

Spectroscopic and Chemometric Approaches to Radiobiological Analyses

Meade A.D.^{1,2}, Byrne, H.J.³, Lyng F.M.^{2,3}.

1. School of Physics, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland
2. Radiation and Environmental Science Centre, Focas Institute, Dublin Institute of Technology, Camden Row, Dublin 8, Ireland
3. Focas Institute, Dublin Institute of Technology, Camden Row, Dublin 8, Ireland

Vibrational spectroscopy is an attractive modality for the analysis of biological samples, providing a complete non-invasive acquisition of the biochemical fingerprint of the sample. It has been demonstrated that this data provides the means to assay multiple functional responses of a biological system at a spatial resolution as low as a micron within the sample. As the interaction of ionizing radiation with biological systems involves chemical reactions between the products of radiation induced damage and various structural and functional units within the cell, the vibrational spectroscopic modalities have received attention as potential measurement platforms for the in-situ examination of the chemistry of biological species in radiobiology. This presents challenges in relation to sample preparation and the construction of suitable analytical methodologies. In this work protocols for sample preparation and approaches to multivariate analysis of vibrational spectra in radiobiological analysis are detailed and the utility of the methodology in analyzing the evolution of biochemical responses to radiobiological damage are highlighted.

Keywords: Vibrational spectroscopy, radiobiology, chemometric modelling, multivariate statistics.

1. Introduction

Ionizing radiation interacts with the cell inducing a complex series of biochemical responses that are dependent on radiation LET, dose, dose rate, and cell type [1-3]. It is thought that the ultimate target for ionizing radiation in the cell is the cellular DNA [1,3]. Direct radiation damage, as a result of interactions between a charged particle or high energy photon and DNA, is reflected in DNA base cleavage reactions, the formation of base adducts, and the creation of single and double strand breaks in addition to sites of clustered damage [4]. It is thought that DNA damage may also be mediated through the generation of free radicals (such as superoxides ($\cdot\text{O}_2^-$), hydroxides ($\cdot\text{OH}$) and nitrous oxides ($\cdot\text{NO}$) etc.) in the water rich cell nucleus, which subsequently react with the DNA [4,5], and may also subsequently be involved in intracellular signalling processes [5]. Damage may also be mediated through cell to cell signalling [6] and membrane-dependent signalling. The damage to DNA is subsequently reflected in the initiation of multiple transduction pathways that attempt to repair the damage [7], resulting in the production of mediatory and regulatory proteins and lipids that are specific to each pathway [1,8,9]. The repair response may subsequently be overwhelmed resulting in apoptosis [2,10], oncosis or mitotic death of the cell [11], or the damage may be repaired with the cell continuing on its cycle, albeit with potential mutations to its DNA that can result in mutagenesis [12].

Much of the radiation damage occurring in biological media after exposure to ionizing radiation is mediated through chemical interactions between the ionization products and sensitive molecular structures within cells [13-15]. Methodologies that examine the total concentration of metabolic species within the cell after exposure to ionising radiation have recently been described [16-18]. Vibrational spectroscopy is a global term that is applied to the complementary techniques of infrared and Raman spectroscopy, modalities that allow the measurement of the concentration and nature of the organic species within a sample. Due to vibrational selection rules biological molecules may contain individual bonds or moieties that will be either infrared or Raman active, and consequently measurement of both spectra assesses the total content of organic chemical species within a biological specimen [19]. Both modalities have in the last decade seen an acceleration of their use in and application to biological research as a result of the development of microscope systems and solid state detection systems that now allow the rapid measurement of the chemical content in cellular and tissue species [20] within laboratory benchtop devices. The spatial resolution of both Fourier Transform Infrared microspectroscopy (FTIRM) and Confocal Raman Microscopy (CRM) is diffraction limited to $\sim 3\mu\text{m}$ to $10\mu\text{m}$ in FTIRM and $\sim 0.5\mu\text{m}$ to $1.5\mu\text{m}$ in CRM depending on the measurement wavelengths and detector characteristics in a given system [21]. The higher signal levels achievable with FTIRM result in increased speed of measurement and consequently higher sample throughput over CRM [21]. The strengths of both are now well established in hyperspectral imaging for non-invasive and label-free histopathology [22-30], while the capabilities of CRM to detect tissue abnormalities in-vivo without the complication of contamination of spectral measurements by water and

atmospheric features has resulted in a move to the development of clinical devices [31].

A significant degree of attention now focusses on the diagnostic and analytical capabilities of FTIRM and CRM for cytometry and functional analysis in biological media. Quantitative spectral changes attributable to the biochemical processes occurring during cell culture and mitosis [32,33], proliferation [34], differentiation, [35,36], adhesion [37], death [38-42], and invasion [43] have previously been observed with FTIRM and CRM. The increased resolution in CRM imaging has also been exploited to enable the detection of cellular mitochondrial distribution [44] and phagosomes [45]. Early applications of both techniques in radiobiology had focussed on studies with Raman spectroscopy of radiation-induced peroxidation in model phospholipid and membrane systems [46,47], structural changes in protein [48] and in nucleic acids [49]. Such studies continue to contribute to the understanding of the interaction mechanisms between radiation and key molecular species and structures within the cell [48-51]. These investigations have since evolved to Raman spectroscopic studies of the effects of in-vitro proton irradiation of excised tissues [52], and radiotherapeutically irradiated (γ -radiation) tissue specimens [53-55]. Other studies with bacteria have investigated the effects of γ -radiation on *Deinococcus radiodurans* [56], *Kocuria rosea* [57] and *Micrococcus luteus* [58] with FTIR spectroscopy. Recently Gault et al have demonstrated the sensitivity of FTIR spectroscopy to apoptotic effects in γ -irradiated lymphocytes [59] and to molecular changes in HaCaT cells exposed to both γ -radiation and hydrogen peroxide [60], together with its sensitivity to biochemical alterations in HaCaT cells irradiated with α -particles delivered by micro-beams [61]. These studies represent proof of the

detection of biochemical effects in irradiated biological species with spectroscopic methods. However, the utility of the technique in the analysis of biochemical changes occurring with radiobiological effect has yet to be established. The present work outlines sample preparation techniques, analytical methodologies and modelling techniques that may be used with FTIRM spectra in the analysis of biochemical changes occurring with dose and time after γ -irradiation of HaCaT keratinocytes.

