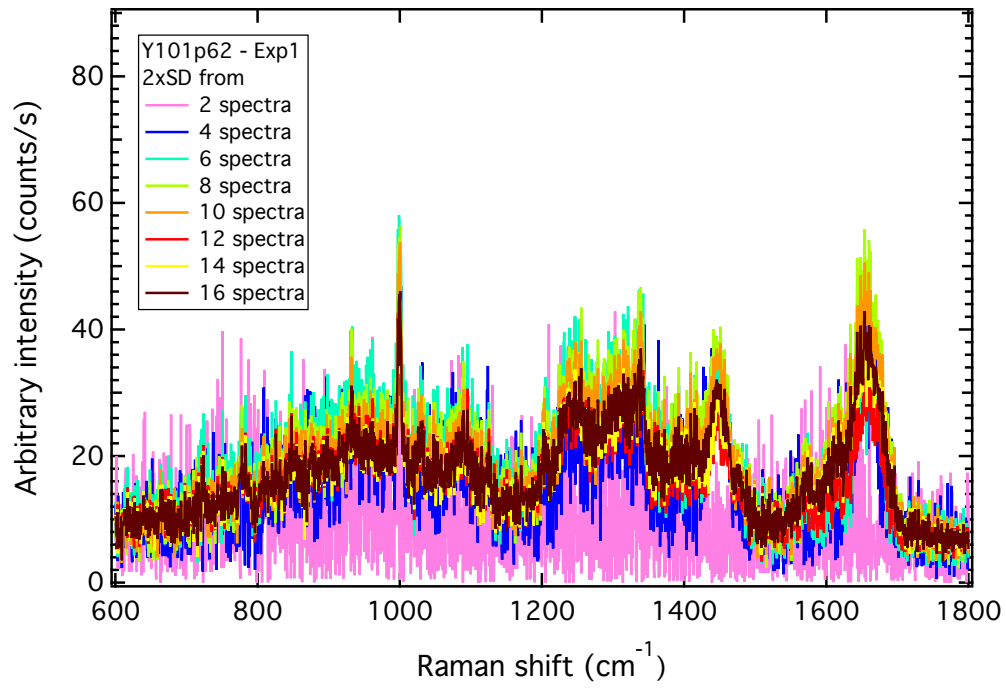
(a)



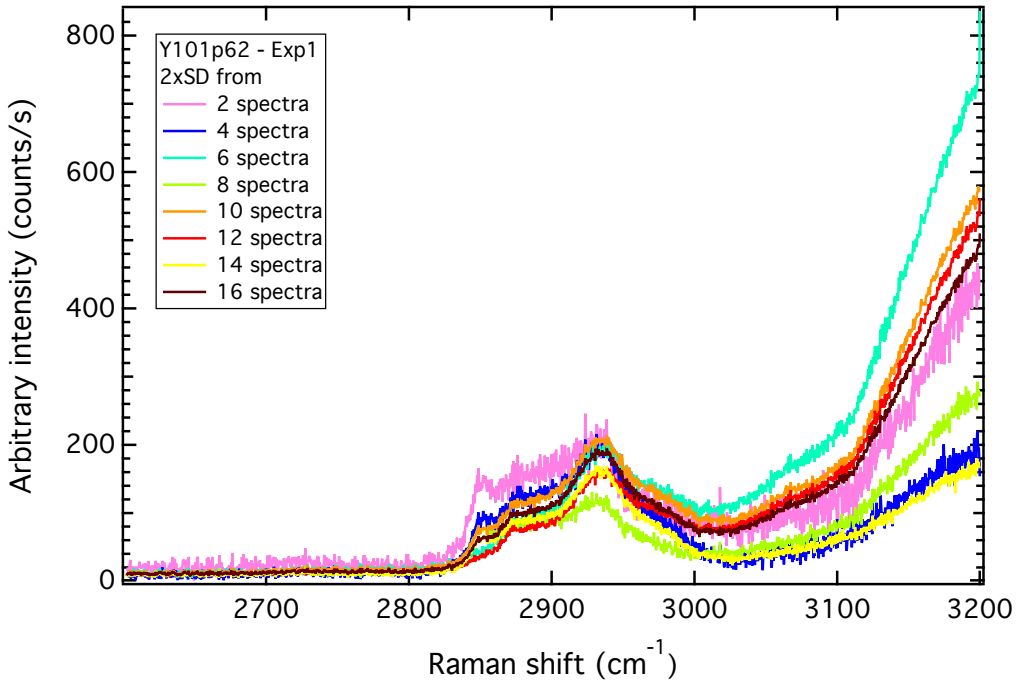
(b)

Figure B.55: SE stacks for live cell analyses of MSC line Y202 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

B.6 Live MSCs – 2x SD convergence tests

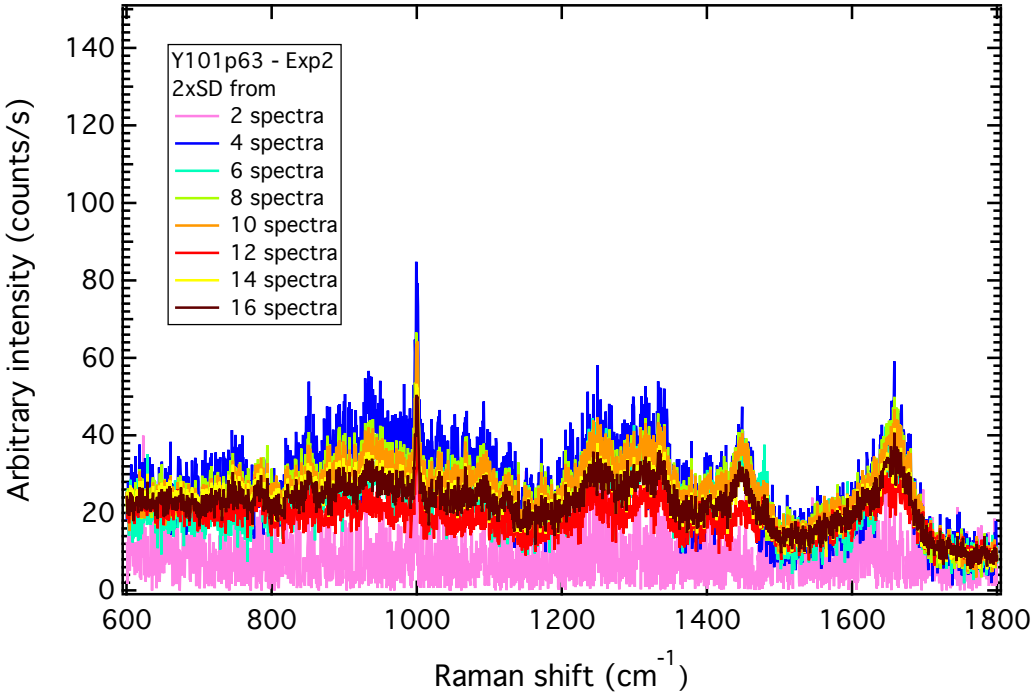


(a)

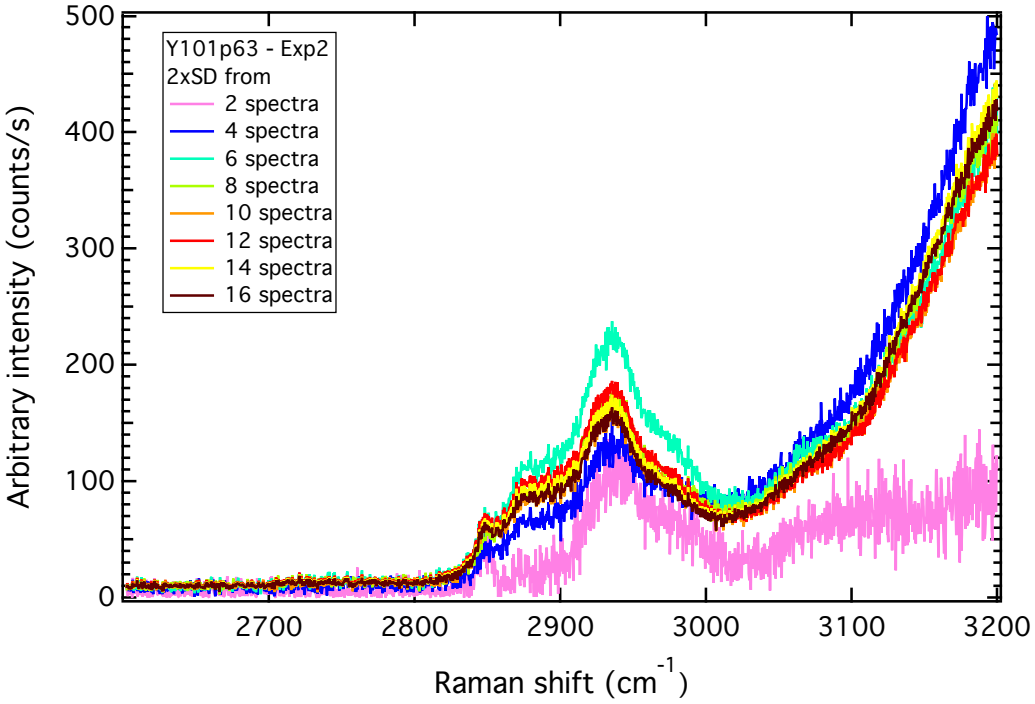


(b)

Figure B.56: 2x SD stacks for live cell analyses of MSC line Y101 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

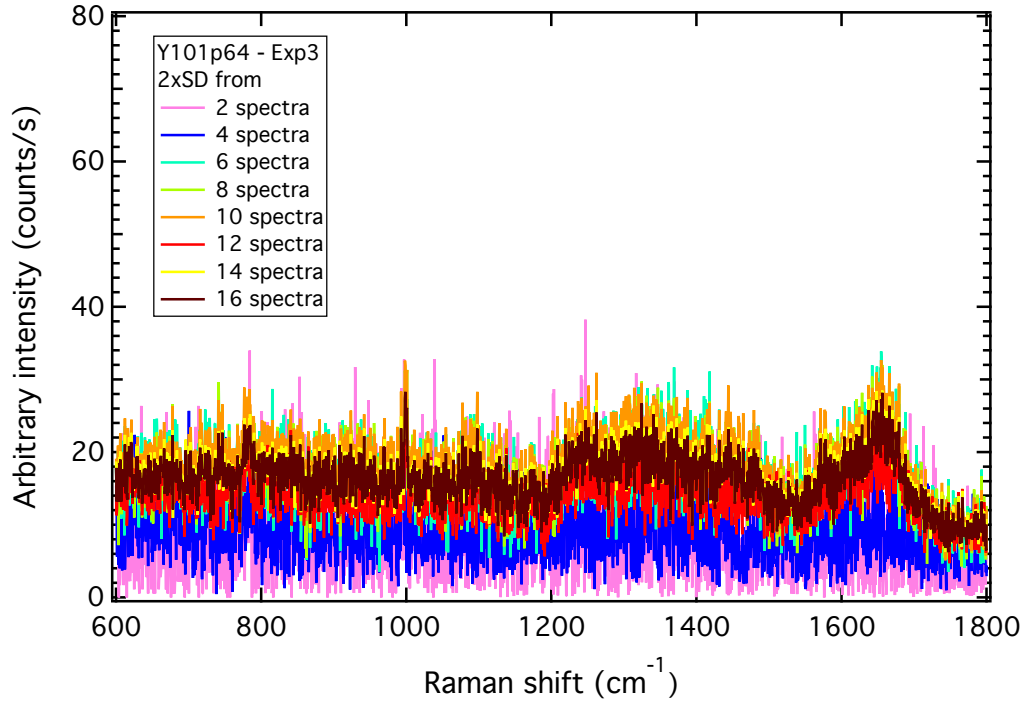


(a)

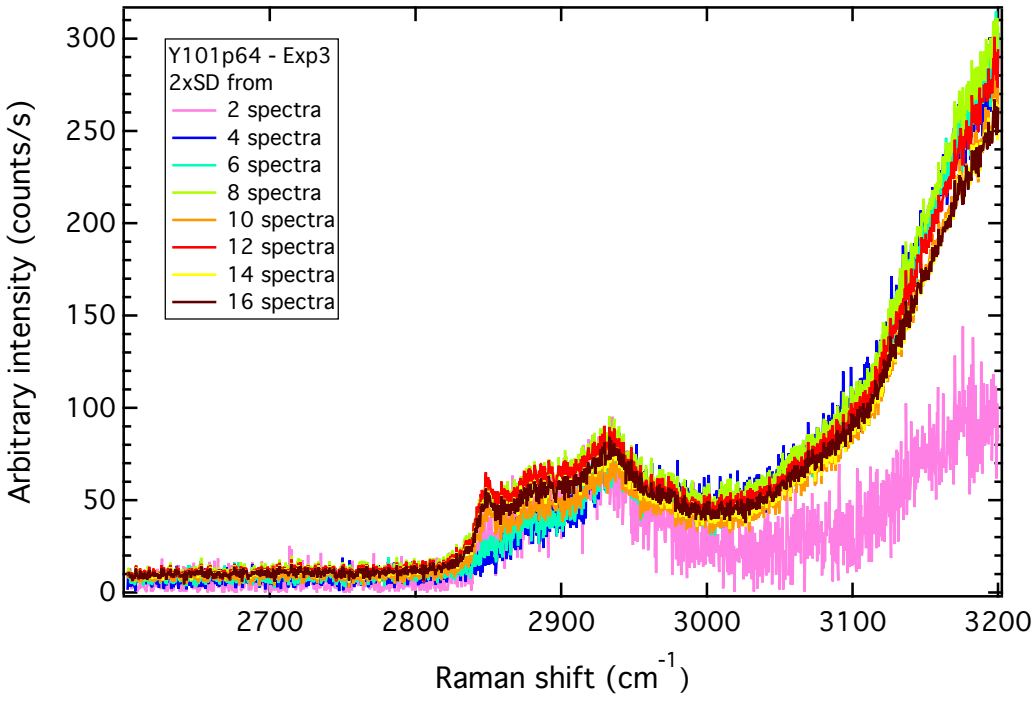


(b)

Figure B.57: 2x SD stacks for live cell analyses of MSC line Y101 experiment 2. (a) Fingerprint and (b) high wavenumber regions.

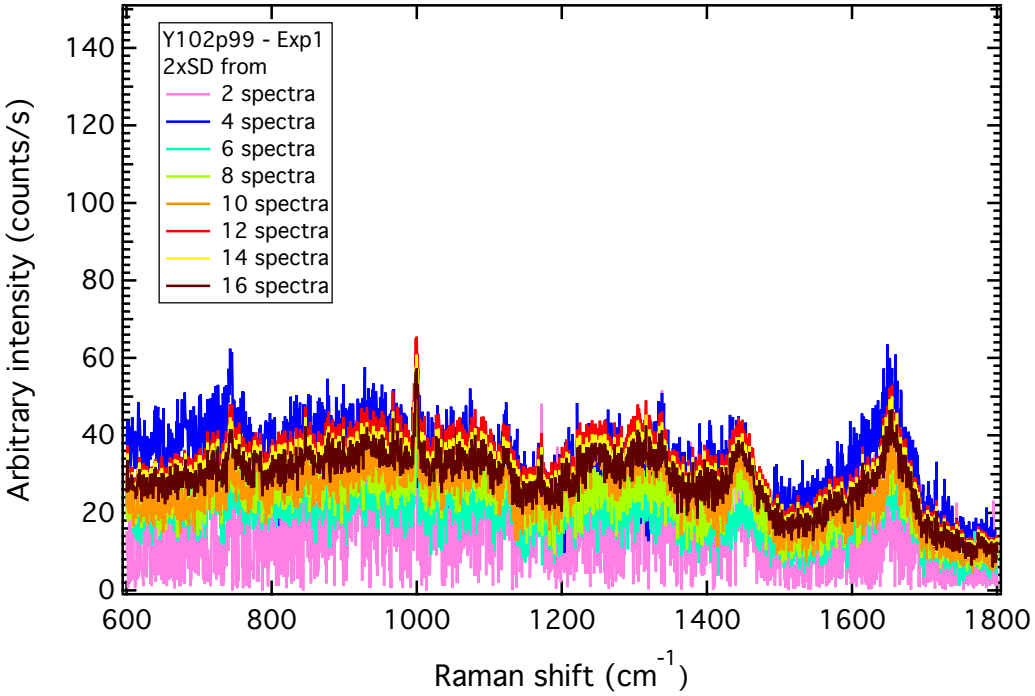


(a)

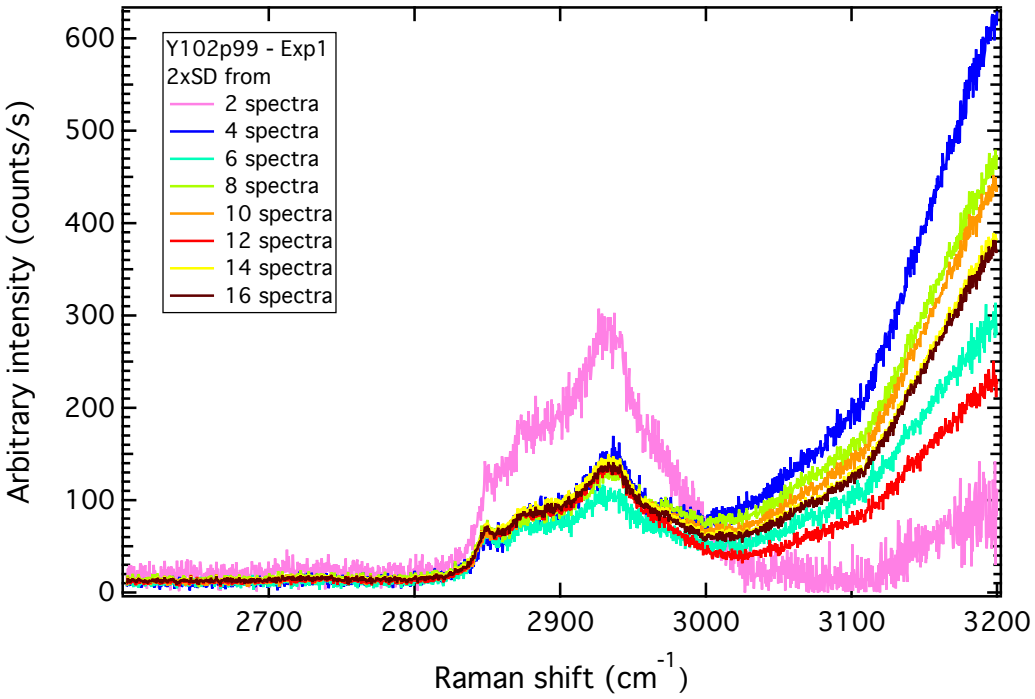


(b)

Figure B.58: 2x SD stacks for live cell analyses of MSC line Y101 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

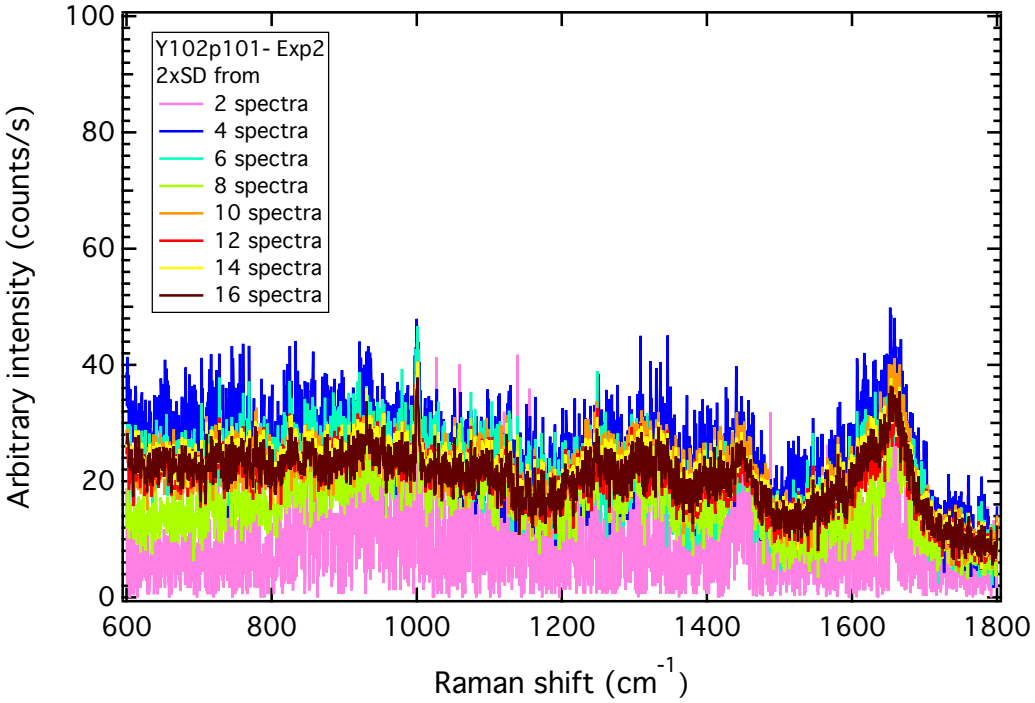


(a)

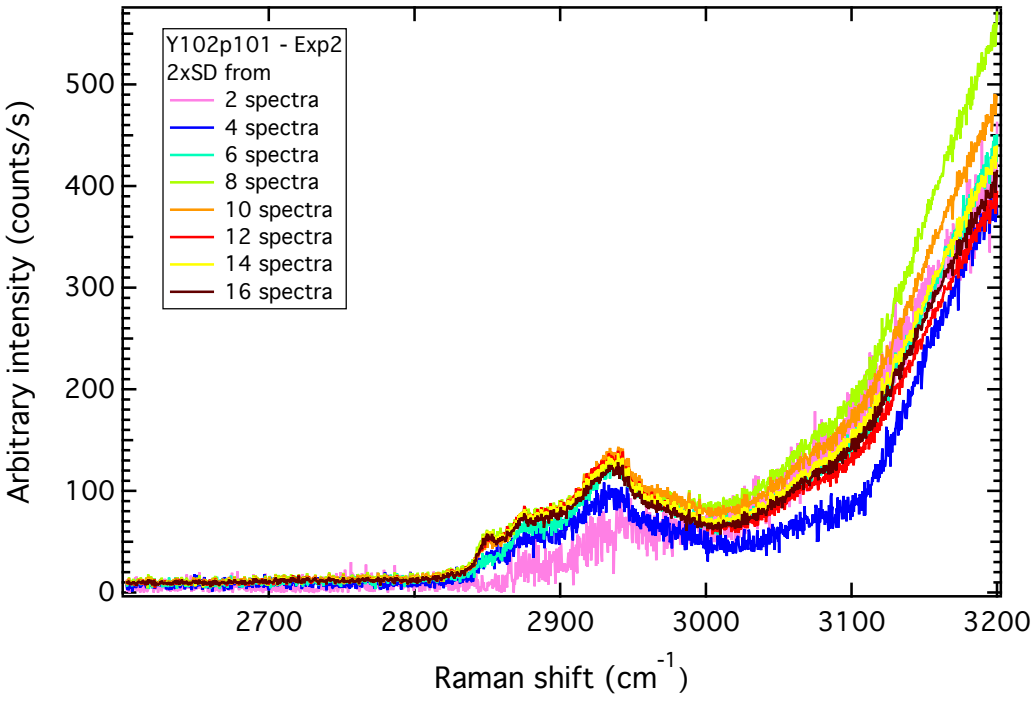


(b)

Figure B.59: 2x SD stacks for live cell analyses of MSC line Y102 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

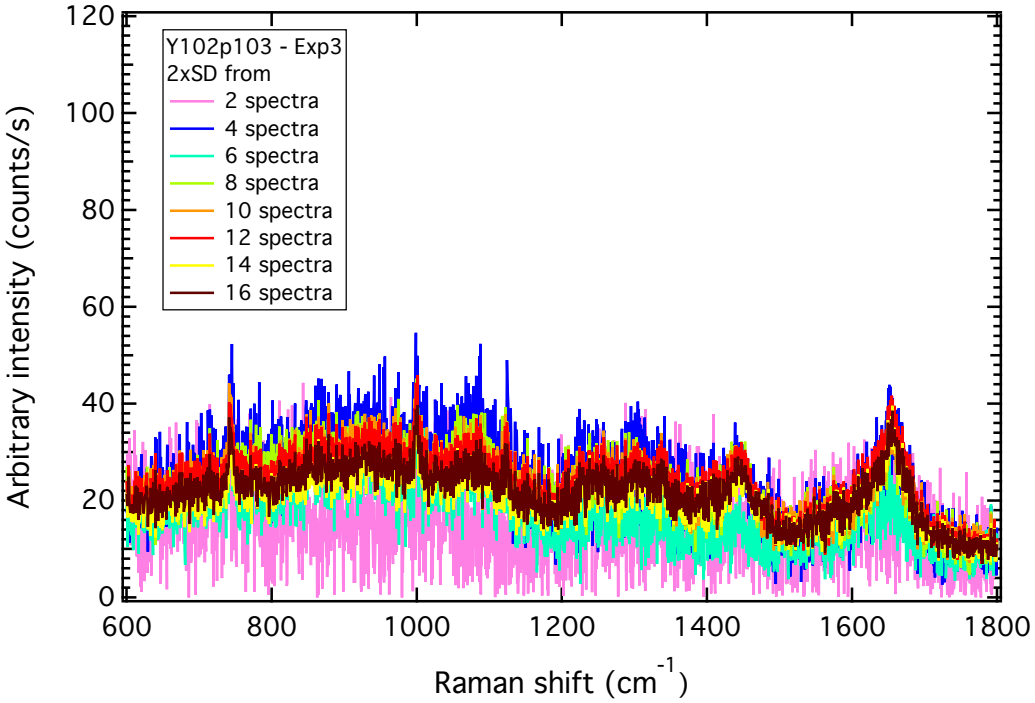


(a)

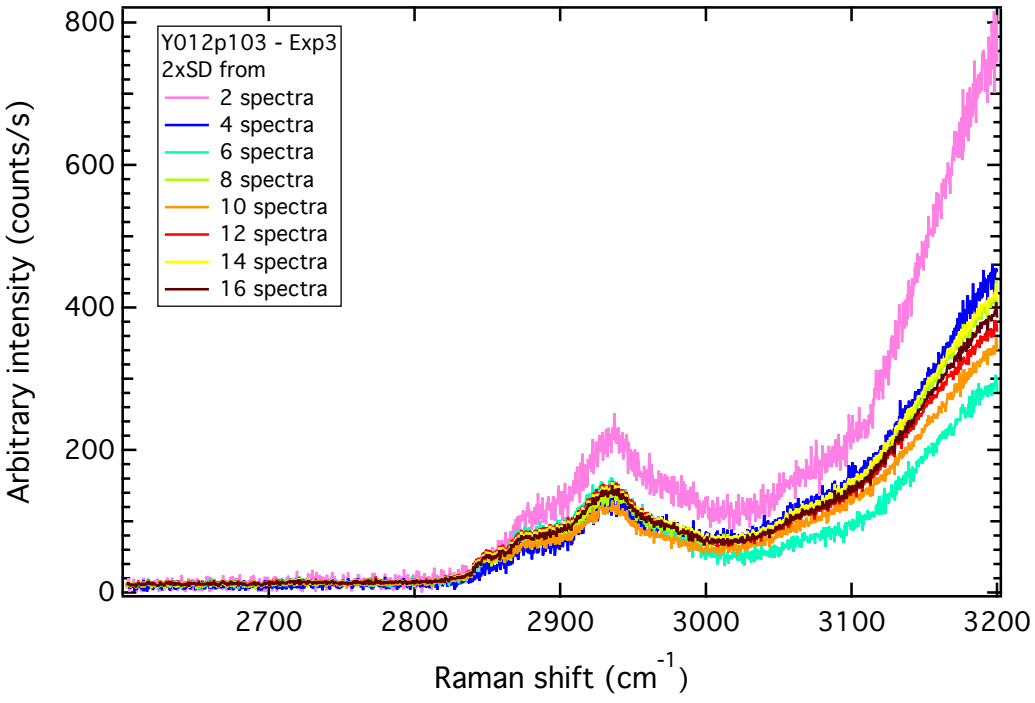


(b)

Figure B.60: 2x SD stacks for live cell analyses of MSC line Y102 experiment 2. (a) Fingerprint and (b) high wavenumber regions.

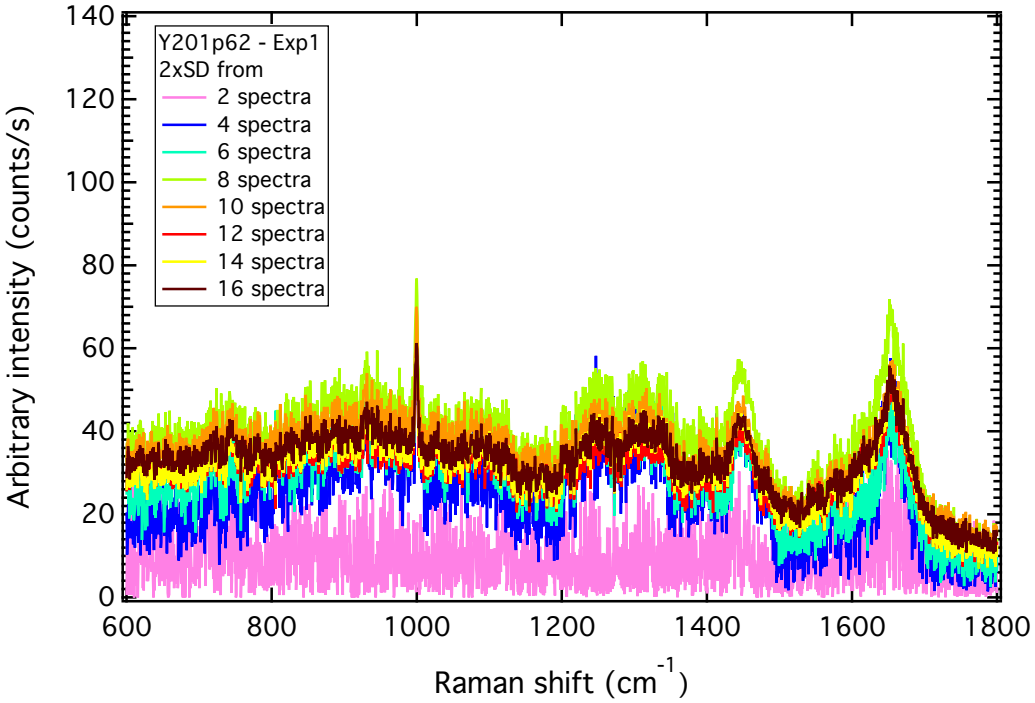


(a)

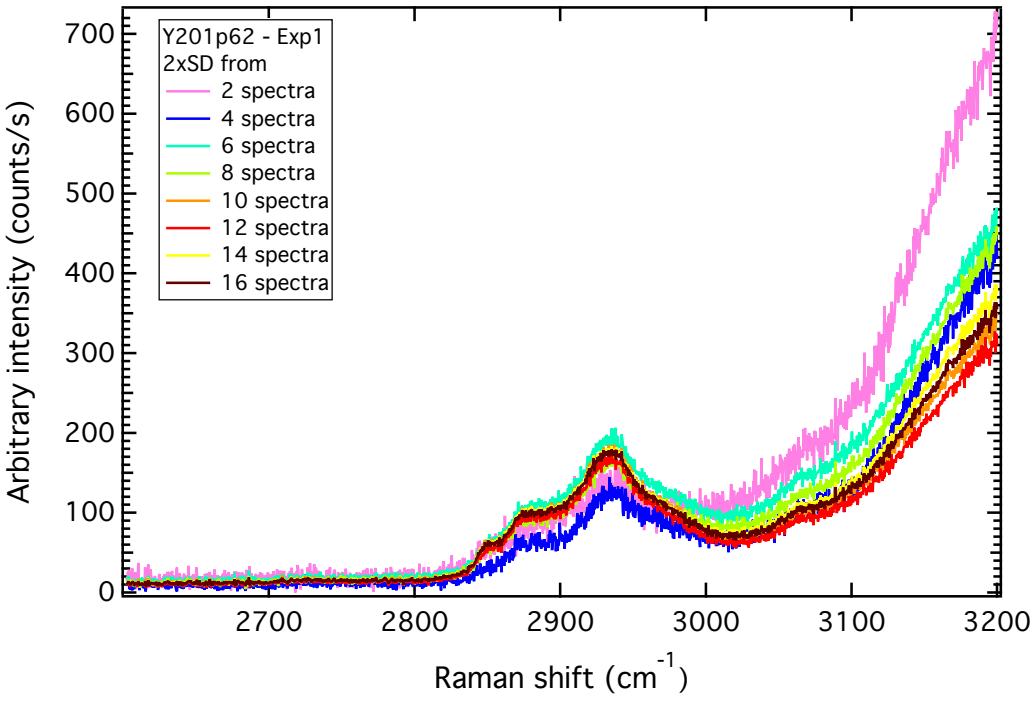


(b)

Figure B.61: 2x SD stacks for live cell analyses of MSC line Y102 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

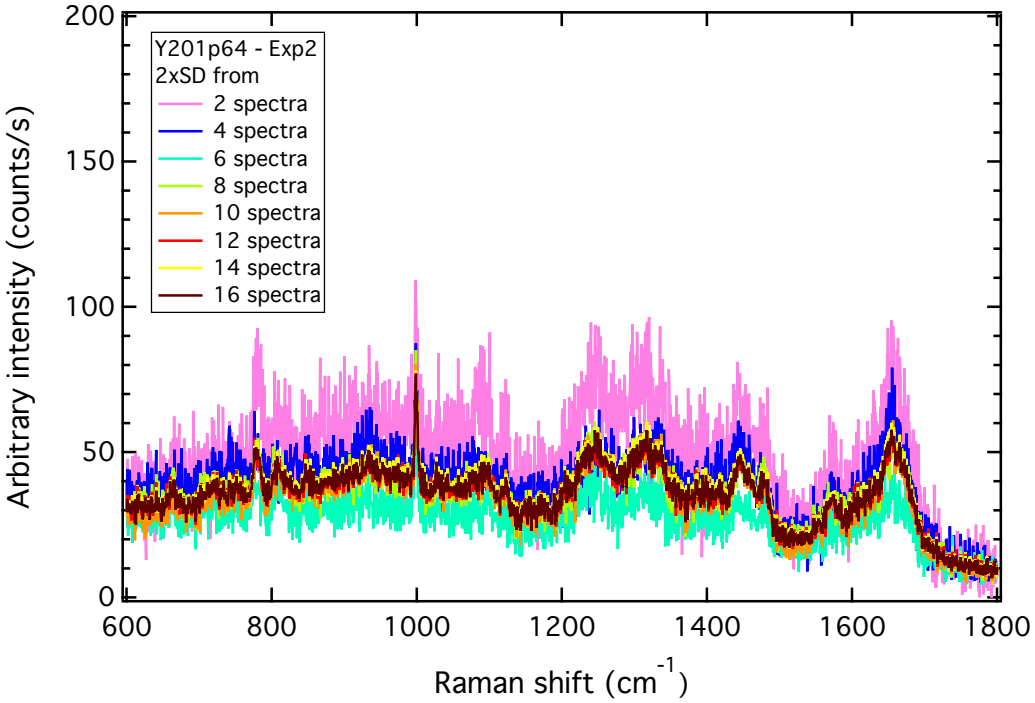


(a)

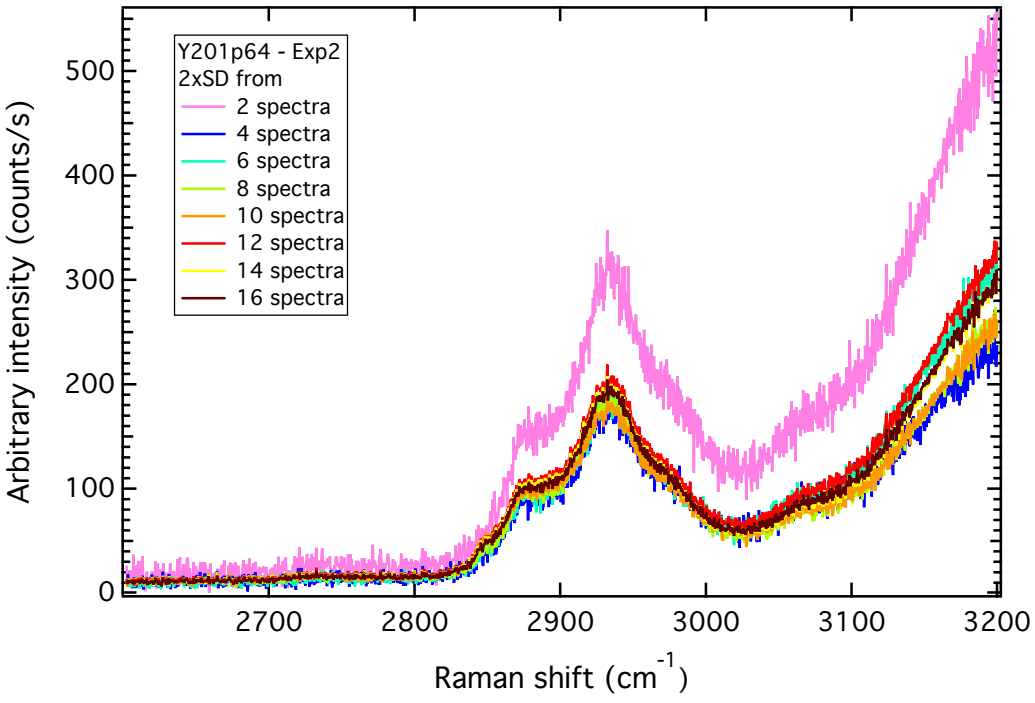


(b)

Figure B.62: 2x SD stacks for live cell analyses of MSC line Y201 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

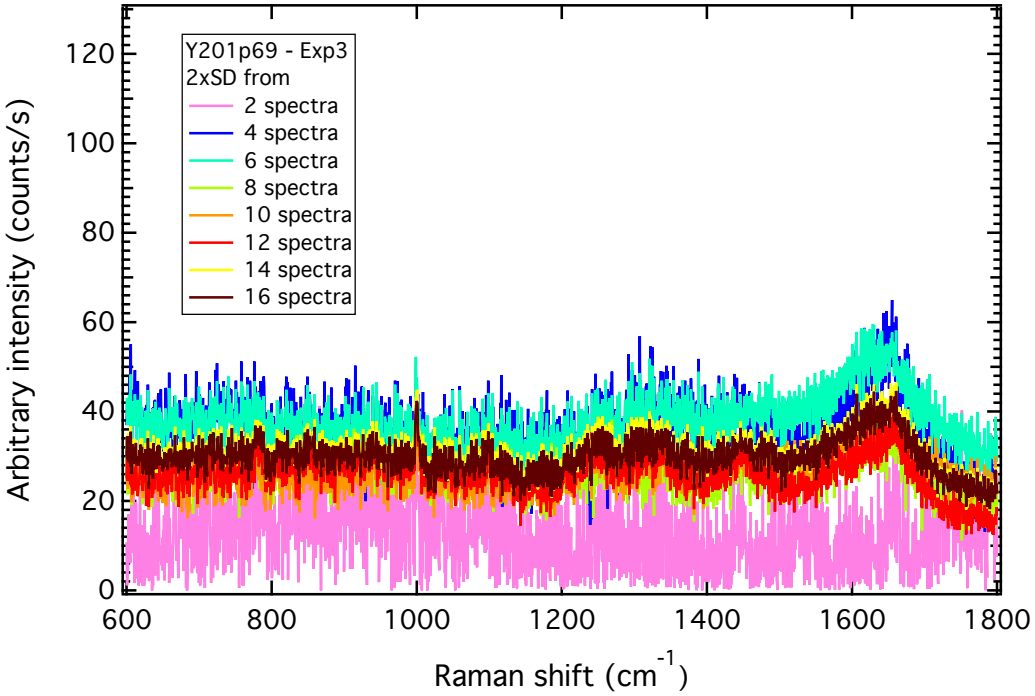


(a)

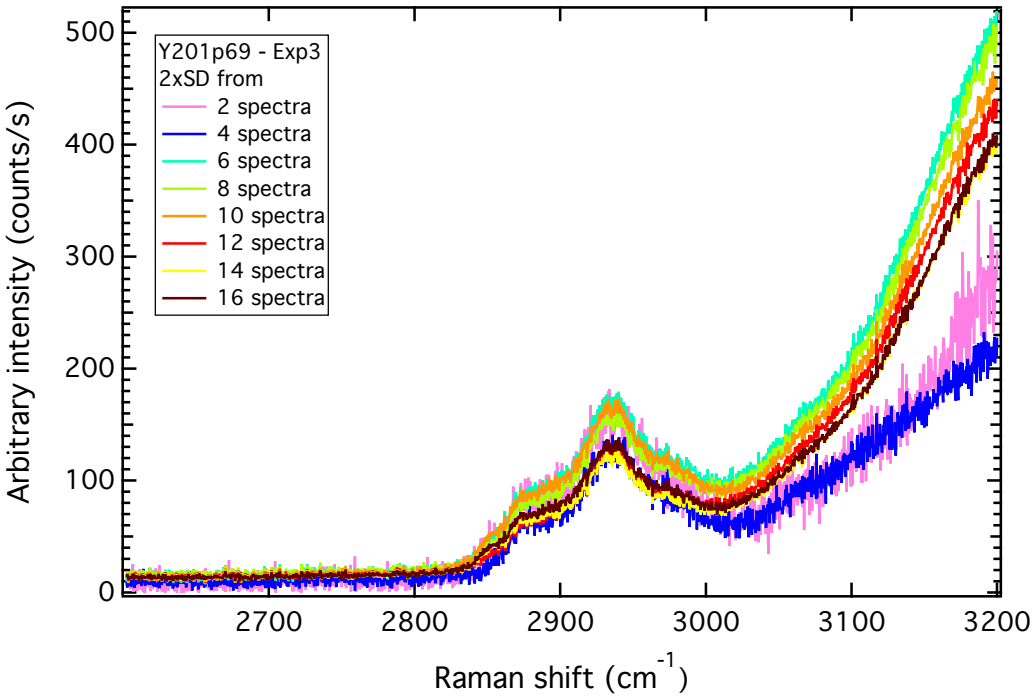


(b)

Figure B.63: 2x SD stacks for live cell analyses of MSC line Y201 experiment 2. (a) Fingerprint and (b) high wavenumber regions.

