Combining metabolic fingerprinting and footprinting to understand the phenotypic response of HPV16 E6 expressing cervical carcinoma cells exposed to the HIV anti-viral drug lopinavir

Dong-Hyun Kim,[a] Roger M. Jarvis,[a] Yun Xu,[a] Anthony W. Oliver,[b] J. William Allwood,[a] Lynne Hampson,[b] Ian N. Hampson[b] and Royston Goodacre*[ac]

Received 3rd November 2009, Accepted 26th March 2010
First published as an Advance Article on the web 14th April 2010
DOI: 10.1039/b923046g

Recently, it has been reported that the anti-viral drug, lopinavir, which is currently used as a human immunodeficiency virus (HIV) protease inhibitor, could also inhibit E6-mediated proteasomal degradation of mutant p53 in E6-transfected C33A cells. In this study, C33A parent control cells and HPV16 E6-transfected cells were exposed to lopinavir at concentrations ranging from 0 to 30 μM. The phenotypic response was assessed by Fourier transform infrared (FT-IR) spectroscopy directly on cells (the metabolic fingerprint) and on the cell growth medium (the metabolic footprint). Multivariate analysis of the data using both principal components analysis (PCA) and canonical variates analysis (PC-CVA) showed trends in scores plots that were related to the concentration of the drug. Inspection of the PC-CVA loadings vector revealed that the effect was not due to the drug alone and that several IR spectral regions including proteins, nucleotides and carbohydrates contributed to the separation in PC-CVA space. Finally, partial least squares regression (PLSR) could be used to predict the concentration of the drug accurately from the metabolic fingerprints and footprints, indicating a dose related phenotypic response. This study shows that the combination of metabolic fingerprinting and footprinting with appropriate chemometric analysis is a valuable approach for studying cellular responses to anti-viral drugs.

Introduction

Each year in the UK, over 2700 women are diagnosed with cervical cancer which leads to approximately 1000 deaths per annum (UK Cervical Cancer Statistics, Cancer Research UK, www.cancerresearchuk.org). Globally an estimated 493 000 women are affected by cervical cancer producing 273 500 deaths from this disease each year (National Cervical Cancer Coalition, www.nccc-online.org). Indeed in many low resource countries it is the greatest cause of women's cancer-related mortality.

It is known that the human papillomavirus (HPV) is the major cause of cervical cancer.[1] There are >100 different types of HPV associated with a variety of clinical lesions with approximately 20 of these being associated with anogenital tract lesions.[2] Of these HPVs, low-risk types (*e.g.*, HPV6 and 11) have low tumourigenic potential and usually cause benign warts. By contrast, the high-risk types (*e.g.*, HPV16 and 18) are more often found in association with pre-malignant cervical lesions and invasive cancers.[1,3] Accounting for >60% of disease, HPV16 and 18 are the most prevalent high-risk types associated with cervical cancer although to date there are 11 other high risk types known.[4,5]

In developed countries national anti-HPV vaccination programs are currently being implemented as a means of preventing HPV infection and subsequent disease. However, these only cover types 16 and 18 which mean that there is still a significant proportion of high risk HPV related cervical disease that will not be protected by this strategy. Furthermore, many women are already persistently infected with high risk types of HPV which drastically reduces the effectiveness of vaccination. Thus, since cervical cancer can take from 10–20 years to develop, this means these women will be at risk from this disease for many years to come.

Surgery is still the preferred treatment of choice for HPV related pre-cancerous cervical intraepithelial neoplasia (CIN).[4,6] However, since most surgical procedures of this type carry an increased risk of infertility, this is usually reserved for so-called high grade CIN (CIN3). This leaves many women diagnosed with low grade disease (CIN1/2) with the unsatisfactory situation of being referred for increased surveillance. It is clear that a simple, preferably self-administered non-surgical therapy would provide several advantages such as better preservation of obstetric function which would enable this treatment to be offered for low grade disease.[7]

Expression of high risk types of the E6 and E7 viral oncoproteins are largely responsible for the oncogenic properties of HPV.[8] One of the most intensively studied properties of the E6 protein is its ability to compromise the function of the p53 tumour suppressor protein.[9] In association with the cellular

[a]*School of Chemistry, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK. E-mail: roy.goodacre@manchester.ac.uk; Fax: +44 (0)161 3064556; Tel: +44 (0)161 3064480*
[b]*The University of Manchester, Gynaecological Oncology Laboratories, School of Cancer & Enabling Sciences, St Mary's Hospital, Manchester, UK*
[c]*Manchester Centre for Integrative Systems Biology (MCISB), Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK*

E3 ubiquitin ligase E6-associated protein (AP), E6 binds to the p53 protein. E6 mediated activation of E6AP then catalyses the ubiquitination and subsequent proteasomal degradation of p53.[10] Indeed this strategy of inappropriate activation of the proteasome is used by many other viruses to subvert the function of a variety of cellular proteins that would prove detrimental to viral persistence.[1,11,12] This implies that selective inhibition of proteasomal function could prove to be an effective strategy for the treatment of HPV infections.

Metabolic fingerprinting is a high-throughput analytical technique which mostly uses spectroscopic methods for the classification of samples on the basis of their origin or biological relevance.[13] In metabolic fingerprinting generating quantitative data on specific metabolites is not the strategy, as this necessitates time consuming sample clean-up steps and chromatography usually linked to mass spectrometry.[14] Physicochemical analytical techniques such as FT-IR spectroscopy, nuclear magnetic resonance spectroscopy, or direct infusion mass spectrometry can provide sufficient resolution to handle a significant quantity of biological information reflecting biochemical differences between samples.[15] Therefore, these methods are often used for screening in functional genomics experiments,[16] for the identification of major alterations in biochemical pathways as well as for diagnostic usage in industrial or clinical applications.[17]

The investigation of intracellular metabolites is a powerful method for the understanding of complex biological systems; but global analysis of the complete set of intracellular metabolites is not yet achievable with a single technique. Furthermore, this approach is time consuming, and rapid metabolite turnover leads to technical difficulties requiring robust methods for the rapid quenching of metabolism and extraction from cells. An alternative yet complementary approach can be the measurement of extracellular metabolites that are secreted and/or excreted from cells or organisms into their growth media. This overall strategy is called metabolic footprinting, which was derived from a study involving the classification of microbial mutants for functional genomics.[18,19]

FT-IR spectroscopy has been used widely to investigate metabolic changes in biological processes as well as to identify the components of biological systems in the field of metabolomics.[13] These techniques make it possible to identify and characterise microbial cells[20] as well as to determine the biochemical differences between normal and diseased cells; as demonstrated for cervical cancer and leukaemia cells in ref. 21 and 22. In addition, FT-IR spectroscopy is a powerful tool for global, sensitive and highly reproducible biochemical analysis and is very rapid due to the automated high-throughput analysis of biological samples and minimal sample preparation requirements.[23] This reagentless, non-invasive and reproducible technique is resulting in manifold applications across a wide range of biosciences[24,25] and has great potential for rapid, low-cost and accurate cancer screening. Since earlier detection of cancer can provide better prognosis, high throughput screening becomes a significant process in cancer therapy. Potential applications in cervical cancer screening are well reviewed in ref. 26. Not only is FT-IR spectroscopy used as a tool for disease diagnosis, but it can also be employed to follow the response of patients to chemotherapeutic drugs, in addition to identifying drug resistant phenotypes such that the best drug or combinations of drugs can be chosen for each patient.[27,28]

Hampson *et al.* have recently reported that the anti-viral drug lopinavir, which is currently used as a human immunodeficiency virus (HIV) protease inhibitor, could also inhibit E6-mediated proteasomal degradation of p53 and selectively kill E6-dependent cervical carcinoma cells *in vitro*.[29] In order to contribute to an understanding of the mechanism of this drug against HPV on human cervical cell lines, in this study we investigate the level and compositional changes in intra- and extra-cellular metabolites of control and HPV16 E6 expressing cervical carcinoma cells upon exposure to a series of physiological relevant lopinavir drug concentrations.