2. Sample Preparation and Spectral Treatment

2.1 Sample Preparation Considerations

Various cytological sample preparation techniques exist that often involve the centrifugal deposition of cells onto spectroscopic mounting substrates [62], or their culture in-situ on the substrate. Cell culture on such substrates has been demonstrated to be toxic to the cell, although use of molecular biochemical coatings (such as gelatin) have been shown to ameliorate such effects [37]. Cells may be studied in their live form with CRM, but may also be preserved with chemical fixation for both CRM and FTIRM (which will also require sample dessication), and this has been demonstrated to adjust both sets of spectra such that experimental controls preserved in the same manner are generally required [63].

2.2 Spectral Preprocessing Considerations

As FTIRM in cytometry requires the transmission of the IR radiation through the cell, mounting substrates such as zinc-selenide (ZnSe) and calcium-fluoride (CaF₂) disks

are commonly used for ‘transmission mode’ measurements and glasses coated with materials reflecting in the IR are commonly used for ‘transflection mode’ measurements (such as *MirrIR* (Kevley Technologies)) [64]. Spectra may, due to optical transmission effects exhibit a broad oscillating baseline that is attributable to Mie scattering [62,65]. This effect arises since the nucleus and subcellular organelles may act as non-absorbing dielectric spheres and scatter, in a wavelength and organelle-size dependent manner, the incident IR radiation. However, this fails to account for the derivative-like line shapes (termed a ‘dispersion artefact’) seen in various positions in the spectrum, which contribute to both a background oscillation, and a shifting of certain peak positions. The resonant behaviour of this phenomenon (where scattering and absorption by structures within the cell) has recently been shown to produce these effects simultaneously [66]. Corrections for this effect may be provided with the extended multiplicative scatter correction (EMSC) [67,68]. Other contaminating effects in FTIRM spectra occur due to absorptions of water vapour and CO₂, which may also be removed by existing embedded machine-dependent software corrections or machine-dependent in-house correction algorithms [69], while simple filtering and normalization procedures may correct for spectral noise and remove point-to-point variations in sample concentration.

Raman spectra from cellular species are also characterized by baseline variability that originates in interactions of the excitation laser with the complex chemistry and optical characteristics of the sample. These include resonant effects that may dominate the spectrum with excitation wavelengths in the visible and UV regions, and a highly variable background that has its origin in sample fluorescence with visible and UV excitation, but not with near IR excitation. The utilisation of excitation

wavelengths in the near IR removes the potential for photobleaching effects [70] and with short exposure times is also non-cytotoxic and therefore suitable for use in studies on live cells or in-vivo [71]. Raman spectra may be corrected for the presence of known contaminants, and the broad background signal may be removed using EMSC or subtraction of fitted polynomials. As the optical window used as a sample support during CRM is itself Raman active (in the case of ZnSe and quartz) this signal is generally measured as part of the acquisition protocol, and is often combined with other instrumental corrections in preprocessing [72].

3. Quantitative Multivariate Analytical Methodologies

Multivariate statistical models and chemometric techniques for relating spectral signatures to analyte or agents of interest through multivariate regression have been developed extensively for the past three decades, and have seen a wide range of applications in fields from proteomics to the petroleum industry [73]. With the increased proliferation of powerful computational facilities, these algorithms have also been coupled with feature selection techniques that allow the decomposition of an optimal set of variables (spectral features) from the spectral data which maximize the predictive capacity and speed of implementation of the modelling algorithms, together with simplifying them and rendering them more interpretable scientifically [74]. All of the algorithms seek to calibrate a model that relates the spectral dataset (*X-matrix*) to a target of interest (*Y-matrix*, eg. concentration of reaction product or analyte) [73]. The most extensively applied multivariate regression algorithm is the partial least squares regression (PLSR) technique, which involves the decomposition of the X-matrix and Y-matrix into their eigenvectors and eigenvalues and the

selection of an optimal model with minimal complexity that predicts Y from X [73], with minimized residual error, E .

$$Y = XB + E \tag{1}$$

This model assumes a linear relationship between the structure of the data in X and that in Y , although many experiments involve non-linear relationships between the two sets of variables [73]. Quadratic PLSR (Q-PLSR) has therefore been developed to allow the study of quadratic relationships between spectral variables and the predictand, Y , while more complex variations in X that relate to Y may be modelled using the generalized regression neural network (GRNN), which can model any non-linear regression problem [75]. A method of feature selection that provides readily interpretable results in PLSR involves the method of *Jack-knifing* developed by Martens and colleagues, whereby the uncertainty variance in the regression coefficients, B , in the PLS model is estimated for a model with optimal complexity, while variables which do not possess regression coefficients that are statistically significant at a certain level of confidence are eliminated using a t-test [76]. Other methods that employ evolutionary computational techniques such as genetic algorithms are less readily interpretable, but very rapidly decompose the spectral dataset into a small number of variables that maximize the predictive capacity of the models [77], and can be used with any regression algorithm. These algorithms are particularly suited to this type of minimization problem since the search space (i.e. the number of combinations of variables that may be input to the regression models) is vast, and standard exhaustive search techniques are therefore unrealistic [77]. Genetic algorithms are one such option that have been successfully utilised in multivariate

regression problems [73]. They employ concepts from biological evolution where spectral variables are encoded as ‘genes’ (which are often encoded as binary variables, 1 and 0, such that 1 represents a variable that is selected and vice versa), a group of genes is termed a ‘chromosome’ and the algorithm evolves by means of mutation (changing the values of individual genes randomly) and crossover (exchanging genes between chromosomes) to test different chromosome selections and to escape local minima. Individual chromosomes are selected and rejected by virtue of their ‘fitness’, i.e. the value of a multivariate selection criterion that must either be maximised or minimised depending on the evolution criteria [73,77].

4. Application to Radiobiological Analysis

4.1 Sample Preparation and Data Acquisition

FTIRM and multivariate regression techniques were used to analyse the spectra of a human keratinocyte cell line (HaCaT) with γ -radiation dose and with time post-irradiation, extensive details of which are available elsewhere [78]. Cells were cultured on low emissivity silver oxide coated glass slides (MirrIR, Kevley Technologies), on which a 2% gelatin coating was deposited to enable attachment of the cell and improve cell viability [37]. Doses (ten points) ranging from 0 Gy to 5 Gy were delivered to the samples, and they were fixed in 4% neutral-buffered formalin at time points ranging from 6 hours to 96 hours post-irradiation.