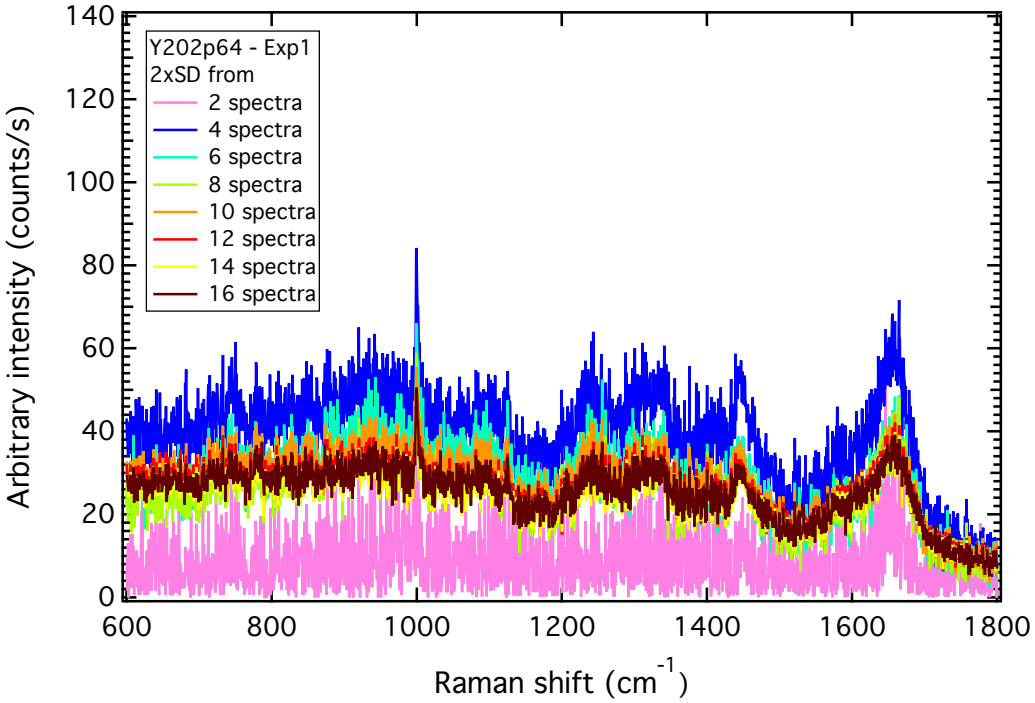


(a)

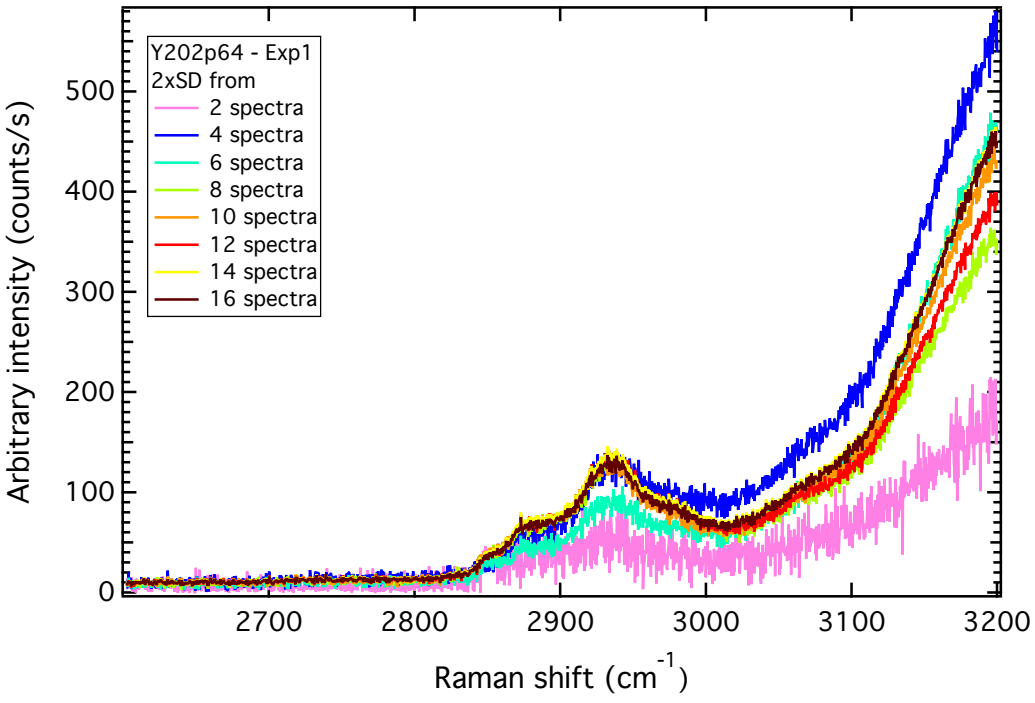


(b)

Figure B.64: 2x SD stacks for live cell analyses of MSC line Y201 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

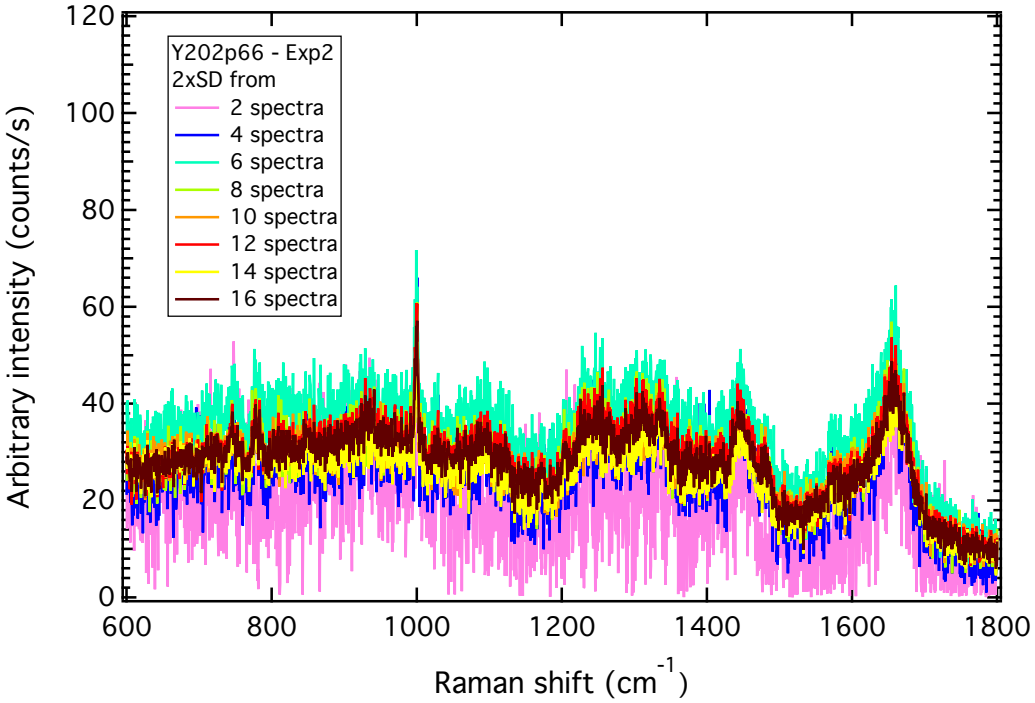


(a)

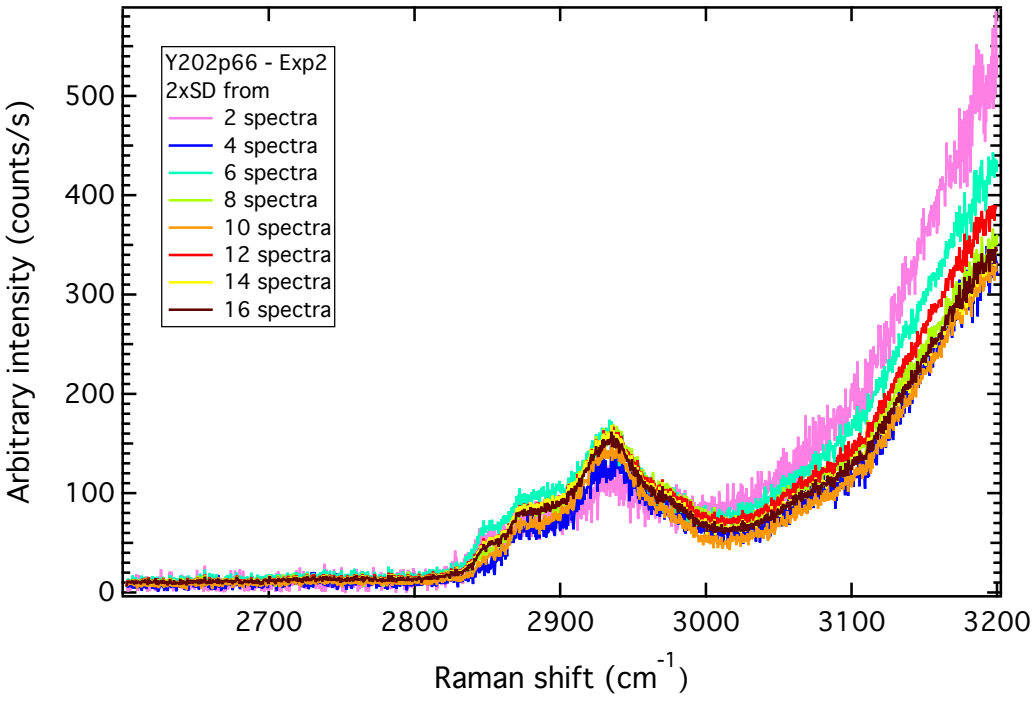


(b)

Figure B.65: 2x SD stacks for live cell analyses of MSC line Y202 experiment 1. (a) Fingerprint and (b) high wavenumber regions.

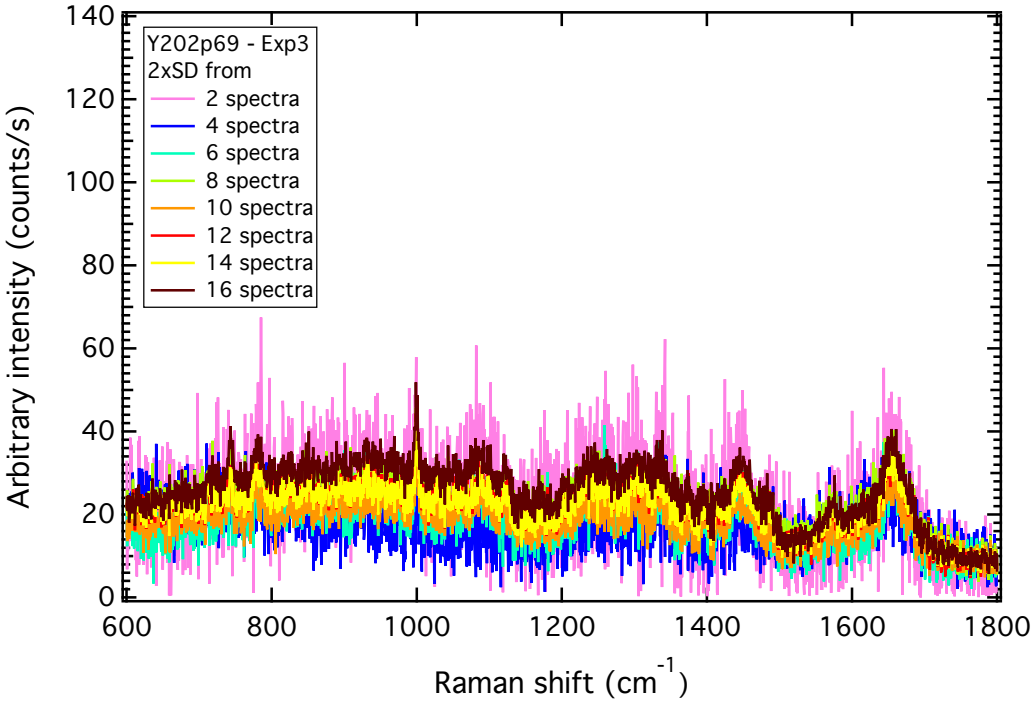


(a)

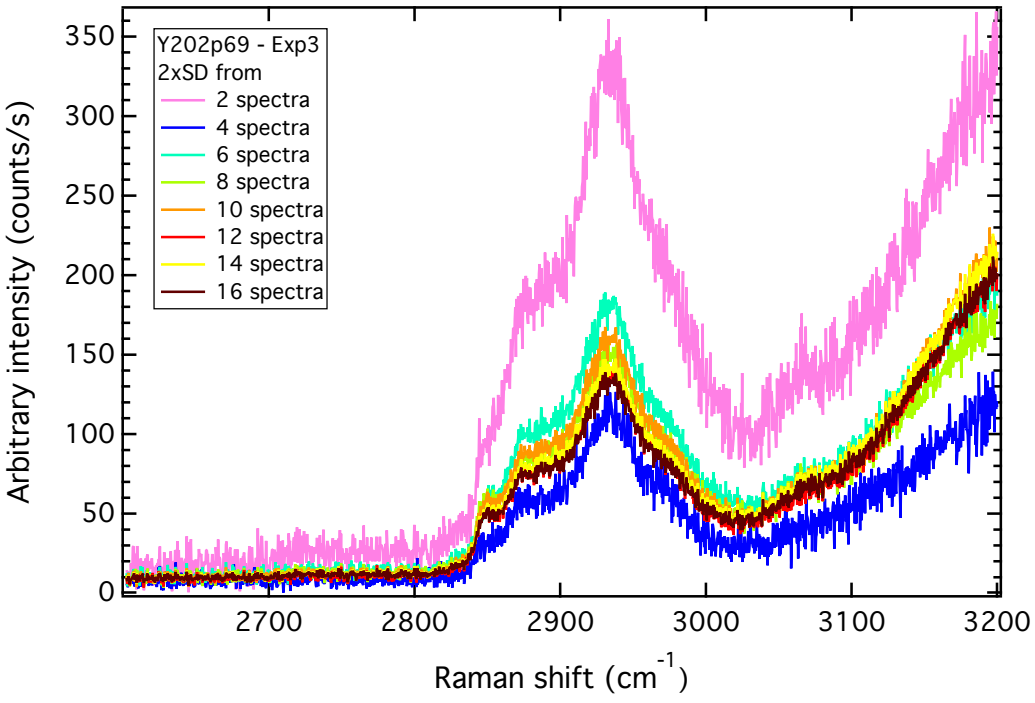


(b)

Figure B.66: 2x SD stacks for live cell analyses of MSC line Y202 experiment 2. (a) Fingerprint and (b) high wavenumber regions.



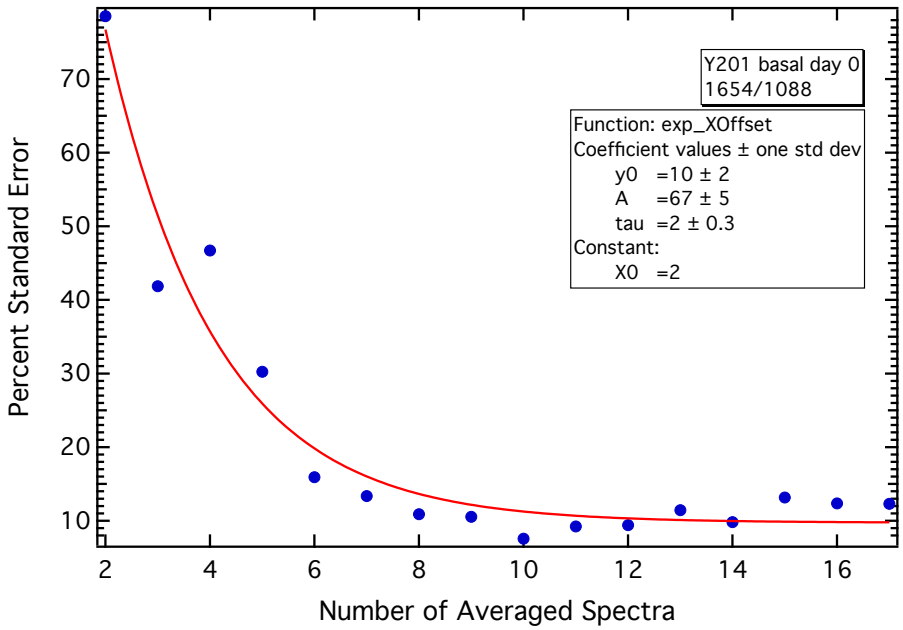
(a)



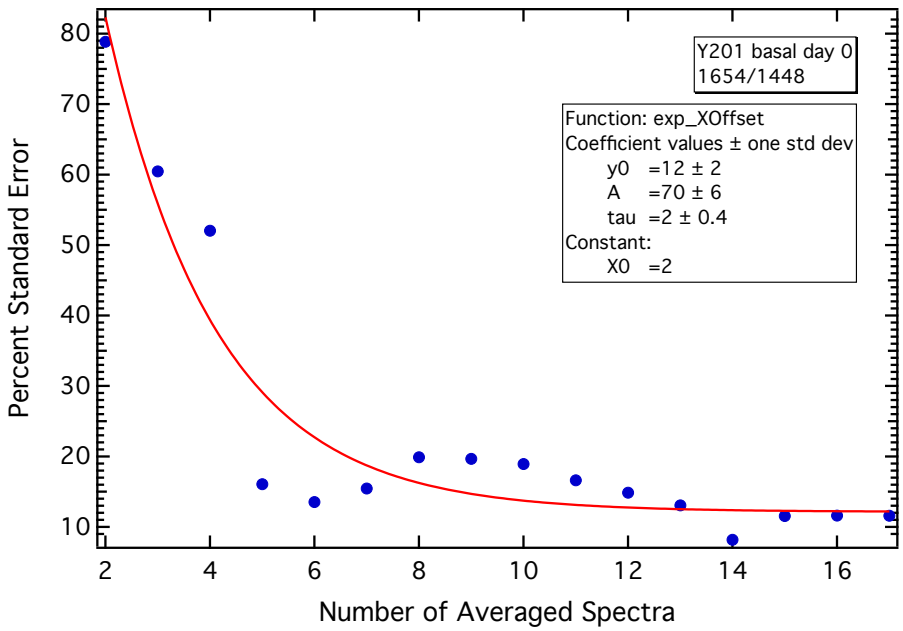
(b)

Figure B.67: 2x SD stacks for live cell analyses of MSC line Y202 experiment 3. (a) Fingerprint and (b) high wavenumber regions.

B.7 Y201 MSC line in basal medium – %SE convergence tests

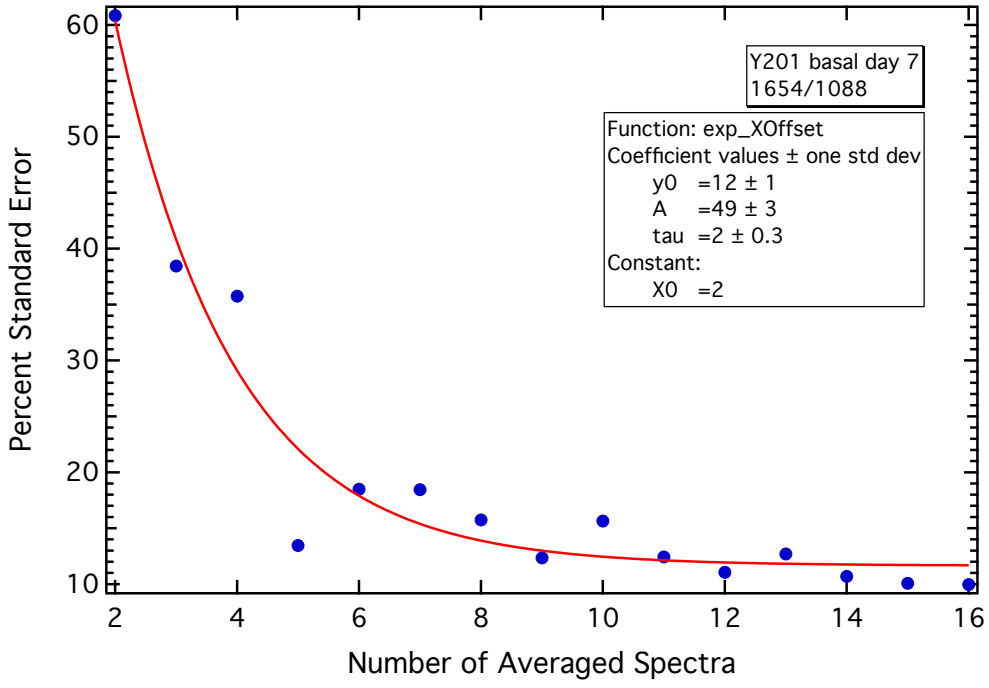


(a)

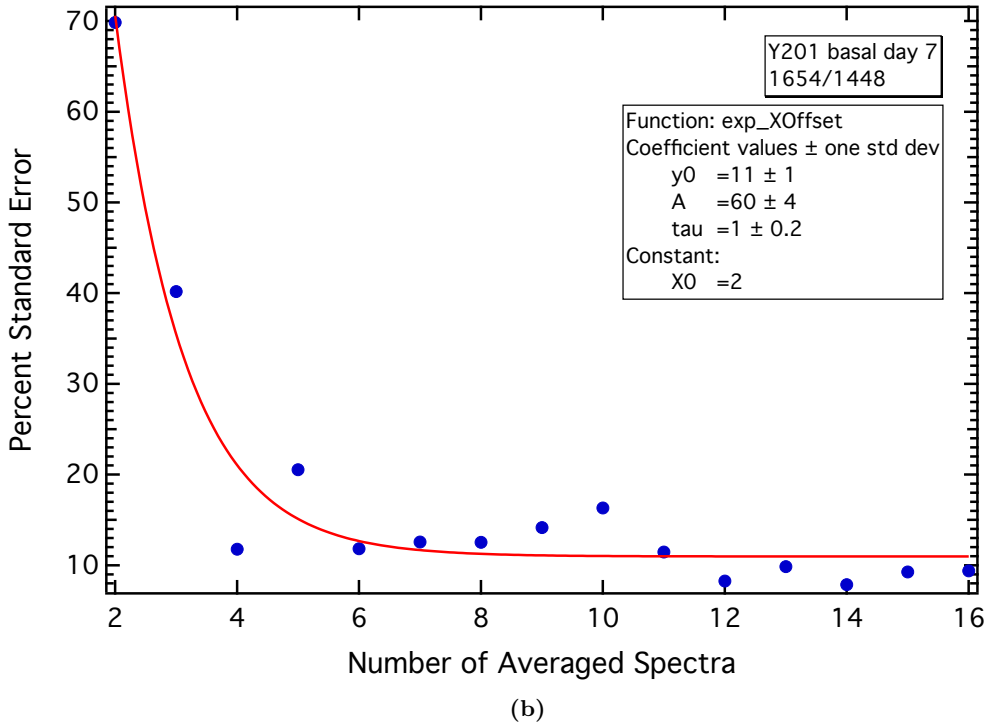


(b)

Figure B.68: %SE convergence tests for MSC line Y201 in basal medium at day 0 for PIRs (a) 1654/1088 and (b) 1654/1448.

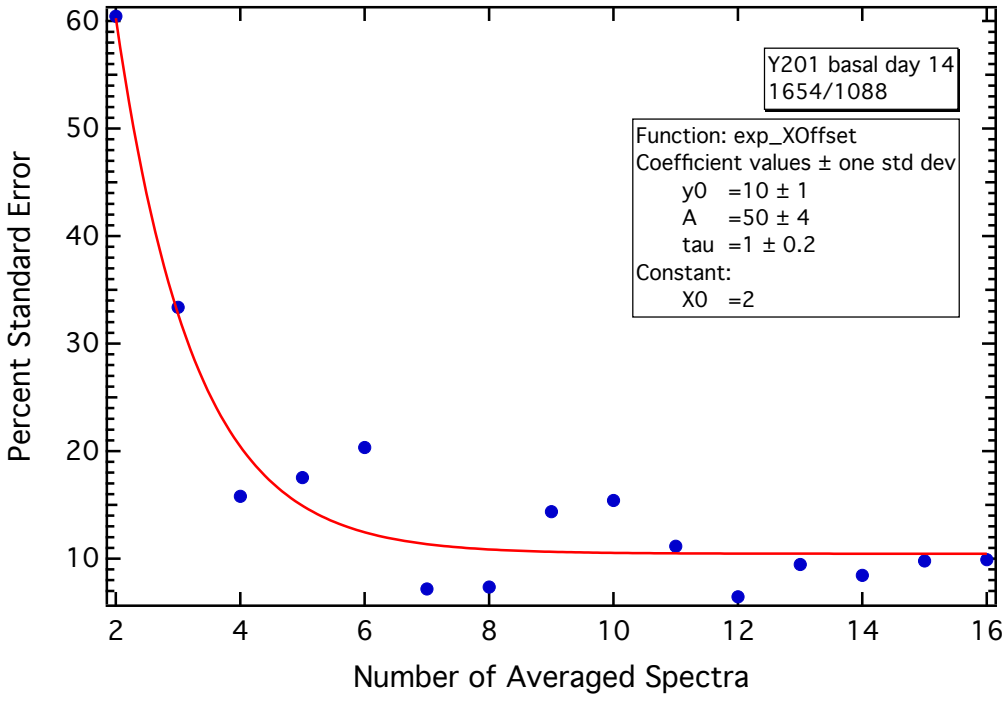


(a)

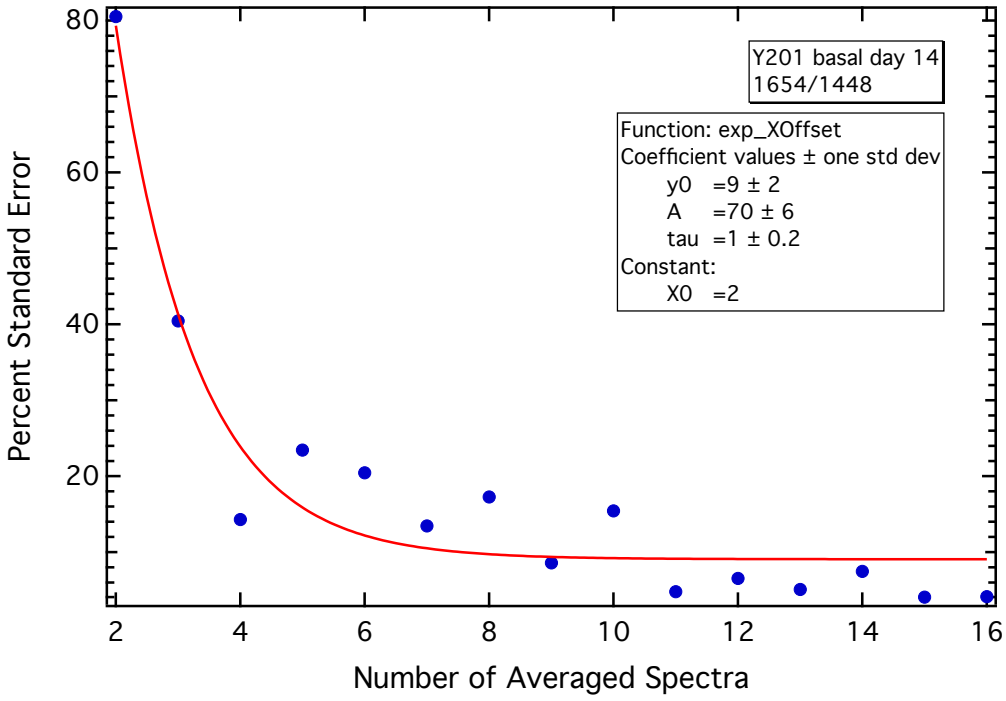


(b)

Figure B.69: %SE convergence tests for MSC line Y201 in basal medium at day 7 for PIRs (a) 1654/1088 and (b) 1654/1448.

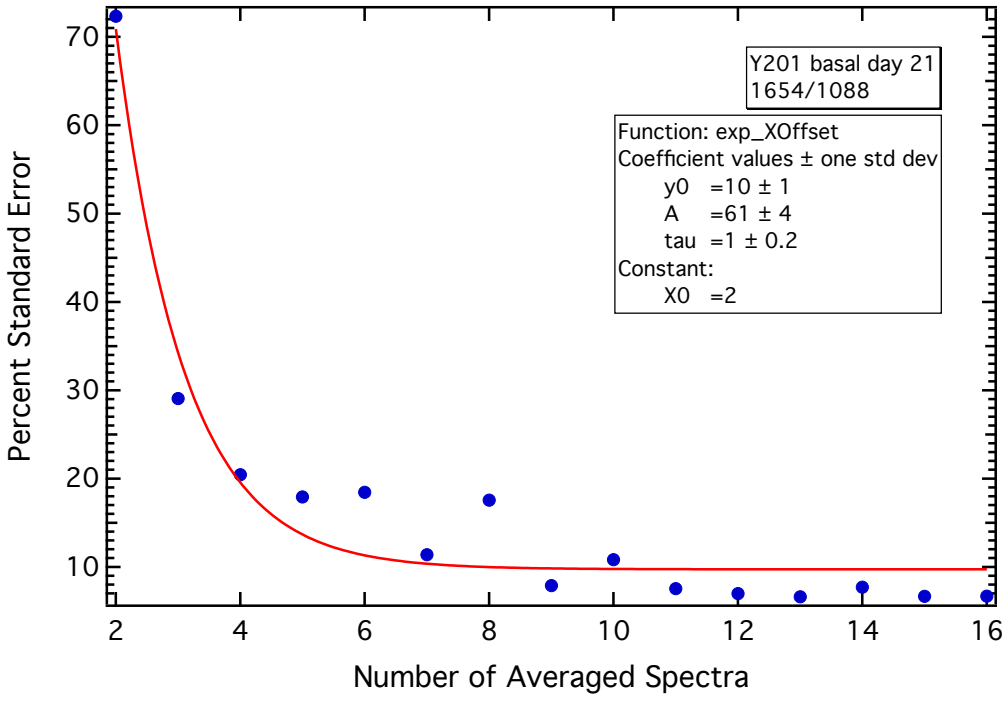


(a)

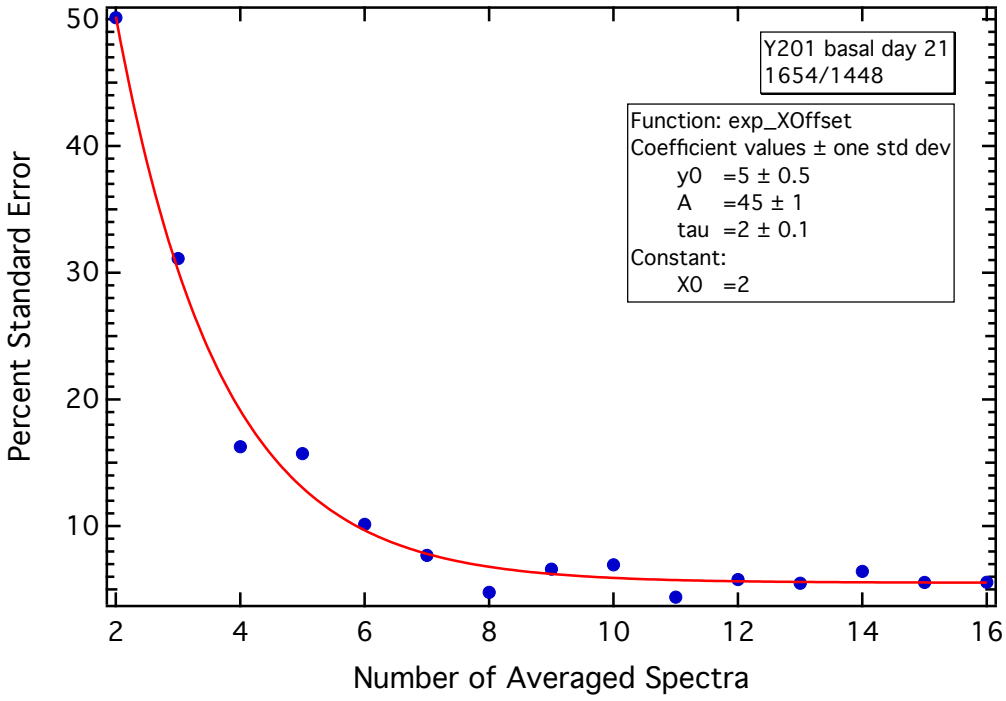


(b)

Figure B.70: %SE convergence tests for MSC line Y201 in basal medium at day 14 for PIRs (a) 1654/1088 and (b) 1654/1448.



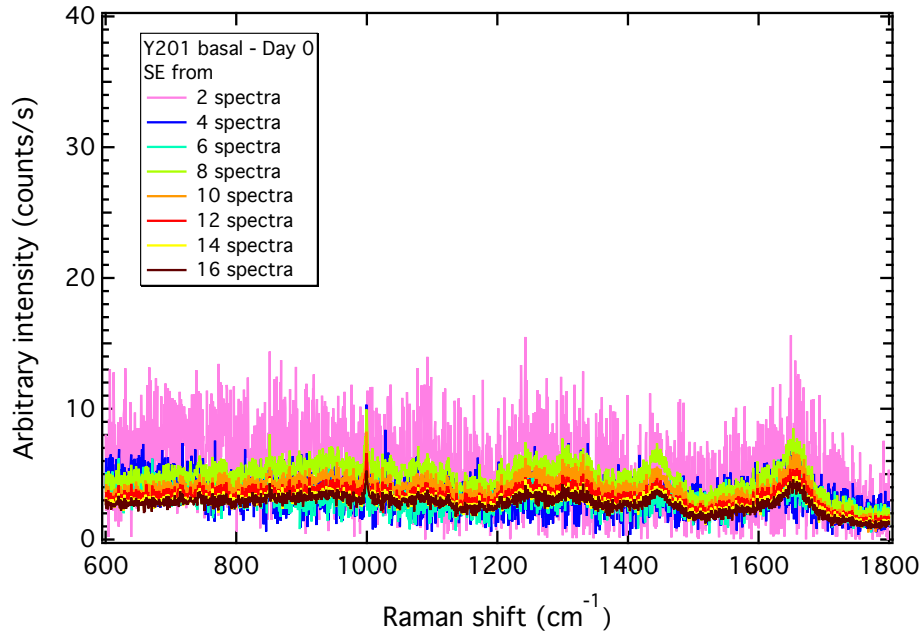
(a)



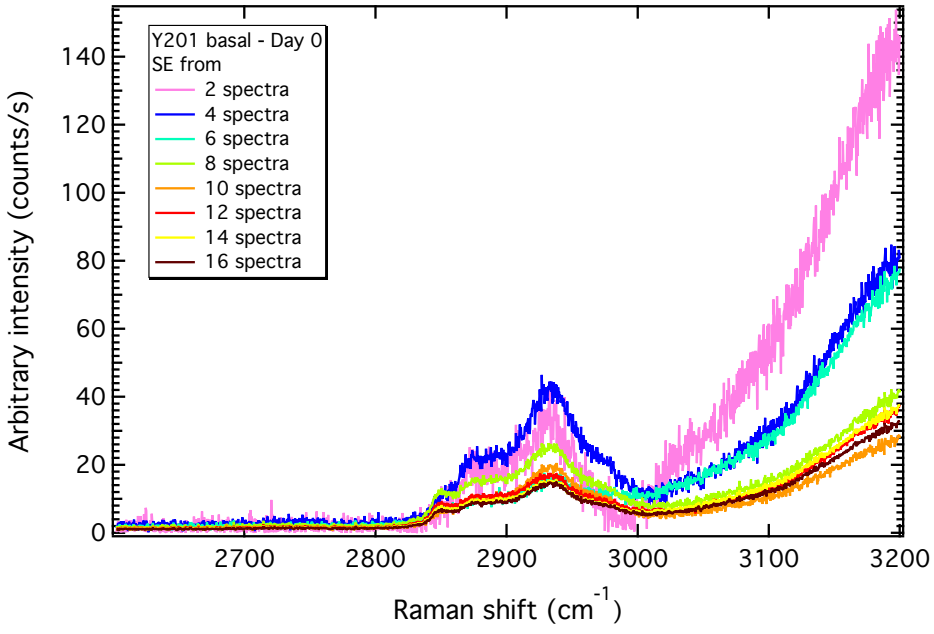
(b)

Figure B.71: %SE convergence tests for MSC line Y201 in basal medium at day 21 for PIRs (a) 1654/1088 and (b) 1654/1448.