Materials and methods

Cell line and culture medium

HPV-negative human C33A cervical carcinoma cells were maintained in RPMI 1640 medium (Invitrogen, Paisley, UK) supplemented with 10% fetal bovine serum (FBS) and 2 mM L-glutamine (Sigma-Aldrich Company Ltd, Dorset, UK) (complete medium) at 37 °C, 5% CO_2. C33A cells stably transfected with HPV16 E6 and pcDNA3.1 were derived and cultured as previously described.[30]

Protease inhibitor

Lopinavir was provided as a generous gift from Abbott Laboratories, Park Road, Abbott Park, IL 60064-6187, USA. Lopinavir was dissolved in DMSO (Sigma-Aldrich Company Ltd, UK) at a working stock concentration of 20 mM.

Sample preparation

C33A parent and E6-transfected cells (E6 cells) were seeded in T75 culture flasks and then incubated at 37 °C, 5% CO_2. Before cells reached confluence, the complete medium was removed and 800 μl of trypsin were added in order to detach the adherent cells from the flask. Cells were then incubated for 3 min at 37 °C, 5% CO_2. After this incubation period, cells were resuspended in 10 ml of the complete medium and were counted.

For fingerprint analysis, 1×10^6 cells were seeded into 33 T75 culture flasks to give three biological replicates at each drug concentration and allowed to adhere overnight at 37 °C, 5% CO_2. 0, 3, 6, 9, 12, 15, 18, 21, 24, 27 and 30 μM of lopinavir, 0 = control DMSO, were added to the relevant flasks and cells were incubated for 24 h at 37 °C, 5% CO_2. The growth medium was decanted and cells were washed once with 5 ml of PBS which was warmed to 37 °C in advance. Next 500 μl of ice cold MeOH (−48 °C) were added for quenching metabolism and cells were scraped from the surface of the culture flasks and then transferred to 1.5 ml microcentrifuge tubes. For FT-IR spectroscopy (see below) 20 μl of each sample were pipetted onto a 96 well zinc selenide (ZnSe) plate (Bruker Ltd., Coventry, UK) and dried for 10 min at 60 °C.

For footprint analysis, 1×10^5 cells were seeded into 12 6-well culture dishes to give three biological replicates and allowed to adhere overnight at 37 °C, 5% CO_2. 0, 3, 6, 9, 12, 15, 18, 21, 24, 27 and 30 μM of lopinavir, including control DMSO, were added to the relevant well and cells were incubated for 24 h at 37 °C, 5% CO_2. After this incubation period, 1 ml of the growth medium

was collected and centrifuged at 3000 × *g* for 10 min at 4 °C to remove any cells and cell debris. For FT-IR spectroscopy the supernatants were collected and 40 μl of each were pipetted onto a ZnSe plate and dried for 20 min at 60 °C.

High throughput screening (HTS) FT-IR spectroscopy

FT-IR analysis was carried out on Bruker Equinox 55 infrared spectrometer equipped with a motorised microplate module HTS-XT™ utilising a deuterated triglycine sulfate (DTGS) detector (Bruker Ltd.) as detailed in ref. 23. FT-IR spectra were recorded directly from the dried cell solutions and the growth media in transmission mode. The blank zero position of the plate was used to collect the reference spectrum for the background before each measurement. All spectra were obtained in the 4000–600 cm^{-1} range using 64 co-added scans at 4 cm^{-1} resolution. A total of 396 spectra from fingerprinting and footprinting were collected and each spectrum took 60 s to obtain. These analytical settings were maintained during all measurements.

Data analysis

ASCII data were exported from the OPUS 5.5 FT-IR operation software package and imported into Matlab version 7 software (The Mathworks, Inc., Natick, MA, USA) for further spectral analysis. To reduce non-biological variances arising from CO_2 vibrations, the CO_2 peaks at 2400–2275 cm^{-1} and 680–660 cm^{-1} were removed and filled with a set trend. FT-IR spectra were next scaled *via* extended multiplicative scatter correction (EMSC).[31]

Matlab was programmed using in-house routines and was then used to perform principal components analysis (PCA), principal components-canonical variates analysis (PC-CVA) and partial least squares regression (PLSR) analysis on FT-IR spectral datasets from both the finger- and footprinting of C33A cells. Briefly, PCA is one of the oldest and most widely used multivariate techniques, it is employed to reduce the dimensionality of spectroscopic data whilst maintaining the majority of its variance, and is often utilised as an initial step prior to cluster or discriminant analysis.[32–34] Matlab was used to carry out PCA according to the NIPALS (nonlinear iterative partial least squares) algorithm.[35] Canonical variates analysis (CVA, also known as discriminant function analysis (DFA)) was then used to discriminate the samples on the basis of the retained principal components (PCs) and *a priori* information on the experimental class structure.[36,37] Since the class structure within this experiment was based on the analytical replicates (see below) this does not bias the analysis and allows any natural trend differences between drug level treatments or cell types to be observed. Validation of the PC-CVA model for finger- and foot-print analyses was performed according to the methods described in ref. 16,38 and 39. Briefly, the training set, which consisted of the first two biological replicates including all three analytical replicates, was used to construct a PC-CVA model. After which the test dataset was projected first into the PCA space used in the first model, subsequently the retained PCs were projected into CVA space. For comparison between the training and test datasets, the independently generated PCA and PC-CVA score plots were superimposed, co-clustering of the respective class groups within the test and training set indicated the validity of the PCA and PC-CVA models. Finally, 95% confidence circles were constructed around the PC-CVA group means using the χ^2 confidence intervals using two degrees of freedom[37] in order to investigate the statistical significance and to visualise the close superimposition of test and training dataset clusters.

To investigate the relationships between the spectra and the series of drug concentrations, partial least squares regression (PLSR) analysis was used[40,41] as detailed in ref. 42–44. PLSR analysis is generally performed for predictive linear modelling, and in this research was used to model the relationship between the spectral variations of C33A cervical cancer cells and the variable concentrations of anti-viral drug, lopinavir. Validation was achieved using a *k*-fold cross-validation with the samples of each concentration level being left out once except the lowest and the highest concentration.