Spectra were recorded in transfection mode using a Perkin-Elmer GXII spectrometer in the 4000 to 720 cm^{-1} wavenumber range with an aperture size of 100 $\mu\text{m} \times 100 \mu\text{m}$,

a spectral resolution of 4 cm^{-1} and with 64 scans per spectrum. Approximately 300 spectra were recorded at each dose and time point. Outliers were removed using a multivariate test of the dissimilarity between the spectral data within a given dose and time class, and the contributions of water and carbon dioxide were removed from each spectral set [69]. Each spectral set was then corrected using EMSC [68], and were vector normalised.

4.2 Multivariate Modelling of Dose Dependencies in FTIRM Spectra

All data processing and analysis was performed in the Matlab 7.2 environment (The MathWorks Inc., USA) with PLS Toolbox 5.0.3 (Eigenvector Research, Wenatchee, WA, USA). PLSR, Quadratic PLSR (QPLSR) and GRNN models were constructed for the spectral datasets at each time point. Each model was constructed on 60% of the spectral data randomly selected at each time point, with the remaining 40% of the spectral data retained as an unseen testing set. The PLSR and QPLSR models performance were subsequently evaluated for 10 separate randomizations of the datasets, while the GRNN was evaluated for 50 randomizations of the data to prevent data bias. The root mean squared errors of calibration (RMSEC) and prediction (RMSEP) were used as measures quantifying the performance of the models in regressing the spectral data against the radiation dose to which the cells were exposed.

Table 1 depicts the performance of each of the models in predicting radiation dose at time point. The large disparity between the RMSEC and RMSEP values obtained using the GRNN, PLSR and QPLSR algorithms suggests that the biochemical fingerprint of the cell varies in a non-linear manner with radiation dose, and such

variations are not adequately described by simple non-linear regressions such as QPLSR. In addition, the predictive efficiencies of the PLSR and QPLSR models are relatively consistent with time post-irradiation, while that of the GRNN varies with time post-irradiation, being best at 6 hours and 96 hours post-irradiation. The variation may be the result of the differences in the degree of variability in the biochemical fingerprint with dose at each time point, an effect that may be the result of the initiation of a variety of DNA repair processes and other responses.

In order to identify spectral features that vary linearly and non-linearly with dose, a study with feature selection methods was conducted. Variable selection was performed on the PLSR and QPLSR regression coefficients according to the Jack-knifing procedure described by Martens and Naes, with $p < 0.05$ for t-testing. Variable selection was also performed with PLSR and QPLSR using a genetic algorithm (GA) constructed in Matlab. Initial selection of variables was performed according to the method of Yoshida et al [73,79], where a number of short GA runs (with evolution for 20 generations and with 30 crossovers ($p=0.9$) and 50 mutations ($p=0.05$) per generation) were executed to minimize the number of features selected by the algorithm, and to prevent overfitting that has been demonstrated to occur when using GA's with a large search domain [73]. Subsequently, a more extensive evaluation was performed with the GA, using the features most often selected during the initial evaluation (the numbers of which are detailed in Table 2). In this case the GA was executed on fifty separate occasions, with evolution for 50 generations at each execution (and with 100 crossovers ($p=0.9$) and 100 mutations ($p=0.05$) per generation). Variables for selection were encoded with binary digits, and a population of 25 individual chromosomes was randomly created for evolution at the beginning of each independent execution of the algorithm. One hundred spectra from the pooled

dataset at each time point were selected randomly for calibration of the multivariate models with each GA chromosome, and a separate one hundred spectra were selected for testing, with the minimisation of the RMSEP being the fitness criterion adopted for the evolution of the algorithm. It was found that the number of features selected by PLS Jack-knifing was quite large (see Table 2), and was not minimised further with repeated evaluations, whereas the number selected by PLS-GA after a number of rapid evaluations is much less and is thus more readily interpretable. The GA was subsequently adopted for feature selection with both the PLS and QPLS models.

In figure 1, a spectrum of a control sample of the HaCaT cell line is provided. The IR spectrum contains broad peaks due to strong overlap between vibrations of distinct biochemical components, although characteristic group vibrations enable the identification of specific species (such as carbohydrate OH bond stretching from $3520\text{-}3100\text{ cm}^{-1}$, protein amide vibrations from $1654\text{-}1530\text{ cm}^{-1}$, and phosphate bond stretch of the DNA and RNA backbone from $1090\text{-}1084\text{ cm}^{-1}$). A full list of characteristic vibrations is contained elsewhere [78]. In figure 2, the spectrum of the sham-irradiated control sample is provided at each time point to provide a reference for the interpretation of the features selected by the GA. The features selected by the regression at each time point, and for both the PLS and QPLS models are highlighted as bars at the spectral wavenumbers that were selected by the model, with the length of the bar reflecting the frequency with which it is selected by the GA over all the independent runs of the GA. It is apparent that there are significant differences between the features selected by the linear (PLS) and quadratic (QPLS) models, particularly at early time points after irradiation, which signify biochemical

differences in the cellular response to more complex photon deposition and repair process (such as occurs after double strand breakage and clustered base damage).

Interpreting only those features with a high frequency of selection (i.e. frequency greater than 0.9, where a frequency of 1 is equivalent to that feature being selected on each of the independent runs of the GA) one consistent range over which the features are selected is between $\sim 2600\text{ cm}^{-1}$ to 1850 cm^{-1} , another being from 3720 cm^{-1} to 3500 cm^{-1} , which do not contain features of biochemical origin, but rather may be assigned to changes in the optical characteristics of the cell [62,65] due to organelle swelling that occurs during the cell cycle, and during apoptotic and necrotic cell death.