B.8 Y201 MSC line in basal medium – SE convergence tests

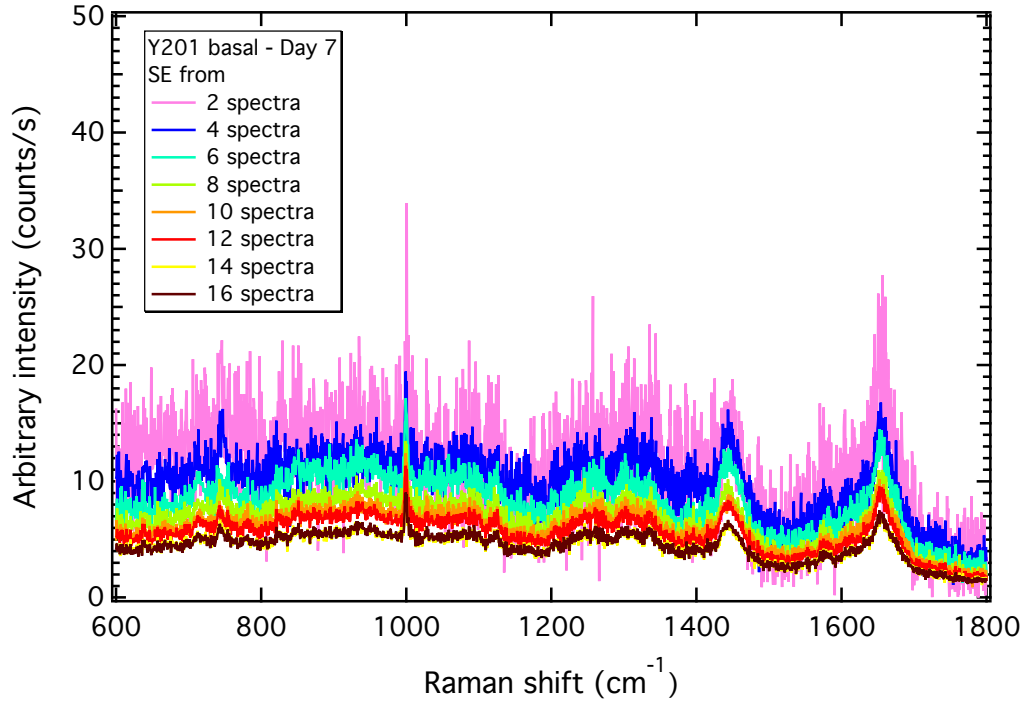


(a)

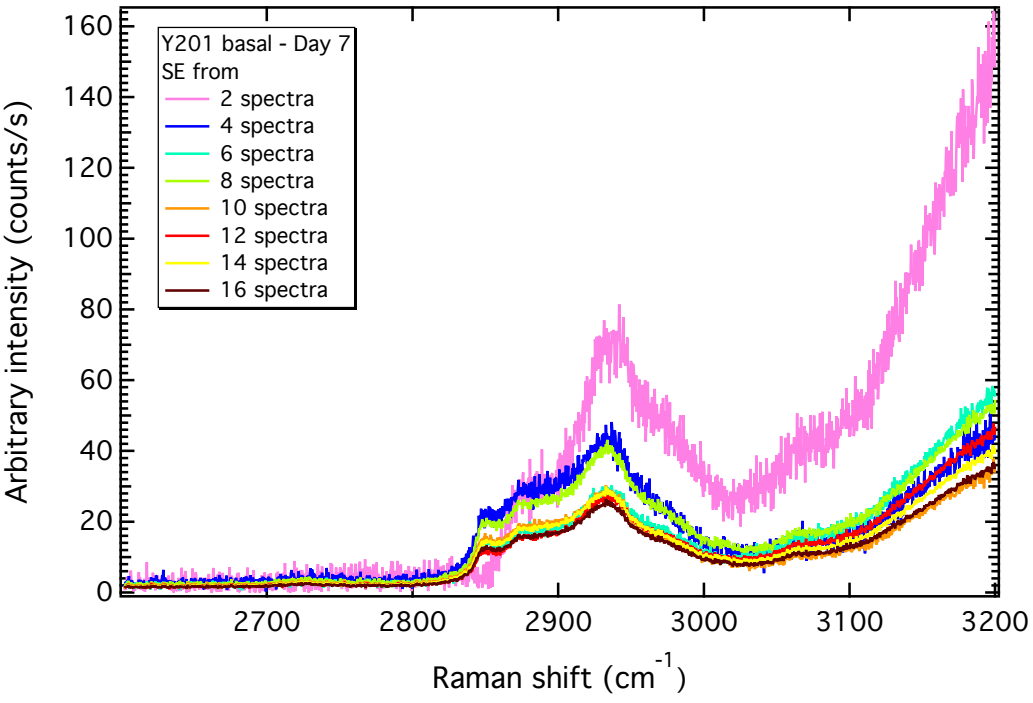


(b)

Figure B.72: SE stacks for Y201 in basal medium at day 0. (a) fingerprint and (b) high wavenumber regions.

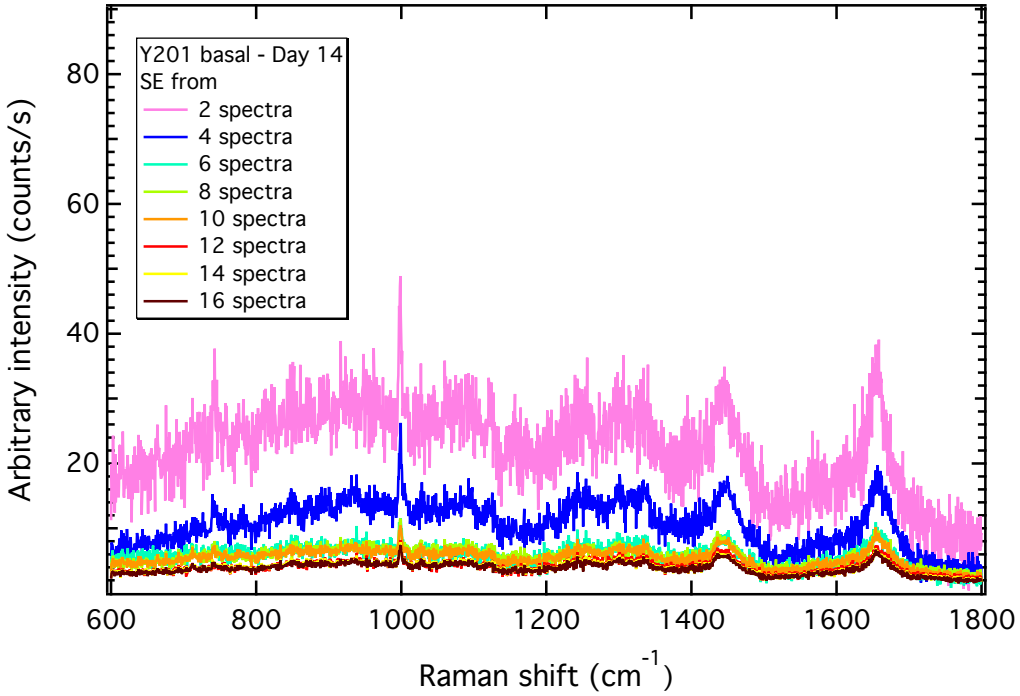


(a)

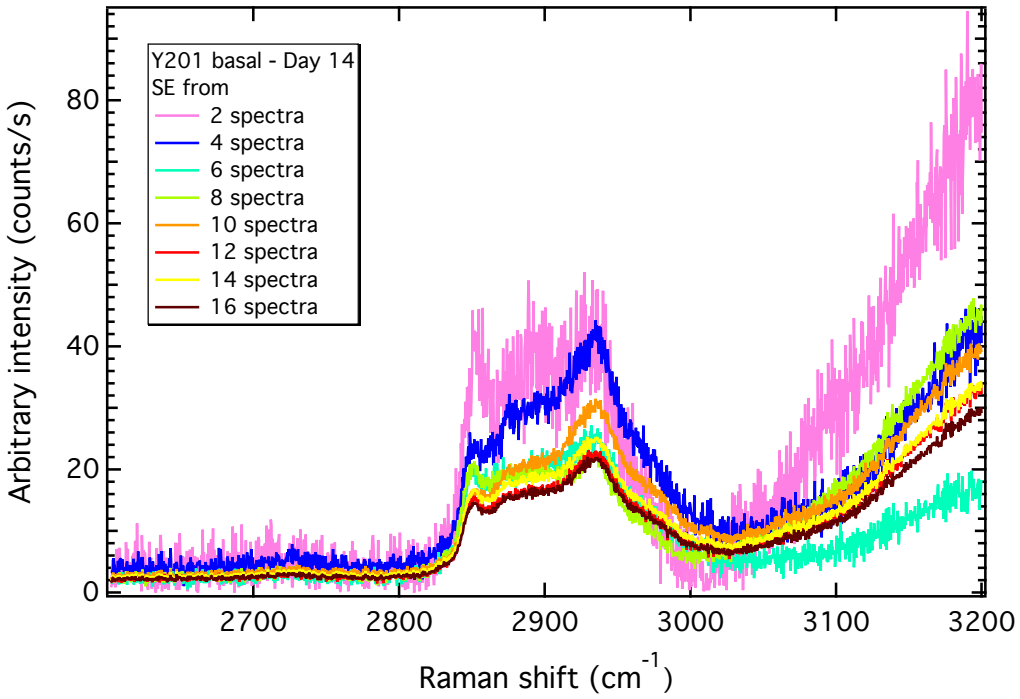


(b)

Figure B.73: SE stacks for Y201 in basal medium at day 7. (a) fingerprint and (b) high wavenumber regions.

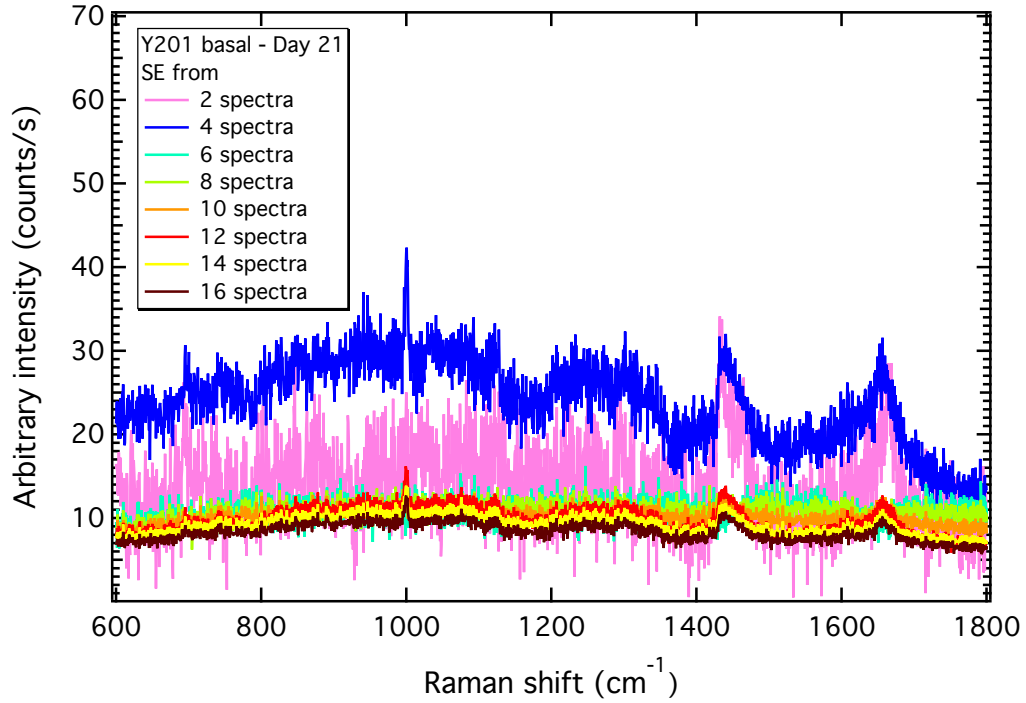


(a)

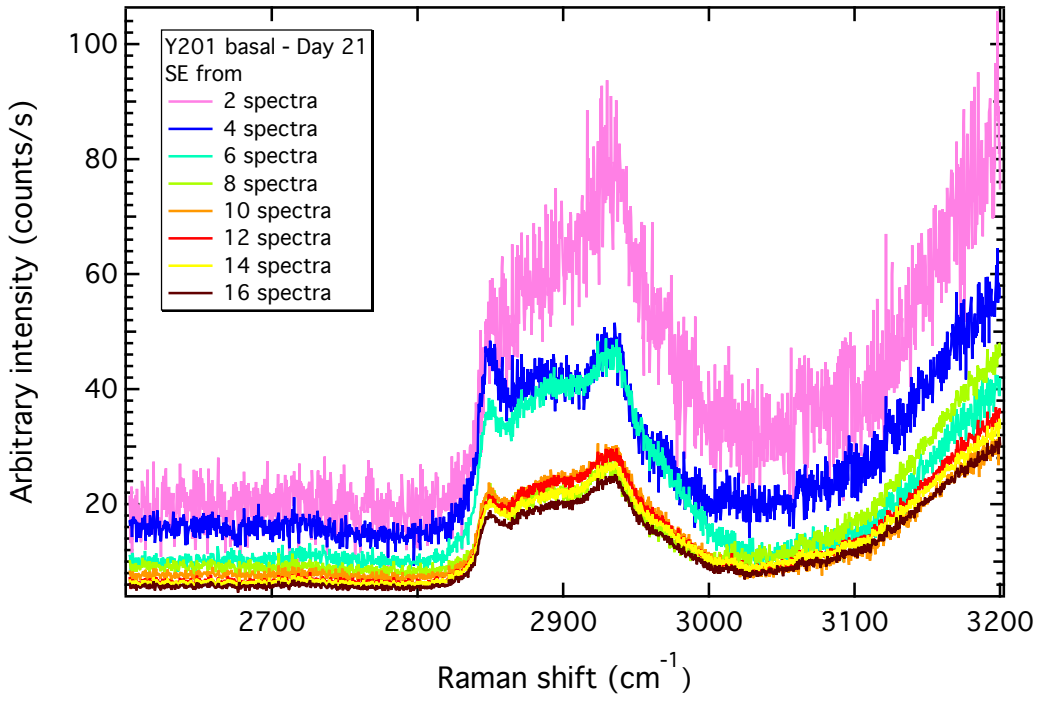


(b)

Figure B.74: SE stacks for Y201 in basal medium at day 14. (a) fingerprint and (b) high wavenumber regions.



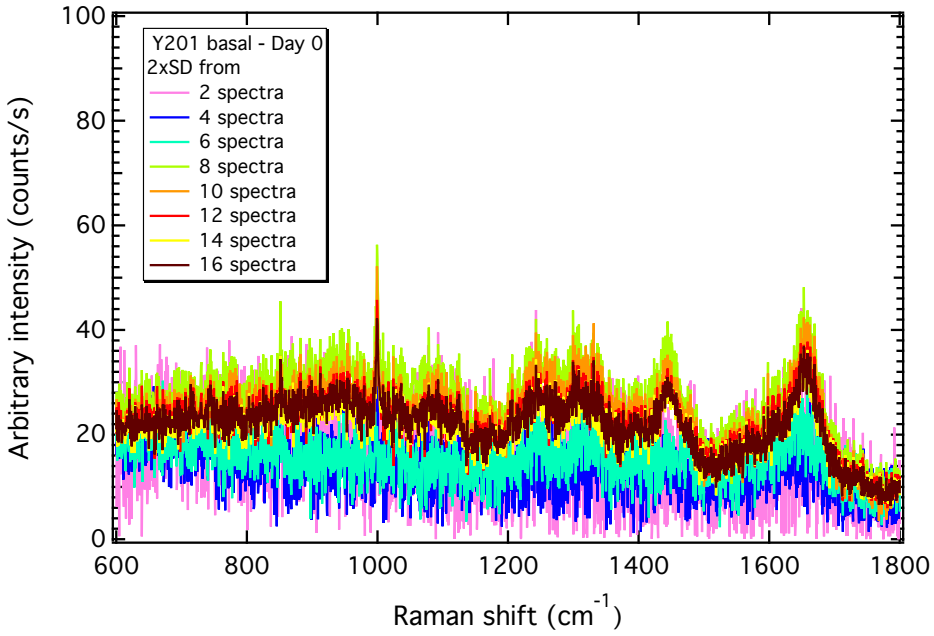
(a)



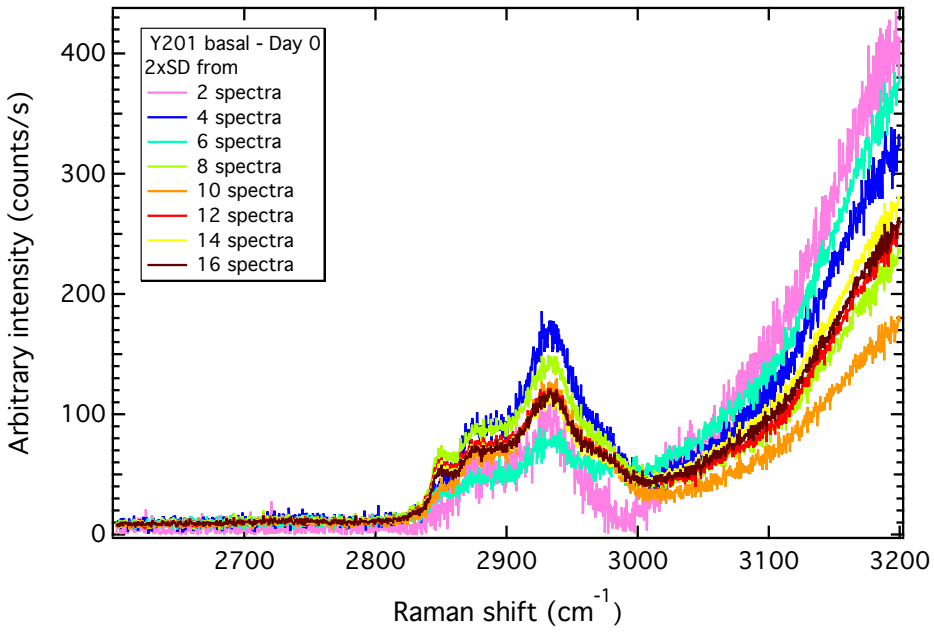
(b)

Figure B.75: SE stacks for Y201 in basal medium at day 21. (a) fingerprint and (b) high wavenumber regions.

B.9 Y201 MSC line in basal medium – 2x SD convergence tests

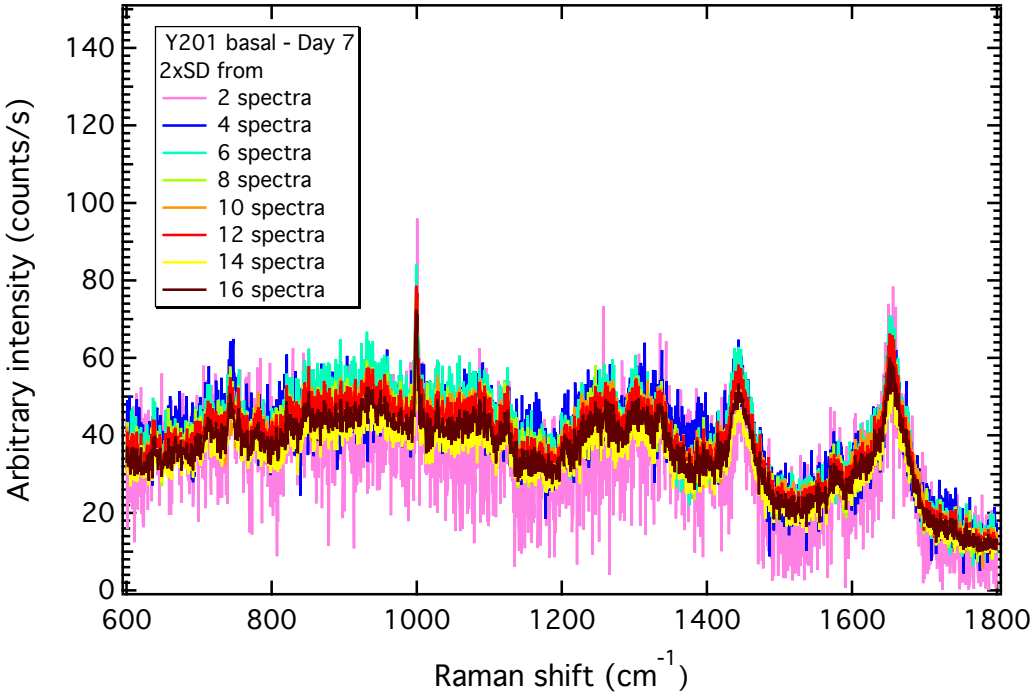


(a)

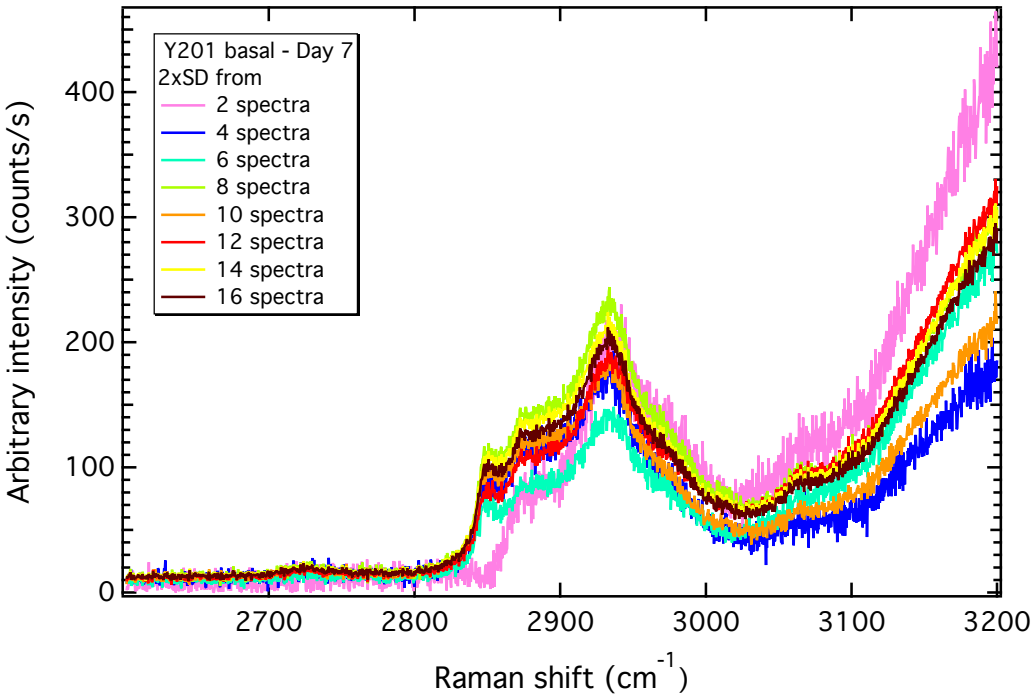


(b)

Figure B.76: 2x SD stacks for Y201 in basal medium at day 0. (a) fingerprint and (b) high wavenumber regions.

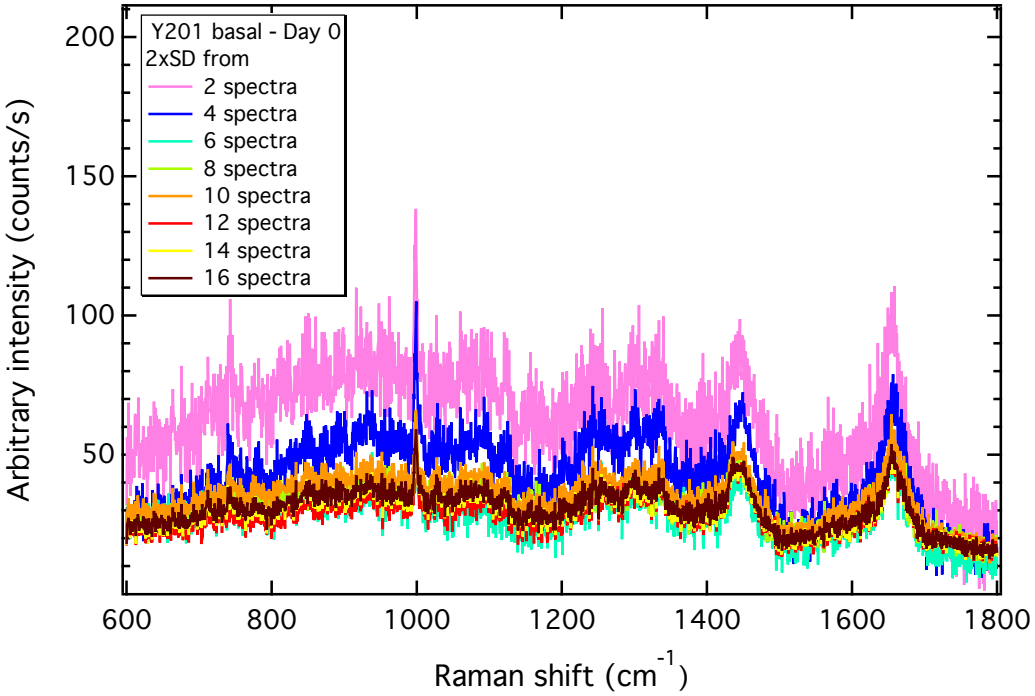


(a)

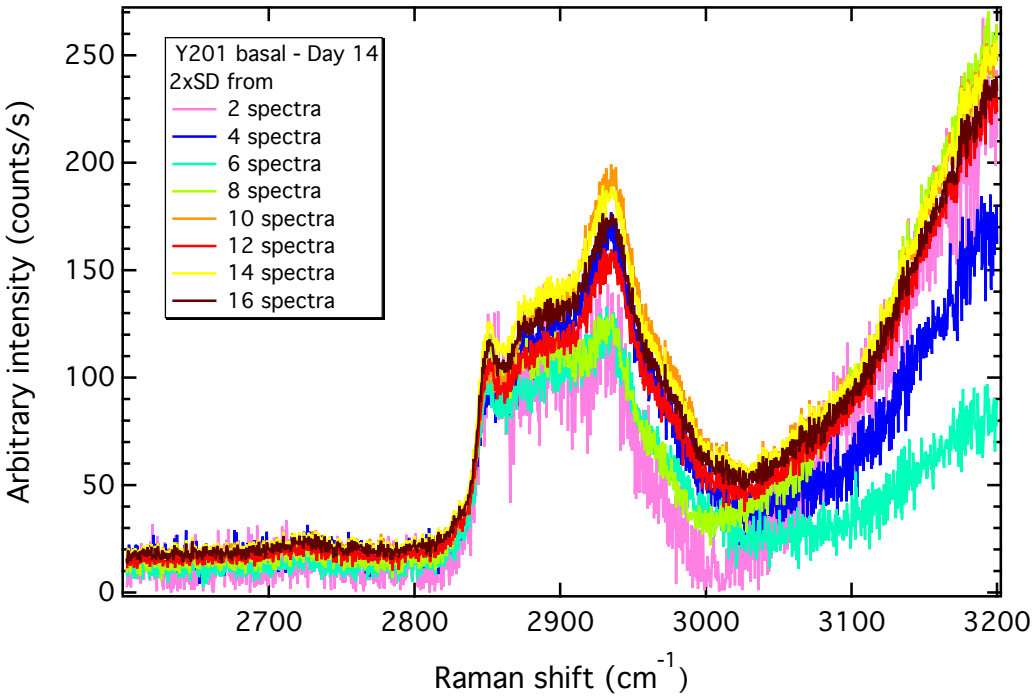


(b)

Figure B.77: 2x SD stacks for Y201 in basal medium at day 7. (a) fingerprint and (b) high wavenumber regions.

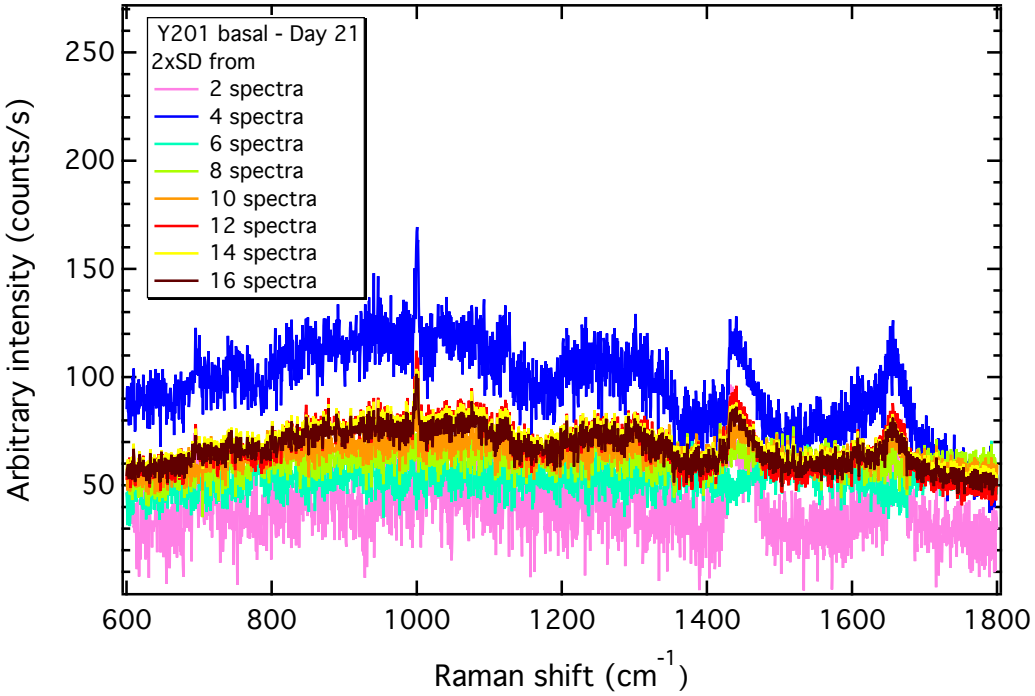


(a)

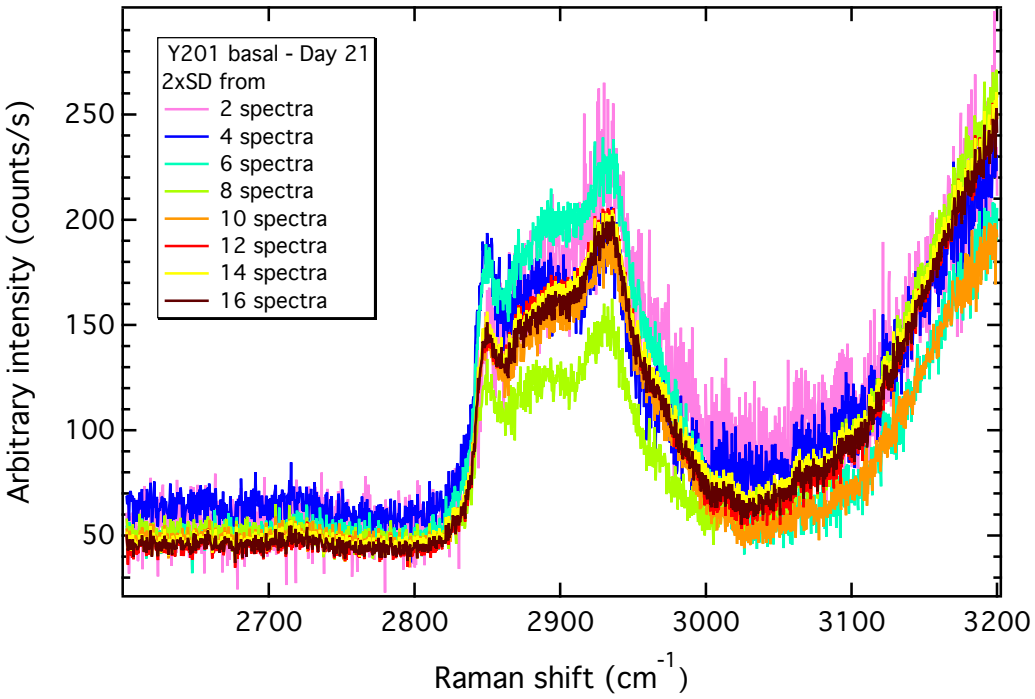


(b)

Figure B.78: 2x SD stacks for Y201 in basal medium at day 14. (a) fingerprint and (b) high wavenumber regions.



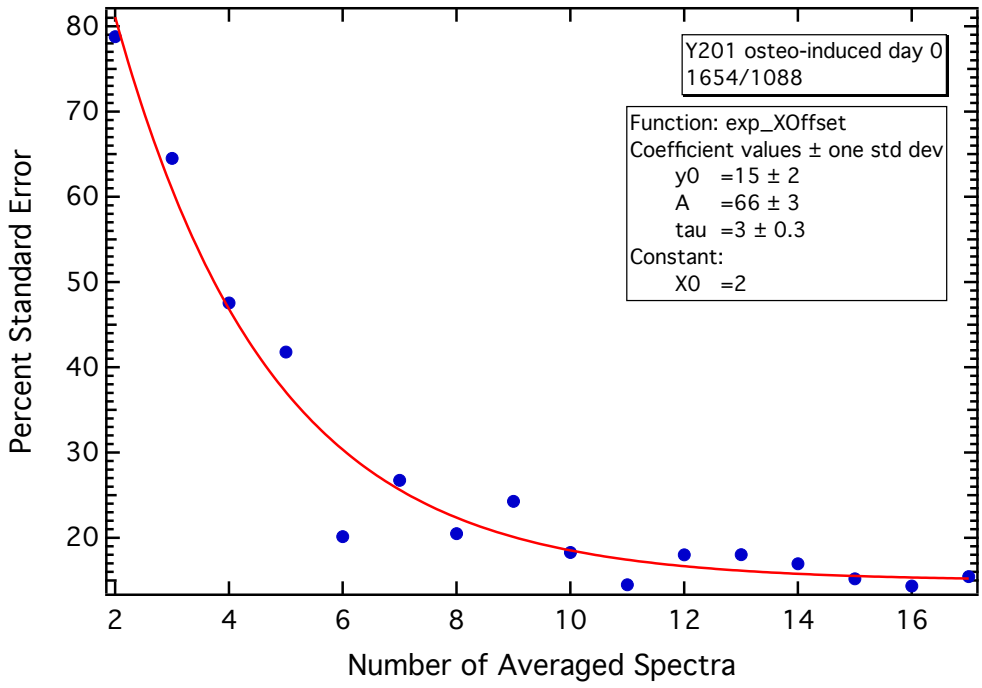
(a)



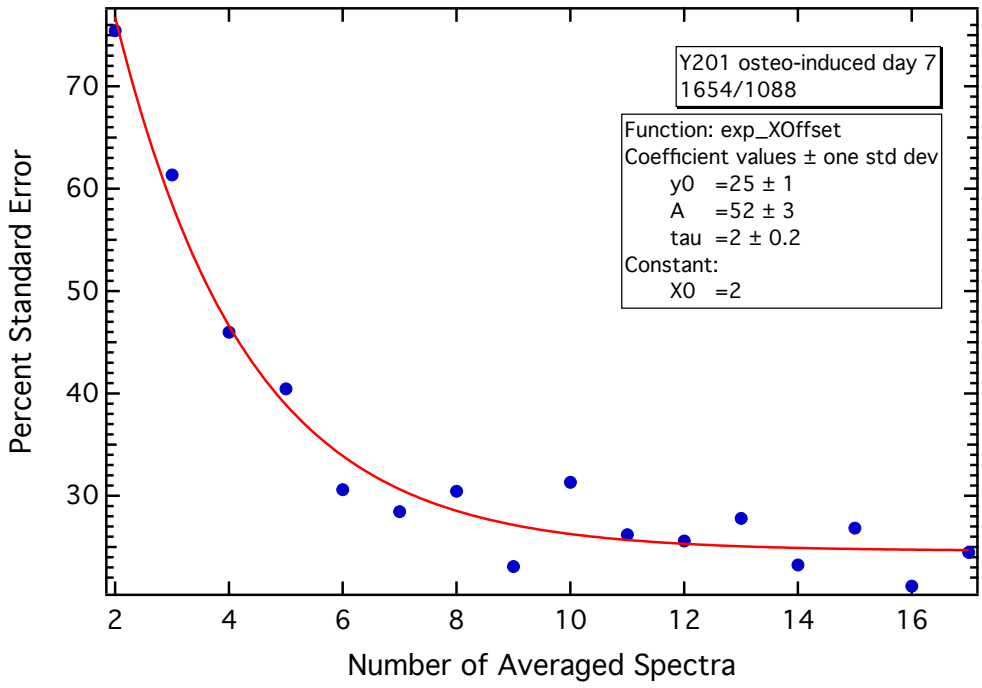
(b)

Figure B.79: 2x SD stacks for Y201 in basal medium at day 21. (a) fingerprint and (b) high wavenumber regions.