Results and discussion

Fingerprint analysis

The C33A parent and E6-transfected cells, grown in both the absence and presence of lopinavir, were analysed by FT-IR spectroscopy. Typical transmission-generated FT-IR spectra of C33A human cervical cancer cells exposed to lopinavir for the two different cell lines are illustrated in Fig. 1A. All components of the cells such as a range of amino acids, lipids and nucleic acids contribute to the FT-IR spectra. The band assignments of FT-IR

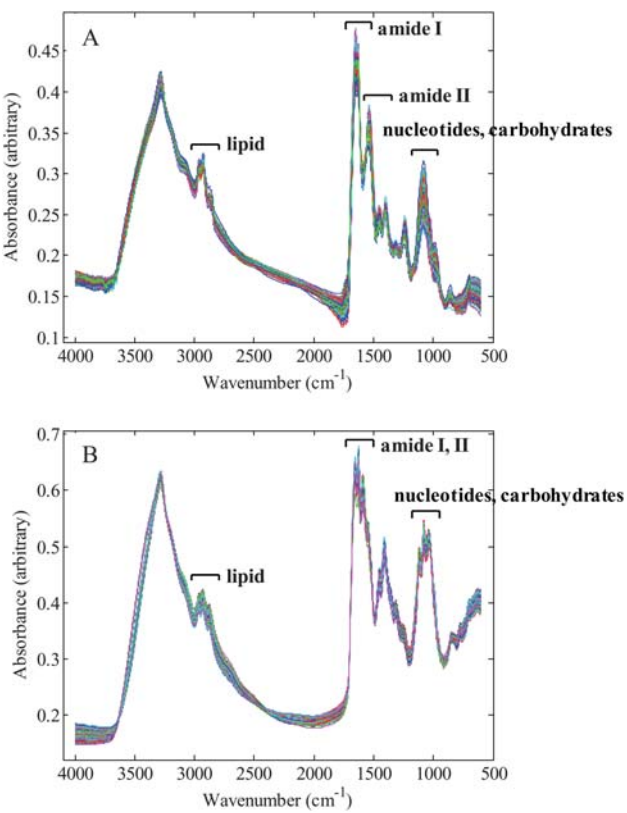

Fig. 1 Typical transmission-generated FT-IR spectra of C33A parent and E6-transfected cells and their growth media that have been grown in both the absence and presence of lopinavir: (A) fingerprint and (B) footprint.

Table 1 Tentative FT-IR band assignment for C33A cells

Wavenumber/cm^{-1}	Assignment	FT-IR vibrational modes
Protein		
~3400 to ~3300	Amide A	Str. mode of N–H
~1690 to ~1620	Amide I	ν(C═O)
~1590 to ~1530	Amide II	δ(N–H) and ν(C–N)
~1340 to ~1240	Amide III	
Nucleotide		
~1225		PO_2^- (asymmetric phosphate)
~1080		PO_2^- (symmetric phosphate)
Lipid		
~2924 and ~2850	Membrane lipids	Symmetric CH_2 str. mode of CH_2 chains
~2958 and ~2873	Membrane lipids	Asymmetric CH_3 str. mode of CH_3 end groups
~1746		C═O
~1463		CH_2 bending
~1165		Ester CO–O–C str.
Carbohydrate		
~1200 to ~900		ν_s(C–O) coupled to the δ(C–O–H)
		ν_s(C–O–C)
		ν_s(C–O)
~1030	Sugar (glycogen)	

spectra were obtained from available literatures[26,45–48] and are shown in Table 1. We attempted to collect the spectrum of lopinavir at the maximum concentration used in this work (30 μM); however, no infrared absorbance was seen and so no FT-IR spectrum was recovered so this confirms that there was no direct contribution of the drug to the spectra of cell solutions (data not shown).

Due to the qualitative similarity and complexity of the spectra from the two cell lines, a visible inspection of differences would be very difficult and so an initial multivariate analysis was carried out with PCA, as described above, to investigate basic biological differences between the two different cell types. The resultant PCA scores plots and loadings plots are shown in Fig. 2. As can be seen in Fig. 2A and B, a clear separation between the two cell types is observed, indicating a high feasibility for an investigation of the inherent differences between the parent and E6 cells in terms of their biochemistry. Since the control and transfected cells are isogenic, and thus have the same genetic background, the only difference between the two cell types is the presence of the E6 oncogene and its protein. Therefore, this result from PCA clearly illustrates that the metabolic fingerprinting approach has great enough resolution to detect clear metabolic changes caused

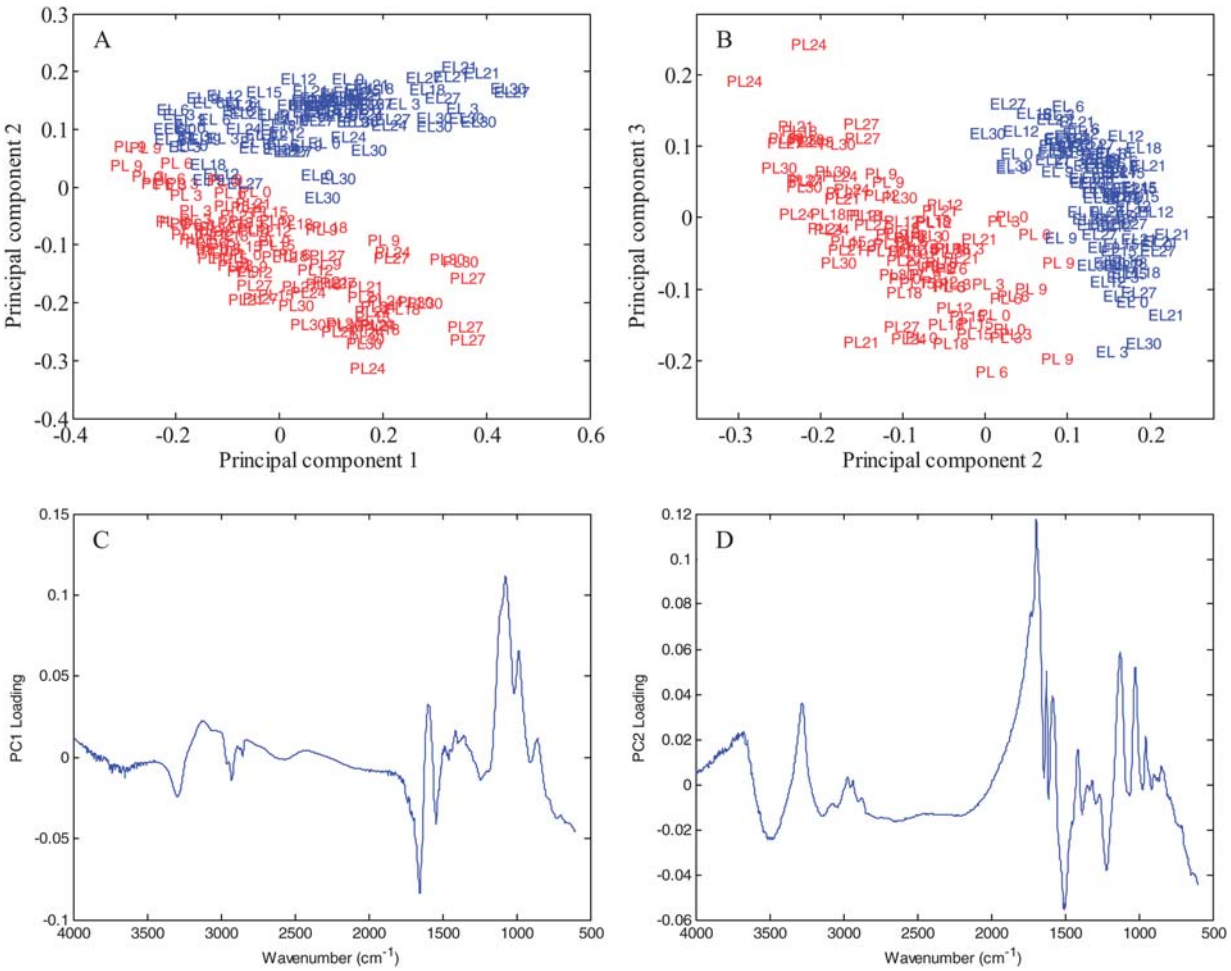

Fig. 2 PCA scores plots (A, B) from PC1 (44.9%), PC2 (31.1%) and PC3 (13.1%), and loadings plots (C: PC1), (D: PC2) extracted from scores plot (A) showing the differences between the spectra of C33A parent and E6-transfected cells; red: parent cells, blue: E6 cells; first two letters indicate the type of cells and the name of the drug, lopinavir, last number represents the drug concentrations (μM).