Various investigations with FTIR spectroscopy have identified spectral features that are changed during apoptosis and necrosis [38,39,41,42]. Evidence of early dose dependent apoptotic and necrotic cell death are seen here at 6 hours after irradiation with the selection of the spectral features by the GA-PLS algorithm at 1155 cm^{-1} (assigned to the -C-OH bond stretch of serine, threonine and tyrosine residues [42] signifying protein denaturation) and by the GA-PLS and GA-QPLS algorithms at 1639 cm^{-1} (assigned to the amide I bond vibration within β -sheet structures in protein [38,39]). At 12 hours post irradiation the GA-PLS algorithm selects features at 1713 cm^{-1} (assigned to base paired DNA) and 3012 cm^{-1} (assigned to the olefinic bond in lipid) that are changed with apoptosis [39], while the GA-QPLS algorithm selects features at 1161 cm^{-1} (again assigned to the -C-OH bond stretch of serine, threonine and tyrosine residues [42]) and 968 cm^{-1} (assigned to C-C stretch in the deoxyribose sugar moiety [39,41]) that have also been shown to be characteristically changed

during apoptosis. At 24 hours the amide I bond vibration at 1619 cm^{-1} is again selected by the GA-PLS algorithm together with the $-\text{CH}_2$ stretching vibrations in lipid and protein (2917 cm^{-1} and 2848 cm^{-1}) that have been shown to be apoptotic markers reflecting structural changes in membrane lipid [41]. Further apoptotic markers are not selected by either algorithm after 24 hours.

Throughout the spectrum, at each time point after irradiation, the GA-PLS and GA-QPLS algorithms select spectral features of protein which may signify the formation of protein-protein crosslinks and proteolytic reactions occurring during cell death. Features that reflect the changing concentrations of carbohydrate during cellular proliferation and apoptosis are selected by the GA-PLS algorithm in the spectral ranges from $3520\text{-}3100\text{ cm}^{-1}$ (OH bond stretch), $1290\text{-}1030\text{ cm}^{-1}$ (OH bond stretch), $1200\text{-}1000\text{ cm}^{-1}$ (osidic (C-O) bond stretch), and $960\text{-}730\text{ cm}^{-1}$ (overlapping C-H stretching vibrations and ring vibrations in α and β -pyranose carbohydrates) at each time point from 6 to 96 hours post-irradiation. Contrastingly, the GA-QPLS algorithms select features over each of these ranges only at 48 and 96 hours after irradiation, the selected spectral features such as osidic bond stretching vibrations and CH deformation vibrations at 6, 12 and 24 hours possibly signifying structural changes to nucleic acids rather than carbohydrate metabolism. Evidence of the initial effects of irradiation on the conformation of the nucleic acid backbone are the selection of the feature at 1227 cm^{-1} ($-\text{PO}_2^-$ asymmetric stretch in DNA and RNA) by the GA-PLS algorithm at 6 hours after irradiation. Similar features ($-\text{PO}_2^-$ symmetric stretch in DNA and RNA at 1243 cm^{-1}) are again selected by the GA-PLS algorithm at 24 hours, and by the GA-QPLS algorithm at 24 hours ($-\text{PO}_4^-$ stretch (964 cm^{-1}), $-\text{PO}_2^-$ symmetric (1084 cm^{-1}) and asymmetric (1234 cm^{-1}) stretch) and at 48 hours ($-\text{PO}_2^-$

stretch in DNA/ RNA at 1084 cm^{-1}). Each of these features may signify the early and late production of single and double strand breaks through initial energy deposition events and potentially as a result of long-term elevation of ROS [3,5] with some differences in the relationship between the production of such damage and the initial radiation dose. The selection of the osidic bond stretching vibration in DNA and RNA (in the region of 1050 cm^{-1}) at 24, 48 and 96 hours after irradiation may also signify the occurrence of base cleavage reactions that may be due to ROS damage or DNA repair.

These results demonstrate the capacity of FTIRM coupled with suitable data mining techniques to capture and elucidate the complex nature of radiobiological responses at the molecular level, and to elucidate the evolution of the molecular changes with time after irradiation. Further development of the technique, through correlation of spectral changes with biochemical and biological endpoints, is necessary to fully exploit the richness of the data and establish it as a modality for routine use in radiobiological analysis.

5. Conclusion

Vibrational spectroscopy coupled with multivariate chemometric algorithms represent a new and powerful approach to the analysis of the total biochemical fingerprint after radiobiological damage. Suitable data mining techniques coupled with parallel physiological data can elucidate molecular aspects of the cellular radiobiological response mechanisms. Future studies will develop the methodology outlined here and

demonstrate the capability of the measurement to separate and analyse various radiobiological responses of the cell in-vitro.