B.10 Osteo-induced Y201 MSC line – %SE convergence tests

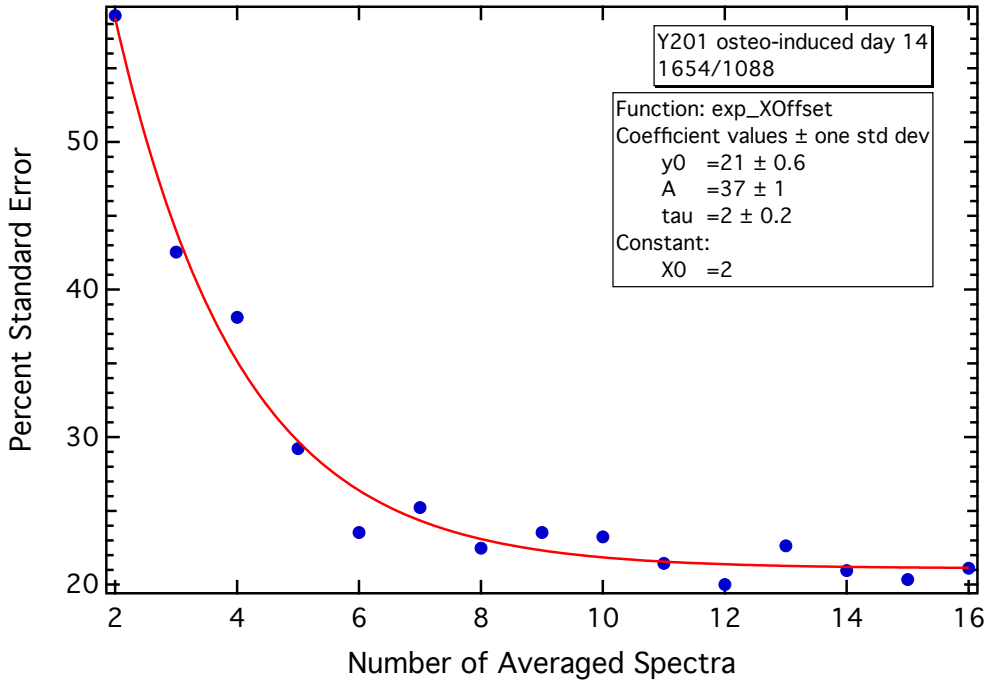


(a)

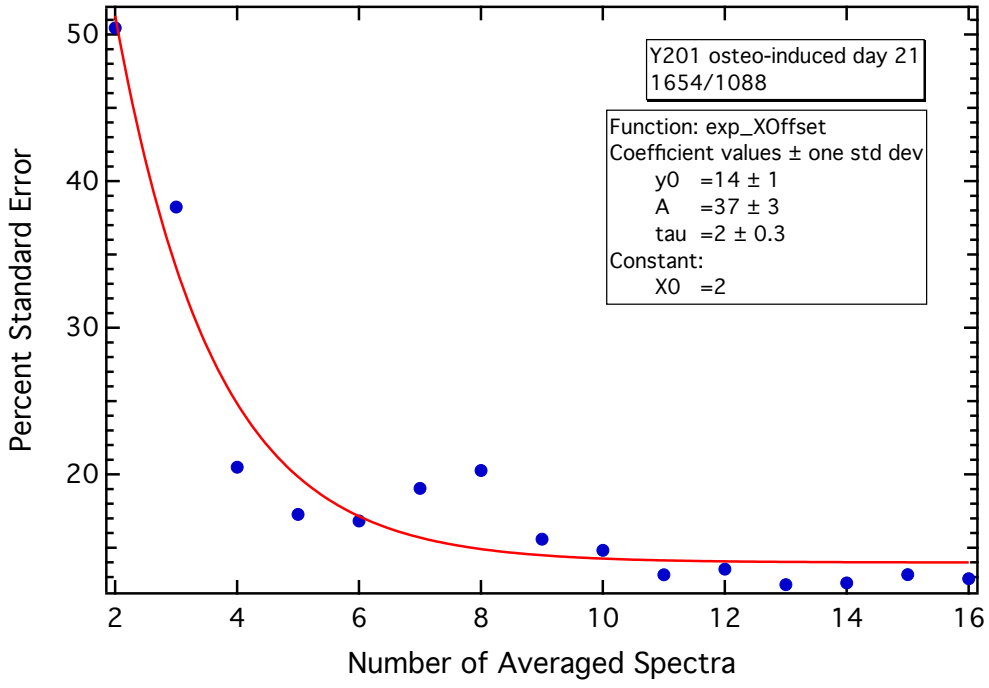


(b)

Figure B.80: %SE convergence tests for osteo-induced Y201 at days (a) 0 and (b) 7 for PIR 1654/1088.



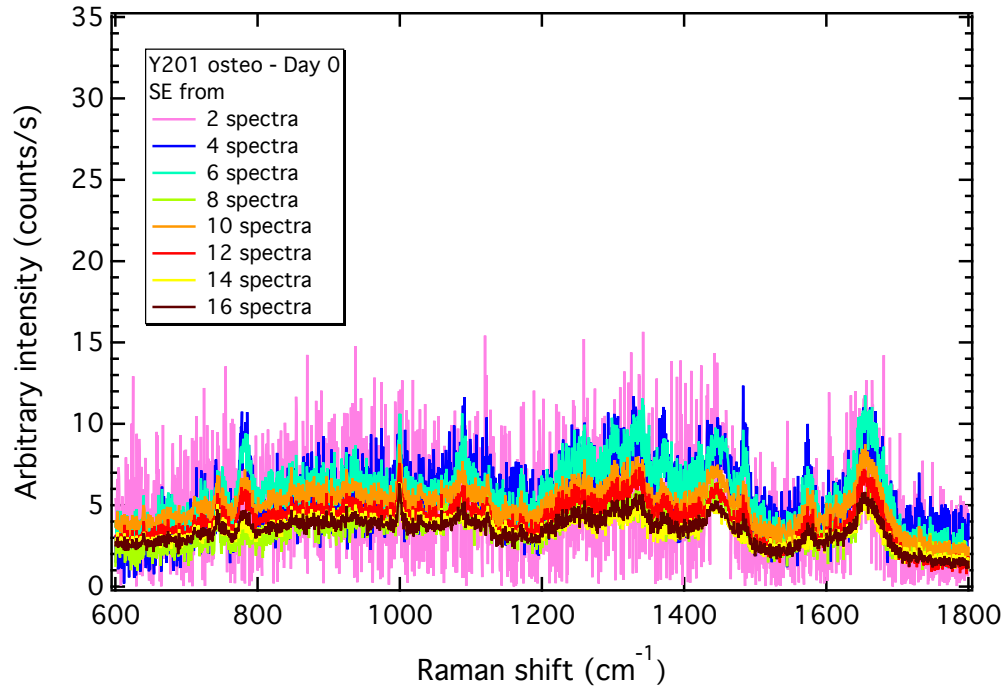
(a)



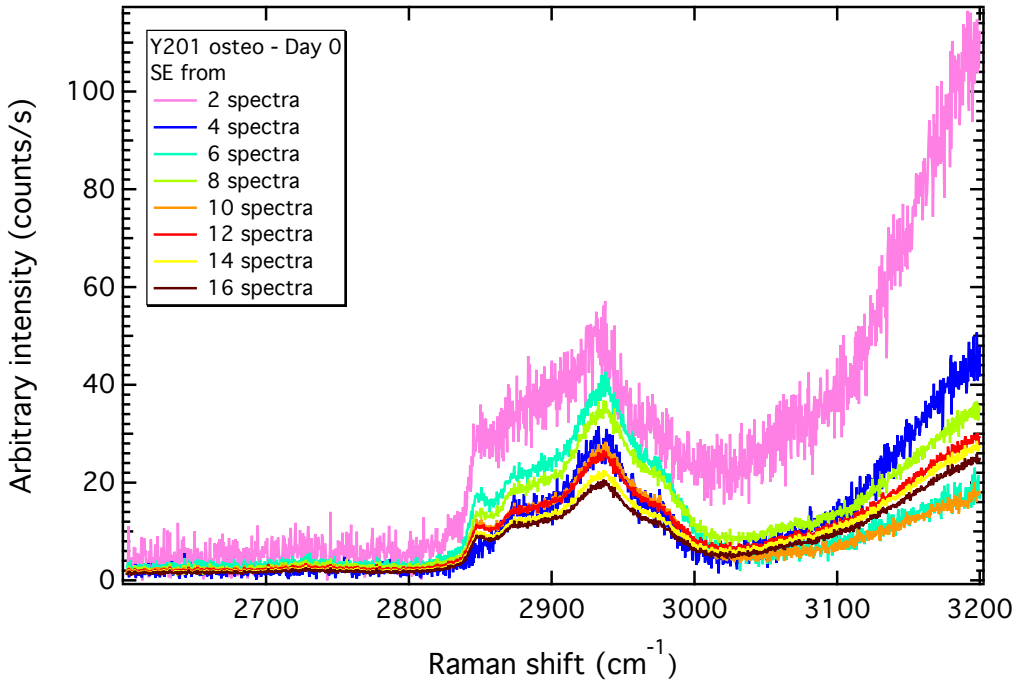
(b)

Figure B.81: %SE convergence tests for osteo-induced Y201 at days (a) 14 and (b) 21 for PIR 1654/1088.

B.11 Osteo-induced Y201 MSC line – SE convergence tests

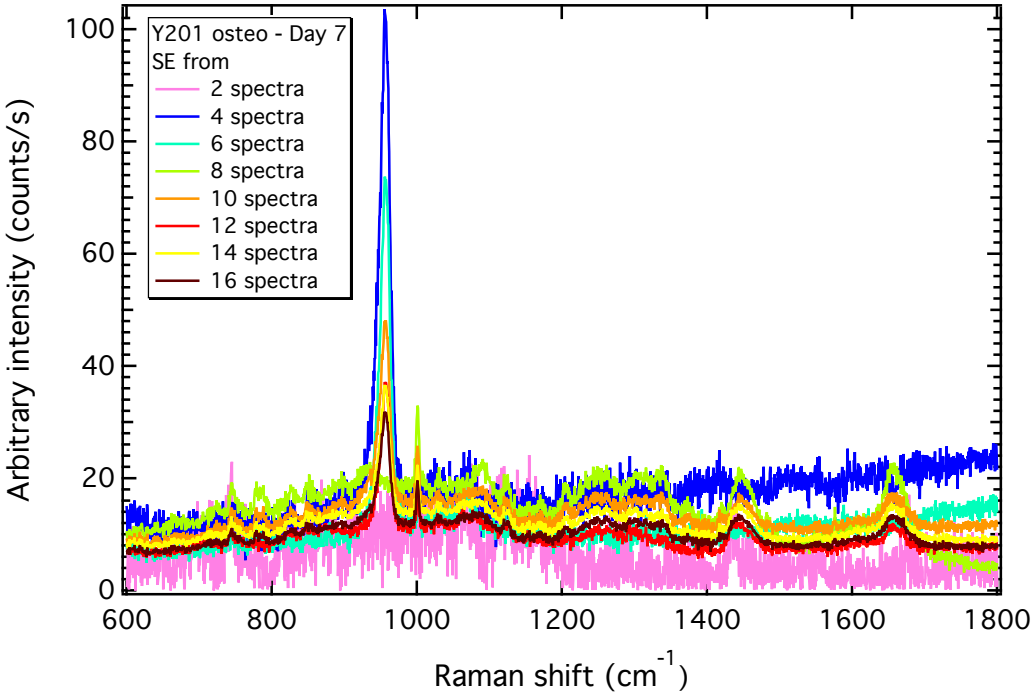


(a)

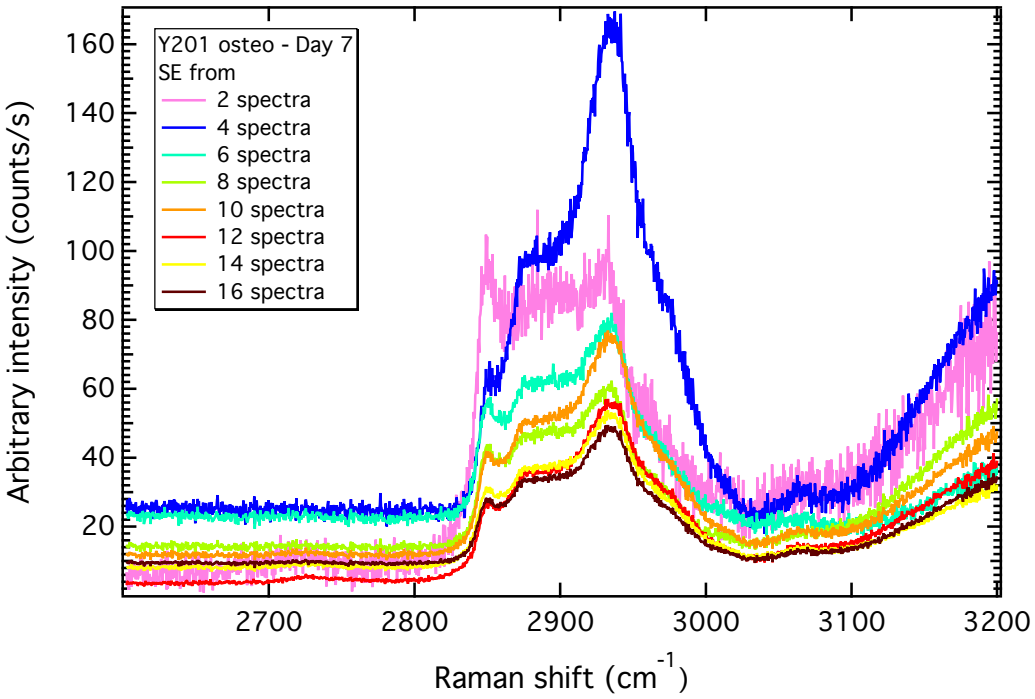


(b)

Figure B.82: SE stacks for osteo-induced Y201 at day 0. (a) fingerprint and (b) high wavenumber regions.

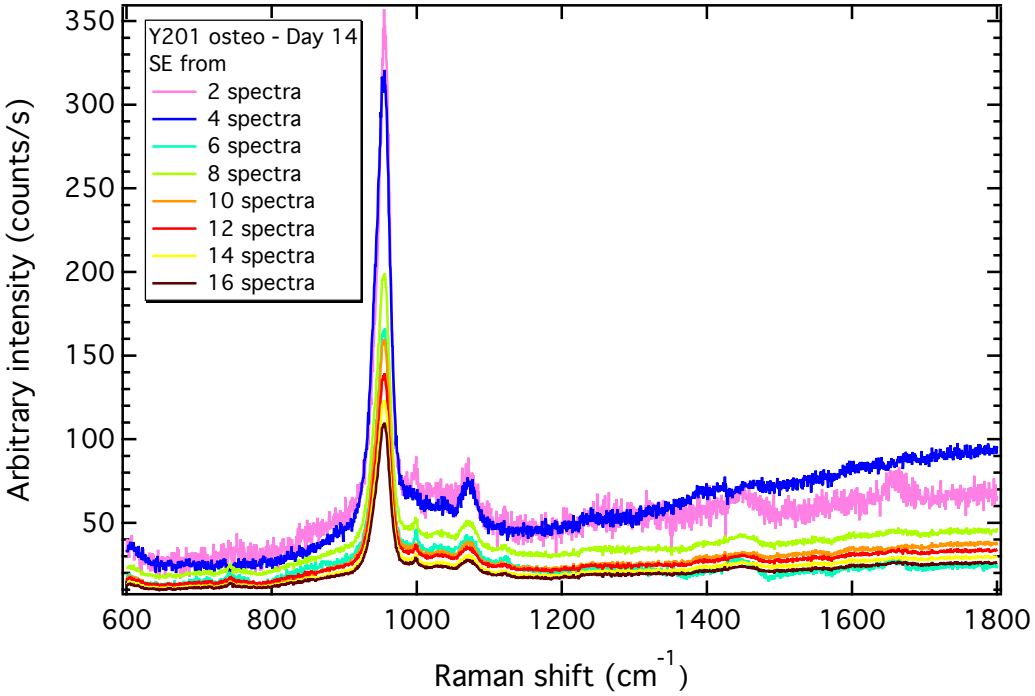


(a)

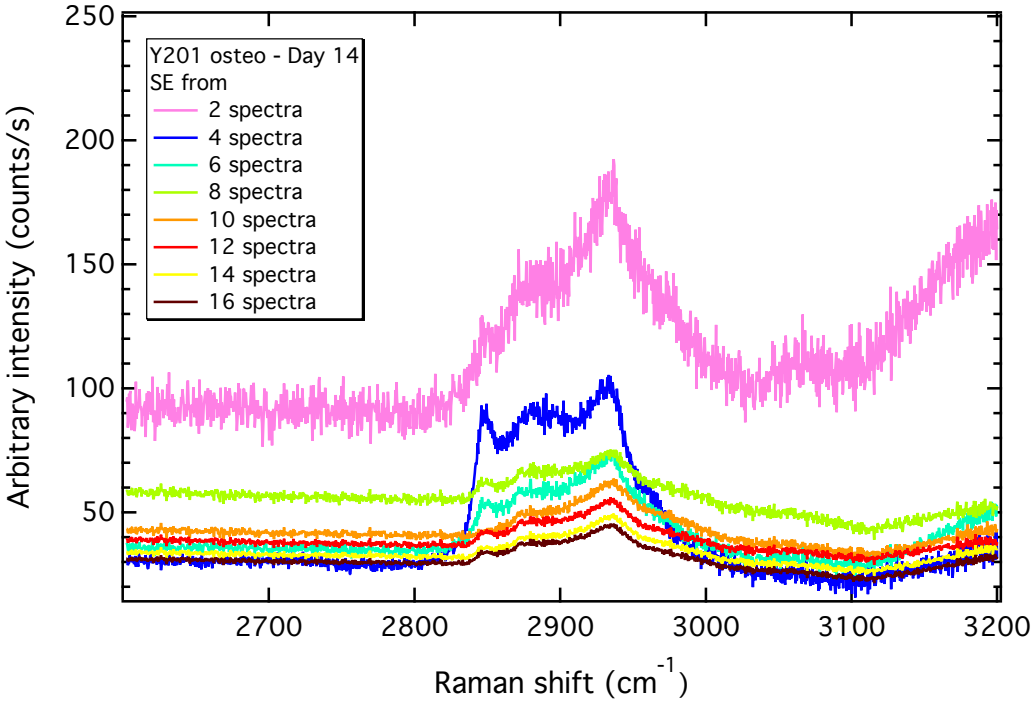


(b)

Figure B.83: SE stacks for osteo-induced Y201 at day 7. (a) fingerprint and (b) high wavenumber regions.

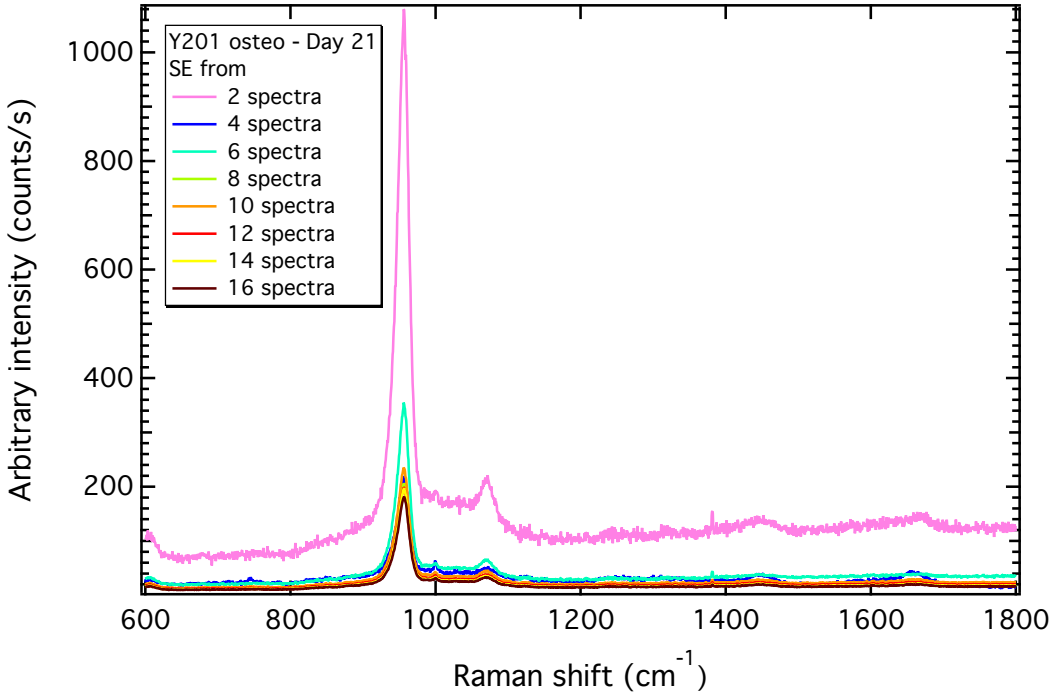


(a)

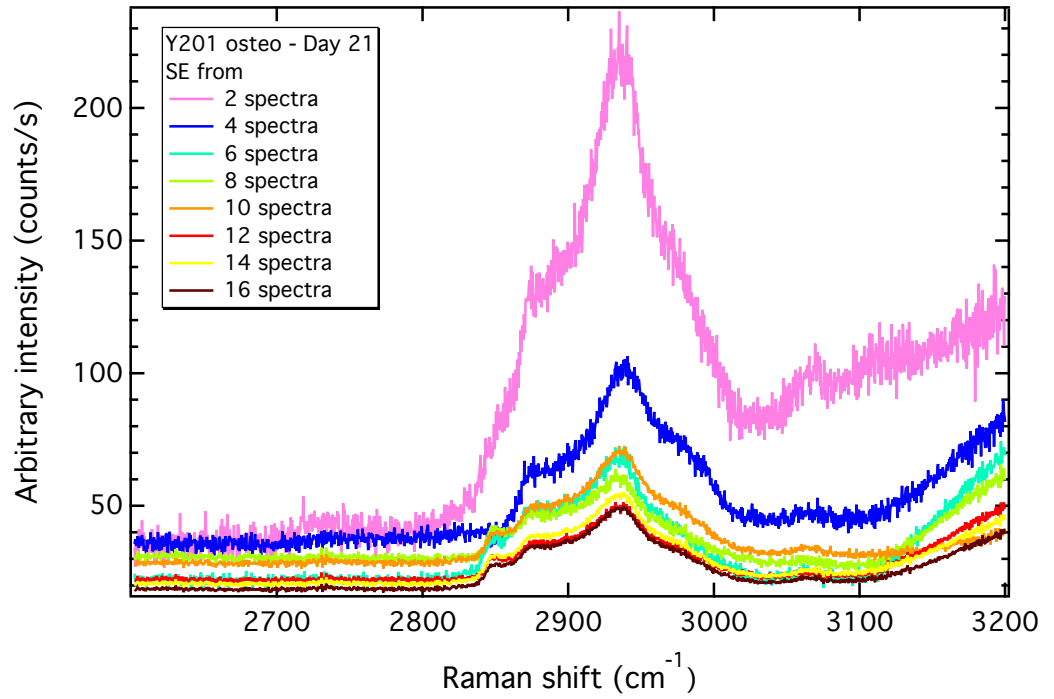


(b)

Figure B.84: SE stacks for osteo-induced Y201 at day 14. (a) fingerprint and (b) high wavenumber regions.



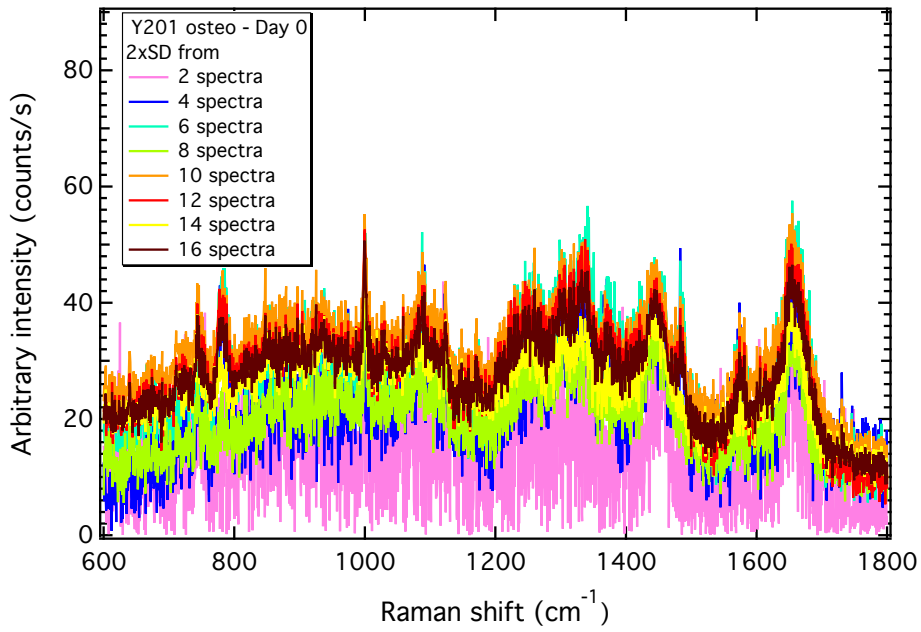
(a)



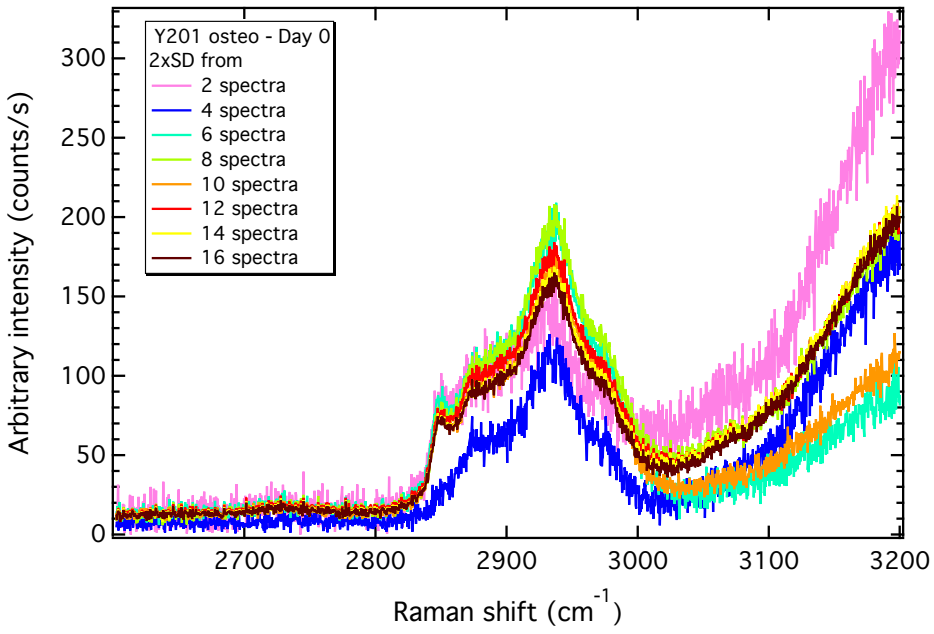
(b)

Figure B.85: SE stacks for osteo-induced Y201 at day 21. (a) fingerprint and (b) high wavenumber regions.

B.12 Osteo-induced Y201 MSC line – 2x SD convergence tests

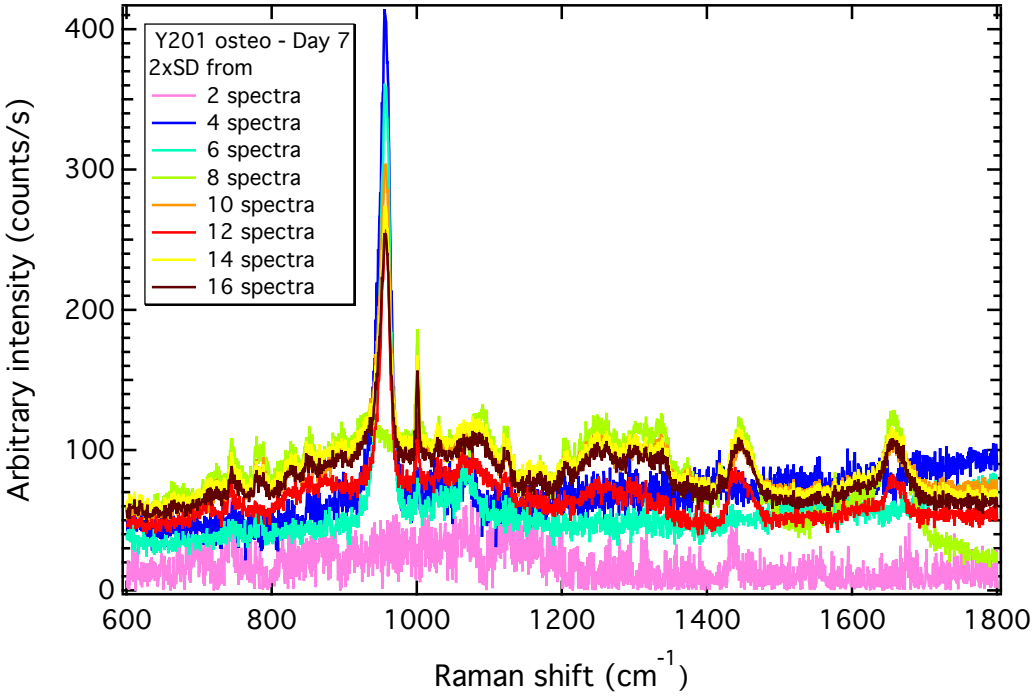


(a)

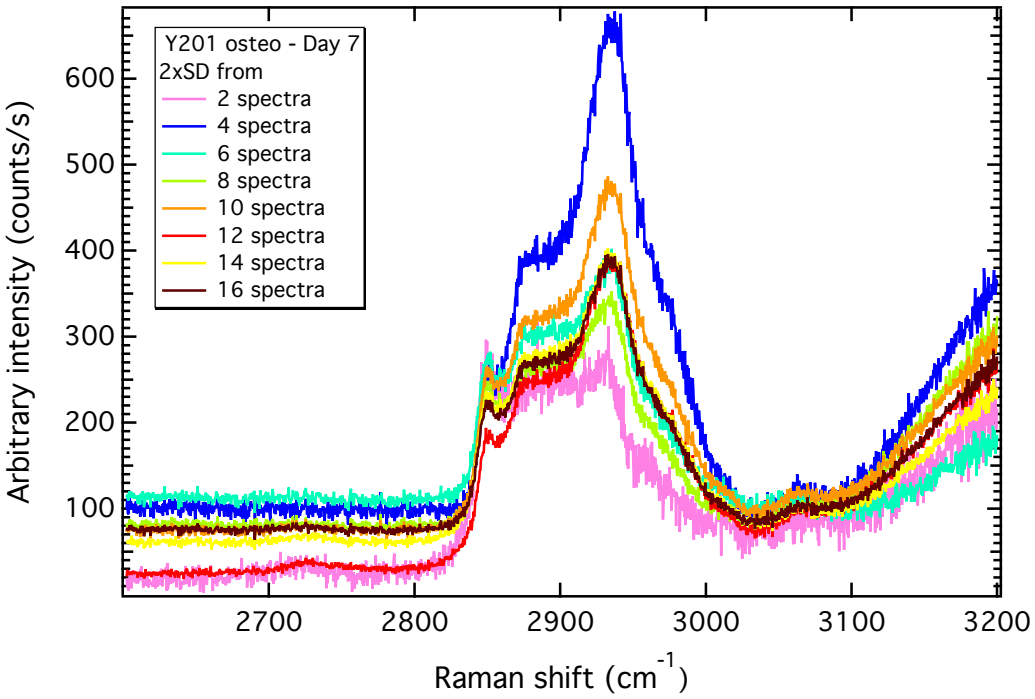


(b)

Figure B.86: 2x SD stacks for osteo-induced Y201 at day 0. (a) fingerprint and (b) high wavenumber regions.

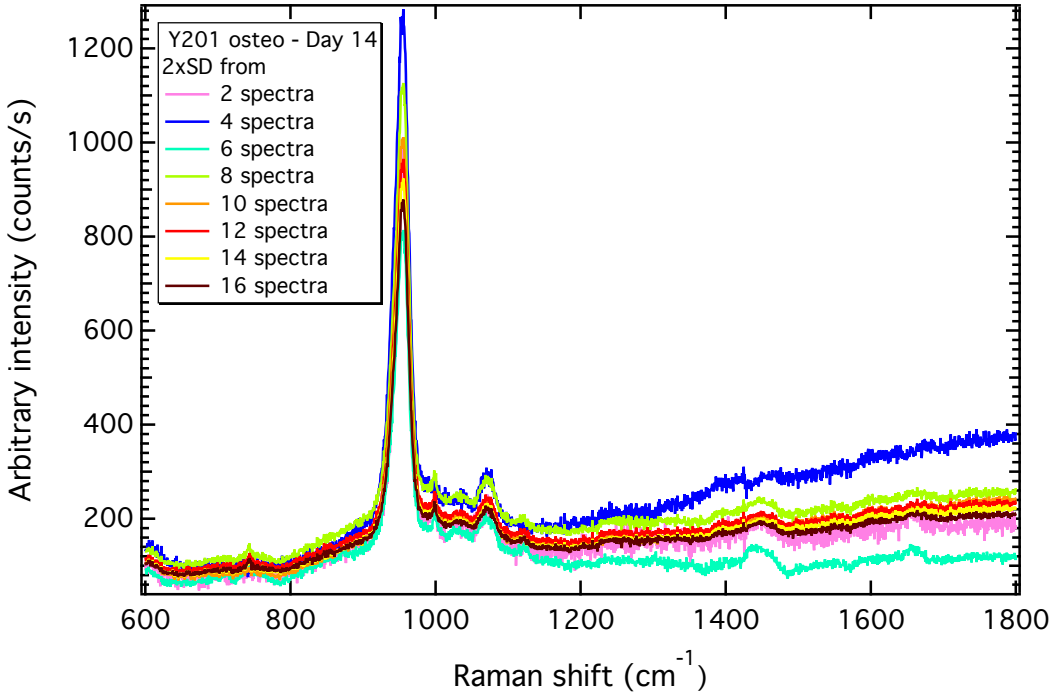


(a)

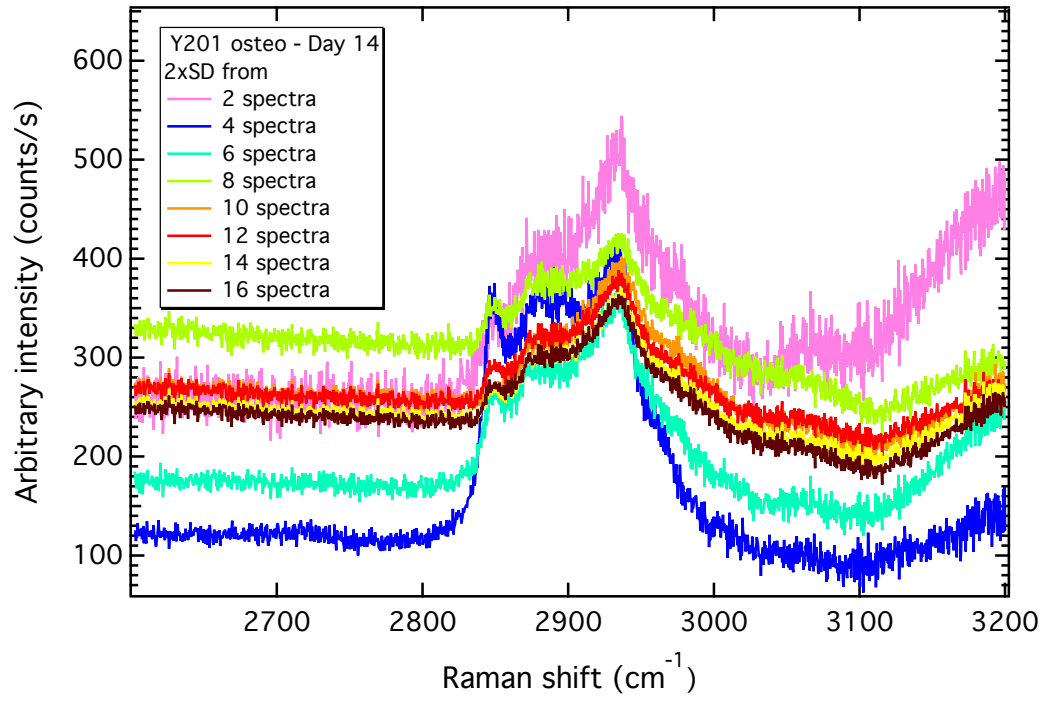


(b)

Figure B.87: 2x SD stacks for osteo-induced Y201 at day 7. (a) fingerprint and (b) high wavenumber regions.