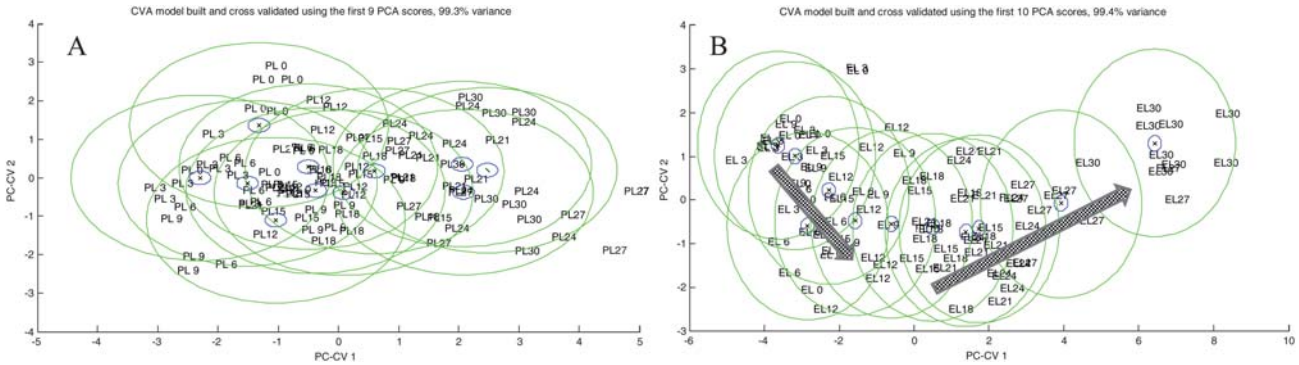

Fig. 3 PC-CVA ordination plots (parent cells, A; E6 cells, B) from E6-transfected cells showing the anti-viral effect on C33A parent and E6-transfected cells exposed to different concentrations of lopinavir. 9 PCs for parent cells and 10 PCs for E6 cells, accounting for 99.3% and 99.4% of the TEV, respectively, were used to construct the ordination plots. Green circles represent the 95% confidence region about the group sample population, and the blue circles represent the 95% confidence interval about the group centroid. Arrows are drawn as a visual guide indicating the relationship between the spectra and the concentrations of lopinavir.

by E6; that is to say the effects of E6 oncoproperties are captured in the metabolome of a cancer cell.

To investigate which spectral regions discriminated between parent and E6 cells, PCA loadings vectors were calculated and plotted (Fig. 2C and D). These loadings plots revealed that the PC1 loadings (Fig. 2C), which account for the greatest variation within the IR dataset and which revealed a drug concentration related trend, predominately revealed a band at $\sim$1080 cm^{-1} as making the greatest contribution, with some contribution from a sharper band at $\sim$970 cm^{-1}. In the PC2 loadings (Fig. 2D), which largely discriminates between C33A parent (red) and E6-transfected (blue) cells, the bands at $\sim$1750, $\sim$1125 and $\sim$1025 cm^{-1} positively contributed to the PCA scores plot clustering. As can be confirmed in Table 1, the bands at $\sim$1080 cm^{-1} and $\sim$970 cm^{-1} can be attributed to phosphate vibrations from nucleotides and the peaks at $\sim$1125 and $\sim$1025 cm^{-1} could arise from a range of vibrations (Table 1) from carbohydrates including sugars within the cells. These large variations in nucleotides and carbohydrates content between the two cells are due to the biological differences which occur by virtue of the expression of the E6 oncogene in the C33A cervical carcinoma cell line. As discussed above, E6 oncogene encoded high risk E6 proteins are approximately 150 amino acid polypeptides, which bind and functionally inactivate the p53 tumour suppressor proteins and also facilitate the degradation of other cellular proteins. Although one of the major roles of E6 is to reduce the effective protein levels of p53 in order to abolish its tumour suppressor function, it is clear that it also directs the degradation of many other cellular proteins such as E6TP1, Scribble and the DNA-repair protein O6-methylguanine DNA methyltransgerase[3,11,49] which are adenosine triphosphate (ATP) dependent processes. Indeed the E6 protein is known to effect the function of many cellular targets[50] with the conclusion that this will cause large biological perturbations in the level of various cellular components which result in the large spectral changes observed between the parent and E6 cells.

In successive multivariate analyses, individual PC-CVA was carried out as described above to investigate the anti-viral drug effect on each cell line separately. The first 9 PCs for parent cells and 10 PCs for E6 cells, which accounted for 99.3% and 99.4% of the total explained variance (TEV), respectively, were used by the CVA algorithm, with knowledge of the biological replicates, cluster analysis validation was applied as detailed above. The resultant PC-CVA ordination plots are shown in Fig. 3A and B for parent and E6 cells, respectively. As can be clearly seen, trends in the individual PC-CVA scores plots reveal that metabolic relationships between drug concentrations and spectra are more significant for the E6 cell line, but less so for the parent. As seen in Fig. 3B, the scores of E6 cells exposed to low concentrations of lopinavir are observed in the top left hand corner, as the concentration of the anti-viral drug increases, the cluster spreads from left to right. It is therefore valuable in the sense that this trend is directly relative to the lopinavir concentrations. This result clearly shows that the metabolic fingerprinting data from C33A human cervical cancer cell lines using FT-IR spectroscopy contain valuable information for studying the phenotypic effect of the anti-viral drug.

Again, to determine whether any specific IR bands accounted for this trend, the PC-CV1 loadings vector from the spectra of E6 cells was calculated and plotted in Fig. 4. As can be seen in this

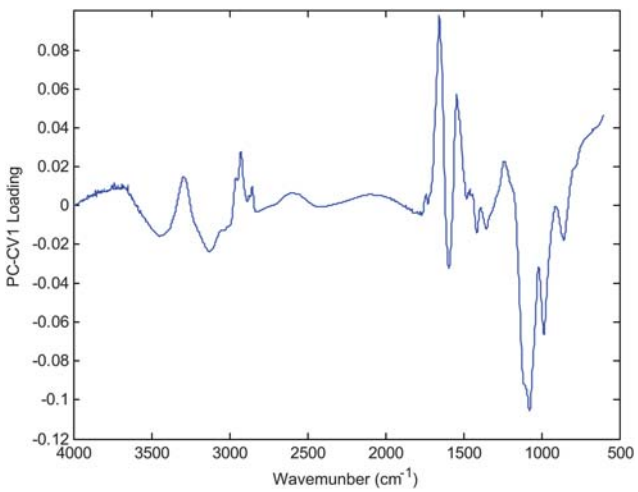

Fig. 4 Loadings plot for PC-CV1 of E6 cells (see Fig. 3B for respective scores plot).