6. References

- [1] K.M. Prise, G. Schettino, M. Folkard and K.D. Held New insights on cell death from radiation exposure, *Lancet Oncol* 6 (2005) 520-528.
- [2] B.G. Wouters Cell death after irradiation: how, when and why cells die, in: M.C. Joiner and A. van der Kogel (Eds.), *Basic Clinical Radiobiology*, Hodder Arnold, London, 2009, pp. 27-39.
- [3] B.G. Wouters and A.C. Begg Irradiation-induced damage and the DNA damage response, in: M.C. Joiner and A. van der Kogel (Eds.), *Basic Clinical Radiobiology*, Hodder Arnold, London, 2009, pp. 11-26.
- [4] P. O'Neill and P. Wardman Radiation chemistry comes before radiation biology, *Int J Radiat Biol* 85 (2009) 9-25.
- [5] R.B. Mikkelsen and P. Wardman Biological chemistry of reactive oxygen and nitrogen and radiation-induced signal transduction mechanisms, *Oncogene* 22 (2003) 5734-5754.
- [6] E.I. Azzam, S.M. de Toledo and J.B. Little Stress signaling from irradiated to non-irradiated cells, *Curr Cancer Drug Targets* 4 (2004) 53-64.
- [7] P. Dent, A. Yacoub, J. Contessa, R. Caron, G. Amorino, K. Valerie, M.P. Hagan, S. Grant and R. Schmidt-Ullrich Stress and radiation-induced activation of multiple intracellular signaling pathways, *Radiat Res* 159 (2003) 283-300.
- [8] J. Bartek, C. Lukas and J. Lukas Checking on DNA damage in S phase, *Nat Rev Mol Cell Biol* 5 (2004) 792-804.
- [9] L. Brugmans, R. Kanaar and J. Essers Analysis of DNA double-strand break repair pathways in mice, *Mutat Res* 614 (2007) 95-108.
- [10] R.C. Taylor, S.P. Cullen and S.J. Martin Apoptosis: controlled demolition at the cellular level, *Nat Rev Mol Cell Biol* 9 (2008) 231-241.
- [11] M. Castedo, J.L. Perfettini, T. Roumier, K. Andreau, R. Medema and G. Kroemer Cell death by mitotic catastrophe: a molecular definition, *Oncogene* 23 (2004) 2825-2837.
- [12] M.M. Vilenchik and A.G. Knudson Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer, *Proceedings of the National Academy of Sciences of the United States of America* 100 (2003) 12871-12876.
- [13] A.P. Breen and J.A. Murphy Reactions of oxyl radicals with DNA, *Free Radical Biology and Medicine* 18 (1995) 1033-1077.
- [14] P. O'Neill Radiation-induced damage in DNA, in: C.D. Jonah and B.S. Madhava Rao (Eds.), *Radiation Chemistry - Present status and future trends*, Elsevier, London, 2001, pp. 585-622.
- [15] C. von Sonntag and H.P. Schuchmann Radiation Chemistry of the Nucleobases, in: C.D. Jonah and B.S. Madhava Rao (Eds.), *Radiation Chemistry - Present status and future trends*, Elsevier, London, 2001, pp. 585-622.
- [16] J.B. Tyburski, A.D. Patterson, K.W. Krausz, J. Slavik, A.J. Fornace, Jr., F.J. Gonzalez and J.R. Idle Radiation metabolomics. 1. Identification of minimally invasive urine biomarkers for gamma-radiation exposure in mice, *Radiat Res* 170 (2008) 1-14.
- [17] A.D. Patterson, H. Li, G.S. Eichler, K.W. Krausz, J.N. Weinstein, A.J. Fornace, Jr., F.J. Gonzalez and J.R. Idle UPLC-ESI-TOFMS-based metabolomics and gene expression dynamics inspector self-organizing metabolomic maps as tools for understanding the cellular response to ionizing radiation, *Anal Chem* 80 (2008) 665-674.

- [18] G. Rainaldi, R. Romano, P. Indovina, A. Ferrante, A. Motta, P.L. Indovina and M.T. Santini Metabolomics using ¹H-NMR of apoptosis and Necrosis in HL60 leukemia cells: differences between the two types of cell death and independence from the stimulus of apoptosis used, *Radiat Res* 169 (2008) 170-180.
- [19] D. Naumann FT-Infrared and FT-Raman Spectroscopy in Biomedical Research, in: H.U. Gremlich and B. Yan (Eds.), *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York, 2001, pp. 323-377.
- [20] M. Diem, M. Romeo, S. Boydston-White, M. Miljkovic and C. Matthaus A decade of vibrational micro-spectroscopy of human cells and tissue (1994-2004), *Analyst* 129 (2004) 880-885.
- [21] M.D. Schaeberle, I.W. Levin and E.N. Lewis Biological Vibrational Spectroscopic Imaging, in: H.U. Gremlich and B. Yan (Eds.), *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York, 2001, pp. 231-258.
- [22] P. Crow, A. Molckovsky, N. Stone, J. Uff, B. Wilson and L.M. WongKeeSong Assessment of fiberoptic near-infrared raman spectroscopy for diagnosis of bladder and prostate cancer, *Urology* 65 (2005) 1126-1130.
- [23] D.C. Fernandez, R. Bhargava, S.M. Hewitt and I.W. Levin Infrared spectroscopic imaging for histopathologic recognition, *Nat Biotechnol* 23 (2005) 469-474.
- [24] I.W. Levin, Bhargava, R. Fourier Transform Infrared Spectroscopic Imaging: Integrating Microscopy and Molecular Recognition, *Annu. Rev. Phys. Chem.* 56 (2005) 429-474.
- [25] F.M. Lyng, E.O. Faolain, J. Conroy, A.D. Meade, P. Knief, B. Duffy, M.B. Hunter, J.M. Byrne, P. Kelehan and H.J. Byrne Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool, *Exp Mol Pathol* 82 (2007) 121-129.
- [26] G. Shetty, C. Kendall, N. Shepherd, N. Stone and H. Barr Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus, *Br J Cancer* 94 (2006) 1460-1464.
- [27] M.J. Baker, E. Gazi, M.D. Brown, J.H. Shanks, P. Gardner and N.W. Clarke FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer, *British Journal of Cancer* 99 (2008) 1859-1866.
- [28] F. Bonnier, S. Rubin, L. Venteo, C.M. Krishna, M. Pluot, B. Baehrel, M. Manfait and G.D. Sockalingum In-vitro analysis of normal and aneurismal human ascending aortic tissues using FT-IR microspectroscopy, *Biochim Biophys Acta* 1758 (2006) 968-973.
- [29] C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker and N. Stone Vibrational spectroscopy: a clinical tool for cancer diagnostics, 2009, pp. 1029-1045.
- [30] C. Krafft, G. Steiner, C. Beleites and R. Salzer Disease recognition by infrared and Raman spectroscopy, *Journal of Biophotonics* 2 (2009) 13-28.
- [31] A. Nijssen, S. Koljenovic, T.C.B. Schutt, P.J. Caspers and G.J. Puppels Towards oncological application of Raman spectroscopy, *Journal of Biophotonics* 2 (2009) 29-36.
- [32] S. Boydston-White, M. Romeo, T. Chernenko, A. Regina, M. Miljkovic and M. Diem Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: principal component and artificial neural network analysis, *Biochim Biophys Acta* 1758 (2006) 908-914.
- [33] C. Matthaus, S. Boydston-White, M. Miljkovic, M. Romeo and M. Diem Raman and infrared microspectral imaging of mitotic cells, *Appl Spectrosc* 60 (2006) 1-8.
- [34] K.W. Short, Carpenter, S., Freyer, J. P., and Mourant, J. R. Raman Spectroscopy Detects Biochemical Changes Due to Proliferation in Mammalian Cell Cultures, *Biophysical Journal* 88 (2005) 4274-4288.
- [35] D. Ami, T. Neri, A. Natalello, P. Mereghetti, S.M. Doglia, M. Zanoni, M. Zuccotti, S. Garagna and C.A. Redi Embryonic stem cell differentiation studied by FT-IR spectroscopy, *Biochim Biophys Acta* 1783 (2008) 98-106.