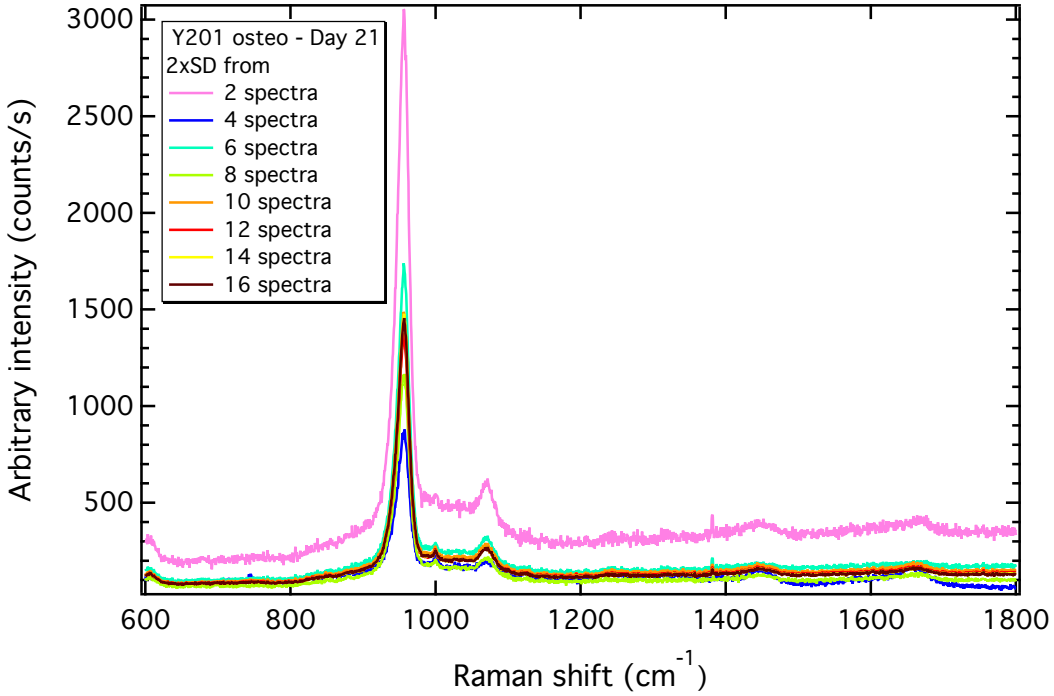


(a)

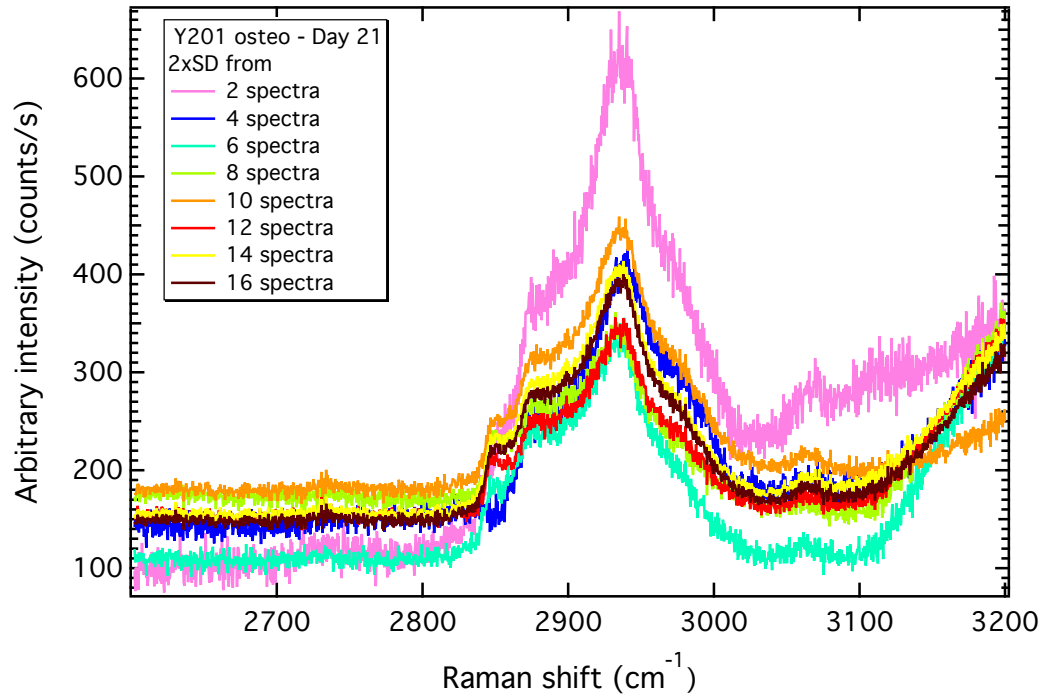


(b)

Figure B.88: 2x SD stacks for osteo-induced Y201 at day 14. (a) fingerprint and (b) high wavenumber regions.



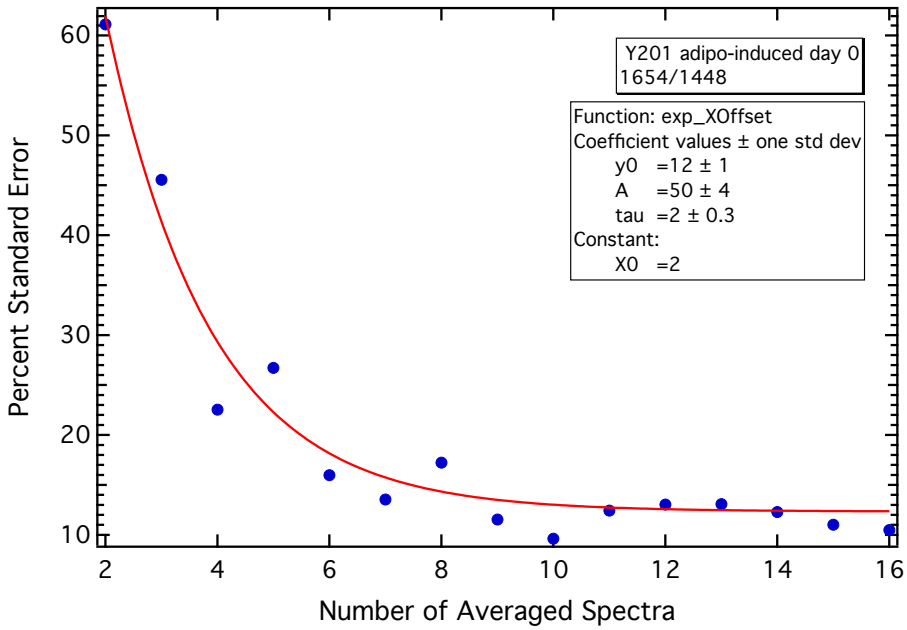
(a)



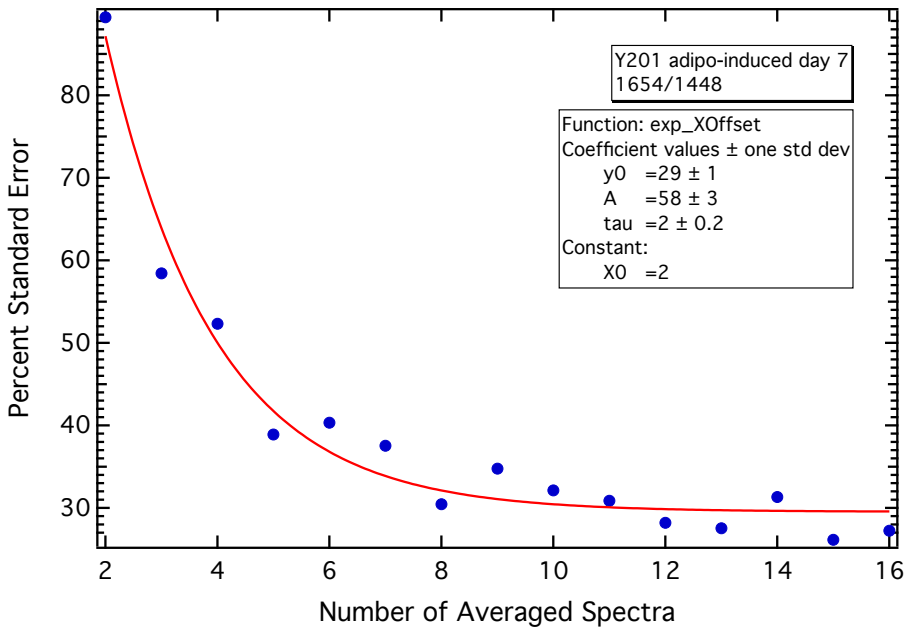
(b)

Figure B.89: 2x SD stacks for osteo-induced Y201 at day 21. (a) fingerprint and (b) high wavenumber regions.

B.13 Adipo-induced Y201 MSC line – %SE convergence tests

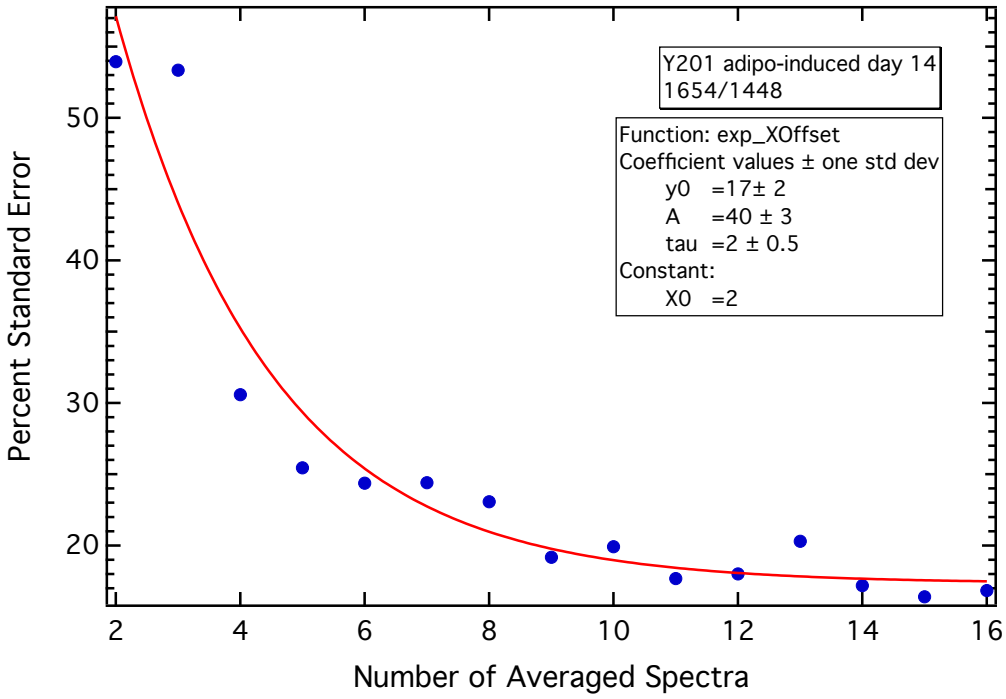


(a)

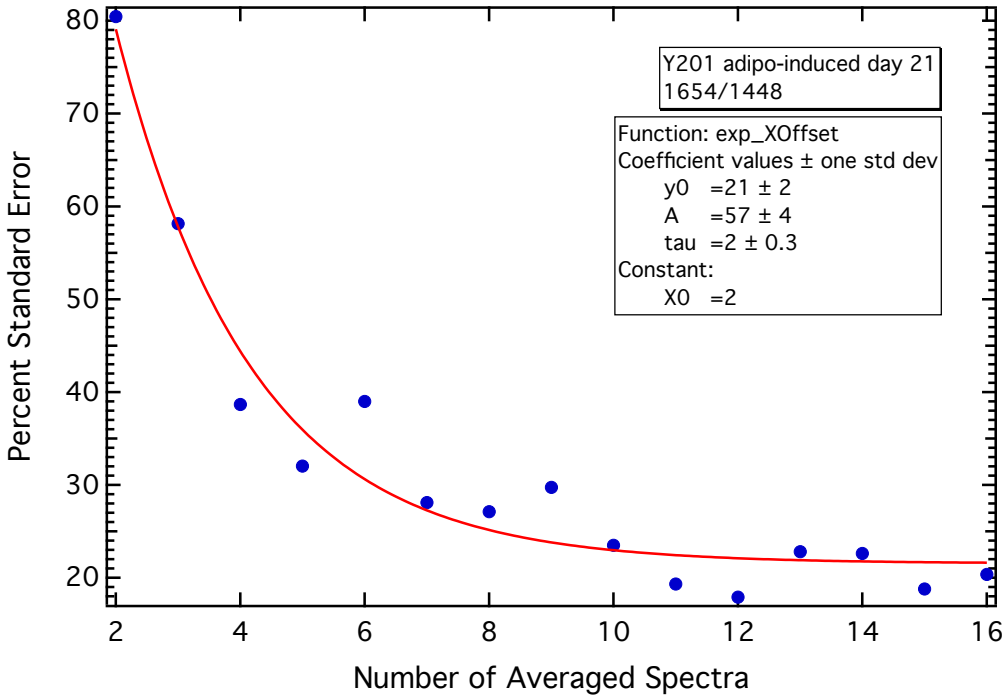


(b)

Figure B.90: %SE convergence tests for adipo-induced Y201 at days (a) 0 and (b) 7 for PIR 165/1448.



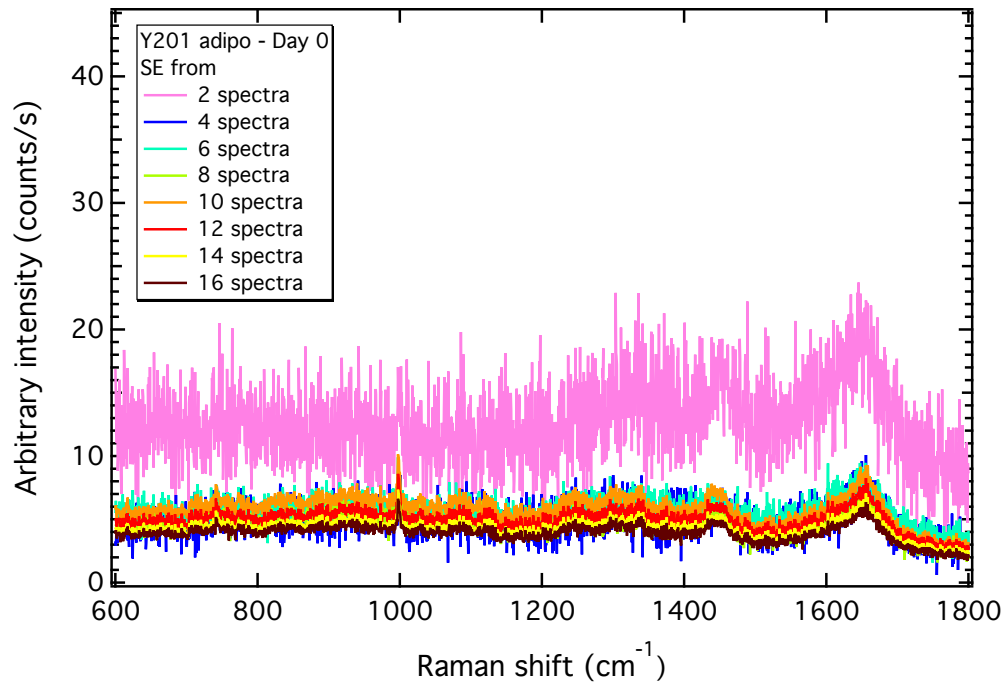
(a)



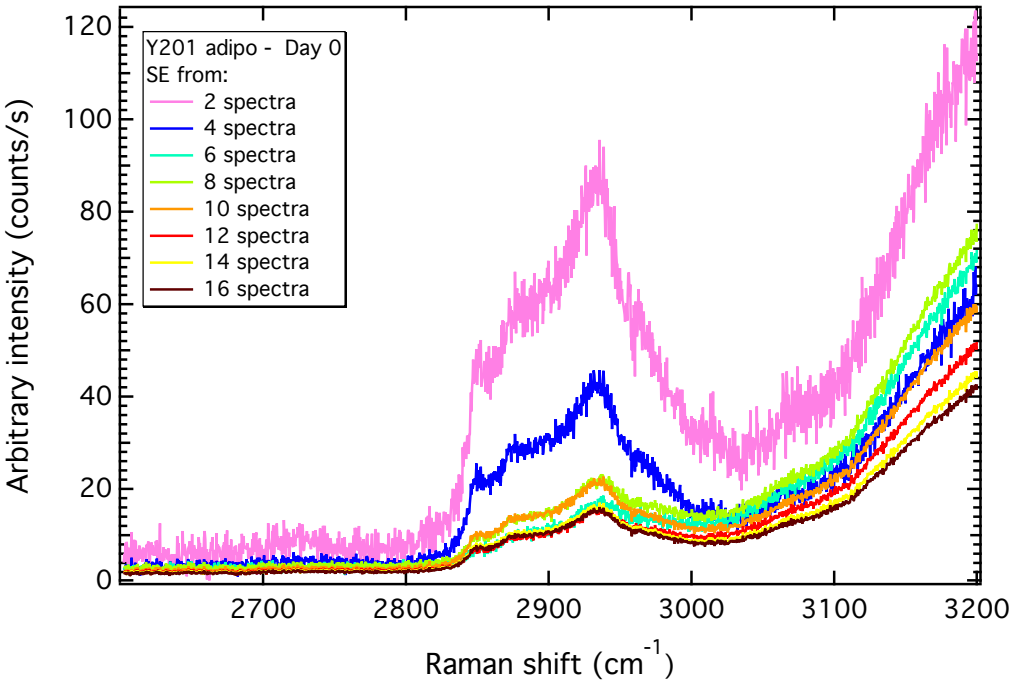
(b)

Figure B.91: %SE convergence tests for adipo-induced Y201 at days (a) 14 and (b) 21 for PIR 165/1448.

B.14 Adipo-induced Y201 MSC line – SE convergence tests

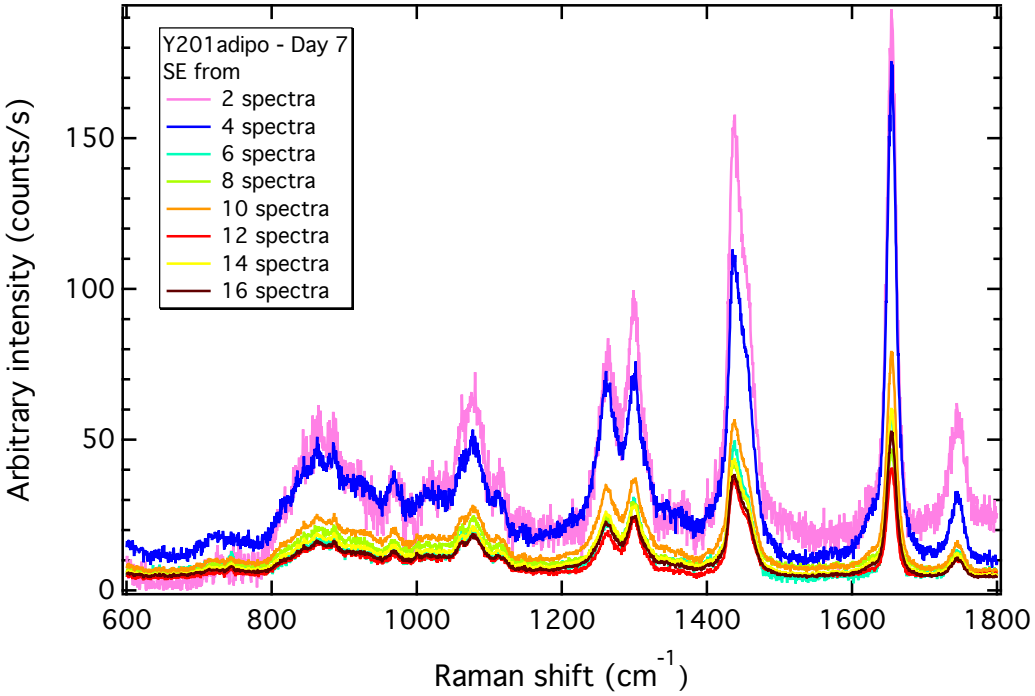


(a)

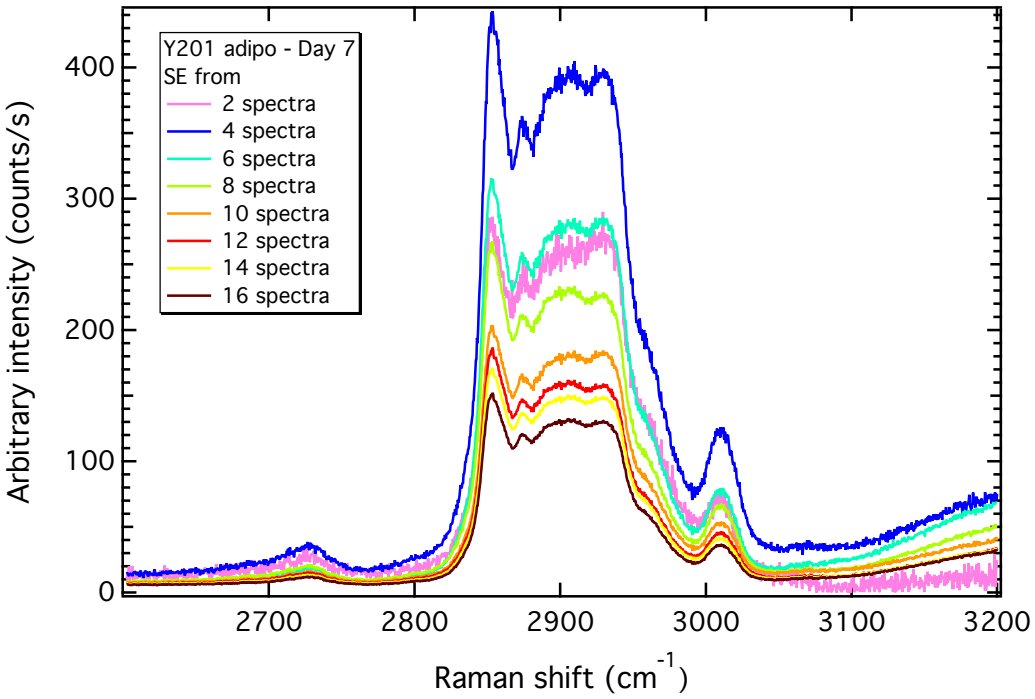


(b)

Figure B.92: SE stacks for adipo-induced Y201 at day 0. (a) Fingerprint and (b) high wavenumber regions.

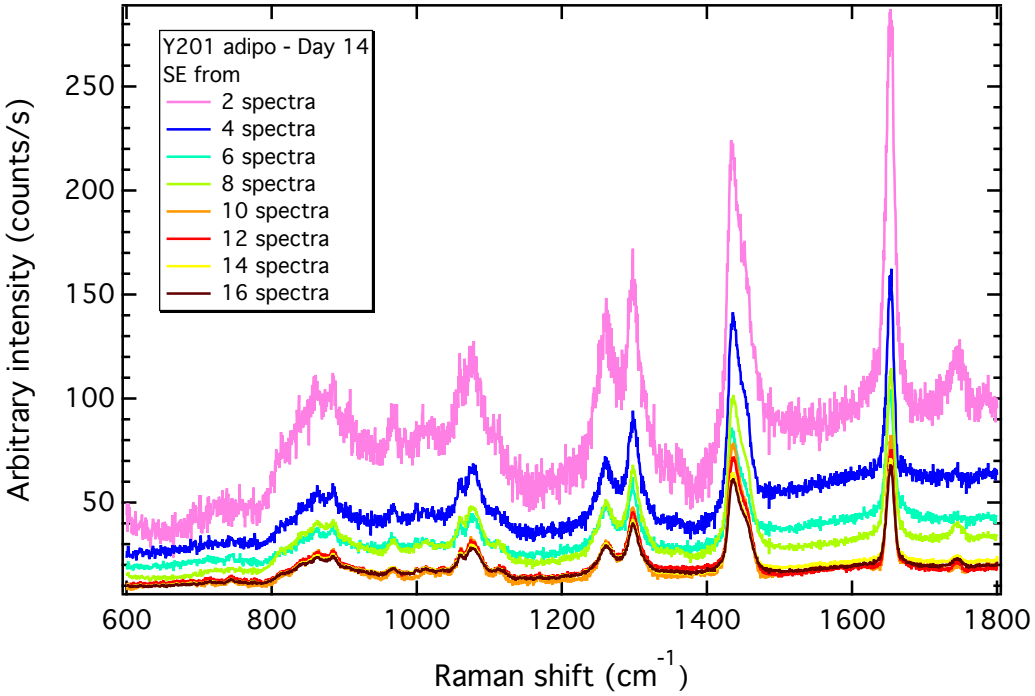


(a)

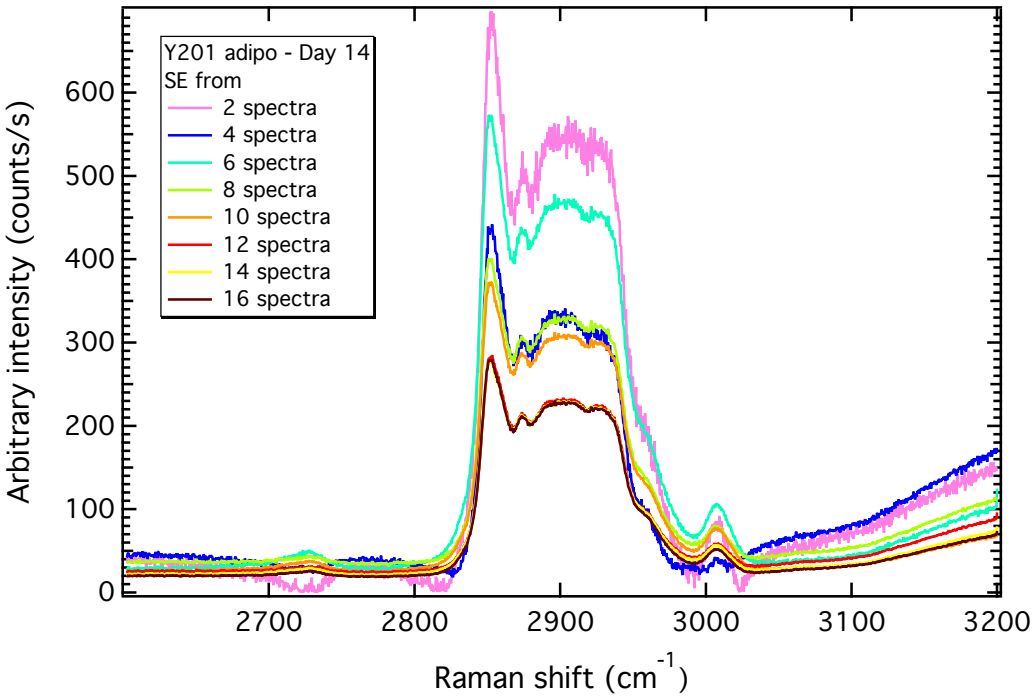


(b)

Figure B.93: SE stacks for adipo-induced Y201 at day 7. (a) Fingerprint and (b) high wavenumber regions.

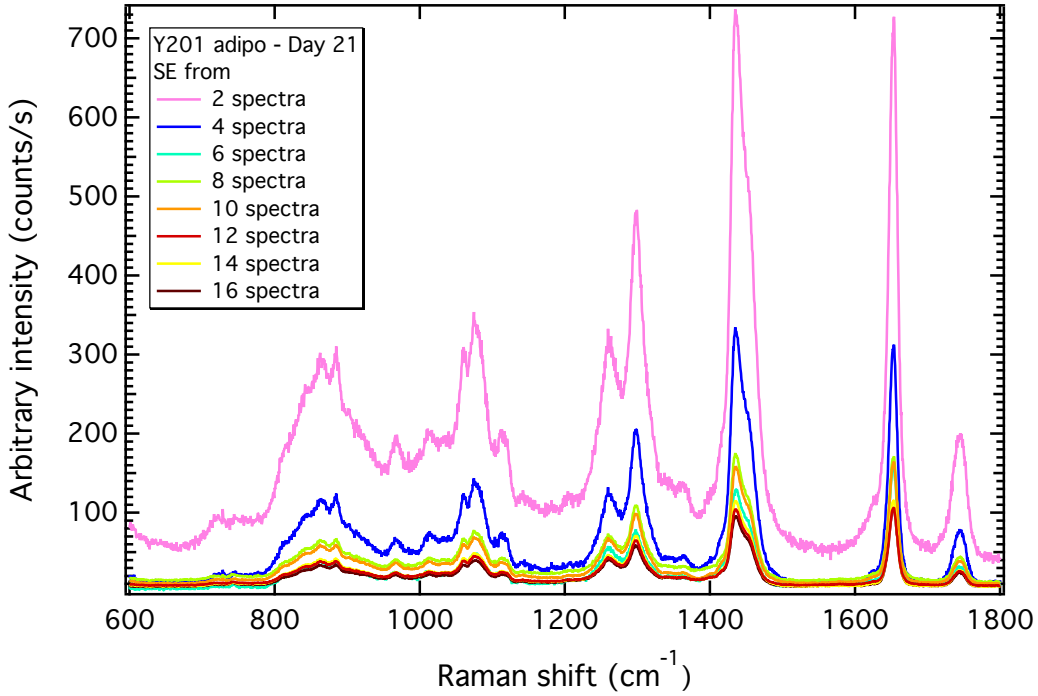


(a)

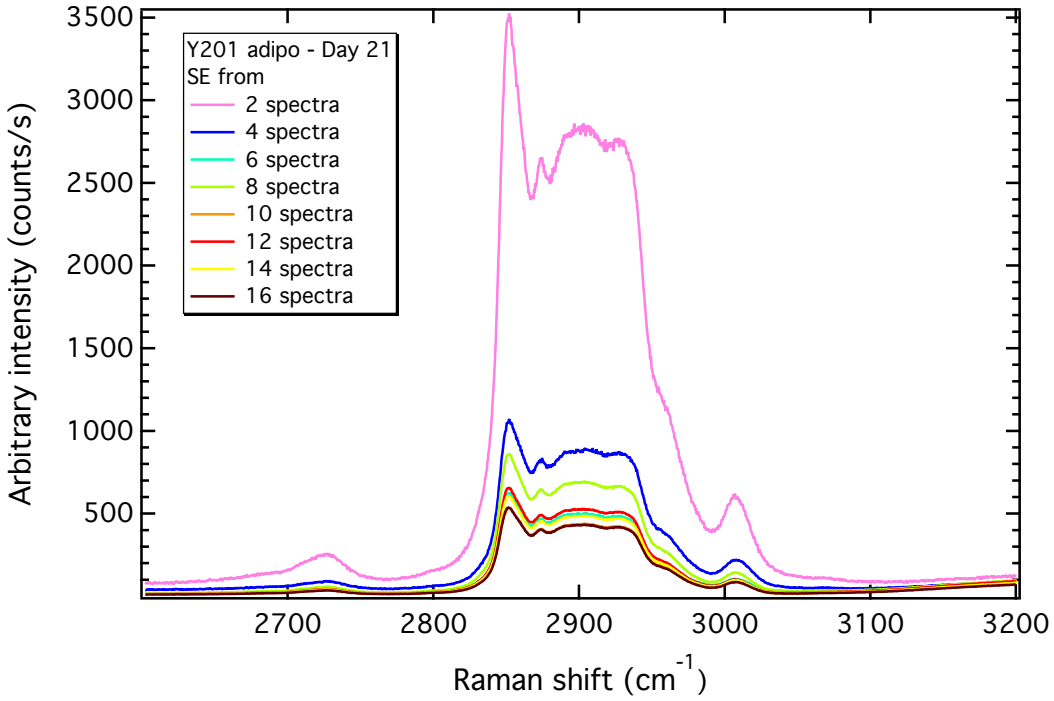


(b)

Figure B.94: SE stacks for adipo-induced Y201 at day 14. (a) Fingerprint and (b) high wavenumber regions.



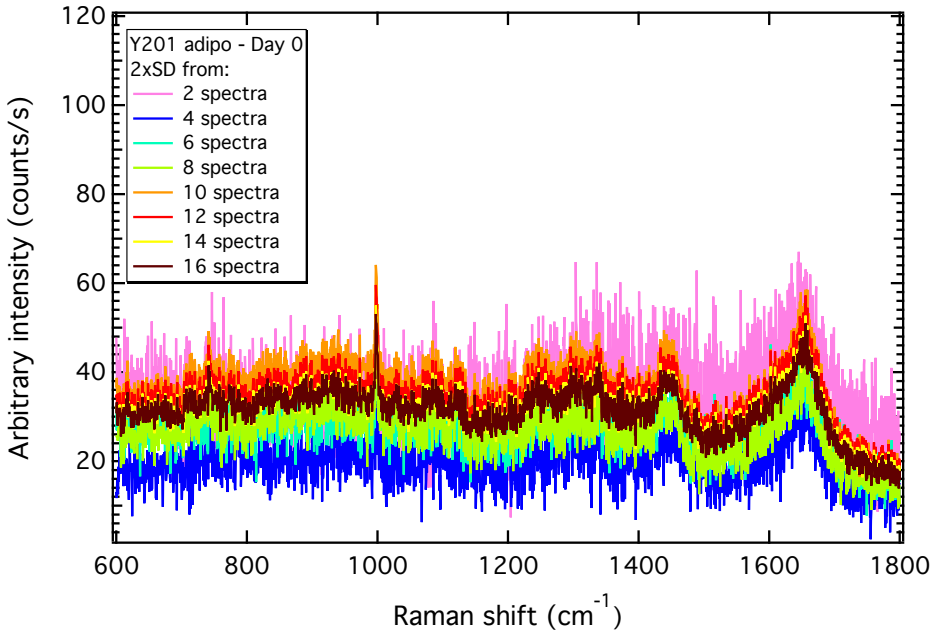
(a)



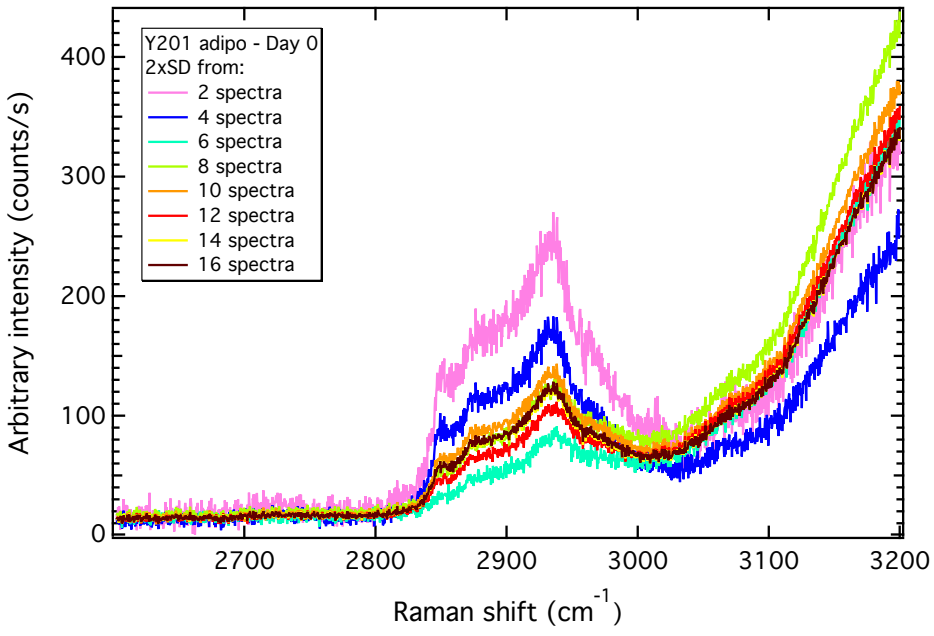
(b)

Figure B.95: SE stacks for adipo-induced Y201 at day 21. (a) Fingerprint and (b) high wavenumber regions.

B.15 Adipo-induced Y201 MSC line – 2x SD convergence tests

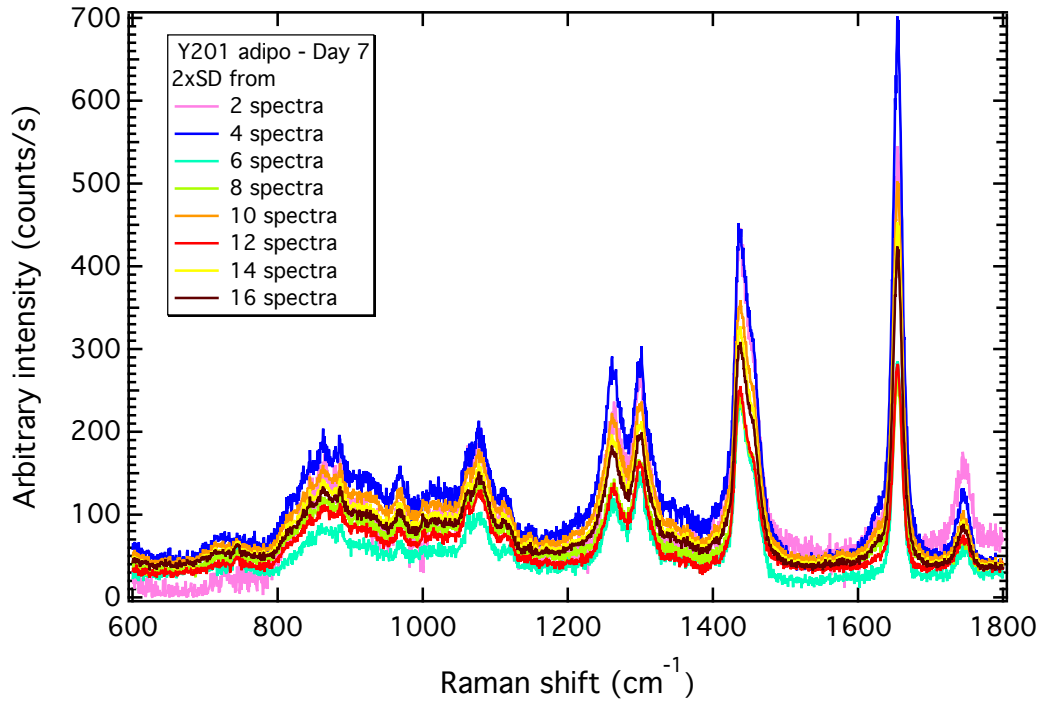


(a)

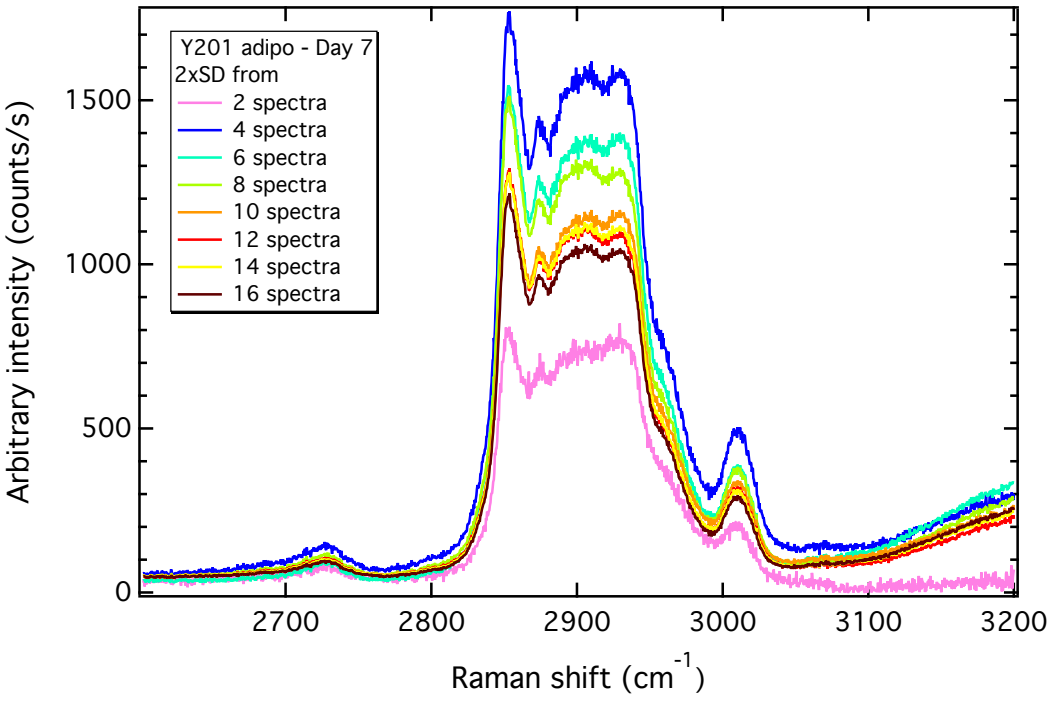


(b)

Figure B.96: 2x SD stacks for adipo-induced Y201 at day 0. (a) Fingerprint and (b) high wavenumber regions.