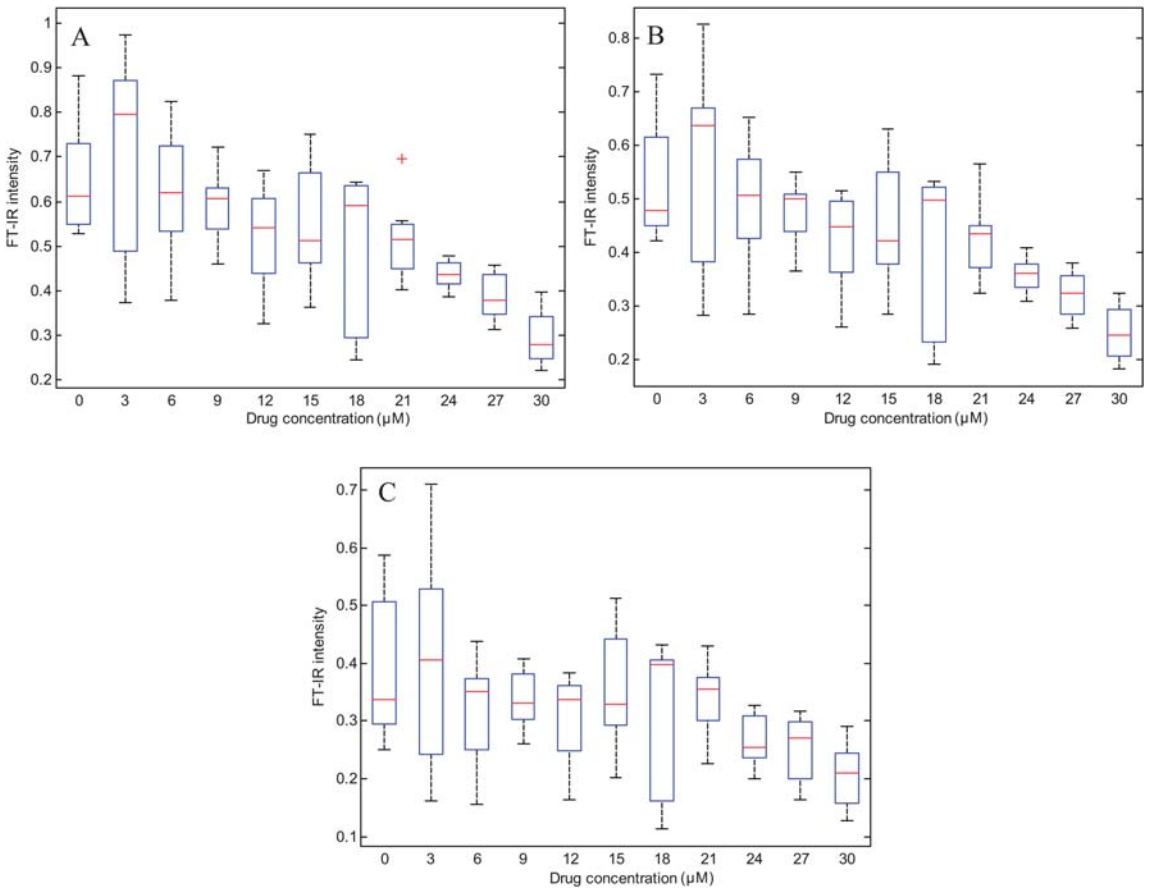

Fig. 5 Box plots of proteins (amide I; A: amide II; B) and, nucleotides and carbohydrates (C) against drug concentrations (μM) for E6-transfected cells. Amide I and II regions are ~1653 and ~1543 cm^{-1}, respectively, nucleotide and carbohydrate regions are ~1170 to 1020 cm^{-1}, respectively.

figure, the band from ~1170 to ~1020 cm^{-1} is predominant within the PC-CV loadings and is negatively correlated to the anti-viral drug concentration. Also, the loadings vectors highlighted other peaks that made a significant contribution at ~1653 and ~1543 cm^{-1}, that are positively correlated to the anti-viral concentration. As discussed above, the band from ~1170 to ~1020 cm^{-1} can be related to nucleotides and carbohydrates, and also peaks at ~1653 and ~1543 cm^{-1} can be attributed to protein (amides I and II) content within the cells.

There could be several explanations for the observed correlation between the level of some intracellular components and the concentration of lopinavir. As can be seen in Fig. 5A and B, which plots the level of amides I and II against drug concentration there is a negative correlation with respect to protein contents. From this observation we can hypothesise that as the drug concentration increases and the proteolytic functions of E6 are inhibited, more p53 and other unidentified E6 target proteins are stabilised in E6 expressing cells. The increase in the level of the p53 protein produced by inhibition of the effects of E6 by lopinavir has already been confirmed by Hampson *et al.*[29] and it is very unlikely that this will be the only E6 targeted protein to be stabilised by this treatment. However, since the comparative cellular level of p53 is very small,[51] it will clearly be below the threshold of detection by the instrument. In addition it is also known that HPV oncoproteins can indirectly promote the stabilisation of specific proteins (*e.g.*, $p16^{INK4a}$) which implies that the net effect of blocking the function of E6 with lopinavir will consist of a balance between virally stabilised and destabilised proteins that are present. This provides a potential, explanation for the observed drop in total cellular proteins in lopinavir treated E6 cells where the overall effect of the drug produces a drop in cellular protein levels. Another intriguing possibility could be that lopinavir is inducing down regulated expression of proteins associated with enhanced toxicity of the drug in the E6 cells. Indeed, it is known that stable ectopic expression of a single oncogene in carcinoma cell lines can cause "oncogene addiction" whereby the cells become dependent on the expression and continued function of the oncogene exhibiting enhanced sensitivity to agents that specifically target its functions.[52]

A potential explanation for the negative correlation between the carbohydrates and nucleotides band (1170–1020 cm^{-1}) against drug concentration shown in Fig. 5C could be differences in carbohydrate metabolism. For example, it has been described that lopinavir can induce insulin resistance *in vitro*[53] which inhibits glucose and 2-deoxyglucose uptake into primary rat adipocytes *in vitro*.[54] Thus the documented influence of lopinavir on carbohydrate metabolism could produce a reduction in the levels of carbohydrate based nucleotides which coincides with the up-regulation of cellular proteins such as p53.