- [36] I. Notingher, Bisson, I, Bishop, A., E., Randle, W., L., Polak, J.M.P. Hench, L., L. In Situ Spectral Monitoring of mRNA Translation in Embryonic Stem Cells during Differentiation in Vitro, *Anal. Chem.* 76 (2004) 3185-3193.
- [37] A.D. Meade, F.M. Lyng, P. Knief and H.J. Byrne Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in in-vitro cultured human keratinocytes, *Anal Bioanal Chem* 387 (2007) 1717-1728.
- [38] F. Gasparri and M. Muzio Monitoring of apoptosis of HL60 cells by Fourier-transform infrared spectroscopy, *Biochem J* 369 (2003) 239-248.
- [39] K.Z. Liu, L. Jia, S.M. Kelsey, A.C. Newland and H.H. Mantsch Quantitative determination of apoptosis on leukemia cells by infrared spectroscopy, *Apoptosis* 6 (2001) 269-278.
- [40] I. Notingher, S. Verrier, S. Haque, J.M. Polak and L.L. Hench Spectroscopic study of human lung epithelial cells (A549) in culture: living cells versus dead cells, *Biopolymers* 72 (2003) 230-240.
- [41] U. Zelig, J. Kapelushnik, R. Moreh, S. Mordechai and I. Nathan Diagnosis of Cell Death by Means of Infrared Spectroscopy, *Biophysical Journal* 97 (2009) 2107-2114.
- [42] J. Zhou, Z. Wang, S. Sun, M. Liu and H. Zhang A rapid method for detecting conformational changes during differentiation and apoptosis of HL60 cells by Fourier-transform infrared spectroscopy, *Biotechnol Appl Biochem* 33 (2001) 127-132.
- [43] Y. Yang, J. Sule-Suso, G.D. Sockalingum, G. Kegelaer, M. Manfait and A.J. El Haj Study of tumor cell invasion by Fourier transform infrared microspectroscopy, *Biopolymers* 78 (2005) 311-317.
- [44] 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, *Biophys J* 93 (2007) 668-673.
- [45] H.J. van Manen, Y.M. Kraan, D. Roos and C. Otto Single-cell Raman and fluorescence microscopy reveal the association of lipid bodies with phagosomes in leukocytes, *Proc Natl Acad Sci U S A* 102 (2005) 10159-10164.
- [46] K. Sailer, S. Viaggi and M. Nüsse Kinetics of radiation- and cytochrome c-induced modifications in liposomes analysed by FT-Raman spectroscopy, *Biochim Biophys Acta* 1329 (1997) 259-268.
- [47] S.P. Verma and N. Sonwalkar Structural changes in plasma membranes prepared from irradiated Chinese hamster V79 cells as revealed by Raman spectroscopy, *Radiat Res* 126 (1991) 27-35.
- [48] A. Synytsya, P. Alexa, J. de Boer, M. Loewe, M. Moosburger, M. Wirkner and K. Volka Raman spectroscopic study of serum albumins: an effect of proton- and gamma-irradiation, *Journal of Raman Spectroscopy* 38 (2007) 1646-1655.
- [49] K. Sailer, S. Viaggi and M. Nüsse Radiation-induced structural modifications in dsDNA analysed by FT-Raman spectroscopy, *Int J Radiat Biol* 69 (1996) 601-613.
- [50] A. Synytsya, P. Alexa, J. de Boer, M. Loewe, M. Moosburger, M. Wirkner and K. Volka Raman spectroscopic study of calf thymus DNA: an effect of proton- and gamma-irradiation, *Journal of Raman Spectroscopy* 38 (2007) 1406-1415.
- [51] A. Torreggiani, M. Tamba, I. Manco, M.R. Faraone-Mennella, C. Ferreri and C. Chatgililoglu Investigation of radical-based damage of RNase A in aqueous solution and lipid vesicles, *Biopolymers* 81 (2006) 39-50.
- [52] A. Synytsya, P. Alexa, J. Besserer, J. De Boer, S. Froschauer, R. Gerlach, M. Loewe, M. Moosburger, I. Obstova, P. Quicken, B. Sosna, K. Volka and M. Wirkner Raman spectroscopy of tissue samples irradiated by protons, *Int J Radiat Biol* 80 (2004) 581-591.
- [53] R. Jyothi Lakshmi, V.B. Kartha, C. Murali Krishna, R.S. JG, G. Ullas and P. Uma Devi Tissue Raman spectroscopy for the study of radiation damage: brain irradiation of mice, *Radiat Res* 157 (2002) 175-182.