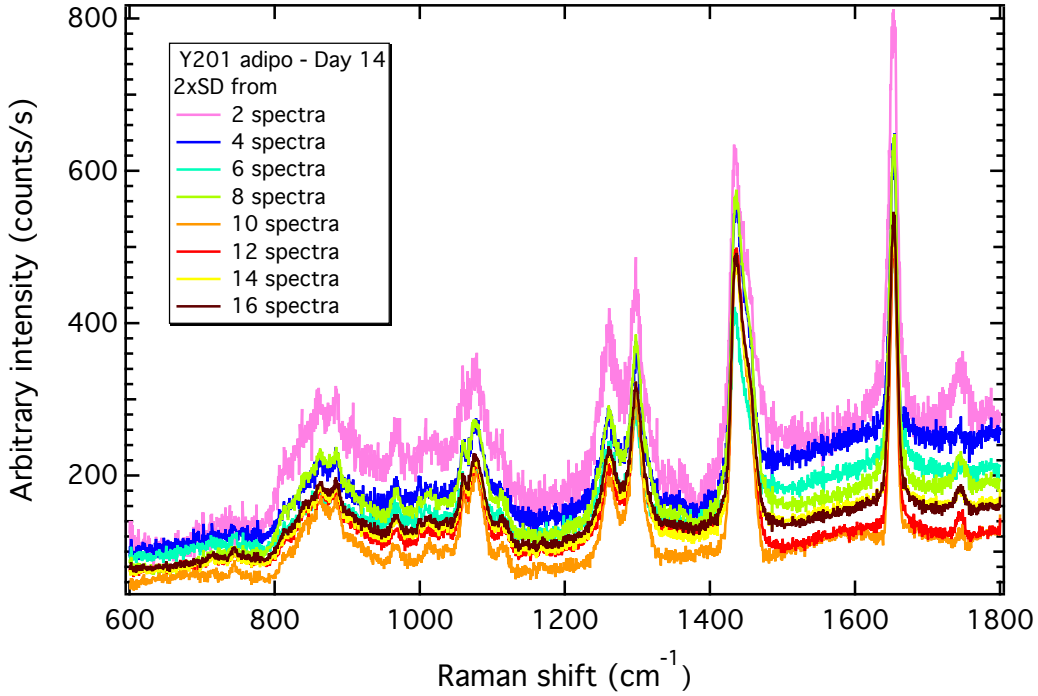


(a)

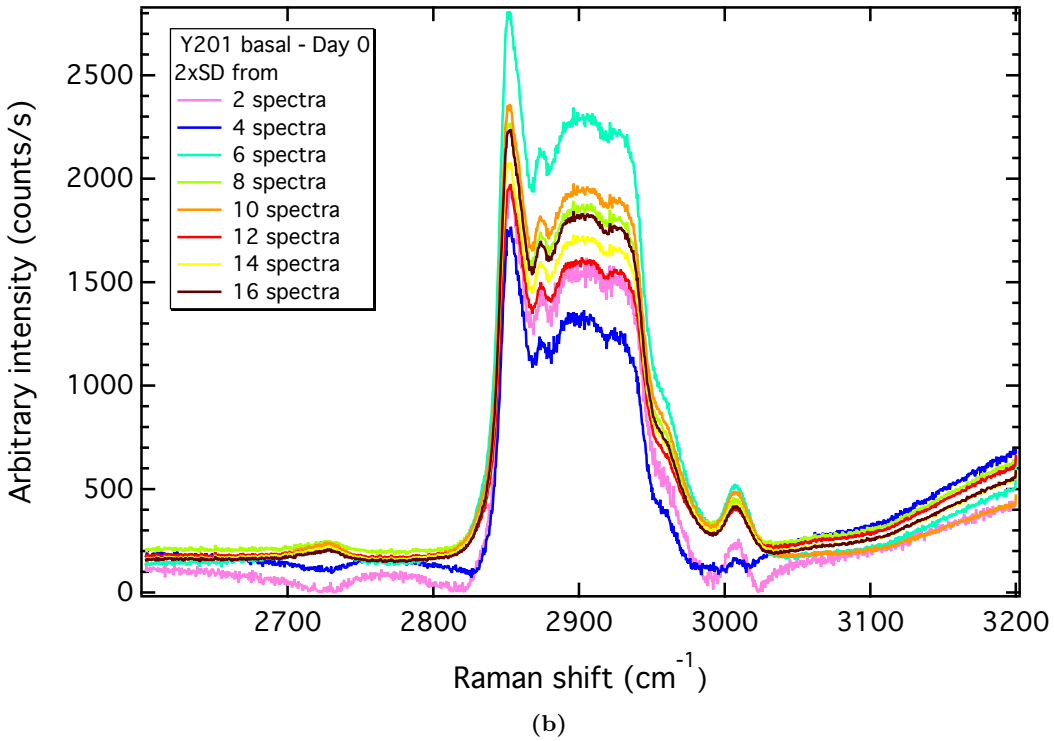


(b)

Figure B.97: 2x SD stacks for adipo-induced Y201 at day 7. (a) Fingerprint and (b) high wavenumber regions.

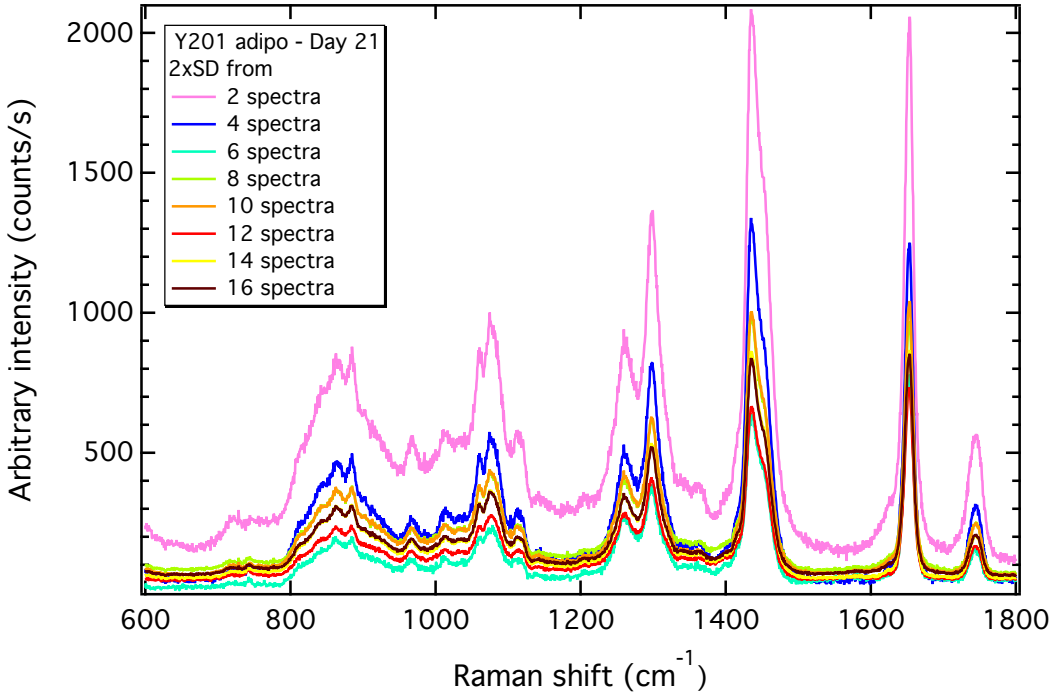


(a)

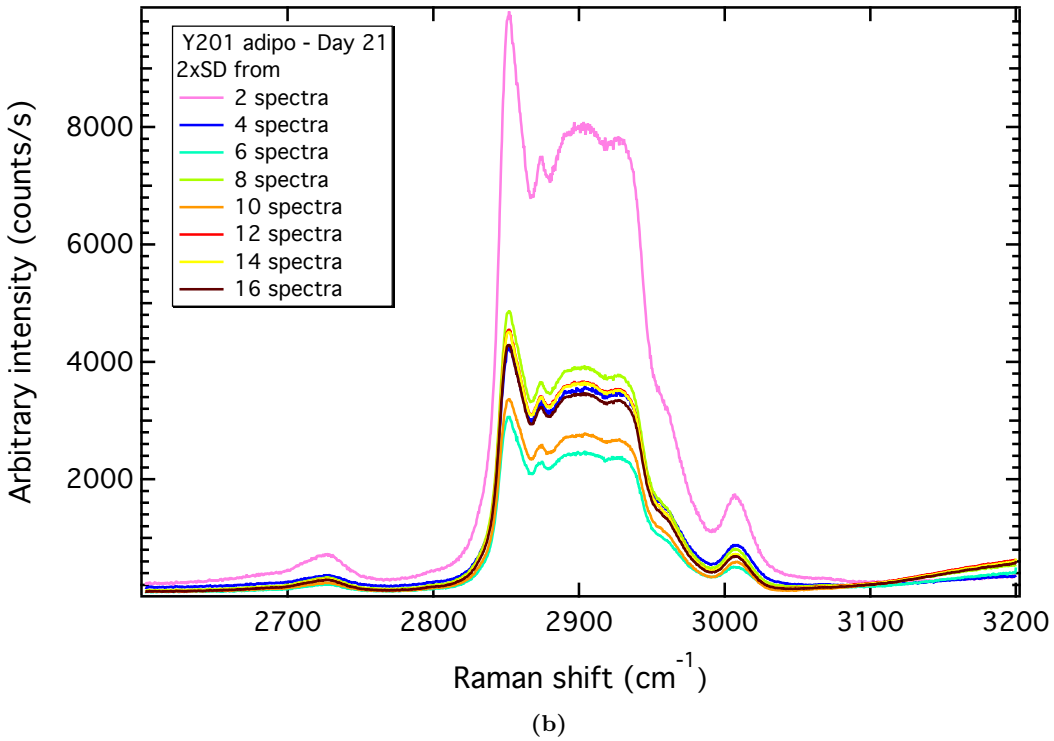


(b)

Figure B.98: 2x SD stacks for adipo-induced Y201 at day 14. (a) Fingerprint and (b) high wavenumber regions.



(a)

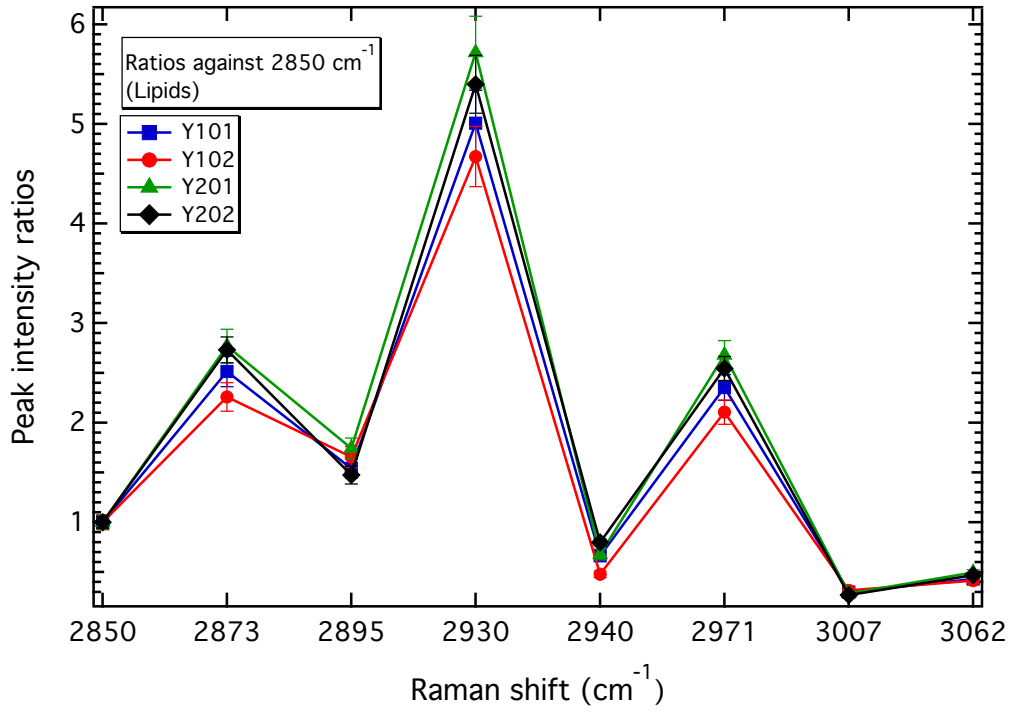


(b)

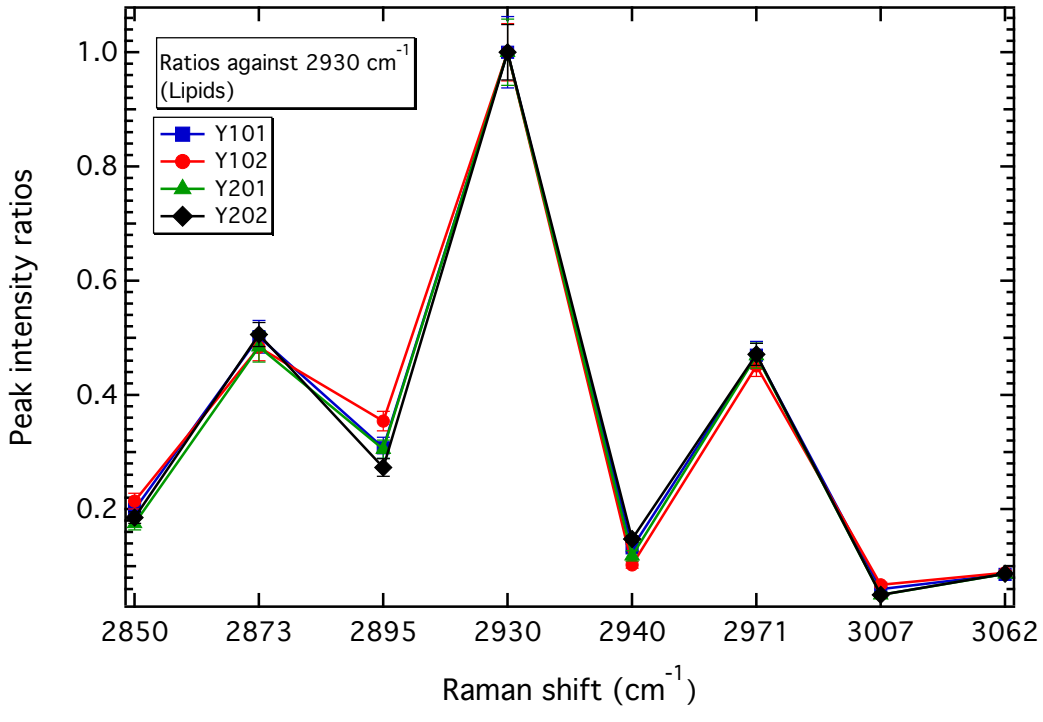
Figure B.99: 2x SD stacks for adipo-induced Y201 at day 21. (a) Fingerprint and (b) high wavenumber regions.

Appendix C

High-wavenumber PIRs of the live MSC lines

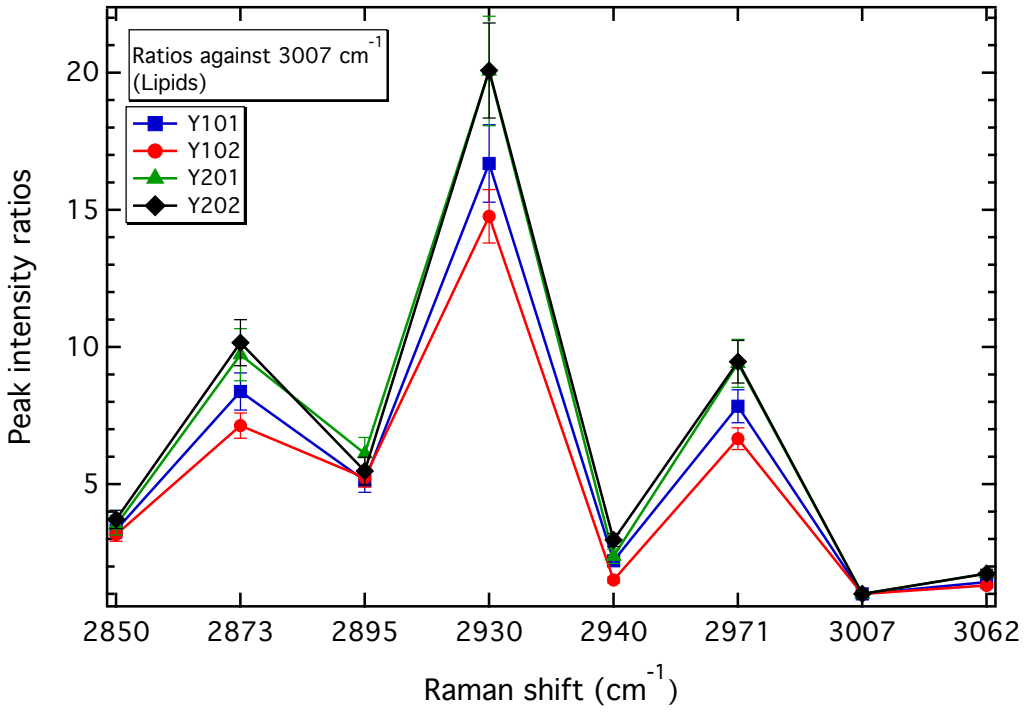


(a)

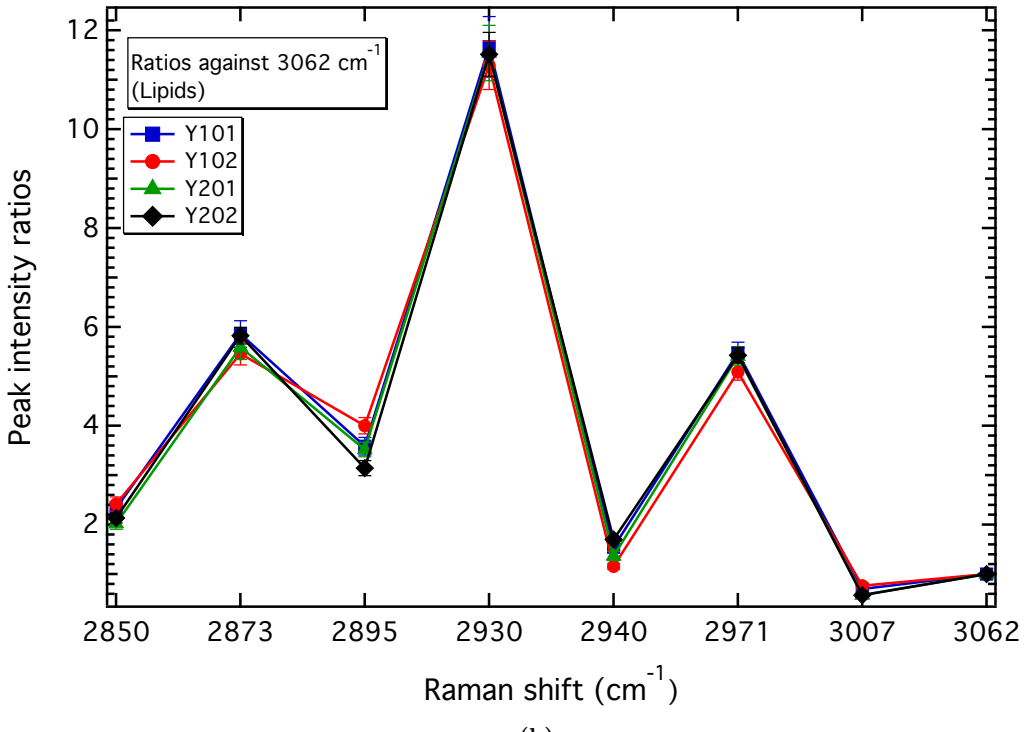


(b)

Figure C.1: PIR comparison for the live MSCs against the (a) 2850 cm^{-1} and (b) 2930 cm^{-1} peaks.

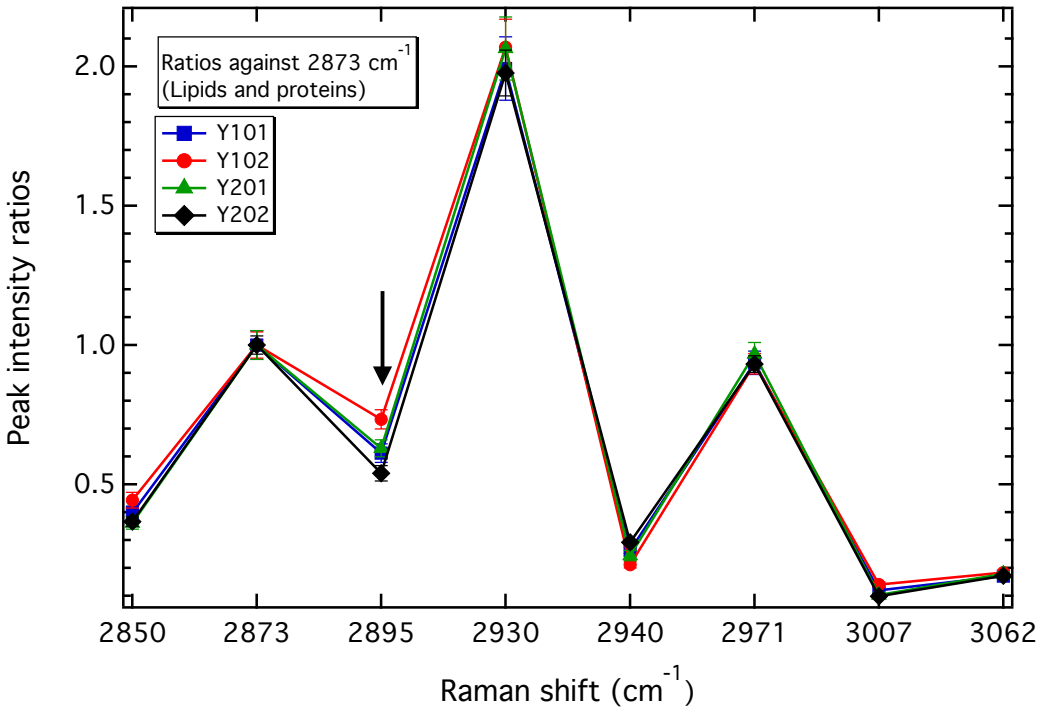


(a)

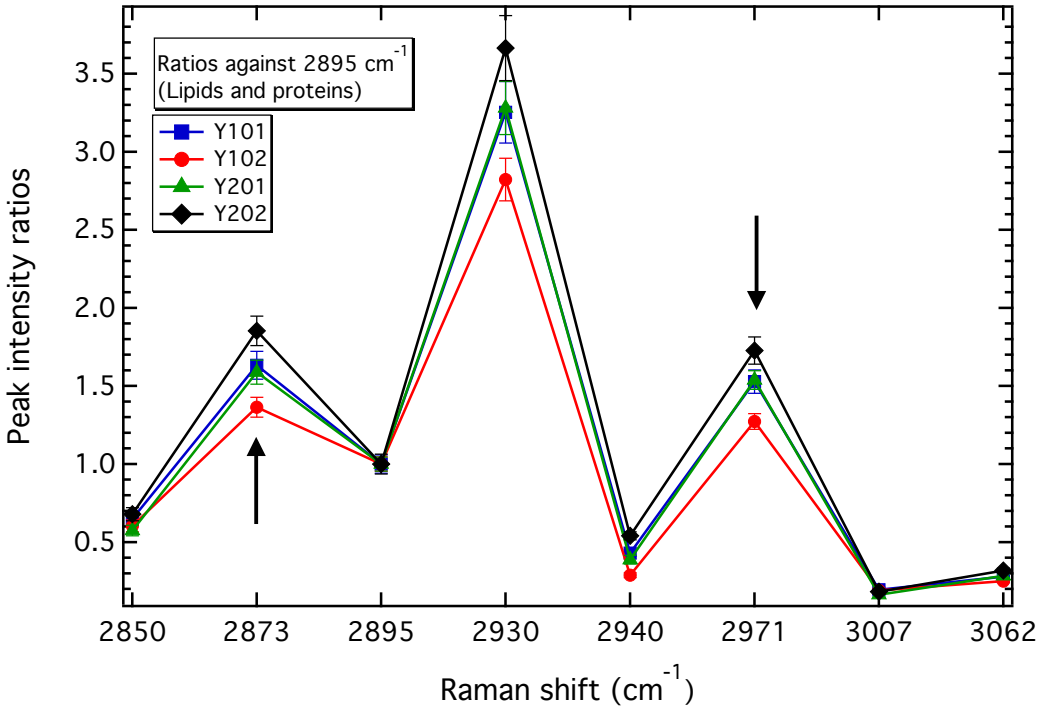


(b)

Figure C.2: PIR comparison for the live MSCs against the (a) 3007 cm^{-1} and (b) 3062 cm^{-1} peaks.

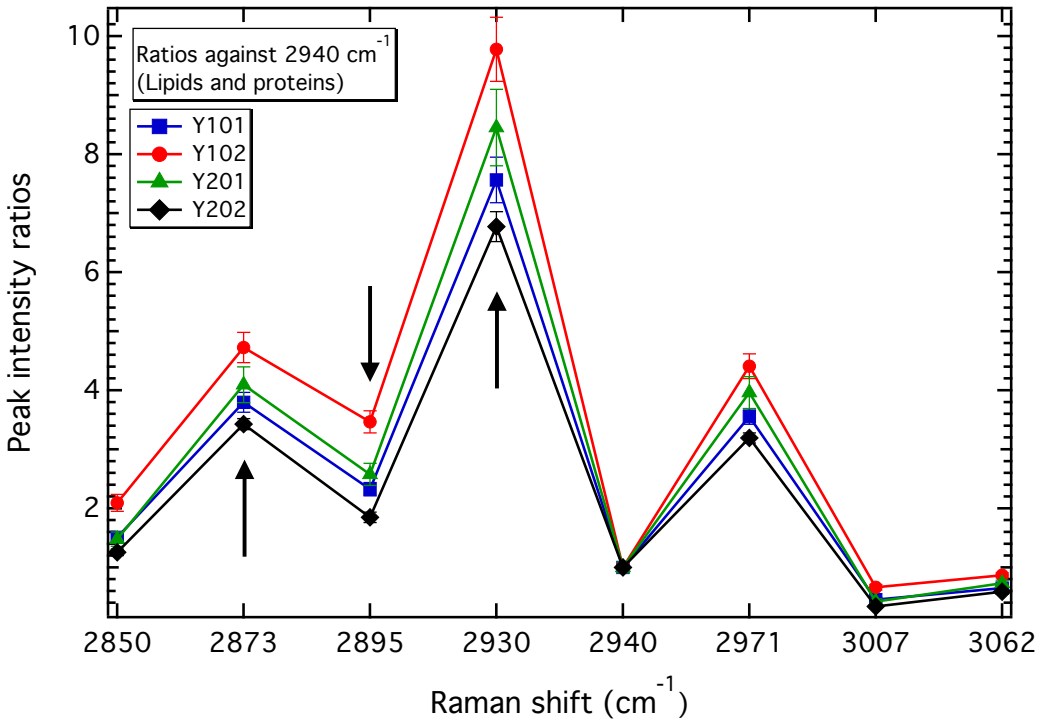


(a)

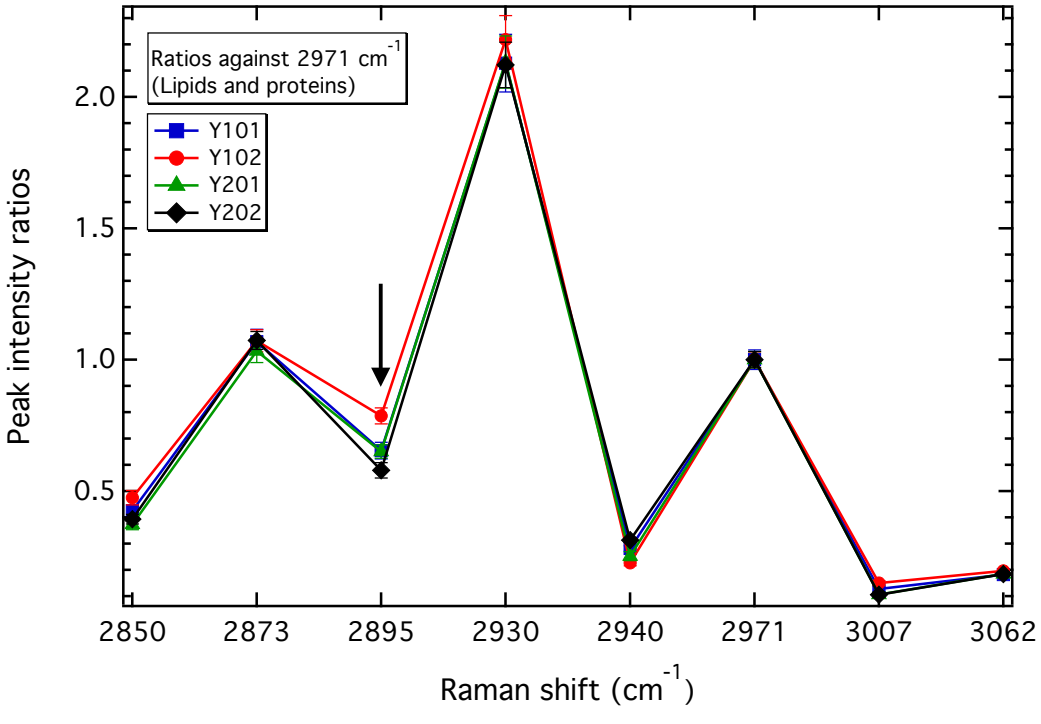


(b)

Figure C.3: PIR comparison for the live MSCs against the (a) 2873 cm^{-1} and (b) 2895 cm^{-1} peaks. The arrows show PIRs that discriminated 01s from 02s.



(a)

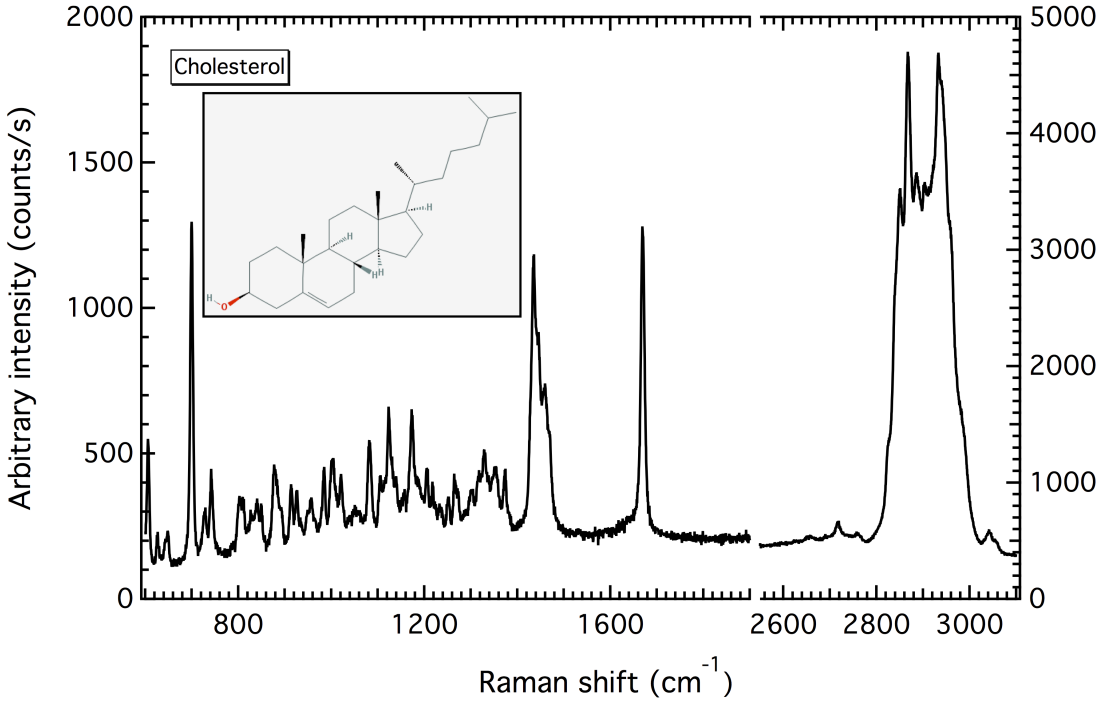


(b)

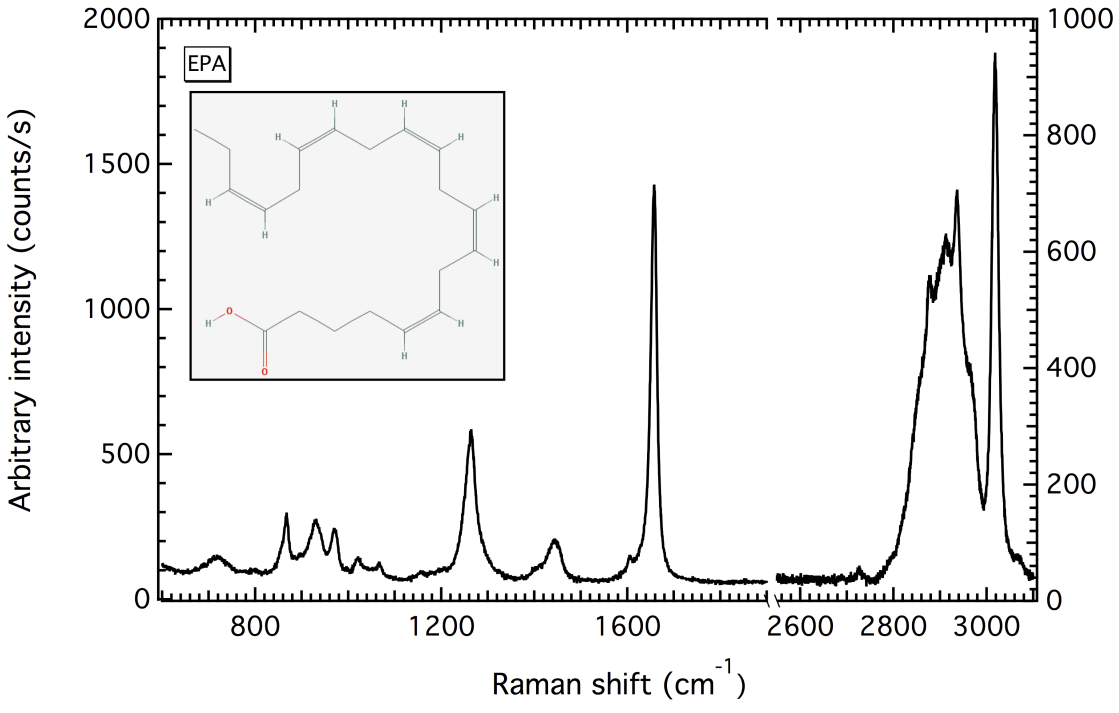
Figure C.4: PIR comparison for the live MSCs against the (a) 2940 cm^{-1} and (b) 2971 cm^{-1} peaks. The arrows show PIRs that discriminated 01s from 02s.

Appendix D

Raman spectra of pure lipids

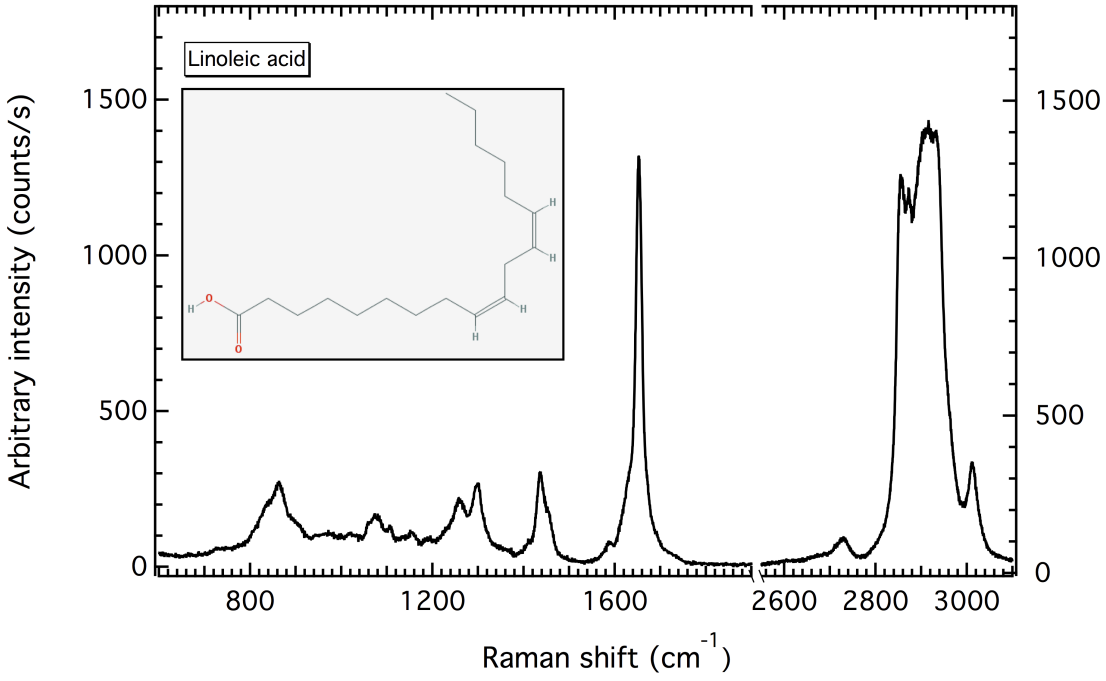


(a)

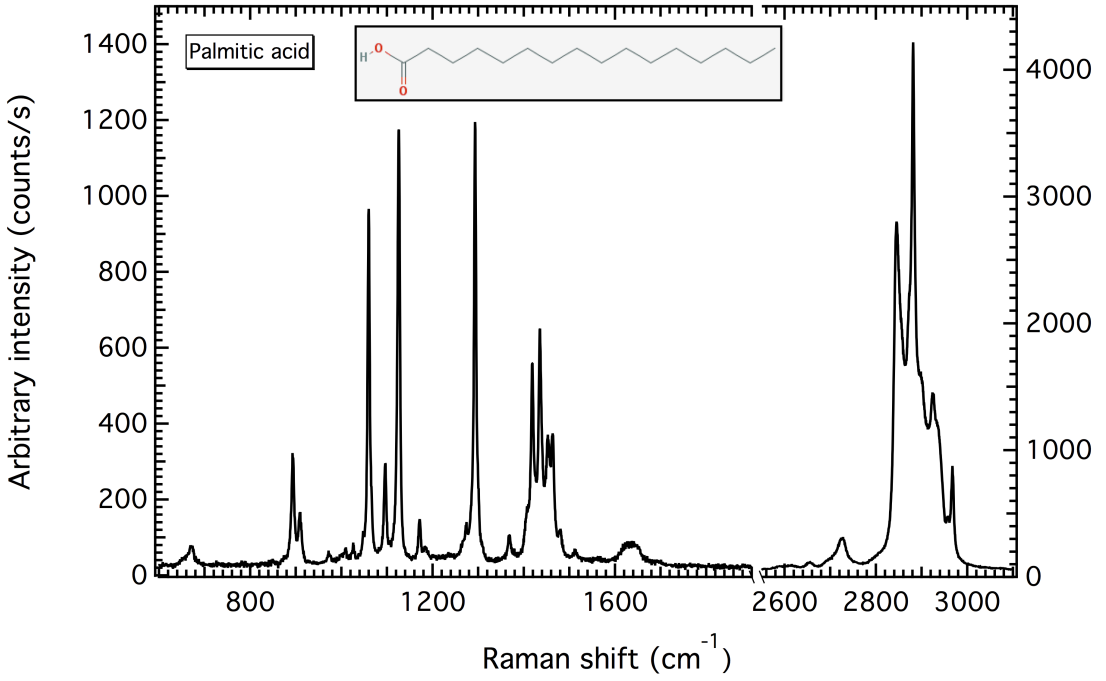


(b)

Figure D.1: Raman spectra of (a) cholesterol and (b) eicopentaeonic acid (EPA). Note the different range on the left and right Y-axes. The insert window shows the chemical structure of each lipid, as per [200].