From these PC-CVA results, it is confirmed that the FT-IR spectra from E6 cells contain biological information related to the concentration of lopinavir. Therefore, supervised learning

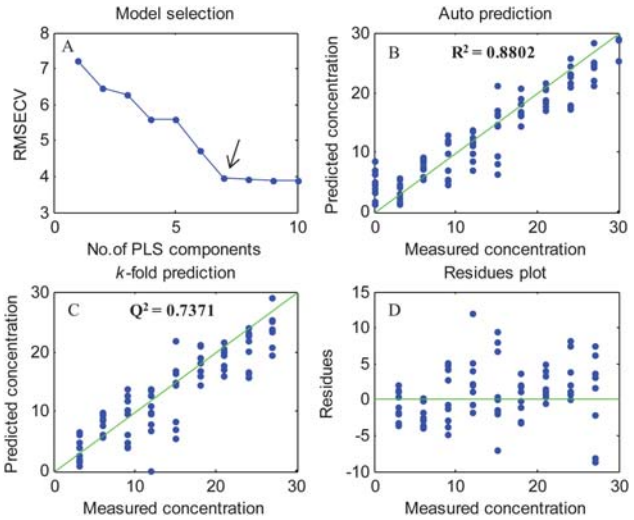

Fig. 6 Estimates from PLS *versus* the actual concentration of lopinavir present in the growth media. (A) Model selection, number of PLS components against root mean square error of cross-validation (RMSECV); seven latent variables were used for the best model with the lowest prediction error in a test set. (B) Auto prediction of the training set, R^2, the percentage of the variations explained by the module based on training sets. (C) Prediction of the test set, Q^2, the percentage of the variations explained by the module based on cross-validation sets. (D) Residues plot, actual concentration minus prediction.

analysis using PLSR was used to attempt to quantify the level of the drug. PLSR was conducted using *k*-fold cross-validation as described in the Data Analysis section. For the best model with the lowest prediction error in a test set (the arrow indicated in Fig. 6A (model selection)), seven latent variables were used. The plot of PLS predicted *versus* actual concentrations of the drug shows relatively good prediction for both the training set ($R^2 = 0.88$) and test set ($Q^2 = 0.74$) as shown in Fig. 6B and C, and the residual errors are randomly distributed (Fig. 6D) suggesting a stable model. Like for the PC-CVA result (Fig. 3), it was also found that the predictions of the anti-viral drug concentrations from spectra using PLS are much better for the E6 cell line than the parent (data not shown). Thus it could be inferred that the spectra of E6 cells have a more consistent response against increasing doses of the anti-viral drug.

Footprint analysis

Growth media from C33A parent and E6 cells grown in the absence and presence of lopinavir were also analysed by FT-IR spectroscopy. Typical footprint spectra for these are shown in Fig. 1B. Clearly all the components of RPMI 1640 medium such as a range of amino acids, vitamins and sugars (formulation of RPMI 1640 medium, American Type Culture Collection. www.atcc.org) contribute to the FT-IR spectra. However, it has been shown for yeast and bacterial cells that the metabolic footprint contains useful information relating to the overflow of metabolism[16,18,19] and so we decided to investigate the metabolic footprint from E6 and parent cells in the presence and absence of lopinavir.

Due to the low concentrations of metabolites secreted into the growth medium relative to the high levels of normal medium components, it is difficult to detect metabolic changes by simple visual inspections of the data. Therefore, PCA was carried out for initial analysis of the spectra produced from the growth medium of both the parent and E6 cells. The resultant PCA scores plot and loadings plot are shown in Fig. 7. Fig. 7A shows that separate clusters from the media of parent and E6 cells are revealed which indicates metabolic differences in terms of the level and compositional changes of extracellular metabolites secreted from these two cell types, irrespective of drug level exposure. It is also notable that the E6 cells are more tightly clustered that the parent cells and this could be due to the addition of the E6 oncogene and its protein causing a common response to the addition of the lopinavir. As can be seen from the PC1 loadings plot in Fig. 7B this separation is related to the variations of protein (amide II) and carbohydrate contents originating from ~1590 and ~1200 to ~1000 cm^{-1}, respectively. These data are the first demonstration of the overall effects of the HPV16 E6 protein on the secreted metabolic signature of cancer cells.

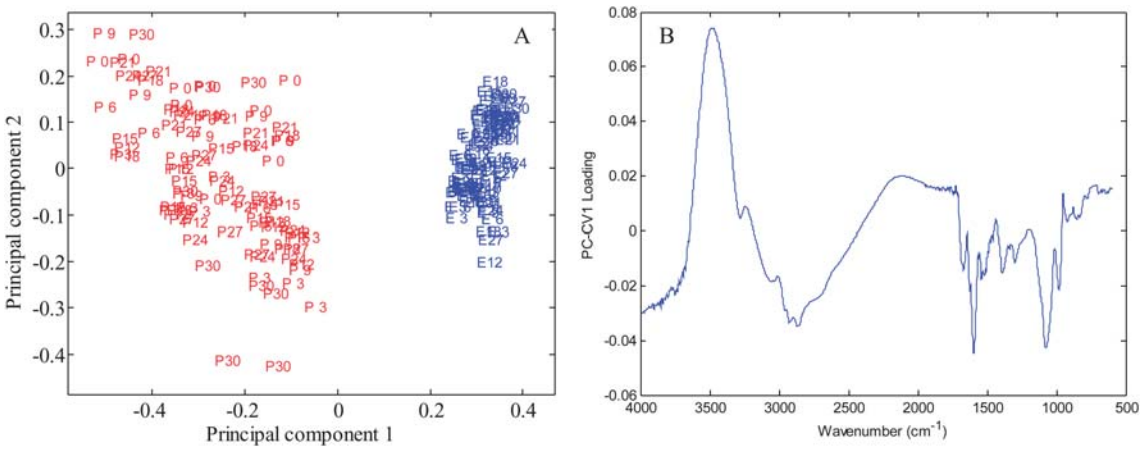

Fig. 7 PCA scores plot (A) from PC1 (75.9%) and PC2 (12.9%), and loadings plot (B) showing the differences between the spectra from footprint media of the growing C33A parent and E6-transfected cells; red: parent cells; blue: E6 cells; first two letters indicate the type of cells and the name of the drug, lopinavir, last number indicate the drug concentrations (μM).

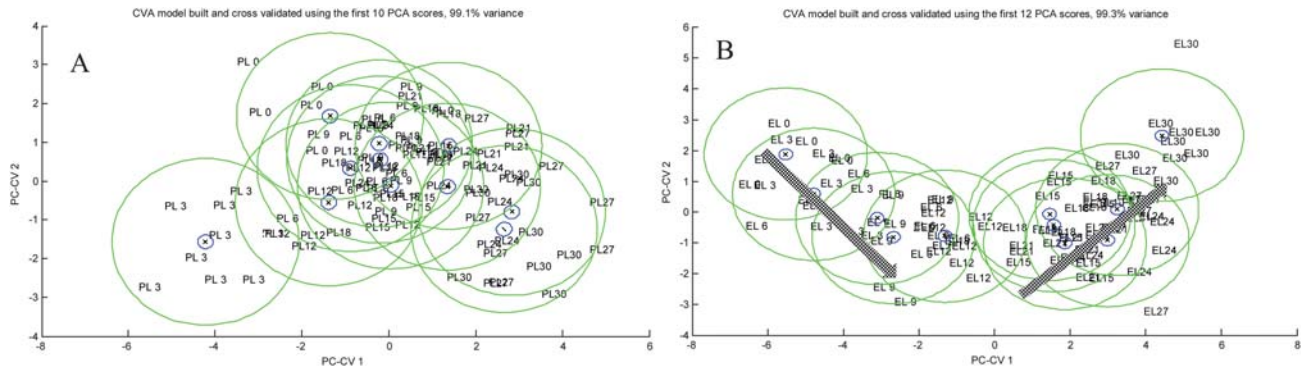

Fig. 8 PC-CVA ordination plots (the medium of parent cells, A; the medium of E6-transfected cells, B) showing the anti-viral effect on extracellular metabolites secreted from C33A parent and E6-transfected cells exposed to different concentrations of lopinavir. 10 PCs for parent cells and 12 PCs for E6 cells, accounting for 99.1% and 99.3% of the TEV, respectively, were used to construct the ordination plots. Green circles represent the 95% confidence region about the group sample population, and the blue circles represent the 95% confidence interval about the group centroid. Arrows are drawn as a visual guide indicating the relationship between the spectra and the concentrations of lopinavir.