- [54] R.J. Lakshmi, M. Alexander, J. Kurien, K.K. Mahato and V.B. Kartha Osteoradionecrosis (ORN) of the mandible: A laser Raman spectroscopic study, *Applied Spectroscopy* 57 (2003) 1100-1116.
- [55] M.S. Vidyasagar, K. Maheedhar, B.M. Vadhiraja, D.J. Fernandes, V.B. Kartha and C.M. Krishna Prediction of radiotherapy response in cervix cancer by Raman spectroscopy: a pilot study, *Biopolymers* 89 (2008) 530-537.
- [56] A.M. Melin, A. Perromat and G. Deleris Sensitivity of *Deinococcus radiodurans* to gamma-irradiation: a novel approach by Fourier transform infrared spectroscopy, *Arch Biochem Biophys* 394 (2001) 265-274.
- [57] A.M. Melin, A. Perromat, C. Lorin and G. Deleris Gamma irradiation and cellular damage in *Kocuria rosea*: investigation by one- and two-dimensional infrared spectroscopy, *Arch Biochem Biophys* 408 (2002) 211-219.
- [58] A. Perromat, A.M. Melin, C. Lorin and G. Deleris Fourier transform IR spectroscopic appraisal of radiation damage in *Micrococcus luteus*, *Biopolymers* 72 (2003) 207-216.
- [59] N. Gault and J.L. Lefaix Infrared microspectroscopic characteristics of radiation-induced apoptosis in human lymphocytes, *Radiat Res* 160 (2003) 238-250.
- [60] N. Gault, O. Rigaud, J.L. Poncey and J.L. Lefaix Infrared microspectroscopy study of gamma-irradiated and H₂O₂-treated human cells, *Int J Radiat Biol* 81 (2005) 767-779.
- [61] N. Gault, O. Rigaud, J.L. Poncey and J.L. Lefaix Biochemical alterations in human cells irradiated with alpha particles delivered by macro- or microbeams, *Radiat Res* 167 (2007) 551-562.
- [62] M. Romeo, B. Mohlenhoff and M. Diem Infrared micro-spectroscopy of human cells: Causes for the spectral variance of oral mucosa (buccal) cells, Elsevier Science Bv, 2006, pp. 9-14.
- [63] A.D. Meade, C. Clarke, F. Draux, G.D. Sockalingum, M. Manfait, F.M. Lyng and H.J. Byrne Studies of chemical fixation effects in human cell lines using Raman microspectroscopy, *Analytical and Bioanalytical Chemistry* DOI: 10.1007/s00216-009-3411-7 (2010).
- [64] P. Bassan, H.J. Byrne, J. Lee, F. Bonnier, C. Clarke, P. Dumas, E. Gazi, M.D. Brown, N.W. Clarke and P. Gardner Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells, *Analyst* 134 (2009) 1171-1175.
- [65] B. Mohlenhoff, M. Romeo, M. Diem and B.R. Wood Mie-type scattering and non-Beer-Lambert absorption behavior of human cells in infrared microspectroscopy, *Biophys J* 88 (2005) 3635-3640.
- [66] P. Bassan, H.J. Byrne, F. Bonnier, J. Lee, P. Dumas and P. Gardner Resonant Mie scattering in infrared spectroscopy of biological materials - understanding the 'dispersion artefact', *Analyst* 134 (2009) 1586-1593.
- [67] A. Kohler, J. Sule-Suso, G.D. Sockalingum, M. Tobin, F. Bahrami, Y. Yang, J. Pijanka, P. Dumas, M. Cotte, D.G. van Pittius, G. Parkes and H. Martens Estimating and correcting Mie scattering in synchrotron-based microscopic Fourier transform infrared spectra by extended multiplicative signal correction, *Applied Spectroscopy* 62 (2008) 259-266.
- [68] S.N. Thennadil, H. Martens and A. Kohler Physics-based multiplicative scatter correction approaches for improving the performance of calibration models, *Applied Spectroscopy* 60 (2006) 315-321.
- [69] S.W. Bruun, A. Kohler, I. Adt, G.D. Sockalingum, M. Manfait and H. Martens Correcting attenuated total reflection-Fourier transform infrared spectra for water vapor and carbon dioxide, *Appl Spectrosc* 60 (2006) 1029-1039.
- [70] G.J. Puppels, J.H. Olminkhof, G.M. Segers-Nolten, C. Otto, F.F. de Mul and J. Greve Laser irradiation and Raman spectroscopy of single living cells and chromosomes: sample degradation occurs with 514.5 nm but not with 660 nm laser light, *Exp Cell Res* 195 (1991) 361-367.

- [71] I. Notingher, Verrier, S., Romanska, H., Bishop, A.E., Polak, J.M., Hench, L.L. In situ characterisation of living cells by Raman spectroscopy, *Spectroscopy - Int. J.* 16 (2002) 43-51.
- [72] C. Gobinet, V. Vrabie, M. Manfait and O. Piot Preprocessing Methods of Raman Spectra for Source Extraction on Biomedical Samples: Application on Paraffin-Embedded Skin Biopsies, *Ieee Transactions on Biomedical Engineering* 56 (2009) 1371-1382.
- [73] K. Varmuza and P. Filzmoser *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Taylor and Francis Group, Boca Raton, FL, 2009.
- [74] Y. Saeys, I. Inza and P. Larranaga A review of feature selection techniques in bioinformatics, *Bioinformatics* 23 (2007) 2507-2517.
- [75] D.F. Specht A general regression neural network, *IEEE Trans Neural Netw* 2 (1991) 568-576.
- [76] F. Westad and H. Martens Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression, *Journal of near Infrared Spectroscopy* 8 (2000) 117-124.
- [77] C.A. Pena-Reyes and M. Sipper Evolutionary computation in medicine: an overview, *Artif Intell Med* 19 (2000) 1-23.
- [78] A.D. Meade, C. Clarke, H.J. Byrne and F.M. Lyng Fourier Transform Infrared microspectroscopy and multivariate methods for radiobiological dosimetry *Radiation Research* DOI: 10.1667/RR1836.1 (2010).
- [79] H. Yoshida, R. Leardi, K. Funatsu and K. Varmuza Feature selection by genetic algorithms for mass spectral classifiers, *Analytica Chimica Acta* 446 (2001) 485-494.

Tables

Time (hours)	GRNN RMSEC	GRNN RMSEP	PLSR RMSEC	PLSR RMSEP	QPLSR RMSEC	QPLSR RMSEP
6	0.028(0.005)	0.042(0.007)	0.22(0.01)	0.31(0.02)	0.35(0.02)	0.48(0.05)
12	0.170(0.007)	0.405(0.072)	0.59(0.03)	0.79(0.03)	0.56(0.04)	0.76(0.06)
24	0.094(0.006)	0.171(0.024)	0.24(0.02)	0.33(0.02)	0.27(0.02)	0.40(0.04)
48	0.091(0.062)	0.242(0.082)	0.32(0.01)	0.46(0.02)	0.32(0.02)	0.46(0.03)
96	0.003(0.005)	0.005(0.004)	0.27(0.02)	0.37(0.01)	0.39(0.04)	0.52(0.02)

Table 1. Comparative performances of GRNN versus PLSR and NL-PLSR models in prediction of dose from corrected spectral measurements with time point post-irradiation (RMSEC and RMSEP are in the units of dose (Gy). Figures in brackets denote the standard deviations on the mean.

Time (hours)	No. features selected by Jack-knifing	No. features selected by GA
6	849	108
12	880	142
24	669	82
48	819	86
96	783	131

Table 2. Numbers of spectral features selected by PLS Jack-knifing and PLS-GA as described in the text.