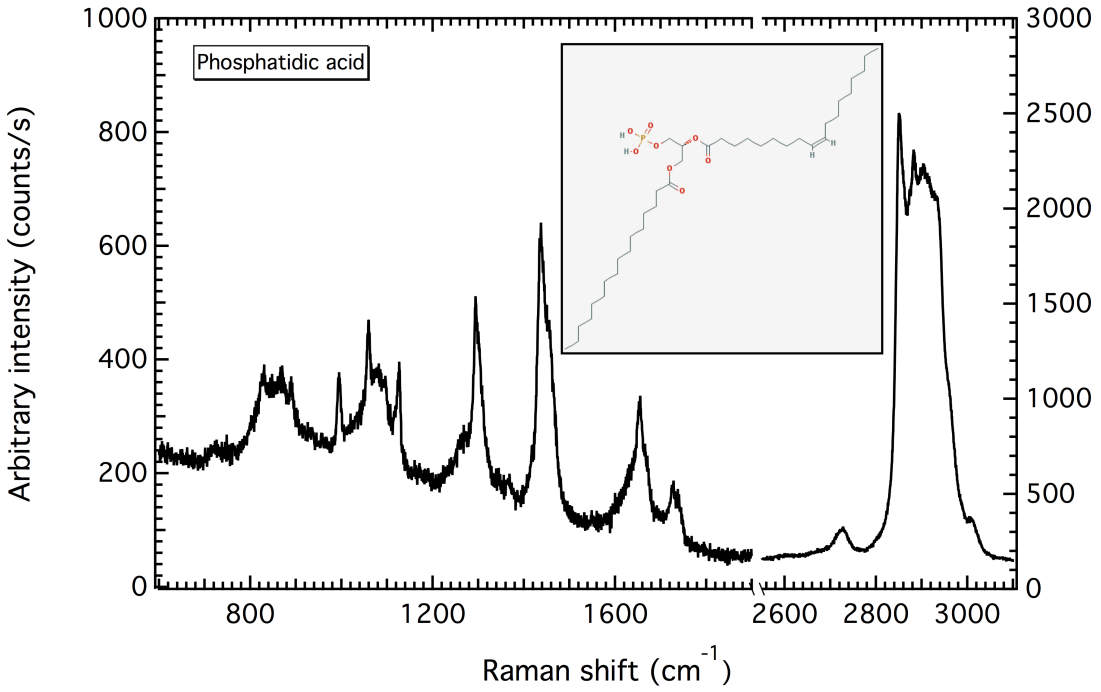


(a)

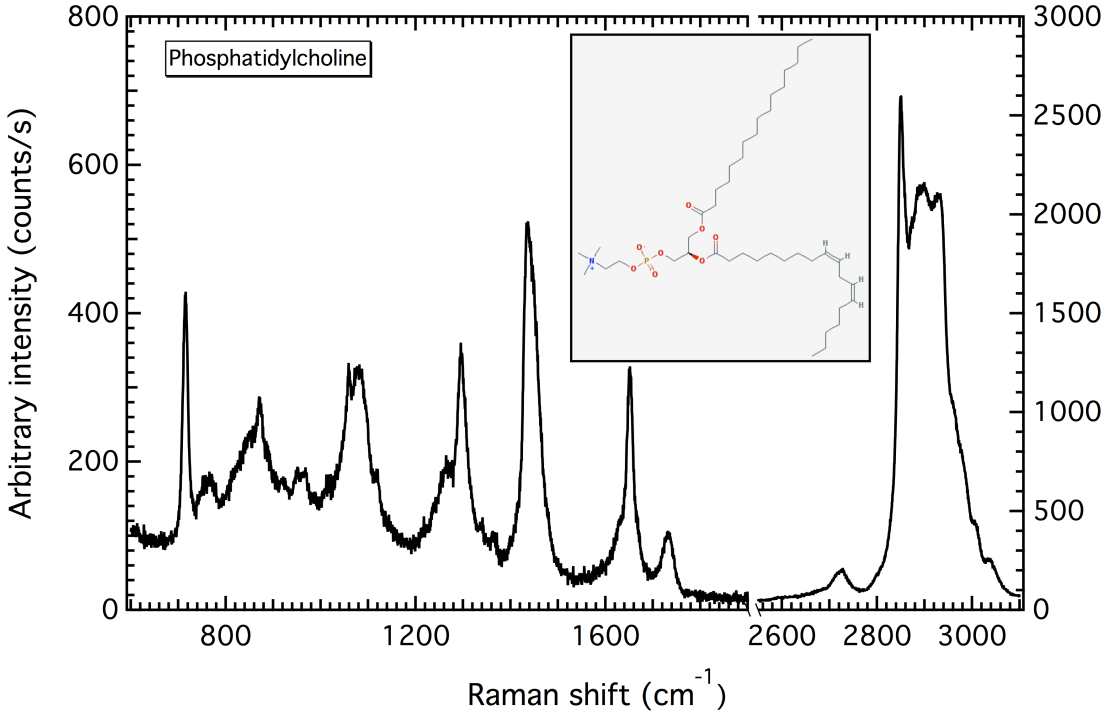


(b)

Figure D.2: Raman spectra of (a) linoleic acid and (b) palmitic acid. Note the different range on the left and right Y-axes. The insert window shows the chemical structure of each lipid, as per [200].

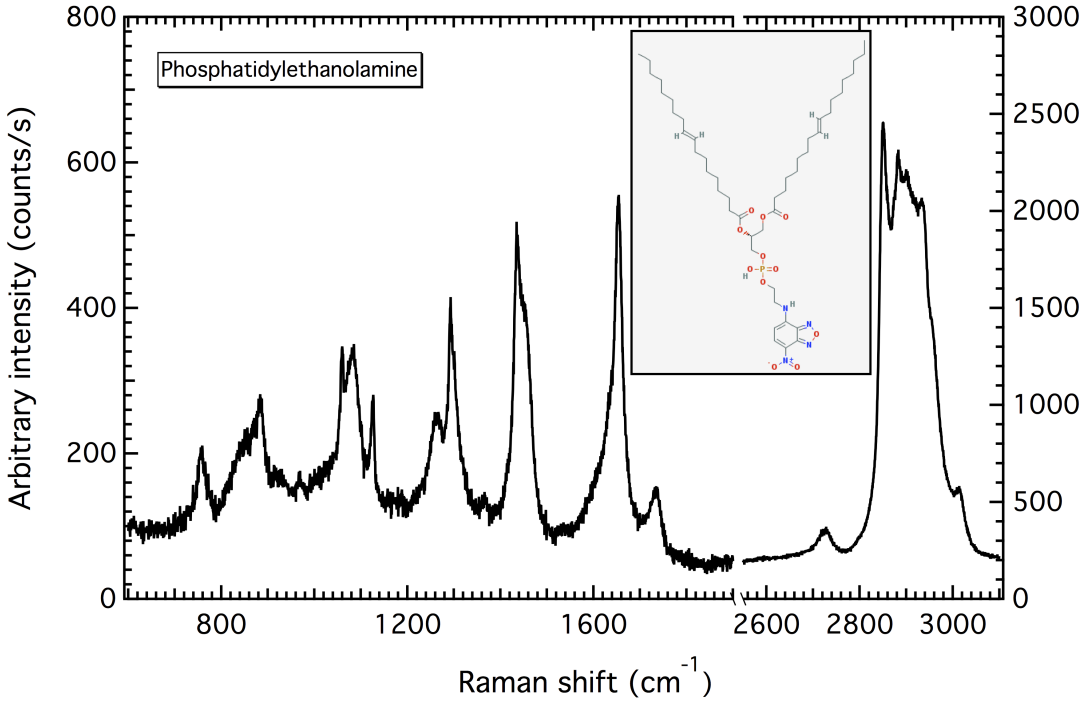


(a)

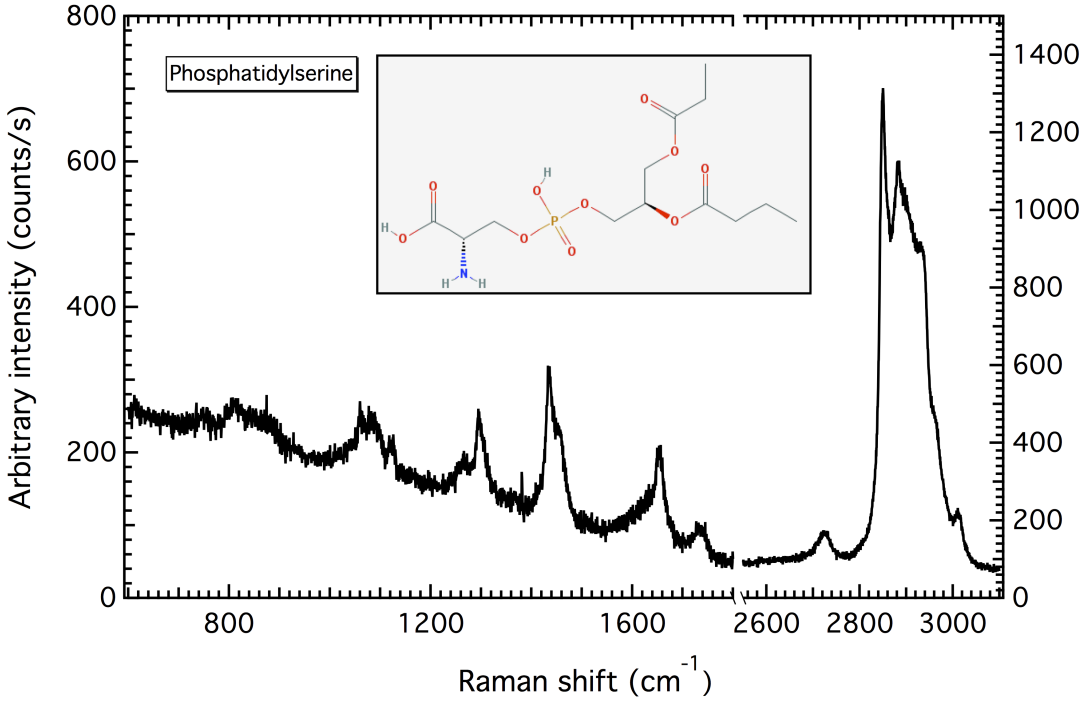


(b)

Figure D.3: Raman spectra of (a) phosphatidic acid and (b) phosphatidylcholine. Note the different range on the left and right Y-axes. The insert window shows the chemical structure of each lipid, as per [200].

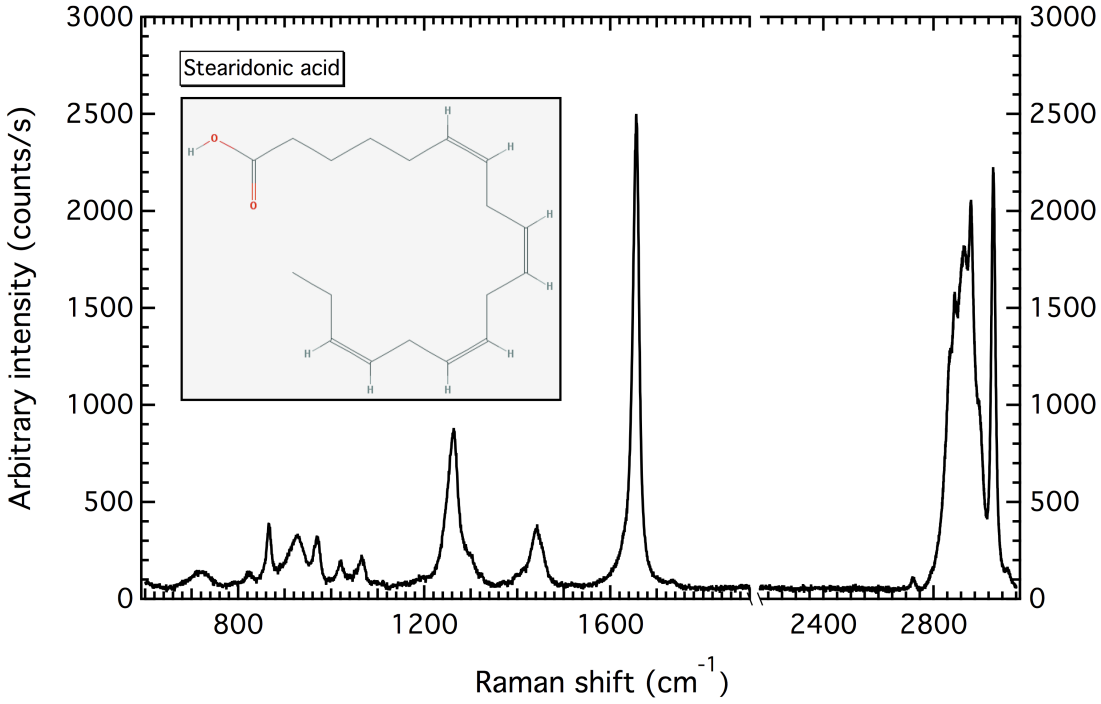


(a)

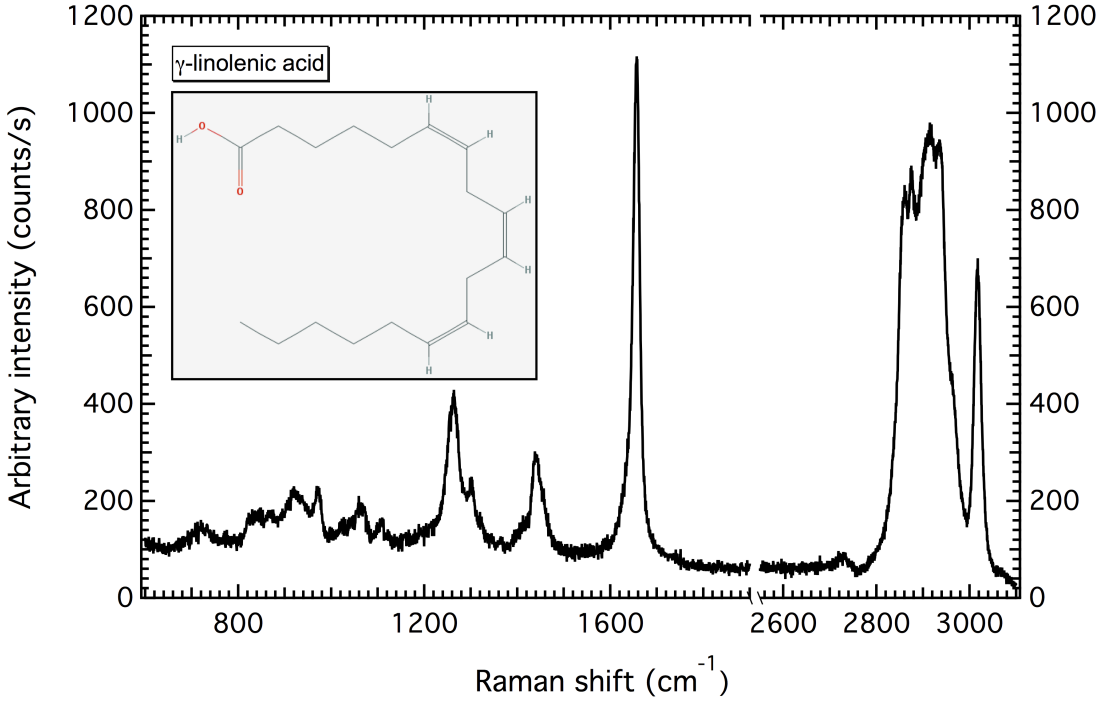


(b)

Figure D.4: Raman spectra of (a) phosphatidylethanolamine and (b) phosphatidylserine. Note the different range on the left and right Y-axes. The insert window shows the chemical structure of each lipid, as per [200].



(a)



(b)

Figure D.5: Raman spectra of (a) stearidonic acid and (b) γ-linolenic acid. Note the different range on the left and right Y-axes. The insert window shows the chemical structure of each lipid, as per [200].

Appendix E

Nuclear area, nuclear perimeter and cell ploidy

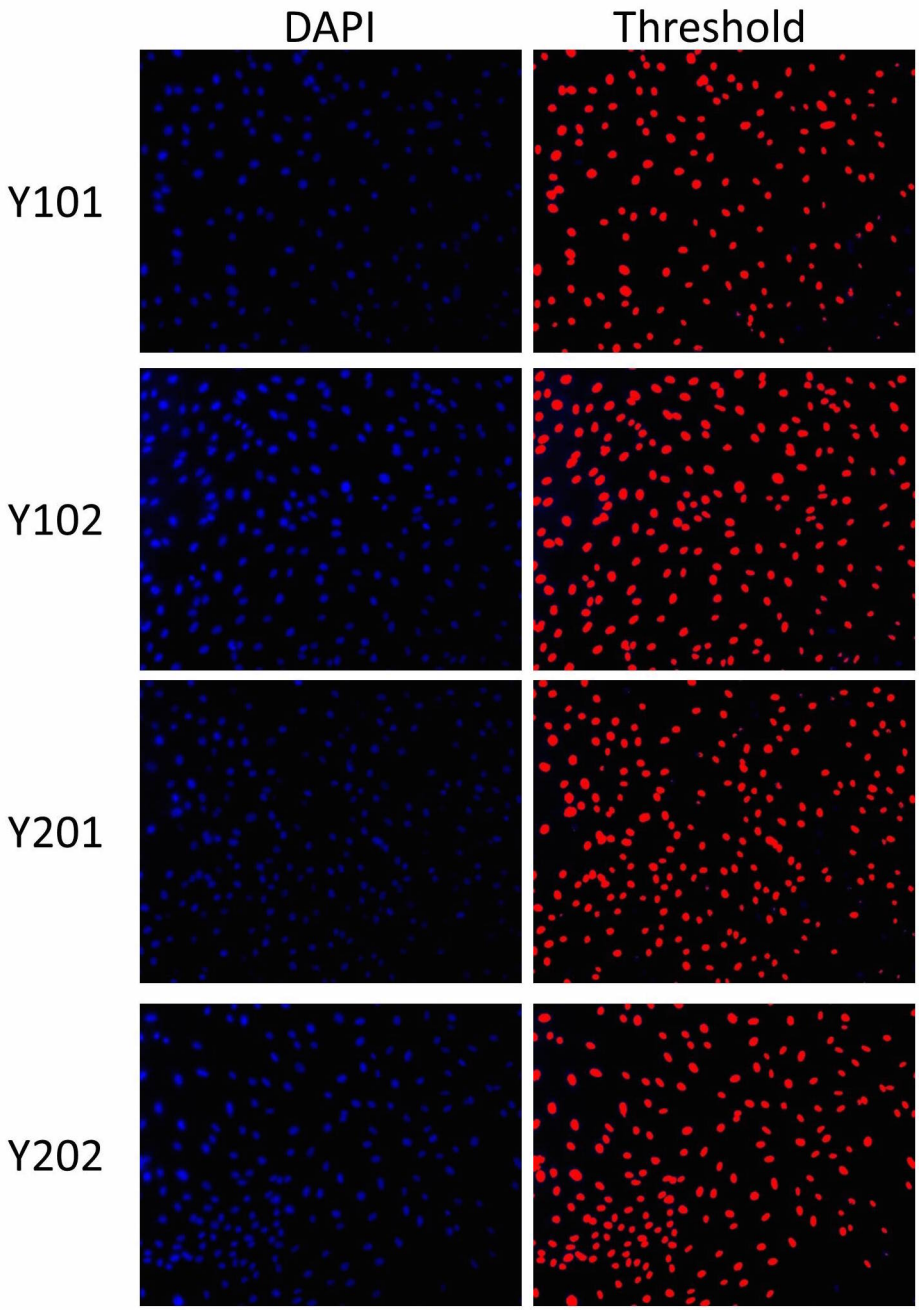
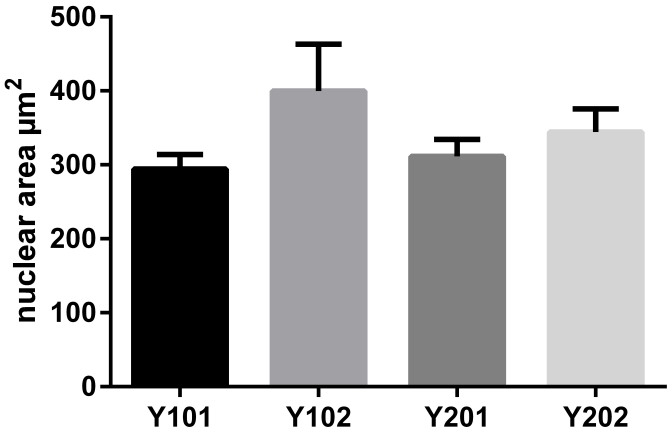
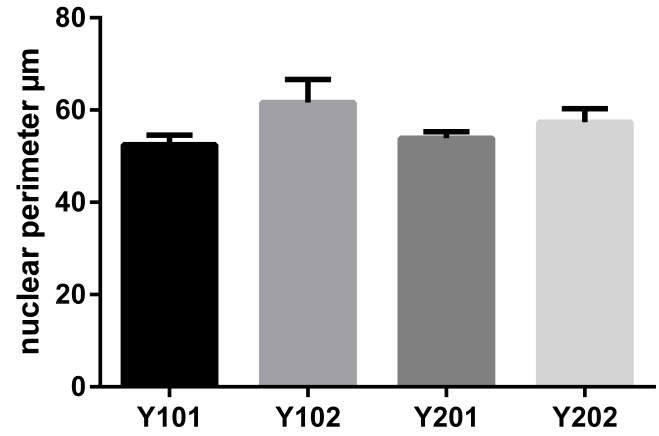


Figure E.1: Representative fluorescent images and Image J thresholded counterparts of DAPI counter-stained MSC lines. The figure shows a representative image of DAPI counter-stained nuclei of each MSC line alongside its Image J thresholded counterpart. The Image J figure selects each object that it considers to be nuclei and gives it a number. Images were screened for nucleus validity: partial-objects or objects which had been incorrectly discriminated as nuclei were excluded from area and perimeter analyses. This experiment was performed by Dr. James Fox.



(a)



(b)

Figure E.2: Comparisons of the area (a) and perimeter (b) of the nuclei of MSC lines. Nucleus-area and -perimeter of the four MSC lines were calculated using Image J processing of fluorescent images of cells counter-stained with DAPI. Average data ± standard deviation from at least 95 cells per image in triplicate is displayed. This experiment was performed by Dr. James Fox.

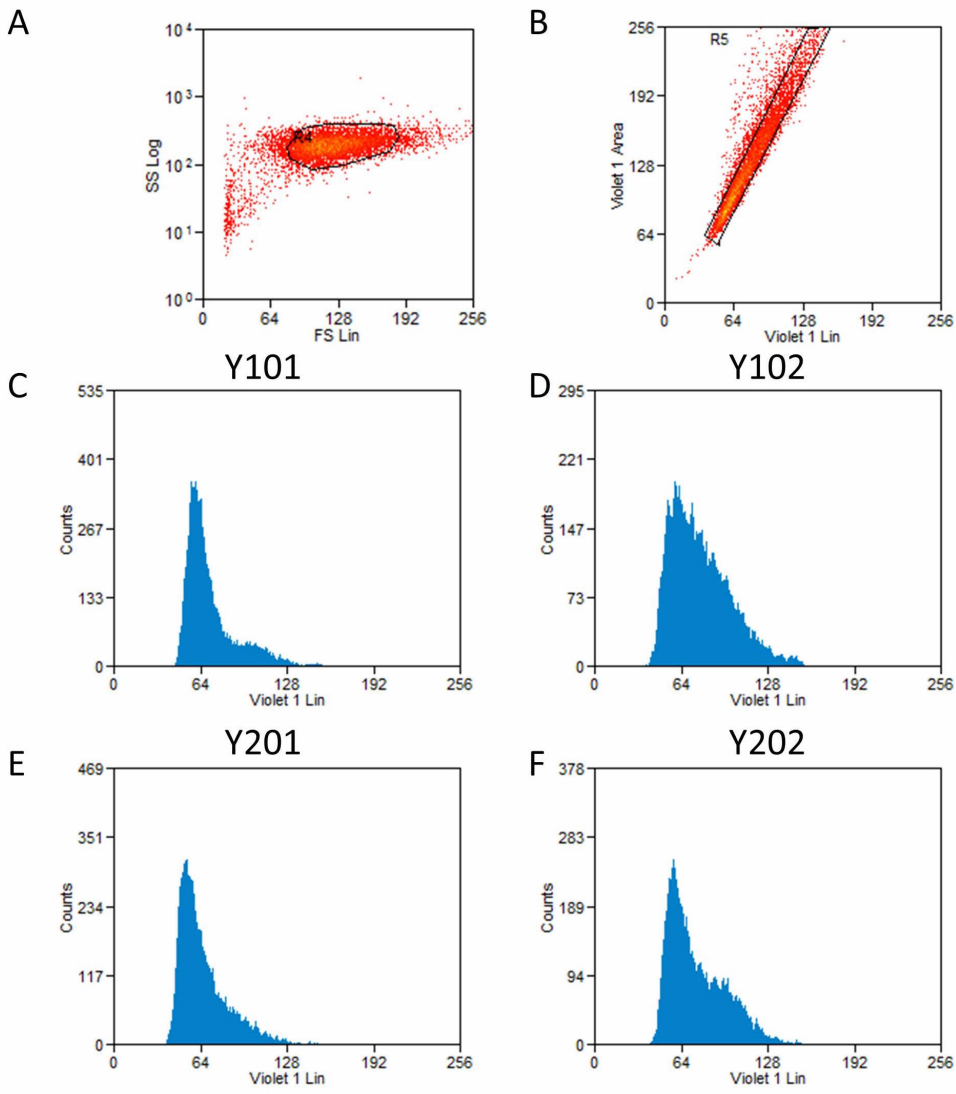


Figure E.3: All four MSC cell lines have a diploid cell cycle. They were stained with Hoechst and analysed on a CyAn flow cytometer; a gating strategy eliminated debris (a) and doublets (b) from the representative single-parameter histograms showing the linear level of fluorescence against the number of events from a population of <10,000 cells for Y101 (c), Y102 (d), Y201 (e) and Y202 (f). The major peak represents cells in G0/G1, the minor peak shows cells in G2M with those in S phase being in-between. This experiment was performed by Dr. James Fox.

Abbreviations

ADSC Adipose-derived Stem Cell

ALP Alkaline Phosphatase

BAT Brown Adipose Tissue

CARS Coherent Anti-Stokes Raman Scattering

CCD Charge-coupled Device

CFU-F Colony-Forming Unit-Fibroblast

CM Chloroform-methanol

DMEM Dulbecco's Moddified Eagle's Medium

ESC Embryonic Stem Cell

ESR Electron Spin Resonance

FBS Fetal Bovine Serum

FWHM Full Width at Half Maximum

GVHD Graft-Versus-Host Disease

GFP Green Fluorescent Protein

HA Hydroxyapatite

HBSS Hank's Balanced Salt Solution

HSC Haemotopoietic Stem Cell

IR Infrared

LDA Linear Discriminant Analysis

LNGFR Low-Affinity Nerve Growth Factor Receptor

MCAM Melanoma Cell Adhesion Molecule

MAPK Mitogen-activated protein kinase

MSC Mesenchymal Stromal Cell

NA Numerical Aperture

Nd:YAG Neodymiun:Yttrium Aluminium Garnet

NK cell Natural KIller cell

NMR Nuclear Magnetic Resonance

PBS Phosphate Buffered Saline solution

PCA Principal Component Analysis

PIR Peak Intensity Ratio

PPAR Peroxisome Proliferator-Activated Receptor

RS Raman Spectroscopy

SE Standard Error of the mean

SEM Scanning Electron Microscopy

SERS Surface Enhanced Raman Spectroscopy

SERRS Surface Enhanced-Resonance Raman Spectroscopy

SRS Stimulated Raman Scattering

TEM Transmission Electron Microscopy

TAG Triacylglycerol

UV Ultraviolet

WAT White Adipose Tissue

References

[1] C. V. Raman. A new radiation. *Indian Journal of Physics*, 2:387–398, 1928.

[2] J. J. Bohning, T. N. Misra, and M. Choudhury. The Raman effect – an international historic chemical landmark. American Chemical Society and Indian Association for the Cultivation of Science, 1998.

[3] J. R. Ferraro, K. Nakamoto, and C. W. Brown. *Introductory Raman Spectroscopy*, volume 2nd edition. Academic Press/Elsevier, Florida, USA, 2002.

[4] T. Dieing, O. Hollricher, and J. Toporski. *Confocal Raman Microscopy*, volume 158. Springer, San Diego, USA, 2011.

[5] P. Larkin. *Infrared and Raman spectroscopy – principles and spectral interpretation*, chapter Basic principles, pages 7–25. Elsevier, 2011.

[6] E. Smith and G. Dent. *Modern Raman Spectroscopy – A practical approach*, volume 1st edition. John Wiley and Sons Ltd., Chichester, England, 2005.

[7] R. L. McCreery. *Raman spectroscopy for chemical analysis*, volume 1st edition. John Wiley and Sons Ltd., New York, USA, 2000.

[8] B. Schrader. *Infrared and Raman spectroscopy*, volume 1st edition. VCH Publishers Inc., New York, USA, 1995.

[9] R. Mukhopadhyay. Raman flexes its muscles - raman instruments are advancing to novel applications and markets. *Analytical Chemistry*, pages 3265–3270, 2007.

[10] L. A. Nafie. Recent advances in linear and nonlinear Raman spectroscopy. Part V. *Journal of Raman Spectroscopy*, 42:2049–2068, 2011.

[11] R. S. Das and Y. K. Agrawal. Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy*, 57:163–176, 2011.

[12] M. Schmitt and J. Popp. Raman spectroscopy at the beginning of the twenty-first century. *Journal of Raman Spectroscopy*, 37:20–28, 2006.

[13] P. R. Bartholomew, M. D. Dyar, and J. B. Brady. The role of intensity and instrumentation sensitivity in Raman mineral identification. *Journal of Raman Spectroscopy*, 46:889–893, 2015.

[14] D. Di Genova, K. U. Hess, M. O. Chevrel, and D. B. Dingwell. Models for the estimation of Fe^{3+}/Fe_{tot} ratio in terrestrial and extraterrestrial alkali-and iron-rich silicate glasses using Raman spectroscopy. *American Mineralogist*, 101:943–952, 2016.

[15] P. Buzzini and E. Suzuki. Forensic applications of Raman spectroscopy for the *in situ* analyses of pigments and dyes in ink and paint evidence. *Journal of Raman Spectroscopy*, 47:16–27, 2016.

[16] C. A. F. de Oliveira Penido, M. T. T. Pacheco, I. K. Lednev, and L. Silveira Jr. Raman spectroscopy in forensic analysis: identification of cocaine and other illegal drugs of abuse. *Journal of Raman Spectroscopy*, 47:28–38, 2016.

[17] H. G. M. Edwards and J. M. Chalmers. *Raman Spectroscopy in Archaeology and Art History*. The Royal Society of Chemistry, Cambridge, UK, 2005.

[18] A. C. Colonese, M. Collins, A. Lucquin, M. Eustace, Y. Hancock, R. A. R. Ponzoni, A. Mora, C. Smith, P. DeBlasis, L. Figuty, V. Wesolowski, C. R. Plens, S. Eggers, D. S. E. Farias, A. Gledhill, and O. E. Craig. Long-term resilience of late holocene coastal subsistence in Southeastern South America. *PlosOne*, 9:e93854, 2014.

[19] I. M. Bell, R. J. H. Clark, and P. J. Gibbs. Raman spectroscopy library of natural and synthetic pigments (pre- 1850 AD). *Spectrochimica Acta Part A*, 53:2159–2179, 1997.

[20] B. Demarchi, S. O'Connor, A. L. L. Ponzoni, R. A. R. Ponzoni, A. Sheridan, K. Penkman, Y. Hancock, and J. Wilson. An integrated approach to the taxonomic identification of prehistoric shell ornaments. *PlosOne*, 9(6), 2014.

[21] J. A. Tunon, A. Sanchez, D. J. Parras, P. Vandenabeele, and M. Montejo. Micro-Raman spectroscopy on Iberian archaeological materials. *Journal of Raman Spectroscopy*, 2015. online version.

[22] B. Brozek-Pluska, J. Musial, R. Kordek, E. Bailo, T. Dieng, and H. Abramczyk. Raman spectroscopy and imaging: applications in human breast cancer diagnosis. *Analyst*, 137:3773–3780, 2012.

[23] P. Crow, N. Stone, C. A. Kendall, J. S. Uff, J. A. M. Farmer, H. Barr, and M. P. J. Wright. The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma *in vitro*. *British Journal of Cancer*, 89:106–108, 2003.

[24] M. C Potcoava, G. L. Futia, J. Aughenbaugh, I. R. Schlaepfer, and E. A. Gibson. Raman and coherent anti-Stokes Raman scattering microscopy studies of changes in lipid content and composition in hormone-treated breast and prostate cancer cells. *Journal of Biomedic Optics*, 19(11):111605, 2014.

[25] Y. Chiang, S. H. Wu, Y. Kuo, H Chen, A. Chiou, and O. K. Lee. Raman spectroscopy for grading of live osteosarcoma cells. *Stem Cell Research and Therapy*, 6, 2015.

[26] I. W. Schie, J. Wu, T. Weeks, M. A. Zern, J. C. Rutledge, and T. Huser. Label-free imaging and analysis of the effects of lipolysis products on primary hepatocytes. *Journal of Biophotonics*, 4(6):425–434, 2011.

[27] J. W. Chan. Recent advances in laser tweezers Raman spectroscopy (LTRS) for label-free analysis of single-cells. *Journal of Biophotonics*, 6:36–48, 2013.

[28] C. Krafft, S. Dochow, I. Latka, B. Dietzek, and J. Popp. Diagnosis and screening of cancer tissues by fiber-optic probe Raman spectroscopy. *Biomedical Spectroscopy and Imaging 1*, pages 39–55, 2012.

[29] A. Paudel, D. Raijada, and J. Rantanen. Raman spectroscopy in pharmaceutical product design. *Advanced Drug Delivery Reviews*, 89:3–20, 2015.

[30] L. Harkness, S. M. Novikov, J. Beermann, S. I Bozhevolnyi, and M. Kassem. Identification of abnormal stem cells using Raman spectroscopy. *Stem Cells and Development*, 21(12):2152–2159, 2012.

[31] M. Pudlas, D. A. C. Berrio, M. Votteler, S. Koch, S. Thude, H. Walles, and K. Schenke-Layland. Non-contact discrimination of human bone marrow-derived mesenchymal stem cells and fibroblasts using Raman spectroscopy. *Medical Laser Application*, 26:119–125, 2011.

[32] P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad, and N. Stone. The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *British Journal of Cancer*, 92:2166–2170, 2005.

[33] H. G. Schulze, S. O. Koronov, N. J. Caron, J. M. Piret, M. W. Blades, and R. F. B. Turner. Assessing differentiation status of human embryonic stem cells nonivasively using Raman microspectroscopy. *Analytical Chemestry*, 82(12):5020–5027, 2010.

[34] A. Hashimoto, L. Chiu, K. Sawada, T. Ikeuchi, K. Fujita, M. Takedachi, Y. Yamaguchi, S. Kawata, S. Murakami, and E. Tamiya. *In situ* Raman imaging of osteoblastic mineralization. *Journal of Raman Spectroscopy*, 45:157–161, 2014.

[35] Pei-San Hung, Y. Kuo, H. Chen, H. Chiang, and O. K. Lee. Detection of osteogenic differentiation by differential mineralised matrix production in mesenchymal stromal cells by Raman spectroscopy. *Plos One*, 08, 2013.

[36] J. W. Chan, D. K. Lieu, T. Huser, and R. A. Li. Label-free separation of human embryonic stem cells and their cardiac derivatives using Raman spectroscopy. *Analytical Chemistry*, 81:1324–1331, 2009.

[37] F. C. Pascut, H. T. Goh, N. Welch, L. D. Buttery, C. Denning, and I. Notingher. Nonivasive detection and imaging of molecular markers in live cardiomyocytes derived from human embryonic stem cells. *Biophysical Journal*, 100:251–259, 2011.

[38] S. Klob, B. Kampe, S. Sachse, P. Rosch, E. Straube, W. Pfister, M. Kiehntopf, and J. Popp. Culture independent Raman spectroscopy identification of urinary tract infection pathogens: a proof of principle study. *Analytical Chemistry*, 85:9610–9616, 2013.