To determine the specific anti-viral drug effect on extracellular metabolites secreted from each cell line, individual PC-CVAs were carried out separately and are shown in Fig. 8A and B for parent and E6 cells, respectively. Similar to the fingerprint analysis, trends in the individual PC-CVA score plots show that metabolic relationships between the footprinting spectra and anti-viral drug dose are more obvious for the E6 cell line and less so for the parent cell line. Based upon the previous fingerprint analysis, it is suggested that the changes in intracellular metabolites are in line with those observed for extracellular components in terms of the increasing dose of lopinavir. As can be seen in Fig. 8B, the scores of the medium from E6 cells exposed to low concentrations of lopinavir are located in the top left hand corner and as the dose of the anti-viral drug increases the cluster spreads from left to right. As previously mentioned in the fingerprint analysis, we hypothesise that the trend in the footprint from E6 cells is attributed to the inhibitory activity of anti-viral drug, lopinavir against the functions of E6, rather than the drug itself since the previous analysis of this drug by FT-IR produced no IR absorbance spectrum.

PLSR on the IR spectra from the medium of E6 cells was also conducted using *k*-fold cross-validation as described in the Data Analysis section. Eight latent variables were selected for the best model (Fig. 9A) with the lowest prediction error in a test set. The plot of PLS predicted *versus* actual concentrations of the drug shows very good prediction for both the training set ($R^2 = 0.95$) and test set ($Q^2 = 0.86$) as shown in Fig. 9B and C (again with small residuals; Fig. 9D) indicating that consistent responses occurred with increasing concentrations of drug. Moreover, these predictions from the footprint were slightly better than those from the fingerprint (Fig. 6) and this may result from there being no cells within the footprint samples, thus the sample film thickness was more uniform compared to the metabolic fingerprint which contained cells.

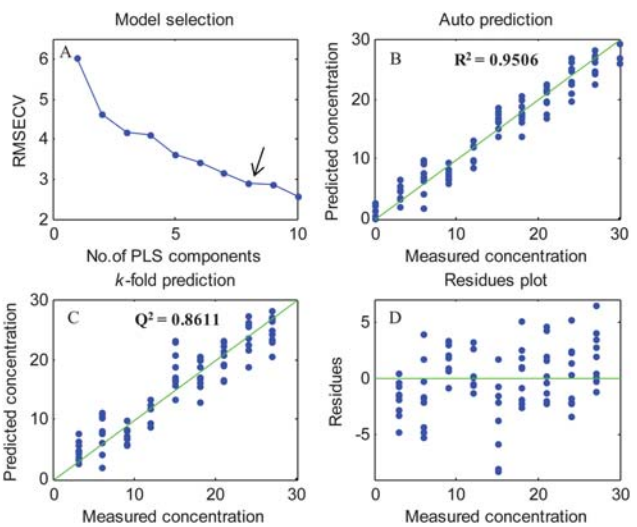

Fig. 9 Estimates from PLS *versus* the actual concentration of lopinavir present in the growth media. (A) Model selection, number of PLS components against root mean square error of cross-validation (RMSECV); eight latent variables were used for the best model with the lowest prediction error in a test set. (B) Auto prediction of the training set, R^2, the percentage of the variations explained by the module based on training sets. (C) Prediction of the test set, Q^2, the percentage of the variations explained by the module based on cross-validation sets. (D) Residues plot, actual concentration minus prediction.

Conclusion

Lopinavir is a well-known highly effective anti-viral drug used routinely for the treatment of HIV. Yet despite there being observed efficacy against HPV the prospective mechanism of action of this drug against HPV infection has not been fully investigated. As discussed it has been reported that the drug has an inhibitory effect on E6-mediated proteasomal degradation of the p53 tumour suppressor protein in C33AE6 cells. Thus, investigation of the changes of intra- and extracellular metabolites caused by lopinavir on C33AE6 and parental control cells could increase understanding of the mode of action of the drug. The objective of this study was to provide valuable data with which to support the prospective new use of lopinavir as a topically applied alternative to surgery for the treatment of HPV related pre-cancerous lesions of the cervix.

Our study shows that the spectral datasets obtained from human C33A cervical cancer cell lines, in combination with

multivariate statistical methods such as PCA, PC-CVA and PLSR, do contain valuable information pertinent to the anti-viral effect of lopinavir. We report that PCA scores plots from both metabolic fingerprint and footprints reveal excellent separation between parent and E6 cells and the PCA loadings vectors show that several regions derived from proteins, nucleotides and carbohydrates contribute to the separation. Trends according to individual PC-CVA scores plots from fingerprinting also reveal that metabolic relationships between drug concentrations and spectra are more significant for the E6 cells and less so for the parent. This indicates that the E6 cells have a bigger drug response. Since the PC-CVA scores plot obtained from the growth media of E6 cells also shows a very similar trend to that from the fingerprint, it is clearly demonstrated that the changes of extracellular metabolites are well correlated to those of intracellular components in terms of response to lopinavir. This is valuable since it means that this approach could be used to analyse cellular response within the footprint in a non-invasive fashion and so the cells do not need to be harvested and dried, thereby allowing time course experiments to be conducted. Finally, PLSR revealed good predictions for the concentration of the drug from the metabolic fingerprinting and footprinting being indicative of a dose-related phenotypic response that was not due to the spectral contribution of the drug itself.

In conclusion we have demonstrated that the combination of metabolic fingerprinting and footprinting with appropriate chemometric analysis is a valuable approach for studying cellular responses to anti-viral drugs.

Acknowledgements

This research was funded by an ORS award to D.-H.K. R.M.J. is very thankful to the UK BBSRC for funding. Y.X. and R.G. are grateful to the EU funded BIOTRACER Integrated Project (www.biotracer.org), and J.W.A. and R.G. are indebted to the EU funded project, Metabolomics for Plants Health and OutReach (META-PHOR: FOOD-CT-2006-036220), for financial support. R.G. is grateful to both the UK BBSRC and EPSRC for financial support of the MCISB. The authors also acknowledge the support of the Humane Research Trust and the Caring Cancer Research Trust.