Figures

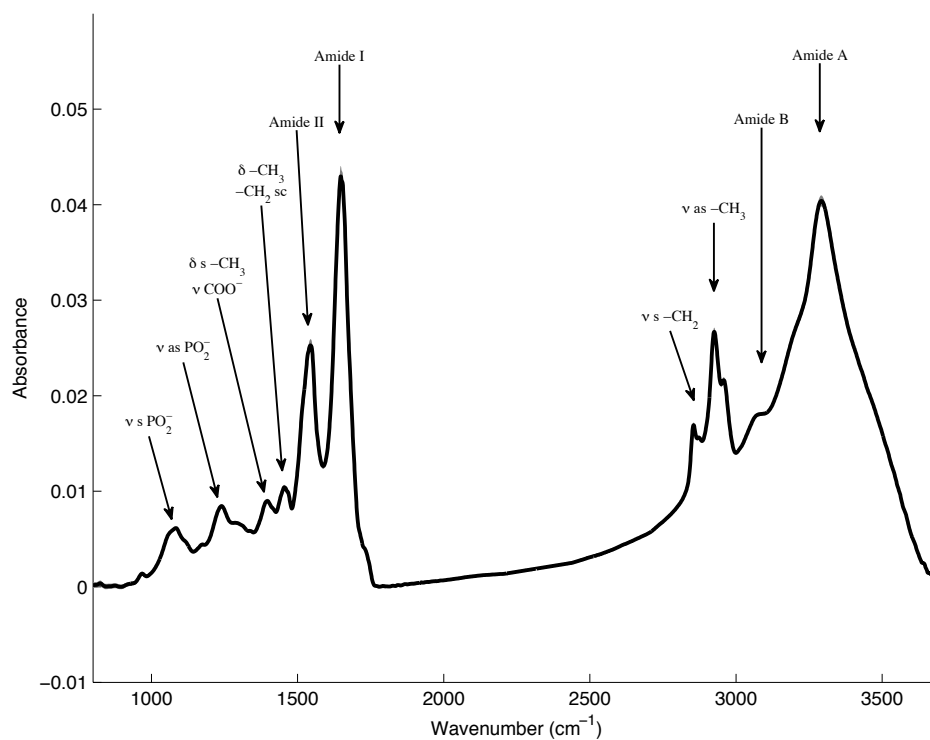


Figure Caption: Figure 1. Reference spectrum of a sham-irradiated control HaCaT cell with labelling of major vibrational features.

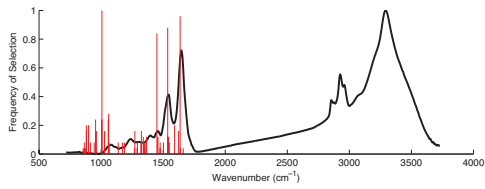
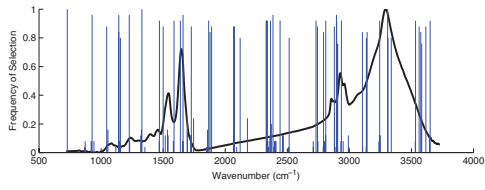


Figure 2(a). Features selected by GA-PLS (top) and GA-QPLS algorithms at 6 hours after irradiation

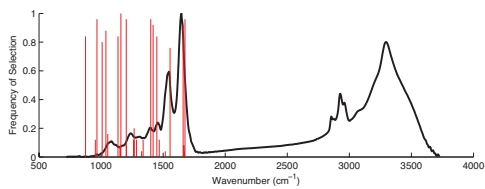
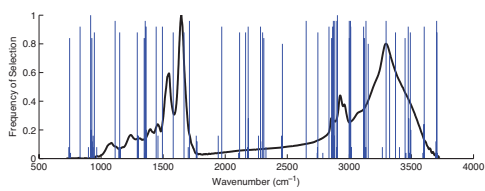


Figure 2(b). Features selected by GA-PLS (top) and GA-QPLS algorithms at 12 hours after irradiation

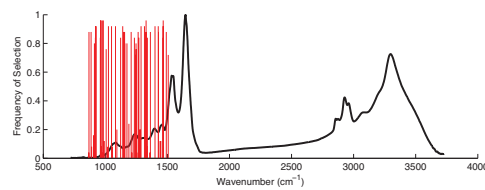
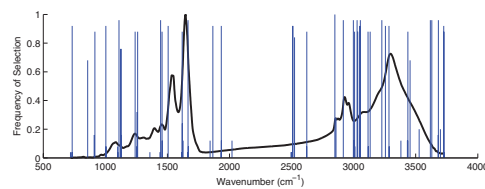


Figure 2(c). Features selected by GA-PLS (top) and GA-QPLS algorithms at 24 hours after irradiation

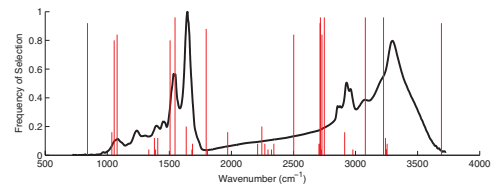
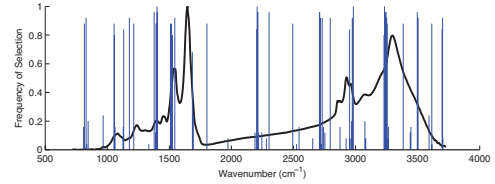


Figure 2(d). Features selected by GA-PLS (top) and GA-QPLS algorithms at 48 hours after irradiation

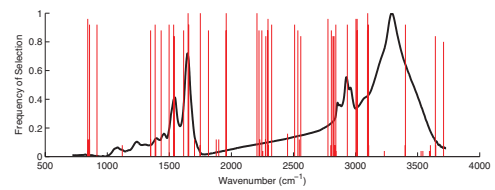
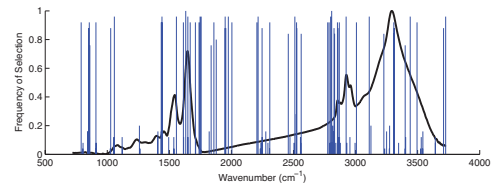


Figure 2(e). Features selected by GA-PLS (top) and GA-QPLS algorithms at 96 hours after irradiation

Figure Caption: Figure 2(a-e). Top, features selected by GAPLS and, bottom, GA-QPLS algorithms in regressing against radiation dose at each time point after irradiation. The frequency with which each feature is selected over several independent runs of the GA is reflected in the length of the bar at each selected feature.