[39] J. W. Chan. Monitoring dynamic protein expression in living *E.coli* bacterial cells by laser tweezers Raman spectroscopy. *Cytometry Part A*, 71A:468–474, 2007.

[40] C. Matthaus, B. Bird, M. Miljkovic, T. Chernenko, M. Romeo, and M. Diem. *Methods in Cell Biology - Biophysical Tools for Biologists*, volume 89, chapter Infrared

and Raman Microscopy in Cell Biology, pages 275–308. Academic Press Elsevier, 2008.

[41] I. Notingher, S. Verrier, H. Romanska, A. E. Bishop, J. M. Polak, and L. L. Hench. *In situ* characterisation of living cells by Raman spectroscopy. *Spectroscopy*, 16:43–51, 2002.

[42] I. Notingher. Raman spectroscopy cell-based biosensors. *Sensors*, 7:1343–1358, 2007.

[43] N. Uzunbajakava, A. Lenferink, Y. Kraan, E. Volokhina, G. Vrensen, J. Greve, and C. Otto. Nonresonant confocal Raman imaging of DNA and protein distribution in apoptotic cells. *Biophysical Journal*, 84:3968–3981, 2003.

[44] I. Notingher, S. Verrier, S. Haque, J. M. Polak, and L. L. Hench. Spectroscopy study of human lung epithelial cells (a549) in culture: Living cells versus dead cells. *Biopolymers*, 72:230–240, 2002.

[45] M. M. Mariani, P. Lampen, J. Popp, B. R. Wood, and V. Deckert. Impact of fixation on in vitro cell culture lines monitored with Raman spectroscopy. *Analyst*, 134:1154–1161, 2009.

[46] J. W. Chan, D. S. Taylor, and D. L. Thompson. The effect of cell fixation on the discrimination of normal and leukemia cells with laser tweezers Raman spectroscopy. *Biopolymers*, 91(2):132–139, 2008.

[47] K. Klein, A. M. Gigler, T. Aschenbrenner, R. Monetti, W. Bunk, F. Jamitzky, G. Morfill, R. W. Stark, and J. Schelgel. Label-free live-cell imaging with confocal Raman microscopy. *Biophysical Journal*, 102:360–368, 2012.

[48] H. G. Schulze, S. O. Konorov, J. M. Piret, M. W. Blades, and R. F. B Turner. Label-free imaging of mammalian cell nucleoli by Raman spectroscopy. *Analyst*, 138:3416–3423, 2013.

[49] C. Krafft, T. Knetschke, R. H. W. Funk, and R. Salzer. Identification of organelles and vesicles in single cells by Raman microspectroscopic mapping. *Vibrational Spectroscopy*, 38:85–93, 2005.

[50] C. Matthaus, T. Chernenko, J. A. Newmark, C. M. Warner, and M. Diem. Label-free detection of mitochondrial distribution in cells by nonresonant Raman microspectroscopy. *Biophysical Journal*, 93:668–673, 2007.

[51] G. P. Smith, C. M. McGoverin, S. J. Fraser, and K. C. Gordon. Raman imaging of drug delivery systems. *Advanced Drug Delivery Reviews*, 89:21–41, 2015.

[52] J. K. Pijanka, N. Stone, A. V. Rutter, N. Forsyth, G. D. Sockalingum, Y. Yang, and J. Sule-Suso. Identification of different subsets of lung cells using Raman microspectroscopy and whole cell nucleus isolation. *Analyst*, 138:5052–5058, 2013.

[53] W. Huang, S. Wu, M. Chen, L. Sun, Y. Li, M. Huang, S. Huang, R. Chen Z. Xu, and H. Zeng. Study of both fingerprint and high wavenumber Raman spectroscopy of pathological nasopharyngeal tissues. *Journal of Raman Spectrosocpy*, 46:537–544, 2015.

[54] C. Nieva, M. Marro, N. Santana-Codina, S. Rao, D. Petrov, and A. Sierra. The lipid phenotype of breast cancer cells characterised by Raman microspectroscopy: towards a stratification of malignancy. *Plos One*, 7, 2012.

[55] I. Notingher, I. Bisson, J. M. Polak, and L. L. Hench. In situ spectroscopic study of nucleic acids in differentiating embryonic stem cells. *Vibrational spectroscopy*, 35:199–203, 2004.

[56] I. Notingher, I. Bisson, A. E. Bishop, W. L. Randle, J. M. P. Polak, and L. L. Hench. In situ spectral monitoring of mRNA translation in embryonic stem cells during differentiation in vitro. *Analytical Chemistry*, 76:3185–3193, 2004.

[57] C. Krafft and J. Popp. The many facets of raman spectroscopy for biomedical analysis. *Analytical and Bioanalytical Chemistry*, 407:699–717, 2015.

[58] S. J. Webb. Laser-Raman spectroscopy of living cells. *Physics Reports*, 60(4):201–224, 1980.

[59] M. Choi, S. J. J. Kwok, and S. H. Yun. In vivo fluorescence microscopy: lessons from observing cell behavior in their native environment. *Physiolgy*, 30:40–49, 2015.

[60] J. W. Lichtman and JA. Conchello. Fluorescence microscopy. *Nature methods*, 2:910–919, 2005.

[61] K. Fujita and N. I. smith. Label-free molecular imaging of living cells. *Molecules and Cells*, 26:530–535, 2008.

[62] M. Okada, N. I. Smith, A. F. Palonpon, H. Endo, S. Kawata, M. Sodoeka, and K. Fujita. Label-free Raman observations of cytochrome c dynamics during apoptosis. *PNAS*, 109:28–32, 2012.

[63] J. Ando, A. F. Palonpon, M. Sodeoka, and K. Fujita. High-speed Raman imaging of cellular processes. *Current Opinion in Chemical Biology*, 33:16–24, 2016.

[64] A. J. Hobro, N. Pavillon, K. Fujita, M. Ozkan, C. Coban, and N. I. Smith. Label-free Raman imaging of the macrophage response to the malaria pigment hemozoin. *Analyst*, 140:2350–2359, 2015.

[65] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, and F. L. Martin. Using Raman spectroscopy to characterize biological materials. *Nature Protocols*, 11(4):664–687, 2016.

[66] A. R. Boyd, G. A. Burke, and B. J. Meenan. Monitoring cellular behaviour using Raman spectroscopy for tissue engineering and regenerative medicine. *Journal of Materials Science*, 2009.

[67] R. J. Swain, G. Jell, and M. M. Stevens. Non-invasive analysis of cell cycle dynamics in single living cells with Raman micro-spectroscopy. *Journal of Cellular Biochemestry*, 104:1427–1438, 2008.

[68] A. Downes, R. Mouras, and A. Elfick. Optical spectroscopy for nonivasive monitoring of stem cell differentiation. *Journal of Biomedicine and Biotechnology*, 2010.

[69] F. C. Pascut, H. T. Goh, V. George, C. Denning, and I. Notingher. Toward label-free Raman-activated cell sorting of cardiomyocytes derived from human embryonic stem cells. *Journal of Biomedical Optics*, 16(4), 2011.

[70] Y. H. Ong, M. Lin, and Q. Liu. Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells. *Optics Express*, 20, 2012.

[71] B. Davidson, A. A. Murray, A. Elfick, and N. Spears. Raman micro-spectroscopy can be used to investigate the developmental stage of the mouse oocyte. *Plos One*, 8:e67972, 2013.

[72] H. Wu, J. V. Volponi, A. E. Oliver, A. N. Parikh, B. A. Simmons, and S. Singh. In vivo lipidomics using single-cell raman spectroscopy. *PNAS*, 108(9):3809–3814, 2011.

[73] M. B. Fenn, V. Pappu, P. G. Georgeiv, and P. M. Pardalos. Raman spectroscopy utilizing Fisher-based feature selection combined with Support Vector Machines for the characterization of breast cell lines. *Journal of Raman Spectrosocpy*, 44:939–948, 2013.

[74] I. Notingher, C. Green, C. Dyer, E. Perkins, N. Hopkins, C. Lindsay, and L. L. Hench. Discrimination between ricin and sulphur mustard toxicity *in vitro* using Raman spectroscopy. *J. R. Soc. Interface*, 1:79–90, 2004.

[75] D. Stockholm, R. Benchaouir, J. Picot, P. Rameau, T. M. Ahn Neildez, G. Landini, C. Laplace-Builhe, and A. Paldi. The origin of phenotypic heterogeneity in a clonal cell population in vitro. *PLoS One*, 4:e394, 2007.

[76] A. A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, L. Cohen, N. Perzov, and U. Alon. Dynamic proteomic of individual cancer cells in response do a drug. *Science*, 322:11511–1516, 2008.

[77] S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M. A. Fischbach, K. Wong, K. Brandstetter,

B. Wittner, S. Ramaswamy, M. Classon, and J. Settleman. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141:69–80, 2010.

[78] M. F. Pittenger, A. M. Mackay, S. C. Beck, R. K. Jaiswal, R. Douglas, J. D. Mosca, M. A. Moorman, D. W. Simonetti, S. Craig, and D. R. Marshak. Multilineage potential of adult human mesenchymal stem cells. *Science*, 284(5411):143–147, 1999.

[79] A. Muraglia, R. Cancedda, and R. Quarto. Clonal mesenchymal progenitors from human bone marrow differentiate in vitro according to a hierarchial model. *Journal of Cell Science*, 113:1161–1166, 2000.

[80] K. C. Russell, D. G. Phinney, M. R. Lacey, B. L. Barrilleaux, K. R. Meyertholen, and K. C. O'Connor. In vitro high-capacity assay to quantify the clonal heterogeneity in trilineage potential of mesenchymal stem cells reveals a complex hierarchy of lineage commitment. *Stem Cells*, 28:788–798, 2000.

[81] T. Okamoto, T. Aoyama, T. Nakayama, T. Nakamata, T. Hosaka, K. Nishijo, T. Nakamura, T. Kiyono, and J. Toguchida. Clonal heterogeneity in differentiation potential of immortalised human mesenchymal stem cells. *Biochemical and Biophysical Research Communications*, 295:354–361, 2002.

[82] S. Altschuler and L. F. Wu. Cellular heterogeneity: do differences make a difference? *Cell*, 141:559–563, 2010.

[83] R. L. Fernandes, M. Nierychlo, L. Lundin, A. E. Pedersen, P. E. P. Tellez, A. Dutta, M. Carlquist, A. Bolic, D. Schapper, A. C. Brunetti, S. Helmark, A. L. Heins, A. D. Jensen, I. Nopens, K. Rottwitt, N. Szita, J. D. van Elsas, P. H. Nielsen, J. Martinussen, S. J. Sorensen, A. E. Lantz, and K. V Gernaey. Experimental

methods and modeling techniques for description of cell population heterogeneity. *Biotechnology Advances*, 29:575–599, 2011.

[84] S. James, J. Fox, F. Afsari, J. Lee, S. Clough, C. Knight, J. Ashmore, P. Ashton, O. Preham, M. Hoogduijn, R. A. R. Ponzoni, Y. Hancock, M. Coles, and P. Genever. Multiparameter analysis of human bone marrow stromal cells identifies distinct immunomodulatory and differentiation-competent subtypes. *Stem Cell Reports*, 4:1004–1015, 2015.

[85] S. V. Avery. Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews*, 4:577–587, 2006.

[86] A. Carstairs and P. G. Genever. Stem cell treatment for musculoskeletal disease. *Current Opinion in Pharmacology*, 16:1–6, 2014.

[87] J. M. Fox and P. G. Genever. Use of adult stem cells for orthopedic regenerative medicine applications. *Cell and Tissue Transplantation and Therapy*, 6:19–25, 2014.

[88] S. Roberts, P. Genever, A. McCaskie, and C. De Bari. Prospects of stem cell therapy in osteoarthritis. *Regenerative medicine*, 6(3):351–66, 2011.

[89] A. J. Friedenstein, I. I. Piatetzky-Shapiro, and K. V. Petrakova. Osteogenesis in transplants of bone marrow cells. *Journal of Embryology and Experimental Morphology*, 16:381–390, 1966.

[90] A. J. Friedenstein, R. K. Chailakhjan, and K. S. Lalykina. The development of fibroblast colonies in monolayer cultures of guinea-pig bone marrow and spleen cells. *Cell Tissue Kinetics*, 3:393–403, 1970.

[91] H. Castro-Malaspina, R. E. Gay, G. Resnick, N. Kapoor, P. Meyers, D. Chiarieri, S. McKenzie, H. E. Broxmeyer, and M. A. S. Moore. Characterization of human

bone marrow fibroblast colony-forming cells (CFU-F) and their progeny. *Blood*, 56(2):289–301, 1980.

[92] A. I. Caplan. Mesenchymal stem cells. *Journal of Orthopaedic Research*, 9:641–650, 1991.

[93] M. Pevsner-Fischer, S. Levin, and D. Zipori. The origins of mesenchymal stromal cell heterogeneity. *Stem Cell Reviews and Reports*, 7:560–568, 2011.

[94] N. Tremain, J. Korkko, D. Ibberson, G. C. Kopen, C. DiGirolamo, and D. G. Phinney. MicroSAGE analysis of 2,353 expressed genes in a single cell-derived colony of undifferentiated human mesenchymal stem cells reveals mRNAs of multiple cell lineages. *Stem Cells*, 19:408–418, 2001.

[95] E. M. Horwitz, K. Le Blanc, M. Dominici, I. Mueller, I. Slaper-Cortenbach, F. C. Marini, R. J. Deans, D. S Krause, and A. Keating. Clarification of the nomenclature for MSC: the International Society for Cellular Therapy position statement. *Cytotherapy*, 7(5):393–395, 2005.

[96] M. Dominici, K. Le Blanc, I. Mueller, I. Slaper-Cortenbach, F. C. Marini, D. S. Krause, R. J. Deans, A. Keating, D. J. Prockop, and E. M. Horwitz. Minimal criteria for defining multipotent mesenchymal stromal cells. the International Society for Cellular Therapy position statement. *Cytotherapy*, 8:315–317, 2006.

[97] E. M. Horwitz and M. Dominici. How do mesenchymal stromal cells exert their therapeutic benefit? *Cytotherapy*, 10(8):771–774, 2008.

[98] E. A. Jones, S. E. Kinsey, A. English, R. A. Jones, L. Straszynkski, D. M. Meredith, A. F. Markham, A. Jack, P. Emery, and D. McGonagle. Isolation and characteri-

zation of bone marrow multipotential mesenchymal progenitor cells. *Arthritis and Rheumatism*, 46(12):3349–3360, 2002.

[99] B. Sacchetti, A. Funari, S. Michienzi, S. Di Cesare, S. Piersanti, I. Saggio, E. Tagliafico, S. Ferrari, P. G. Robey, M. Riminucci, and P. Bianco. Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. *Cell*, 131:324–336, 2007.

[100] C. C. Oliveira, R. M. Dufloth, and K. I. R. Coelho. Absence of wharton's jelly: case report. *Brazilian Journal of Pathology and Laboratory Medicine*, 50(6):452–455, 2014.

[101] R. Hine and E. Martin. *Oxford Dictionary of Biology*. Oxford University Press, 7 edition, 2015. Electronic version.

[102] A. Keating. Mesenchymal stromal cells: new directions. *Cell Stem Cell*, 10:709–716, 2012.

[103] S. Kern, H. Eichler, J. Stoeve, H. Kluter, and K. Bieback. Comparative analysis of mesenchymal stem cells from bone marrow, umbilical cord blood, or adipose tissue. *Stem Cells*, 24:1294–1301, 2006.

[104] W. Wagner, F. Wein, A. Seckinger, M. Frankhauser, U. Wirkner, U. Krause, J. Blake, C. Schwager, V. Eckstein, W. Ansorge, and A. D. Ho. Comparative characteristics of mesenchymal stem cells from human bone marrow, adipose tissue and umbilical cord blood. *Experimental Hematology*, 33:1402–1416, 2005.

[105] P. Bianco and P. G. Robey. Skeletal stem cells. *Development*, 142:1023–1027, 2015.

[106] P. Bianco, P. G. Robey, and P. J. Simmons. Mesenchymal stem cells: revisiting history, concepts and assays. *Cell Stem Cell*, 2:313–319, 2008.

[107] C. Mason and P. Dunnill. A brief definition of regenerative medicine. *Regenerative Medicine*, 3(1):1–5, 2008.

[108] E. Martin. *Oxford Concise Medical Dictionary*. Oxford University Press, 9 edition, 2016. Electronic version.

[109] J. Lackie. *Oxford Dictionary of Biomedicine*. Oxford University Press, 1 edition, 2015. Electronic version.

[110] F. Rastegar, D. Shenaq, J. Huang, W. Zhang, B. Zhang, B. He, L. Chen, G. Zuo, Q. Luo, Q. Shi, E. R. Wagner, E. Huang, Y. Gao, J. Gao, S. H. Kim, J. Zhou, Y. Bi, Y. Su, G. Zhu, J. Luo, X. Luo, J. Qin, R. R. Reid, H. H. Luu, R. C. Haydon, Z. Deng, and T. He. Mesenchymal stem cells: mollecular characteristics and clinical applications. *World Journal of Stem Cells*, 2(4):67–80, 2010.

[111] *https://clinicaltrials.gov*. Accessed: Oct/2016.

[112] A. D. Woolf and B. Pfleger. Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization*, 81:646–656, 2003.

[113] R. Cammack, T. Atwood, P. Campbell, H. Parish, A. Smith, F. Vella, and J. Stirling. *Oxford Dictionary of Biochemistry and Molecular Biology*. Oxford University Press, 2 edition, 2008. Electronic version.

[114] R. Yanez, M. L. Lamana, J. Garcia-Castro, I. Colmenero, M. Ramirez, and J. A. Bueren. Adipose tissue-derived mesenchymal stem cells have in vivo immunosuppressive properties applicable for the control of the Graft-Versus-Host disease. *Stem Cells*, 24:2582–2591, 2006.

[115] Y. Tian, Y. B. Deng, Y. J. Huang, and Y. Wang. Bone marrow-derived mesenchymal stem cells decrease acute graft-versus-host disease after allogenic hematopoietic stem cells transplantation. *Immunological Investigations*, 37:29–42, 2008.

[116] S. Makino, K. Fukuda, S. Miyoshi, F. Konishi, H. Kodama, J. Pan, M. Sano, T. Takahashi, S. Hori, H. abe, J. Hata, A. Umezawa, and S. Ogawa. Cardiomyocytes can be generated from marrow stromal cells *in vitro*. *Journal of Clinical Investigation*, 103:697–705, 1999.

[117] M. Shiota, T. Heike, M. Haruyama, S. Baba, A. Tsuchiya, H. Fujino, H. Kobayashi, T. Kato, K. Umeda, M. Yoshimoto, and T. Nakahata. Isolation and characterization of bone marrow-derived mesenchymal progenitor cells with myogenic and neuronal properties. *Experimental Cell Research*, 313:1008–1023, 2007.

[118] M. F. Pittenger and B. J. Martin. Mesenchymal stem cells and their potential as cardiac therapeutics. *Circulation Research*, 95:9–20, 2004.

[119] D. G. Phinney. Functional heterogeneity of mesenchymal stem cells: Implications for cell therapy. *Journal of Cellular Biochemistry*, 113:2806–2812, 2012.

[120] J. Ylostalo, N. Bazhanov, and D. J. Prockop. Reversible commitment to differentiation by human multipotent stromal cells in single-cell-derived colonies. *Experimental Hematology*, 36:1390–1402, 2008.

[121] B. G. Sengers, J. I. Dawson, and R. O.C. Oreffo. Characterisation of human bone marrow stromal cell heterogeneity for skeletal regeneration strategies using a two-stage colony assay and computational modelling. *Bone*, 46:496–503, 2010.

[122] S. A. Kuznetsov, P. H. Krebsbach, K. Satomura, J. Kerr, M. Riminucci, D. Benayahu, and P. G. Robey. Single-colony derived strains of human bone marrow

stromal fibroblasts form bone after transplantation *in vivo*. *Journal of Bone and Mineral Research*, 12(9):1335–1347, 1997.

[123] A. Banfi, A. Muraglia, B. Dozin, M. Mastrogiacomo, R. Cancedda, and R. Quarto. Proliferation kinetics and differentiation potential of *ex vivo* expanded human bone marrow stromal cells: implications for their use in cell therapy. *Experimental Hematology*, 28:707–715, 2000.

[124] C. M. DiGirolamo, D. Stokes, D. Colter, D. G. Phinney, R. Class, and D. J. Prockop. Propagation and senescence of human marrow stromal cells in culture: a simple colony-forming assay identifies samples with greatest potential to propagate and differentiate. *British Journal of Haematology*, 107:275–281, 1999.

[125] C. C. Moura, R. S Tare, R. O. C. Oreffo, and S. Mahajan. Raman spectroscopy and coherent anti-Stokes Raman scattering imaging: prospective tools for monitoring skeletal cells and skeletal regeneration. *Journal of the Royal Society Interface*, 13:20160182, 2016.

[126] L. L. McManus, G. A. Burke, M. M. McCafferty, P. O'Hare, M. Modreanu, A. R. Boyd, and B. J. Meenan. Raman spectroscopy monitoring of the cell osteogenic differentiation of human mesenchymal stem cells. *Analyst*, 136:2471–2481, 2011.

[127] E. Gentleman, R. J. Swain, N. D. Evans, S. Boonrungsiman, G. Jell, M. D. Ball, T. A. V. Shean, M. L. Oyen, A. Porter, and M. M. Stevens. Comparative materials differences revealed in engineered bone as a function of cell-specific differentiation. *Nature Materials*, 8:763–770, 2009.

[128] M. D. Morris and G. S. Mandair. Raman assessment of bone quality. *Clinical Orthopaedics and Related Research*, 469:2160–2169, 2011.

[129] C. Tarnowski, M. A. Ignelzi Jr, and M. D. Morris. Mineralization of developing mouse calvaria as revealed by Raman microspectroscopy. *Journal of Bone and Mineral Research*, 17(6):1118–1126, 2002.

[130] G. Peled, K. Tai, D. Sheyn, Y. Zilberman, S. Kumbar, L. S. Nair, C. T. Laurencin, G. Gazit, and C. Ortiz. Structural and nanoindentation studies of stem cell-based tissue-engineered bone. *Journal of Biomechanics*, 40:399–411, 2007.

[131] H. K. Chiang, F. Peng, S. Hung, and Y. Feng. *In situ* Raman spectroscopy monitoring of hydroxyapatite as human mesenchymal stem cells differentiate into osteoblasts. *Journal of Raman spectroscopy*, 40:546–549, 2009.

[132] O. Akkus, F. Adar, and M. B. Schaffler. Age-related changes in physicochemical properties of mineral crystals are related to impared mechanical function of cortical bone. *Bone*, 34:443–453, 2004.

[133] G. Penel, G. Leroy, C. Rey, and E. Bres. Micro-Raman spectral study of the po_4 and co_3 vibrational modes in synthetic and biological apatites. *Calcified Tissue International*, 1998.

[134] A. Ghita, F. C. Pascut, V. Sottile, and I. Notingher. Monitoring the mineralisation of bone nodules *in vitro* by space- and time-resolved Raman micro spectroscopy. *Analyst*, 139:55–58, 2014.

[135] A. Hashimoto, Y. Yamaguchi, L. Chiu, C. Morimoto, K. Fujita, M. Takedachi, S. Murakami, and E. Tamiya. Time-lapse Raman imaging of osteoblast differentiation. *Scientific Reports*, 5(12529):1–8, 2015.

[136] A. Carden, R. M. Rajachar, M. D. Morris, and D. H. Kohn. Ultrastructural changes accompanying the mechanical deformation of bone tissue: a Raman imaging study. *Calcified Tissue International*, 72:166–175, 2003.

[137] A. Downes, R. Mouras, P. Bragnaninchi, and A. Elfick. Raman spectroscopy and CARS microscopy of stem cells and their derivatives. *Journal of Raman spectroscopy*, 42:1864–1870, 2011.

[138] A. Mitchell, L. Ashton, X. B. Yang, R. Goodacre, A. Smith, and J. Kirkham. Detection of early stage changes associated with adipogenesis using Raman spectroscopy under aseptic conditions. *Cytometry Part A*, 87A:1012–1019, 2015.

[139] R. Mouras, P. O. Bagnaninchi, A. R. Downes, and A. P. D. Elfick. Label-free assessment of adipose-derived stem cell differentiation using coherent anti-Stokes Raman scattering and multiphoton microscopy. *Journal of Biomedical Optics*, 17(11):116011, 2012.

[140] J. P. Smus, C. C. Moura, E. McMorrow, R. S. Tare, R. O. C. Oreffo, and S. Mahajan. Tracking adipogenic differentiation of skeletal stem cells by label-free chemically selective imaging. *Royal Society of Chemistry*, 6:7089–7096, 2015.

[141] J. K. Pijanka, D. Kumar, T. Dale, I. Yousef, G. Parkes, V. Untereiner, Y. Yang, P. Dumas, D. Collins, M. Manfait, G. D. Sockalingum, N. R. Forsyth, and J. Sule-Suso. Vibrational spectroscopy differentiates between multi potent and pluripotent stem cells. *Analyst*, 135:3126–3132, 2010.

[142] H. E. Keller. *Handbook of biological confocal microscopy*, chapter Objective lenses for confocal microscopy. Plenum Press, 1990.

[143] P. R. Griffiths and E. V. Miseo. *Infrared and Raman spectroscopy imaging*, chapter Infrared and Raman instrumentation for mapping and imaging. Wiley-VCH, 2 edition, 2014.

[144] Ian R. Lewis and Howell G. M. Edwards, editors. *Handbook of Raman Spectroscopy*. Marcel Dekker, 2001.

[145] M. Muller. *Confocal fluorescence microscopy*, volume 2nd edition. SPIE Press, Washington, USA, 2006.

[146] A. Roorda, D. T. Miller, and J. Christou. *Adaptive optics for vision science*, chapter Strategies for high-resolution retinal imaging. Willey - Interscience, 2006.

[147] B&WTek spectrometers. Spectrometer knowledge - slit. `http://bwtek.com/spectrometer-part-1-the-slit`. Online; accessed: October/2016.

[148] V. Achille, M. Mantelli, G. Arrigo, F. Novara, M. A. Avanzini, M. E. Bernardo, O. Zuffardi, G. Barosi, M. Zecca, and R. Maccario. Cell-cycle phases and genetic profile of bone marrow-derived mesenchymal stromal cells expanded in vitro from healthy donors. *Journal of Cellular Biochemistry*, 112:1817–1821, 2011.

[149] D. A. Cook, S. W. Fellgett, M. E. Pownall, P. J. O'Shea, and P. G. Genever. Wnt-dependent osteogenic commitment of bone marrow stromal cells using novel GSK3β inhibitor. *Stem Cell Research*, 12:415–427, 2014.

[150] S. W. Perry, L. G. Epstein, and H. A. Gelbard. *In situ* trypan blue staining of monolayer cell cultures for permanent fixation and mounting. *BioTechniques*, 22(6):1499–1507, 1997.

[151] J. Hodgetts. Raman investigations of graphene and other organics. Master's thesis, University of York, 2013.

[152] P. Candeloro, E. Grande, R. Raimondo, D. Di Mascolo, F. Gentile, M. L. Coluccio, G. Perozziello, N. Malara, M. Francardi, and E. Di Frabizio. Raman database of amino acid solutions: A critical study of extended multiplicative signal correction. *Analyst*, 138:7331–7340, 2013.

[153] P. K. Giri and J. Banerjee. *Introduction to statistics*, volume 6th edition. Academic Publishers, Kolkata, India, 2008.

[154] Z. Movasaghi, S. Rehman, and I. U. Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 2007.

[155] J. W. Chan, D. S. Taylor, S. M. Lane, T. Zwerdling, J. Tuscano, and T. Huser. Nondestructive identification of individual leukaemia cells by laser trapping Raman spectroscopy. *Analytical Chemistry*, 80:2180–2187, 2008.

[156] S. Verrier, I. Notingher, J. M. Polak, and L. L. Hench. *In situ* monitoring of cell death using Raman microspectroscopy. *Biopolymers*, 74:157–162, 2004.

[157] M. Meurens, J. Wallon, J.Tong, H. Noel, and J. Haot. Breast cancer detection by Fourier transformed infrared spectrometry. *Vibrational Spectroscopy*, 10:341–346, 1996.

[158] P. T. T. Wong, R. K. Wong, T. A. Caputo, T. A. Godwin, and B. Rigas. Infrared spectroscopy of exfoliated human cervical cells: Evidence of extensive structural changes during carcinogenesis. *Proc. Natl. Acad. Sci. USA*, 88:10988–10992, 1991.

[159] A. Mitchell, L. Ashton, X. B. Yang, R. Goodacre, M. J. Tomlinson, A. Smith, and J. Kirkham. Aseptic Raman spectroscopy can detect changes associated with the culture of human dental pulp stromal cells in osteoinductive culture. *Analyst*, 2015.

[160] I. T. Jolliffe. *Principal component analysis*, volume 1st edition. Springer Science and Bussiness, New York, USA, 1986.

[161] I. Notingher, G. Jell, U. Lohbauer, V. Salih, and L. L. Hench. In situ non-invasive spectral discrimination between bone cell phenotypes used in tissue engineering. *Journal of Cellular Biochemistry*, 92:1180–1192, 2004.

[162] I. Notingher, G. Jell, P. L. Notingher, I. Bisson, O. Tsigkou, J. M. Polak, M. M. Stevens, and L. L. Hench. Multivariate analysis of Raman spectra for in vitro non-invasive studies of living cells. *Journal of Molecular Structure*, 744:179–185, 2005.

[163] A. Nijssen, T. C. B. Schutt, F. Heule, P. J. Caspers, D. P. Hayes, M. H. A. Neumann, and G. J. Puppels. Discriminating basal cell carcinoma from its surrounding tissue by Raman spectroscopy. *The Journal of Investigative Dermatology*, 119(1):64–69, 2002.

[164] A. B. Zoladek, R. K. Johal, S. Garcia-Nieto, F. Pascut, K. M. Shakesheff, A. M. Ghaemmaghami, and I. Notingher. Label-free molecular imaging of immunological synapses between dendritic and t cells by raman micro-spectroscopy. *Analyst*, 135:3025–3212, 2010.

[165] D. G. Morrison. On the interpretation of discriminant analysis. *Journal of Marketing Research*, 6:156–163, 1969.

[166] C. A. Lieber and M. H. Kabeer. Characterization of pediatric wilms' tumor using Raman and fluorescence spectroscopies. *Journal of Pediatric Surgery*, 45:549–554, 2010.

[167] P. Bianco, M. Riminucci, S. Gronthos, and P. G. Robey. Bone marrow stromal stem cells: Nature, biology and potential applications. *Stem Cells*, 19:180–192, 2001.

[168] J. W. Chan, D. S. Taylor, T. Zwerdling, S. M. Lane, K. Lhara, and T. Huser. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophysical Journal*, 90:648–656, 2006.

[169] B. Prescott, W. Steinmets, and G. J. Thomas Jr. Characterization of DNA structures by laser Raman spectroscopy. *Biopolymers*, 23:235–356, 1984.

[170] I. Puhlev, N. Guo, D. R. Brown, and F. Levine. Desiccation tolerance in human cells. *Cryobiology*, 42:207–217, 2001.

[171] M. Potts, S. M. Slaughter, F-U. Hunneke, J. F. Garst, and R. F. Helm. Desiccation tolerance of prokaryotes: application of principles to human cells. *Integrative and Comparative Biology*, 45:800–809, 2005.

[172] Z. Huang and A. Tunnacliffe. Response of human cells to desiccation: comparison with hyperosmotic stress response. *Journal of Physiology*, 558(1):181–191, 2004.

[173] B. T. Gjertsen and S. O. Doskeland. Protein phosphorylation in apoptosis. *Biochimica et Biophysica Acta*, 1269:187–199, 2009.

[174] R. R. P. Rastogi and R. P. Sinha. Molecular mechanisms and pathogenicity. *EXCLI Journal*, 8:155–181, 2009.

[175] Q. Matthews, A. Jirasek, J. Lum, X. Duan, and A. G. Brolo. Variability in Raman spectra of single human tumor cells cultured *in vitro*: correlation with cell cycle and culture confluency. *Applied spectroscopy*, 64(8):871–887, 2010.

[176] H. Bai, H. Li, Z. Han, C. Zhang, J. Zhao, C. Miao, S. Yan, A. Mao, H. Zhao, and Z. Han. Label-free assessment of replicative senescence in mesenchymal stem cell by Raman microspectroscopy. *Biomedical Optical Express*, 6(11):247933, 2015.

[177] F-M. Boisvert, S. V. Koningsbruggen, J. Navascues, and A. I. Lamond. The multifunctional nucleolus. *Nature Reviews - Molecular Cell Biology*, 8:574–585, 2007.

[178] M. Thiry and D. L. J. Lafontaine. Birth of a nucleolus: the evolution of nucleolar compartments. *Trends in Cell Biology*, 15:194–199, 2005.

[179] A. C. S. Talari, Z. Movasaghi, S. Rehman, and I. Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 50:46–111, 2015.

[180] K. Hamada, K. Fujita, N. I. Smith, M. Kobayashi, Y. Inouye, and S. Kawata. Raman microscopy for dynamic molecular imaging of living cells. *Journal of Biomedical Optics*, 13:044027, 2008.

[181] C. Krafft, L. Neudert, T. Simat, and R. Salzer. Near infrared Raman spectra of human brain lipids. *Spectrochimica Acta Part A*, 61:1529–1535, 2005.

[182] E. Brauchle, S. Thude, S. Y. Brucker, and K. Schenke-Layland. Cell death stages in single apoptotic and necrotic cells monitored by raman microspectroscopy. *Scientific Reports*, 4:4698, 2014.

[183] H-Y. N. Holman, M. C. Martin, E. A. Blakely, K. Bjornstad, and W. R. McKinney. Ir spectroscopy characteristics of cell cycle and cell death probed by synchroton radiation based Fourier transform IR spectromicroscopy. *Biopolymers (Biospectroscopy)*, 57:329–335, 2000.

[184] W. Wagner, R. E. Feldmann Jr., A. Seckinger, M. H. Maurer, F. Wein, J. Blake, U. Krause, A. Kalenka, H. F. Burgers, R. Saffrich, P. Wuchter, W. Kuschinsky, and A. D. Ho. The heterogeneity of human mesenchymal stem cell preparations – evidence from simultaneous analysis of proteomes and transcriptomes. *Experimental Hematology*, 34:536–548, 2006.

[185] F. Shekari, H. Baharvand, and G. H. Salekdeh. Organella proteomics of embryonic stem cells. *Advances in Protein Chemistry and Structural Biology*, 95:215–230, 2014.

[186] D. Nasrabadi, M. R. Larijani, A. Fathi, H. Gourabi, A. V. Dizaj, H. Baharvand, and G. H. Salekdeh. Nuclear proteome analysis of monkey embryonic stem cells during differentiation. *Stem Cell Reviews and Reports*, 6:50–61, 2010.

[187] M. Barthelery, U. Salli, and K. E. Vrana. Nuclear proteomics and directed differentiation of embryonic stem cells. *Stem Cells and Development*, 16:905–919, 2007.

[188] L. J. Foster, P. A. Zeemann, C. Li, M. Mann, O. N. Jensen, and M. Kassem. Differential expression profiling of membrane proteins by quantitative proteomics in human mesenchymal stem cell line undergoing osteoblast differentiation. *Stem Cells*, 23:1367–1377, 2005.

[189] S. Mareddy, J. Broadbent, R. Crawford, and Y. Xiao. Proteomic profiling of distinct clonal populations of bone marrow mesenchymal stem cells. *Journal of cellular biochemistry*, 106:776–786, 2009.

[190] U. Kini and B. N. Nandeesh. *Radionuclide and Hybride Bone Imaging*, chapter Physiology of bone formation, remodeling, and metabolism, pages 29–. Sringer-Verlag, 2012.

[191] L. C. Palmer, C. J. Newcomb, S. R. Kaltz, E. D. Spoerke, and S. I. Stupp. Biomimetic systems for hydroxyapatite mineralization inspired by bone and enamel. *Materials Science and Engineering R*, 108(11):4754–4783, 2008.

[192] F. Nudelman, K. Pieterse, A. George, P. H. H. Bomans, H. Friedrich, L. J. Brylka, P. A. Hilbers, G. de With, and N. A. J. M. Sommerdijk. The role of collagen in bone

apatite formation in the presence of hydroxyapatite nucleation inhibitors. *Nature Materials*, 9(12):1004–1009, 2010.

[193] F. Cui, Y. Li, and J. Ge. Self-assembly of mineralised collagen composites. *Materials Science and Engineering R*, 57:1–27, 2007.

[194] W. Zhang, Z. Huang, S. Liao, and F. Cui. Nucleation sites of calcium phosphate crystals during collagen mineralization. *Journal of American Ceramic Society*, 86(6):1052–1054, 2003.

[195] T. Boenisch, editor. *Immunochemical Staining Methods*. DAKO Corporation, California, US, 3 edition, 2001.

[196] J. De Gelder, K. De Gussem, P Vandenabeele, and L. Moens. Reference database of Raman spectra of biological molecules. *Journal of Raman Spectroscopy*, 2007.

[197] G. S. Mandair and M. D. Morris. Contributions of Raman spectroscopy to the understanding of bone strength. *BoneKEy Reports*, 4:1–8, 2015.

[198] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, and M. Baranska. Raman spectroscopy of lipids: a review. *Journal of Raman spectroscopy*, 46:4–20, 2015.

[199] S. Martin and R. G. Parton. Caveolin, cholesterol and lipid bodies. *Seminars in Cell & Developmental Biology*, 16:163–174, 2005.

[200] NIH U. S. National Center for Biotechnology Information. Pubchem - open chemistry database. `https://pubchem.ncbi.nlm.nih.gov/compound`. Online; accessed: August/2017.

[201] A. Wronska and Z. Kmiec. Structural and biochemical characteristics of various white adipose tissue depots. *Acta Physiologica*, 205:194–208, 2012.

[202] D. Zweytick, K. Athenstaedt, and G. Daum. Intracellular lipid particles of eukaryotic cells. *Biochemica et Biophysica Acta*, 1469:101–120, 2000.

[203] T. Skurk, C. Alberti-Huber, C. Herder, and H. Hauner. Relationship between adipocyte size and adipokine expression and secretion. *The Journal of Clinical Endocrinology & Metabolism*, 92(3):1023–1033, 2007.

[204] D. J. Prockop. Marrow stromal cells as stem cells for nonhematopoietic tissues. *Science*, 276:71–74, 1997.