References

1 H. zur Hausen, *Biochim. Biophys. Acta, Rev. Cancer*, 1996, **1288**, 55–78.
2 E. M. de Villiers, *J. Virol.*, 1989, **63**, 4898.
3 M. Scheffner, B. A. Werness, J. M. Huibregtse, A. J. Levine and P. M. Howley, *Cell*, 1990, **63**, 1129–1136.
4 D. M. Harper, E. L. Franco, C. Wheeler, D. G. Ferris, D. Jenkins, A. Schuind, T. Zahaf, B. Innis, P. Naud, N. S. De Carvalho, C. M. Roteli-Martins, J. Teixeira, M. M. Blatter, A. P. Korn, W. Quint and G. Dubin, *Lancet*, 2004, **346**, 1757–1765.
5 N. Munoz, F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. F. Snijders and C. J. L. M. Meijer, *N. Engl. J. Med.*, 2003, **348**, 518–527.
6 L. L. Villa, R. L. R. Costa, C. A. Petta, R. P. Andrade, K. A. Ault, A. R. Giuliano, C. M. Wheeler, L. A. Koutsky, C. Malm, M. Lehtinen, F. E. Skjeldestad, S. E. Olsson, M. Steinwall, D. R. Brown, R. J. Kurman, R. M. Ronnett, M. H. Stoler, A. Ferenczy, D. M. Harper, G. M. Tamms, J. Yu, L. Lupinacci, R. Railkar, F. J. Taddeo, K. U. Jansen, M. T. Essser, H. L. Sings, A. J. Saah and E. Barr, *Lancet Oncol.*, 2005, **6**, 271–278.
7 M. Kyrgiou, G. Koliopoulos, P. Martin-Hirsch, M. Arbyn, W. Prendiville and E. Paraskevaidis, *Lancet*, 2006, **367**, 489–498.
8 H. zur Hausen, *JNCI, J. Natl. Cancer Inst.*, 2000, **92**, 690–698.
9 J. M. Huibregtse, M. Scheffner and P. M. Howley, *EMBO J.*, 1991, **10**, 4129–4135.
10 J. M. Huibregtse, M. Scheffner and P. M. Howley, *Mol. Cell. Biol.*, 1993, **13**, 775–784.
11 L. Banks, D. Pim and M. Thomas, *Trends Biochem. Sci.*, 2003, **28**, 452–459.
12 M. Thomas, D. Pim and L. Banks, *Oncogene*, 1999, **18**, 7690–7700.
13 D. I. Ellis and R. Goodacre, *Analyst*, 2006, **131**, 875–885.
14 K. Hollywood, D. R. Brison and R. Goodacre, *Proteomics*, 2006, **6**, 4716–4723.
15 O. Fiehn, *Comp. Funct. Genomics*, 2001, **2**, 155–168.
16 N. N. Kaderbhai, D. Broadhurst, D. I. Ellis, R. Goodacre and D. B. Kell, *Comp. Funct. Genomics*, 2003, **4**, 376–391.
17 O. Fiehn, *Plant Mol. Biol.*, 2002, **48**, 155–171.
18 J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver and D. B. Kell, *Nat. Biotechnol.*, 2003, **21**, 692–696.
19 D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic and S. G. Oliver, *Nat. Rev. Microbiol.*, 2005, **3**, 557–565.
20 D. Naumann, D. Helm and H. Labischinski, *Nature*, 1991, **351**, 81–82.
21 M. A. Cohenford and R. Rigas, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 15327–15332.
22 C. P. Schultz, K. Z. Liu, J. B. Johnston and H. H. Mantsch, *Leuk. Res.*, 1996, **20**, 649–655.
23 C. L. Winder, S. V. Gordon, J. Dale, R. G. Hewinson and R. Goodacre, *Microbiology*, 2006, **152**, 2757–2765.
24 G. Déléris and C. Petibois, *Vib. Spectrosc.*, 2003, **32**, 129–136.
25 D. I. Ellis, D. Broadhurst and R. Goodacre, *Anal. Chim. Acta*, 2004, **514**, 193–201.
26 M. J. Walsh, M. J. German, M. Singh, H. M. Pollock, A. Hammiche, M. Kyrgiou, H. F. Stringfellow, E. Paraskevaidis, P. L. Martin-Hirsch and F. L. Martin, *Cancer Lett. (Shannon, Irel.)*, 2007, **246**, 1–11.
27 J. Sulé-Suso, D. Skingsley, G. D. Sockalingum, A. Kohler, G. Kegelaer, M. Manfait and A. J. El Haj, *Vib. Spectrosc.*, 2005, **38**, 179–184.
28 C. M. Krishna, G. Kegelaer, I. Adt, S. Rubin, V. B. Kartha, M. Manfait and G. D. Sockalingum, *Biopolymers*, 2006, **82**, 462–470.
29 L. Hampson, H. C. Kitchener and I. N. Hampson, *Antiviral Ther.*, 2006, **11**, 813–825.
30 L. Hampson, E. El Hady, J. V. Moore, H. Kitchener and I. N. Hampson, *FASEB J.*, 2001, **15**, 1445–1447.
31 H. Martens, J. P. Nielsen and S. B. Engenlsen, *Anal. Chem.*, 2003, **75**, 394–404.
32 I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
33 L. W. Sumner, P. Mendes and R. A. Dixon, *Phytochemistry*, 2003, **62**, 817–836.
34 R. Goodacre, É. M. Timmins, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell and P. J. Rooney, *Microbiology*, 1998, **144**, 1157–1170.
35 H. Wold, *Multivariate Analysis*, Academic Press, New York, 1966.
36 B. F. J. Manly, *Multivariate Statistical Methods: A Primer*, Chapman and Hall, London, 1994.
37 W. J. Krzanowski, *Principles of Multivariate Analysis—A User's Perspective*, Oxford University Press, New York, revised edn, 2000.
38 B. S. Radovic, R. Goodacre and E. Anklam, *J. Anal. Appl. Pyrolysis*, 2001, **60**, 79–87.
39 R. M. Jarvis and R. Goodacre, *Anal. Chem.*, 2004, **76**, 40–47.
40 H. Martens and T. Naes, *Multivariate Calibration*, John Wiley & Sons, Chichester, UK, 1st edn, 1989.
41 S. Wold, A. Ruhe, H. Wold and W. J. Dunn, *SIAM Journal on Scientific Computing*, 1984, **5**, 735–743.
42 R. Goodacre and D. B. Kell, *Metabolic Profiling—Its Role in Biomarker Discovery and Gene Function Analysis*, Kluwer Academic Publishers, London, 2003.
43 D. I. Ellis and R. Goodacre, *Trends Food Sci. Technol.*, 2001, **12**, 414–424.
44 R. Goodacre, *Vib. Spectrosc.*, 2003, **32**, 33–45.
45 H. Fabian, N. A. N. Thi, M. Eiden, P. Lasch, J. Schmitt and D. Naumann, *Biochim. Biophys. Acta, Biomembr.*, 2006, **1758**, 874–882.

46 J. Zhou, Z. Wang, S. Sun, M. Liu and H. Zhang, *Biotechnol. Appl. Biochem.*, 2001, **33**, 127–132.
47 C. Conti, P. Ferraris, E. Giorgini, T. Pieramici, L. Possati, R. Rocchetti, C. Rubini, S. Sabbatini, G. Tosi, M. A. Mariggio and L. L. Muzio, *J. Mol. Struct.*, 2007, **834–836**, 86–94.
48 D. Naumann, *Appl. Spectrosc. Rev.*, 2001, **36**, 239–298.
49 F. Mantovani and L. Banks, *Cancer Biol.*, 1999, **9**, 387–395.
50 H. L. Howie, R. A. Katzenellenbogen and D. A. Galloway, *Virology*, 2009, **384**, 324–334.
51 F. Mantovani and L. Banks, *Oncogene*, 1999, **18**, 3309–3315.
52 C. S. Mitsiades, P. Hayden, V. Kotoula, D. W. McMillin, C. McMullan, J. Negri, J. E. Delmore, V. Poulaki and N. Mitsiades, *J. Clin. Endocrinol. Metab.*, 2007, **92**, 4845–4852.
53 M. A. Noor, O. P. Flint, J. F. Maa and R. A. Parker, *AIDS*, 2006, **20**, 1813–1821.
54 Q. Yan and P. W. Hruz, *J. Acquired Immune Defic. Syndr.*, 2005, **40**, 398–403.

